



9TH INTERNATIONAL CONFERENCE ON KNOWLEDGE, INFORMATION AND CREATIVITY SUPPORT SYSTEMS

6-8 NOVEMBER
LIMASSOL / CYPRUS



PROCEEDINGS

**Proceedings of the 9th International
Conference
on Knowledge, Information
and Creativity Support Systems,
Limassol, Cyprus, November 6-8, 2014**

Editor: George A. Papadopoulos

Editor: George Angelos Papadopoulos

Cyprus Library

**The 9th International Conference on Knowledge, Information and Creativity Support Systems,
University of Cyprus, Nicosia, Cyprus, 2014**

ISBN: 978-9963-700-84-4 (“KICSS’2014 Proceedings”)

**The 9th International Conference on Knowledge, Information and
Creativity Support Systems**

Limassol, Cyprus

November 06-08, 2014

<http://kicss2014.cs.ucy.ac.cy>

A message from the conference chairs

It is our great pleasure to welcome you to the 9th International Conference on Knowledge, Information and Creativity Support Systems (KICSS 2014) in Limassol, Cyprus. Since the 1st KICSS in Ayutthaya in 2006, the International Conference on Knowledge, Information and Creativity Support Systems aims to facilitate technology and knowledge exchange between international researchers and scholars in the field of knowledge science, information science, system science and creativity support systems. After Ayutthaya (2006) KICSS was successfully held in Nomi (2007), Hanoi (2008), Seoul (2009), Chang Mai (2010), Beijing (2011), Melbourne (2012) and Krakow (2013).

Supporting creativity is becoming one of most challenging areas in the 21st century, with research on creativity support systems and relevant fields being emerging. KICSS 2014 covered all aspects of knowledge management, knowledge engineering, intelligent information systems, and creativity in an information technology context, including computational creativity and its cognitive and collaborative aspects.

Putting together KICSS 2014 has been a team effort. First of all, we would like to thank the authors of all submitted papers. Furthermore, we would like to express our gratitude to the program committee and to all external reviewers, who worked very hard on reviewing papers and providing suggestions for their improvements.

The technical program consists of 29 regular papers and 20 short papers. Moreover, we are proud and thankful of having two renowned keynote speakers at the 2014 edition of KICSS:

- **Boris Stilman** on "From Fighting Creative Wars to Making Ordinary Discoveries"
- **Andrzej M.J. Skulimowski** on "The Art of Anticipatory Decision Making"

as well as a Tutorial by **Boris Stilman**.

Finally, we would like to thank our sponsors, Springer, JAIST, JCS, Austrian Airlines, the Cyprus Tourism Organisation and the University of Cyprus for their support of this conference.

We hope that you will find this program interesting and thought-provoking and that it will provide you with a valuable opportunity to share ideas with other researchers and practitioners from institutions around the world.

George A. Papadopoulos
KICSS 2014 Conference Chair
University of Cyprus,
Cyprus

Susumu Kunifuji
KICSS 2014 Conference Chair
JAIST,
Japan

Andrzej M.J. Skulimowski
KICSS 2014 Conference Chair
P&BF and AGH,
Poland

Conference Organizer



The University of Cyprus

In co-operation with



Japan Advanced Institute of Science and Technology



Japan Creativity Society

Financially supported by



Cyprus Tourism Organisation

Patrons



Springer

Sponsors



Save 15% on all applicable fares!

Austrian Airlines

Keynote speakers

Boris Stilman



Dr. Stilman is currently a Professor of Computer Science at the University of Colorado Denver (UC Denver), USA and the Chairman & CEO at STILMAN Advanced Strategies (STILMAN), USA. Boris Stilman received MS in Mathematics from Moscow State University (MGU), USSR in 1972 and two Ph.Ds in Electrical Engineering and Computer Science from National Research Institute for Electrical Engineering (VNIIE), Moscow, USSR in 1984. In 1972-1988, in Moscow, he was involved in the research project PIONEER led by a former World Chess Champion Professor Mikhail Botvinnik. The goal of the project was to discover and formalize an approach utilized by the most advanced chess experts in solving chess problems almost without search. While program PIONEER had never played complete chess games, it solved a number of complex endgames and positions from the games of World Chess Champions. Based on these experiences over a number of years, in Moscow, Dr. Stilman developed experimental and mathematical foundations of the new approach to search problems in Artificial Intelligence. In 1990-91, while at McGill University, Montreal, Canada, based on this approach, he originated Linguistic Geometry (LG), a new theory for solving abstract board games. LG allows us to overcome combinatorial explosion by changing the paradigm from search to construction (from analysis to synthesis). Since 1991, Dr. Stilman was developing the theory and applications of LG at UC Denver. A leap in the development LG was made in 1999, when he (with a group of scientists and engineers) founded STILMAN, LLC. A growing number of applications of LG developed at STILMAN have passed comprehensive testing and are currently being transitioned to the real world command and control systems in the USA and abroad. Since 2010, Dr. Stilman was investigating the structure of the Primary Language of the human brain (following J. von Neumann). According to his hypothesis the Primary Language includes at least two major components critical for humanity, LG and the Algorithm of Discovery. Dr. Stilman published several books (including "Linguistic Geometry: From Search to Construction"), contributions to books, and over 200 research papers. He is a recipient of numerous R&D awards, including the top research awards at University of Colorado, substantial grants from the US government agencies such as major multiple awards from DARPA, US Dept. of Energy, US Army, US Navy, US Air Force, etc.; Ministry of Defence of UK; from the world leading defense companies such as Boeing (USA), Rockwell (USA), BAE Systems (UK), SELEX/Finmeccanica (Italy-UK) and Fujitsu (Japan). More information about Dr. Stilman, history of LG and projects (including several narrated movies) can be found at www.stilman-strategies.com.

Andrzej M.J. Skulimowski

Prof. Skulimowski is a Professor and the Director of the Decision Sciences Laboratory at the Department of Automatic Control and Biomedical Engineering, AGH University of Science and Technology. Since 1995, he has been also the President of the International Progress and Business Foundation, Krakow, where he has led over 80 research, consulting and policy support projects within the EU Framework Programs, ESTO, ETEPS, Interreg, LLP, ERDF, GTD, and other programs. He graduated in Electronics at AGH University, and in Mathematics at the Jagiellonian University, Krakow, Poland. He was awarded a PhD degree in Automatic Control with honors in 1985 and a DSc degree in Operations Research in 1997, both from AGH University. He was a postdoctoral fellow at the Institutes for Automatic Control and for Communication Technology, both at the ETH, Zurich, Switzerland (1987-89), and a visiting professor at the Institute of Information Management at the University of St. Gallen (1990-1995). His main field of expertise is multicriteria decision analysis, cognitive aspects of decision and creativity support systems, foresight, R&D policy, and forecasting. He is the author and editor of 9 books and over 200 scholarly papers on these topics. He invented the anticipatory networks, coined the term mHealth and defined its major challenges, and is one of the pioneers of foresight support systems and advanced approaches to AI foresight. He was the General Chair of the 8th International Conference on Knowledge, Information and Creativity Support Systems (KICSS) held in Krakow, Poland, in November 2013.

Table of Contents

From Fighting Creative Wars to Making Ordinary Discoveries Boris Stilman	1
The Art of Anticipatory Decision Making Andrzej M.J. Skulimowski	14
On Modeling Human-Computer Co-Creativity Oliver Hoffmann	26
Comparison of the Two Companies' Staff by the Group Idea-Marathon Training on Creativity and Motivation Takeo Higuchi	36
Evaluation Indexes to Understand the Creative Problem Solving Process in the Distributed and Cooperative KJ Method Takaya Yuizono and Jun Munemori	48
CARS Context Modelling Christos Mettouris and George A. Papadopoulos	60
Computational Creativity for Intelligence Analysis Robert Forsgren, Peter Hammar and Magnus Jändel	72
Bus Scheduling in Dynamical Urban Transport Networks with the use of Genetic Algorithms and High Performance Computing Technologies V.A. Shmelev, A.V. Dukhanov, K.V. Knyazkov, S.V. Ivanov	86
MCDA and LCSA - A Note on the Aggregation of Preferences João Clímaco, Rogerio Valle	93
A Social Creativity Support Tool Enhanced by Recommendation Algorithms: The Case of Software Architecture Design George A. Sielis, Aimilia Tzanavari, George A. Papadopoulos	100
A Model for Managing Organizational Knowledge in the context of the Shared Services supported by the E-Learning Agostinho Sousa Pinto and Luís Amaral	106
No Space for Emotional Support: Communities of Practice as Arenas of Cognitive Exchange Azi Lev-On, Nili Steinfeld	117
Knowledge Extraction and Annotation Tools to Support Creativity in the Initial Stage of Product Design: Requirements and Assessment Julia Kantorovitch, Ilkka Niskanen, Anastasios Zafeiropoulos, Aggelos Liapis, Jose Miguel Garrido Gonzalez , and Enrico Motta	129
Web Board Question Answering System Based on Problem-Solving Chaveevan Pechsiri, Onuma Moolwat, and Rapepun Piriyaikul	141
Front End System for Personal Health Record with Data Migration Facility from Printed Information Atsuo Yoshitaka , Shinobu Chujiyou, and Hiroshi Kato	153
Effect of Variant Factors on Multi-Document Summarization Nongnuch Ketui and Thanaruk Theeramunkong	165
A Model of Character Affinity for Agent-Based Story Generation Gonzalo Mendez, Pablo Gervas, and Carlos Leon	176

Vicarious: A Flexible Framework for the Creative Use of Sensed Biodata Paul Tennent, Joe Marshall, Brendan Walker, Paul Harter, and Steve Benford.....	184
Rough k Nearest Neighbours for Classification in the Case of Missing Input Data Robert K. Nowicki, Bartosz A. Nowak and Marcin Wozniak.....	196
A Novel Approach to Position Traffic in NoSQL Database Systems by the use of Firefly Algorithm Marcin Wozniak, Marcin Gabryel, Robert K. Nowicki and Bartosz A. Nowak.....	208
An Assessment on the Length of Hospital Stay through Artificial Neural Networks Vasco Abelha, Henrique Vicente, José Machado, and José Neves.....	219
Bio-Inspired Hybrid Intelligent Method for Detecting Android Malware Konstantinos Demertzis, Lazaros Iliadis.....	231
Predicting the Impact of Advertisements on Web Pages Aesthetic Impressions Gianni Fenu, Gianfranco Fadda, and Lucio Davide Spano.....	243
A Process Model of Large-scale Open Idea Generation in Crowdsourcing Marketing Toshihiko Yamakami.....	252
OntoSN - Towards an Integrated Ontology for the Analysis of Online Social Networks Richard Braun and Werner Esswein.....	263
Social Media and Communication Processes at Work: Evidence from China Carol XJ Ou, Louie HM Wong, Robert M Davison and Zhang Cheng.....	277
Educators as Serious Game Designers – A Model-Driven Visual Programming Framework Niroshan Thillainathan, Jan Marco Leimeister.....	289
First Year Students' Algorithmic Skills in Tertiary Computer Science Education Piroska Biró, Mária Csenoch, Kálmán Abari and János Máth.....	301
StalonaLe@rning: A Social Networking Educational System for Teacher Training Pitsikalis Stavros and Lasica Ilona-Elefteyja.....	307
Virtual Environment for Creative and Collaborative Learning Anna Bilyatdinova, Andrey Karsakov, Alexey Bezgodov, Alexey Dukhanov.....	313
A Semantic System to Support Teachers to Create Differentiated Digital Learning Resources Françoise Greffier, Federico Tajariol.....	321
The Significance of 'Ba' for the Successful Formation of Autonomous Personal Knowledge Management Systems Ulrich Schmitt.....	327
The Significance of Memes for the Successful Formation of Autonomous Personal Knowledge Management Systems Ulrich Schmitt.....	339
Computer Creativity in Games - How Much Knowledge is Needed? David C. Moffat and Paul Hanson.....	346
When Engineering and Design Students Collaborate: The Case of Website Development Workshop Meira Levy, Yaron Shlomi, Yuval Etzioni.....	352
Knowledge-Managing Organizational Change Effectors, Affectors and Modulators in an Oil and Gas Sector Environment Anthony Ayoola.....	358

From Computational Creativity Metrics to the Principal Components of Human Creativity Pythagoras Karampiperis, Antonis Koukourikos, George Panagopoulos	364
The Approach to Extension of the CLAVIRE Cloud Platform for the Researchers' Collaboration A.V. Dukhanov, E.V. Bolgova, A.A. Bezgodov, A.V. Boukhanovsky.....	370
iDAF-drum: Supporting Everyday Practice of Drum by Adding an Unperceivable Factor Kazushi Nishimoto, Akari Ikenoue and Masashi Unoki.....	384
Effects of Random Correlation on ANOVA and Regression Gergely Bencsik, Laszlo Bacsardi.....	396
A Proposed Clinical Prediction Rule Register Supporting Primary Care Decision Making Atieh Zarabzadeh, Ronan McDonnell, Derek Corrigan, Rose Galvin, Susan M Smith, Tom Fahey.....	402
Graph Visualization Performed by nVidia CUDA Platform Ondrej Klapka, Antonin Slaby.....	408
Quantitative Evaluation of the Coffee Service Quality and Customer Value by Mining Twitter Shu Takahashi, Ayumu Sugiyama, and Youji Kohda.....	415
Using Wiki as a Collaboration Platform for Software Requirements and Design Irit Hadar, Meira Levy, Yochai Ben-Chaim, Eitan Farchi.....	421
Enhancing Software Architecture via a Knowledge Management and Collaboration Tool Sofia Sherman, Irit Hadar, Meira Levy, Naomi Unkelos-Shpigel.....	427
Prioritization of the Requirements for the Platform to be used for the Employment of Qualified – Disadvantaged Individuals Ahmet Suerdem, Basar Oztaysi.....	433
Visualization of Comparison of Texts from Two Information Sources by Network Integration Ryosuke Saga.....	441
Citizen Science in the Humanities: A Promise for Creativity Milena Dobрева, Daniela Azzopardi.....	446
Academic Achievement in Vocational Secondary Schools: Construction and Validation of a Test to assess potential Learning Performance Metwaly Mabed and Thomas Köhler.....	452

From Fighting Creative Wars to Making Ordinary Discoveries

Boris Stilman^{1,2}

¹ University of Colorado Denver, Denver, CO, USA
Boris.Stilman@UCDenver.edu

² STILMAN Advanced Strategies, Denver, CO, USA
boris@stilman-strategies.com

Abstract. This paper includes results of the research on human creativity in two unlikely related areas of optimal warfighting and making discoveries. According to our hypothesis those algorithms should be the evolutionary components of the Primary Language of the human brain (as introduced by J. von Neumann in 1957). We named those components Linguistic Geometry (LG) and the Algorithm of Discovery. We suggested that both components are mental realities “hard-wired” in the human brain. LG is a formal model of human reasoning about armed conflict, an evolutionary product of millions of years of human warfare. In this paper we focus on discovering the Algorithm of Discovery, the foundation of all the discoveries throughout the history of humanity. This Algorithm is based on multiple thought experiments, which manifest themselves and are controlled by the mental visual streams. This paper reports results of our investigation of the major components of the Algorithm of Discovery with special emphasis on mosaic reasoning as applied to discoveries of the structure of DNA and the theory of Special Relativity.

Keywords: Linguistic Geometry; Primary Language; Artificial Intelligence; Algorithm of Discovery; Game Theory; Mosaic Reasoning

More than 50 years passed since J. von Neumann hypothesized existence of the Primary Language [43]. Unfortunately, the nature of this language is still unknown. Our hypothesis is that the Primary Language is a collection of major algorithms crucial for survival and development of humanity, the underlying “invisible” foundation of all the modern languages and sciences.

1 Creative Wars

What if conventional armed forces were equipped with a tool that changed the way wars were fought forever? What if every military unit had something like an animated X-ray into the future that showed all enemy activities as well as the how best to prevent them? And what if these predictions were updated in real time? Let us make a step further and imagine the future without missiles, nuclear warheads and with limited conventional weapons if at all. What if the outcome of wars is decided without actual fighting like those in a Star Track episode - by computers installed on the hostile planets? It’s not a question of what if, it’s a question of when.

Currently, the U.S. Army is adopting the Linguistic Geometry (LG) software to global intelligence systems to stationary and mobile command posts around the world to command and control systems inside thousands of infantry assault vehicles, and even to soldiers' handhelds. In a couple of years, or even sooner, this 'visionary' software will start saving lives of American and allied soldiers and, maybe, start changing the course of wars around the world.

LG is a type of game theory [12], [14]-[27], [36]-[40] that allows us to solve classes of adversarial games of practical scale and complexity. It is ideally suited for problems that can be represented as abstract board games, for example, military decision aids, intelligent control of unmanned vehicles, simulation-based acquisition, high-level sensor fusion, robotic manufacturing and more. The advantage of LG is that it provides extraordinarily fast and scalable algorithms to find the best strategies for concurrent multi-agent systems. Also, unlike other gaming approaches, the LG algorithms permit modeling a truly intelligent enemy. LG is applicable to the non-zero-sum games and to the games with incomplete information (i.e., imperfect sensors, weather, enemy deception, etc.).

Forty years of development of LG including numerous successful applications to board games and, most importantly, to a highly diverse set of modern military operations from cruise missiles to military operations in urban terrain to ballistic missile defense to naval engagements, led us to believe that LG is something more fundamental than simply yet another mathematical model of efficient wargaming.

After continuous success in proving the applicability and in taking advantage of the power of LG in applications to the modern warfare, over the last 15 years, we decided to investigate if LG would work for the ancient wars. LG was developed originally by generalizing approach utilized by the most advanced experts in playing chess [1]-[4], [27]. Theoretically [40,41], we demonstrated that Alexander the Great and Hannibal, in effect, "used" the LG-like reasoning for their battles, perhaps even consciously, or, most likely, subconsciously. This means that LG or proto-LG "existed" long before the time when chess was invented, which was about fifteen hundred years ago. Chess served just as a means for rediscovering LG. A universal applicability of LG in a variety of military domains, especially, in the domain of the ancient warfare, its total independence of nationality or country, its power in generating human-like strategies permitted to make far reaching conclusions. It is reasonable to suggest that the algorithm of LG utilized by the human brain is "written" or "hard-wired" on the Primary Language and, maybe, even the structure of the language of LG models the structure of the Primary Language. Moreover, the age of the Primary Language must be much greater than the age of human natural languages, and so the age of LG. LG appears to be a part of the human intelligence for probably a million years as the major component essential for survival of the fittest during the constant wars [28]. Moreover, the symbolic part of the human brain, the so-called neo-cortex, did not even exist million years ago. It has been developed later as part of the evolution of the Homo sapiens. This great age suggests that the Primary Language as well the language of LG are not symbolic, i.e., could not utilize strings of symbols, phonetic or written, because symbolic languages have been in existence on Earth no more than a hundred thousand years. The core human skills including the warfighting skills have been developed as hybrid and, mostly, analog algorithms. During several million years of evolution with constant fight for survival of the human species, the human brain perfected those algorithms as well as language(s) utilized by these algorithms. Eventually, this constant perfection has led to the development of multiple symbolic equivalents. We suggest that every human brain

“speaks” the LG language, though, only well trained commanders and, especially, advanced strategists are able to utilize it to full capacity. Most importantly, they are able to translate from the LG language, i.e., from the Primary Language, into the natural languages to describe strategies in the spoken language terms.

2 Towards Ordinary Discoveries

What if discoveries were produced routinely as an output of computer programs? What a leap this would mean for humanity? Approaching this time of the ordinary discoveries is the purpose of our efforts.

In our research on revealing other components of the Primary Language, besides LG, we assumed that they look like LG in some respects. Our contention is that the hypothetical Algorithm of Discovery must be one of such components. In a number of papers, we have been developing a hypothesis that there is a universal Algorithm of Discovery driving all the innovations and, certainly, the advances in all sciences [29]-[35]. All the human discoveries from mastering fire more than a million years ago to understanding the structure of our Solar System to inventing airplane to revealing the structure of DNA to mastering nuclear power in the 20th century utilized this algorithm. The Algorithm of Discovery should be a major ancient item “recorded” in the Primary Language due to its key role in the development of humanity. This line of research involved investigating past discoveries and experiences of construction of various new algorithms, beginning from those, which we were personally involved in [14]-[39], [42]-[44].

Thought experiments allow us, by pure reflection, to draw conclusions about the laws of nature [5]. For example, Galileo before even starting dropping stones from the Tower in Pisa used pure imaginative reasoning to conclude that two bodies of different masses fall at the same speed. The Albert Einstein’s thought experiments that inspired his ideas of the special and general relativity are known even better [13], [6]. The efficiency and the very possibility of thought experiments show that our mind incorporates animated models of the reality, e.g., laws of physics, mathematics, human activities, etc. Scientists managed to decode some of the human mental images by visualizing their traces on the cortex [5]. It was shown that when we imagine a shape “in the mind’s eye”, the activity in the visual areas of the brain sketches the contours of the imagined object; thus, mental images have the analogical nature. It appears that we simulate the laws of nature by physically reflecting the reality in our brain. The human species and even animals would have had difficulty to survive without even minimal “understanding” of the laws of environment. Over the course of evolution and during development of every organism, our nervous system learns to comprehend its environment, i.e., to “literally take it into ourselves” in the form of mental images, which is a small scale reproduction of the laws of nature. Neuropsychologists discovered that “we carry within ourselves a universe of mental objects whose laws imitate those of physics and geometry” [5]. In [28], we suggested that we also carry the laws of the major human relations including the laws of optimal warfighting. The laws of nature and human relations manifest themselves in many different ways. However, the clearest manifestation is in perception and in action. For example, we can say that the sensorimotor system of the human brain “understands kinematics” when it anticipates the trajectories of objects. It is really fascinating that these same “laws continue to be applicable in the absence of any action or perception

when we merely imagine a moving object or a trajectory on a map” [5]. This observation, of course, covers actions of all kinds of objects, natural and artificial. Scientists have shown that the time needed to rotate or explore these mental images follows a linear function of the angle or distance traveled as if we really traveled with a constant speed. They concluded that “mental *trajectory* imitates that of a physical object” [5].

Our main hypothesis is that the Algorithm of Discovery is based not on formal logic but on the so called “visual streams”, i.e., mental imaginary movies which run in our brain. (By the way, LG is highly visual as well.) This is how it may work. Within the brain, the visual streams run consciously and subconsciously and may switch places from time to time (in relation to conscious/subconscious use). We may run several visual streams concurrently, morph them, and even use logic for such morphing, although this use is auxiliary. Then we mentally tag some of the objects shown in the movie and create the so-called symbolic shell around the main visual stream. This shell eventually becomes a standard symbolic algorithm that can be communicated to others employing familiar language, logic, mathematics, etc. I named this approach “visual reasoning”. While the “visual” component (including pattern recognition) is, in general, pretty sophisticated, the reasoning component is relatively simple. Fortunately, the full scale mental visibility is rarely used in discoveries, and, in my opinion, the limited visibility can be simulated with a reasonable effort. The “reasoning” component is certainly within the scope of the modern software development.

Our approach to discovering the Algorithm of Discovery is analogous to an attempt to understand the algorithm of a program while watching its execution. Let us assume that this program’s interface includes color movies on various subjects. In addition to this Algorithm, we are trying to discover the instruction set of the “computer” running this program, i.e., the means of the human brain to running it. With multiple published introspections of great scientists we can recreate clips from various movies, i.e., their imaginary thought experiments. What really helps is the assumption that all those movies were “demonstrated” by the programs running essentially the same algorithm. With our own past developments in LG, we have additional power of asking questions via morphing our own movies and getting answers by watching those morphed movies until the very end. Unfortunately, we do not have this power with the discoveries of other scientists.

3 The Algorithm of Discovery

In this section we briefly summarize the results introduced in [29,30,31,32,33]. The Algorithm of Discovery operates as a series of thought experiments, which interface with the rest of the brain and with external environment via imaginary animated movies (plays), which we named visual streams. These streams may or may not reflect the reality. This interface is constructive, i.e., visual streams could be morphed in the desired direction.

The input to the Algorithm is also a visual stream, which includes several visual instances of the object whose structure has to be understood or whose algorithm of construction has to be developed. Sometimes, the object is dynamic, i.e., its structure is changing in time. Then the input visual stream includes this visual dynamics. As a rule, neither the structure of the object nor the details of the dynamics are present in

the stream. It simply replicates (mimics) the natural or imaginary phenomenon. The task of the Algorithm of Discovery is to understand its structure including dynamics and/or develop an algorithm for reconstructing this object including its changes in time. This understanding happens in several stages. Importantly, it always ends up with the process of actual reconstruction of the object employing the construction set developed by the Algorithm on the previous stages. If the Algorithm investigates a natural real life object this imaginary reconstruction may be totally unrelated to the construction (replication) utilized by the nature. Usually, this reconstruction process is artificially developed by the Algorithm of Discovery with the only purpose to reveal the structure of the object. However, if the algorithm of natural replication is the goal of discovery than the Algorithm of Discovery will employ a set of different visual streams to reveal the relevant components utilized by the nature [35].

All the visual streams are divided into classes, Observation, Construction and Validation. They usually follow each other but may be nested hierarchically, with several levels of depth.

The visual streams operate in a very simple fashion similar to a child construction set. The Construction stream utilizes a construction set and a mental visual prototype, a model to be referenced during construction. This is similar to a list of models pictured in a manual (or a visual guide) enclosed to every commercial construction set. It appears that all the thought experiments in LG related to construction investigated so far, utilized those manuals. Imagine a child playing a construction set. He needs a manual to construct an object by looking constantly at its picture included in this manual. This model comes from the Observation stream as its output. It is not necessarily a real world model. It is not even a model from the problem statement. It is created by the Observation stream out of various multiple instances of the real world objects by abstraction, specifically, by “erasing the particulars”. A final version of the object constructed by the Construction stream should be validated by the Validation stream.

The Algorithm of Discovery initiates the Observation stream, which must carefully examine the object. It has to morph the input visual stream and run it several times to observe (mentally) various instances of the object from several directions. Often, for understanding the object, it has to observe the whole class of objects considered analogous. If the object is dynamic (a process) it has to be observed in action. For this purpose, the Observation stream runs the process under different conditions to observe it in different situations. The purpose of all those observations is erasing the particulars to reveal the general relations behind them. A good example of multiple observations of processes is related to the thought experiments with various objects with respect to the inertial reference frames when discovering the theory of Special Relativity. This includes experiments with uniformly moving ships, trains, experiments with ether as well as experiments for catching a beam of light (Section 5). Once the relations have been revealed, a construction set and a visual model have to be constructed by the Observation stream. Both are still visual, i.e., specific, – not abstract. However, they should visually represent an abstract concept, usually, a class of objects or processes, whose structure is being investigated. For construction, the Observation stream utilizes the Construction stream with auxiliary purpose (which differs from its prime purpose – see below). Note that the model construction is different from the subsequent reconstruction of the object intended to reveal its structure. This model may differ substantially from the real object or class of objects that are investigated. Its purpose is to serve as a manual to be used for references during reconstruction.

When the model and the construction set are ready, the Algorithm of Discovery initiates the Construction stream with its prime purpose. This purpose is to construct the object (or stage the process) by selecting appropriate construction parts of the set and putting them together. If an object has a sequential nature the construction also takes place sequentially, by repetition of similar steps. At some point of construction, the parts are tagged symbolically and, in the end, visual reasoning with symbolic tagging turns into a conventional symbolic algorithm to be verified by the subsequent Validation stream.

Models and construction sets may vary significantly for different problems. Construction of the model begins from creation of the construction set and the relations between its components. Both items should be visually convenient for construction. The Algorithm of Discovery may utilize a different model for the same object if the purpose of development is different. Such a different model is produced by a different visual stream.

In many cases the Algorithm of Discovery employs “a slave” to visually perform simple tasks for all types of visual streams. This slave may be employed by the Construction stream to “see” construction parts and put them together. More precisely, imagine a child playing a simplistic construction set. To avoid offending children, I had named this personality a Ghost. This Ghost has very limited skills, knowledge and, even, limited visibility. The Observation stream may utilize the Ghost to familiarize itself with the optional construction set, to investigate its properties. Next, the Construction stream may use the Ghost to perform the actual construction employing those properties. Eventually, the Validation stream may use the Ghost to verify visually, if properties of the constructed object match those revealed by the Observation stream. In all cases, the Ghost is guided by the Algorithm of Discovery or, more precisely, by the respective visual streams.

As was already discussed, the initial visual model is usually guided by a very specific prototype, where the Observation stream has actually erased the particulars. However, this specificity does not reduce generality in any way. This sounds like a paradox. Essentially, every component of this model carries an abstract class of components behind it. This way visual reasoning about the model drives reasoning about abstract classes, which is turned eventually into the standard formal reasoning. This happens as follows. A visual model drives construction of the formal symbolic model so that the key items in a visual model have tags representing the respective formal model. At first, the formal model is incomplete. At some stage, a running visual stream is accompanied by a comprehensive formal symbolic shell. Running a shell means doing formal derivation, proof, etc. synchronized with a respective visual stream. While the shell and the stream are synchronized, the visual stream drives execution of the shell, not the other way around. For example, a formal proof is driven by animated events within the respective visual stream. The visual streams, usually, run the creation of the visual model, the construction set and the final construction of the object several times. During those runs as a result of persistent tagging the symbolic shell appears. Multiple runs utilize the same visual components but during initial runs the synchronization of the stream and the shell is not tight. Further on, synchronization is tightened by morphing the visual model and/or adjusting symbolic derivation if they initially mismatch. Eventually, the stream and the shell switch their roles. In the end, it appears that the stream becomes the animated set of illustrations, a movie, driven by the running symbolic shell. For example, during the final runs (and only then), the visual streams, presented in [29-34], are driven by the constraints of the abstract board game, the abstract set theory and/or the productions of the

controlled grammars. At this point the visual stream and the symbolic shell can be completely separated, and the visual stream can be dropped and even forgotten.

A stream may schedule other streams by creating almost a “program with procedure calls”. Essentially, it may schedule a sequence of thought experiments to be executed in the future. These experiments will, in their turn, initiate new visual streams. In this case, the purpose, the nature, and the general outcome of those experiments should be known to the stream created this sequence. However, this sequence is different from the list of procedure calls in conventional procedural (or imperative) programming. The algorithms of those “procedures”, i.e., the algorithms to be produced by the respective thought experiments are generally unknown. The experiments are not programmed – they are staged. The actual algorithm should be developed as a result of execution of such experiment. In a sense, this is similar to the notion of declarative programming when a function is invoked by a problem statement while the function’s body does not include an algorithm for solving this problem.

The ability of a visual stream to schedule a sequence of thought experiments permits to create a nested top-down structure of visual streams with several levels of depth. Though, we suspect that the actual depth of the nested programmed experiments never exceeds two or three.

Proximity reasoning as a type of visual reasoning was introduced due to the need for approaching optimum for many discoveries. It is likely that all the technological inventions and discoveries of the laws of nature include “optimal construction” or, at least, have optimization components [13]. Thus, various construction steps performed by the Algorithm of Discovery require optimization, which, certainly, makes construction more difficult. As the appearance of this Algorithm is lost in millennia, for its main purpose, it could not certainly utilize any differential calculus even for the problems where it would be most convenient. For the same reason, it could not utilize any approximations based on the notion of a limit of function. Those great components of differential calculus could certainly serve as auxiliary tools. In that sense, in order to reveal the main optimization components, the most interesting problems to be investigated should lack continuity compelling the Algorithm of Discovery to employ explicitly those components. Based on several case studies [34], we suggested that this optimization is performed by the imaginary movement via approaching a location (or area) in the appropriate imaginary space. Having such space and means, the Algorithm employs an agent to catch sight of this location, pave the way, and approach it. Contrary to the function based approach, which is static by its nature, the Algorithm operates with dynamic processes, the visual streams. Some of those streams approach optimum (in a small number of steps); other streams show dynamically wrong directions that do not lead to the optimum and prevent the Algorithm from pursuing those directions. Both types of streams represent proximity reasoning. We suggested that proximity reasoning plays a special role for the Algorithm of Discovery as the main means for optimization. Proximity reasoning is a type of visual reasoning. This implies that the Algorithm should reason about the space where distances are “analogous” to the 3D Euclidian distances. Roughly, when we approach something, the distance must be visually reduced, and this should happen gradually. The space for proximity reasoning should provide means to evaluate visually if the animated images representing various abstract objects approach each other or specific locations [34]. Construction of those spaces is the key component of the Algorithm of Discovery.

Mosaic reasoning as a type of visual reasoning was introduced due to the analogy of the Construction stream operation with assembling a mosaic picture of small colorful tiles. Another, maybe, even more transparent analogy is known as a jigsaw puzzle when a picture is drawn on a sheet of paper and then this paper is cut into small pieces, mixed up, to be assembled later into the original picture. As Sir G. Thompson [42] pointed "... the progress of science is a little like making a jig-saw puzzle. One makes collections of pieces which certainly fit together, though at first it is not clear where each group should come in the picture as a whole, and if at first one makes a mistake in placing it, this can be corrected later without dismantling the whole group". Both analogies, the pictorial mosaic and the jigsaw puzzle, represent well the key feature of the Algorithm of Discovery construction set. However, we prefer the former because the jigsaw puzzle looks more like an assignment in reassembling a construct, a picture, which has already been created and, certainly, well known. In that sense, a tile mosaic is created from scratch, including choosing or even creating necessary tiles. In addition, a jigsaw puzzle is reassembled out of pieces based on random cuts. On the contrary, in pictorial mosaic, in many cases, every tile should have unique properties; it should be shaped and colored to match its neighbors precisely. A similar specificity is related to a group of adjacent tiles, the aggregate.

In the following sections we will utilize discoveries of the structure of DNA and Special Relativity to demonstrate mosaic reasoning for objects and processes, respectively.

4 Mosaic Reasoning for Discovering Objects

For many discoveries, the components of the construction set should be developed with absolute precision, in the way that every part should be placed to its unique position matching its neighbors. We will use the same name, the tiles, for those construction parts. If precision is violated the final mosaic will be ruined and the discovery will not happen. Though a group of tiles, an aggregate, may be configured properly, its correct placement in the mosaic may be unclear and requires further investigation. Moreover, a tile itself may have complex structure which may require tailoring after placement in the mosaic. In some cases, a tile is a network of rigid nodes with soft, stretchable links.

Mosaic reasoning may stretch through the observation, construction, and validation steps of the Algorithm of Discovery operating with tiles and aggregates of tiles. Overall, mosaic reasoning requires tedious analysis of the proper tiles and their matching rules. Investigation of the matching rules is the essential task of the Observation stream. Multiplicity of those rules and their specificity with respect to the classes of construction tiles make the actual construction very complex. Selecting a wrong tile, wrong tailoring, choosing a wrong place, or incompatible neighbors may ruin the whole mosaic. The matching rules are the necessary constraints that control the right placement of the tiles. Missing one of them, usually, leads to the wrong outcome because the Algorithm of Discovery is pointed in the wrong direction.

Some of the matching rules impact mosaic locally while other rules provide global constraints. The global matching rules include the requirement of the top-down analysis and construction, the global complementarity rule, certain statistical rules, the transformation rules, etc. For many if not all natural objects and processes, their structure is not reducible to a combination of the components. Large groups of tiles,

i.e., large aggregates, may obey the rules which are hardly reducible to the rules guiding placement of singular tiles. This matching rule must be understood globally first, implemented in the mosaic skeleton construction, and, only then, reduced to the placement of the specific tiles. An example of the global matching rule for the discovery of the structure of DNA is the choice of the helical structure of the DNA molecule including the number of strands [35], [44]. The rule of the global complementarity means that placement of one aggregate may determine precisely the adjacent aggregate. In case of DNA, one strand of the helix with the sequence of the base tiles attached to it determines the unique complementary second strand with the corresponding sequence of the base tiles. The global statistical rules related to the whole mosaic may reflect the relationship between major structural components, the large aggregates. If understood and taken into account by the Observation stream, they may focus the Construction stream and lead to a quick discovery. In the case of DNA, the so-called Chargaff rules reflect the structural relationship between the base tiles of the complementary strands of the double helix [35], [44]. Yet another class of global matching rules is called transformation rules. This is an algorithm for reconstructing an aggregate out of another aggregate and placing this aggregate in the proper location. Applied sequentially, such a rule permits to turn an aggregate, the so-called generator, into the set of adjacent aggregates. This way the whole mosaic could be constructed. For example, the whole mosaic of the DNA molecule could be constructed if the generator and the singular transformation are defined. Over the course of four experiments, the double helix generator was constructed. It includes a pair of nucleotides with sugar-phosphate backbone and purine-pyrimidine base. The transformation is a combination of translation and rotation. Interestingly, this type of construction may be utilized by the Algorithm of Discovery as a convenient procedure to reveal the structure of an object, e.g., the DNA molecule, while the nature may use a totally different algorithm for producing the same object.

The local matching rules include the local complementarity rule, the interchangeability rule, etc. The local complementarity means, roughly, that a protrusion on one tile corresponds to the cavity on the complementary adjacent tile. For the DNA molecule this is usually a hydrogen bond of a base tile (a protrusion) that corresponds to a negatively charged atom of the adjacent tile (a cavity). The local complementarity often expresses itself in the requirement of various kinds of symmetry within the pairs of matching construction tiles. The whole class of the local matching rules is based on interchangeability. In simple terms, if two aggregates that include several tiles are not identical but interchangeable, their internal structure may be unimportant. There are several levels of interchangeability. Two aggregates could be essentially the same, i.e., their skeletons coincide. Importantly, those skeletons must include nodes which serve as the attaching points of the aggregates to the rest of the mosaic. The notion of an internal skeleton depends on the problem domain and is specific for different types of mosaic. For example, two different aggregates for the DNA mosaic may have identical ring structures but the atoms and respective bonds that do not belong to those structures may be different. Another lower level of interchangeability of the aggregates does not require their skeletons to coincide. The only requirement is that the attaching points of those aggregates are identical. In all cases interchangeability means that the stream can take one aggregate off the mosaic and replace it with another. This will certainly change the picture but the whole structure will stand. We named those aggregates plug-ins. It appears that plug-ins played crucial role in the discovery of the structure of DNA because such a plug-in was the key component of the helical generator, a purine-pyrimidine base [35], [44].

Besides mosaic structural components that include tiles, aggregates, global and local matching rules, there is an unstructured component that we named a mosaic environment. Such environment may impact the structure of tiles, aggregates, application of matching rules, and the whole mosaic while being relatively unstructured itself. In case of DNA, this was the water content whose lack or abundance could seriously impact the structure of the whole mosaic.

5 Mosaic Reasoning for Discovering Processes

A different type of mosaic, the mosaic of processes, was constructed by Einstein while discovering his theory of Special Relativity [6]. In reality, this was not a construction from scratch – it was a reconstruction of the Galileo-Newton mosaic into new one, the Einstein mosaic. Both mosaics consist of moving tiles, the inertial frames, i.e., those frames moving along straight lines with constant velocities with respect to each other. Contrary to the static mosaic of objects considered above, the inertial frames mosaics represent processes developing in time. Moreover, various entities like human beings or water waves could be moving within those frames. Essentially, these are processes of processes. Mathematically, all those frames should be considered as those in the 4D space with time as the fourth dimension. For the Galileo-Newton mosaic this is a 4D Euclidian space, while for the Einstein mosaic this is a Minkowski space. Note that none of those mathematical constructs were actually used by Einstein for his discovery [6]. Visualization of the 4D spaces is impossible; however, visualizing those mosaics as sets of the 3D processes developing in time supports fully various thought experiments and related visual streams. Typically, one of the frames is chosen as “static” like the one associated to the platform with the Ghost standing on it while the other frame is “moving” and is associated to a train passing by this platform and another Ghost walking inside a car of the moving train. There is no notion of an adjacent tile. However, there is still a notion of the transformation matching rules utilized by the Construction stream for transforming a generator into the whole mosaic of tiles. In those mosaics all the inertial frames are equivalent (or indistinguishable in terms of the laws of Physics), hence, any tile could serve as a generator and a plug-in simultaneously. For the Galileo-Newton mosaic the required transformation is just the Galileo transform while for the Einstein mosaic it is the Lorentz transform.

The Galileo-Newton mosaic has been around for several hundred years and was able to explain numerous experiments. There was no need for any reconstruction. Only during a couple of decades, before the Einstein’s discovery in 1905, several questions were raised. Reconstruction of mosaics was preceded by two series of thought experiments. Some of them were pure thought experiments while many others were replayed in real world, though, initially, they were certainly conceived as the thought ones.

The first series has led to revealing the principle of invariance of the laws of physics, the foundation of both mosaics. These laws should be the same in all the inertial reference frames, i.e., for all the tiles. This series could be traced back to the Galileo experiments in the main cabin of a large swimming ship below its deck [8]. Uniform movement of entities inside the ship (also moving uniformly) is indistinguishable from those on the land. However, assuming that the ship is transparent their velocities would look different from the land due to addition of the

ship's velocity (as a vector). While the Galileo's experiments dealt with mechanical movements the same principle should have covered the laws of electromagnetism by Maxwell-Lorentz. This meant, in particular, that there should not be any distinction in how the induction occurs in both cases, whether the magnet or the conducting coil is in motion. However, according to the classic theory this experiment was interpreted differently for those cases. This meant different laws for different tiles. It was noted by several scientists, including Föppl [7], and emphasized by Einstein [6].

The second series of thought experiments revealed special nature of light or, more precisely, electromagnetic waves. This series could be traced back to the experiments with water and sound waves as traveling disturbances in a medium. As is the case with movements of other entities, velocities of the waves inside the Galileo's ship are indistinguishable from those over the land (for the sound waves) or near the sea shore (for the water waves). Analogously to other moving entities on the transparent ship their velocities would look different from the land due to addition of the ship's velocity. According to the Maxwell's theory, light as well as all types of electromagnetic waves travels at a speed of approximately 186,000 miles per second. This includes AM and FM radio signals, microwaves, visible light, ultraviolet, X-rays, gamma rays, etc. Beginning from Maxwell himself, scientists believed that these waves propagate as disturbances of the invisible medium called the ether, and their velocity was registered relative to this ether. The ether should have had interesting properties. It should spread through the entire universe and should not affect big bodies like planets and stars as well as the smallest ones like specks of dust. In addition it has to be stiff for the light wave to vibrate at a great speed. Numerous thought experiments were intended to demonstrate that the ether waves are going by the Ghost at a faster speed if he is moving through the ether towards the light source. Some of those experiments have actually been implemented in real world. These include experiments by Fizeau, Michelson and Morley as well as those contemplated by Einstein himself. The most influential was the thought experiment of the Ghost riding uniformly at the speed of light alongside a light beam and observing "frozen" light. The 13 real life experiments refuted all their thought prototypes by registering no difference to the speed of light. Multiple executions and morphing (over the period of ten years) of the riding light experiment led Einstein to conclusion that this was not a real effect – the light would not freeze but would run at the same speed according to the same Maxwell equations as for the Ghost standing on the land.

The visual streams utilized in the two series of thought experiments considered above led conclusively to adoption of the two principles [6]. The first was the rigorous spreading of the principle of relativity to all the physical systems that undergo change meaning that the laws governing this change are the same for all the inertial frames of reference. The second principle stated the constancy of the velocity of light c , whether the ray is emitted by a stationary or by a moving body.

Adoption of those new principles led in its turn to the construction of the Einstein mosaic of the processes. The first matching rule was the rule of simultaneity which was the algorithm based on the relation between time and signal velocity. The second matching rule was the length measuring rule which was the algorithm for applying the measuring-rod. Those rules utilized by the Construction stream for constructing processes involving rigid bodies moving at the speed close to the speed of light demonstrated visually (within visual stream) and mathematically the effects of time dilation and length contraction. The major matching rule derived from the above principles was the Lorentz transform. It could be visualized as a hyperbolic rotation.

Applying it to the generator, an arbitrary inertial frame, permitted to construct the whole Einstein mosaic of processes, the universe of inertial frames.

Our research demonstrated that the Algorithm of Discovery does not search for a solution in the search space. Instead, it constructs the solution out of the construction set employing various tools and guides. The right choices of the construction tiles and the matching rules by the Observation stream permit focusing the Construction stream to produce the desired mosaic, i.e., to make a discovery.

The maturity of our recent results on the Algorithm of Discovery is not even close to the maturity of our results on LG. All the results on the Algorithm of Discovery are still hypothetical and have to be verified by software implementations. The very first implementations have been initiated at the University of Colorado Denver.

References

1. Botvinnik, M., Chess, Computers, and Long-Range Planning, Springer, New York, 1970.
2. Botvinnik, M., Blok-skema algoritma igry v shahmaty (in Russian: A Flow-Chart of the Algorithm for Playing Chess), Sov. Radio, 1972.
3. Botvinnik, M., O Kiberneticheskoy Celi Igry (in Russian: On the Cybernetic Goal of Games), Sov. Radio, Moscow, 1975.
4. Botvinnik, M., Computers in Chess: Solving Inexact Search Problems, Springer, 1984.
5. Dehaene, S., A Few Steps Toward a Science of Mental Life, Mind, Brain and Education, Vol. 1, No. 1, pp. 28-47, 2007.
6. Einstein, A., On the Electrodynamics of Moving Bodies, Annalen der Physik, 17:891, 1905 (transl. from German).
7. Föppl, A., Introduction to Maxwell's Theory of Electricity, B. G. Teubner, Leipzig, 1894 (transl. from German).
8. Galilei, G., Dialogue Concerning the Two Chief World Systems, 1632 (tranl. by Stillman Drake).
9. Isaacson, W., Einstein: His Life and Universe, Simon and Shuster, 779 p., 2007.
10. S. Kosslyn, W. Thompson, I. Kim, and N. Alpert, Representations of mental images in primary visual cortex, Nature, 378 , 496-498, 1995.
11. Kott, A., McEneaney, W., (editors), Adversarial Reasoning: Computational Approaches to Reading the Opponent's Mind, Chapman & Hall/CRC, 2007.
12. Linguistic Geometry Tools: LG-PACKAGE, with Demo DVD, 60 pp., STILMAN Advanced Strategies, 2010. This brochure and 8 recorded demonstrations are available online: www.stilman-strategies.com.
13. Miller, Insights of Genius: Imagery and Creativity in Science and Art, Copernicus, an imprint of Springer-Verlag, 1996.
14. Stilman, B., Formation of the Set of Trajectory Bundles, Appendix 1 to [3], 1975.
15. Stilman, B., Ierarhia formalnikh grammatik dla reshenia perebornikh zadach (in Russian: Hierarchy of Formal Grammars for Solving Search Problems), Tech. Report, 105 p., VNIIE, Moscow, 1976.
16. Stilman, B., The Computer Learns, in: 1976 US Computer Chess Championship, by Levy, D., Computer Sci. Press, Woodland Hills, CA, pp. 83-90, 1977.
17. Stilman, B., Fields of Play, Appendix 1 of [4], 1979.
18. Stilman, B., Tsfasman, M., Positional Value and Assignment of Priorities, Appendix 2 of [4], 1979.
19. Stilman, A Formal Language for Hierarchical Systems Control, Int. J. Languages of Design, Vol. 1, No.4, pp. 333-356, 1993.
20. Stilman, B., A Linguistic Approach to Geometric Reasoning, Int. J. of Computers & Math. with Appl, Vol. 26, No. 7, pp. 29-58, 1993,

21. Stilman, B., Network Languages for Complex Systems, *Int. J. of Computers & Math. with Appl.*, Vol. 26, No. 8, pp. 51-80, 1993.
22. Stilman, B., Linguistic Geometry for Control Systems Design, *Int. J. of Computers and Their Applications*, 1(2): 89-110, 1994.
23. Stilman, B., Translations of Network Languages, *Int. J. of Computers & Math. with Appl.*, Vol. 27, No. 2, pp. 65-98, 1994.
24. Stilman, B., Linguistic Geometry Tools Generate Optimal Solutions, *Proc. of the 4th Int. Conference on Conceptual Structures - ICCS'96*, pp. 75-99, Aug. 19-22, 1996, Sydney, Australia.
25. Stilman, B., Managing Search Complexity in Linguistic Geometry, *IEEE Trans. on Syst., Man, and Cybernetics*, 27(6): 978-998, 1997.
26. Stilman, B., Network Languages for Concurrent Multi-agent Systems, *Int. J. of Computers & Math. with Appl.*, 34(1): 103-136, 1997.
27. Stilman, B., *Linguistic Geometry: From Search to Construction*, Kluwer Acad. Publishers (now Springer), 416 p., 2000.
28. Stilman, B., Linguistic Geometry and Evolution of Intelligence, *ISAST Trans. on Computers and Intelligent Systems*, Vol. 3, No. 2, pp. 23-37, 2011.
29. Stilman, B., Thought Experiments in Linguistic Geometry, *Proc. of the 3d Int. Conf on Advanced Cognitive Technologies and Applications – COGNITIVE'2011*, Sep. 25-30, 2011, pp. 77-83, Rome, Italy.
30. Stilman, B., Discovering the Discovery of Linguistic Geometry, *Int. J. of Machine Learning and Cybernetics*, (DOI) 10.1007/s13042-012-0114-8, 20 p., 2012; Printed in 2013, Vol. 4, No. 6, pp. 575-594.
31. Stilman, B., Discovering the Discovery of the No-Search Approach, *Int. J. of Machine Learning and Cybernetics*, (DOI) 10.1007/s13042-012-0127-3, 27 p., 2012.
32. Stilman, B., Discovering the Discovery of the Hierarchy of Formal Languages, *Int. J. of Machine Learning and Cybernetics*, Springer, (DOI) 10.1007/s13042-012-0146-0, 25 p., 2012.
33. Stilman, B., Visual Reasoning for Discoveries, *Int. J. of Machine Learning and Cybernetics*, Springer, (DOI) 10.1007/s13042-013-0189-x, 23 p., 2013.
34. Stilman, B., Proximity Reasoning for Discoveries, *Int. J. of Machine Learning and Cybernetics*, Springer, (DOI): 10.1007/s13042-014-0249-x, 31 p., 2014.
35. Stilman, B., Mosaic Reasoning for Discoveries, *Int. J. of Artificial Intelligence and Soft Computing*, 36 p., 2014.
36. Stilman, B., Yakhnis, V., Umanskiy, O., Winning Strategies for Robotic Wars: Defense Applications of Linguistic Geometry, *Artificial Life and Robotics*, Vol. 4, No. 3, 2000.
37. Stilman, B., Yakhnis, V., Umanskiy, O., Knowledge Acquisition and Strategy Generation with LG Wargaming Tools, *Int. J. of Comp. Intelligence and Applications*, Vol 2, No.4, Dec. 2002, pp. 385-409.
38. Stilman, B., Yakhnis, V., Umanskiy, O., Chapter 3.3: Strategies in Large Scale Problems, in [11], pp. 251-285, 2007.
39. Stilman, B., Yakhnis, V., Umanskiy, O., Linguistic Geometry: The Age of Maturity, *J. of Advanced Computational Intelligence and Intelligent Informatics*, Vol 14, No. 6, pp. 684-699, Sep. 2010.
40. Stilman, B., Yakhnis, V., Umanskiy, O., Revisiting History with Linguistic Geometry, *ISAST Trans. on Computers and Intelligent Systems*, Vol. 2, No. 2, pp. 22-38, Oct. 2010.
41. Stilman, B., Yakhnis, V., Umanskiy, O., The Primary Language of Ancient Battles, *Int. J. of Machine Learning and Cybernetics*, Vol. 2, No. 3, pp. 157-176, 2011.
42. Thomson, G., *The Inspiration of Science*, Oxford U. Press, London, 1961.
43. Von Neumann, J., *The Computer and the Brain*, Yale U. Press, 1958.
44. Watson, J.D., *The Double Helix: A Personal Account of the Discovery of the Structure of DNA*, Atheneum, New York, 1968. [Scribner Classics Edition, New York, 1996.]

The Art of Anticipatory Decision Making

Andrzej M.J. Skulimowski^{1,2}

¹AGH University of Science and Technology
Department of Automatic Control and Biomedical Engineering
Decision Science Laboratory, al. Mickiewicza 30, PL-30-050 Kraków, Poland
and

²International Centre for Decision Sciences and Forecasting
Progress & Business Foundation, ul. Lea 12B, 30-048 Kraków, Poland
ams@agh.edu.pl

Abstract: This paper presents the recent advances of the theory of anticipatory networks and its applications in future-oriented decision-making. Anticipatory networks generalize earlier models of consequence anticipation in multicriteria decision problem solving. This theory is based on the assumption that the decision maker takes into account the anticipated outcomes of future decision problems linked in a prescribed manner by the causal relations with the present problem. Thus arises a multigraph of decision problems linked causally (the first relation) and representing one or more additional anticipation relations. Such multigraphs will be termed anticipatory networks. We will also present the notion of a superanticipatory system, which is an anticipatory system that contains a future model of at least one anticipatory system beside itself. It will be shown that non-trivial anticipatory networks are superanticipatory systems. Finally, we will discuss an application of anticipatory networks to establish efficient collaboration of teams consisting of humans as well as of robots.

Keywords: anticipatory networks; superanticipatory systems; multicriteria decision making; anticipatory collaboration; foresight scenarios; backcasting

1 Introduction

This paper presents the theory of anticipatory networks, which generalizes the ideas related to anticipatory models of consequences in multicriteria optimization problems presented in [11], [12], and [16]. It is assumed that when making a decision, the decision maker takes into account the anticipated outcomes of each future decision problem linked by the causal relations with the present problem. In a network of linked decision problems the causal relations are defined between time-ordered nodes. The future scenarios of the causal consequences of each decision are modelled by multiple edges starting from an appropriate node. The network is supplemented by one or more relations of anticipation, or anticipatory feedback, which describes a situation where decision makers take into account the anticipated results of some future optimization problems while making their choice. They then use the causal dependences of future constraints and preferences on the choice just made to influence future

outcomes in such a way that they fulfill the conditions contained in the definition of the anticipatory feedback relations.

Both types of relations as well as forecasts and scenarios regarding the future model parameters form an information model called an anticipatory network [16]. In Sec. 2 we will show the basic properties of anticipatory networks as well as a method for reducing and computing them.

Following [11] and [16], in Sec. 3 we will present an application of anticipatory networks to select compromise solutions to multicriteria planning problems for anticipatory trees and general networks. We propose a more general notion of preference structure as compared to [15] and [16] that allows us to separate the preferences included in the anticipatory network from those used in present-time decision making. The study of properties of the anticipatory networks led us to introduce the notion of superanticipatory systems in Sec. 4. By definition, a superanticipatory system is a system that is anticipatory in the Rosen sense [10], or weakly anticipatory in the Dubois sense [2], and contains a future model of at least one other anticipatory system whose outcomes may influence its current decisions by a so-called anticipatory feedback relation. This notion is idempotent, i.e. the inclusion of other superanticipatory systems into the model of the future does not yield an extended class of systems, but we can classify them according to a grade that counts the number of nested superanticipations. We will observe that most anticipatory networks can be regarded as superanticipatory systems if we assume that future decisions can be based on similar anticipatory principles as the present decision. Finally, we will discuss an application of anticipatory networks to establish efficient collaboration of teams consisting of humans as well as of robots.

The above theory arose from the need to create an alternative approach to selecting solutions to multicriteria optimization problems taking into account direct multi-stage models of future consequences of the decision made, which was presented in [11]. The anticipatory behavior of decision makers correspond to the definition of an anticipatory system proposed by Rosen [10] and developed further by Dubois [2,3] and other researchers [7],[9]. A bibliographic survey of these ideas can be found in [7]. The ability to create a model of the future (of the outer environment and of itself), which characterizes an anticipatory system, is also the prerequisite for an anticipatory network, where nodes model anticipatory systems that can influence each other according to causal order. In this paper, anticipatory networks are restricted to model decisions made in so-called networks of optimizers, where each node models an optimization problem [16]. This class of information processing systems has been introduced in [14].

It will be pointed out that most anticipatory networks are model-based so that their nodes correspond to weak anticipatory systems [2]. The networks of anticipatory agents can be constructed by applying scenarios of anticipated consequences provided by a foresight project. In the final Sec.5, we will outline an example of such a construction applied in Information Technology (IT) foresight. Similarly to anticipatory networks of optimizers, we can construct networks with nodes modelling Nash equilibria, set choice problems, random or irrational decision makers, or hybrid networks containing nodes of all types [15].

2 Anticipatory networks as generic causal models

The idea behind introducing anticipatory networks as models of consequences was formulated in [11],[14],[16]. The basic principle is to use forecasts and foresight scenarios to estimate the parameters of future decision-making agents and to build a network of them. The anticipated future consequences of a decision made are modelled as changes in constraints and/or preference structures of future decision problems. The nature of these changes is assumed a priori known. It may result from model-based forecasts or foresight as well. Then, the anticipated outcomes of future decision-making problems that – of course – depend on constraints and preference structures, serve as a source of additional information that can be used to solve the current problem. In addition, future decision-making agents may use the same principle to make their decisions and this must be taken into account at the preceding decision stages.

Constructive algorithms for computing the solutions to the current multicriteria decision making problem taking into account the above anticipatory preference information feedback may be applied if we know that:

- All agents whose decisions are modelled in the network are rational, i.e. they make their decisions complying with their preference structures.
- An agent can assess whether the outcomes of some or all future decision problems causally dependent on the present one are more or less desired. This dependence is described as relations (usually multifunctions) between the decisions to be made now and the constraints and/or preference structures of future problems.
- The above assessments are transformed into decision rules for the current solution choice problem, which affect the outcomes of future problems in such a way that they comply with the agent's assessments. The decision rules so derived form an additional preference structure for the decision problem just considered.
- There exists a relevance hierarchy in the network; usually the more distant in the future an agent is, the less relevant the choice of solution. However, this rule is not a paradigm.

Anticipatory networks which contain only decision-making agents solving optimization problems are termed *optimizer networks*. According to [16], an *optimizer* O is a function that acts on a set of feasible decisions U and on the preference structure P and selects a subset $X \subset U$ according to P and to a fixed set of optimization criteria F with values in an ordered space E that define the optimizer. Throughout this paper we will assume that the optimization problems solved by the optimizers have the form

$$(F: U \rightarrow E) \rightarrow \min(\theta), \quad (1)$$

where E is a vector space with a partial order \leq_θ defined by a convex cone θ , i.e. iff

$$x \leq_\theta y \Leftrightarrow y - x \in \theta \text{ for each } x, y \in E.$$

The solution to (1) is the set of nondominated points defined as

$$\Pi(U, F, \theta) := \{u \in U : [\forall v \in U: F(v) \leq_\theta F(u) \Rightarrow v = u]\}.$$

Most frequently, the decision maker's aim is to select and apply just one nondominated solution to (1). Thus the role of the preference structure P that occurs in the definition of an optimizer is to restrict the set of nondominated points in the solution pro-

cess. Without a loss of generality we can assume that P is defined explicitly by pointing out for each $u \in U$ which elements of U dominate u . These are termed *dominating sets* and form a *domination structure* [1] which models the way the decision maker takes into account additional information about preferences when making the decision. Thus P can be defined as a family of subsets of U in the following way

$$P := \{\pi(u) \subset U : u \in \pi(u) \text{ and if } v \in \pi(u) \text{ and } w \in \pi(v) \text{ then } w \in \pi(u)\}_{u \in U},$$

i.e. for each $u \in U$ $\pi(u)$ is the set of elements preferred to u .

As in the case of spaces ordered by convex cones, the element $u \in U$ is *nondominated* with respect to P iff $\pi(u) \cap U = \{u\}$, which means that no element of U is preferred to u . The set of nondominated points with respect to P will be denoted by $\Pi(U, F, P)$. In the most common case, where an additional preference structure P is defined by a convex cone ζ ,

$$\pi(u) := \pi(u, \zeta) = \{v \in U : F(v) \leq_{\zeta} F(u)\} \quad (2)$$

and $\Pi(U, F, P) = \Pi(U, F, \zeta)$. Conversely, in problem (1) $\Pi(U, F, \theta) = \Pi(U, F, P_{\theta})$ with P_{θ} defined by (2). Now we can formulate the following:

Definition 1. A *free multicriteria optimizer* O is a mapping that selects a solution u_0 from U that is nondominated with respect to θ in (1) and P , i.e. if u_0 is an element of

$$\Pi(U, F, \theta, P) := \{u \in U : [\forall v \in U : F(v) \leq_{\theta} F(u) \Rightarrow v = u]\} \cap \Pi(U, F, P). \quad (3)$$

O is characterized by U , F , θ , and P and will be denoted by a 4-tuple $O := (U, F, \theta, P)$. ■

If, beyond the criteria F , the ordering θ , and the preference structure P , an optimizer O realizes certain decision rules R , such as a heuristics or random choice from $\Pi(U, F, \theta, P)$ then the admissible solution set returned by this optimizer may be different from $\Pi(U, F, \theta, P)$ and equal to $X \subset \Pi(U, F, \theta, P)$. In such cases we will use the notation $O := X(U, F, \theta, P)$, where X is interpreted as the set of solutions to the optimization problem (1) actually selected from U . For the sake of brevity, whenever no ambiguity arises, free multicriteria optimizers will be referred to as *optimizers*.

Observe that if for a convex cone $\zeta \subset E$ $P := P_{\zeta}$ and $\theta \subset \zeta$ then, of course,

$$\begin{aligned} \Pi(U, F, \theta, P) &= \\ &= \{u \in U : [\forall v \in U : F(v) \leq_{\theta} F(u) \Rightarrow v = u]\} \cap \{u \in U : [\forall v \in U : F(v) \leq_{\zeta} F(u) \Rightarrow v = u]\} = \Pi(U, F, \zeta). \end{aligned}$$

However, in such cases the resulting preference structure represented by the cone ζ is usually the result of an iterative process of gradually restricting the set of nondominated points to (1). This technique is referred to as the *contracting cone method* (cf. [4]), since the dual cones to an increasing sequence of ordering cones $\theta \subset \zeta_1 \subset \zeta_2 \dots \subset \zeta$ contract as do the sets $\Pi(U, F, \theta)$, $\Pi(U, F, \zeta_1)$, ..., $\Pi(U, F, \zeta)$. Here, we refer to this methodology to show its similarity to the anticipatory network technique described in the Anticipatory Decision-Making Problem (ADMP). Indeed, it can be seen [16] that the more anticipatory feedbacks taken into account in an anticipatory network with the starting node, the more opportunities exist to confine the choice in problem (1) to a smaller subset of the set $\Pi(U, F, \theta, P)$. If for all $u, v \in U$

$$F(v) \leq_{\theta} F(u) \Rightarrow v \in \pi(u)$$

then we will say that P conforms to the criteria F and order θ ; in short P is *conforming*. Observe that this is the case if $P := P_{\zeta}$ and $\theta \subset \zeta$. If P is conforming then to select an $X \subset \Pi(U, F, \theta)$ the action of the optimizer can be stretched on the whole set U , without computing $\Pi(U, F, \theta)$, otherwise it must be restricted to $\Pi(U, F, \theta)$. However, the computation or even an approximation of $\Pi(U, F, \theta)$ can be a hard task.

As already mentioned, besides their optimizing capabilities, optimizers may form networks with several new properties compared to the theory of sequential decision problems. In particular, in feed-forward networks of optimizers constraints and preference structures in some optimizers are causally linked to the results of solving other problems and may depend on their preference structures. Thus, in a network of optimizers, the parameters of the actual instances of optimization problems to be solved vary as the results of solving other problems in the network.

Definition 2. If $O_1 := X_1(U_1, F_1, \theta_1, P_1)$ and $O_2 := X_2(U_2, F_2, \theta_2, P_2)$ are multicriteria optimizers then a constraint influence relation r between is defined as

$$O_1 r O_2 \Leftrightarrow \exists \varphi: X_1 \rightarrow 2^{U_2}: X_2 = \varphi(X_1). \quad (4)$$

Acyclic r are termed causal constraint influence relations –in short, causal relations. ■

Causal relations are represented by a (causal) *network of optimizers*. Def. 2 models the situation where the decision maker anticipating a decision output at a future optimizer can react by creating or forbidding decision alternatives, described by influencing the constraints by multifunctions φ depending on the outputs from the preceding problems. As in [16] and [15], from this point on the term *causal network* will refer to the graph of a causal constraint influence relation. To complete the definition of anticipatory networks, we will define the anticipatory feedback relation.

Definition 3. Suppose that G is a causal network consisting of free optimizers and that an optimizer O_i in G precedes another optimizer, O_j , in the causal order r . Then the *anticipatory feedback* between O_j and O_i in G is information concerning the model-based anticipated output from O_j , which serves as an input influencing the choice of decision at optimizer O_i . Such a relation will be denoted by $f_{j,i}$. ■

By the above definition, the existence of an anticipatory information feedback between the optimizers O_n and O_m means that both condition below apply:

- the decision maker at O_m is able to anticipate the decisions to be made at O_n ,
- the results of this anticipation are to be taken into account when selecting the decision at O_m .

The anticipatory information feedback relation does not need to be transitive. As in the case of causal relations, there may also exist multiple types of anticipatory information feedback in a network, each related to the different way the anticipated future optimization results are considered at optimizer O_m . The multigraph of r and one or more anticipatory feedbacks define an anticipatory network of optimizers:

Definition 4. A causal network of optimizers with the starting node O_0 and at least one additional anticipatory feedback relation linking O_0 with another node in the network will be termed an *anticipatory network* (of optimizers). ■

In [16], the anticipatory information feedback in causal networks of optimizers was applied to selecting a solution to an optimization problem modelled by the starting element in an anticipatory optimizer network G . Specifically, while making the decision, the decision maker takes into account the following information contained in G :

- forecasts concerning the parameters of future decision problems represented by the decision sets U , criteria F , and the ordering structure of the criteria values θ
- the anticipation concerning the behavior of future decision makers acting at optimizers, represented by the preference structures P ,
- the forecasted causal dependence relations r linking the parameters of optimizers in the network,
- the anticipatory relations pointing out which future outcomes are relevant when making decisions at specified nodes and the anticipatory feedback conditions.

We will now present a few key definitions that refer to solving multicriteria decision problems using an anticipatory network of optimizers as a source of additional preference information.

Definition 5. An anticipatory network (of optimizers) is said to be *solvable* if the process of considering all anticipatory information feedbacks results in selecting a non-empty solution set at the starting problem. ■

Definition 6. A causal graph of optimizers G that can be embedded in a straight line will be called a *chain* of optimizers. If it contains at least one anticipatory feedback $f_{i,0}$ then G will be termed an *anticipatory chain* (of optimizers) ■

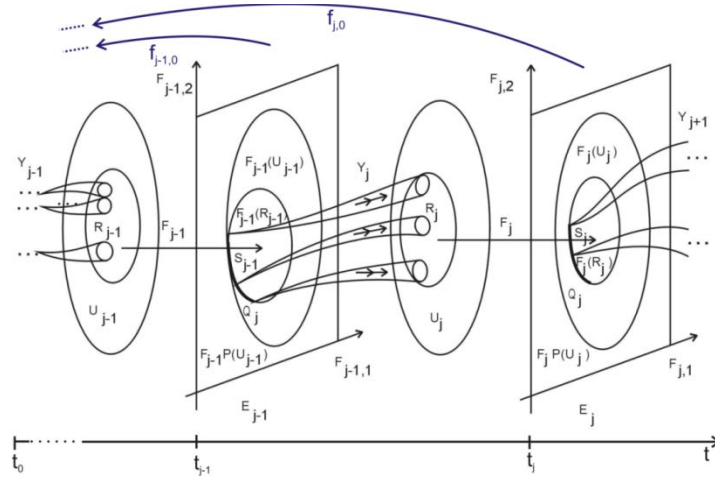


Fig. 1. A chain of optimizers with two anticipatory feedbacks $f_{j-1,0}$ and $f_{j,0}$ linking O_{j-1} and O_j with the starting node O_0 , respectively. The temporal order complies with the causal relations defined by multifunctions $\varphi_{j-1,j} := Y_j \circ F_{j-1}$.

Example 1. Fig.1 presents an example of an anticipatory chain of optimizers, where a decision made at the optimizer $O_0 = (U, F, IR_+^2)$ will take into account the anticipated outcomes at $O_{j-1} = (U, F, IR_+^2)$ and $O_j = (U, F, IR_+^2)$. The causal constraint influence

relations $\varphi_{i,j}$ (cf. Def.2) are defined as $\varphi_{i,j} := Y_j \circ F_i$, where the multifunctions $Y_j: F_i(U_i) \rightarrow U_j$ model the dependence of the scope of decisions available at O_j on the optimization outcomes of the problem O_i . Following [11], the total restriction of the decision scope at O_j generated by Y_j is denoted by R_j , i.e. $R_j := Y_j(F_i(U_i))$. The resulting restriction in the set of nondominated outcomes at O_j is denoted by S_j (in a chain, as exemplified in Fig.1, i can be replaced by $j-1$). By definition, the causal relation represented by $\varphi_{i,j}$ is *non restrictive* iff $S_j = \Pi(U_j, F_j, \theta_j)$. We will say that $\varphi_{i,j}$ *complies with* O_j iff $S_j \subset \Pi(U_j, F_j, \theta_j)$ - this is the case shown in Fig.1.

Definition 7. A causal graph of optimizers G that is a tree and contains at least one anticipatory feedback $f_{i,0}$ will be termed an *anticipatory tree (of optimizers)* ■

A simple tree of optimizers is shown in Fig. 2.

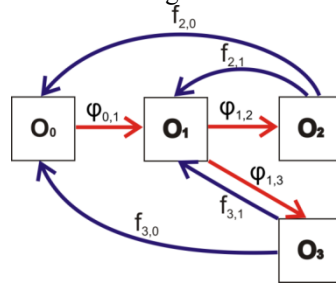


Fig. 2. An example of a simple tree of optimizers, where O_1 is the bifurcation optimizer [16] for O_2 , and O_3 . Causal relations are defined by the multifunctions $\varphi_{i,j} := Y_i \circ F_{i-1}$. Four anticipatory feedback relations are denoted by $f_{k,m}$, $k=0,1$, $m=2,3$.

3 Decision making problems in general anticipatory networks

In a non-trivial anticipatory network the following problem can be formulated:

Anticipatory Decision-Making Problem (ADMP). For an anticipatory network G with finite decision sets, for all chains of optimizers find the set of all admissible sequences of decisions (u_0, \dots, u_n) that maximize the function

$$g(u_0, \dots, u_n) := \sum_{i \in J(0)} h(u_i, q(0, i)) w_{0,i} \quad (5)$$

and such that for all i , $1 \leq i \leq n$, the truncated decision chain (u_i, \dots, u_n) maximizes

$$g(u_i, \dots, u_n) := \sum_{j \in J(i)} h(u_j, q(i, j)) w_{i,j} \quad (6)$$

where $J(i)$, $i=0,1,\dots,n$, denote the sets indices of decision units in G , which are in the anticipatory feedback relations with O_i . The function h is defined as

$$h(u_i, q(i, j)) := \|F_{i-1}(u_i) - q(i, j)\|, \quad (7)$$

and $w_{i,j}$ are positive coefficients corresponding to the relevance of each anticipatory feedback relation between the optimizers O_i and O_j . ■

From the formulation principle in the above decision-making problem it follows that the decision maker at O_0 , while selecting a decision $u_0 \in U$ that is the first element of an admissible decision sequence, uses the anticipatory network G and the function g as an auxiliary preference structure to solve the problem (1). The key notion in this section can now be defined as follows:

Definition 8. A solution to the ADMP, a family of decision sequences $u_{0,m(0)}, \dots, u_{N,m(N)}$ minimizing (4)-(6), will be called *anticipatory chains*. ■

Constructive solution algorithms for solving the ADMP take into account the information contained in an anticipatory network G . These have been proposed in [16] (Algs.1 and 2) for a class of anticipatory networks with discrete decision sets U_i , when the graph of causal relation r is either a chain or a tree. The anticipatory feedback conditions have been defined there as a requirement of O_i that the decisions at O_j , for j from a certain index set $J(i)$ such as O_i precedes O_j in causal order r are selected from the subsets $\{V_{ij}\}_{j \in J(i)}$, $V_{ij} \subset U_j$. Usually, this means that the values of criteria F_j admitted on V_{ij} are of special importance to the decision makers and can be defined as reference sets [13]. The general principles behind these algorithms are as follows:

- Decompose the anticipatory network into causal chains of optimizers linked by causal relations.
- Identify *elementary cycles* in each chain in the anticipatory network, i.e. cycles which do not contain other such cycles except themselves, consisting of causal relations along chains and anticipatory feedback relations.
- Solve the decision problem for each chain, by eliminating the elementary cycles.
- Use the logical conditions that defined the anticipatory requirements to bind the solution sequences to the common parts of the anticipatory chains.

Thus it is possible to reduce the analysis of anticipatory trees to a recursive analysis of anticipatory chains in the tree. Moreover, a general network can be decomposed into trees or chains, which makes it possible to apply solution rules for chains iteratively, gradually eliminating solved trees and chains. However, the solution procedures for anticipatory trees cannot be directly adopted for the solution of the problems where in a network of optimizers there may exist units that are influenced causally by two or more predecessors without taking into account synchronization problems.

Such networks can model problems where multiple resources, provided as outcomes of a number of different and independent decision processes, determine the scope of a later-stage decision. For example, to optimize the decisions in a potential future joint venture created to develop a new product (so-called NPD problem), the outputs provided by the potential future partners of this joint venture should be considered. It can be shown that taking into account the possibility of creating future production alliances and representing such relations in an anticipatory network results in a competitive advantage over agents optimizing their own future outputs only. An example of a general anticipatory network is shown in Fig.3.

To analyze general networked optimizers, it will need to be assumed that if an optimizer O_p is directly influenced by more than one predecessor then the aggregation rules are defined for each subset of influencing factors generated by the preceding optimizer (e.g. as an intersection or a union of the sets of feasible alternatives, each one imposed by a different preceding optimizer).

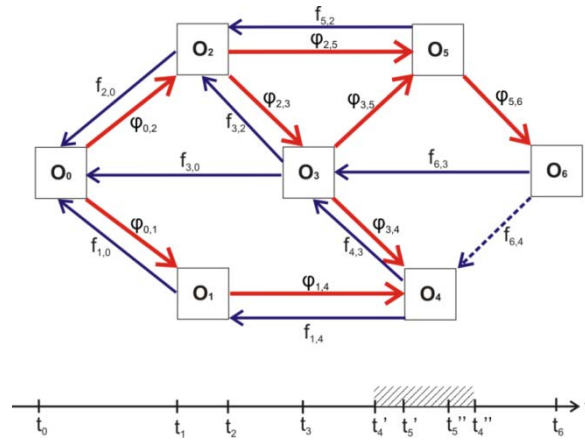


Fig.3. A causal network of seven optimizers, where O_2 and O_3 are bifurcation optimizers, while O_4 and O_5 are each influenced by the two predecessors. The shadowed area between t_4' and t_4'' on the time axis denotes the synchronization interval for the simultaneous influence of O_1 and O_3 on the decisions of O_4 . An analogous interval for O_5 is contained in $[t_4', t_4'']$. The dotted arrow between O_6 and O_4 denotes an irrelevant anticipatory feedback, because there is no causal relation between these optimizers.

In addition, these rules must take into account the synchronization of influence that was not necessary in the case of anticipatory trees. Specifically, the simultaneous action of predecessors on O_p may be restricted to the prescribed time intervals. This is depicted above in Fig.3, where t_i' and t_i'' denote the start and end of a synchronization time interval for the i -th optimizer.

In the most common situation, where the influence of preceding optimizers imposes a logical product of individual influences, the synchronization problem reduces to analysing the time conditions when the intersection of constraints resulting from multiple influencing multifunctions can still yield a feasible solution. However, in general, all combinations of logical conditions binding independent influences should be considered, including the situation where one agent's influence results in removing another agent's constraints. The analysis of such cases requires further studies, which, however, can be based on the solution scheme presented above and in [16].

4 Anticipatory networks as superanticipatory systems

Let us observe that in the above presented approach to solving anticipatory networks, we have assumed that anticipation is a universal principle governing the solution of optimization problems at all stages. In particular, future decision makers modelled at the starting decision node O_0 can in the same way take into account the network of their relative future optimizers when making their decisions. Thus, the future model of the decision maker at O_0 contains models of future agents including their respective future models. This has led us to introduce the notion of superanticipatory systems,

which directly generalize anticipatory systems in the Rosen sense [10] and weak anticipation in the Dubois sense [2]:

Definition 9. A *superanticipatory system* is an anticipatory system that contains at least one model of another future anticipatory system. ■

Since a superanticipatory system is required to contain a model of *another* system, the above definition excludes the case where an anticipatory system models itself recursively. This is discussed later in this section.

By definition, this notion is idempotent, i.e. the inclusion of other superanticipatory systems in the model of the future of a superanticipatory system does not yield an extended class of systems since every superanticipatory system is also anticipatory.

Superanticipatory systems can be classified according to a grade that counts the number of nested superanticipations.

Definition 10. A superanticipatory system S is of *grade n* if it contains a model of a superanticipatory system of grade $n-1$. An anticipatory system which does not contain any model of another anticipatory system is defined as superanticipatory of grade 0.

Let us note that the actual grade n of a superanticipatory system S depends on the accuracy of the model of other systems used by S . In addition, when constructing its model of the environment, S may underestimate the actual content of the other system models. Then, according to Def.10, the grade of superanticipation of S should be regarded as a grade of the model, rather than the actual grade of the physical system.

It may be conjectured that if a superanticipatory system uses an empirical and rational modelling approach then it is more likely that the other systems will have models of a higher grade than S has estimated based on experiments. Thus the grade of the rational system S , when determined based on the information coming solely from the same system, can be regarded as the lower bound of an actual grade. Perfect knowledge of the grade can be attributed to a hypothetical ideal external observer only.

When referring to an anticipatory network, which is always a result of a certain modelling compromise, the following statement can be formulated

Theorem 1. Let $G=(O,r,f)$ be an anticipatory network, where O is the (finite) family of optimizers, r is the causal influence relation, and f is the anticipatory feedback relation. If G contains an anticipatory chain C such that there exist exactly n optimizers in C , $\{O_{C,1}, \dots, O_{C,n}\} \subset C=(O,r,f)$, $N \geq n$, with the following property:

$$\forall i \in \{1, \dots, n\} J_C(i) \neq \emptyset \text{ and } (\exists j \neq i: O_{C,j} r O_{C,i} \text{ and } i \in J_C(j)), \quad (8)$$

where $J_C(i)$ is the set of indices of optimizers in G , which are in the anticipatory feedback relation with O_i and no other chain in G has the property (8) with $m > n$ then G is a superanticipatory system of a grade of at least n .

The **proof** of the above Theorem 1 follows directly from the definitions of anticipatory networks (Def.4) and superanticipatory systems (Defs. 9 and 10).

It is easy to see that an anticipatory network containing a chain on n optimizers, each one linked with O_0 and with all its causal predecessors with an anticipatory feedback is an example of a superanticipatory system of grade n .

The notion of superanticipation is obviously related to the general recursive properties of anticipation. By definition, superanticipation makes sense only when the

anticipation of the future is based on a predictive model. Problems to be solved that arise in a natural way are related to the accuracy of such models and to the grade of superanticipation. They are also related to the relation between *internal* (system) time, when the model is built and analyzed, and *external* real-life time, when the modelled objects evolve. Other recursive approaches related to anticipation include meta-anticipation defined by Dubois [2] and information set models in multi-step games.

A recursive anticipation can be applied in n -stage games, when one player anticipates the behavior of the others, cf. e.g. [6]. From the point of view of player G_1 , anticipation is defined here for $k(G_1)$ steps forward and includes the anticipatory models for the other players G_2, \dots, G_N . Each player can also possess a model of themselves (G_i) and of some or all the remaining participants with an anticipation horizon of $k(G_i)$ moves, $i=2, \dots, N$. Player G_1 thus fulfills the definition of a superanticipatory system and the game can be represented as an evolving anticipatory network. However, when the future moves of the other players result from a deterministic algorithm rather than from a decision-making model, anticipation may be based on the knowledge of the (deterministic) function identified with the operation of that algorithm. This may happen when a human player plays a deterministic game with a computer, or when a machine-machine interaction is modelled. Such games could be modelled by the master-slave (or driver-response) structure of the coupled system analogous to the leader-follower relation in multi stage Stackelberg games [5],[8].

5 Conclusions

This paper examined the principle ideas concerning anticipatory networks as a new tool to model the multicriteria decision problems and the basic methods for solving them. Their extension, so-called superanticipatory systems, was also presented. Anticipatory networks may be applied to model and solve a broad range of problems. Apart from the above-mentioned potential uses in foresight, roadmapping, and socio-econometric modelling, there are further potential fields of application, such as:

- Anticipatory modelling of sustainable development: the underlying assumption of the anticipatory network theory, namely that the present decision maker wants to ensure that future decision makers have the best possible opportunities to make satisfactory decision corresponds to the 'future generation' paradigm of sustainability theory. These 'future generations' are modelled by other network nodes.
- Anticipatory planning based on results of foresight studies, such as development trends, scenarios, and relevance rankings of key technologies, strategic goals, etc. Such planning can use deterministic as well as stochastic planning techniques and include multi-step game models.

Further applications are discussed in [16]. Anticipatory networks can also contribute to implementing the knowledge contained in foresight scenarios in a clear, formal way. Specifically, the development of the theory outlined above has been motivated by the problem of modelling the process of finding feasible foresight scenarios based on the identification of future decision-making processes and on anticipating their outcomes. Scenarios, such as those defined and used in foresight and strategic planning [4], may depend on the choice of a decision in one of the networked optimization

problems and can be external-event driven. When included in a causal network of optimizers, the anticipation of future decisions and alternative external events would allow us to generate alternative structures of optimizers in the network.

Anticipatory networks, those that contain optimizers as well as hybrid ones [15], extend the plethora of modelling tools that can be used to formulate and solve decision making problems taking into account new future-dependent preference structures. When regarded as a class of world models for robotic systems, anticipatory networks provide a flexible representation of the outer environment, while superanticipation allows us to model collective decision phenomena in autonomous robot swarms. Further studies on this class of models may also contribute to the general theory of causality and lead to discovering surprising links to physics and neural decision mechanisms.

References

1. Bergstresser, K., Charnes, A., Yu, P. L.: Generalization of domination structures and nondominated solutions in multicriteria decision making. *J. Optim. Theory Appl.* 18(1), 3-13 (1976)
2. Dubois D.M.: Theory of Incurive Synchronization and Application to the Anticipation of a Delayed Pearl-Verhulst Chaotic System, *Systems Science*, 27(1), 71-86(2001)
3. Dubois D.M.: Mathematical Foundations of Discrete and Functional Systems with Strong and Weak Anticipations. In: Butz, M.V. et al. (eds.) *Anticipatory Behavior in Adaptive Learning Systems, State-of-the-Art Survey*. LNAI 2684, Springer, pp. 110-132 (2003)
4. Godet M.: *Creating Futures – Scenario Planning as a Strategic Management Tool*. Economica, London (2001)
5. Kaliszewski I.: *Quantitative Pareto analysis by cone separation technique*. Kluwer Academic Publishers, Boston (1994)
6. Kijima, K., Xu, Ch.: Incentive Strategies Dealing with Uncertainty about the Followers MCDM Behavior. *Int. J. Systems Sci.* 25(9), 1427-1436 (1994)
7. Nadin M.: Annotated Bibliography. *Anticipation. Int. J. General Sys.* 39(1), 35-133 (2010)
8. Nishizaki I., Sakawa M.: *Cooperative and Noncooperative Multi-Level Programming*. OR/CS Interfaces Series Vol. 48, Springer, Dordrecht, Heidelberg, London, New York (2009)
9. Poli, R.: An introduction to the ontology of anticipation. *Futures* 42, 769-776 (2010)
10. Rosen R.: *Anticipatory Systems - Philosophical, Mathematical and Methodological Foundations*. Pergamon Press, London, (1985), 2nd Ed. Springer, 2012
11. Skulimowski A.M.J.: Solving Vector Optimization Problems via Multilevel Analysis of Foreseen Consequences. *Found. Control Engrg.* 10(1), 25-38 (1985)
12. Skulimowski A.M.J.: Foreseen Utility in Multi-Stage Multicriteria Optimization. In: J. Kacprzyk (ed.), *Seminar on Nonconventional Problems of Optimization*, Warsaw, May 9-11, 1984. *Proceedings, Part III, Prace IBS PAN* 136, pp. 365-386 (1986)
13. Skulimowski A.M.J.: Methods of Multicriteria Decision Support Based on Reference Sets, In: Caballero R., Ruiz F., Steuer R.E. (eds.) *Advances in Multiple Objective and Goal Programming*, LNEMS 455, Springer, Berlin-Heidelberg-New York, pp. 282-290 (1997)
14. Skulimowski A.M.J.: Anticipatory Networks and Superanticipatory Systems. Paper presented at the 10th International Conference CASYS, Liège, 5-9 August 2011
15. Skulimowski A.M.J.: Hybrid anticipatory networks. In: L. Rutkowski et al. (eds.), *Proc. ICAISC 2012*, LNAI 7268, Springer, Berlin-Heidelberg, pp. 706-715 (2012)
16. Skulimowski A.M.J.: Anticipatory Network Models of Multicriteria Decision-Making Processes. *Int. J. Systems Sci.* 45(1), 39-59 (2014), doi:10.1080/00207721.2012.670308

On Modeling Human-Computer Co-Creativity

Oliver Hoffmann

Center for Medical Statistics, Informatics, and Intelligent Systems,
Medical University of Vienna,
oliver@hoffmann.org

Abstract. Do we have a scientific model of creativity as emerging from contributions of computer users, computer systems and their interaction? Such a model would require describing the creative process in general, conditions for human creativity, the added value of human-computer cooperation as well as the role and power of computing. All of these topics have been the subject of research, but they have been addressed in different research communities. Potential obstacles for combining research results from research fields such as knowledge engineering and creativity research and properties of a general model of Human-Computer Co-Creativity are discussed.

Creativity Support Tools (CST) are a relatively recent topic in Human-Computer Interaction (HCI) [20]. But the goal of extending not just the limits of human productivity but also creativity has occupied a prominent place in the earlier history of computing. Computer-supported creativity had to be rediscovered after decades of scientific neglect. Is creativity a particularly hard challenge for computer science? On the one hand, creativity is a difficult topic in itself. On the other hand, we have developed a number of powerful computer technologies with the explicit goal of augmenting the human intellect, but we do not know whether these technologies have increased creativity. More than once, computer science has abandoned the topic of creativity as soon as a path towards extending computing power in itself was identified.

Augmenting Human Creativity

As early as 1945, aiding human creativity with computing machinery was proposed: “...creative thought and essentially repetitive thought are very different things. For the latter there are, and may be, powerful mechanical aids.” [1] Computers were expected to increase human creativity by alleviating the routine elements of scientific work. Did we reach this goal? With data-intensive science, we now have a research paradigm that relies completely on computational support [7]. But has research become more creative as a result? When Chaomei Chen developed CiteSpace and assessed scientific creativity, he still had to “identify principles that appear to be necessary for creative thinking from a diverse range of sources” [2]. Seven decades after Vannemar Bush, scientific creativity and its relationship to computer support are still poorly understood.

In 1960, J.C.R. Licklider extended the vision of supporting creative thought beyond the domain of scientific work as a generalized goal for achieving *man-computer symbiosis*: “The hope is that, in not too many years, human brains and computing machines

ON MODELING HUMAN-COMPUTER CO-CREATIVITY

will be coupled together very tightly, and that the resulting partnership will think as no human brain has ever thought and process data in a way not approached by the information-handling machines we know today.” [8] When Doug Engelbart followed up on the research into *augmenting human intellect*, he already emphasized the central role of *knowledge*: “The capabilities of prime interest are those associated with manipulating symbols and concepts in support of organizing and executing processes from which are ultimately derived human comprehension and problem solutions.” [5].

This line of research was exceptionally successful: With the comparatively recent exception of touch devices, the combination of keyboard, graphical monitor, computer mouse and hypertext have been the standard computer interface over the last three decades. One might argue that our intellect was indeed augmented and our creativity supported by these technologies. If that is the case, we should in fact call them Creativity Support Tools. Why don’t we? Because even though creativity support has been a recurrent research goal, a scientific method for identifying successful Human-Computer Co-Creativity (HC³) was never developed. A recent systematic review of design theories for creativity support tools for instance states:

Much work on creativity support systems provides design guidelines that inform about what set of features the class of information system should encompass in order to be successful [...] literature does not yet provide a wide range of design research efforts that test the effects of particular creativity support system design features.

[21]. In other words: We can evaluate computer system properties, but we do not have a reliable method for identifying a process of creative human-computer co-operation.

Research Fields Fragmentation

Human-Computer Co-Creativity is a process resulting in a creative outcome and involving one or more human individual(s) and one or more computer system(s). A model of HC³ would therefore have to account for the role of each of these three elements as well as the relationships between them (Figure 1). HCI has recently rediscovered Creativity Support Tools. But as a computing discipline, HCI is (again) focusing its efforts on the properties of the computer *tool*, rather than the creative process. Computational Creativity (CC) is investigating the entire creative process, but aims for developing creative computer systems as opposed to supporting human creativity. Human creativity itself is typically investigated independent of any computer interaction in Creativity Research (CR).

HC³ research is fractured. While there are several research communities working on some of the relevant topics, there is no research community dedicated to understanding all of them in conjunction. The obvious task would then be combining research results from these communities. But trans-disciplinary research is facing fundamental differences in research culture. CR has grown out of psychology and its methods. HCI and CC are sub-disciplines of computer science. As engineering disciplines, the latter two are primarily concerned with working towards clearly defined outcomes, while the

ON MODELING HUMAN-COMPUTER CO-CREATIVITY

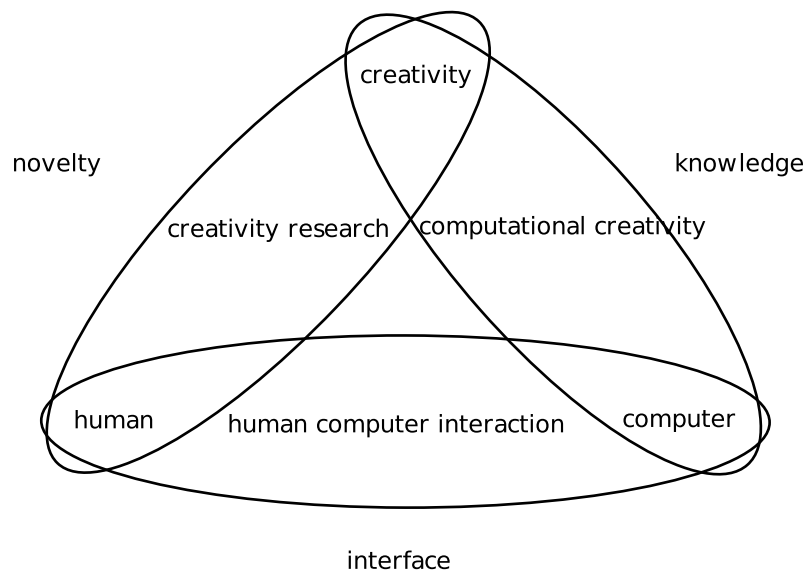


Fig. 1. research topics for Human-Computer Co-Creativity

former places more emphasis on understanding a pre-existing phenomenon. When research is synthesized across these disciplines, key notions and their meaning have to be accommodated.

Knowledge in Computer Science

CC has its roots in Artificial Intelligence (AI). As in all AI-related disciplines, *knowledge* plays a key role in CC. In every day language, knowledge is understood as something an individual has or does not have. Whether or how this knowledge is or can be represented or transmitted through the use of symbolic representations is secondary. This is also the way *knowledge* is viewed in CC.

Computers do not develop and learn like human individuals, they are constructed by computer engineers. To give computers knowledge, AI researchers have invented various types of *knowledge representation*, leading to extended discussions on what constitutes the best type of knowledge representation.

Alan Newell has tried to settle this controversy within AI and his proposal is very illustrative of the content and method of AI research. He formulated a hypothesis claiming knowledge would reside in a virtual computer system level called the *knowledge level*. According to Newell [10] the levels in (Figure 2) all share the following properties: Higher levels do not have to know about the specific implementation of lower levels. Each level can be reduced to the level below. And each level has its unique medium.

ON MODELING HUMAN-COMPUTER CO-CREATIVITY

knowledge level: knowledge
symbol level: software
register transfer sublevel: integrated circuits
logic circuit level: Boolean functions
circuit level: electric circuits
device level: individual transistors

Fig. 2. computer levels according to Newell

According to Newell's hypothesis, knowledge is a medium of computation just like electrons or software symbols and it consists of goals, actions and bodies, governed by the principle of rationality. By delegating knowledge to a new computer system level, Newell removed the need for deciding on the best form of knowledge representation: Any representation can implement knowledge in its own way.

But there is no knowledge without symbols, a view consistent with the earlier *physical symbol systems* hypothesis [11]. A discussion of competing views such as *subsymbolic AI* would be beyond the scope of this article, but as Newell correctly points out, his levels metaphor follows the practice of AI and computer science research in general.

The primary goal of computer science is to make computers more powerful and to ensure that computed results are correct at all times. Then how is correctness verified on the knowledge level? The highest directly observable system level contains symbols that might or might not encode some specific knowledge, but knowledge itself would be removed from direct observation. Newell proposed a mechanism of indirect verification: If some (human or artificial) *agent A* can detect the impact of some specific knowledge in the actions of another agent B, agent A can verify the presence of such knowledge in agent B.

Objectivist/Realist World View

The world view underlying the computer science notion of knowledge can be described as *objectivist* or *realist*, which can be detected by discussing those aspects that authors like Newell do not discuss, and the questions that are not answered. A good example is: "Who is the agent that will verify the presence of knowledge in another agent?" More than one agent could take the role of the observing agent A in Newell's knowledge verification mechanism and the different As could come to different conclusions on whether knowledge is present in agent B (Fig. 3).

Newell does not explain how the A agents would be able to come to any consistent conclusion on agent B's knowledge, and since their symbol levels might be mutually incompatible, it is unclear how they could even communicate their different views on agent B. The only conceivable way Newell's knowledge verification might yield consistent results is by relying on a standardized observer agent A that serves as the absolute

ON MODELING HUMAN-COMPUTER CO-CREATIVITY

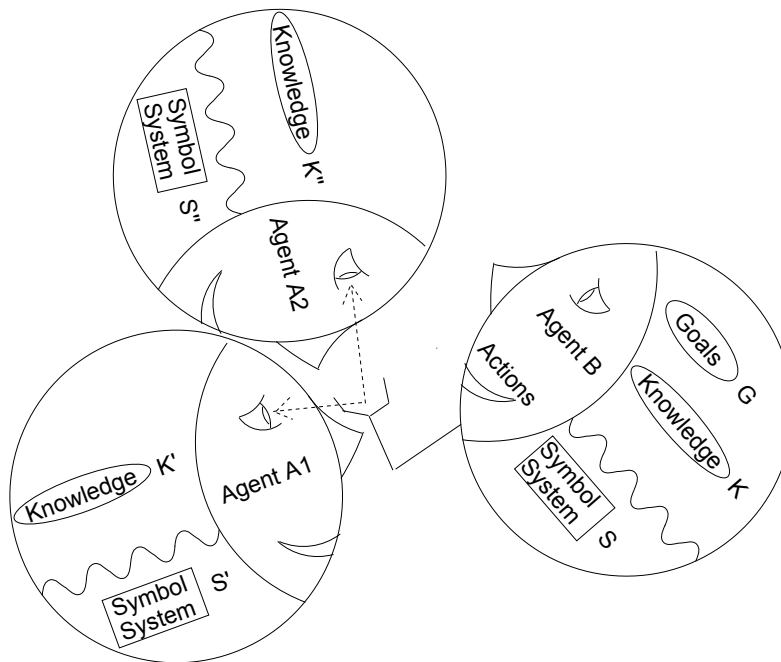


Fig. 3. how do different observing agents agree on the presence of knowledge?

reference on verifying knowledge. Newell implicitly assumes *objective knowledge* that would be the basis, rather than the result of attempts at creating artificial intelligence.

Such assumptions are quite common in information technology. In fact, Shannon's definition of *information* [19] only works under the assumption of an objective observer for assessing the probability of the next symbol. Shannon actually used an object in the form of the English language dictionary as a means for replacing subjective meaning with an objective type of measurement.

For good reason, the objectivist/realist view has a long tradition in computer science. Computer systems are first of all objects which are required to work correctly and independent of the subjects that created them. The assumption (or implicit goal) of objective knowledge and objective correctness is therefore in general useful for this research field. But this view is in stark contrast to the more *subjectivist* or *constructivist* view in psychology and CR. Realist and constructivist world views each have their role in science and engineering [13], but the constructivist view is a better fit for topics such as learning or creativity. And some key notions from CR can only be fully appreciated within the constructivist view.

ON MODELING HUMAN-COMPUTER CO-CREATIVITY

Divergent Thinking in Creativity Research

The standard definition of *creativity* in CR is bipartite: Creativity requires both *novelty* (also called *originality*) and *effectiveness* [17]. Effectiveness is quite compatible with the realist/objectivist world view adopted in CC, but novelty poses some challenges. Where does novelty come from?

In human creativity, two different modes of thinking have been observed: *Convergent thinking* and *divergent thinking* [6]. Problem-solving as search [9] is a typical objectivist approach in AI and a good example of convergent thinking: First all relevant constraints are determined and then possible solutions are checked for their consistency with these constraints. The end result will then be the solution with the least amount of divergence from what was required.

Divergent thinking does the exact opposite: exploring multiple paths of thought seemingly irrelevant and even inconsistent with the task at hand and only later merging some of the paths into a new, potentially surprising, and eventually also consistent result. Divergent thinking temporarily ignores all the criteria of correctness and objectivity that are at the heart of computer engineering. And divergent thinking has been shown to be an integral part of human creativity [16].

Convergent thinking is important for creativity as well. But the normal, convergent mode of thinking is regarded as a given in CR, and the divergent mode as a special addition required for true creativity. Divergent thinking is seen at the core, divergent thinking at the periphery of creativity. The four P's of creativity [15, 18] are a widely adopted framework of research consisting of Product, Person Process and Press. The Press portion stands for environmental and social constraints, which of course can only be accommodated with the appropriate amount of convergent thinking.

So the Press aspect of creativity can be seen as mandating a certain amount of realism/objectivism. But external constraints can also be interpreted from a radical constructivist point of view. In his highly influential study, Csikszentmihalyi [4] has for instance argued, that creativity can only be evaluated against its social frame of reference, denying any realist aspect of creativity.

Radical constructivism is of course not suited for describing computer systems, since computer systems are expected to work correctly and autonomously. The kind of radical realism adopted in AI on the other hand is ill suited for understanding human creativity which relies on the subject's ability of creating new and unexpected meaning. This does not imply that computer systems cannot be creative. It does however imply that the world view typically adopted in computing is inherently limited in its ability to describe and understand creativity. Divergent thinking is the one aspect of human creativity highlighting the shortcomings of the standard view of knowledge and meaning in computing: Any symbol, any meaning and the relevance of any information or assumption are up for re-interpretation in the divergent thinking stage of the creative process. HC³ involves both human actors and computer systems, therefore a scientific model of HC³ will have to accommodate both world views.

ON MODELING HUMAN-COMPUTER CO-CREATIVITY

What a model could look like

In spite of the objectivist/realist notions for describing computer systems, AI researchers have repeatedly argued that their models do not have to conform to the objectivist requirement of falsifiability formulated by Popper [14]. Which assumptions from computing have been validated and should be incorporated in an integrated model? Newells knowledge level hypothesis was for instance meant to be validated via the success of strong AI. Apart from rather technical implementation details on all the levels below the knowledge level, classical strong AI had only one falsifiable scientific hypothesis: That it would be able to build intelligent robots and that it would be able to build them in a certain time frame. This one falsifiable hypothesis has turned out to be false and therefore the knowledge level hypothesis could be considered falsified. But the research approach described by this hypothesis is still applied today, so his hypothesis should rather be regarded as a description of research culture. Rather than documenting all of their failures, computer engineers tend to modify their computer models until they arrive at a working prototype. Falsification has a very low priority in computing, to the degree that the criterion of falsifiability is argued to be irrelevant for instance in cognitive science [3] While such a form of scientific discovery might be appropriate for the development of purely artificial systems, it certainly cannot be applied to a phenomenon involving human actors. A scientific model of HC³ has to incorporate falsifiable hypotheses.

As opposed to the typical CC model, a model of HC³ cannot be a computer model. The human and computer side of such an interaction are different and will have to be described in different ways. The objectivist/realist world view will be more suitable for describing the computer system, the subjectivist/constructivist view more suitable for describing the human partner. Their interaction will have to be described from both the human and computer perspective as well. The model should accommodate all the key notions from relevant research fields and describe compatible computer system properties. It should describe the interactions between humans and computers and the creative process. It should incorporate both an objectivist and a subjectivist understanding of knowledge and should explain the role of knowledge in the creative process. It should make the role of computer and human partner in divergent and convergent thinking explicit and it should be able to predict the impact of computer use on creativity.

What does divergent and convergent thinking have to do with knowledge? I will introduce two new notions: Diversion and consolidation. Diversion of knowledge involves selectively disconnecting symbols, meaning and information and re-creating connections irrespective of external and internal constraints. Consolidation of knowledge involves merging previously inconsistent symbols, meaning and information into a new consistency. Concept blending [12] can for instance be seen as a type of knowledge consolidation.

The following hypotheses on HC³ are at least in principle falsifiable:

1. HC³ exists as a unique phenomenon
2. Creativity as resulting from HC³ can be measured in the same way as human creativity
3. HC³ is more than the sum of its parts: The creative co-operation of computer(s) and person(s) yield results that are either more creative or creative in a different way

ON MODELING HUMAN-COMPUTER CO-CREATIVITY

4. HC³ The HC³ process has observable properties and has general properties independent of application domain
5. the process of creative co-operation is different from other types of HCI
6. HC³ requires both convergent and divergent thinking
7. divergent thinking involves the diversion of knowledge
8. convergent thinking depends on the previous consolidation of knowledge
9. Diversion and consolidation can be achieved by human or computer, but in fundamentally different ways and never at the same time
10. Each step in the HC³ process can be characterized by the flow of control between human and computer

By such a deep integration of key research notions across disciplines, system properties such as openness or playfulness [21] can be analyzed in a systematic way, for instance as resulting from the availability of knowledge diversion tools. HC³ is an iterative process, and in each of the process stages, human or computer can perform one of the following functions:

1. diverge: multiple options are explored with little regard for consistency or correctness
2. compare: intermediate results are compared against original requirements
3. consolidate: previously inconsistent knowledge is blended into new consistent knowledge
4. converge: consistent solutions are generated

And finally the hypothesis:

- Both human and computer can be the driving agent for either convergent or divergent thinking

If this model is correct, human and computer can be creative in very different ways. One scenario would be the diverging computer partner generating seemingly inconsistent material with the human partner selecting and finally merging some of this material into something new. Or vice versa: The human partner could be providing the role of divergent thinking and generating candidate solutions which are then evaluated or consolidated by the computer partner. The properties for the design of creativity support systems will be different for each of these scenarios, but if these hypotheses are correct, some properties will be relevant for all of them. Neither human nor computer would have to be creative on their own. But if we want to identify Human-Computer Co-Creativity, then we should be able to identify the stages of such a process as well as the overall creative result.

Conclusion

Human-Computer Co-Creativity is an old research goal which has seen renewed interest in the form of Creativity Support Tools in the last years, but a general scientific model of the process of HC³ is still missing. Elements for such a model can be found in research fields for different aspects of HC³. Therefore a synthesis of research results is

ON MODELING HUMAN-COMPUTER CO-CREATIVITY

an appropriate option for building a model. But the research communities are exhibiting fundamental differences in research culture and epistemology, which have to be acknowledged and considered. Divergent thinking and knowledge are key concepts and a model of creative co-operation has to describe which roles are played by which partner in relation to these concepts. I am formulating the hypothesis that both human and computer can be the agents for either convergent or divergent thinking. I am discussing properties of a general model of HC³ and a path towards consolidation of models and research approaches across the existing research fields. Both may serve as examples of the challenges as well as the opportunities that still lie ahead of us.

References

- [1] Bush, V.: As we may think. Atlantic Monthly (July 1945),
- [2] Chen, C.: Turning Points : The Nature of Creativity. Springer (2012),
- [3] Cooper, R.P.: The role of falsification in the development of cognitive architectures: Insights from a lakatosian analysis. Cognitive Science 31(3), 509–533 (2007),
- [4] Csikszentmihalyi, M.: Creativity - Flow and the Psychology of discovery and invention. HarperPerennial, NY, USA (1997)
- [5] Engelbart, D.C.: Augmenting human intellect - a conceptual framework. Tech. Rep. AFOSR-3233, Stanford Research Institute, Menlo Park, CA, USA (1962),
- [6] Guilford, J.P.: Creativity. American Psychologist 5, 444–454 (1950)
- [7] Hey, T., Tansley, S., Tolle, K.: The Fourth Paradigm: Data-Intensive Scientific Discovery. Microsoft Research (2009),
- [8] Licklider, J.C.R.: Man-computer symbiosis. IRE Transactions on Human Factors in Electronics HFE-1, 4–11 (March 1960),
- [9] Newell, A.: Human Problem Solving. Prentice-Hall, Inc., Upper Saddle River, NJ, USA (1972)
- [10] Newell, A.: The knowledge level. Artificial Intelligence 18(1), 87–127 (1982),
- [11] Newell, A., Simon, H.A.: Computer science as empirical inquiry: Symbols and search. Commun. ACM 19(3), 113–126 (Mar 1976)
- [12] Pereira, F.C.: Creativity and Artificial Intelligence: A Conceptual Blending Approach. Walter de Gruyter and Co. Hawthorne (2007)
- [13] Peschl, M.F.: Constructivism, cognition, and science: An investigation of its links and possible shortcomings. Foundations of Science 6(1-3) (2001),
- [14] Popper, K.: Logik der Forschung. Springer (1935)
- [15] Rhodes, M.: An analysis of creativity. The Phi Delta Kappan 42(7), 305–310 (1961)
- [16] Runco, M.A., Acara, S.: Divergent thinking as an indicator of creative potential. Creativity Research Journal (2012)
- [17] Runco, M.A., Jaeger, G.J.: The standard definition of creativity. Creativity Research Journal 24(1), 92–96 (Feb 2012)
- [18] Runco, M.A., Pritzker, S.R.: Encyclopedia of creativity. Elsevier (1999)
- [19] Shannon, C.E.: A mathematical theory of communication. Bell System Technical Journal 27, 379–423 and 623–656 (1948),
- [20] Shneiderman, B., Fischer, G., Czerwinski, M., Resnick, M., Myers, B., Candy, L., Edmonds, E., Eisenberg, M., Giaccardi, E., Hewett, T., Jennings, P., Kules, B., Nakakoji, K., Nunamaker, J., Pausch, R., Selker, T., Sylvan, E., Terry, M.: Creativity Support Tools: Report from a U.S. National Science Foundation sponsored workshop. International Journal of Human-Computer Interaction 20(2) (2006),
- [21] Voigt, M., Niehaves, B., Becker, J.: Towards a unified design theory for creativity support systems. In: Design Science Research in Information Systems: Advances in Theory and Practice. Springer (2012)

Comparison of the Two Companies' Staff by the Group Idea-Marathon Training on Creativity and Motivation

Takeo Higuchi, PhD (Knowledge Science)

Idea-Marathon Institute, Tokyo, Japan

info@idea-marathon.net ,

Abstract. This paper presents the results of a comparative analysis of different levels of motivation for the R Company staff (R Group) and the P Company staff (P Group) through the three month Group Idea-Marathon training (G-IMS) on Creativity.

Various observations during the training seminars and the support system show that the R Group was less serious in the course than the P Group. In this paper, the two companies' group training were analyzed by comparing creativity effects based on their respective motivation.

The R Group participants were mainly forced by their department heads to participate in this Idea-Marathon training or just nominated to participate..

Many staff of the R Group did not appear at the second or third training workshop within 3 months. They did not fulfill the recommended assignment of thinking and writing in their notebook every day.

On the other hand, the P Group participants, however, mainly consisted of staff who voluntarily participated in the Idea-Marathon seminar.

All the members of both groups were given the same creativity tests of the Torrance Tests of Creative Thinking (TTCT) as Pretest and Posttest with three months training of the G-IMS.

The T-test analysis of both group scores for five Norms Reference Measures showed that the P Group improved more than the R Group in creativity while the Mann-Whitney U-test for 13 items Criterion Measures showed that the R Group had more variety of improvement in strength than the P Group.

But the two-factor factorial ANOVA analysis between the R and P Groups, and within and Pretest and Posttest show that there were no difference statistically in TTCT main creativity Items. Reviewing the result of ANOVA and ETS score, the Idea-Marathon was found to be effective in improving creativity in the R & P Groups and we noticed that the R Group could be more creative if they had stronger motivation in their self-enlightenment.

Keywords: Idea-Marathon System (IMS), Torrance Tests of Creative Thinking (TTCT)Figural tests, Support Systems, motivation

1 Introduction

Today, any companies and laboratories in Japan and in the world are seeking higher creativity for their staff and researchers [1]. Many seminars with various kinds of newest popular creativity thinking methods are being offered to their staff, and the staff might voluntarily attend the seminar or attend with the feeling to be forced to.

Many major Japanese companies have been challenged by China, Korea and other globalized companies. Sony, Sharp and Panasonic have been in a painful situation in recent years. The only way to escape from this adverse situation is to create unique new products, for which before anything group creativity is required.

But through Lost Decades in Japan, the staff of many companies have been said to become less curious, more silent, non-confident and non-creative. Many staff might have been caught by a kind of learned helplessness, which was warned by Seligman [2].

When the author was giving IMS training in the two R & P Companies, it was observed that the enthusiasm in the P Group was stronger than that in the R Group.

(1) The general attitude of the participants of both companies was almost the same and also both were fairly devoted during the training seminar period.

(2) The rate of absentees of the R Group (66.7%) at the third training seminar (three months later) was much higher than that for the P Group(4.5%).

(3) One of the Support Systems of G-IMS training, ETS (see 2.2) had shown that the R Group recorded much less than the P Group.(Table 1)

(4) It was estimated that the P Group had stronger motivation than the R Group.

Almost all the original members of the P Group attended the third training seminar while many of the R Group were absent in their third training seminars. Therefore, we consider that the high rate of absentees in the R Group was a primary evidence of low motivation.

From these facts, we hypothesize that, if the participants had a high level of motivation, the G-IMS support system would be well followed, and also the result of the TTCT score could be higher as well, and if the participants had a lower level of motivation, or a kind of helplessness, the support system would be less followed and the TTCT test result would be affected accordingly.

In this paper, two companies' group experimental results were studied.

In this Chapter1 Introduction, we explain the origin of our hypothesis of these two companies' staff. In Chapter 2, we explain the IMS and the support system of ETS and supply of thinking hints. In Chapter 3, we present the TTCT tests. In Chapter 4, we present the comparative analysis of two companies' staff. In Chapter 5, we discuss the results. We conclude in Chapter 6 and discuss future research, followed by references.

2 Idea-Marathon and Its Support Systems

2.1 Basic Concept of the Idea-Marathon (IMS)

In 1984, the Idea-Marathon System (IMS) was developed and implemented by the author, who published the first book on IMS in Japanese in 1992 [3]. Since 1984, the author has continued to practice IMS almost every day, thus for the last 30 years. IMS books were published also in English [4], Korean, Chinese, Thai, Hindi and Nepali.

The principles of IMS are defined as follows:

- (1) Use notebooks of the same kind.
- (2) Generate new ideas daily and write them in the notebook chronologically, numbering each idea regardless of any categories.
- (3) Draw pictures to illustrate your ideas as often as possible.
- (4) Talk to your neighbors.
- (5) Review your ideas.
- (6) Implement the best ideas [5]

IMS is a process which involves daily idea creation and immediate writing in notebooks individually following a chronological order. The ideas expressed are not limited to any specific area or topic but may encompass any subject area. As IMS becomes a daily habit, the practitioner will be encouraged by experiencing the power of continuity, instantaneity and self-confidence that stem from increased creativity. A synergistic effect can also be expected to emerge among those who practice IMS in groups if they discuss the ideas in their notebooks.

Since the Idea-Marathon Institute was established in 2004, the institute administered the group training of many companies and laboratories.

To make wider IMS effects in companies and laboratories in Japan, IMS has been studied quantitatively for its potential to enhance creativity. TTCT Figural tests were implemented for studies to prove the creative effect of IMS in details. Prior to this experiments at two companies, the creativity effect of IMS for researchers at one major food manufacturer's laboratory near Tokyo was already tested and analyzed by Higuchi et al in 2011[6].

And the P Group, one of the two companies discussed in this paper, has been tested, analyzed and discussed in the author's dissertation [7]

After we made another group training of IMS for 3 months at R Group staff with Pretest and Posttest of TTCT, we made our comparative study between R& P staff using the TTCT figural test Pretest and Posttest.

2.2 ETS (e-Training System) and Its Weekly e-Hints

(1) ETS (e-Training System).

ETS is a support system for IMS practitioners using e-mails. During the IMS training period, the participants are supposed to send regularly the current numbers of ideas written in their notebooks to the IMS Institute, and they receive individual comment with advice, encouragement, useful references, thinking hints from the Institute.

This is a hand-made training support system or monitoring system by humans for IMS, though the actual content of participants' ideas is not checked nor reviewed.

After starting IMS, if IMS participants are not advised and encouraged regularly and properly before the next training seminar taking place a few months later, many participants will be naturally and gradually getting not steady, slowing down or stopping practicing IMS as their motivation decreases.

These comments from the IMS Institute will keep positively encouraging participants to continue or accelerate IMS. This is named ETS (e-Training System). During 3 months of IMS training, 5 to 6 times ETS were implemented. In these experiments at two companies, we tried to study the correlation between the participants score of staff of both companies of creativity test and number of ideas recorded in ETS.

(2) ETS Data of R & P Groups As Idea-Numbers, Average Per Person.

We checked ETS data and found that the ETS result of both those companies greatly differed.

Both the R & P Groups are large, busy electronic and software companies. However, in the case of the R Group, the rate of attendance to the third IMS lecture/workshop within a 3 month period was exceptionally low (Ref 1 (2)).

Of course this rate of absence includes unavoidable business assignments for the staff to stop attending the third lecture/workshop. But still we understand many avoided the third IMS lecture/workshop intentionally and without any business reason. These absentees thought that there would be others attending in the training. Therefore, if they could be sleepers, the training seminar would go on without them. This is a kind of irresponsible idleness that each of these absentees would like to be a sleeper as they knew that they will not be officially complained by the management of their absence in the training through their past experience.

In the case of the P Group, at the first lecture/workshop (with TTCT tests Type A), the number of the participants was 25 persons in total, but at the third lecture/workshop (with TTCT tests Type B) after 3 months, the participants decreased to 21 persons, which was an 84% attendance rate of the total number of participants.

If we are supposed to create and write one idea per day, after almost 3 months, the balance of ideas can indicate how the participants are creating every day.

Participants ID	5th ETS No. of days	5th ETS No. of Ideas	Balance of Ideas
1	87	30	-57
2	87	110	23
3	87	140	53
4	87	145	58
5	87	90	3
6	87	121	34
7	87	55	-32
8	87	142	55
9	87	27	-60
10	87	43	-44
11	87	87	0
12	87	175	88
13	87	88	1
14	87	54	-33
15	87	87	0
16	87	65	-22
17	87	36	-51
18	87	90	3
19	87	45	-42
20	87	30	-57
21	87	98	11
22	87	96	9
23	87	91	4
Idea no. Average per person		84.60	
Idea no. Average per person per day		0.97	

Participants ID	6th ETS No. of days	6th ETS No. of Ideas	Balance of Ideas
1	92	192	100
2	92	131	39
3	92	114	22
4	92	273	181
5	92	184	92
6	92	163	71
7	92	98	6
8	92	115	23
9	92	135	43
10	92	212	120
11	92	392	300
12	92	111	19
13	92	308	216
14	92	152	60
15	92	98	6
16	92	142	50
17	92	239	147
18	92	158	66
19	92	101	9
20	92	101	9
21	92	292	200
Idea no. Average per person		175.30	
Idea no. Average per person per day		1.91	

Table 1. Comparison of R Group (left) and P Group (right) and Numbers of Ideas After Almost Three Months

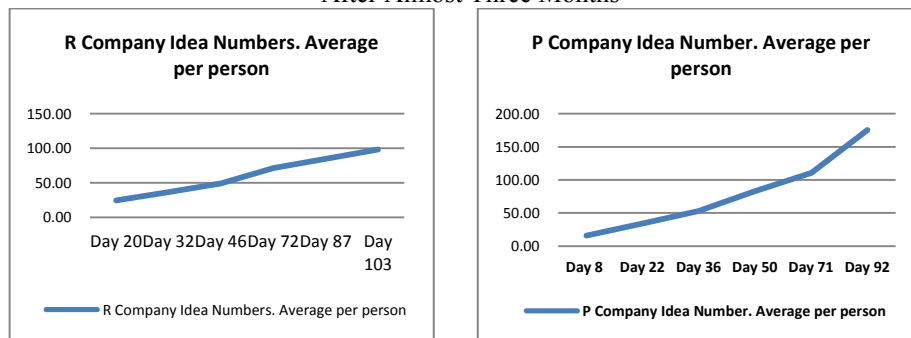


Fig. 1. ETS Data of R Group (Left) and P Group (Right) and Numbers of Ideas Per Person

In Figure 1 above, it seems that both the R & P Groups remarked increasing ideas. However, it is clear that the number of ideas per person per day for the R Group (left) is slowly going down below “one idea per day per person”, while it is going up to almost “two ideas per day per person” in the case of the P Group (right) (Figure 2).

“Below one” means that the members of the R Group were not creating one idea per day per person but they were skipping some days, without any ideas. However, the P Group was creating nearly two ideas per day per person.

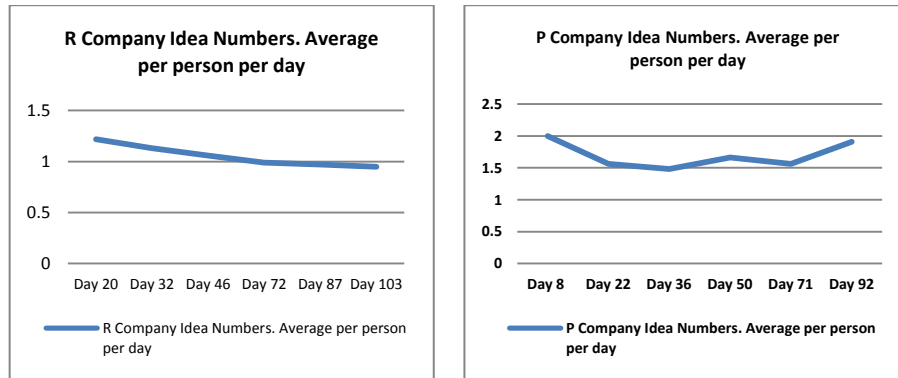


Figure 2 .ETS Data of R Company Staff (Left) and P Company Staff (Right) As Numbers of Ideas Per Person Per Day

(3) Weekly e-Hints.

At the beginning stage of the IMS training, it is important for beginners to find some thinking tips and hints for creating ideas in order to continue practicing the IMS every day.

To improve this situation, participants of IMS training can receive weekly hints from the Idea-Marathon Institute by e-mail (e-Hints), from which they may create new ideas until they get used to creating hints themselves. Some examples of e-Hints are: “Create a new Sushi recipe”, “Think of a new type of vending machine” or “Think of new functions and uses for a calculator”, and so on.

3 Creativity Measurement Using TTCT

To evaluate the creative effect of IMS training on both companies, we used the Torrance Tests of Creative Thinking (TTCT) Figural tests. We carried out the TTCT figural tests for groups of both companies by the Pretest in the first lecture and the Posttest in the third lecture to obtain quantitative creativity factors of the IMS training. There are two types of TTCT Figural tests (Type A as the Pre-test and Type B as the Post-test) which were used to measure the creativity effects of continuous training of the Idea-Marathon for the 3 months period.

TTCT Figural tests consist of the following five Norm-referenced measures [8]:

(1) Fluency: The number of expressed ideas which meaningfully utilize the stimulus. Fluency is the gatekeeper of TTCT analysis.

(2) Originality: The uniqueness of the ideas in terms of statistical infrequency.

(3) Elaborations: The imagination and exposition of detail in the pictures.

(4) Abstractness of Titles: The level of synthesis and organization evidenced in the titles in order to express the essence of the pictures.

(5) Resistance to Premature Closure (RPC): The ability to consider all relevant information and resist impulsive, premature conclusions.

In addition to the above-mentioned 5 Norms Reference Measures, there are 13 more Creative Strengths as Criterion-Referenced Measures in TTCT analysis to check the existence or nonexistence of: Emotional Expressiveness, Storytelling Articulate-ness, Movement and Action, Expressiveness of Titles, Synthesis of Incomplete Figures, Synthesis of Lines and Circles, Unusual Visualization, Internal Visualization, Extending or Breaking Boundaries, Humor, Richness of Imagery, Colorfulness of Imagery and Fantasy.

To make the inter-rater scoring, the first author obtained the official certificate for scoring TTCT tests Figural Type A test and Type B test at a scoring seminar held by the TTCT Center of the University of Georgia, from October 1 to October 3, 2012.

4 Comparative Analysis of the Two Companies

4.1 *t*-Test Analysis of TTCT Score of Companies R and P

For the R Group, the *t*-Test for five Norm-Referenced Measures of TTCT Figural tests between Pretest and Posttest scores was studied. A statistically significant difference was found on Fluency $t(23)=2.144, p<0.05$. No other significant difference was found in the component of the Total Score, Originality, Elaborations, Abstractness of Titles and Resistance to Premature Closure. Figures in Table 2 depict average scores.

For the P Group, the *t*-test for TTCT Figural test Pretest and Posttest scores for five Norm-Referenced Measures showed statistically significant results in Score Total $t(20)=2.187, p<0.05$ and Originality $t(20)=3.371, p<0.01$. There was no significant result in Fluency, Elaborations, Abstractness of Titles and RPC.

R Company TTCT Pre-Post test t-Test				
Measurement	Change of Scores between Pre-Post tests		<i>t</i> value	Sig
	Pre-test	Post-test		
Measurement	M(SD)	M(SD)		
Total Score	582 (59)	574 (67)	0.79	<i>n.s.</i>
Fluency	89 (15)	99 (17)	2.144	$p<0.05$
Originality	108 (22)	117 (20)	1.927	<i>n.s.</i>
Elaborations	150 (12)	150 (26)	0.219	<i>n.s.</i>
Abstractness of Titles	110 (26)	111 (25)	0.103	<i>n.s.</i>
RPC	105 (20)	98 (17)	1.198	<i>n.s.</i>
M=Score Means SD=Standard Deviation df=23, $p<0.05$ $p<0.01$ (Two sided t-test)				

P Company TTCT Pre-Post Score t-Test				
Measurement	Pre-test	Post-test	<i>t</i> -value	Sig
	M(SD)	M(SD)		
Measurement	M(SD)	M(SD)		
Score Total	557(97)	591(85)	2.187	$p<0.05$
Fluency	99(19)	110(25)	1.99	<i>n.s.</i>
Originality	111(27)	126(23)	3.371	$p<0.01$
Elaborations	145(18)	152(11)	1.878	<i>n.s.</i>
Abstractness of Titles	106(24)	106(34)	0.018	<i>n.s.</i>
RPC	95(22)	97(17)	0.576	<i>n.s.</i>
M=Score Means, SD=Standard Deviation df=20, $p<0.05$, $p<0.01$ (Two sided t-test)				

Table 2. *t*-test results of TTCT Scores of R Company (Left) and P Company (Right)

4.2 13 Items Criterion-Referenced Measures Scores of TTCT of R and P Company Staff

13 Items Criterion-Referenced Measures Scores of the TTCT test Pretest and Post-test in both R & P staff were analyzed with the Mann Whitney U-test:

R Company 13 Items Criterion-Referenced Measures of TTCT Score					
		Pre-test	Post-test	t-Value	Sig.
	Criterion reference measure	Mean Average	Mean Average		
1	Emotional Expressiveness	23	26		<i>n.s</i>
2	Story-telling Articulatenees	21	28		<i>n.s</i>
3	Movement and action	23	26		<i>n.s</i>
4	Expressiveness of Titles	23	26		<i>n.s</i>
5	Synthesis of Incomplete Figures	25	21		<i>n.s</i>
6	Synthesis of Lines or Circles	18	31		<i>p<.01</i>
7	Unusual Visualization	18	31		<i>p<.01</i>
8	Internal Visualization	24	25		<i>n.s</i>
9	Extending or Breaking Boundaries	22	27		<i>n.s</i>
10	Humor	22	27		<i>n.s</i>
11	Richness of Imagery	25	24		<i>n.s</i>
12	Colorfulness of Imagery	23	26		<i>n.s</i>
13	Fantasy	20	30		<i>p<.01</i>
14	Score Total of 13 Creative Strength M(SD)	3.4 (9)	3.5 (13)	5.03	<i>p<.01</i>
Item 1-13 Mann-Whitney U-test. Item 14 t-test					

P Company 13 Items of Criterion-Referenced Measures by TTCT Scores					
		Pre-test	Post-test	t-Value	Sig.
	Criterion Reference Measure	Mean Average	Mean Average		
1	Emotional Expressiveness	21	22		<i>n.s.</i>
2	Story-telling Articulatenees	20	23		<i>n.s.</i>
3	Movement and action	21	22		<i>n.s.</i>
4	Expressiveness of Titles	20	23		<i>n.s.</i>
5	Synthesis of Incomplete Figures	20	23		<i>n.s.</i>
6	Synthesis of Lines or Circles	17	26		<i>p<.01</i>
7	Unusual Visualization	21	22		<i>n.s.</i>
8	Internal Visualization	21	22		<i>n.s.</i>
9	Extending or Breaking Boundaries	21	23		<i>n.s.</i>
10	Humor	21	20		<i>n.s.</i>
11	Richness of Imagery	18	25		<i>p<.05</i>
12	Colorfulness of Imagery	22	21		<i>n.s.</i>
13	Fantasy	20	23		<i>n.s.</i>
14	Score Total of 13 Creative Strength M(SD)	10(6)	13(6)	2.568	<i>p<.05</i>
d=16, Item 1-13 Mann-Whitney U-test. Item 14 t-test					

Table 3. Comparison of Both R & P Staff on 13 Items Criterion-Referenced Measures Score of TTCT Figural Test Between Pretest and Posttest with the Mann Whitney U-test

4.3 ANOVA analysis between Companies P & R and Pretest and Posttest

(1) Score Total

Then, the analysis of the Score Total by two factor factorial ANOVA within Pretest and Posttest and between R&P staff indicated a significant main effect for Pre-Post test ($F(1,43)=5.222$, $p<.05$), the significance of which was confirmed by Bonferroni multiple comparison ($p<.05$). No interaction was found.

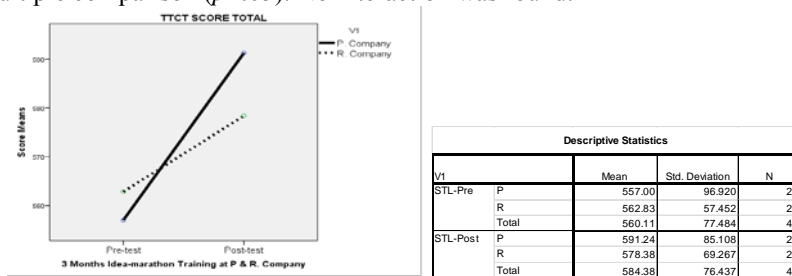


Fig. 3 ANOVA Comparison between R & P Staff of TTCT Score Total in Pretest and Posttest

(2) Fluency

The analysis of Fluency by two factor factorial ANOVA within Pretest and Posttest and between R & P staff indicated a significant main effect for Pre-Post test ($F(1,43)=9.186$, $p<.01$), the significance of which was confirmed by Bonferroni multiple comparison ($p<.05$). No interaction was found.

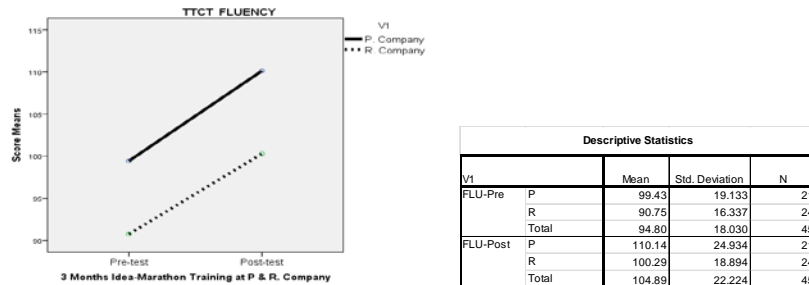


Fig. 4. ANOVA Comparison between R & P Staff of TTCT Fluency in Pretest and Posttest

(3) Originality

The analysis of Originality by two factor factorial ANOVA within Pretest and Posttest and between R & P staff indicated a significant main effect for Pre-Posttest ($F(1,43)=15.089$, $p<.01$), the significance of which was confirmed by Bonferroni multiple comparison ($p<.05$). No interaction was found.

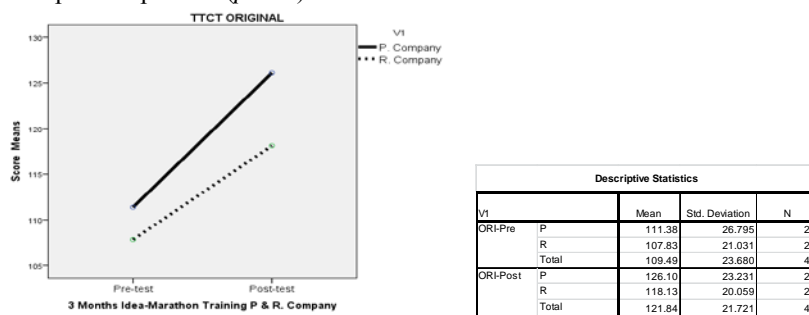


Fig. 5. ANOVA Comparison between R & P Staff of TTCT Originality in Pretest and Posttest

(4) Elaborations

The analysis of Elaborations by two factor factorial ANOVA within Pretest and Posttest and between R & P staff did not indicate any significant main effect for Pre-Posttest. No interaction was found.

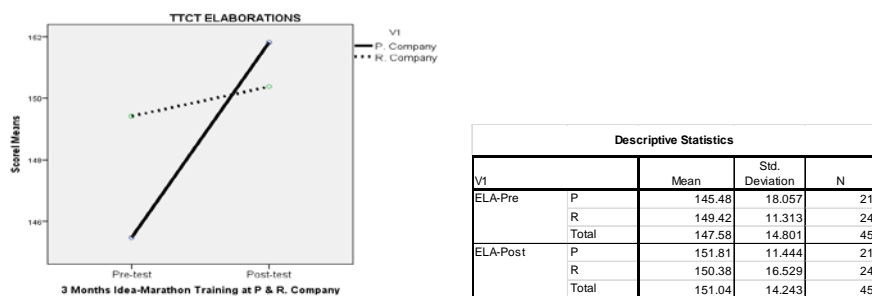


Fig. 6. ANOVA Comparison between R & P Staff of TTCT Elaborations in Pretest and Posttest

(5) Abstractness of Titles

The analysis of Abstractness of Titles by two factor factorial ANOVA within Pre-test and Posttest and between R & P staff did not indicate any significant main effect for Pre-Posttest. No interaction was found.

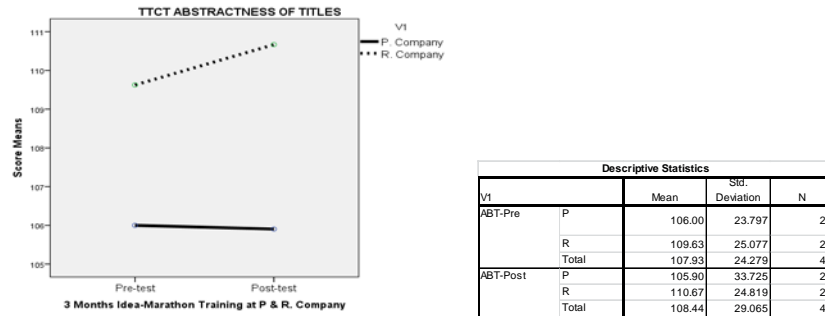


Fig. 7.ANOVA Comparison between R & P Staff of TTCT Abstractness of Titles in Pretest and Posttest

(6) Resistance to Premature Closure

The analysis of Resistance to Premature Closure by two factor factorial ANOVA within Pretest and Posttest and between R & P staff did not indicate any significant main effect for Pre-Posttest. No interaction was found.

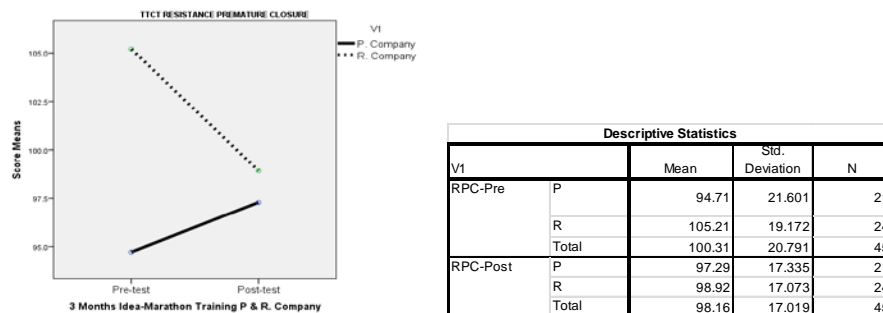


Fig. 8.ANOVA Comparison between R & P Staff of TTCT RPC in Pretest and Posttest

4.4 Analysis of Correlation Between TTCT Scores and ETC Number of Ideas in R & P Staff

In the case of the R staff, the correlation between the number of ideas recorded after three months in the ETS and each of the five Norm-Referenced Measures and Score Total was checked with the Pearson's product-moment correlation coefficient method. ETS's number of ideas correlated with the TTCT's Abstractness of Titles ($r=.534$, $p<.01$). Other Norm items of Fluency, Originality, Elaborations, RPC and Score Total were not found to be correlated.

In the case of the P staff, the correlation between the number of ideas recorded after three months in the ETS and each of the five Norm-Referenced Measures and Score Total was checked with the Pearson's product-moment correlation coefficient method.

od. ETS's number of ideas correlated with the TTCT's Originality ($r=.516, p<.05$) and Elaborations ($r=.522, p<.05$) ETS's number of ideas correlated with the TTCT's. Other Norms of Fluency, Abstractness of Titles, RPC and Score Total were found not to be correlated.

5 Discussion

The discussion from the above-mentioned results is as follows:

(1) Judging from the data of ETS (Figures 1 & 2), the P Group was showing better results than the R Group. Actually, according to Figure 2, the R Group was not creating ideas written in a notebook everyday while the P Group was creating very high daily rate of idea creation (Ref. 2.2 (2)).

From this situation, prior to our Posttest of TTCT, we made one hypothesis that the R Group should have achieved lower scores of the TTCT Pretest and Posttest than the P Group.

(2) The result of *t*-Tests to analyze the TTCT Pre-Posttest score of both the R & P staff showed a significant difference for the R Group for only Fluency with the P Group for Score Total and Originality. But this did not clearly show that the R Group was inferior in score to the P Group (Ref. 4.1).

(3) From the viewpoint of the 13 items Criterion-Referenced Measures, the R Group had more significant items than the P Group. (Ref. 4.2).

(4) Then, we proceeded to two factors factorial ANOVA analysis between the R and P Groups with Pre-Posttest. The ANOVA test confirmed that there were no significant differences nor interactions between the R & P Groups, although there were significant results between Pretest and Posttest (Ref. 4.3).

(4) Between the numbers of ETS ideas and the TTCT Five Norm-Referenced Measure Result, Abstractness of Titles was found to be significantly correlated in the R Group while Originality and Elaborations are found to be significantly correlated in the P Group. (Ref 4.4).

(5) From all of the items (1) to (4), we may probably say that both the R & P Groups showed their creative improvement whatever the ETS results were.

(6) The R Company is one of the major long-standing engineering companies in Japan and its business has been expanding steadily. The staff of the R Company is reputed to be very capable. However, reviewing the increasing number of absentees for the second and third lectures, the management was afraid that their R&D and other engineers were getting narrow-minded, interested in their own present assigned scope of work, especially lack of new business motivations. As the company is quite large, none of the staff imagines of going bankrupt. As a result, the creativity of these engineers has been getting stale.

6 Conclusion and Future Research

From our above-mentioned discussion, we can say that the training of Idea-Marathon was effective in improving the creativity of both the R & P Groups.

Though the ETS of the P Group had significant correlations with TTCT in Score Total and Originality, the R Group had a significant correlation with TTCT in Abstractness of Titles only.

We tried to make our reasoning of what the reason for the large difference in the ETS result between the R & P Groups is. We estimated that the study of the R Group ETS and TTCT Scores was connected with the degree of their back-scratching, falling-in, taking-things-easy, living under the same roof with the same food, settling always amicably and self-estimating that someone else will solve problems. Ken-ichi Omae pointed out these natures [9].

We hypothetically found the reason for a poorer ETS result as the R Group was less motivated about self-enlightenment, and the R Group has much more real ability to improve in TTCT tests if its members get serious. They may have suffered from learned helplessness of Seligman [1].

In future training seminars for any companies, we must make our survey to estimate the motivation and willpower level of the participant individual, and also we must make our investigation about the company business atmosphere so that we can improve our Idea-Marathon group training method for a staff like that of the R Group with more contents of encouraging motivation for self-enlightenment.

REFERENCES

1. Drucker, P., *The Age of Discontinuity: Guidelines to Our Changing Society*, Butterworth-Heinemann Ltd, 1969 (Reprinted in 1992, Transaction Pub.)
2. Seligman, M. E. P, Maier, S. F. Failure to Escape Traumatic Shock, Washington, DC, USA, *Journal of Experimental Psychology*, Vol. 74, No.1, American Psychological Society, 1967
3. Higuchi, Takeo, "You can get affluent ideas from this book". Diamond Inc, Tokyo, 1992 (in Japanese)
4. Higuchi, Takeo, *Ideas in Action, Digital Achievement of Idea Marathon System*, Adarsh Books, 2001
5. Higuchi, Takeo, *Idea-Marathon System*, *Encyclopedia of Creativity, Invention, Innovation, and Entrepreneurs*, Springer, UK, 2013
6. Higuchi, Takeo et al., *Creativity Improvement by Idea-Marathon Training, Measured by Torrance Tests of Creative Thinking (TTCT) and Its Applications to Laboratories*, KICSS 2012
7. Higuchi Takeo, Dissertation "Enhancement Effects of the Idea-Marathon System on Creativity," Chapter 6-2, Japan Advanced Institute of Science and Technology, Ishikawa, Japan
8. Torrance, E. Paul, and Ball Orlow E., *Torrance Tests of Creative Thinking, Streamlined Scoring Guide for Figural Forms A and B*, Scholastic Testing Service, Inc, USA
- Omae, Ken-ichi, *Cycle of Devil or Japanese Nature leaning to others*, Shincho Bunko, 1987 (in Japanese)

Evaluation Indexes to Understand the Creative Problem Solving Process in the Distributed and Cooperative KJ Method

Takaya Yuizono¹ and Jun Munemori²

¹ Japan Advanced Institute of Science and Technology, 1-1 Asahidai, Nomi, Ishikawa 923-1211, Japan

² Wakayama University, Sakaedani 930, Wakayama City, 640-8510, Japan

yuizono@jaist.ac.jp, munemori@sys.wakayama-u.ac.jp

Abstract. Evaluation indexes have been proposed to understand group intelligence when using the distributed and cooperative KJ method (the DC-KJ method) with groupware for new idea generation. The DC-KJ method consists of three steps: brainstorming ideas, grouping similar ideas, and writing conclusion sentences. In this research, the number of kinds of words in ideas (as an index of divergent thinking: D-index), number of kinds of words in ideas reflected in the conclusion sentences (as an index of convergent thinking: C-index), and multiplication of both numbers (as an index of group intelligence: M-index) are investigated as indexes to understand the creative problem solving process. These three indexes were applied to data from meetings that integrated collective intelligence and the DC-KJ method. The results showed that the C-index displays the highest correlation with the general satisfaction value of the conclusion sentences.

Keywords: Groupware, KJ method, Idea generation, Divergent thinking, Convergent thinking, Creative problem solving process.

1 Introduction

Collaboration technologies are expected to assist human intelligence or knowledge creation [1-4]. A goal of collaboration research is the realization of electronic conferencing systems to utilize the participants' combined intelligence, or group intelligence. Research on groupware for new idea generation has been carried out since the 1990s [3, 4]. Such research in Japan in particular has tended to focus on the KJ method [5-7]¹, which is a well-known idea generation method in Japan [3].

Currently, group techniques for idea generation, such as brainstorming [8,9] and the KJ method [5-7], are well known and used in meetings. A well-known element of

¹ The KJ method is a registered trademark of Kawakita Research Institute.

the KJ method is the affinity diagram, which is applied in contextual design [10, 11]. These techniques have also been connected to creative problem solving and applied in industry development in various kinds of organizations. Moreover, design firms, such as IDEO, activate idea generation not only using brainstorming, but also through idea finding after observation in the field [12].

Idea generation systems that support the KJ method have been well studied in Japan [3], for example KJ editor [13], D-Abductor [14], GUNGEN [4], and KUSANGI [15, 16]. The KJ method can be described as the technique of summarization after brainstorming. The method has four steps: brainstorming, grouping ideas, making charts, and writing conclusion sentences. The grouping of ideas and the making of charts are result in affinity diagrams. The inventor of the KJ method, Jiro Kawakita, thought that a good meeting process displayed divergent thinking, convergent thinking, and evaluation [17]. These three thinking styles are known to be a part of human operation in Guilford's intelligence model [18]. In particular, a combination of divergent thinking and convergent thinking is important in creative problem solving [19]. These basic thinking styles are supported when using the KJ method with an idea generation support system.

A groupware tool for new idea generation, named GUNGEN [4], has been used to evaluate data resulting from cooperative work. This data consisted of number of ideas, number of idea groups (islands), and number of conclusion sentences, in addition to satisfaction value of conclusion sentences as a reflection of the meeting result. However, the grouping process has an impact on the result of electronic conferencing, but the relationships between the value of a meeting result and the results of each step of the grouping process are not clear. Therefore, it is not possible to evaluate group intelligence within the creative problem solving process.

In this research, we proposed and investigated three indexes, automatically evaluated in electronic conferencing systems, to evaluate the contribution of the grouping process to a good meeting result. By investigating the three indexes with past successful data from the DC-KJ method, we revealed an index that can be used to evaluate group intelligence within the creative problem solving process.

In the next section, we describe related work on the DC-KJ method. In section 3, the proposed indexes are described. In section 4, the target data to investigate the indexes and the calculation method are described. Finally, we present and discuss the results in section 5 and conclude in section 6.

2 Related work

2.1 The distributed and cooperative KJ method

The distributed and cooperative KJ method (the DC-KJ method) is an abbreviated version of the original KJ method and is used in the groupware tool GUNGEN [4]. GUNGEN supports the DC-KJ method with multiple computers over a computer network. The DC-KJ method has three steps: generating ideas through processes such

as brainstorming, grouping ideas (concept formation), and writing conclusion sentences. This process is similar to a process supported by Cognoter, a groupware tool used in Xerox Colab Projects [1]. The process of the DC-KJ method omits the step of making charts from the original KJ method.

The three steps of the DC-KJ method are depicted in Figure 1. In the step of idea generation, participants present as many ideas as possible following the rules of brainstorming. In the next step, they combine similar ideas into groups. Each group (or island) represents a concept or relationship and given a title that reflects its content. Related groups can be placed together. In the final step, participants write conclusion sentences by reviewing all the ideas, groups and group titles. In the DC-KJ method, they are guided to use the island names in the conclusion sentences.

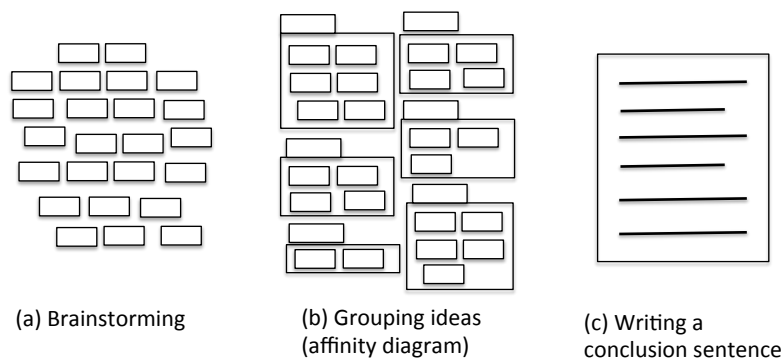


Fig. 1. Three steps of the DC-KJ method.

2.2 General satisfaction value to evaluate the results of the DC-KJ method

A method for evaluating the conclusions reached using the DC-KJ method was proposed by Yagishita and Munemori [20, 21]. The value of the conclusions is called the general satisfaction value. The evaluation method is based on the analytic hierarchy process, a decision-making method, and is divided into two steps.

The first step is to determine the weight of six factors of evaluation of creative artifacts; these factors are originality, convenience, individual attractiveness, correctness, feasibility, and applicability. Each weight value is calculated after paired comparison of all six factors by an evaluator. Following this, the evaluator evaluates the conclusion sentences based on the same six factors.

The second step is evaluation of the conclusion sentences. The evaluator judges the conclusion sentence according to the degree of either satisfaction or dissatisfaction of the six factors. The satisfaction value of each factor of the conclusion sentence is the degree of satisfaction multiplied by the weight of each factor. The sum of all satisfaction values is defined as the s-value. Similarly, the sum of all dissatisfaction values is defined as the ds-value. After two summations, the general satisfaction value of the conclusion sentence is calculated as the s-value divided by the ds-value.

2.3 The DC-KJ method with four conditions of collective intelligence [16]

The concept of collective intelligence, which is estimated as the effects of a web, is expected to have the application of improving group intelligence. Therefore, a meeting process could be improved by considering the four conditions of collective intelligence, diversity, independence, distribution, and aggregation, as proposed by Surowiecki [22]. In particular, two conditions (diversity and independence) are supported when using the DC-KJ method with collective intelligence, and the condition of aggregation is supported when using the DC-KJ method with selected ideas.

In the first step of the DC-KJ method with collective intelligence, independent multiple groups of more than ten people follow the DC-KJ method separately. This step supports independence of collective intelligence. In the second step, each group follows the DC-KJ method with the collected ideas generated during the first step. This step supports diversity of collective intelligence. Two types of idea collection are considered. One type is execution of the DC-KJ method with all ideas from all the groups, the other is execution of the DC-KJ method with the best ideas from each group as selected by the group.

A case study of the DC-KJ method with collective intelligence showed the DC-KJ method with selected ideas, in which a mechanism of idea selection is added to support aggregation, yields better general satisfaction values of conclusion sentences than both the usual DC-KJ method and the DC-KJ method with all ideas.

By including diversity, which is similar to divergent thinking, and aggregation, which is similar to convergent thinking, in the meeting process, the result of the meeting improved.

3 Evaluation indexes of group intelligence

In the pursuit of electronic meeting systems that can harness group intelligence, the creative problem solving process has been explored. The creative problem solving process for meetings can be divided into two steps: one is idea generation and the other is reaching a conclusion from many ideas. In research on creativity in human intelligence, Guilford proposed two thought styles: divergent thinking to generate many and various ideas and convergent thinking to lead to the answer to a problem [18]. The process of creative problem solving, as refined mainly by Osborn and Persones, emphasizes the importance of interaction between divergent thinking and convergent thinking [19].

Therefore, the ideal meeting will use group intelligence to generate many ideas with divergent thinking, and select the best ideas through evaluation with convergent thinking. From the above description, we define three indexes needed to understand the creative problem solving process in the DC-KJ method: the D-index is the number of divergent ideas, the C-index is the number of convergent ideas, and the M-index is the measure of group intelligence.

Specifically, the M-index is considered an index of measurement of the relationships between divergent thinking and convergent thinking. In this research, a

formula to measure group intelligence (M-index) in the creative problem solving process is defined as the multiplication of the number of divergent ideas (D-index) by the number of convergent ideas (C-index). Therefore, group intelligence reflects both divergent and convergent thinking.

4 Application of evaluation indexes

4.1 Target data

Data resulting from application of the DC-KJ method to investigate the effects of conditions for collective intelligence, Table 1 [16], were used to investigate the proposed three indexes (D-index, C-index, and M-index). The general satisfaction value, explained in section 2.2, of conclusions was assumed as the result of the creative problem solving process. The general satisfaction value of conclusion sentences in the case of the DC-KJ method with selected ideas was significantly superior to those of the DC-KJ method and the DC-KJ method with all ideas. Therefore, that data was chosen to investigate the proposed index to understand creative problem solving.

In the investigation experiments, four groups (Group A, B, C, and D) of three participants executed the DC-KJ method. These four executions were treated as the first step of the DC-KJ method with collective intelligence. Following this, two groups (Group A and B) of three participants followed the DC-KJ method with all the ideas, and the four groups (Group C, D, E, and F) of three participants followed the DC-KJ method with the best selected ideas. The Group E and F participated in only this step. When all steps were completed, the general satisfaction value of each conclusion sentence was calculated after evaluation by six persons. The meeting theme was “How to solve global environmental problems”.

All steps of the DC-KJ method with collective intelligence were executed via groupware named KUSANGI [15, 16], which supports the DC-KJ method with multiple computers connected via the Internet. An example screen showing the DC-KJ method being used with all ideas using the groupware is shown in Figure 1. An example screen showing the DC-KJ method being used with selected ideas using the groupware is shown in Figure 2.

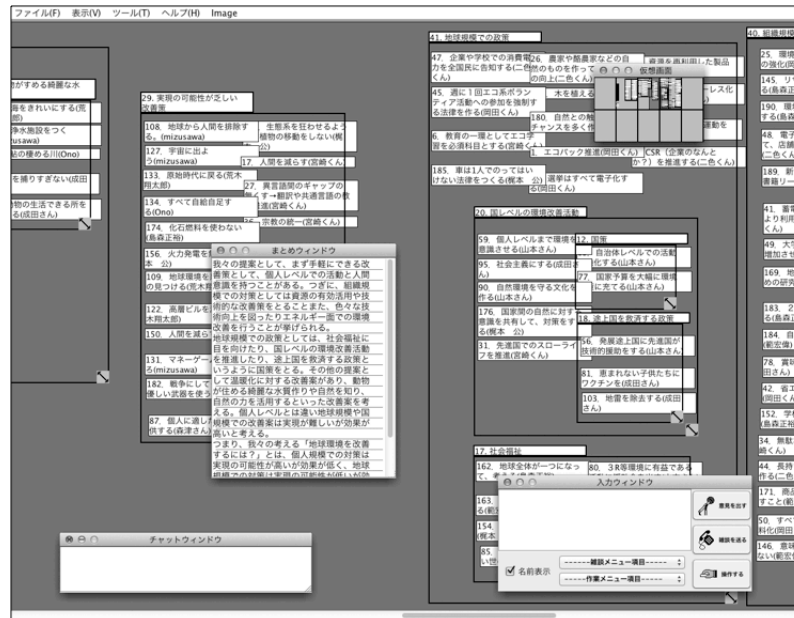


Fig. 1. A screenshot of the DC-KJ method with all ideas.

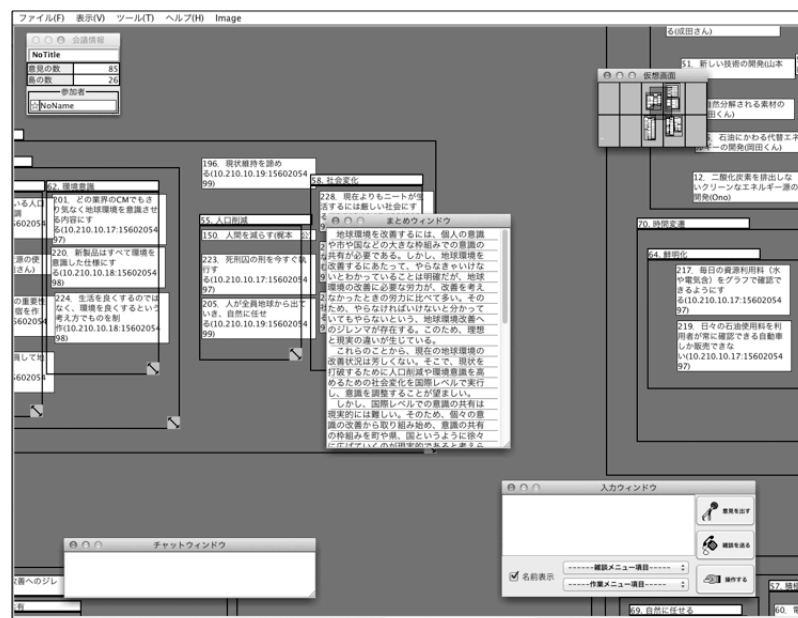


Fig. 2. A screenshot of the DC-KJ method with selected ideas.

Table 1. Results of the DC-KJ method with four conditions of collective intelligence [16].

		General satisfaction value	Number of ideas	Number of groups	Number of characters of conclusion sentences
The DC-KJ method	Group A	1.2	35	6	508
	Group B	1.0	51	3	129
	Group C	1.6	50	20	339
	Group D	0.6	52	12	106
The DC-KJ method with all ideas	Group A	2.6	188	18	402
	Group B	1.6	199	9	172
The DC-KJ method with selected ideas	Group C	5.1	62	26	860
	Group D	3.2	52	4	303
	Group E	4.4	67	5	409
	Group F	5.4	85	26	798

4.2 Word processing

The java library lucene-gosen (<http://code.google.com/p/lucene-gosen/>), which supports morpheme analysis of the Japanese language, was utilized to count the number of words that appeared during the meeting. The version of lucene-gosen was the IPA dictionary version. In this research, the number of kinds of words means the number of different words used.

A function that counts the number of kinds of words in the results (ideas, island names, and conclusion sentences) has been implemented in the groupware KUSANGI, which supports the DC-KJ method and has been developed with the Java programming language. A function to count the number of kinds of words in ideas, or in island names, that reflected in the conclusion sentences was added.

The parts of speech that were counted are nouns, verbs, adjectives, and unknowns. Excluded words are particle words, such as “Ga” and “Wa” in Japanese, auxiliary words, and prefix words. For example, in the idea “Red bicycle named SEKITOBA”, the number of kinds of words is 4: “bicycle” is a noun, “named” is a verb, “Red” is an adjective, and “SEKITOBA” is unknown.

5 Results and discussion

5.1 Results

The results of number of kinds of words in all target data are shown in Table 2. The number of kinds of words is considered as index of variety. The target data groups were: all ideas, island names, conclusion sentences, ideas and island names, and all

data. The final values from all ideas, island names, and ideas and island names provide the D-index.

The DC-KJ method with all ideas utilized the most kinds of words, while the number of kinds of words using the DC-KJ method with selected ideas was lower than that of the DC-KJ method with all ideas but more than that of the usual DC-KJ method.

Table 2. The number of kinds of words in the results of the DC-KJ method with the four conditions of collective intelligence.

		Ideas*	Names of islands*	Conclusion sentences	Ideas and names of islands*	All data
The DC-KJ method	Group A	101	23	93	120	184
	Group B	161	4	38	162	183
	Group C	184	12	54	189	225
	Group D	162	31	27	178	189
The DC-KJ method with all ideas	Group A	463	48	73	485	504
	Group B	484	20	35	492	504
The DC-KJ method with selected ideas	Group C	225	38	117	240	285
	Group D	181	4	59	182	220
	Group E	236	8	89	238	277
	Group F	303	46	113	320	365

*D-index: an index of divergent thinking

Next, the number of kinds of words in ideas or island names reflected in the conclusion sentences is shown in Table 3. In the case of “ideas and names of islands”, the number of kinds of words takes into account the words in the conclusion sentences that are seen in both ideas and island names. This number reflects the convergence of ideas and so can be seen as the C-index. This number is generally high in the case of the DC-KJ method with selected ideas.

The calculated results indicating group intelligence, the M-index, from both data in Table 2 and data in Table 3 are shown in Table 4.

For Group A (the DC-KJ method), group intelligence is calculated as follows. The value of “Ideas” is 2121 (101×21), which is the number that appears in the “Ideas” column in Table 2 multiplied by the number that appears in the “Ideas” column in Table 3. The value of “Ideas and names of islands” is 3480 (120×29), which is the number that appears in the “Ideas & name of islands” column in Table 2 multiplied by the number that appears in the “Ideas & name of islands” column in Table 3. The value of “Max value” of group intelligence is 17112 (184×93), which is the number that appears in the “Conclusion sentences” column in Table 2 multiplied by the number that appears in the “All data” column in Table 2.

In most cases in which the DC-KJ method with collective intelligence was used, group intelligence was larger than when the usual DC-KJ method was used. Group D was an exception, with a small value.

Table 3. The number of kinds of words reflected in the conclusion sentences (C-index).

		Ideas	Names of islands	Ideas and names of islands
The DC-KJ method	Group A	21	12	29
	Group B	17	1	17
	Group C	13	12	18
	Group D	10	14	16
The DC-KJ method with all ideas	Group A	34	46	54
	Group B	22	10	23
The DC-KJ method with selected ideas	Group C	59	36	72
	Group D	20	4	21
	Group E	49	6	50
	Group F	51	45	68

Table 4. Group intelligence (M-index).

		Ideas	Ideas and names of islands	Max value
The DC-KJ method	Group A	2121	3480	17112
	Group B	2737	2754	6954
	Group C	2392	3402	12150
	Group D	1620	2848	5103
The DC-KJ method with all ideas	Group A	15742	26190	36792
	Group B	10648	11316	17640
The DC-KJ method with selected ideas	Group C	13275	17280	33345
	Group D	3620	3822	12980
	Group E	11564	11900	24653
	Group F	15453	21760	41245

5.2 Correlation coefficient with conclusion of meeting

The correlation coefficients between the general satisfaction value of conclusions (Table 1) and the various parameters are shown in Table 5. The various parameters are the basic data from the DC-KJ method: number of ideas, number of islands, and number of characters in conclusion sentences, from Table 1, the D-index from Table 2, the C-index from Table 3, and the M-index from Table 4.

The correlation coefficient with the number of kinds of words in ideas reflected in the conclusion sentences is 0.93, and that with ideas and island names is 0.87. These two values high correlation with the general satisfaction value. Next, the number of characters in conclusion sentences, number of kinds of words in conclusion sentences,

and the maximum size of group intelligence have a high correlation coefficient of around 0.8.

As mentioned above, the number of convergent ideas showed high correlation along with the number of divergent ideas and the magnitude of group intelligence. This result reflects the importance of convergent thinking in the process of creative problem solving.

Table 5. Correlation coefficients between indexes and the general satisfaction value of the DC-KJ method with collective intelligence.

		Correlation coefficient value with general satisfaction value	P value (two-tailed)	
Basic data (Table 1)	Number of ideas	0.00	0.995	
	Number of islands	0.52	0.125	
	Number of characters in conclusion sentences	0.82	0.004	**
The number of kinds of words (Table 2)	Ideas	0.18	0.612	
	Island names	0.34	0.330	
	Conclusion sentences	0.82	0.004	**
	Ideas and island names	0.18	0.610	
	Total	0.25	0.480	
The number of kinds of words reflected in the conclusion sentences (Table 3)	Ideas	0.93	0.000	**
	Island names	0.54	0.104	
	Ideas and island names	0.87	0.001	**
Group intelligence (Table 4)	Ideas	0.74	0.014	*
	Island names	0.65	0.040	*
	Max value	0.80	0.005	**

*P < 0.05, **P < 0.01

5.3 Future research

The results show that the highest correlation with the general satisfaction value was the number of kinds of words, which indicates the number of convergent ideas, reflected in conclusion sentences. This suggests a guideline for the DC-KJ method: good ideas should be reflected in the final conclusion sentences.

To support this guideline with groupware, we plan to implement a function to check whether ideas that were generated in the brainstorming step are reflected in the conclusion sentences. In this research, the word processing is realized using the Java programming language. Therefore, the source code for the word processing will be

embedded into the groupware KUSANGI [15,16], developed with the Java programming language.

The data from the DC-KJ method will further be analyzed to try to understand how certain parameters affect the results. For example, the effect of number of persons will be investigated by comparison of results from one person and those from three persons. In addition, the effect of educational experience will be investigated by comparing the results from undergraduate students with those from graduate students.

6 Conclusion

In pursuit of collaboration technology to build group intelligence into the creative problem solving process, we investigated three indexes to evaluate the DC-KJ method. The indexes were the number of divergent ideas (D-index), the number of convergent ideas (C-index), and multiplication of both numbers as a measure of group intelligence (M-index).

These indexes were investigated by analyzing the results of the DC-KJ method with collective intelligence. In this case, the number of kinds of words in ideas was utilized as the D-index and the number of kinds of words in ideas reflected in conclusion sentences was utilized as the C-index. The result revealed that the index with a high correlation with the result of creative problem solving was the number of kinds of words in ideas reflected in the conclusion sentences. Therefore, we recommend that this number can be utilized as an index to check the state of group intelligence in the creative problem solving process with collaboration technology.

In future work, the groupware will be improved by adding a function to guide group intelligence. Moreover, understanding of indexing of group intelligence will be improved by investigating past data from the DC-KJ method.

Acknowledgement

This research was partially supported by the Japan Society for the Promotion of Science (JSPS) and the Grant-in-Aid for Scientific Research (C) 24500143, 2014.

References

1. Stefik, M., Foster, G., Bobrow, D.G., Kahn, K., Lanning, S., Suchman, L.: Beyond the Chalkboard: Computer Support for Collaboration and Problem Solving in Meetings. *Communications of the ACM*, 30, 1, 32-47 (1987)
2. Ellis, C.A., Gibbs, S.J., Rein, G.L.: GROUPWARE: Some Issues and Experiments. *Communications of the ACM*, 34, 1, 38-58 (1991)
3. Kunifuji, S.: A Survey on Creative Thinking Support Systems and the Issues for Developing Them. *Journal of The Japanese Society for Artificial Intelligence*, 8, 5, 552-559 (Sept. 1993) (In Japanese)
4. Munemori, J., Nagasawa, Y.: GUNGEN: Groupware for a New Idea Generation Support System. *Inf. and Soft. Technology*, 38, 3, 213-220 (1996)

5. Kawakita, J.: Idea Generation Method. Chuokoron-sha, Tokyo (1967) (In Japanese)
6. Kawakita, J.: KJ Method, Chuokoron-sha, Tokyo (1986) (In Japanese)
7. Kawakita, J.: The Original KJ Method, Kawakita Research Institute (1991)
8. Osborn, A.F.: Applied Imagination, Third Revised Edition, Charles Scribner's Son (1963)
9. Wilson, C.: Brainstorming and Beyond, Morgan Kaufmann (2013)
10. Rogers, Y., Shrap, H., Preece, J.: Interaction Design: Beyond Human-Computer Interaction (3rd ed.), Wiley (2011)
11. Harboe, G., Minke, J., Ilea, I., Huang, E.M.: Computer Support for Collaborative Data Analysis: Augmenting Paper Affinity Diagrams, Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work, 1179-1182, ACM (2012)
12. Kelley, T.: The Art of Innovation, Crown Business (2001)
13. Ohiwa, H., Takeda, N., Kawai, K., Shimomi, A.: KJ Editor: a Card-Handling Tool for Creative Work Support, Knowledge-Based Systems, 10, 1, 43-50 (1997)
14. Misue, K., Nitta, K., Sugiyama, K., Koshiba, T., Inder, R.: Enhancing D-ABDUCTOR towards a Diagrammatic User Interface Platform. Proceedings of KES 1998, 359-368 (1998)
15. Yuizono, T., Munemori, J., Shigenobu, T.: Effects of Groupware for and Idea Generation with a Large Collaboration Interface. IPSJ Journal, 49, 7, 2574-2588 (2008) (in Japanese)
16. Yuizono, T., Munemori, J.: Investigation of the Cooperative KJ method for Collective Intelligence with Groupware KUSANAGI for a New Idea Generation. IPSJ Journal, 53, 11, 2635-2648 (2012) (in Japanese)
17. Kawakita, J.: Teamwork. Kobun-sha, Tokyo (1966) (In Japanese)
18. Guilford, J.P.: The Nature of Human Intelligence, McGraw-Hill (1967)
19. Treffinger, D. J., Isaken, S.G, Stead-Dorval, K.B.: Creative Problem Solving - An Introduction (4th ed.), Prufrock Press (2006)
20. Yagishita, K., Munemori, J., Sudo, M.: A Proposal and Application of an Evaluation Method for Sentences of B Type KJ Method Based on Contents and Structures. Journal of Information Processing Society of Japan, 39, 7, 2029-2042 (1998) (in Japanese)
21. Munemori, J., Yagishita, K., Sudo, M.: Evaluation of an Idea Generation Method and Its Supporting Groupware. Proceedings of KES 1999, 54-57 (1999)
22. Surowiecki, J.: The Wisdom Of Crowds, Anchor books (2004)

CARS Context Modelling

Christos Mettouris and George A. Papadopoulos

Department of Computer Science, University of Cyprus, Nicosia, Cyprus
1 University Avenue, P.O. Box 20537, CY-2109
{mettour, george}@cs.ucy.ac.cy

Abstract. Most context-aware recommender systems in the literature that use context modelling have the tendency to develop domain and application specific context models that limit, even eliminate any reuse and sharing capabilities. Developers and researchers in the field struggle to design their own context models without having a good understanding of context and without using any reference models for guidance, often resulting in overspecialized, inefficient or incomplete context models. In this work we build upon prior work to propose an enhanced online context modelling system for Context-Aware Recommender Systems. The system supports CARS developers in the process of building their own context models from scratch, while it supports at the same time sharing and reuse of the models among developers. The system was tested with a real dataset with positive results, as it was able to support context model development with instructions to the developer, model comparison, useful statistics, recommendations of similar models, as well as alternative views of context models to aid the developer's task.

Keywords: Context Modelling System, Context-Aware Recommender Systems, Application Context Model, Context Instance Model, Context variables, Context Dimensions

1 Introduction

A well-known and effective solution to the information overload modern life experiences at all fields is the usage of Recommender Systems (RS). Information overload refers to the vast amount of information users have to access nowadays: users can get lost, disappointed and frustrated for failing to retrieve the desired and needed information at a given time. RS use a variety of filtering techniques and recommendation methods to provide personalized recommendations to their users, mostly by using information retrieved from the user profile, from user's usage history, as well as item related information [1, 2]. However, traditional RS use limited or none contextual information to produce recommendations, as opposed to the Context-Aware Recommender Systems (CARS) that focus in using contextual information to enhance recommendations [3]. Context was first utilized into the recommendation process by Adomavicius by proposing three approaches: the Pre-filtering approach, the Post-filtering approach and the Multidimensional Contextual Modelling approach [3, 4].

Context modelling is important for modelling the contextual information to be used during the recommendation process.

An important contextual modelling issue in CARS is the development of domain specific and application specific context models that only represent information on the particular application domain (e.g. recommendation of movies). Our review on RS [5] had revealed that most CARS and semantic RS in the literature are domain and application specific, meaning that they cannot be applied in other domains. By designing domain and application specific context models, many different and very specific models are produced with no reuse and sharing capabilities.

Another problem is that developers and researchers attempt to design their own models based on their own knowledge and skills and more importantly without using any reference model, without any guidance and strictly focused on the application at hand, often resulting in overspecialized, inefficient or incomplete contextual models.

We had partially addressed the above contextual modelling problem in prior works, at first by proposing a generic, abstracted contextual modelling framework for CARS which developers and researchers can use theoretically to be guided through the process of properly defining the context for their application [5], and later by developing an online “Context Modelling System and Learning Tool” [6] based on [5], which is able to teach and guide developers towards a more efficient, effective and correct selection and usage of context attributes for building their own application model, allowing at the same time for sharing and reuse of context models among applications, regardless of the domain they belong to.

The modelling framework in [5] was essentially a model template in UML built in the Eclipse Modelling Framework (EMF) [7]. Although this framework was developed as a UML class diagram, it was mainly a theoretical tool rather than a modelling tool since: (i) it was not an easy and straight-forward procedure for developers to extend or instantiate a UML class diagram in order to build their own context models, (ii) it was time consuming, (iii) it required programming knowledge and skills and (iv) it did not offer guidance and learning of important concepts. The online “Context Modelling System and Learning Tool” [6] aimed at solving the aforementioned problems by specifying an easy to use UI for the developers and researchers to be able to effectively and efficiently build their models, share them with others, as well as reuse models of others. The above can be accomplished without any programming skills being required by the user. The system focuses also on learning, being able to introduce developers and new researchers with modern concepts from CARS research, as well as their role in a context model and a recommendation process [6].

In this work, we have extended and finalized the work conducted in [6] based on feedback received by experts, aiming to advance the functionality of the system towards a more effective context model development, sharing and reuse. The work was extended with important system functionality and the new *CARS Context Modelling System* [8] was tested by us in real settings by using a dataset released in the framework of the 2nd International Workshop on Information Heterogeneity and Fusion in Recommender Systems - HetRec 2011 [9]. More to the point, we have designed and implemented: (i) the validation of application contexts through context instances, (ii) the comparison of application context models regarding their common context varia-

bles, (iii) the recommendation of similar application context models to the user and (iv) the inclusion of the *context dimensions* concept [3] in the system and the enablement of a “context dimensions” view of the context models.

In sections 5 to 7 we discuss the above important additions, after presenting related work in section 2, an introduction to the system concept in section 3 and a presentation of its functionality in section 4. Section 8 discusses system testing and section 9 closes the paper with conclusions and future work.

2 Related Work

During our research [5] we have reviewed a number of CARS and semantic RS to opine whether the context models used were application/domain specific or generic. While domain specific models focus extensively on a particular domain, generic models do not and focus on being able to facilitate *any* application specific domain. This research revealed that most CARS and semantic recommenders in the literature use domain specific models [5], meaning that they cannot be applied for usage in other domains. A number of generic recommenders also exists that either apply to some generic application area, or can be applied to more than one domain by linking domain specific ontologies to their own data and knowledge pool in order to gain domain-aware knowledge and provide domain-aware functionality. Although some of the semantic and contextual models attempt to be more generic, the majority represent information that either concern a particular application domain (e.g. movies), or a more abstracted domain (such as products in general, web services, e-learning).

To the best of our knowledge no attempts have been made towards developing a context modelling tool that could facilitate the development of truly generic contextual models for CARS and the definition of their contextual entities, so that CARS developers be able to extend/update and reuse them to construct application specific models for their needs. Such a tool would simplify the process of contextual modelling in CARS and enable context uniformity, share and reuse; this is the motivation for this work.

3 System Concept

In our work we have followed the representational view of context [3], [10], meaning that, as in most CARS, the context of an application is defined through a predefined set of observable context attributes of static (not dynamic) structure which does not change significantly over time (as opposed to the interactional view of context where context is not necessarily an observable feature of an interaction [11]). Please note that by static structure we do not mean that the context itself is static, e.g. the context attribute *user location* has a well-defined static structure but is itself a dynamic context since it continuously changes values as the user changes locations. Therefore, we assume that there is a predefined, finite set of contextual attributes in a given CARS

application and that each contextual attribute is defined in our system as a context variable. As an example of the interactional view of context the reader may refer to [11] in which the user is modelled based on human memory models proposed in psychology, where user preference models for previous interactions are stored within the user's long term memory, while the current user's model is stored in the user's short term memory. Then, the short term memory is used to retrieve information from the long term memory in order to be used to generate recommendations for the user.

The CARS Context Modelling System is presented online at [8]. On the right side of the main webpage, an image of the context modelling framework is presented for reference [5], while on the left side various options are presented. The red colour text throughout the system [8] is clickable and provides information on important system concepts that have to do with CARS research, as well as information on how to use the system. The information is provided in pop-up text boxes, as well as on the context modelling framework image on the right side for reference and easier comprehension of the information.

Many complicated concepts related to CARS research are used by the system in order to construct context models [8]. Concepts like "the multidimensional context-aware recommendations: $Users \times Items \times Context \rightarrow Ratings$ ", "context variables", "application contexts" and "context instances" need to be well understood by CARS developers and researchers in order to be able to use the modelling system to create their own application context models. To assist users on this difficult task the text provided within the text boxes is carefully selected from important published research papers, while references are provided as links wherever needed for further reading.

The fundamental concept of CARS research is to include the context in the recommendation process to result from the 2D un-contextual RS: $Users \times Items \rightarrow Ratings$ to the multidimensional CARS: $Users \times Items \times Context \rightarrow Ratings$ [3]. The latter represents a single complete recommendation process and is the main idea behind this CARS Context Modelling System. For each recommendation attempt, a RS must examine whether each item is suitable for a user *in a certain context*. This can be depicted through the question: "*what is the rating a particular user would assign to a particular item under a certain context?*" This rating score is what a recommender must calculate. Therefore, in order to examine whether an item is suitable for a user, the recommendation scheme must have exactly one *user*, exactly one *item* but one or more *context* entities, for each of which a *rating* score can be assigned.

The *context variable* concept contains the actual contextual information to be inserted into a context model [8]. Each context variable has a name and a value to describe both the context parameter and its particular value, e.g. "Temperature" is a context parameter and "high" is its value. Therefore, to define the context: "temperature can be high, medium or low", a total of three context variables will be needed. Via a weight property developers may denote a particular importance for their context variable. The "static" property refers to whether the context variable is static (cannot change dynamically, e.g. user's date of birth) or dynamic (can change, e.g. weather).

An *Application Context* is a context model for a particular RS, e.g. a movies RS. It is built by a CARS developer in order to model the context for this particular RS. An application context model contains all context variables that the developer will select,

along with their values. Since each context variable has a name and a specific value ("Temperature: high"), a developer must select all variables with a particular name for the model to be accurate and complete (e.g. all of the following: "Temperature: high", "Temperature: medium", "Temperature: low").

Although an application context model is built by a CARS developer to model the context for a particular recommender, the system enables sharing and reuse of such models by supporting other developers that want to build similar models in using the same context model and enhance/update it as needed. The idea is: since all RS of a specific type/field (e.g. online movie recommenders) interact in similar context settings, why having one (often incorrect or incomplete) context model for each recommender built by each developer, when we can have just one (correct and complete) context model for all recommenders. We argue that application context models of recommenders of the same or similar fields should be identical, or in any case similar to a great extent. The system suggests to developers to use pre-existing application context models of similar applications (if any) and build upon them, instead of building a model of their own. Eventually, only one application context model for each type of recommender system will exist in the system that should be able to satisfy all developers.

A *Context Instance* is a "screenshot" of the context during an event of interaction between the user and the item that is involved in the recommendation process. For example, for a movie recommender, a context instance is the set of context variables that constitute the context during the event of a particular user (user = "Tom") watching a particular movie (item = "Rambo") at a particular time (we assume the first time that Tom watches Rambo). Such a context instance may have title: "Tom-Rambo C1" (C1 results from "Context1") and may be consisted of the context variables: the time of day, the day of the week, the IMDB ratings of the movie watched, whom did the user watch the movie with, etc. In a similar way, the context instance around the second time Tom watches Rambo will have a title "Tom-Rambo C2" and will again be consisted of a number of context variables. Another definition of the context instance is that it is the set of context variables with their corresponding values that constitute the context at the time a single recommendation is requested. In addition, a context variable may participate in a number of context instances, each of which is characterized by that context variable. For example, the context variable "time: morning" can participate in many context instances, all of which refer to morning time.

Based on the above, the context instances define all valid contextual information around a particular fact or event (in the example above around user Tom watching movie Rambo at a certain time). Ideally, a context instance should be automatically created using context sensing and retrieval during the occurrence of a fact/event and stored in the system. E.g. around the event of a user watching a movie the system should be automatically aware of: the time of day, the day of the week, the IMDB ratings of the movie watched, whom did the user watch the movie with, and any other context variables that could participate in the process. This is a very difficult process, in some way impossible to achieve with the current technology available (how is the system going to know whom did the user watch the movie with, unless the user states it?) and it is beyond the scope of this work. In this CARS Context Modelling System

we provide the ability for CARS developers to create their own context instances for modelling purposes, in order to closely observe and study whether their application contexts are able to "catch" *any* context instance that may occur, and in that way *validate* their models. More on modelling validation follows in Section 5.

4 Building Context Models

In the main page in [8] an input form is provided in order for developers to add context variables. Context variables contain the actual contextual information that developers need to insert into their context model to populate it [5]. The set of all context variables currently available in the system constitute the *Generic Context Model*¹. The application context models, the context instance models and the generic context model constitute the three types of context models supported by the system. The generic context model defines the basic contextual entities of RS, as well as their properties and associations in order for CARS developers to be able to extend it to construct application specific models for the needs of the application at hand. This generic context model simplifies the process of context model development and enables context uniformity, sharing and reuse.

All context models are based on the abstracted contextual modelling framework for CARS (refer to Fig. 2 in [5] and main page in [8]). These models constitute the "context" entity of this framework, as well as all the context related entities and relationships from that level downwards. In the generic context model¹, as in all context models, the four context categories are presented: "Item Context", "User Context", "System Context" and "Other Context" [8]. The rectangles depict the context variables (both name and value) and their parameters. Note that a context variable can belong to more than one context category; such decisions are made by the developer at the time of creation of the context variable or the context model. The "itemContext", "userContext", "systemContext" and "otherContext" constitute the four main context classes in the system and are meant to be perceived as the main context entities for any contextual model of CARS; any context information of any CARS should be able to be represented as a context property of one (or more) of the main context classes, as a context variable.

A simple and straightforward way is developed for CARS developers to build their own context models. They can either add a new context variable that is not currently included in the system, or use the generic context model and simply select/unselect the context variables (already in the system) that are of interest to them by clicking on them. Before adding a new context variable in the system, developers are advised to first check whether the context variable they would like to add already exists in the system as part of the generic context model; if it does, developers are asked to use the

¹ Due to limited space in the paper, context models are not shown in figures; instead, we provide hyperlinks to the models on the online tool in footnotes for reference: <http://www.cs.ucy.ac.cy/~mettour/phd/CARSContextModellingSystem/genericContextModel.php>

existed variable in order to avoid redundant information in the system and confusion to other developers. In this way, the context variables created are universal among many applications and domains and hence they can be shared and reused in many context models of various CARS. For more information on the features and functionality offered by the developed online system, the reader is referred to the system [8].

5 Validating Application Contexts Through Context Instances

The system supports the validation of an application context model through one or more context instance models. As already stated in Section 3, an application context is a context model for a particular recommender application, e.g. a movies recommender system, while a context instance defines the context “screenshot” at the time a single recommendation is requested. It is evident that an application context model should be able to support any context instance related to the particular recommender; otherwise, the application context model is incomplete. E.g. an application context model for a movie recommender should be able to support any movie recommender related context instance, such as “Tom-Rambo C1” (see Section 3) which can be consisted of a number of context variables such as the time of day, the day of the week, the IMDB ratings, etc. This means that the application context model must include all context variables of the context instance model.

The possibility to validate an application context model through context instance models can be a useful tool for the CARS developer who has just created her application context model in the system (let’s suppose an application context model for a movie recommender) and wants to ensure that her application context is able to properly model the context of movie RS (i.e. model any context instance of these recommenders). The context instance can be created preferably by a RS user who will reflect her experiences regarding the activity of watching movies in the context instance model. If this is not feasible, the context instance model can be created by another developer. Following, the CARS developer will be able to access the context instance and by clicking a button, validate this model against her application context model. The system then provides a justified answer whether the context instance model was validated against the application context. If yes, then the application context is also validated. If not, the system provides information as to why the model was not validated. A context instance may not be validated against an application context either because the context instance is incorrect or because the application context is incomplete.

6 Recommendation of Application Context Models and Model Comparison

During the creation of a new application context model by the CARS developer, the system is able to recommend the top N (currently N=5) most similar application con-

text models regarding the percentage of common context variables (refer to Section 8 Fig. 2). This is very important for CARS developers who are in the process of creating their application context model and would like to be informed about similar context models in the system, as well as the level of similarity. As soon as a new application context is created, the system automatically provides recommendations.

Moreover, the system provides an easy way to compare two specific application context models. Through the usage of colours, the system depicts the common context variables of the two application contexts, as well as the context variables that belong only to one of the two application contexts. Besides common context variables, the system also provides statistics regarding the percentage that each application context participates in the other application context. For an example on the above you may refer to Section 8.

If two application context models have many common variables, then the system proposes a merge. It is certainly a situation where the CARS developers must decide whether both application contexts are needed. This is especially interesting in the case where each application context concerns a different type of recommender system, e.g. a movie recommender and a book recommender, as context models of recommenders of different fields are very rarely similar.

7 Including Context Dimensions in the System

The basic concept of CARS research is to include the context in the recommendation process so that to result from the 2D un-contextual recommenders: $Users \times Items \rightarrow Ratings$ to the multidimensional context-aware recommenders: $Users \times Items \times Context \rightarrow Ratings$ [3]. The term *Context* appears to be a single dimension itself, but in essence it represents the *Context Dimensions*, i.e. all the additional context-related dimensions that are being used in the recommendation process besides the user and the item (refer to Fig. 2 in [3] for an example of a three dimensional model for the recommendation space: $User \times Item \times Time$).

In our work, a context dimension can only be a context variable with all of its possible values. We use the following definition of the context dimension: *each individual context variable with a unique name and with all of its values can be perceived as a context dimension*. Therefore a context dimension name is a unique context variable name.

The system provides the “context dimensions view” option for each context model in the system: the generic model, an application context model or a context instance model. The “context dimensions view” is an alternative view of a model, besides the default “context variables view” of the system presented in Section 4. The “context dimensions view” is of particular importance for: i. observing the context dimensions of a model and ii. observing the context variables of a model with all of its values aggregated. Fig. 1 presents the “context dimensions view” of the application context model Movie Recommender (refer to Section 8).

8 System Testing

The CARS Context Modelling System was tested by us in real settings by using a dataset aimed for usage with movie RS which was released in the framework of the 2nd International Workshop on Information Heterogeneity and Fusion in Recommender Systems - HetRec 2011 [9]. The dataset is an extension of MovieLens10M dataset, which contains personal ratings and tags of users about movies. In the dataset, the movies are linked to Internet Movie Database (IMDb) and RottenTomatoes (RT) movie review systems. The dataset includes more than 2000 users, more than 10000 movies, more than 800000 ratings, as well as many related data such as movie genres, directors, actors, countries, locations and tags. It also includes information on user ratings on movies, as well as on user tagging of movies.

For system testing purposes we have played the role of a CARS developer who would like to model the context for building a movie recommender system based on this dataset. Similar datasets are frequently used by RS developers and researchers to develop their systems, both for commercial as well as research purposes. By using this dataset for building a context model in our system we can closely observe how a developer can be assisted in real settings.

We have created an application context model named Movie Recommender in the system based on this dataset and then, by using the available system functionality, we have compared this dataset with other application context models in the system to find similarities, as well as opine whether the context model built from the dataset was adequate to be used in a CARS for movies.

The first step was to (manually) extract the context attributes from the dataset. The context attributes we have identified are: “movie genre”, “movie director”, “movie actor”, “movie country”, “movie location”, “movie tag”, “user (previous) rating” and “user (previous) tagging”. The first 6 context attributes were assigned under the context category “item context”, while the final two under the context category “user context”. Each context attribute is inserted into the system as a context variable in the form: “name: value”. E.g. “movie country: France”. Since the amount of raw data included in the dataset is huge (10000 movies) at this point we have decided to insert only a sample of the data in the system’s database. This is adequate for our “proof of concept” type of testing. The context variables of the created application context model Movie Recommender can be seen via the online system², while Fig. 1 shows the context dimensions of the same model (there are 34 context variables and 8 context dimensions in this model).

As we can observe from the online model² and Fig 1, we were able to extract only 8 context dimensions from the dataset, populating the application context model with a sample of 34 context variables. If we had populated the model with all the information of the dataset, the context variables in the model would be hundreds more; however, the distinct context variable names (and hence the context dimensions) would still be only 8. Another important observation is that the application context

²<http://www.cs.ucy.ac.cy/~mettour/phd/CARSContextModellingSystem/displayAppInstancesModel.php?appCont=Movie%20Recommender>

model² includes information in only two of the four context categories. This is certainly a limitation of the model, since it does not utilize the context entirely.

After creating the application context model, we have been provided with system recommendations of similar application context models already in the system (created earlier by us, by colleagues of ours and by experts on context-aware recommenders). The system provides the top 5 most similar models to our model. The most similar application context model is one of another CARS for movies (named Default Movie Recommender³) with 58% similarity, following other models of other recommenders that are similar to ours in percentages between 8%-2% (Fig. 2). The similarities of the two application context models can be seen via the online system⁴. In red colour the common context variables are depicted. In green are the context variables that belong only to the Movie Recommender context model, while in blue are shown the context variables that are included only in the Default Movie Recommender context model.

It is quite easy to observe that the context model for the Movie Recommender is inferior to the Default Movie Recommender application context model, since it only includes a percentage of the context attributes of the Default recommender in all context categories. In Item Context category there is 65% of common context, in User Context category there is 62% of commons context, while in the System Context and Other Context categories the common context is 0%.

Switching to the “context dimensions view” (on the online system), we observe that the Movie Recommender context model includes the context dimension “movie location” consisted of 4 context variables which is not included in the Default Movie Recommender system. This is a deficiency of the Default model. However, there is a number of context dimensions that were included in the Default model but not in the Movie Recommender application context model: movie title, imdb ratings (the imdb ratings of a movie), movie duration, company (with whom the user watches the movie with), user location (GPS), network (describes the network connectivity/bandwidth) and time (time of day of watching the movie). We can observe that, while our context model includes only 8 context dimensions, the Default model includes 14.

There are 8 Context Dimensions:

movie genre	movie tag	movie country	movie director	movie actor	movie location	user previous rating
values	values	values	values	values	values	values
comedy	earth	USA	John Lasseter	Philip Proctor	USA	1
drama	police	Australia	Oliver Parker	Darryl Henriques	Italy	2
action	boxing	Canada	Siddharth Randeria	Rick Garcia	Canada	3
romantic	almodovar	France	Chris Noonan		Uganda	4
	futurish					5
	time travel					
	excellent characters					

user previous tagging
values
earth
police
boxing

Fig. 1. Movie Recommender Application Context Model: “Context Dimensions View”

³<http://www.cs.ucy.ac.cy/~mettour/phd/CARSContextModellingSystem/displayAppInstancesModel.php?appCont=Default%20Movie%20Recommender>

⁴[http://www.cs.ucy.ac.cy/~mettour/phd/CARSContextModellingSystem/compareApplicationContexts2.php?sentData=\\$Default%20Movie%20Recommender\\$Movie%20Recommender](http://www.cs.ucy.ac.cy/~mettour/phd/CARSContextModellingSystem/compareApplicationContexts2.php?sentData=$Default%20Movie%20Recommender$Movie%20Recommender)

There are existing Application Contexts similar to **Movie Recommender**. The 5 most similar are:

- Comparing **Movie Recommender** with **Default Movie Recommender**: **58% similarity** (30 Common Vars out of 52) [see details](#)
 The Application Context **Movie Recommender** participates in the Application Context **Default Movie Recommender** with 30 out of 34 context variables (88%).
 The Application Context **Default Movie Recommender** participates in the Application Context **Movie Recommender** with 30 out of 48 context variables (63%).

- Comparing **Movie Recommender** with **Car recommender**: **8% similarity** (4 Common Vars out of 51) [see details](#)
 The Application Context **Movie Recommender** participates in the Application Context **Car recommender** with 4 out of 34 context variables (12%).
 The Application Context **Car recommender** participates in the Application Context **Movie Recommender** with 4 out of 21 context variables (19%).

- Comparing **Movie Recommender** with **Shopping recommender**: **7% similarity** (4 Common Vars out of 57) [see details](#)
 The Application Context **Movie Recommender** participates in the Application Context **Shopping recommender** with 4 out of 34 context variables (12%).
 The Application Context **Shopping recommender** participates in the Application Context **Movie Recommender** with 4 out of 27 context variables (15%).

- Comparing **Movie Recommender** with **Book recommender**: **5% similarity** (3 Common Vars out of 61) [see details](#)
 The Application Context **Movie Recommender** participates in the Application Context **Book recommender** with 3 out of 34 context variables (9%).
 The Application Context **Book recommender** participates in the Application Context **Movie Recommender** with 3 out of 30 context variables (10%).

- Comparing **Movie Recommender** with **Colloquium Room**: **2% similarity** (1 Common Vars out of 48) [see details](#)
 The Application Context **Movie Recommender** participates in the Application Context **Colloquium Room** with 1 out of 34 context variables (3%).
 The Application Context **Colloquium Room** participates in the Application Context **Movie Recommender** with 1 out of 15 context variables (7%).

Fig. 2. Top 5 Recommendations of Similar Application Context Models

9 Conclusions and Future Work

The purpose of the CARS Context Modelling System presented in this work is to serve as a tool for CARS developers (and researchers), enabling them to efficiently, effectively and correctly select and use context attributes for building their own application models, allowing at the same time for sharing and reuse of context models and information among applications, regardless of the domain they belong to. We have presented the system concept and its functionality, as well as discussed additional important features such as the validation of application contexts through context instances, the comparison of application context models, the recommendation of similar application context models to the user and the inclusion of the context dimensions concept in the system (via the “context dimensions view” of context models).

The system testing we have conducted in real settings by using a real dataset [9] showed that the system was able to depict important limitations of the application context model created based on the dataset, in comparison with other pre-existing CARS application context models in the system (built by us, colleagues and experts).

Based on the above, we conclude that if CARS developers were to define the context for a movie RS based on this dataset (which is a valid dataset for movie RS used by many developers) and without using the CARS Context Modelling System, the result would be an inefficient and incomplete context model. Instead, by using the CARS Context Modelling System we were able to build our application context model with the aid of system tools, compare our model with other models, as well as be recommended with similar models that could assist us in developing and enhancing our application context model.

Currently, the system provides all context models as visual images and will support the extraction of the models in xml and txt formats. As future work, we aim to support CARS developers in incorporating the context models within their RS by providing

appropriate tools. An idea is to extent the system by providing automatic transformation of the context models to code. The idea is to support CARS developers in attaching their context models to their recommendation methods via an easy and straightforward way that will also support context model sharing and reuse.

References

1. Deshpande, M., Karypis, G.: Item-based top-n recommendation algorithms. *ACM TRANSACTIONS ON INFORMATION SYSTEMS*, vol. 22 (2004) 143–177
2. Karypis, G.: Evaluation of Item-Based Top-N Recommendation Algorithms. *Proceedings of the tenth international conference on Information and knowledge management* (2000) 247–254
3. Adomavicius, G., Tuzhilin, A.: Context-aware recommender systems. In *Recommender Systems Handbook*, F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor (2011) 217–253
4. Adomavicius, G., Sankaranarayanan, R., Sen, S., Tuzhilin, A.: Incorporating contextual information in recommender systems using a multidimensional approach. *ACM Transactions on Information Systems (TOIS)*, vol. 23 (2005) 103–145
5. Mettouris, C., Papadopoulos, G. A.: Contextual Modelling in Context-Aware Recommender Systems: a generic approach. In: Haller, A., Huang, G., Huang, Z., Paik, H.-Y., Sheng, Q.Z. (eds.) *WISE 2011 and 2012 Combined Workshops. LNCS*, vol. 7652 Springer, Heidelberg (2013) 41–52
6. Mettouris, C., Achilleos, A. P., Papadopoulos, G. A.: A Context Modelling System and Learning Tool for Context-Aware Recommender Systems. In *Scaling up Learning for Sustained Impact*, D. Hernandez-Leo, T. Ley, R. Klamma, and A. Harrer, Eds. *LNCS*, Springer Berlin Heidelberg, Volume 8095 (2013) 619–620
7. Eclipse Modeling Framework Project (EMF) Online: <http://www.eclipse.org/modeling/emf/>
8. “CARS Context Modelling System”. Online: <http://www.cs.ucy.ac.cy/~mettour/phd/CARSContextModellingSystem/>
9. hetrec2011-movielens-2k. Dataset released in the framework of the 2nd International Workshop on Information Heterogeneity and Fusion in Recommender Systems (HetRec 2011) at the 5th ACM Conference on Recommender Systems (RecSys 2011). Online: <http://ir.ii.uam.es/hetrec2011/datasets.html>.
10. Dourish, P.: What we talk about when we talk about context. *Personal and ubiquitous computing* (2004) 8(1):19–30
11. Anand, S. S., Mobasher, B.: Contextual recommendation. *WebMine, LNAI* (2007) 4737:142–160

Computational Creativity for Intelligence Analysis

Robert Forsgren, Peter Hammar and Magnus Jändel

Swedish Defence Research Agency
164 90 Stockholm, Sweden
{robert.forsgren, peter.hammar, magnus.jaendel}@foi.se

Abstract. We describe a decision support system for hypothesis assessment in which exploration is supported by computational creativity. A software tool for morphological hypothesis analysis and evidence handling is extended with a creative assistant that on demand suggests hypotheses that the analyst should consider. Suggested hypotheses are chosen so that they are far from hypotheses that the analyst previously has paid attention to but nevertheless are supported by evidence in an interesting way. For the purpose of providing thought-provoking suggestions, the creative assistant employs ensembles of novelty and value assessment methods and proposes hypotheses that stand out in this multi-ensemble analysis. Preliminary experiments investigate the system's potential for infusing novel and valid ideas into the decision making process.

Keywords: Decision support system, hypothesis assessment, situation assessment, computational creativity

1 Introduction

Situation assessment is the core of intelligence analysis. Once the analyst understands what really is going on it is comparatively easy to come up with reasonable actions and evaluate costs, risks and likely outcomes. Classical intelligence analysis means that the space of all relevant hypotheses is defined and that the analyst relates each piece of evidence to each hypothesis as for example in the *Analysis of Competing Hypotheses* (ACH) method [1]. Realizing that such classical methods are impractical in face of the profusion of hypotheses and evidence provided by present day information handling systems, Gustavi et al. [2] pioneered that evidence should be connected to hypothesis attributes rather than to each hypothesis per se thus greatly reducing the number of evidence links and hence simplifying the analysis process.

The *Multi-Hypothesis Management and Analysis* (MHMA) method of Gustavi et al. [2] extends legacy *Morphological Analysis* tools in use by intelligence analysts. Morphological analysis is a three-phase qualitative analysis methodology in which a first phase maps out the hypothesis space, a second phase identifies consistency constraints and a third phase applies judgment to analyse and select preferred hypotheses. The method was pioneered by the astrophysicist and polymath Fritz Zwicky [3] and has been applied in diverse fields including future studies, policy analysis, law and

adfa, p. 1, 2011.

© Springer-Verlag Berlin Heidelberg 2011

technology. There is a vast literature on morphological analysis¹ for decision support that we will not attempt to review here (for a recent monograph see [4]).

Situation assessment is, however, not just about mechanically evaluating how the weight of pre-existing evidence is distributed over some hypothesis space. Based on what hypotheses that are found to be interesting, analysts will look for more evidence and re-evaluate existing evidence thereby following what essentially is a scientific research methodology. The creativity of this process is as important as the formal representations and information processing tools. Human creativity is, however, a fickle resource at best and succumbs easily to group-think and prejudice. In this paper we investigate how computational creativity can contribute to the analysis process.

Computational creativity is a new and burgeoning branch of computer science. For a recent review see [5]. In [6] we outline a research program for computational creativity in decision making where six different options for incorporating computational creativity in decision processes are described. One of these options pertains to using computational creativity for situation assessment i.e. coming up with the main hypothesis about what kind of situation that is at hand. The present paper describes a system and initial experiments that address this issue and provides hence a first step towards realizing the outlined research program. [6] provides also a brief review of the literature on computational creativity in decision support. As a harbinger of nascent interest in the decision analysis community, note the recent article [7].

In this paper we take a first step towards computational creativity for supporting hypothesis assessment in intelligence analysis by exploring the design space and design principles as well as experimentally comparing the simplest possible implementation with a few selected more complex designs. Section 2 introduces the intelligence analysis methods from the user's point of view and describes how computational creativity is integrated in the user interface. Section 3 defines and motivates the computational creativity algorithms whereas section 4 describes the implemented system that is explored experimentally in section 5.

2 The User Perspective

We introduce evidence supported morphological analysis according to [2] by an example using the same scenario as used in our experiments in section 4. The scenario is about the 2001 anthrax attacks [8] in which, shortly after the 9/11 events, letters containing lethal anthrax spores were mailed killing several infected victims. Both al-Qaeda and Iraq were suspected but B.E. Ivins, a U.S. biodefense scientist was declared to be the sole perpetrator although this conclusion is disputed [8]. Suppose now that we want to make a situation assessment in this scenario.

The hypothesis space is set up by morphological analysis. We define the *parameters* or conceptual dimensions of the problem and a domain of discrete *values* for each parameter. To facilitate presentation in readable figures and tables we use a small

¹ Note that “morphological analysis” has different meanings in decision support and linguistics. In this paper we use the term only in the former meaning.

morphological chart with only three parameters and a handful of values per parameter. *Culprit*, *Motive* and *Source* (of the anthrax spores) are chosen as parameters. The set of values for each parameter is shown in Fig. 1. This morphological chart was developed in [2] where a detailed discussion of how it connects to the scenario is provided. A hypothesis is formed by picking a specific value (cell) from each of the parameters (columns), the hypothesis space being the set of all such combinations.

Culprit	Motive	Source
Al-Qaeda	9/11 inspired	Gov. special lab
Iraq	Domestic politics	Simpler lab U.S.
B.E. Ivins	Personal gain	Simpler lab Afg.
S.J. Hatfill	Target recipients	Iraqi special lab
People from U.S. gov.		

Fig. 1. Morphological chart with evidence weight represented by shade of colour. The cells with darker colour have twice as much supporting evidence as the lightly coloured cells. White cells are not related to any evidence.

Using the MHMA tool, analysts enter evidence and connect evidence as having a positive, neutral or negative impact on each value cell. Hypotheses are scored based on the aggregated evidence weight on all the values that compose the hypothesis. The essential simplification of MHMA compared to ACH is that evidence is linked to cells rather than hypotheses thus avoiding the combinatorial explosion of evidence links in a large hypothesis space. To illustrate the method of coupling evidence, we have in Fig. 1 entered two pieces of evidence supporting each of the value cells *al-Qaeda*, *9/11 inspired* and *Simpler Afghan lab*. One piece of evidence supports each of the cells *B.E. Ivins*, *Target recipients* and *Gov. special lab*.

Assume now that a forensics team is strongly biased in favour of the *al-Qaeda* hypothesis maybe to the degree that it is considered disloyal to explore any alternatives to the current orthodoxy. A good computational creative assistant could help overcoming the ingrained bias of the group by proposing that analysts should look into the Ivins hypothesis. Given that the creative assistant knows that only the *al-Qaeda* hypothesis has been considered by the analysts, it would look for other hypotheses that both has evidential support and are different from most from the well-known hypotheses. The Ivins hypothesis fulfils both these criteria although some variations of the dominating hypothesis actually are better supported by the evidence.

The user interface of our tool consists of four fields as shown in Fig. 2. The *morphological chart* shows the hypothesis space. Analysts can enter evidence in the *evidence list* and connect each piece of evidence by positive or negative links to selected value cells of the morphological chart. The *hypotheses and evidence list* collates the hypotheses according to evidential support and allows the analyst to sort hypotheses according to evidence weight. By selecting a specific hypothesis the analyst can explore all evidence related to the hypothesis. The fourth field is the only new user interface component in our *Computational Creativity Supported Multi-Hypothesis Management and Analysis* tool (CC-MHMA). If the user clicks on the light bulb icon, the system suggests a creative hypothesis.

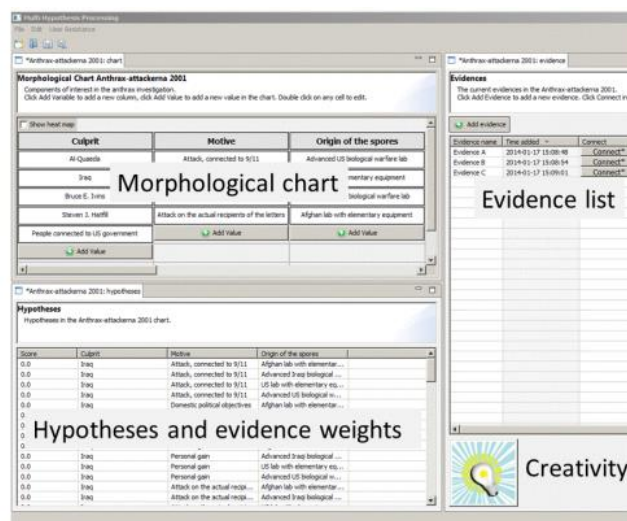


Fig. 2. Annotated user interface of the CC-MHMA tool consisting of the MHMA interface enhanced with the light bulb icon. Legacy MHMA features are the morphological chart, the hypothesis and evidence weight list and the evidence list.

The purpose of the computational creativity function is to provide useful creative input to the analysts. Creative means in this context that the suggestions should be novel from the point of view of the analysts and germane to human goals and the situation at hand. The creativity of a suggestion can hence only be evaluated by human experts. One of the main problems is to provide the computational creativity agent with sufficient knowledge for producing novel and valuable suggestions. Novelty requires insight into what the analyst has considered before and might be aware of. Value calls for knowledge about the situation, the objectives of the analyst and general background information about the application domain. It is often much more difficult to compile the required knowledge in a computer readable form than to come up with suitable processing algorithms. The main hurdle for practical applications of computational creativity is that users cannot be bothered with extra work or cumbersome sensors in order to provide a computational creativity agent with the necessary knowledge. Therefore, we have in this work designed the system to harvest all inputs to the computational creativity from the pre-existing user interface and to include the light bulb as the single minimalistic and unintrusive extension of the user interface. The underlying computational creativity algorithm strives to suggest a hypothesis that has not yet been considered by the analyst and that is supported by the evidence in an interesting way. The next section will explain exactly what this means and how it is done. From the user perspective, the suggestion is shown by a colour-coded selection of values in the morphological chart, highlighting of the proposed hypothesis in the hypothesis list and highlighting of the evidence that impacts on the proposed hypothesis. The data provided by the tool to the computational creativity agent is the morphological chart, the evidence, the links between evidence and parameter values and the history of hypotheses selected for examination by the user. Note that any hypothesis

that is suggested by the computational creativity agent is put into this list of known hypotheses to the effect that the computational creativity agent will not suggest something similar again.

3 The Algorithms

The first subsection describes the formal representation of the computational creativity agent while the following subsections flesh out the details. We discuss a broad range of possible algorithms leaving the definition of what has been explored in our preliminary tests to the next section.

3.1 Hypothesis Space and Formal Creativity Model

The hypothesis space H is the set of all hypotheses combined with a distance measure. The distance measure $d(h_1, h_2)$ is a real-valued function of the two hypotheses h_1 and h_2 . The distance matrix D is formed by indexing the hypotheses according to an arbitrary order and defining matrix elements according to $D_{ij} = d(h_i, h_j)$.

A variable that is associated with a specific hypothesis and furthermore is defined for all hypotheses in the morphological chart forms a scalar field in hypothesis space. Any such scalar field will in the following be called a *charge field* and be denoted by a bold face symbol.

The next hypothesis to be recommended by the creative assistant is computed from by some multi-objective optimization algorithm Ω ,

$$h_c = \Omega(\mathbf{n}, \mathbf{v}), \quad (1)$$

where \mathbf{n} and \mathbf{v} are the novelty and value² charge fields respectively. Multi-objective optimization algorithms (see [9] for a review) typically includes (1) finding the set of (Pareto) efficient hypotheses such that no other hypothesis have better novelty charge without having a worse value charge or vice versa and (2) selecting a solution among this set.

The *novelty charge* \mathbf{n} of a hypothesis is a predictor for how novel the hypothesis appears to the user. It depends on how aware the analyst is of the hypothesis, the distance of the hypothesis to, and the level of awareness for, other hypotheses that the analyst is aware of. Formally we express this as,

$$\mathbf{n} = \Psi(\mathbf{a}, D), \quad (2)$$

in which \mathbf{a} is the awareness charge field and D is the distance matrix. The function Ψ should be designed to allocate high novelty to hypotheses with low awareness charge

² We suffer from a terminology collision regarding the term *value* between the fields of morphological analysis and computational creativity. Where the distinction is not obvious from the context a *cell in the morphological chart* is referred to as *value cell* whereas *usefulness* is called *value charge*.

that are far from any hypotheses with high awareness charge. The awareness charge represents how conscious the user is of the hypothesis.

The value charge \mathbf{v} of a hypothesis is a predictor for how well the hypothesis describes the real-world situation that the analyst is interested in. Formally, the value charge field depends on the evidence E according to some function,

$$\mathbf{v} = \Xi(E) . \quad (3)$$

Both the novelty charge and the value charge are real-valued variables in the domain $[0, 1]$. Ascending numerical value means rising novelty and value respectively.

3.2 Hypothesis Distance Measures

If the creative assistant has no insight in the semantics of the parameters and values of the morphological chart, the most obvious choice of distance measure $d(h_1, h_2)$ seems to be the Hamming distance [10] i.e. count the number of value cell substitutions that is required for transforming one of the hypotheses to the other and use the result as our distance measure.

Many possible enhancements of the distance measure depend on that the user can be coaxed to input more information. An example would be that some parameters in the table could be given more importance than others, creating a weighted Hamming distance where value substitutions would contribute to the distance in proportion to the weight of the parameter. In certain domains values could be ordinal, creating additional structure for a distance measure. Furthermore, we could ask the user to express the relations between the values in the domain of a given parameter as a graph. The contribution to the distance from each parameter could be counted as the number of edges that connect the value cells in each hypothesis. Also such a graph could be enhanced by weighing the edges. Other options for improving the distance measure with additional user input can also be envisaged.

3.3 Modelling Value Charge

We model the value charge of a hypothesis with a sigmoid form,

$$v_h = (1 + \exp(-\sum_{i=1}^{N_e} w_{ih}))^{-1} , \quad (4)$$

where the evidence weight w_{ih} describes how the piece of evidence with index i supports or contests the hypothesis h whereas N_e is the number of evidences. We take w_{ih} to be real-valued with positive values denoting degrees of support.

The sigmoid in (4) is symmetric, and saturates as the amount of (positive or negative) evidence increases, i.e. for a large collection of evidence the value v_h will change very little by the addition of one more piece of evidence. This captures the assumption that the belief (or lack thereof) in a hypothesis that is already well supported (or refuted)

ed) by a massive amount of evidence will not change much if another piece of evidence is added.

The MHMA software invites the analyst to define a negative, neutral or positive relation between each value cell in the morphological chart and each piece of evidence. By this device, MHMA implicitly connects each piece of evidence to many hypotheses in one fell swoop. To simplify, we will in the following only consider the resultant connections between evidence and hypotheses. The user input describing the relation between hypothesis h and the evidence with index i , is henceforth called the evidence impact factor and is denoted e_{ih} . In general, e_{ih} is real-valued with positive values representing levels of support.

The evidence weighting model relates the evidence weights to the evidence impact factors,

$$w_{ih} = \begin{cases} \gamma e_{ih} & \text{if } e_{ih} \geq 0 \\ \gamma \alpha_- e_{ih} & \text{if } e_{ih} < 0 \end{cases}, \quad (5)$$

in which γ is used for defining how much evidence that is needed for near certainty. The sigmoid function in (4) saturates when the absolute value of the sum of weights is about three which means that γ should be set so that this happens when the analyst would judge that the total weight of evidence is such that any further corroboration makes little difference for the conclusion. The parameter α_- is used for controlling the importance of negative evidence. The MHMA tool has primed the analyst to assume that $\alpha_- = 1$. The creative assistant could for example explore $\alpha_- = \infty$ according to which hypotheses that are disfavoured by any negative evidence have zero value or $\alpha_- = 0$ where the impact of negative evidence is disregarded. The creative assistant can furthermore use different perspectives on the relation between the evidence and the hypotheses by using several different evidence weighting models each represented by a choice of values of γ and α_- .

3.4 Capturing Awareness

To provide novelty, the computational creativity agent needs to know what hypotheses that the analyst already have considered. User interface actions, the timing of user interface actions and biometry are the three main means for learning about this. We represent the analyst's awareness of hypothesis h by a single real-valued variable a_h with values in the domain $[0, 1]$ where $a_h = 0$ means no knowledge of h and $a_h = 1$ means maximum awareness of h .

Relevant user interface actions include selecting a hypothesis, selecting a group of hypotheses and examining evidence related to selected hypotheses. For each such action A and each hypothesis h that is associated with the action, the system should update the awareness charge according to some function $a_h = F(A, h, a_h)$ which operates on the present value of a_h and outputs an updated value of a_h . We can for example initialize all awareness charges to zero, allocate a baseline charge $a_h = 0.5$ the first

time h is selected by the user and further increase a_h if the user scrutinizes related evidence or selects the hypothesis again.

The timing of user interface actions related to a hypothesis could also be taken into account which means that a_h is updated at regular time intervals $\{t_1, t_2, \dots\}$ by an updating function that incorporates the history of user interactions according to,

$$a_h(t_n) = F(\mathbf{A}, h, a_h(t_{n-1}), t_n), \quad (6)$$

where t_n is the current time, t_{n-1} is the time of the previous update and $\mathbf{A} = \{A(t_1), A(t_2), \dots\}$ is the history of user actions. In (6) we could for example model the forgetfulness of analysts by letting a_h decay over time according to $a_h(t_n) = c(t_n - t_h)^{-\mu}$ where $c, t_h < t_n$ and μ are parameters. Psychological research indicates power laws for forgetting [11].

Timing analysis of user interactions is complicated by the lack of knowledge about what the user is doing while a given hypothesis is selected. The analyst may be vigorously ruminating over the hypothesis or alternatively be on a coffee break. Biometric methods, including for example video analysis or eye tracking, could provide crucial information about user behaviour to be encoded as special types of actions in (6).

3.5 Modelling Novelty Charge

Although there are many possible algorithms for computing the novelty charge field according to (2), we shall presently only define a simple baseline method and a somewhat more generic parameterized model.

According to the baseline method, the novelty charge of a hypothesis is zero if the awareness charge of the hypothesis is positive i.e. the user is considered to be aware of it. Otherwise, the novelty charge is proportional to the distance to the closest other hypothesis with positive awareness charge. Finally, we normalize so that novelty charges fall in the domain $[0, 1]$.

The parameterized model is based on computing a *familiarity potential*,

$$\varphi(h) = \sum_{h' \in H} \frac{a_{h'}}{d^\kappa(h, h')}, \quad (7)$$

where the sum is taken over all hypotheses and κ is a positive integer. To handle the self-potential issue, we define $d(h, h) = \delta$ where δ is another parameter. In this equation, awareness charge is similar to electric charge and φ is analogous to electric potential. Consequently, the familiarity potential decreases as the distance to known hypotheses increases. The novelty charge is a sign-reversed normalized version of φ according to,

$$n_h = \frac{\varphi_{\max} - \varphi(h)}{\varphi_{\max} - \varphi_{\min}} \quad (8)$$

in which φ_{\max} and φ_{\min} are the maximum and minimum values of the familiarity potential respectively.

3.6 Putting It All Together: Creative Suggestions

The main assumption driving the design of our first prototype system is that users may be able to second-guess deterministic algorithms and will then feel that the creative clout wanes as they gain increasing experience of the tool. To avoid this, the creative assistant should choose randomly from an ensemble of different algorithms. In future experiments we intend to test if this assumption is valid. A generic creative suggestion method consists of, (A) a distance measure $d_i(h_1, h_2)$, (B) an algorithm for computing the novelty charge field Ψ_j (see (2)), (C) an algorithm for computing the value charge field Ξ_k (see (3)), (D) a multi-objective optimization algorithm $\Omega_m(\mathbf{n}, \mathbf{v})$ (see (1)). Each of these are selected from a corresponding ensemble of distance measures, novelty or value charge fields or optimization algorithms, respectively. The method for capturing the user awareness is considered to be fixed. At each new round of suggestion production, the creative assistant will select a creative suggestion method comprised of a randomly collection of components from these four aspects.

4 The Implementation

This section describes the experimental setups used in our initial explorative experiments. The computational creativity agent is implemented as an extension to the MHMA tool described in [2].

4.1 Baseline Implementation

The baseline implementation of the creative assistant is intended to investigate computational creativity in the simplest possible setting thereby providing a reference point for more complex implementations. The user awareness charge is initiated to zero for all hypotheses and is changed to one the first time that the user selects a hypothesis for examination. There is no time decay, and no other relations between awareness charge and user actions. Hence there are two distinct set of hypotheses: known hypotheses with $a_h = 1$ and unknown hypothesis with $a_h = 0$. This simplistic model of the user state was selected because it only uses information that is available in the MHMA tool.

The Hamming distance is the only distance measure employed by the baseline implementation. There is just one novelty charge algorithm according to which the novelty charge (n_h) of a known hypothesis is zero and the novelty charge of an unknown hypothesis is proportional to the distance to the closest known hypothesis. Furthermore, we employ one single value charge algorithm according to (4) and (5) with $\gamma = 1$ and $\alpha_- = 1$.

The multi-objective optimization algorithm applies the utility function,

$$f(v_h, n_h) = \beta \frac{n_h - \hat{n}}{\sigma_n} + (1 - \beta) \frac{v_h - \hat{v}}{\sigma_v}, \quad (9)$$

for selecting the hypothesis to suggest. In (9), \hat{n} and σ_n are the average and standard deviation of the novelty charge with the corresponding notation for the value charge. Only unknown hypotheses are considered as candidates for selection and for calculating the averages and standard deviations. The parameter β balances the influences of novelty and value and is in our initial experiments ad hoc selected to 0.75. Since the optimum of this particular utility function always is Pareto efficient there is no need to explicitly compute the efficient set.

5 The Experiments

Experiments are performed on the morphological chart presented in the introduction. Three models are tested, representing a baseline implementation, an enhanced value model and an enhanced novelty model.

5.1 Testing the Baseline Implementation

The experiments use a collection of evidence comprising 35 pieces of information as further described in [2] and summarized in Fig. 3. Note that this set of evidence has been selected only for the purpose of testing the decision support tool and is not claimed to accurately represent the factual circumstances. A brief look at Fig. 3 suggest that there are much evidential support of Ivins being involved as well as domestic politics and the 9/11 events as motives. There is more evidence rejecting than supporting an Afghan lab as the source of the anthrax spores, whereas an Iraqi lab have an equal amount of positive and negative evidence.

Culprit	Motive	Source
Al-Qaeda	9/11 inspired	Gov. special lab
Iraq	Domestic politics	Simpler lab U.S.
B.E. Ivins	Personal gain	Simpler lab Afg.
S.J. Hatfill	Target recipients	Iraqi special lab
People from U.S. gov.		

Fig. 3. Morphological chart showing the total weight of the evidence considered in our experiments. Positive evidence weight is indicated by solid green shading with darker tone representing more evidential support. Negative evidence weight is indicated by striped cells with the level of red tone indicating the absolute value of the evidence weight.

We will first consider a situation in which the user just keeps pressing the light bulb icon and does not select any hypotheses other than those suggested by the creative assistant. This is not the normal mode of usage but serves to illustrate how the creative assistant works. The first ten hypotheses suggested by the creative assistant

are shown in Table 1. We use abbreviated versions of the value cell names in the tables.

The first suggestion in Table 1 is the hypothesis with highest total value charge. For the next three suggestions, the assistant selects the hypothesis with the highest value charge that does not include any value cells from hypotheses already selected (i.e. they are maximally novel). The fifth hypothesis to be suggested is the one with highest evidence support but not sharing more than one value cell with any of the already known hypotheses.

Table 1. Hypotheses in the order suggested by the baseline creative assistant assuming that the user is aware only of previously suggested hypotheses. The Evidence column shows the number of evidences supporting and refuting the hypothesis as well as the summed evidence weight.

Culprit	Motive	Source	Evidence
Ivins	Domestic	Simple U.S.	+21, -7 = +14
U.S. gov.	9/11	Gov. special	+11, -2 = +9
Al-Qaeda	Recipients	Iraqi special	+9, -3 = +6
Hatfill	Personal	Simple Afg.	+5, -1 = +4
Ivins	Recipients	Gov. special	+18, -7 = +11
Hatfill	9/11	Simple U.S.	+12, -2 = +10
Hatfill	Domestic	Gov. special	+12, -2 = +10
Ivins	9/11	Iraqi special	+18, -8 = +10
U.S. gov.	Recipients	Simple U.S.	+11, -2 = +9
U.S. gov.	Domestic	Iraqi special	+11, -3 = +8

Note that the wide scope of the hypothesis space combined with the simple MHMA evidence handling model occasionally give high value to intuitively rather unlikely hypotheses such as al-Qaeda targeting the individual mail recipients or Ivins using spores from an Iraqi lab. The user could, however, use even unrealistic suggestions as creative stepping stones rather than literally as candidates for the most likely solution. Although Ivins may not have had access to spores from Iraqi labs, investigators could be inspired to consider the possibility of a U.S. perpetrator producing a strain that appears to come from an Iraqi lab.

As we keep generating suggestions, the creative assistant eventually runs out of creativity at a point when all not known hypotheses have the same distance to the closest known hypothesis and thus the same novelty. The value charge is then the only decisive factor in (9) which makes any further suggestions trivial since the analysts have other tools for sorting hypotheses according to evidential support. Note that this effect is a consequence of using the Hamming distance as novelty measure and that analysts in practice will reach this state only in scenarios with quite small hypothesis spaces.

The fourth suggestion in Table 1 has a rather low value; only three out of 80 hypotheses have a lower value. Its high position in the list is caused by the combination of using the coarse-grained Hamming distance as novelty measure combined with the strong bias towards generating novel hypotheses engendered by the choice of $\beta = 0.75$

in (9). We tried setting $\beta = 0.25$ with the result that the fourth suggestion in Table 1 was removed from the top ten suggestions while all other hypotheses remained in the same order with a new hypothesis $\{al-Qaeda, Personal gain, Simpler lab U.S.\}$ with evidence sum +8 appearing at the bottom of the list.

5.2 Enhanced Value Model

The sigmoid function in (4) saturates for most of the hypotheses in the baseline implementation. This means that differences in the creative utility function (9) is dominated by differences in novelty. Reducing the parameter γ should extend the range of value charge explored by (4). By running a series of experiments with different γ and β we found that the combination $\gamma = 0.3$ and $\beta = 0.3$ appears to give a reasonable balance of novelty and value. Table 2 provides an example of the output indicating a higher preference for value compared to Table 1.

Table 2. Hypotheses in the order suggested by creative assistant using the baseline implementation enhanced with the value model $\gamma = 0.3$, $\alpha_- = 1$ and using utility function parameter $\beta = 0.3$.

Culprit	Motive	Source	Evidence
Ivins	Domestic	Simple U.S.	+21, -7 = +14
U.S. gov.	9/11	Gov. special	+11, -2 = +9
Ivins	Recipients	Gov. special	+18, -7 = +11
Hatfill	9/11	Simple U.S.	+12, -2 = +10
Hatfill	Domestic	Gov. special	+12, -2 = +10
Ivins	9/11	Iraqi special	+18, -8 = +10
Al-Qaeda	Recipients	Simple U.S.	+11, -2 = +9
Ivins	Domestic	Gov. special	+20, -7 = +13
Ivins	9/11	Simple U.S.	+20, -7 = +13
U.S. gov.	Domestic	Iraqi special	+11, -3 = +8

As suggested in the discussion of modelling value charge, different value models can be obtained by varying, α_- . We have briefly investigated the case of $\alpha_- = 0$ which requires additional adjustment of γ in order to obtain results substantially differing from the baseline implementation. Setting $\gamma = 0.1$ and $\beta = 0.5$ produces a new and different set of creative suggestions.

5.3 Enhanced novelty model

The enhanced novelty model according to (7) and (8) is controlled by parameters δ and κ . The self-potential (δ) has no effect in the present implementation since only hypotheses with zero awareness charge are candidates for creative suggestions. Using the baseline implementation with an enhanced novelty model with either $\kappa = 1$ or $\kappa = 2$ is, however found to make a significant difference in the output of the creative assistant. Table 3 shows the resulting hypotheses. Comparing this enhanced novelty model with the baseline implementation shows that the first four suggestions are the

same but that the following suggestions are quite different reflecting that the potential model will differ more from the Hamming distance as the inventory of known hypotheses accumulates.

Table 3. Hypotheses in the order suggested by creative assistant using the baseline implementation enhanced with the novelty model of (7) and (8) in which both $\kappa = 1$ and $\kappa = 2$ give the same output.

Culprit	Motive	Source	Evidence
Ivins	Domestic	Simple U.S.	+21, -7 = +14
U.S. gov.	9/11	Gov. special	+11, -2 = +9
Al-Qaeda	Recipients	Iraqi special	+9, -3 = +6
Hatfill	Personal	Simple Afg.	+5, -1 = +4
Iraq	Domestic	Gov. special	+10, -2 = +8
Hatfill	9/11	Simple U.S.	+12, -2 = +10
Ivins	Recipients	Simple Afg.	+14, -6 = +8
U.S. gov.	Personal	Iraqi special	+8, -3 = +5
Al-Qaeda	Personal	Simple U.S.	+10, -2 = +8
Iraq	9/11	Iraqi special	+8, -3 = +5

6 Discussion and Conclusions

Our initial experiments have focused on exploring key aspects of the generative algorithm and in particular the effect of varying selected components. The general impression from the experiments is that the creative assistant shows some promise and should be properly evaluated. This means that we must use much larger scenarios than in the present experiments. We cannot expect users to get a genuine eureka experience from creative suggestions unless the hypothesis space is so large that humans find it impossible to systematically consider all alternatives. Furthermore, we need a rich and complex evidence situation. In order to judge the creativity of the hypotheses suggested, and the value of the creative assistant, an experiment would use subject matter experts working on a realistic case for which they do not know the solution beforehand. The evaluation should be done in the process as well as post mortem.

We primarily regard the computational creativity as a tool for inspiring human analyst to explore a wider range of ideas and consider more alternatives rather than as generator of optional solutions per se. Furthermore, the infusion of creativity can stimulate users to expand the analysis model with new data, such as adding new value cells or looking for further evidence.

We recognize that morphological analysis may not be the ideal platform for building computational creativity. One could argue that a morphological chart spans a limited and static hypothesis space so that the only scope for creativity is to explore a predefined domain. However, we hold forth that the contribution of the computational creativity can only be evaluated by real users and that access to a user community of a perhaps less than ideal tool is a better starting point for researching computational creativity than to build a perhaps theoretically better tool with no practical opportuni-

ty for real-life evaluation. Likewise, we understand that a much better model of the user state can be built with state of the art behavioural research methods but we also know that more intrusive probes would deter professional analysts. By unobtrusively extending an existing, actively used baseline system for evidence supported morphological analysis we will in future more comprehensive experiments benefit from the crucial resource of the existing user community.

Although some of the suggestions produced by the creative assistant may seem to be obvious to a detached viewer, such ideas may still be useful in real-life decision making. Human free-thinkers may find it hard to get attention in a team that is locked into group-think. Originators of dissident ideas could be accused of disloyalty or having ulterior motives. That a divergent suggestion originates from a supposedly impartial and objective machine may help to make analysts consider it seriously and perhaps take it as a stepping stone for further exploration. Application of computational creativity to situation assessment in intelligence analysis, as described here, has not been explored before. Despite the limitations of the present implementation and the experimental scenario we feel that the novelty of the approach makes it interesting as a basis for further investigations and in particular full-fledged user trials.

Acknowledgements. Christian Mårtensson suggested that we should consider how to apply computational creativity in morphological analysis. We thank Tove Gustavi, Maja Karasalo and Christian Mårtensson for making the MHMA tool available for our research. This research is commissioned by FMV, the Swedish Defence Materiel Administration.

References

1. Heuer, R.J.: Psychology of Intelligence Analysis. US Gov. Printing Office (1999)
2. Gustavi, T., Karasalo, M., Mårtensson, C.: A Tool for Generating, Structuring, and Analyzing Multiple Hypotheses in Intelligence Work. In Intelligence and Security Informatics Conference (EISIC), 23–30 (2013)
3. Zwicky, F.: Discovery, Invention, Research - Through the Morphological Approach. The Macmillian Company, Toronto (1969)
4. Ritchey, T.: Wicked Problems – Social Messes: Decision Support Modelling with Morphological Analysis. Springer (2011)
5. Colton, S., Wiggins, G.A.: Computational Creativity: The Final Frontier?. Proc. 20th European Conference on Artificial Intelligence (2012)
6. Jändel, M.: Computational Creativity in Naturalistic Decision-Making. In Proceedings of the Fourth International Conference on Computational Creativity, 118–122 (2013)
7. Varshney, L.R.: Surprise in Computational Creativity and Machine Science. Decision Analysis Today 32: 25–28 (2013)
8. Cole, L.A.: The Anthrax letters. Skyhorse publishing (2009)
9. Ehrgott, M.: Multiobjective Optimization. AI Magazine 29 (4): 47-57 (2009)
10. Hamming, R.W.: Error Detecting and Error Correcting Codes. Bell System Technical Journal 29: 147–160 (1950)
11. Wixted, J.T., Ebbesen, E.B.: Genuine Power Curves in Forgetting: A Quantitative Analysis of Individual Subject Forgetting Functions. Memory & Cognition 25 (5): 731–739 (1997)

Bus Scheduling in Dynamical Urban Transport Networks with the use of Genetic Algorithms and High Performance Computing Technologies

V.A. Shmelev¹, A.V. Dukhanov¹, K.V. Knyazkov¹, S.V. Ivanov¹

¹ITMO University, Saint Petersburg, Russian Federation
vad1611@yandex.ru, dukhanov@niuitmo.ru,
constantinvk@gmail.com, sivanov@mail.ifmo.ru

Abstract. Public transport is one of the main infrastructures in any city. It facilitates the smooth running of everyday life for ordinary people. Public transport services require constant improvement and current methods of problem-solving are not sufficient for dealing with high traffic congestion. In this paper, we present a genetic algorithm to optimize bus routes. We achieved a reduction of passengers' waiting times at bus stops.

Keywords: public transport scheduling, genetic algorithm, urban transport networks

1 Introduction

Public transport is a vital infrastructure for common citizens. Every day, lots of people use different kinds of public transport to get to their work, university or school. The use of public transport contributes to the reduction of air pollution and prevents traffic jams. Public transport facilities could be improved. Our goal was to develop an effective bus schedule in order to increase public transport use.

The scientific and business communities have taken action toward the optimization of public transport [1]. There is a lot of work to be done in this area as it is an urgent problem.

In the current paper, we describe an algorithm to reduce passengers' waiting time and driving time, taking into consideration the variations of passenger density and traffic conditions. We offer a solution that optimizes public transport dispatch and the routes between bus stops.

2 Related works

As mentioned before, public transportation is a real problem. There are many approaches to optimize public transportation performance.

adfa, p. 1, 2011.

© Springer-Verlag Berlin Heidelberg 2011

In the paper, [2] traffic light span optimization was produced. The authors tried to implement a genetic algorithm's approach to resolve the problem of controlling traffic lights.

S. Amoroso and M. Migliore [3] used a multiobjective function to design bus networks for small towns, like Trapani. The applied approach gave a more efficient and effective bus network.

The bus network optimization problem also was solved in work [4]. It proposed a heuristic approach to solve bus transportation network optimization problems. Their method involved genetic operators and a number of additional ingredients to aggregate a number of performance indicators to calculate a fitness function.

Work [5] solved the problem of buses allocation among routes. This problem is solved in two steps. Firstly, the frequency of buses required on each route is minimized by considering each route individually. Secondly, the fleet size of the first step is taken as an upper bound, and fleet size is again minimized by considering all routes together and using GAs.

3 Criterion of Efficiency and Genetic Algorithm

In the current paper, we try to optimize bus dispatch to make public transport more convenient, which means less bus waiting time, less driving time, and lower cost of providing public transport activity. Thereby, we have the multiple-criteria optimization problem, where each parameter should be minimized. We use a weighted sum model to compare solutions. Our criteria can be determined by the following:

$$F = w_0 \sum_{i=0}^n waiting_time_i + w_1 \sum_{i=0}^m driving_time_i + w_2 \sum_{i=0}^{routes_number} \sum_{j=0}^{f(i)} bus_driving_time_j^i \quad (1)$$

w_i —relative weight of importance of the i criterion

n — number of came passengers.

m — number of transported passengers.

$waiting_time_i$ — waiting time by i passanger.

$driving_time_i$ — driving time by i - passenger.

$bus_driving_time_i^j$ — time, that needs j - bus on route i .

Let's calculate the described parameters. Suppose we have r routes and on each route there are $f(i)$ flights. Our bus schedule can be represented as set of tuples, each of them containing bus dispatching times on separate routes:

$$\{(t_1^1, t_2^1, t_3^1, \dots, t_{f(1)}^1), (t_1^2, t_2^2, t_3^2, \dots, t_{f(2)}^2), \dots, (t_1^r, t_2^r, t_3^r, \dots, t_{f(r)}^r)\}$$

where t_i^j — dispatch time of i -th bus on route j . The passengers' arrival times and traveling times are described on input data, which we get from a multi-agent model

[6] of the city. Routes information, fleet size and bus information are given and not changed during optimization process.

For each route, there is given order of bus stops. Example for route number 0:

$$(bus_stop_0^0, bus_stop_1^0, bus_stop_2^0, \dots, bus_stop_{bus_stop_count(0)}^0)$$

Where $bus_stop_count(i)$ - number of bus stops on route i , $bus_stop_k^i$ - number of bus stop k on i -th route. i -th bus arrival time on route k to bus stop j :

$$bus_arriving_time_{k_i}^j = bus_arriving_time_{k_i}^{j-1} + elapsed_time_{j-1}^j(bus_arriving_time_{k_i}^{j-1})(2)$$

Where $elapsed_time_p^q(time_moment)$ - time, required to pass path $p \rightarrow q$ in the $time_moment$ moment. But $bus_arriving_time_{k_i}^0 = t_i^k$

Time, required to i -th bus to pass j route:

$$bus_driving_time_i^j = bus_arriving_time_{k_i}^{bus_stop_count(j)} - bus_arriving_time_{k_i}^0(3)$$

Define i -th passenger waiting time. Suppose that our passenger comes to bus stop with number bus_stop at $pass_coming_time(i)$ time moment and takes the bus with bus number p and follows route k . Then, passenger waiting time can be represented as follows:

$$waiting_time_i = bus_arriving_time_{k_p}^{bus_stop} - pass_coming_time(i)(4)$$

Time required by passenger to drive to bus stop $destination_bus_stop$ on route k in bus p can be represented as follows:

$$driving_time(i) = bus_arriving_time_{k_p}^{destination_bus_stop} - (pass_coming_time(i) + waiting_time_i)(5)$$

Genetic and evolutionary methods are usually used for the solution of the schedule's problems. In this paper, these methods are also used to find such schedules that allow for the reduction of passenger waiting and driving time. First of all, it is necessary to represent our schedule as an individual of the population. In our realization, each individual contains chromosomes and each of these chromosomes contributes to the buses' route. A chromosome is an ordered set of the departure time t_j^i , where i - number of route, j - number of bus. Genes take their values from route working interval. There r - number of routes, $f(i)$ - number of buses on i -th route.

$$\{((t_1^1, t_2^1, t_3^1, \dots, t_{f(1)}^1), (t_1^2, t_2^2, t_3^2, \dots, t_{f(2)}^2), \dots, (t_1^r, t_2^r, t_3^r, \dots, t_{f(r)}^r)), \dots, ((t_1^1, t_2^1, t_3^1, \dots, t_{f(1)}^1), (t_1^2, t_2^2, t_3^2, \dots, t_{f(2)}^2), \dots, (t_1^r, t_2^r, t_3^r, \dots, t_{f(r)}^r))\},$$

$$\{(t_1^1, t_2^1, t_3^1, \dots, t_{f(1)}^1), (t_1^2, t_2^2, t_3^2, \dots, t_{f(2)}^2), \dots, (t_1^r, t_2^r, t_3^r, \dots, t_{f(r)}^r)\}$$

The current realization used a classical genetic algorithm scheme. It used one-point crossing, tournament selection with the best individual transferring to a new population.

To estimate individuals, the fitness function was entered (1). Discrete event simulation was used to measure passenger numbers in buses, stopping places, and times.

The fitness function calculation takes the most time, and individual estimating procedures are independent. Therefore, we can calculate that they are parallel. The parallel scheme of calculations is presented in Figure 1. A master-slave [7] paradigm was used to make parallel calculations. Using such an approach, we have automatic, dynamic balancing which is important when using heterogeneous networks

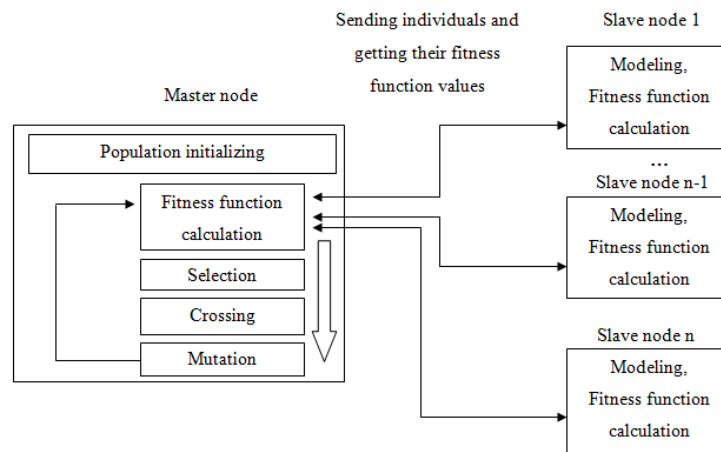


Fig. 1. Parallel calculations' scheme.

To analyze real data, we need big computing capacities. This problem can be solved by using cloud implementation. We use CLAVIRE [8], which allows us to make data flow among the parts of optimization package programs. The order of data flows is presented on fig. 2.

The *Transport_genpairs* package creates pairs of all possible points that buses visit. The *Transport_findpath* package receives pairs and searches optimal paths between them. The resulting data is sent to the schedule optimization program.

4 Scheduling algorithm examination

To check the algorithm, we performed an experiment on the data from the Petrograd district of Saint-Petersburg. Parameters for the genetic algorithm were found by a series of executions with parameters variations. The following parameters were used:

- Probability of mutation: 1%

- Probability of crossing: 60%
- Period of mutation: 15 iterations
- Population size: 3000 individuals
- Iteration number: 700

Fig. 3 shows the mean waiting time, the passengers arriving density and the number of passengers waiting at bus stops.

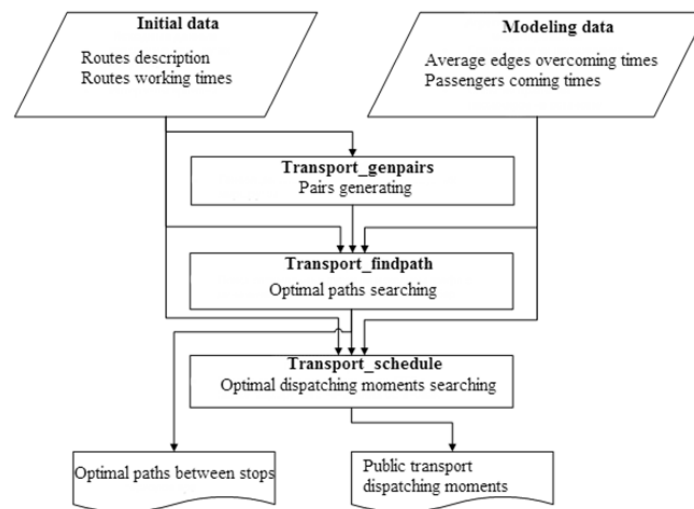


Fig. 2.Data flow order.

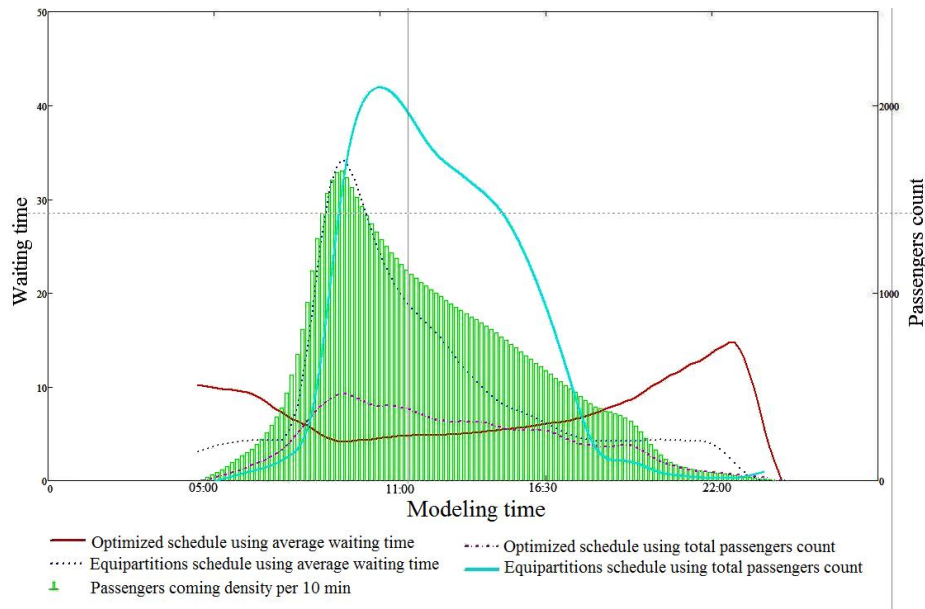


Fig. 3. Average waiting time of optimized schedule and evenly dispatching using.

As shown on figure 3, the developed algorithm and relevant software generate a more effective schedule compared to the traditional one, with equal headways. This difference is very noticeable during rush hour periods (from 8:30 to 11:00) when the mean waiting time for passengers was reduced by 3 to 5 times and the number of passengers waiting for a bus was reduced by 2 to 4,5 times. This advantage influenced the beginning and the end of the day, but it is normal if passengers know that public transport runs less frequently.

5 Conclusion

In this paper, we have described the method of bus scheduling in a dynamic transport network with the use of genetic algorithm and high-performance technologies. The effectiveness criteria were based on waiting times and transportation times of all passengers and costs of bus movements. We examined the received method and relevant software of the Petrograd District of Saint Petersburg. The experiment showed that the components of the effectiveness criteria are two or more times better for an optimized schedule against a traditional schedule with equal headways.

6 Acknowledgement

This paper is supported by Russian Scientific Foundation, grant #14-21-00137 "Supercomputer simulation of critical phenomena in complex social systems".

The research is done in Advanced Computing Lab (ITMO University), which is opened in frame of 220 Decree of Russian Government, contract #11.G34.31.0019.

References

1. M. Berlingerio, F. Calabrese, G. Di Lorenzo, R. Nair, and M. L. Sbodio, "AllAboard : a System for Exploring Urban Mobility and Optimizing Public Transport Using Cellphone Data," *Machine Learning and Knowledge Discovery in Databases*, pp. 663-666 2013
2. H. Ceylan and M. G. . Bell, "Traffic signal timing optimisation based on genetic algorithm approach, including drivers' routing," *Transp. Res. Part B Methodol.*, vol. 38, no. 4, pp. 329-342, May 2004.
3. S. Amoroso, M. Migliore, M. Catalano, and F. Galatioto, "A demand-based methodology for planning the bus network of a small or medium town," *European Transport \ Trasporti Europei n.* vol. 44 (2010): 41-56
4. M. Bielli, M. Caramia, and P. Carotenuto, "Genetic algorithms in bus network optimization", *Transportation Research Part C: Emerging Technologies* vol. 10, no. June 1998, pp. 19-34, 2002.
5. F. A. Kidwai, "A genetic algorithm based bus scheduling model for transit network," *Proceeding of the Eastern Asia Society for Transportation Studies* vol. 5, pp. 477-489, 2005.
6. A. B. Бухановский, "Моделирование и оптимизация движения городского общественного транспорта в среде облачных вычислений CLAVIRE," *программные системы: теория и приложения* vol. 3, no. 17, pp. 1-11, 2013.
7. K.-C. Huang, F.-J. Wang, and J.-H. Tsai, "Two design patterns for data-parallel computation based on master-slave model," *Inf. Process. Lett.*, vol. 70, no. 4, pp. 197-204, May 1999.
8. K. V. Knyazkov, S. V. Kovalchuk, T. N. Tchurov, S. V. Maryin, and A. V. Boukhanovsky, "CLAVIRE: e-Science infrastructure for data-driven computing," *J. Comput. Sci.*, vol. 3, no. 6, pp. 504-510, Nov. 2012.

MCDA and LCSA - A Note on the Aggregation of Preferences

João Clímaco¹, Rogerio Valle²

¹ INESC/University of Coimbra, Coimbra, Portugal
jclimaco@fe.uc.pt

² SAGE/Universidade Federal do Rio de Janeiro, Brazil
rogerio.valle@sage.coppe.ufrj.br

Abstract. Cost and social dimensions are now being added to the existing environmental Life Cycle Assessment (LCA), leading to Life Cycle Sustainability Assessment (LCSA). LCSA is very complex with deep uncertainties and, generally, involves several stakeholders. Therefore, the analysis and interpretation of the outputs of LCSA is a difficult and complex task, requiring aggregation of preferences. The work in progress here described deals with a comparative study regarding the use of open exchange interactive software packages dedicated to Multi-Criteria Decision Aiding in the context of LCSA outputs analysis and interpretation.

Keywords: Life Cycle Sustainability Assessment· Complexity, Multi-Criteria Analysis, Aggregation of Preferences Procedures.

1 Introduction

In recent years, Multi-Criteria models have undergone great development. As it is well known the concept of “Efficient Solutions”/Pareto Optimal Solutions is central in Multi-Criteria Analysis. A solution is Efficient/Pareto Optimal if there is no other feasible solution equal or better than this one in all the criteria, and strictly better in at least one of them. In multi-criteria generating methods, the aggregation of preferences is made a posteriori by the decision makers/stakeholders/actors. So, this class of methods consists in the calculation of the whole set of Efficient Solutions. The other two families of multi-criteria methods involve the use of techniques to aggregate preferences, as the value/utility function approaches, where the aggregation of preferences is made a priori, and the interactive methods, based on a progressive and selective definition of preferences. Roughly speaking, the aggregation of preferences, in most of the cases, includes one or several of the following procedures: optimization of weighted sums of the criteria (or other form of utility function), pairwise comparison of alternatives, or minimization of a distance to the ideal point, or to other reference point. Interactive procedures, especially those rooted in constructivism, avoid the aggregation of the preferences of decision agents in a single criterion, in some cases proposing the combination of algorithmic protocols with the experience and intuition of decision agents in the process of preference aggregation. We call them *open exchange* interactive procedures. In any case, aggregation implies loss of information, which means special care is needed. Ethical issues are very relevant in these circumstances... Having this in mind, it seems interesting to make a parallel between multi-criteria approaches and the informational basis of the different branches of the theory of justice, dealt with by A. Sen in [1]. We believe it helps understanding the potentialities and the limitations of the multi-criteria type of aggregation procedures. A. Sen summarizes, as follows, the main informational characteristics of these branches:

- Utilitarianism – based on the individual independent desire fulfillment, therefore closely related to *utilitarianism* ethics and to the rational choice theory. Of course, this is related to the use of *Value/Utility Functions* in multi-criteria aggregation.
- Rawls theory of Justice – it is rooted on *equity*, on a contractual framework and it is based on a transcendental ethics. Of course, it is related to the use of the Chebychev distance metric for looking for a solution as close as possible to some *reference point* in multi-criteria aggregation.
- Comparative theory of Justice (Sen) – comparing pairs of feasible alternatives. Using interpersonal information in a comparative framework involving real behavior of persons in different societies.

Based on a *comparative ethics*. Of course, related to those multi-criteria procedures based on pairwise comparison of alternatives, as, for instance, the ELECTRE methods in the framework of multi-criteria discrete/multi-attribute problems.

It is clear, from an ethical point of view, that the comparative theory is more related to the improvement of justice in real life conditions. It is based on a face to face evaluation and cross-fertilization of practical, and, so, necessarily imperfect conditions and behaviors. Moreover utilitarianism is closely related to the so called *rational choice theory of economy*, and Rawls theory is based on the idea of looking for solutions as close as possible to some pre-defined ideal system.

Taking into account the complexity of the problem (as it will be outlined next), it seems that the most interesting multi-criteria approaches in the framework of LCSA are those closer to the comparative theory of Sen. So, we privilege the interactive approaches, and among them learning oriented tools, which are easier to accommodate to practical situations involving several stakeholders, and in many cases are also able to take also into account public opinion points of view.

Preference aggregation is a widespread, day-to-day requirement. But perhaps the contemporary real life domain where it emerges in wider range and with greater intensity is in making decisions related to a sustainable way of life. Even though concerns with such decisions are not necessarily recent (for instance, still in the 19th century Marx and Engels referred to the finitude of resources and on the environmental challenge in the production of goods), sustainability remains a disputed concept. Yet, its association with development, as proposed by the Brundtland Report, sustainable development being the one that “meets the needs of the present without compromising the ability of future generations to meet their own needs” [2], has gathered an overwhelming acceptance. Later, the notion has been incorporated into business and governmental decisions through an often mentioned “triple bottom line” accounting framework that evaluates social, environmental and economic performance. Therefore, decisions about sustainability are inherently multi-criteria, raising some theoretical and practical controversies. It is consensual that sustainability assessment will be progressively more associated with the Life Cycle of the products. Here we are considering the set of social and environmental impacts not only associated to the production, but also to the use and discard of goods, services and events, from the extraction of the raw materials, through the several stages of transport and storage, to the recycling, recovering or disposal in landfills. The tool to evaluate all these impacts is the Life Cycle Sustainability Assessment (LCSA), which may help in public or private decisions on design options, transportation, end life use, etc. [3]. So, LCSA includes (environmental) Life Cycle Analysis (LCA), Social Life Cycle Analysis (SLCA) and Life Cycle Costing (LCC). Hence, it is very complex, involving much diversified uncertainty issues and also actors/stakeholders with diverse backgrounds, interests, and points of view on the subject. The methodological framework of LCA, formalized in ISO norms from 2006, has been tentatively extended to LCSA not yet with full success. In summary, four phases must be considered in LCA:

- Assessment objectives definition (ISO 14040 norm)
- Inputs and outputs inventories of the associated processes (ISO 14041 norm)
- Impact evaluation associated with each of the life cycle phases (ISO 14042 norm)
- Results interpretation (ISO 14043 norm)

In the context of this work, the following three aspects must be taken into account:

- The application of LCSA models implies a compromise between the detail of the data used and its operational restrictions. This issue has to be evaluated case by case.
- Presently, the use of LCSA is limited by the lack of ISO norms, except with regards to LCA. There are some recommendations regarding SLCA and LCC, but they are not norms. As expected SLCA is particularly complex due to lack of data, due to the involved qualitative and many times subjective aspects, and finally due to the existence of multiple actors/stakeholders, etc. [4].
- Concerning LCC some doubts remain on its inclusion as part of sustainability. Some authors put in question the relevance of LCC in sustainability assessment. For instance, [5] consider that, according to the sustainability definition of Brundtland report, only cost and benefits associated with poverty (based on the notion of decent life) should be accounted in LCSA. Other economic interests should be out of the scope of sustainability and its consideration would lead to biased results of a LCSA.

In conclusion, LCSA is much more complex than LCA and it poses preference aggregation problems that reflect and deepen the ones usually found in other sustainability approaches.

Summarizing, this paper aims to discuss the potentialities and limitations of some multi-criteria approaches that can be used to aggregate preferences in a particular, very complex situation: Life Cycle Sustainability Assessment problems. It is based on the experience of the authors concerning the difficulties involved in LCSA as well as in the use of multi-criteria approaches. Still, taking into account that this is a very recent issue, more definitive conclusions have to wait for ongoing research.

2 Preference Aggregation in LCSA – Summary of the Principal Difficulties...

In a paper by John Reap et al, the most relevant difficulties regarding Life Cycle Assessments are outlined and discussed [6]. On the one hand, they justify why this important methodology is still not very relevant in practice and on the other hand they give important indications on how to improve the available tools, namely concerning the aggregation of preferences. In summary:

1. *Impact Indicators* definition (characterizing the inventoried data in the different impact categories) is not a simple task. Furthermore, it must be remarked that when several *stakeholders/actors* exist they may have different opinions on the relevance of indicators.
2. Impacts aggregation (usually based on the use of *weights*) raises many controversies. This will be dealt with in more detail later.
3. *Uncertainty* associated with decision processes, and the eventual existence of *multiple actors* conditions the choice of adequate procedures of aggregation.
4. Increasing *time horizons* leads to a bigger difficulty in measuring the socio-economic impacts (for instance, public health impacts, etc.). Moreover, sometimes there are interconnections (positive and negative synergies, etc.) among impact indicators; as well as thresholds that, when surpassed lead to disruptions in the outputs. It is very difficult to deal adequately with these aspects in the models.
5. The interpretation of the results depends on the LCA/LCSA goals: from the analysis of processes to the support of policy definitions. So, its characterization is a priority in the beginning of the implementation process.
6. The use of adequate interfaces, namely enabling a holistic evaluation, is also an important issue.

3 MCDA Tools and Aggregation of Outputs in LCSA

In the framework of LCA, the use of multi-criteria tools is methodologically discussed in [7] and [8]. The phases of characterization and modelling of the process, i.e. the definition of the goals of the study, the boundaries of the study definition, the data harvest and management, the definition of the impacts, indicators/criteria and alternatives are crucial for the success of LCA and LCSA in practice. In [7] and [8] the potentialities of using *problems structuring techniques* in the LCA framework are discussed. It must be remarked that these issues are still more relevant in LCSA. Furthermore, it must be remarked that data harvest and management, leading to a good description and representation can be improved by using *information systems* and *knowledge management systems*. A very interesting project on the subject of *environmental cartography* is presented in [9].

The aggregation of outputs in LCA is usually done using weighted sums of normalized outputs obtained by direct elicitation procedures. The use of weights raises several questions or even errors, still more critical in LCSA, as mentioned in several recent papers. It must be remarked that compensatory procedures are very problematic under these circumstances.

Moreover, it must be remarked that the direct use of weights still involves normalization of the terms, which also contributes to the distortion of the results in many situations.

Multi-Attribute Utility Theory (MAUT) can help doing weights elicitation in a more scientific way, but cannot completely respond to our concerns. See, for instance, [6, 10, 11]. More recently, taking into account the limitations of the additive model and trying to avoid total compensation approaches, some authors propose the use of Outranking Approaches, such as the methods of the ELECTRE family, building the partial order outranking relations using the concepts of concordance and discordance, followed by some kind of aggregation and exploitation of the results. Although avoiding the complete compensation, these approaches require fixing a large number of parameters, such as concordance and discordance thresholds, weights (note that, in this case, they are less problematic than in the additive

model because they just represent coefficients of importance of the criteria, and not trade-offs as in additive model), and in many cases indifference and preference thresholds to take into account uncertainty associated to the criteria scores. It must be remarked that changes in these parameters can influence drastically the results (see for instance [12]). Furthermore, simple non-compensatory procedures can also be useful in the aggregation of outputs in LCSA.

Note that in the required aggregation to get each partial indicator/criteria, it is reasonable to use a direct elicitation of weights, when the terms are commensurable. On the other hand, the inter-indicators/criteria aggregation is much more delicate, particularly in LCSA, considering not only quantitative data, but also qualitative data.

Moreover, dealing with uncertainty and lack of information, especially in LCSA, can be dealt with by using diversified decision aiding tools using several types of uncertainty representations. Sensitivity analysis and robustness analysis are very important approaches in the framework of LCSA and can be efficiently associated with multi-criteria modelling. Finally, the classical LCA is a “*steady-state*” assessment. However, the life cycles under study involve inter-connections and dynamic interactions. So, *dynamic systems* tools should be associated with MCDA in many cases (see for instance [13]).

4 On the potentialities of Open Exchange Interactive Multi-Criteria Procedures in the aggregation of the outputs of LCSA

The use of Multi-Criteria Analysis in the aggregation of outputs in LCSA is still very limited. Taking into account the difficulties associated with the use of many well-known methods in this context, we are doing tests using flexible learning oriented multi-criteria packages better adapted to the difficulties associated with this problem. Reinforcing all previous remarks, we believe that the following issues should be considered carefully: the involvement of several stakeholders/actors (cooperating and/or negotiating) in this area; the desirable public participation in many situations, where public and private spheres and the evolution of their borders are important issues; and the inevitability of coping with large uncertainties (see for instance, [6] and [14]). Before summarizing the tools we are using, it must be remarked that two theses on LCSA were recently finished in UFRJ, Brazil, which have been an excellent stimulus for the present research [15, 16].

Here, we outline the principal characteristics of the *open exchange* (the meaning of this concept can be seen in 23) interactive tools we are using in this work and we sketch some preliminary conclusions.

4.1 VIP-Analysis [17]

Main characteristics: It is an interactive platform dedicated to the choice problematic dedicated to the evaluation of a discrete set of alternatives according to a multi-attribute additive value function. It does not require precise values for the weights. Rather, it can accept imprecise information (i.e. intervals and linear constraints) on these values. It enables the discovering of robust conclusions, i.e. those that hold for every feasible combination of the weights/scaling constants, and to identify what is the variability of the results resulting from the imprecision in the parameter values. This software is free available by its authors: lmcaldas@fe.uc.pt or jclimaco@fe.uc.pt.

Preliminary conclusions: As VIP Analysis does not requires accurate values for the weights, some of the drawbacks of using the additive model are really mitigated: the decision makers/stakeholders only need to identify linear constraints for the weights, normally by indirect ways (one example is by comparing equivalent swings); it is very appreciated, in the context of LCSA, the possibility of identifying which are the robust conclusions compatible with the use of incomplete information regarding the weights; this tool seems very adequate to support a group of stakeholders/actors meeting face to face around a computer, which is also very adequate in the LCSA context. However, a proficient use of VIP-Analysis requires the sharing of knowledge about the tool potentialities and limitations with the stakeholders, so a facilitator is required.

4.2 A Non-Compensatory Software Package Integrating an Interactive Dashboard with an extension of the Conjunctive Method[18]

Main characteristics: The interactive package, based on an extension of the Conjunctive Method, integrates an interactive multidimensional dashboard, in order to open options of analysis. It must be emphasized that we do not follow the mainstream aggregation frameworks using weighted sums of normalized data regarding the considered dimensions, as in many situations when aggregating outputs of LCA/LCSA. We opted by a non-compensatory tool thence avoiding the most negative aspects of the additive model. The software of support to our proposal [19] is based on an interactive implementation of the Conjunctive Method, enabling the consideration of up to three performance thresholds, having in mind to classify the objects under evaluation. Quantitative and qualitative criteria are admitted. For details see [17]. This software is free available by its authors: sergio.fernandes@estsetubal.ips.pt, jclimaco@fe.uc.pt, mecaptivo@fc.ul.pt

Preliminary conclusions: Confronting this aggregation process with the additive model, we realize that in those LCSA situations where providing (even incomplete) information on weights is not a comfortable requirement for the stakeholders/decision makers, this can be a very adequate alternative tool: since, in the case of the proposed methodology, there is no inter-criterion aggregation, there is no need to reduce the various evaluation dimensions into the same scale thereby avoiding all possible associated distortions which we have referred to above; the described aggregation process is non-compensatory thence avoiding the problem that a weak performance in one dimension may always be compensated by a strong performance in another dimension, as in an additive model; there is no need to assume additive independence among the various attributes, an adequate property because that is a requirement too strong in the context of LCSA. In conclusion, the main limitations of the additive model are surpassed.

The characteristics of this tool seem designed to fulfill some requirements in LCSA, as requested in the scientific literature. Namely: the simplicity of the user-oriented proposed tools, both conceptually and operationally, being supported by graphics in order to promote their intuitive use, has been very well received in the complex environment of LCSA interventions; in general, the selection of the indicators to judge sustainability (so, economic, social and environmental indicators) should involve multidisciplinary scientific actors, several political actors and social actors (including institutions and citizens), therefore, we are dealing with a group decision involving, in many cases, negotiation processes; it is very relevant that besides the classification of alternatives/scenarios, the system enables a holistic view of them, where the combination of an analytic process with visual holistic views constitute a remarkable added value, giving an important global feedback from the system to the stakeholders contributing to a global perception of the sustainability phenomena. The system provides a very flexible and intuitive sensitivity analysis of the thresholds, together with an impressive visual interactive sensitivity analysis and these are very important issues for the stakeholders understanding and validation of the process (note that difficulties in fixing adequately parameters is one of the principal reasons for the small practical impact of other approaches...); pairs of alternatives /scenarios, identified as belonging to the same class, can be holistically compared; the system also enables a visual inspection on the robustness of the classification concerning each dimension.

Note that, while LCC is dominated by a short-term rationality, LCA needs to take into account longer term effects. On the other hand, SLCA is very qualitative and subjective in most of the situations. In these circumstances, the use of a tool not requiring inter-criteria information is the only way to escape to a big mess...

For details on the application of this tool in other type of applications see [20] and [21].

4.3 ELECTRE Methods

This type of approach has not yet been tested. The idea is using a new package involving a control panel with slide bars enabling individual or combined “continuous” variation of the parameters of the model, such as the threshold values.

It is very adequate in our case because, as was told before, the uncertainty is very high. Note that it enables by visual inspection a real time evaluation of the variability of the outranking relations graph with parameter variations.

The choice of the adequate ELECTRE method depends on the problematic of the case under study. For instance, in the cases presented in [15] and [16], we have choice problematic problems. For details on ELECTRE methods see [22].

Further planned research will use the software MATRIX, developed at the Civil Engineering Department, University of Coimbra, Portugal by Luís Alçada Almeida (see <http://moth.dec.uc.pt:8000/matrix>).

5 Conclusions

In this paper we have discussed the potentialities and limitations of some open exchange interactive multi-criteria approaches dedicated to the aggregation of preferences, when applied to very complex problems, as the Life Cycle Sustainability Assessment ones. Our proposal is based on the experience of the authors concerning the difficulties involved in LCSA as well as in the use of the most important multi-criteria approaches, potentially adequate to the above referred situations. More detailed research is still in progress, but up to this moment they seem to confirm the ideas above expressed. These practical experiments will be detailed in a future work, when conclusions and recommendations for potential users will be presented and discussed.

Acknowledgement. This work has been partially supported by the Portuguese Foundation for Science and Technology (grant PEst-OE/EEI/UI308/2014).

6 References

1. Sen, A, "Idea of Justice", Harvard University Press, 2011.
2. Our Common Future –Report of the World Commission on Environment and Development, United Nations, 1987.
3. Zamagni, A., H.L.Pesonen, T.Swarr, "From LCA to Life Cycle Sustainability Assessment: concept, practice and future directions", International Life cycle Assessment, 18, pp 1637-1641, 2013.
4. Valdivia, S., C.Ugaia, J. Hildenbrand, M. Traverso, B. Mazijn, G. Sonnemann, A UNEP/SETAC Approach Towards a Life Cycle Sustainability Assessment – our contribution to Rio+20. International Journal of Life Cycle Assessment, vol18, pp1673-1685, 2013.
5. Jorgensen, A., I.T. Herrmann, A. Bjorn, "Analysis of the link between a definition of sustainability and the life cycle methodologies", International Journal of Life Cycle Assessment, 18, pp1440-1449, 2013.
6. Reap, J., F. Roman, S. Duncan, B. Bras, "A survey of unresolved problems in life cycle assessment –part 2: impact assessment and interpretation", International Journal of Life Cycle Assessment, 13, pp374-388, 2008.
7. Mazri, C., A. Ventura, A. Jullien, D.Bouyssou, "Life Cycle Analysis and Decision Aiding: An Example of Roads Evaluation", 2001, <http://sciencestage.com/d/1093635/life-cycle-analysis-and-decision-aiding-an-example-for-roads-evaluation-.html>.
8. Miettinen, P., R.P. Hamalainen, "How to Benefit from Decision Analysis in Environmental Life Cycle Assessment (LCA). EJOR 102, pp279-294, 1997.
9. Raherimandimby, H. R., D. Le Boulch, D. Bouyssou and M. Grundstein, "Environmental Notions Representation and Description: Towards a Redefinition of the Relationships between Information Systems Development and Individual Cognition", 2008.
10. Benoit, V. and P. Pousseaux, Aid for Aggregating the Impacts in Life Cycle Assessment. International Journal of Life Cycle Assessment, vol.8, pp 74- 82, 2003.
11. G. Huppes, L. von Oers, U. Pretato, D. Pennington, Weighting Environmental Effects: Analytic Survey with Operational Evaluation Methods and a Meta-Method. International Journal of Life Cycle Assessment, vol17, pp 876- 891, 2012.
12. Cinelli, M., S. Coles, K. Kirwan, "Use of Multicriteria Decision Analysis to Support Life Cycle Sustainability Assessment: An Analysis of the Appropriateness of the Available Methods", In the 6th International Conference on Life Cycle in Gothenburg 2013.
13. Halog, A. and Y. Manik, "Advancing Integrated Systems Modeling Framework for Life Cycle Sustainability Assessment". In Sustainability, vol3, pp 469 – 499, 2011.
14. Benetto, E. and C. Dujet, "Uncertainty Analysis and MCDA; A Case Study in Life Cycle Assessment (LCA) Practice". In: Proceedings of the 57th Meeting of the European Working Group on Multicriteria Decision Aiding Viterbo (ITALY) 27-29 March 2003. (Oral presentation)

15. Pontes, A, "Evaluation of Scenarios of Reverse Logistics of Household Pharmaceutical Waste in Rio de Janeiro using Life Cycle Sustainability Assessment. Tese de Doutorado em Engenharia da Produção (Ph.D. Thesis in Production Engineering), COPPE, UFRJ, Rio de Janeiro, May 2013.
16. Gabbay de Souza, R. , Modelling and Sustainability Assessment of Waste Management Systems Based on Multiple Stakeholders Perspectives, Tese de Doutorado em Engenharia da Produção (Ph.D. Thesis in Production Engineering), COPPE, UFRJ, Rio de Janeiro, April 2014.
17. Dias, L. and J. Clímaco, "Additive Aggregation with interdependent Parameters: the VIP Analysis Software", JORS vol.51, pp1070-1082, 2000.
18. Clímaco, Fernandes and Captivo 2011 Clímaco, "Classificação MultiAtributo Suportada por uma Versão Interactiva do Método Conjuntivo" (An Interactive version of the Conjunctive Method dedicated to Multiattribute Classification Problems), CIO – Working Paper 9/2011 (original in Portuguese).
19. Clímaco, J. , Maria Eugénia Captivo, Sérgio Fernandes, "Classificação Multi-Atributo Suportada por uma Versão Interactiva do Método Conjuntivo CIO – Working Paper 9/2011 (in Portuguese).
20. Valle, Rogerio and João Clímaco, "Green Economy in the State of Rio de Janeiro – A Non-Compensatory Multidimensional Interactive Evaluation", presented at the XXII MCDM Conference, Malaga, June 2012.
21. Clímaco, J., J. Craveirinha, "Multi-Actor Multidimensional Quality of Life and Sustainable Impact Assessment – Discussion Based on a New Interactive Tool". Proceedings of GDN 3013, Sweden June 2013.
22. Figueira, J., V.Mousseau , B.Roy, ELECTRE Methods, Multiple Criteria decision Analysis: State of the Art Surveys, Chapter1, eds J. Figueira, M. Ehrgot and S. Greco, Springer 2005.
23. Feyerabend, "Against the Method", New Left Books, London 1975.

A Social Creativity Support Tool Enhanced by Recommendation Algorithms: The Case of Software Architecture Design

¹George A. Sielis, ²Aimilia Tzanavari, ¹George A. Papadopoulos

¹Department of Computer Science, University of Cyprus, P.O.Box 20537, 1678,
Nicosia, Cyprus
{sielis, george}@cs.ucy.ac.cy

²Department of Design and multimedia, University of Nicosia, P.O.Box, 24005, 1700,
Nicosia, Cyprus
Tzanavari.a@unic.ac.cy

Abstract. Reusability of existing knowledge for the design and development of new ideas is a key principle of the definition of creativity as a process. Professionals to create the Architecture design of innovative software tools use software Architecture Design as such process. This work proposes a framework/tool by which the design of SAD can be directed based on several types of recommendations and in particular of well-known design patterns. This work approaches the recommendations the use of contextual factors in combination with recommendation methods and known recommendation algorithms.

1 Introduction

A creativity outcome is a sequence of thoughts and actions that lead to a novel adaptive production [6]. Plucker and Beghetto [7] define creativity as the interplay between ability and process by which an individual or a group produces an outcome or product that is both novel and useful as defined within some social context. Having assumed that creativity is an attribute that we all have, we reach the conclusion that it is necessary to find ways and means to assist in outsourcing this property. Creativity is a characteristic that can be cultivated and at the level of its development is different for each person. Cultivating and expressing the creative ability can be achieved through help and guidance. One approach to achieve that is through the use of Creativity Support Tools (CSTs). Creativity Support Tools are software systems that can emulate a realistic creative process by offering the users the ability to record thoughts and ideas with the use of ICT means. The utilization of the existing CSTs is bounded to the replacement of the conventional means of collecting ideas, such as the paper or the white board, with corresponding virtual environments, but they also offer the users the possibility to collaborate with other people from a distance.

This research aims to reinforce the importance of these tools with the addition of more advanced functionality. Generally, CSTs are the basis for strengthening the

adfa, p. 1, 2011.

© Springer-Verlag Berlin Heidelberg 2011

creative capacity and they can also be used for the cultivation of the creative ability. At this stage, these tools lack the mechanisms that can create the necessary stimulus to the user, which would make the user more productive in the process of creativity [8]. Such mechanisms are the recommendation systems for the support of the creativity process like for example the recommendations of people to collaborate with, related problems or solutions, related ideas and finally related resources such as articles, images or videos. Recommendation systems are systems that belong to the information filtering systems family. In general Recommendation systems seek to predict the rating or the preference that a user would give to an item. Thus the common methodology that the recommendation systems follow is to find the correlation between three types of modeling, user, rating and item modeling in order to produce recommendations. The use of recommendation systems to support a creativity process is a challenging task due to the multi-dimensionality of factors that influence the process and the complexity in identifying the type of recommendation that can foster creativity at each phase of the creativity process within a Creativity Support Tool.

The motivation for this work is driven from the notion that Creativity and Innovation are keys to success for any business. The development of a Social Creativity Support Tool (SCST) can become the means of bringing professionals and experts together for the exchange of ideas and collaborate together through creativity support spaces supported by user aiding functionality.

For example at the existing CSTs noticed that they lack of methods for finding people while formulating a collaboration team. The functionality they support regarding the group composition for collaboration is only the invitation of person to join a creativity session without analyzing her knowledge background and the level of competence related to the project. The use of a social network designed for creativity will give the opportunity to a practitioner to get recommendations of experts in a particular topic who are (or not) in the same company or organization and can add value to a creative project that is processed. A social creativity network can be beneficial for businesses in two ways. On the one hand it will be used as means for collaborations between companies (public social web network). On the other hand it will be used as a tool that connects employees of large scale international companies (intranet application). The supporting recommendations can vary according to the context of the creative project and the type of actions needed for the successful completion of a creative project. In that case the interfaces, the tools and the supporting functionality can be different based on the context of the project.

2 Problem Statement

In this research it is stated that social creativity tools integrated with context aware recommender systems can positively influence the creativity process. Context aware recommendations such as recommendations of users for collaboration or resources related to the context of a creative project but also other recommendations related to particular creativity phases can facilitate the creativity process. Therefore the problem approached has two parts: first the development of a Social Creativity Support Tool

for the development of professional networks targeted in solving problems and producing innovative ideas and second the development of a context aware recommendation algorithm, which will have the flexibility to analyze the most important contextual factors for each phase of the creativity process and produce useful recommendations to foster the professionals' creative ability.

3 Approach

As mentioned above, the creativity process is not a straight-forward process. On the contrary, it is characterized by dimensions that may vary depending on the domain it is applied to. According to the problem statement, the expected outcomes of this work are related to creativity as the main research area. The wideness of the creativity topic though makes the design of a research plan very complex. For example a creativity process may be an art session, a book conversation or the development of a web site. So to study the research hypothesis a particular topic was selected in order to model creativity and to extract valid conclusions. Therefore we framed the problem into a particular creativity-related problem, that of the Software Architecture Design (SAD) process. SAD is by default a creativity process since it is a process that is executed in phases by individuals or groups of experts. The participants of such a process must have an important knowledge background in the topic and be able to exchange expert opinions/ideas for each phase of the SAD lifecycle process. For the study of SAD as a creativity process and following the problem statement, a Creativity Support Tool has been developed for the support of SAD. The SCST will be the tool on which the users will be able to create professional networks for collaborative creativity sessions. The development of the creativity tool and the integration with the proposed recommendation algorithms will be used as prototype for the examination of the current work's research questions. In particular, the research goals that were set for the current research work are listed in the rest of this section.

3.1 Research Goals and Methods

To assess the impact that Context Aware Recommendations have in creativity process within a Social Creativity network, the following set of objectives have been determined:

- Define the creativity context elements and design of a creativity model that will be applied in a social creativity network.
- Define the process of SAD through the designed creativity tool and identify the types of recommendations that must be applied at each step.
- Define the recommendation algorithms and in particular specific recommendation methods for recommending SAD oriented types of recommendations such as, recommendation of Design Patterns, for the examination of how such recommendations are influencing the SAD as a creative process.
- Design the evaluation plan to lead to evaluation results concerning:

- The usability of the SCST
- The accuracy and validity of the proposed recommendations
- User satisfaction and experience by using types of recommendations such as recommendation of users to collaborate, resources and most importantly the recommendation of Design Patterns in order to facilitate users in designing Software Architecture independently to the knowledge level they currently have.

The outcomes of this research are expected to be taken by applying the prototype in a large scale international IT company whose employees will be asked to create profiles in the prototype application. Then a group of Software Architect Analysts will be asked to design new projects with it. They will be asked to formulate groups of users to collaborate with, from branches of the company that are in other countries and use the recommender system to collect resources related to a given project. With the design of a prototype for a domain specific group of users will lead to the collection of results regarding the creativity impact as well as the overall users' satisfaction and the prototype usefulness. The results of the test will be used for further improvements of the prototype the improvement of the proposed recommendation algorithm and also lead to the design of new methods and variations of the recommendation algorithms that will be applied.

4 SAD Recommendations - Related Work

The last years Design Patterns for Software Architecture design are increasing. New patterns are appearing and applied widely. Gueheneuc et.al [2] proposed a methodology of recommending design patterns through the textual analysis of each pattern into the most important words and computing the similarity distance between those words and the words of the query given by the user. Gomes et al [3] proposed a Case Based Reasoning Recommendation system for the recommendation of Design Patterns based on previous experiences using a Design Patterns Knowledge Base and related taxonomies. A similar system developed by [4] that recommends patterns using the Implicit Culture Framework (ICF). The recommendations are produced based on the users previous actions, based on conventional Information Retrieval and CBR methods. Palma et al. in [5] propose a DPR framework which recommends patterns based on predefined questions that the designers have to answer, and based on the given answer the framework has a weighting mechanism for the selection of the appropriate pattern. The initial identification of patterns that can be used through the DPR framework [5] are selected through LUCENE indexing and TF-IDF filtering of the query given and the intent description of each pattern.

5 Current Status and Future Work

The research work described through the objectives definition in section 3 is a work which is still in progress. Most of the objectives are currently met in the current status

of the work, but some are still in progress. In particular, at the time being we defined the contextual model of creativity which is designed and represented in the form of an ontology. The Contextual model is an extension of the proposed ontology proposed in previous work of ours in [9]. The new ontology is taking into account additional conceptual elements related to the phases of the creative process as a SAD process. With the use of the designed ontology it became possible to bring data from different data sources such as existing software design repositories like GitHub.com and Sourceforge.net and combine them with the new data that is created in the local database of the SCST. Based on this model we managed to define the most important relations that can be used as input data for the recommendation algorithms that we use.

Having the Creativity Contextual Model defined in the form of ontology, and taking into account the semantic data and relations of the model, we developed a Knowledge Base tool by which a user is able to browse through the existing SAD projects, the participants of the projects and the metadata related to each particular entity. The knowledge Base tool is part of the overall SCST which in its whole contains a second sub-tool which is used for the design of Software Architecture Diagrams. The flow that can be followed by end users in order to open an existing Design project or create a new one defines the proposed workflow process for the SAD as creative process. Thus at each phase of the process we defined specific Context Aware Recommendation types that vary based on the process phase and the available context information. In addition to that social networking functionality is also added in particular phases such as professional network connections between designers; *Follow* a professional's activities or view the progression of selected SAD projects, which are in progress by other professionals.

A lot of effort at the current state of the research is given in producing recommendations on Design Patterns during the SAD in the form of diagrammatic design. The recommendation of Design Patterns consists one of the most important tasks of this work and to meet that objective we developed a mechanism by which, we use three different recommendation methodologies. The methodologies used are the following: 1. *Lucene indexing, TF-IDF [10] and Cosine Similarity*, 2. *Graph based recommendations using Design Patterns rules* and 3. *Recommendation of Design Patterns Based on Related Projects*.

Each methodology is used in based on the phase of the design and the available contextual information but in that case the selection of the methodology is determined only within the design editor sub-tool of the prototype. The Design Patterns were selected from the GoF [11] list of patterns and modelled in a separate ontology that is queried according to needed recommendation methodology input.

The remaining work for the completion of this research work is mostly related to the Evaluation Objectives. We currently organize the evaluation sessions with professionals Software Architecture Designers from corporate organizations as well as academic ones. The evaluation will be done in a controlled environment, using specific use case scenarios and specific problem statements. That way we aim to collect feedback from the experts regarding the experience and usability, the accuracy of recommendations, the impact of recommendation in the creativity design and finally whether the social attributes of the prototype facilitated overall process. By the completion

of this work we will have the following accomplished tasks: 1. The creativity ontology model, which will be available to be used for other CSTs, The Design Patterns Ontology, a Semantic Web Application having a SAD knowledge Base and a graph editor supported by recommendation types that facilitate the creative process of SAD and finally three new methodologies for the recommendation of Design Patterns.

References

1. Alexander, Christopher, Sara Ishikawa, Murray Silverstein. A Pattern Language: Towns, Buildings, Construction. New York: The Oxford University Press, 1977.
2. Y. G. Gueheneuc, and M. Rabih. A Simple Recommender System for Design Patterns, Proceedings of the 1st EuroPLOP Focus Group on Pattern Repositories (2007)
3. P. Gomes, F. C. Pereira, P. Paiva, N. Seco, P. Carreiro, Jos e Luis Ferreira and Carlos Bento 2002. Using CBR for Automation of Software Design Patterns. In Proceedings of the 6th European Conference on Advances in Case-Based Reasoning, pp:534-548.
4. WEISS, M. AND BIRUKOU, A. 2007. Building a pattern repository: Benefitting from the open, lightweight, and participative nature of wikis. In Workshop on Wikis for Software Engineering at ACM WikiSym, 2007 International Symposium on Wikis (WikiSym), Montre'al, Que'bec, Canada, October 21-23.
5. F. Palma, H. Frazin, Y. Gueheneuc, and N. Moha, "Recommendation System for Design Patterns in Software Development: An DPR Overview", in Proceeding of 2012 Third International Workshop on Recommendation Systems for Software Engineering (RSSE), Switzerland, pp. 1-5, 2012.
6. Lubart T.: Models of the creative process: Past, present and future. Creativity Research Journal 13, 3-4, 295-308, 2000
7. Plucker, J. A., Begheto, R. A.: Why creativity is domain general, why it looks domain specific, and why the distinction does not matter. Creativity From Potential to Realization, 153-167, 2004.
8. Sielis A. G., Tzanavari A., Papadopoulos A. G., Enhancing the Creativity Process By Adding Context Awareness in Creativity Support Tools, HCI 09 International, San Diego, 2009
9. Sielis A. G., Mettouris C., Papadopoulos G. A. , Tzanavari A., Dols, R.M.G, Siebers Q. A Context Aware Recommender System for Creativity Support Tools, JJUCS, 2011, volume 17, no. 12, p. 1743-1763
10. Pasquale, L., Gemmis, M., Semeraro, G. (2011). Content-based Recommender Systems: State of the Art and Trends. Book Section in Recommender Systems Handbook, Ricci, Francesco, Rokach, Lior, Shapira, Bracha, Kantor, Paul B., 978-0-387-85819-7, http://dx.doi.org/10.1007/978-0-387-85820-3_3, Springer US, pp.73-15.
11. Gamma, E., Helm, R., Johnson, R., Vlissides, J. *Design Patterns: Elements of Reusable Object-Oriented Software*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA. 1995

A Model for Managing Organizational Knowledge in the context of the Shared Services supported by the E-Learning

Agostinho Sousa Pinto and Luís Amaral

IPP/ISCAP/CEISE, Portugal
Universidade do Minho / Centro Algoritmi, Portugal
apinto@iscap.ipp.pt / amaral@dsi.uminho.pt

Abstract. Lifelong learning has become a fundamental element in any organization, as a tool for managing development, sustainability and innovation of knowledge capabilities. In this paper, we start by describing the combination of two phenomena - Shared Service Centre (SSC) and E-Learning - that expand knowledge capabilities. Also this intentional combination provides for the emergence of new knowledge. Then we produce a conceptual model for developing a continuous learning organization. The Delta model assumes - in essence - the characteristics of complex adaptive systems (CAS) and incorporates concepts of crowdsourcing. This model conceptualizes and integrates SSC for managing organizational knowledge and memory repository and E-Learning, as the enhancer for a push-pull communication process and the emergence of knowledge. Then, from the conceptual model, we mathematically deduct a Δ (Delta) factor. This factor presents characteristics that allow it to be used for determining the viability of continuous learning from the sharing of knowledge capabilities.

Keywords: knowledge management, conceptual model, lifelong learning, shared services, e-learning, complex adaptive systems.

1 Introduction

Learning organization is induced as a concept defining a structure, where knowledge is fully utilized to increase capacity, behaviour change and increased competition. As knowledge management evolved and developed, learning organization conceptualisation has become an interesting metaphor for the sustainability of contemporary organizations. It focuses on the importance and plausibility of the relationship between sustainability and knowledge, as a basis for innovation and business performance. This process of building a learning organization is founded primarily to be sustained by the availability of organizational memory. With the development of Information and Communication Technology (ICT), the conditions for increasing the amount of

adfa, p. 1, 2011.
© Springer-Verlag Berlin Heidelberg 2011

organizational memory, its availability and its relational power were created. The basic conditions were the interoperability of systems and knowledge of communication tools.

The Shared Services Centre (CSC) and E-learning characterise this evolution, respectively, by promoting interoperability and communication of knowledge. These two ICT developments are the focus for two reasons. The first reason is the adoption of the Shared Services Centre by the organization to build a common organizational memory. A Shared Services Centre in its concept is a repository of memories of different organizations, which integrates diversity. The second reason is the adoption of e-learning as a tool for the communication of knowledge. E-learning is in its concept a communication tool for knowledge. These two developments of ICT overcome two of the facilitators of knowledge management - the existence of diversity and communication skills. These facilitators can be expanded, so that organizations can have access to other experiences that could develop unique knowledge. Moreover, if this knowledge is available to be communicated it may allow an organization what to choose to learn.

In this paper, we conceptually explore knowledge management organization and, consequently, its ability for continuous learning. This results in the development of lifelong learning, as a dependent from the Shared Services Centre (CSC) and E-learning variables. We focus on the question "Does using a SSC and E-learning conceptual model positively influence knowledge management?" Our work is based on the resources of the organization (RBV), where there is a live, independent and combined memory supported by resource stored in the systems of information technology and communication organizational memory. These features develop capabilities to support knowledge sharing. We are on the threshold of the definition of the Δ (Delta) factor for these knowledge sharing capabilities. This factor, initiated for SSC and E-Learning conceptual model for knowledge sharing can produce a positive effect on knowledge management and support the sustainability of the learning organization.

The sustainability of an organization relies mainly on the result of its responsiveness and adaptability to the environment that surrounds it. This environment is a competitive environment and each organization requires adaptability and the adoption of sustainability policies, which are supported by learning and innovation. This work shows that this dimension integrates continuous learning from a perspective of knowledge management. As a "function", knowledge management is able to promote the goal of developing and maintaining an organization of continuous learning. Two elements of knowledge management are pointed out for promoting continuous learning: first, diversity management and, second, the communication of that diversity. These two elements are presented in any environment.

This work begins with an approach of the theoretical assumptions. Section 2 next discusses the conceptualization of the structure that can support the learning organization, while Section 3 presents the Research Design. Section 4 then describes the conceptual model, the Δ (Delta) factor and other results and, finally, Section 5 presents the conclusions and future work.

2 Conceptual background

2.1 Knowledge management

Organizations recognise the need for collaborative innovation, as it improves how they can create, accumulate and exploit knowledge that enhances their competitiveness and the sustainability of their organization. Additionally it not only revolutionises working ways and creation but it also promotes organizational learning. Moreover, it engenders the management of knowledge and interactions with the environment. Knowledge management should be used to stimulate organizational knowledge for the optimization of diversity and interoperable resources.

Through globalisation, the success of organizations depends on their ability to interact with the environment and their ability to operate globally [1]. There is a need to identify and define technological models that can effectively support this interaction for pursuing an innovative and entrepreneurial approach. In this pursuit knowledge management has a central role.

The creation of knowledge occurs in different forms. According to Davenport and Prusak [2], [3] there are five ways to generate knowledge; acquisition, dedicated resources, fusion, adaptation, and knowledge networks. Generated knowledge can then be analysed and this generates knowledge internalisation, if its analysis determines its usefulness for the organization. After verifying its usefulness, it is then systematised and filed, through a codification and co-ordination of knowledge. This usefulness aims to make knowledge accessible to those who need it. To determine how it should be encoded, the knowledge is defined into a tacit or explicit way.

Tacit knowledge is the knowledge that individuals or groups have but it is not consciously accessible, since it is not yet part of the reality [4]-[6]. This knowledge acquiring requires an effort of understanding by processes that are not directly controlled by the learner. Explicit knowledge is the knowledge that is at the conscious level. Thus, not only the person or group recognise it, as they can convince others of this knowledge [7].

An important component of knowledge management is the existence of organizational memory. This memory enhances organizational knowledge gathering, organization, dissemination and re-use of the knowledge created inside the organization. Organizational memory is then a system capable of storing the resulting perceptions of experience or of keeping the abstract memory registers construction. This organization memory should be persistent in time and recoverable [8].

Organizational Knowledge Management is a field of multi-disciplinary research that cuts across areas such as information systems, computer science, human resource management and organizational sciences. It focuses on the sharing and re-use of knowledge. It includes individual and group skills, for improving quality, efficiency, increased customer and employee satisfaction and reduction of risks. Additionally, it improves knowledge development, through imagination, experience and experimentation.

The process of knowledge management can be structured into four fundamental areas [7]:

- Knowledge creation;
- Retention and retrieval of knowledge;
- Sharing and knowledge transfer; and
- Application of knowledge.

These four areas sustain the continuous learning, which is essential for keeping up to date human resources in relation to technological innovations and work practices. Additionally, information communication technology makes learning easier.

2.2 Shared services

The promise of SSC comes from a hybrid conception of traditional models aimed at capturing the benefits with centralised and decentralised arrangements. By unbundling and centralising activities, the basic premise for SSC seems to be that services provided by one local department can be replicated in others with relatively few difficulties. This is a conceptualisation that is only possible through the support of information and communication technologies.

Centralised governance structures are characterised by substantial economies of scale and scope, because procurement of assets and services is done on the broadest scale possible within the organization. A centralised staff can eliminate redundant functions and improve the clarity of strategic alignment. Moreover, the organization evolves a simpler communicating structure. However, centralised governance makes response time slower and often there is a higher distance from customers.

In the decentralised governance structure business units respond faster and with more flexibility to necessary changes, since they have the knowledge and choice about the usable resources to support business priorities and the costs allocated to business unit initiatives. However, the company, as a whole, will have higher costs due to natural inefficiencies related to the duplication of services.

A Shared Services Centre should combine ideally the advantages of the structures of the two worlds, the centralised and the decentralised. On the one side, this should result in economies of scale, scope and standardisation. On the other side, this should result in a flexible and effective alignment of IT with the needs of business. Additionally, synergy and mutual learning will increase, while SSC will provide a clear management focus. [9], [10]

2.3 E-Learning

E-Learning is one way for enterprises to improve their process of information flow for knowledge improvement and achievement. Its web-based system nature removes the users or learners time restrictions or geographic limitations. Moreover, availability and flexibility are often presented as advantages, when compared with traditional face-to-face systems. However, too many projects have high failure costs or users find them difficult to adopt.

E-Learning development followed two pathways. One pathway was as a distance learning tool and the other was as assisted computer learning. Common to the two

pathways is the need for technological support with bottlenecks in the communication networks (availability) and in the experience of use (usability). These two pathways are subsumed under E-Learning, as the Internet becomes the integrating technology.

This adds pressure in the development of the learning materials - a task, which is most of the time unplanned but which is fundamental for the learning experience that reflects in the usability. The Internet allows for the instant widespread dissemination of the learning content simultaneously and at anytime and anywhere.

An additional strength of E-Learning is its use as a tool for standardisation of content, when compared with site formation of the different sectors of one organization. E-Learning can have control points that make it possible to have assessment of user interactions and outcomes.

In these scenarios E-Learning has properties of potential usefulness that goes far beyond the delivery of learning. The technology of E-Learning has characteristics for evolving into a structure for supporting organization knowledge management. It allows learning by relating new knowledge with past experiences, through the linking of learning to needs, and then by practically applying the learning. This can potentially develop a more user-oriented and effective deployment of knowledge availability and the development experiences. The adoption of E-Learning for knowledge management develops an environment of interactivity and it promotes efficiency, motivation, cognitive effectiveness and flexibility of the learning style. [11]-[13]

2.4 Complex adaptive system (CAS)

The paradigm that has been used to model organization development has changed. This change can be found in a fundamental core of articles and books that deal with enterprise dynamics [14]-[15].

Now, organizations are centred on developing and maintaining dynamic core structures in order to support the unpredictability of change on the edge of chaos. However, for supporting this in a viable manner, that is, with efficiency and effectiveness, information systems become a core structure.

In this context, information systems, as a fundamental element supported within an attached reductionist vision, will never have relevance. A paradigm shift should be achieved and developed. Prosperity in this new world demands dynamics for complexity and CAS is a great way to provide those [16].

Complex adaptive systems are systems with a great number of components, sometimes called agents, that interact, adapt and learn and many contemporary problems are under the theory of a complex adaptive system [17].

The dependency of complex systems on initial conditions can be expressed in a very exponential level by the butterfly metaphor, whereby a butterfly flapping its wings in South America can affect the weather in Central Park [18]. Complex systems exist on the edge of chaos presenting a regular and predictable behaviour but, suddenly, they can start a mass change response to what can be seen as a minor change [19].

1 Research Design

Aims and objectives

The primary aim of this study is the evaluation of what can be obtained by using a SSC with E-learning for developing shared knowledge capabilities.

We develop this aim through a conceptual modelling of SSC and E-Learning sustained in the evaluation of the initial motives and expectations for adopting a SSC and E-Learning for knowledge management. This conceptual model evaluation results in the development of a mathematical deduction. This mathematical deduction validates whether knowledge increases or not, through the use of the conceptual model.

This study results into a conceptual model for sharing knowledge capabilities. Moreover, it has developed the identification of its knowledge sharing capabilities, through the measuring of its Δ (Delta) factor. This conceptual model is sustained under the concepts of SSC and E-Learning.

This analysis can be used by the organization to support a decision-making process related to the introduction of such a shared service centre. This research should contribute to the limited body of research on SSCs available and the use of E-Learning as a knowledge management tool.

This research focuses on the question “Is it possible to add value, through sharing knowledge capabilities by using a SSC and E-learning conceptual model for knowledge management?” We start by trying to define the Δ (Delta) factor for having a SSC for sharing knowledge. This Δ (Delta) factor will positively affect the management decisions regarding the use of Shared Services Centres with E-Learning for developing the sharing of knowledge capabilities.

Research methodology

Qualitative research can take a positivist, interpretive or critical approach. The present work will take an interpretive approach supported by a critical approach. Absence from the outset is that there is no hypothesis to confirm or refute that the positivist approach is not appropriate.

The methods used tend to be qualitative, interpretive and constructive but do not exclude any quantitative analysis. For triangulation and the ability to be used on occasion, a quantitative approach is considered.

For convenience our contacts and references are organizations at national and international levels. Four organizations were contacted - one in the area of postal services, another service in the area of banking and two in the national government and regional administration. It was considered important for the sample that interviews were conducted in organizations that acted in different branches of activity.

The collection of materials analysed for this study used qualitative methods, which were carried out using the techniques of documentary analysis, participant observation and interviews (unstructured or semi-structured).

From the point of view of information processing techniques, all materials collected in the interviews were recorded initially using manuscripts and then recorded in a

second step using audio recording. For the application of Grounded Theory several empirical materials and data were collected by using interviews in large Portuguese organizations. After the data was systematically collected it was coded according to the three encoding types: open, axial and selective.

From the selective coding emerged a set of concepts or classes that led to the theory presented here.

2 Conceptual Model

In a conceptual model for SSC use for knowledge management it is necessary to go beyond the process systematisation, standardisation and cost reduction that are the fundamental philosophical assumptions for SSC management. In this case SSC should developed unique and new combined knowledge emerging from the combination of sharable organizations. On one side, by adopting a SSC an organization has shareable knowledge available with potential for being integrated into other organizations. When this process occurs we have a process of sharing knowledge. On the other side, the SSC can combine different organizational memories and potentially develops new patterns. These two sides are conceptualized into the model of Fig. 1. This conceptual model describes how the SSC is a live, combined organizational memory and develops unique sharable combinations of knowledge using E-Learning communication flow.

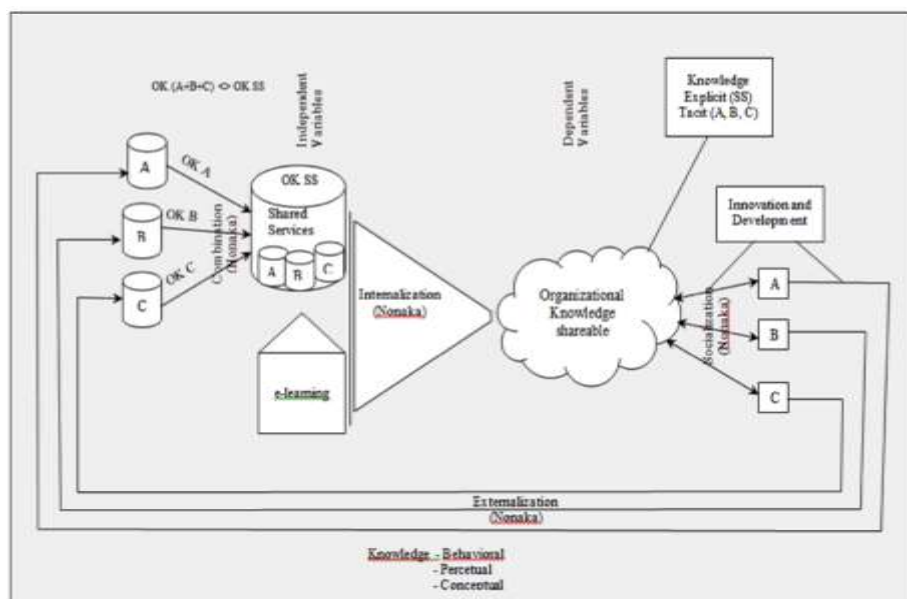


Fig. 1 - Conceptual modeling for knowledge management using SSC and E-learning.

In this conceptual model A, B and C represent organizations that use and integrate the shared services centre and where the shared services centre holds the larger part of the individual organizational knowledge. The resulting Organizational Knowledge of the Shared Service (OK SS) comes from the addition of the Organization Knowledge (OK) that is sharable in each organization. The organizational knowledge of shared services exists by itself and also in one of the independent organizational memories (OK). The total amount of available knowledge will always be different from that existing individually $[OK(A+B+C) <> OK SS]$. The E-Learning is the artefact of technological nature that allows the communication of the organizational knowledge and also part of the organizational memory. The property of internalisation of Nonaka,[14] [15], which is fundamental for learning, becomes a dependent variable from the independent variables of SSC and E-Learning. The organizational learning is then dependant on these variables and results for its combinatory development and communication of learning.

The shared organizational knowledge results from the composition of two types of knowledge. First, from individual organizational knowledge, which is available at the SSC and which can be used and integrated by other organizations. Second, from the set of organizational knowledge that emerges from the combination of individual organizational knowledge that is only made possible by its co-existence in the SSC.

The conceptual model reflects also the concepts [20] that organizational knowledge can be arranged into three large classes:

- Behavioural knowledge - characterized as know how;
- Perceptual knowledge - characterized by the knowing; and
- Conceptual knowledge - characterized by its use.

Moreover, the Delta Model assumes characteristics of crowdsourcing because it is able to make a contact between different agents who seek and consume the best solutions, also contributing to the decision taken by the organization. In this case, the organization is the Shared Services Centre and the public the various collaborators, framed in their organizations and spread over large geographical areas.

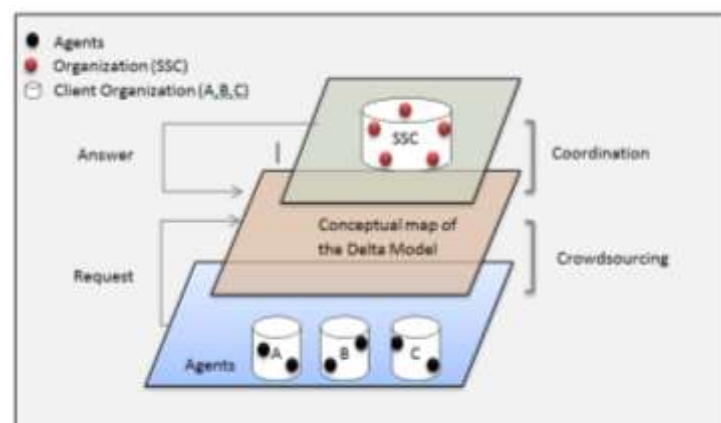


Fig. 2 – The Delta Model as a crowdsourcing solution. Based on Interagency map of Harnessing the Crowdsourcing Power of Social Media for Disaster Relief.

The Shared Services Centers can then take actions, share and receive information and coordinate with each other using the information on the map those results from applying the Delta Model.

3 Results

It is fundamental, due to the component of joining SSC and E-Learning that the developed conceptual model intends to add value, through demonstrating that there exists a possible Δ (Delta) factor. This Δ (Delta) factor result from the difference between every organizational memory that the SCC has and the resulting combined knowledge. The Δ (Delta) factor is described according to Equation 1. In Equation 1, K_{ssc} represents the organizational shareable knowledge resulting from the sum of all the different organizational memories.

$$\Delta = K_{ssc} - \sum_{i=1}^n K_i \quad (1)$$

Equation 2 describes the composition of the K_{ssc} . In Equation 2, K represents the amount of sharable knowledge; P represent the amount of perceptual knowledge; C represents the amount of conceptual knowledge; Ce represents the amount of explicit behavioural knowledge and K' represents the combination of knowledge. This combination results from the existing in the individual organizational memories of the SSC according to Equation 3.

$$K_{ssc} = \sum_{i=1}^n K_i (P + C + Ce) + K' \quad (2)$$

$$K' = C(K, p) \quad (3)$$

Equation 3 characterises K' . K' represents the patterns of knowledge that result from the combination, C , of the different knowledge, K , of different organizations, p . Replacing in Equation 1 the K_{ssc} and Kp by the equivalents in Equations 2 and 3, respectively, results in Equation 4.

$$\Delta = K' \quad (4)$$

This measure, the Δ (Delta) factor, allows the SSC manager to develop strategies for justifying and improving SSC. The Δ (Delta) factor determines that this is possible through the adding of the amount of diversity that exist at the SSC and the efficiency of E-Learning, as a communication for the learning tool.

It is assumed and demonstrated that the from the delta factor emerging from the application of the delta model is, thus, demonstrated the possibility of creating

knowledge using the application of this complex adaptive system. Alongside comes a theoretical reference model, which is represented by a set of classes for improvement, which should be applied in organizations.

From an analysis of the data, there emerges a set of classes for improvement, which serve as a reference for organizations to improve knowledge management and thereby promote their efficiency.

4 Conclusion

This conceptual model should be seen and evaluated, as a model focused on the management of organizational knowledge. When we talk about shared services, we focus on economy of scale, standardisation, re-engineering of processes and control. When we think about E-Learning after overcoming the convenient temptation to associate it with traditional education, we think in technological solutions for continuous learning.

With the present model and its adoption in a context of integrating SSC and E-Learning, we are able to integrate the process of organizational knowledge management. At the same time this incorporates the process of knowledge discovery that happens in first place, through the sharing of diversity and in second place, as a result of the explicit activities of standardisation and re-engineering.

It can also support the discovery of new knowledge by developing the potential of infinite growth, through the use of analytical tools that are used to discover patterns. This discovery often happens and is sustained in the model by the elicitation of knowledge, in particular behavioural knowledge, which by other ways would be no more than tacit knowledge.

Thereby the amount of knowledge that potentially exists in the knowledge bases and organizational memory of the SSC and which evolves continuously for the stated reason becomes sharable knowledge. This knowledge can then be used and re-used by all those intervening, thus sustaining innovation and promoting the development of the organization.

References

1. S. P. Bradley, J. A. Hausman, and R. A. Nolan, "Globalization, technology, and competition: the fusion of computers and telecommunications in the 1990s," p. 392, 1993.
2. T. H. Davenport, "Conhecimento empresarial," 1998.
3. T. H. Davenport and L. Prusak, "Working knowledge: How organizations manage what they know," 2000.
4. K. R. Popper, "Conhecimento objetivo: uma abordagem evolucionária," 1975.
5. K. Popper, "Three worlds," *bengin.net*, 1979.
6. K. R. Popper, "Truth, rationality and the growth of scientific knowledge," 1979.
7. M. Y. Santos and I. Ramos, "Business Intelligence: Tecnologias da informação na gestão de conhecimento," 2006.
8. R. Maier, "Knowledge management systems: information and communication technologies for knowledge management," 2004.

9. H. Strikwerda and J. Zee, "Fostering Execution-How to Leverage Leadership in a Confused Business Environment," 2003.
10. N. Su, R. Akkiraju, N. Nayak, and R. Goodwin, "Shared Services Transformation: Conceptualization and Valuation from the Perspective of Real Options," *Decision Sciences*, vol. 40, no. 3, pp. 381–402, Aug. 2009.
11. Y. Liu, Y. Bai, and Z. Chen, "Research on E-Learning Supporting System: A Review from the Perspective of Knowledge Management," presented at the E-Learning, E-Business, Enterprise Information Systems, and E-Government, 2009. IEEE '09. International Conference on, 2009, pp. 329–332.
12. N. Capuano, S. Miranda, F. Orciuoli, and S. Vassallo, "e-Learning at Work in the Knowledge Virtual Enterprise," presented at the Complex, Intelligent and Software Intensive Systems, 2008. CISIS 2008. International Conference on, 2008, pp. 507–512.
13. Z. Chunhua, "E-learning: The New Approach for Knowledge Management (KM)," presented at the Computer Science and Software Engineering, 2008 International Conference on, 2008, vol. 5, pp. 291–294.
14. I. Nonaka, "Perspective--Tacit Knowledge and Knowledge Conversion: Controversy and Advancement in Organizational Knowledge Creation Theory," *Organization Science*, 2009.
15. I. Nonaka and R. Toyama, "The knowledge-creating theory revisited: knowledge creation as a synthesizing process," *Knowledge Management Research & Practice*, vol. 1, no. 1, pp. 2–10, 2003.
16. Schneider, M., & Somers, M. (2006). Organizations as complex adaptive systems: Implications of Complexity Theory for leadership research. *The Leadership Quarterly*(17), 351–365.
17. Sutherland, J., & van den Heuvel, W. J. (2002). Enterprise application integration and complex adaptive systems. portal.acm.org.
18. Mazzocchi, F. (2008). Exceeding the limits of reductionism and determinism using complexity theory. EMBO reports.
19. Brownlee, J. (2007). Complex Adaptive Systems. Complex Intelligent Systems Laboratory.
20. J. A. Carvalho and M. P. Morais, "Sistemas Informáticos e Conhecimento Organizacional: Uma Reinterpretação dos Papeis Desempenhados pelos Sistemas Informáticos nas Organizações," pp. 1–16, Apr. 2004.

No Space for Emotional Support: Communities of Practice as Arenas of Cognitive Exchange

Azi Lev-On¹, Nili Steinfeld¹

¹Ariel University, Ariel, Israel
{azilevon,nilisteinfeld}@gmail.com

Abstract. This study examines whether interactions between members of communities of practice typically have cognitive or social character. Content analysis of more than 7000 posts, automatic words frequencies analyses as well as interviews with community members demonstrate that the interactions between members of the Israeli Ministry of Social Affairs' communities of practice, the subject of the present study, emphasize the cognitive rather than social aspects. This emphasis is reflected in the content of posts, the avoidance from discussing personal cases or offering emotional support and more. The findings are particularly interesting given the nature of these communities as a space for social workers whose work requires and is characterized by a high degree of social and emotional interactions.

Keywords: Online Communities of Practice, Content Analysis, Digital Government, Organizational Systems, Social Services, Cognitive Exchange.

1 Introduction

Knowledge has always been perceived as an essential resource for the progression and effectiveness of an organization, but in the past decades its importance grew even more, and it is perceived by many as a central mean for achieving competitive advantage and long-term success [1].

The field of knowledge management was traditionally dominated by the focus on technology, but there has been an increased understanding of the role of people in the process of knowledge management, and today it is commonly accepted that the key to successful knowledge management within organizations lies in the relationships between individuals, and an organization's ability to leverage its knowledge is largely dependent on the people who work within it and the flow of information between them [5]. Therefore, social relations between members are significant in understanding the success and operations of the community of practice.

Online communities of practice are groups of professionals who share a common interest in the examination of professional issues related to their work [9,14]. In some communities, the social aspect of communication between members is particularly important, and fulfills needs that are impossible to similarly satisfy in offline contexts. For example, people with unique characteristics, or those who have social difficulties,

adfa, p. 1, 2011.

© Springer-Verlag Berlin Heidelberg 2011

may find it difficult to create peer groups in their offline social environment. People who belong to marginalized groups, as well, typically seek connections to people similar to them, who may not always be found in their immediate environment. The Internet creates novel opportunities for social contacts in such cases [12]. Due to intensive and stressful work, both physically and emotionally, social workers often require support and sympathy [13]. A closed online group where colleagues and partners converge can function as place for ventilation, expression of support, solidarity, and creation of social ties among participants.

However, the literature about user behavior in online communities focuses less on the social aspects and more on the cognitive aspects of the use of the communities, such as the amount and quality of the information shared, information overload, uses of and gratifications from the information in the communities [6,8]. The focus on the cognitive aspects is not surprising since these are online communities of practice, i.e. groups of professionals who are interested in a joint examination of professional issues related to their work. For this they share professional knowledge, tools, resources and experiences. The amount of shared knowledge in communities, its flow and its uses are therefore important aspects of such environments.

The research on the cognitive aspects of online communities of practice focuses on the various motivations to contribute information [7,10,11,15]. Yet, although motivations to contribute information have been investigated in depth, there are no prospective studies that analyzed the content of communities of practice to learn whether cognitive uses are indeed central in terms of the content uploaded to the communities. This is the main contribution of the current study.

2 Research Environment and Methodology

This study analyzes the case of Israeli social workers' online communities of practice established by the Ministry of Social Affairs.

The research on communities of practice presents a unique case in Israel, where a government office established online discussion groups to allow interaction between office staff and the wider community of practitioners and professionals in related fields. Such communities may very well have advantages in terms of exposure of tacit and local knowledge, improving the flow of knowledge between professionals and even improving professional acquaintanceship and solidarity among workers [2].

The communities of practice project of the Ministry of Social Affairs was established in 2006, aiming to promote the development of methodologies of organizational learning among groups of social workers in the Israeli welfare system [4,14]. Since the establishment of the project, more than 7,700 members have joined one or more of 31 communities. Registration to communities requires the approval of the communities' managers. All communication within the communities is identified using the real names of the members. The list of members is available to all members of the community, and they know who might read the content they upload and comment on. Community management is voluntary, and the managers receive a small reward in the form of vouchers [14]. While the communities are hosted on governmental platforms,

less than 30% of their members are government employees, and the rest are employed in NGOs, municipalities and elsewhere [4].

This study used three research methods. First, content analysis of 11 communities of practice that were selected for the study was performed. The communities were chosen to constitute a representative sample of the various communities in the project in terms of the date of establishment (older communities vs. newer ones), the scope of activity within the community (measured by the percentage of active members out of all members of the community), the size of the community (measured by the number of members in the community) and the areas of practice of the community (therapeutic communities compared to communities engaged in formal issues and procedures). We categorized communities as "therapeutic" if the topic of the community relates directly to work with clients or specific populations that receive welfare services, and "non-therapeutic" if the topic of the community is more related to organizational needs, programs creation and general work related issues. Therefore, the communities included in the study were composed of two groups.

- The first group included communities that were categorized as therapeutic: Intellectual Disability, Children at Risk, Immigrants and Inter-Cultural Issues, Blind and the Visually Impaired, Domestic Violence, Foster care and Juvenile Delinquency.
- The second group included non-therapeutic communities, which are more organizational in nature (less concerned with clients and more concerned in estimations and programs creation): Community Work, Policy and Performance, Welfare Management at Municipalities and Organizational Learning.

In each community of practice all posts available at the time of data collection (beginning of 2012) were analyzed. A total of 7248 posts were coded using a coding sheet developed for the study.

The study involves two units of analysis: Single posts and threaded discussions (a first post and at least one additional comment related to it). Thus, some of the categories in the coding sheet relate to posts and others to discussions.

Altogether, the coding sheet comprised 24 quantitative categories, which included the identity of the author (manager or community member), time of publication, the name of the community where the post was published, the post's order (is it a first post, opening a new discussion or a response to a previous post), and a number of categories relevant to the content of the post, as will be described later on. The posts were coded by 13 coders after an intensive training and a reliability test, which repeated until an agreement rate of 90% between the coder and the leading researcher was obtained in each of the categories of the coding book. In addition, the first set of posts each coder has coded (usually 50-120 posts) were examined by the lead researcher, who went over the coding to make sure that reliability is maintained.

The rationale behind looking at first posts distinctively is the assumption that when community members initiate a new discussion, they dictate the topic of discussion. Reply to messages being raised in previous posts will mainly deal with the issues brought up in the first post. Therefore examining the first posts only allows us to study the issues that community members decide to initiate discussions around.

In particular, the following categories were used to code the content of messages:

- Practical advice, which is directly related to daily work with clients, for example: what is the impact of certain kinds of interventions?
- Organizational advice, related to employees' daily work unrelated to working with clients, for example concerning forms, procedures, programs and courses.
- Statements about the community's theme, which are statements that relate not to employees' daily work, but to more general issues related to the community's main theme, for example: How to improve service for patients? How to improve the status of blind people in the Israeli society?
- Emotional support- addressing community members' manifestations of charged emotions (anger, frustration, fear, sadness, etc.) that are related to their work.
- Additional categories were: academic advice (references to academic literature, relevant research etc.); informing on an event or conference; greetings and gratitude; publication of a project or organization, submitting contact details, and finally- other topics.

Another section in the coding sheet examined whether the message included sharing of personal experiences. It was also examined whether posts included provision and/or supply of assistance, and at the level of the entire thread, it was examined whether questions raised in the discussion were answered.

Secondly, word frequency analysis (conducted with Corsis, formerly Tenka Text, an open-source corpus analysis class library (Corsis),¹ and visualized using Wordle- an online open-source tag cloud generator (Wordle)²) was used for linguistic analysis of texts in the discussions of the various communities, to examine what are the main issues discussed within the communities and whether they stay "on-topic", dealing with the main topic of the community.

Finally, interviews were conducted with 71 community members. The interviewees selected were sampled based on the level of involvement measured by the number of logins to the community and the number of times they contributed content within the community, so that the study involved members with low, medium and high involvement in communities. The interviews were conducted by five interviewers across the country, and the average length of an interview was about 45 minutes.

3 Hypotheses

The communities of practice of the Ministry of Social Affairs is a professional project, established and managed by the Ministry, hosted on the Ministry's website and its stated purpose is to promote organizational learning processes among employees. Consequently, we expect that the perception of the project as will be reflected in the approaches of members of the communities and the content of their posts, would be that it is a place of cognitive interactions between colleagues, rather than a place of

¹ We thank the developer, Cetin Sert, for the use of the tool.

² We thank the developer, Jonathan Feinberg, for the use of the tool.

social, emotional or personal engagements, in spite of the social and even emotional character of the profession.

Accordingly, the following hypotheses are:

3.1 Subjects of Posts

H1. Posts will include mostly professional and organizational advice and involve less emotional support.

H2. Posts that open new discussions will focus mostly on organizational and professional advice and less on emotional support.

H3. In therapeutic communities, due to the nature of work in the community, and the mental challenges members of these communities are faced with, a higher percentage of posts will contain emotional support. Still, they will be less prominent than professional and organizational advice.

3.2 Subjects of Discussions

H4. The discourse in discussions of the various communities tends to remain within the practice field of the community (on-topic) and doesn't drift to other areas, personal or social. The common words in each community will be words relevant to the community's field of activity.

3.3 Sharing Personal Experiences

H5. Community members will rarely share personal experiences from their work. The subjects of posts will be more general and members will rarely discuss particular cases.

H6. This tendency will also characterize the first posts in discussions. Community members will rarely initiate discussions dealing with personal work cases.

H7. In therapeutic communities a higher percentage of messages containing sharing of personal cases will be found. But such posts will still be rare.

4 Findings

4.1 General Findings

The dataset includes 7,248 posts. 22.1% of the posts in the dataset contained messages requesting assistance from community members and 42.1% of the posts contained provision of assistance. This indicates that for any question, a few answers were typically provided. In more than half of the discussions (55.1%) a relevant response to a question was received during a discussion. When looking only at discussions in which questions were raised, in 76.7% a relevant response was given. This portrays the communities as having lively dynamics of knowledge circulation, where questions are raised and answers provided.

4.2 Subjects of Posts

What are the issues raised in discussions in the various communities of practice? Table 1 presents the percentage of posts containing reference to the ten topics included in the coding book. Note that many posts include a number of topics, therefore the percentages total to more than 100%.

The findings support hypotheses H1-H3.

Of all the posts in the sample, the main topic (found in 39.3% of the posts) is professional advice. The second most common topic (25.7%) is organizational advice. Third prominent (23%) was statements about the community's theme.

With the prominence of these issues in the posts, it is noticeable that emotional support is uncommon among community members' posts. Only 3.4% of the posts in the sample included emotional support.

Further analysis referred only to first posts that were published in the community. There are 1509 first posts in the sample. The findings indicate similarities between the first posts and all posts in the sample, where the two prominent topics in the first posts are professional advice (35.1%) and organizational advice (26%). The third most prominent topic among 16.8% of the first posts is a publication of a project or organization.

The percentage of the first posts that included emotional support is identical to the percentage of all posts in the sample that included this topic- 3.4%. This is the least common topic among all posts as well as among first posts.

Due to the therapeutic nature of some of the communities of practice, it can be expected that a higher percentage of messages containing emotional support will be found in communities dealing with more therapeutic areas in nature, in comparison to communities who's main concern is more organizational. In communities engaged in therapeutic areas, where workers are faced with complex personal cases, we would expect a higher demand from the community members for ventilation of feelings, sharing or leaning on other members' shoulders for support.

A chi-square test was performed to examine the relation between the type of community (therapeutic or non-therapeutic) and degree of emotional support in members' posts. The relation between these variables was significant ($\chi^2 = 22.35$, $p < 0.01$), supporting hypothesis H3. The effect size was calculated using Cramer's v and was found to be weak ($r = 0.06$). In non-therapeutic communities, only 1.6% of the posts included emotional support, while in therapeutic communities, 4% of the posts included such aspect. Still, the percentage of only 4% within therapeutic communities strengthens the conclusion that emotional support is rarely present in posts published in the communities.

Table 1. Distribution of the posts' and first posts' subjects in the sample

Topic of post	% of all posts	% of first posts
Practical advice	39.3%	35.1%

Organizational advice	25.7%	26%
Academic advice	6.4%	10.4%
Emotional support	3.4%	3.4%
Informing on an event or conference	8.1%	12.4%
Greetings and gratitude	10.9%	3.7%
Publication of a project or organization	8.5%	16.8%
Submitting contact details	7.6%	9.3%
Statements about the community's theme	23%	12.5%
Other topics	8.6%	7%

Interviews with community members complement the findings suggesting that the main use of the community is cognitive, and not social or emotional. Interviewees emphasized the exchange of knowledge within the framework of the discussions in the communities as the most fundamental aspect to them, an exchange of knowledge which allows them to learn from the experience of others. An additional cognitive use that was brought up is stimulation of creativity and original thought, with group discussions helping to come up with creative ways to solve professional problems.

Interviewees also see differences between communities in terms of emotional support, but insist that emotional support is not a main function of most communities. Interviews revealed differences between respondents about the role of the communities as a place of support among members. One respondent expressed a firm stand regarding emotional support:

"Not in any form! It should be something unmediated-by personal conversations and not by interactive means. It gets lost through the computer [...] something emotional has to be personal and not interactive".

Another member sees emotional support in contrast to the professional goals of the communities:

"Communities need to be professional, and if it becomes the source of support-that's a sign that the situation is difficult..."

On the other hand, other members see the importance of support within the community:

"It's ventilation of emotions... Someone brings up something that happened to him [...] and then it triggers reactions and sharing around emotions. Or someone who was attacked [...] she was reinforced by all of us and... A lot of support".

However, the opinion expressed in the last quote does not represent the majority of interviewees, and indeed we see that in some cases, and especially in therapeutic communities, there was a small degree of emotional support in the posts analyzed, but

these were rare in comparison to posts discussing professional or organizational matters.

4.3 Subjects of Discussions

Word frequency analysis was used to determine the most common words in each of the communities.³ The word clouds below show examples of the most frequent words that appeared in all posts published in the communities analyzed, with stop-words in Hebrew (such as "I", "it", "not", "of" etc.) excluded as customary. The size of the word indicates the extent of its incidence in the texts, so that the most common words are also the largest. Data from the analysis software, Corsis, includes information on the frequency and percentage of each word in the text. The analysis clearly shows that the texts that comprise the posts tend to deal with issues relevant to the community, supporting hypothesis H4:

1. In the **Foster Care** community the words "the child" (appears 123 times in the texts, 0.66% of the text) or "children" (51 times, 0.27% of the text), "the family" (65 times, 0.35% of the text, and "family"- 51 times, 0.27%), "the foster" (86 times, 0.46%, and "foster"- 81 times, 0.44%, or "in foster"- 59 times, 0.32%, and "foster care taker"- 53 times, 0.29%), "adoption" (48 times, 0.26%), "the guardian" (43 times, 0.23%) and "the parents" (41 times, 0.22%) were the most frequent words in the text.
2. In the **Domestic Violence** community the most common words were "violence" (appears 910 times in the texts, 0.62% of the text, or "the violence"- 344 times, 0.23%, and "in violence"- 325 times, 0.22%), "women" (732 times, 0.50%), "for therapy" (380 times, 0.26%, or "therapy"- 280 times, 0.19% of the text) and "in the family" (349 times, 0.24%).
3. In the **Children at Risk** community common words were variations of "children" (such as "children"- 108 times, 0.43%, "the child"- 60 times, 0.24%, "the children"- 50 times, 0.20%, "for children"- 47 times, 0.19%, "child"- 43 times, 0.17% etc.) and "parents" (such as "parents"- 56 times, 0.22%, "the parents"- 51 times, 0.20%, "the family"- 44 times, 0.17%).
4. The community **Policy and Performance** is characterized by a high frequency of the words "the welfare" (27 times, 0.28%) and "organization" in various forms (such as "organizational"- 30 times, 0.31%, "organizations"- 18 times, 0.19%, "the office"- 16 times, 0.17%, "organization"- 13 times, 0.13%).
5. In the **Blind and the Visually Impaired** community especially prevalent are variations of the word "blind" ("for the blind"- 34 times, 0.25%, "blind" in plural form- 28 times, 0.21%, "the blind" in plural form- 22 times, 0.16%,), variations of "community" ("the community"- 23 times, 0.17%, "in the community"- 18 times, 0.13%,) and "the service" (25 times, 0.19%).

³ The texts are originally in Hebrew and the frequent words were translated to English by the researchers.

6. In the community **Organizational Learning** most common were variations of the word "knowledge" ("the knowledge"- 117 times, 0.52%, "knowledge"- 109 times, 0.48%,), "learning" ("the learning"- 55 times, 0.24%, "learning"- 48 times, 0.21%), and "community" ("the community"- 92 times, 0.41%, "communities"- 55 times, 0.24%, "in the community"- 44 times, 0.20%, "for the community"- 43 times, 0.19%, "community"- 38 times, 0.17% etc).
7. The community **Welfare Management at Municipalities** was characterized by the prevalence of the word "welfare" (51 times, 0.51%), variations of the word "manager" ("manager"- 22 times, 0.22%, "the managers"- 17 times, 0.17%, "managerial"- 14 times, 0.14%), and the words "social" (22 times, 0.22%), "services" (21 times, 0.21%) and "departments" (17 times, 0.17%).
8. In the **Community Work** community most common were variations of the words "work" ("work"- 229 times, 0.26%, "for work"- 196 times, 0.22%, "the work"- 160 times, 0.18%, "working"- 155 times, 0.17%, "the workers"- 114 times, 0.13%), "community" ("community" (female)- 428 times, 0.48%, "community" (male)- 200 times, 0.22%, "community" (plural)- 140 times, 0.16%, "the community"- 104 times, 0.12% etc), "activists" (119 times, 0.13%), and it is interesting to note the frequency of the words "thank you" in this community (241 times, 0.27%).
9. The community **Juvenile Delinquency** is characterized by the prevalence of variations of the word "youth" ("The youngster"- 236 times, 0.24%, "youth"- 140 times, 0.14% "the youth"- 138 times, 0.14%, "teens"- 113 times, 0.11% etc), and "probation" ("the probation"- 398 times, 0.40%, "probation"- 218 times, 0.22%), probably as part of the pair of words "probation officer", while the words "officer" and "officers" were also frequent- 143 times and 104 times respectively".
10. In the **Immigrants and Inter-Cultural Issues** community, the most common words were "immigrants" (11 times, 0.39%), "Ethiopia" (10 times, 0.35%), "in Israel" (8 times, 0.28%), variations of the word "social" ("the social"- 8 times, 0.28%, "social" (plural)- 5 times, 0.18%, "and the social"- 4 times, 0.14%), and it is interesting to note the frequency of the words "good luck" in this community (9 times, 0.32%).
11. The community **Intellectual Disability** was characterized by the prevalence of the words "retardation" (545 times, 0.41%), "mental" (394 times, 0.30%) "in the community" (261 times, 0.20%), "the tenants" (220 times, 0.17%, or "tenants"- 158 times, 0.12%), "at the framework" (201 times, 0.15%) and "at the center" (179 times, 0.14%).



Foster Care



Community Work

Fig. 1. Word clouds of the 100 most common words in the communities *Foster Care* and *Community Work* (Source: Wordle).

4.4 Sharing Personal Experiences

The findings support hypotheses H5-H7:

The communities of practice at the Ministry of Social Affairs were created to serve as a safe space to raise professional dilemmas, and as a platform for discussions on methods of conduct that should be performed with different clients. However, it appears that the members of the communities rarely share information related to personal cases they encounter in their work. Among all posts in the sample, only 6.6% share a personal case related to work. A slightly higher percentage can be found in first posts- 9.7% include sharing of a personal case, but this is still a fairly low percentage considering that these are communities of practice of social workers.

In this case as well, it is natural to assume that therapeutic communities will include a higher percentage of messages containing sharing of personal cases than communities revolving around issues that are more organizational or principle in nature. Therefore, a chi-square test was performed to examine the relation between the type of community (therapeutic or non-therapeutic) and degree of sharing personal cases in posts published in the community. The relation between these variables was significant ($\chi^2 = 7.39$, $p < 0.01$). The effect size was calculated using Cramer's v and was found to be weak ($r = 0.03$). In non-therapeutic communities, 5.1% of the posts included sharing of a personal case, while in therapeutic communities, 7% of the posts included this aspect. Still, 7% is a pretty low degree of sharing when dealing with communities of practice of care-givers in the social services.

5 Discussion

The findings clearly indicate that the communities of practice are first and foremost a place of professional advice, exchange of knowledge and a place of general discussion of organizational, practical and general community-related issues. The study suggests that the discourse within communities of practice tends to remain within the occupational field of the community-as expressed in the analysis of the most common words in each community.

Despite the importance of social interactions for the success of online communities, as indicated by the literature, analysis of the content of the posts in the communities suggests that community members deal primarily with practical and organizational issues, both when they are initiating new discussions and when they are participating in existing discussions, and they rarely deal with more personal or emotional issues. Sharing personal experiences and cases is also quite rare.

The interviews with community members as well as content analysis of the posts show that communities are perceived and used primarily as a space of cognitive and professional interactions, and as an organizational learning platform. Although these are communities of practice that serve the employees of the Ministry of Social Affairs, which by virtue of their role deal with emotional challenges and complex social situations which obviously can lead sometimes to distress, it is clear that they do not see these spaces as an appropriate place for emotional support or social interactions. Even within communities dealing with therapeutic issues they do not often turn to

community members for assistance, sharing or reinforcement on a personal level, and from the interviews it is also apparent that the idea of these communities being a platform for sharing and support was seen by many of the members as misplaced. When members need a comforting shoulder, they will choose other platforms for this purpose and maintain the communities of practice as areas of professional discussions among colleagues.

References

1. Castells, M.: *The Rise of the Network Society: The Information Age: Economy, Society and Culture*. Blackwell, Malden (1996)
2. Cook-Craig, P.G., Sabah, Y.: The role of virtual communities of practice in supporting collaborative learning among social workers. *Brit. J. Soc. Work.* 39, 725-739 (2009)
3. Corsis, <http://sourceforge.net/projects/corsis/>
4. Fein, T.: *Online Communities of Practice in the Social Services* (in Hebrew). MA Thesis. The Hebrew University. Israel (2011)
5. Ipe, M.: Knowledge sharing in organizations: A conceptual framework. *Hum. . Resource Dev. Rev.* 2, 337-359 (2003)
6. Johnson, C.M.: A survey of current research on online communities of practice. *Internet. High. educ.* 4, 45-60 (2001)
7. Kankanhalli, A., Tan, B.C., Wei, K.K.: Contributing knowledge to electronic knowledge repositories: An empirical investigation. *MIS. Quart.* 29, 113-143 (2005)
8. Kraut, R., Wang, X., Butler, B., Joyce, E., Burke, M.: Beyond information: Developing the relationship between the individual and the group in online communities. *Inf. Syst. Res.* 10 (2010)
9. Lave, J., Wenger, E.: *Situated Learning: Legitimate Peripheral Participation*, Cambridge University Press, Cambridge (1991)
10. Lev-on, A., Hardin, R.: Internet-based collaborations and their political significance. *Journal of Information Technology & Politics* 4, 5-27 (2007)
11. Ma, M., Agarwal, R.: Through a glass darkly: Information technology design, identity verification, and knowledge contribution in online communities. *Inform. Syst. Res.* 18, 42-67 (2007)
12. McKenna, K.Y., Green, A.S.: Virtual group dynamics. *Group. Dyn.-Theory Res. Pract.* 18, 116-127 (2002)
13. Meier, A.: An online stress management support group for social workers. *J. of Technology in Human Services.* 20, 107-132 (2002)
14. Sabah, Y.: *Online learning groups and communities of practice in the social services*. *Society and Welfare* (in Hebrew). 30, 111-130 (2010)
15. Wasko, M.M., Faraj, S.: Why should I share? Examining knowledge contribution in electronic networks of practice. *MIS. Quart.* 29, 1-23 (2005)
16. Wordle, <http://www.wordle.net/>

Knowledge Extraction and Annotation Tools to Support Creativity in the Initial Stage of Product Design: Requirements and Assessment

Julia Kantorovitch¹, Ilkka Niskanen¹, Anastasios Zafeiropoulos², Aggelos Liapis²
Jose Miguel Garrido Gonzalez³, and Enrico Motta⁴

¹ VTT – Technical Center of Finland, Vuorimiehentie 3, Espoo, Finland
{julia.kantorovitch, ilkka.niskanen@vtt.fi}

² Intrasoft International, Markopoulou-Peania Avenue, Athens, Greece
{anastasios.zafeiropoulos, aggelos.liapis }@intrasoft-intl.com

³ Atos, Albarracín, Madrid, Spain
jose.garridog@atos.net

⁴ The Open University, Walton Hall, Milton Keynes, UK
enrico.motta@open.ac.uk

Abstract. The initial stage of the conceptual product design is the most creative and knowledge-exploration intensive process. To facilitate design process and knowledge exploration tasks the availability of annotated resources is essential which set the requirements to the knowledge extraction tools. In this paper, the current state of this technology is assessed against the demand identified to support designer work.

Keywords: Conceptual product design, creativity, knowledge extraction and annotation, semantic technology

1 Introduction

Over the years semantic technologies and tools has shown significant development. The value of these technologies has been demonstrated in several application fields like healthcare, logistics, retail, transport and finance [6]. Other field which may potentially benefit from modelling, reasoning and decision support ability of semantic technology is the knowledge-intensive process of product design in particular considering the initial stage of design. During the early stages of the product design process, whenever the question is about medical device, coffee machine or web interface, designers rely on existing practices and resources available from a local company's databases such as documents and sketches produced in the course of previous designs, and on external information sources to help in the generation of new ideas. The external resources may include electronic books, images, music, online design journals and sources of background and domain specific information. During

the initial stage of the design process designers are often required to solve ill-defined problems. This means that the requirements and the needs to be addressed are not completely known. There is no definitive solution to the problem. Different solutions can be equally valid as response to the initial problem. There is no objective true-or-false evaluation of a solution [4]. However solutions can be assessed as good or bad, appropriate, innovative, creative, and so on, among others in the process of collaboration. The designers therefore need to go through the further process of data gathering and exploration to clarify the vision they have in mind to deconstruct and specify the problem and solution for the task in hand. Accordingly the initial stage of the conceptual design is the most creative and knowledge-exploration intensive process. Extending French's design model [13] the role of knowledge exploration in the overall process of conceptual design is framed in Fig.1. As the designer is proceeding from the stage of problem analysis to working conceptual design, the knowledge search space is narrowing and becoming more structured and domain specific. To facilitate the process of knowledge exploration by designers, the availability of large quantities of annotated resources is essential.

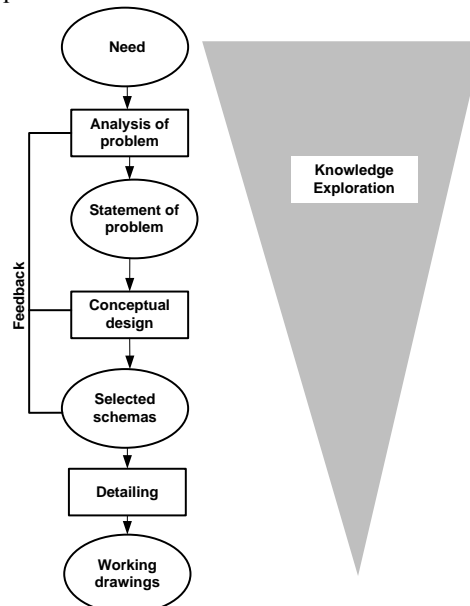


Fig.1. French model of design process extended with knowledge exploration process

Methods for knowledge exploration based on semantic annotation using ontologies are recognised as a powerful approach, which can make the processing of information resources more “intelligent” i.e. machine interoperable, effective and meaningful [2], [14], [17], [18]. Ontologies may provide elegant mechanisms to organise multimedia content in logically contained groups while linking them with other distant but still related concepts [16]. Annotations with well-defined semantics are a requirement to ensure the interoperability of available information supporting knowledge sharing and collaboration across design teams. Semantic technology based annotation is potentially beneficial enhancing the Information Retrieval (IR) or the ability to

perform effective customised searches exploring the knowledge contained within annotations in order to access the heterogeneous content. Recently introduced Linked Data technology [1] has the potential to facilitate the interlinking of unconnected documents and other data sources (e.g. internal enterprise data assets) to more interlinked ecosystems of data.

While knowledge extraction (KE) and semantic annotation research has been undertaken for several decades and an extensive number of various tools and frameworks have resulted, our analysis shows that there is still effort required before semantic tools will be able to address the knowledge exploitation needs of designers towards the collaboration and creativity support. While semantic approaches are potentially promising, we argue that future developments to be put to address particular use cases, application tasks and user groups.

In the following section, the identified requirements for knowledge extraction and annotation which are essential from the perspective of applications in order to support design teams are discussed. The identified requirements are utilised as a basis for the assessment of semantic tools which is presented in the section 3. Conclusions are presented in section 4.

2 Requirements to Knowledge Extraction and Annotation tools

The requirements has been acquired analyzing the dimensions of creativity, performing literature review focusing on what is needed to support the collaboration and creativity across the design team as well as utilizing an extensive discussion with product designers from CONCEPT consortium¹. This has been an iterative process. Attention has been paid to quality of extraction and annotation means, type of resources to be annotated (web based resources and company's local databases/resources (images, documents, sounds, articles, sketches, etc.)), and the types of annotation that might be required to support the designer e.g. visual features (colour, shape, texture) and content level (tags, location, description/meaning...), functional (who, where, who to access) as well as the format of annotations and requirements for the interoperability of the system. The requirements have also been assessed analysing the role of the designer in the knowledge extraction process and the possible support that might be needed. The compatibility with existing resources on the internet (multimedia databases, virtual museums, etc.) as well as support needed for the porting of annotation tools to the existing company databases, has been considered. Accordingly several main requirements are identified and discussed in the following.

2.1 Versatility and conceptual expressivity of semantic annotations

When considering products and the role of the designer in its creation, it is useful to refer to Offenbach's Theory of product language. Offenbach makes the distinction between a product's function (i.e. functions and services) and the product's semantics

¹ <http://www.concept-fp7.eu/>

(product language), which refers to the way in which the product is understood by users (see Fig. 2).

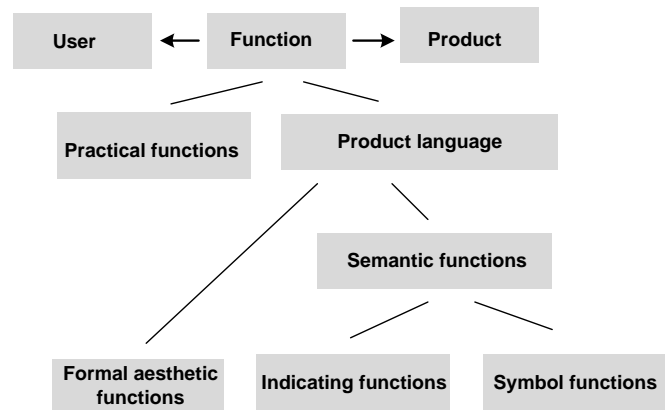


Fig.2. Conceptual model of Offenbach Theory of Product Language [21]

Looking further on the product language, the formal aesthetic functions of an object are referred to as the “grammar” of the design concept, which includes shapes, colour, texture, material etc. The indicating functions indicate the way in which the product could be used. Symbolic functions refer to impressions and associations within the mind of the observer (designer and end user) using the product, for example historical, economic aspects as well as associations such as cold/warm, strong/weak. From Offenbach’s point of view, the product language and the communication of the product’s message and formal meaning is considered to be within the domain of the designer, while engineers, ergonomists, etc. are responsible for the product’s practical aspects (in other words, technical and technological functions) [21]. New technology can provide a source of innovation for new products. An innovation in product language means the creation of semantically coherent product concepts where product technical functions are framed by communicative functions of product. The design of a semantically coherent product concept is the great source of innovation where the creativity is most needed.

In order to support innovation, to help designers access useful resources for their creative tasks, these resources need to be identified, which may include the identification of properties for the product acquired in requirements phase, which can be potentially mapped to some visual resources, styles, using for example, colour, shape or behaviour and any other properties discussed above. The mapping and ‘context’-aware knowledge delivery indicates the level of semantic annotation required and consequently the knowledge extraction technology used by the system. Levels of expression are achieved by introducing a versatility of concepts and richness of relationships between them. Moreover problem formalization requires changing the context and viewpoints surrounding the problem in hand, where the associated knowledge plays a critical role. Knowledge organization and annotation patterns should be able to make the information search and delivery in a context-

aware and effective way, supporting the designers whilst considering the task in hand from alternative perspectives.

2.2 Support for various types of knowledge extraction

The wide variety of information available on the web and locally in company's repositories can spark creativity and support collaboration. Required information can be available as part of text embedded in documents, or delivered by image or sketch or music lyric or composed of multimedia content of various types. Knowledge extraction tools may provide the required information using subject-based metadata of documents as well as the metadata for the multimedia content such as images (key words, tags and also low level visual properties).

Looking at a typical model for the creative design, this process consists of several iterative steps such as incubation, creative insight and evaluation [5]. During the incubation phase, association, analogy, metaphors, etc., are techniques which are often used to facilitate the process of creative thinking and inspiration later transferred into new product design. Metaphor is an association between two concepts or artefacts when their properties are common. Analogous thinking is often triggered by visual stimulation using images to stimulate creativity. Many studies have demonstrated that designers make intensive use of analogies, adapting design features from other fields to their own design problem [19]. Moreover it was shown that often the most creative analogies are those that are made between the most distant domains [7], [20]. Because multimedia content based knowledge extraction plays an important role in creativity support, this sets the requirement to multimedia knowledge extraction and annotation tools to bring together low-level visual features and high level meaning of multimedia objects.

The system should also consider the location of required content which may reside in enterprise databases, personal collections and catalogues as well as existing on the Web. Ideally the knowledge extraction and annotation tools should cope with both situations.

2.3 Open standards

Any IT and in particular a knowledge management system can only benefit by making use of open standards as far as possible. Open standards facilitate the porting of various applications to knowledge extraction platforms where guidance, rules and interfaces are known. Moreover the integration of various components developed by different parties (e.g. IE, ontology and annotation management) is easier resulting in smooth interoperability with one another. For annotation systems in particular, open standards may efficiently bridge heterogeneous resources to be accessed by collaborating users and organizations to share annotations [22]. It's important for semantic knowledge extraction tools to be compatible with the most used open Semantic Web standards such as RDF and OWL (e.g. OWL Lite) metadata model and RDF schema as a language for knowledge representation, SPARQL as a knowledge query language, FOAF-like ontologies to create relationships between described

resources and Linked Data concepts as a “control vocabularies” (i.e. collection of classes and properties used to describe resources in a particular domain) if available to tag resources unambiguously.

Knowledge extraction components and extensions should be preferably developed based on such programming language that will ensuring the usability of the same software in different environments as well as being compliant with widely used middleware and communication standards such as Web Services, REST-like interfaces and data formats.

2.4 Support for the extendability of semantic concepts

While the product design from a process point of view can be considered as domain independent, from a knowledge management perspective, there are still aspects that need to be considered such as: various creativity techniques can be applied which may require different knowledge organization patterns available; various aspects in the product design domain may require a differing depth of knowledge (marketing, technology, application domain) and accordingly making use of different knowledge bases; the product being designed may belong to different application domains (e.g. sport, health, work, leisure), which would require domain specific ‘control vocabularies’ to ensure sharing the understanding of domain concepts.

It is obvious that knowledge extraction tools can’t address all the needs of product designers for various application domains, thus a knowledge extraction and annotation frameworks should support clear separation between domain/task independent and domain/task specific knowledge and to allow easy-to-use creation of knowledge bases and patterns towards task specific needs (e.g. facilitate different existing creativity-supporting techniques from knowledge management point of view).

2.5 Support for the automation of knowledge extraction and annotation

In general, automatic approaches in knowledge extraction and annotation are desirable compare to manual, as they promise scalability to annotate the considerable amount of Web resources and organisation’s databases. Once they are available, they are extremely useful to the designer. Resources such as documents and various types of multimedia content can be used by the system. Thus we anticipate that the applications to support the designer work will need various types of annotation (technical, content- and visual-related). While automatic document annotation systems are more mature in their state of development, the image annotation methods remain a challenge for many real world applications [3], [8], [9].

In addition, we can’t expect designers to expend effort on the annotation of material they use during the design process, unless of course a natural and intuitive interface is offered. Therefore to become useful, the annotation tools should provide as much as possible a degree of automation for knowledge extraction in order to reduce the burden of annotation of any resources in use.

2.6 Usability and collaboration support

Taking into account that end users are designers who do may not necessarily have a 'technical' background or in-depth experience of knowledge management systems, the usability aspect which is related to human involvement in the generation of semantic metadata, is an important consideration in the design of the user interface.

Ideally the complexity of creating semantic data should be hidden behind an intuitive user interface and the vocabularies used for annotation as much as possible should be naturally aligned with designers' world. In addition the user interface technology should be designed to be portable and interoperable with existing web based interface standard (such as HTML5, JavaScript, etc.) and web applications. As regards to offering annotation concepts, some learning features might be useful to facilitate the suggestion of earlier used concepts by the designer, and aligned with widely agreed categories such as for example defined by open LD community.

The design and deployment of advanced collaborative functions is considered crucial to support creativity. Knowledge sharing and re-use of existing content among the product designers has to be based on interactions between individuals that work on a specific project. Interaction between designers making use of shared design vocabularies supported by KE technology should be carefully considered. In addition interconnection with available resources or relevant ideas online as well as options for tools that may be used per case and ensuring seamless activities flow (e.g. collaborative editing and content annotation, content synchronization etc.) should be empowered through the application of a semantic backbone.

3 Discussion

Semantic knowledge extraction and annotation tools have been searched through the literature in several domains including knowledge management, information retrieval, and semantic resource annotations using the digital libraries of ACM, IEEE and Springer. Analysing designer' needs, several KE systems have been found to be the most relevant and potentially promising candidates for further assessment. A selection was made based on the initial test experience and reported functionalities to support knowledge management tasks. Apart from functions provided by tools, the ease of installation, and the availability of suitable APIs in addition to web applications, as well as suitable licencing options and active development community around tools have been considered.

Various simulated design-briefs, HTML pages and documents are used to assess the tools. The results of the assessment in relation to the requirements discussed early are summarised in the Table 1. The Quality Function Deployment (QFD) concept is used to express the overall impression about how the requirements are supported in each particular case [11]. According QFD model, in the process of defining the product concept the design requirements (WHAT's) served as input to establish components characteristics (HOW's) of product design. Then HOW MUCH represents the amount that can be achieved. Adopting the QFD approach, in our case, WHAT's refers to the requirements specified in section 2 and the HOW's are the

knowledge extraction and annotation tools that have been assessed. HOW MUCH is rated as the following: **S** – requirement is Strongly supported; **M** - requirement is Moderately supported; **W** - requirement is Weakly supported; **N** - requirement is NOT supported.

Table1. The assessment of tools against requirements: 2.1-conceptual expressivity; 2.2-heterogeneous content; 2.3-open standards; 2.4-extendability of semantic concepts; 2.5-automatic knowledge extraction; 2.6-usability and collaboration

KE tools	2.1	2.2	2.3	2.4	2.5	2.6
LMF ²	M	M	S	M	M	M
Stanbol ³	M	M	S	M	M	W
KIM ⁴	S	M	S	M	S	W
OpenCyc ⁵	M	W	M	W	M	N
OpenCalais ⁶	S	W	M	N	M	M
Pundit ⁷	S	M	S	M	M	M
GoNTogle ⁸	M	W	S	M	S	W
Annnotation ⁹	M	W	M	M	M	W
SugerTube ¹⁰	W	M	M	M	W	W
Annotator ¹¹	S	W	S	M	S	W
DBPediaSpotlight ¹²	S	W	S	M	S	W
TextRazor ¹³	S	W	S	M	S	W
FRED ¹⁴	M	W	S	M	S	N
Zemanta ¹⁵	M	W	S	M	S	N

² <http://code.google.com/p/lmf/>

³ <http://dev.iks-project.eu:8081/>

⁴ <http://www.ontotext.com/kim>

⁵ <http://www.cyc.com/platform/opencyc>

⁶ <http://www.opencalais.com>

⁷ <https://thepund.it/>

⁸ <http://web.imis.athena-innovation.gr/projects/gontogle/>

⁹ <http://annnotation.open.ac.uk/annnotation>

¹⁰ <http://sugartube.open.ac.uk/>

¹¹ <http://annotatorjs.org>

¹² <https://github.com/dbpedia-spotlight/dbpedia-spotlight/wiki>

¹³ <https://www.textrazor.com/>

¹⁴ <http://wit.istc.cnr.it/stlab-tools/fred/>

¹⁵ <http://www.zemanta.com/>

It can be seen from the table that most of tools being compliant with open and widely used standards (2.3). Technically speaking most of the tools provide a support for Web Service and REST-like interfaces and data formats as well as RDF/OWL and linked data to frame KE in an unambiguous manner. The exceptions are only the KIM and the OpenCyc tools. In addition, in spite of conceptual richness and expressiveness of annotations, the encyclopaedic ambitions of the OpenCyc knowledge base and the fact that it can only be extended by hand, increases the complexity of the system (2.4; 2.6).

It was found that many tools also fail to support collaborative activity (i.e. reuse and sharing) in the process of knowledge extraction and annotation and the usability of knowledge extraction and annotation tools is the area which is still in infancy (2.6). The available technology is mostly developed keeping in mind application developer needs or in the case of semi-automatic annotation, providing support for individuals performing web authoring tasks or simple tag based functions in a non-professional context. Considering that product designers are not data exploitation professionals or either application developers, existing user interfaces need considerable effort to hide the complexity of existing knowledge extraction technology. Moreover many tools are stand-alone technology. Their integration with popular content management platforms, such as Drupal or WordPress, would facilitate the collaboration and sharing content among designers (2.6).

As evidenced by the assessment undertaken for this report the support for heterogeneous resource formats is not as developed as one may anticipate (2.2). A majority of tools focuses on knowledge extraction from text. This area of technology is more advanced compared with multimedia content processing. This is in line with other analyses made earlier [9]. Whilst prototype deployments as well as research initiatives are available and a plethora of image and video retrieval tools are also released, the process of information extraction and annotation is still much reliant on human involvement, and is semi-automatic in nature (2.5) while the support for annotation interoperability utilising ontologies and open linked data is widely provided [12], [15]. On the other hand, it should be acknowledged that these tools have facilitated the providers and community of web based online multimedia collections with the ability to semantically annotate their content (e.g. Europeana¹⁶), hereby enhancing knowledge extraction outcome.

Overall, keeping in mind that knowledge extraction from text is much more mature technology, it's reasonable to take this into consideration in developing applications to support collaboration and creativity in particular. It's not a surprise that during the past years a number of so called Semantic Web start-ups such as Zemanta, OpenCalais and others have emerged to take advantage of knowledge extraction technology from the text, and provide APIs to enable bloggers and businesses to link writing text with publicly available content.

From the tools considered, only LMF framework addresses the needs for multimedia based knowledge extraction and annotation extending the functionality developed in Stanbol. It was found that the LMF platform development was an attractive technology because of its interoperability and content interlinking features

¹⁶ <http://www.europeana.eu/>

and ease of integration of knowledge management functions with enterprise databases. Like LMF, Stanbol allows for the production of KE and annotation results in different output formats including RDF and JSON. An additional advantage of Stanbol is that its components are already integrated with CMS systems such as Drupal which is popular content management platform. However, compared to LMF, Stanbol does not support annotation of multimedia content. Perhaps the biggest deficiencies of both tools are related to their complexity and their ability to analyze text fragments. In order to efficiently exploit the functionalities of the systems, developers have to simultaneously understand many technologies including OSGI and RDF. Additionally, the ability to configure the frameworks to better serve domain specific use cases (2.4) requires versatile skills from developers. Finally, semantic content processing in LMF and Stanbol is still behind some of more established text analysis tools [10]. Whilst LMF is more advanced in the knowledge extraction and annotation from multimedia, still images and videos can be automatically annotated with only technical information (date, owner, etc.) (2.5).

TextRazor and DBPedia Spotlight are light and an efficient solution for Entity Recognition from text. While it's not open source software, Zemanta and TagIt are well developed and very inspiring services. Providing portable and interoperable APIs, tools can potentially facilitate the content enrichment functions. It should be mentioned that FRED is the only tool able to automatically produce RDF/OWL ontologies and link data from natural language sentences. It can be potentially used to facilitate the analysis of design-briefs or to support the creation of control vocabularies as well as to enrich syntactic keyword based searches that may be used by designers. These would however require an additional effort to extend the existing knowledge base with domain specific knowledge and ontologies as well as with reasoning functions depending on objectives of use cases targeted by application and in particular, user interface and related knowledge components which will be required in order to address the needs of designers.

The brainstorming sessions with designers from CONCEPT consortium and literature review [20] have revealed that designers are the active users of classical web and such tools as Google, Getty, and Flickr and specific design portal such as Fotolia, Corbis as well as mood boards (e.g. Pinterest) and other online resources (e.g. co.design, yatzer, designboom, designobserver, etc.) are used by designers on daily basis to search for new product concepts and inspiration. In addition, to support application developer work, various APIs are provided to utilize the content of the web. For example, with recent extensions, Google APIs allow for the extraction of images based on various parameters which relate to image features such as colour or face recognition as well as retrieved from particular web sites¹⁷. Flickr APIs¹⁸ can be used through REST or Web Service interfaces and public photos can be extracted using various arguments such as a combination of tags, text as well as technical metadata such as location, date, owner, etc. In addition, there are many open datasets and online collections along with APIs available for developers, which are provided by museums and galleries nowadays¹⁹. There are various configuration options

¹⁷ <https://developers.google.com/image-search/v1/devguide>

¹⁸ <https://www.flickr.com/services/api/>

¹⁹ <http://museum-api.pbworks.com/w/page/21933420/Museum%C2%A0APIs>

offered for developers to perform context-aware content retrieval. One example is Victoria and Albert museum's APIs, which allow users to filter retrieved content based on its type, materials, techniques and categories related to colour and the semantic content of the image and similarity to particular concepts/artefacts.

The advances of Internet tools discussed above, combined with recent developments in technology of information extraction from text, used for example to analyse textural design-briefs or the result of brainstorming session, can bring real benefit for product designers in background information search, inspiration and overall process of products' requirements assessment. Accordingly better alignment of semantic knowledge extraction and annotation tools with advances of classical web and application specific tasks of designer is desirable. Zemanta service developed to support bloggers discussed earlier is one example of such application-specific enhancement. The relevant support for initial stage of product design such as the means to bring associations to stimulate creativity or tools to facilitate the development of various mood boards are of great help.

4 Conclusion

Semantic knowledge extraction and annotation technology has a potential to support designer in knowledge exploration tasks during the initial stage of conceptual product design. While extensive number of various KE tools has resulted from a research community, our analysis shows that there is still effort required before the tools will be able to address the needs of designers towards the collaboration and creativity support. The areas of essential improvements include the usability of knowledge extraction technology. Also taking into account that multimedia is one of the main sources of inspiration for designer, further work on knowledge extraction and annotation of images is required. Moreover, we argue that future developments to be put to address particular use cases/application tasks (e.g. definition of requirements, brainstorming, creative thinking, creation of design-views /visualizations) and user groups (designer, developer). In addition the alignment of knowledge extraction technology with classical web from end user- as well as from application developer point of view is very desirable.

Acknowledgments. This research is funded by the European Commission 7th Framework ICT Research Programme. Further details can be found accessed at: <http://www.concept-fp7.eu>

References

1. Bizer, C., Heath, T., Berners-Lee, T.: Linked Data—The Story So Far. *International Journal on Semantic Web and Information Systems* 5 (3): 1–22 (2009)
2. Carbone, F., et al.: Enterprise 2.0 and Semantic Technologies for Open Innovation Support. *Trends in Applied Intelligent Systems Lecture Notes in Computer Science* Volume 6097, 2010, pp 18-27 (2010)

3. Caputa, B., et al.: ImageCLEF 2013: the vision, the data and the open challenges. Information Access Evaluation. Multilinguality, Multimodality, and Visualization Lecture Notes in Computer Science Volume 8138, 2013, pp 250-268 (2013)
4. Cross, N.: Engineering Design Methods Strategies for Product Design. John Wiley & Sons. Ltd (2000)
5. Csikszentmihalyi, M.: Creativity: Flow and the psychology of discovery and invention. New York: HarperCollins (1996)
6. Blomqvist, E.: The Use of Semantic Web Technologies for Decision Support - A Survey.: Semantic Web Journal, IOS Press (2012)
7. Bonnardel N., Marmèche E.: Towards Supporting Evocation Processes in Creative Design: A Cognitive Approach. International Journal Human-Computer Studies, 63, 422-435 (2005)
8. Datta, R., et al. (2008):. Image Retrieval: Ideas, Influences and Trends of New Age. ACM Computing Surveys, Vol.40, No.2, Article 5 (2008)
9. Dasiopoulou, S., Giannakidou, E., Litos, G., Malasioti, P., & Kompatsiaris, Y.: A Survey of Semantic Image and Video Annotation Tools., in Georgios Paliouras; Constantine D. Spyropoulos & George Tsatsaronis, ed., 'Knowledge-Driven Multimedia Information Extraction and Ontology Evolution', Springer, , pp. 196-239 (2011)
10. Gangemi, A.: A Comparison of Knowledge Extraction Tools for the Semantic Web. ESWC 2013: 351-366 (2013)
11. Govers, C. P.: What and how about quality function deployment (QFD).International journal of production economics, 46, 575-585 (1996)
12. Grassi, M., Morbidoni, C., Nucci, M., Fonda, S., and Di Donato, F.: Pundit: Creating,Exploring and Consuming Semantic Annotations. In Proceedings of the 3rd International Workshop on Semantic Digital Archives, Valletta, Malta, September 26 (2013)
13. French, M.J.: Conceptual design for engineers. Springer (1999)
14. Hare, J.S., Lewis, P.H., Enser, P.G.B., Sandom, C.J.: Mind the gap:another look at the problem of the semantic gap in image retrieval. In:Multimedia Content Analysis, Management and Retrieval. Vol.6073., SPIE (2006) 607309-1 (2006)
15. Haslhofer, B., Momeni, E., Gay, M., and Simon, R.: Augmenting Europeana Content with Linked Data Resources. 6th International Conference on Semantic Systems (I-Semantics), September 1-3, 2010, Graz, Austria (2010)
16. Halaschek-Wiener, C., Schain, A., Grove, M., Parsia, B., Hendler, J.: Management of digital images on the semantic web. In: Proceedings of the International Semantic Web Conference (2005)
17. Hollink, L., Worring, M.: Building a visual ontology for video retrieval. In: Proceedings of the 13th International ACM Conference on Multimedia (MM), New York, NY, USA, ACM Press (2005) 479-482 (2005)
18. Kobilarov, G., et al.: Media Meets Semantic Web – How the BBC Uses DBpedia and Linked Data to Make Connections. The Semantic Web: Research and Applications. Lecture Notes in Computer Science Volume 5554, pp 723-737 (2009)
19. Leclercq P., and Heylighen A.: 5.8 analogies per hour - A designer's view on analogical reasoning. AID'02 Artificial Intelligence in Design, Cambridge, July 15-17 (2002)
20. Mougenot C., Bouchard C., Aoussat A.: Creativity in design - How designers gather information in the "Preparation" phase. In ASDR 2007, Hon-Kong, Design Research Society, 11-15 November (2007)
21. Steffen, D.: Design semantics of innovation, product language as a reflection on technical innovation and socio-cultural change. In Proceedings of Design Semiotics in Use workshop, held as a part of World Congress in Semiotics "Communication: Understanding/ Misunderstanding, June 6-8, Helsinki, Finland (2007)
22. Uren, V et al.: Semantic annotation of knowledge management: Requirements and a survey of the state of the art. Wb Semantics: Science, Services and Agents on the world Wide Web, Vol.4 Issue 1 (2006)

Web Board Question Answering System Based on Problem-Solving

Chaveevan Pechsiri¹, Onuma Moolwat¹, and Rapepun Piriyakul²

¹ Dept. of Information Technology, Dhurakij Pundit University, Bangkok, Thailand

² Dept. of Computer Science, Ramkhamhaeng university, Bangkok, Thailand
{ itdpu@hotmail.com, moolwat@hotmail.com, rapepunnight@yahoo.com }

Abstract. This paper aims to work on the Question Answering (QA) system, especially the Why and how question types approach for solving problems, through the online web boards. The research QA system benefits for the online communities in solving their problems, especially on health-care problems. Both question and answer expressions are based on multiple EDUs (Elementary Discourse Units) where each EDU is equivalent to a simple sentence or a clause. Our research contains two problems: the first is how to identify the reasoning questions. The second is how to determine the corresponding answer from the knowledge source after the question focuses have been solved. Where the knowledge source contains several question-answer pairs express in terms of EDU vectors for both questions and answers. Therefore, the research applies different machine learning techniques, Naïve Bayes and Support Vector Machine, to solve the reasoning question type identification. The research also proposes an answer-determination method based on two levels of determining similarity scores of questions from the knowledge source. Finally, the research achieves 83% correctness of the answer determination.

Keywords: Why question, How question, similarity score, problem solving

1 Introduction

The objective of this research is to develop a Question Answering (QA) system based on explanation knowledge or reasoning knowledge for primary problem solving, especially on the health care problems, through the online community web boards (e.g. <http://haamor.com/webboard/>). When people have problems of their disease symptoms, some of them direct to a certain clinic or hospital whilst some others with their own reasons (e.g. a time conflict, a tight budget, a location problem, and etc.) prefer to post their problems of disease symptoms including queries or questions on the health-care-community web boards. Then, they wait for several minutes to a couple days depending on their topics of queries to receive the recommended answers of solving their problems on the web boards posted by the experts. However, it is time consuming for them to know how to solve their problems. Moreover, some common accidents can be occurred on trip, in family, or at work as burns, bleeding wounds,

adfa, p. 1, 2011.

© Springer-Verlag Berlin Heidelberg 2011

dislocations, or sprains. If someone knows how first aid from the health-care-community web boards through the QA system, it can limit the damage, maybe even save the lives of victims. Thus, it is necessary to develop an automatic web board QA system, based on reasoning questions such as Why and How questions, to solve their problems. The corresponding answers of the web board QA system can be determined from the knowledge source containing question-answer pairs of the previous extracted Symptom-Treatment Relation from the medical-care-consulting documents [1] and also the extracted causality from the health-care documents [2]. Thus, our research of the web board QA system can provide the explanation knowledge or the reasoning knowledge for people to make their own decisions. In addition, the posted health care problems of disease symptoms on the web boards are expressed in the form of several Elementary Discourse Units or several EDUs (where each EDU is defined as a simple sentence or a clause, [3]) And, the posted problems always consist of multiple EDUs of the symptom expression followed by 1-3 questions as shown in the following question-pattern.

$$EDU_{ct-1} EDU_{ct-2} \dots EDU_{ct-n} EDU_{q-1} \dots EDU_{q-end} \quad \text{where } 0 < end < 4$$

EDU_{q-i} is the question EDU containing a question word (qw_{q-i}) and $qw_{q-i} \in QW$ where QW is a question-word set.

$QW = \{ \text{'ทำไม/Why'} \text{'อย่างไร/How'} \text{'อะไร/What'} \text{'ขอคำวินิจฉัย/RequestDiagnosis'} \text{'ใช่หรือไม่/yes-no'} \dots \}$

EDU_{ct-a} is a content EDU expressing a problem content of EDU_{q-i} based on 'Why-Question' (Why-Q), 'How-Question' (How-Q), and 'ขอคำวินิจฉัย/RequestDiagnosis' where $a=1,2,\dots,n$ and $i=1,2,3$. n is an integer number and is greater than 0.

EDU_{ct-a} and EDU_{q-i} have the following Thai linguistic patterns where the problem content of EDU_{q-i} for this research is based on a symptom event expressed by a verb phrase.

$EDU_{ct-a} \rightarrow NP1 \text{ VP } | \text{ conj } NP1 \text{ VP}$

$VP \rightarrow V_{sym} \text{ NP2}$

$EDU_{q-i} \rightarrow qw_{q-i} \text{ NP1 } V_q \text{ NP2 } | \text{ NP1 } V_q \text{ NP2 } qw_{q-i}$

(where V_{sym} is a symptom verb concept set having $v_{ct-a} \in V_{sym}$; V_q is a verb set expressed on EDU_{q-i} having $v_{q-i} \in V_q$; conj is a conjunction word; NP1 and NP2 are noun phrases.)

For example :

Example1

EDU_{ct-1} : “เมื่อผมไอออกมา/**When I cough,**”

(เมื่อ/When ผม/I ไอออกมา/cough)

EDU_{ct-2} : “เสมหะเป็นเลือด/**phlegm contains blood.**”

(เสมหะ/phlegm (sputum) เป็นเลือด/contain blood)

EDU_{ct-3} : “แต่ว่า[ผม]ไม่เป็นไข้/**But [I] have no fever.**”

(แต่ว่า/But [ผม/I] ไม่เป็นไข้/have no fever)

EDU_{ct-4} : “[ผม]เป็นมาประมาณ 2 วันครับ/[**I have the symptoms about 2 days.**”

([ผม/I] เป็นมา/have the symptoms ประมาณ 2 วันครับ/about 2 days)

EDU_{q-1}: “ทำไม[ผม]ไอออกมาเป็นเลือดครับ/**Why do [I] cough up blood ?**”

(ทำไม/Why [ผม/I] ไอออกมาเป็นเลือดครับ/cough up blood)

EDU_{q-2}: “ [ผม/I] ควรจะทำอย่างไรครับ/**How should [I] remedy ?**”

([ผม/I] ควรจะ/should ทำ/remedy อย่างไร/How)

(where [...] means ellipsis.)

This research emphasizes on the problem-solving questions especially the disease-symptom questions on the health-care domain. According to the knowledge source, the Symptom-Treatment relation is represented by a symptom-treatment vector pair (which consists of a symptom-EDU vector, $\langle \text{EDU}_{\text{sym-1}} \text{EDU}_{\text{sym-2}} \dots \text{EDU}_{\text{sym-m}} \rangle$, as a question, and a treatment-EDU vector, $\langle \text{EDU}_{\text{treat-1}} \text{EDU}_{\text{treat-2}} \dots \text{EDU}_{\text{treat-p}} \rangle$, as an answer for How-Q). And, the causality relation is also represented by a cause-effect vector pair (which consists of a cause-EDU vector, $\langle \text{EDU}_{\text{cause-1}} \text{EDU}_{\text{cause-2}} \dots \text{EDU}_{\text{cause-s}} \rangle$ as an answer for Why-Q of solving causes, and an effect-EDU vector or the symptom-EDU vector, $\langle \text{EDU}_{\text{sym-1}} \text{EDU}_{\text{sym-2}} \dots \text{EDU}_{\text{sym-m}} \rangle$, as a question).

In addition to the question-pattern, EDU_{ct-a} expresses the problem content such as a symptom event based on a verb phrase. Thus, EDU_{ct-a} can be represented by a word co-occurrence of two adjacent words (called “contentWord-CO”, $v_{\text{ct-a}} w_{\text{ct-a}}$) after stemming words and skipping the stop words in between two adjacent words. The first word of contentWord-CO is a verb expression, $v_{\text{ct-a}}$, where $v_{\text{ct-a}} \in V_{\text{sym}}$ and V_{sym} approaches to the symptom concept. The second word of contentWord-CO is a co-occurred word, $w_{\text{ct-a}}$ ($w_{\text{ct-a}} \in W_{\text{sym}}$ where W_{sym} is a co-occurred word set inducing the $v_{\text{ct-a}} w_{\text{ct-a}}$ co-occurrence to have the symptom concept [1], based on Wordnet and MeSH after using Thai-To-English dictionary).

$V_{\text{sym}} = \{ \text{‘ถ่าย/defecate’}, \text{‘นั่ง/push’}, \text{‘ปวดท้อง/have an abdomen pain’}, \text{‘ปวด/pain’}, \text{‘อึดอัด/be uncomfortable’}, \text{‘รู้สึกไม่สบาย/be uncomfortable’}, \text{‘มี[อาการ]/have [symptom]’}, \}$

$W_{\text{sym}} = \{ \text{‘ ’}, \text{‘ยาก/difficultly’}, \text{‘ถ่าย/stools’}, \text{‘เชื้อ/germ’}, \text{‘เหลว/liquid’}, \text{‘ประจำเดือน/period’}, \text{‘แน่นท้อง/fullness’}, \text{‘ท้องเฟ้อ/flatulence’}, \text{‘ไข้/fever’}, \dots \}$

EDU_{q-i} can also be represented by a word co-occurrence of two adjacent words where the first word is v_{q-i} and the second word is the co-occurred word, w_{q-i} after stemming words and the stop word removal.

There are several techniques ([4],[5],[6],[7],[8]) having been used for reasoning QA system especially on a Why and How QA system (see section 2). However, the Thai documents have several specific characteristics, such as zero anaphora or the implicit noun phrase, without word and sentence delimiters, and etc. All of these characteristics are involved in two main problems. The first problem is how to classify the reasoning question type, e.g. ‘Why’ ‘How’ ‘Request-Diagnosis’, from EDU_{q-i} having the $q_{w_{q-i}}$ ambiguity where Request-Diagnosis requires the answer from both Why-Q and How-Q. The second problem is how to determine a corresponding answer from the knowledge source for each question, EDU_{q-i}, after the question focuses of EDU_{q-i} have been solved, where the knowledge source contains several the question-answer pairs of the symptom-treatment vector pairs and cause-symptom vector pairs (see section 3.2). According to these problems, we need to develop a framework

which combines the machine learning technique and the linguistic phenomena to learn the several EDU expressions of the question-pattern on the health-care-community web boards. Therefore, we apply Naïve Bayes (NB) and Support Vector Machine (SVM) to learn the posted-problem question types especially the Why, How, Request-Diagnosis, and Other questions from two EDUs, EDU_{ct-n} and EDU_{q-i} where $i=1,2,3$. We also apply the IR (Information Retrieval) techniques to determine the corresponding answer from the knowledge source by applying two levels of determining similarity scores between the content-EDU vector, $\langle EDU_{ct-1} EDU_{ct-2} \dots EDU_{ct-n} \rangle$, of the posted-problem question on the web board and each symptom-EDU vector, $\langle EDU_{sym-1} EDU_{sym-2} \dots EDU_{sym-m} \rangle$, which is equivalent to contents of a question in the knowledge source.

Our research is separated into 5 sections. In section 2, related work is summarized. Problems in web board QA system are described in section 3 and section 4 shows our framework for web board QA system. In section 5, we evaluate and conclude our proposed model.

2 Related Work

Several strategies ([4], [5], [6],[7],[8]) have been proposed to solve their QA systems including Why questions or How questions.

Girju, 2003, [4] worked on the Why question with the answer based on the lexico-syntactic pattern as ‘NP1 Verb NP2’ (where NP1 and NP2 are the noun-phrase expressions of a causative event and an effect event, respectively), i.e. “What causes Tsunami? \rightarrow Earthquake causes Tsunami”. However, it is not suitable for our research mostly based several effect-event explanations which express by verbs/verb phrases.

Schwitler et al., (2004) [5] worked on the procedural questions/How questions with their answers being extracted from technical documents by the ExtrAns system. Their procedural answer is often expressed in a procedural writing style with guidelines. The high performance in their QA system is best achieved through logic-based and pattern-matching techniques.

Verberne et al., (2007)[6] proposed using RST (Rhetorical Structure Theory) structures to approach Why questions by matching the question topic with a nucleus in the RST tree while yielding the answer from the satellite. The RST approach to the Why-QA system achieved the answer correctness of 91.8% and a recall of 53.3%.

Baral et al., (2012)[7] developed a formal theory of answers to why and how questions by developing the biological-graph model having event nodes and compositional edges as the knowledge-base with corresponding to why and how questions on the biology domain. Their questions are based on the forms: “How are X and Y related in the process Z?” and “Why is X important to Y.

Oh et al., (2013)[8] used intra- and inter- sentential causal relations between terms or clauses as evidence for answering Why-questions. They ranked their candidate answers (from documents retrieval Japanese web texts) with the ranking function

including re-ranking the answer candidates done by a supervised classifier (SVM). Their why-QA system achieve 83.2% precision.

However, most of previous researches on a Why-QA system / a How-QA system [4][5][7] are based on a single sentence/one EDU of a Why question / a How question, except [6] and [8] based on two EDUs of Why question, whereas our Why-Q and/or How-Q is based on several EDUs (see section 1).

3 Problems of Web Board QA System

To develop the web board QA system, there are two main problems that must be solved: how to identify the reasoning questions (especially a Why question and a How question) from the posted problems with multi-questions having the qw_{q-i} ambiguity and how to determine the corresponding answer of each question from the knowledge source.

3.1 How to Identify Why Question and How Question with qw_{q-i} Ambiguity

The problem of identifying the question expression without having the question mark symbol ('?') is solved by using a question-word set, QW, {'ทำไม/Why', 'อย่างไร/How', 'อะไร/What', ..} having $qw_{q-i} \in QW$. Where a 'ทำไม/Why' function is a reasoning question, a 'อะไร/What' function is asking for information about something (<http://www.englishclub.com/vocabulary/wh-question-words.htm>). However, there is a question word's function ambiguity, e.g. 'อะไร/What' as in reasoning (as shown in the following example on EDU_{q-1})

Example2

EDU_{ct-1}: "[ผม] ไม่มีน้ำมูก/ [I] do not have mucus."

EDU_{ct-2}: "[ผม] ไม่ไอ/[I] do not cough."

EDU_{ct-3}: "แต่[ผม] คัดจมูกบ้าง /But [I] have a congested nose."

EDU_{ct-4}: "[ผม] เจ็บคอ/[I] have sore throat ."

EDU_{q-1}: "เป็นเพราะอะไร/What are the reasons?"

EDU_{q-2}: "มีวิธีรักษาอาการอย่างไรคะ/ Show how to treat symptoms."

Therefore, we solve the qw_{q-i} ambiguity by apply NB and SVM to learn the posted-problem question type from EDU_{q-i} and EDU_{ct-n} where EDU_{ct-n} is the last content EDU and $i=1,2,3$. There are three features used for NB learning and SVM learning, $v_{ct-n} w_{ct-n}$ co-occurrence of EDU_{ct-n}, qw_{q-i} from EDU_{q-i}, and another verb-noun co-occurrence, $v_{q-i} w_{q-i}$, of EDU_{q-i} after stemming words and eliminating the stop words.

3.2 How to Determine Corresponding Answer from Knowledge Source

The knowledge source contains the question-answer pairs of several symptom-treatment vector pairs and several cause-symptom vector pairs. Each symptom-EDU of a symptom-EDU vector and each treatment-EDU of a treatment-EDU vector are represented by the $v_{sym-i} w_{sym-i}$ co-occurrence (with the symptom concept, called

‘symptomWord-CO’, where $v_{\text{sym}-i} \in V_{\text{sym}}$ and $w_{\text{sym}-i} \in W_{\text{sym}}$ and the $v_{\text{treat}-j} w_{\text{treat}-j}$ co-occurrence (with the treatment concept) respectively [1]. Where $i=1,2,\dots,m$; $j=1,2,\dots,p$; m is the number of symptom-EDUs on the symptom-EDU vector; p is the number of treatment-EDUs on the treatment-EDU vector (see section 1); $v_{\text{treat}-j} \in V_{\text{treat}}$; $w_{\text{treat}-j} \in W_{\text{treat}}$; V_{treat} is the treatment-verb concept set; W_{treat} is co-occurred word sets inducing the $v_{\text{treat}-j} w_{\text{treat}-j}$ co-occurrence to have treatment concept. Then, $v_{\text{sym}-i} w_{\text{sym}-i} \in VW_{\text{sym}}$ (where VW_{sym} is a symptomWord-CO set) and $v_{\text{treat}-j} w_{\text{treat}-j} \in VW_{\text{treat}}$ (where VW_{treat} is a treatmentWord-CO set). A contentWord-CO of $\text{EDU}_{\text{ct}-a}$ is also an element of VW_{sym} ($v_{\text{ct}-a} w_{\text{ct}-a} \in VW_{\text{sym}}$). Thus, the content-EDU vector, $\langle \text{EDU}_{\text{ct}-1} \text{EDU}_{\text{ct}-2} \dots \text{EDU}_{\text{ct}-n} \rangle$, of the question-pattern can be represented by a contentWord-CO vector, $\langle v_{\text{ct}-1} w_{\text{ct}-1} v_{\text{ct}-2} w_{\text{ct}-2} \dots v_{\text{ct}-n} w_{\text{ct}-n} \rangle$ where $\text{SubSym}, \{v_{\text{ct}-1} w_{\text{ct}-1} v_{\text{ct}-2} w_{\text{ct}-2} \dots v_{\text{ct}-n} w_{\text{ct}-n}\}$, is a set containing all elements of the contentWord-CO vector. SubSym can occur on several symptom-treatment vector pairs/several cause-symptom vector pairs of the knowledge source (where $\text{SubSym} \subseteq VW_{\text{sym}}$) as shown in the following example.

Example3

$\text{EDU}_{\text{ct}-1}$: “Yesterday, I felt dizzy.”

$\text{EDU}_{\text{ct}-2}$: “I vomited frequently.”

$\text{EDU}_{\text{ct}-3}$: “Now I am very nauseated.”

$\text{EDU}_{\text{q}-1}$: “What can I do?”

$\text{EDU}_{\text{q}-2}$: “What could have caused this?”

From Example 3, $\text{SubSym}, \{\text{‘feel_dizzy’ ‘vomit_frequently’ ‘be_nauseated’}\}$, occurs on the symptom-treatment vector pairs of the following diseases, Brain Tumor, Food Poisoning, and Vestibular Imbalance. It is challenge to determine the answer of this question of the posted problem on the web board. Therefore, we propose using two levels of determining similarity scores among the content-EDU vector (the posted problem) and the symptom-EDU vectors from the knowledge source after stemming words and skipping the stop words. The first level of determining similarity scores is at the EDU level of each EDU pair ($\text{EDU}_{\text{ct}-a} \text{EDU}_{\text{sym}-b}$; $a=1,2,\dots,n$; $b=1,2,\dots,m$). The EDU pair is marked ‘match’ if its similarity score is ≥ 0.5 otherwise ‘un-match’. The final similarity scores is determined in the second level as the vector level based on the ‘match’ marks of each content-symptom vector pair, $\langle \text{EDU}_{\text{ct}-1} \text{EDU}_{\text{ct}-2} \dots \text{EDU}_{\text{ct}-n} \rangle$ $\langle \text{EDU}_{\text{sym}-1} \text{EDU}_{\text{sym}-2} \dots \text{EDU}_{\text{sym}-m} \rangle$. The answer of How-Q/Why-Q is selected from the symptom-treatment vector pairs or the cause-symptom vector pairs having the top 3 ranked final similarity scores.

4 A Framework for Web Board Question Answering System

There are four steps in the web board QA system, Question-Corpus Preparation, Question-Type Learning, Question-Type Identification, and Answer Determination as shown in Fig.1.

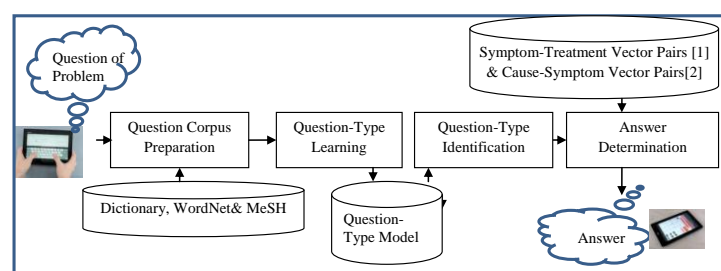


Fig.2 System Overview

4.1 Question-Corpus Preparation

```

<EDUct1> “[ผม] รู้สึก คลื่นไส้” ( [I] feel nauseated.)
[φ = ผม/I/NP <Qfocus> [<WordCo concept= be nauseated><vct1 =feel> รู้สึก/vi</vct1><wct1=
nauseated> คลื่นไส้/vi </wct1></WordCo>]/VP</Qfocus></EDUct1>
<EDUct2> “[ผม] อาเจียน” ( [I] vomit.)
[φ = ผม/I/NP <Qfocus> [<WordCo concept= vomit><vct1 =vomit> อาเจียน/vi</vct1><wct1=null>
</wct1> </WordCo>]/VP</Qfocus></EDUct2>
<EDUct3> “[ผม] เวียนศีรษะ” ( [I] am dizzy.)
[φ = ผม/I/NP <Qfocus> [<WordCo concept= be dizzy><vct1 =be dizzy> เวียนศีรษะ/vi</vct1><wct1=
null></wct1> </WordCo>]/VP</Qfocus></EDUct3>
<EDUct4> “[ผม] นอนไม่หลับ” ( [I] can't sleep.)
[φ = ผม/I/NP <Qfocus> [<WordCo concept= cannot sleep><vct1 =sleep>นอน/vi </vct1>
<wct1=cannot>ไม่หลับ/adv </wct1> </WordCo>]/VP</Qfocus></EDUct4>
<EDUq1> “[ผม] ควรจะทำอย่างไร” (How should [I] do?)
[ φ=ผม/I/NP [should-ควร/prev <vq1 : concept=do>ทำ/vt </vq1><wq1=null> </wq1>
<Qword concept=how>how-อย่างไร/print </Qword>]/VP </EDUq1>

```

Where: a 'Qfocus' tag is a question focus tag. A WordCo tag is a word co-occurrence tag. A 'vct#' tag is a verb tag of a content EDU and the other verb concept set. A wct# tag is a co-occurred word tag. A 'vq#' tag is a verb tag of an EDU containing the question word. A 'Qword' tag is a question word tag. An EDUct tag is an EDU content tag. An EDUq tag is a tag of an EDU having the question word. And, φ stands for a zero anaphora or ellipsis.

Fig.3 Examples of question annotation

This step is the preparation of the question corpus in the form of EDU from the medical-care consulting documents on the hospital's web-boards of the Non-Government-Organization (NGO) website. The step involves with using Thai word segmentation tools [9], including Name entity [10]. After the word segmentation is achieved, EDU segmentation is then to be dealt with [11]. These annotated EDUs will be kept as an EDU corpus. This corpus contains 6000 EDUs of 580 questions from four disease categories, gastrointestinal tract diseases, brain-never diseases, ear-nose-throat diseases, and bone diseases including arthritis. The corpus is separated into 2 parts; one part is 400 questions for learning the question type concepts based on ten folds cross validation. And, the other part of 180 questions is for testing the question type identification. In addition to this step of corpus preparation, we semi-automatically annotate the question-word concepts, the question-focus group, the contentWord-CO concepts of symptoms, as shown in Fig.3. All concepts of annotation are referred to Wordnet (<http://word-net.princeton.edu/obtain>) and MeSH after translating from Thai to English, by using Lexitron (the Thai-English dictionary) (<http://lexitron.nectec.or.th/>).

4.2 Question-Type Learning

This step is using Weka (<http://www.cs.waikato.ac.nz/ml/weka/>) to learn Why-Q, How-Q, Request-Diagnosis, and Other-Q by NB, and SVM from EDU_{q-i} and EDU_{ct-n} (where i=1,2,3) of the question-pattern from the annotated-question corpus. The

features used for these learning consist of three feature sets; VW_{sym} , QW , and $VW_{question}$. $VW_{question}$ is a questionWord-CO set where each element of the set is formed by a verb and a co-occurred word from a question EDU, EDU_{q-i} after stemming words and eliminating stop words. Where EDU_{ct-n} contains $v_{ct-n}w_{ct-n} \in VW_{sym}$, and EDU_{q-i} contains $qw_{q-i} \in QW$ and $v_{q-i}w_{q-i} \in VW_{question}$ after stemming words and skipping the stop words

Naïve Bayes (NB) According to [12], the NB learning is a generic classification to determine the feature probabilities of four classes (class1, class2, class3, and class4) of the question types; Why-Q, How-Q, Request-Diagnosis, and Other-Q respectively. The features of NB classifiers consist of three features, $v_{ct-n}w_{ct-n}$, qw_{q-i} , and $v_{q-i}w_{q-i}$ of VW_{sym} , QW , and $VW_{question}$ respectively, from the annotated corpus of two EDUs (EDU_{q-i} and EDU_{ct-n} where $i=1,2,3$).

Support Vector Machine The linear binary classifier, SVM, applies in this research to classify the four question types, Why-Q, How-Q, Request-Diagnosis, and Other-Q based on pairwise classification. According to [13] this linear function, $f(x)$, of the input $x = (x_1 x_2 \dots x_n)$ assigned to the positive class if $f(x) \geq 0$, and otherwise to the negative class if $f(x) < 0$, can be written as:

$$\begin{aligned} f(x) &= \langle w \cdot x \rangle + b \\ &= \sum_{j=1}^n w_j x_j + b \end{aligned} \quad (1)$$

where x is a dichotomous vector number, w is weight vector, b is bias, and $(w, b) \in \mathbb{R}^n \times \mathbb{R}$ are the parameters that control the function. The SVM learning results are w_j and b for x_j which consists of three feature sets, VW_{sym} , QW , and $VW_{question}$, for the corresponding elements, $v_{ct-n}w_{ct-n}$, qw_{q-i} , and $v_{q-i}w_{q-i}$ where $i=1,2,3$.

4.3 Question-Type Identification

All probabilities or weights from the previous learning step by NB and SVM are used to identify the question types

Naïve Bayes According to [12], equation 2 and the feature-probabilities determined by the previous step of NB are used to identify the question type class of the question-pattern.

$$\begin{aligned} QuestionClass &= \underset{class \in Class}{\operatorname{argmax}} P(class | v_{ct-n}w_{ct-n}, v_{q-i}w_{q-i}, qw_{q-i}) \\ &= \underset{class \in Class}{\operatorname{argmax}} P(v_{ct-n}w_{ct-n} | class) P(v_{q-i}w_{q-i} | class) P(qw_{q-i} | class) P(class) \end{aligned} \quad (2)$$

where $v_{ct-n}w_{ct-n} \in VW_{sym}$; VW_{sym} is a symptom Word – CO set

$v_{q-i}w_{q-i} \in VW_{question}$; $VW_{question}$ is a Word – CO set on EDU_{q-i} and $i=1,2,3$

$qw_{q-i} \in QW$; QW is a question word set

Support Vector Machine The results from SVM learning are weight, w_j , and bias, b , of each feature (x_i). According to equation (2), the input vector (x) consisting of contentWord-CO features, $v_{ct-n}w_{ct-n}$, questionWord-CO features, $v_{q-i}w_{q-i}$, and question-word features, qw_{q-i} , including their weights, w_j , and bias, b , are used to determine the question types based on pairwise classification.

4.4 Answer Determination

According to the correct question-type identification from the previous step, we randomly select 120 correct questions which consist of 30 correct questions of each disease category, gastrointestinal-tract diseases, brain-never diseases, ear-nose-throat-diseases, and bone-arthritis-diseases. The selected questions consist of 50 questions of Why-Q, 50 questions of How-Q, and 20 questions of Request-Diagnosis. Since the focuses of Why-Q, How-Q, and Request-Diagnosis from the content-EDU vector of the question-pattern are based on the symptom events expressed by verb phrases, the question focuses of this research can be solved by the contentWord-CO vector with $v_{ct-a} w_{ct-a} \in VW_{sym}$. The answers can be solved by using two levels of determining similarity scores between the content-EDU vector (which contains the contentWord-CO vector as focuses) and the symptom-EDU vector of the following vector pairs of the knowledge source:

- a symptom-treatment vector pair for How-Q where the symptom-EDU vector is equivalent to the question and the treatment-EDU vector is the answer.
- a cause-symptom vector pair for a Why-Q where the symptom-EDU vector is equivalent to the question and the cause-EDU vector is the answer.
- a cause-symptom vector pair and a symptom-treatment vector pair for a question of Request-Diagnosis where the symptom-EDU vector is equivalent to the question and the cause-EDU vector and the treatment-EDU vector are the answers.

The first level of determining similarity scores as shown in equation 3 [14] is at the EDU level between EDU_{ct-a} and EDU_{sym-b}

$$Similarity_Score = \frac{|S1 \cap S2|}{\sqrt{|S1| \times |S2|}} \quad (3)$$

where $S1$ is an EDU_{ct-a} of the content-EDU vector (having $a=1,2,...,n$) after stemming words and skipping stop words.

$S2$ is an EDU_{sym-b} of the symptom-EDU vector (having $b=1,2,...,p$) after stemming words and skipping stop words.

All word concepts of a $S1$ EDU and a $S2$ EDU are based on WordNet and MeSH after using the Thai-to-English dictionary. The number of words in the $S1$ EDU and the number of words in the $S2$ EDU are not significantly different. According to the similarity score determination at EDU level, both EDU_{ct-a} and EDU_{sym-b} are marked 'match' if its similarity score is ≥ 0.5 otherwise 'un-match'. The final similarity scores based on the vector level is calculated by equation 4 from the content-symptom vector pair, $\langle EDU_{ct-1} EDU_{ct-2} \dots EDU_{ct-n} \rangle \langle EDU_{sym-1} EDU_{sym-2} \dots EDU_{sym-m} \rangle$ based on the 'match' marks. After ranking the final similarity scores, the answer of How-Q/Why-

Q is selected from the symptom-treatment vector pairs or the cause-symptom vector pairs from the knowledge source having the top 3 ranked similarity scores at the vector level as shown in Fig. 4.

$$\text{Similarity_Score} = \frac{|Vec1 \cap Vec2|}{\sqrt{|Vec1| \times |Vec2|}} \quad (4)$$

where Vec1 is a content-EDU vector of the question-pattern on a certain web-board
Vec2 is a symptom-EDU vector of a symptom-treatment vector pair or a cause-symptom vector pair

Assume that each EDU is represented by (NP V P). L is a list of EDUs. S1 is an EDU _{ct-a} of the content-EDU vector. S2 is an EDU _{sym-b} of the symptom-EDU vector. Vec1 is a content-EDU vector of the question-pattern on a certain web-board. Vec2 is a symptom-EDU vector of a symptom-treatment vector pair or a cause-symptom vector pair. D is matrix of Vec2s.	
ANSWER_DETERMINATION(D _{Vec2} L _{s1} L _{s2})	
1	x=1
2	while x<= row_of_matrix(D) do
3	{ j=1,count_intersect_Vector=0
4	while j<=length[L _{s2}] do
5	{ i=1;flag=true;
6	while i<=length[L _{s2}] and flag do
7	{ sim_score1 = similarity_score(S _{1-i} ,S _{2-j}) (eq.3)
8	if sim_score1>0.5 then
9	{ count_intersect_Vector++;
	intersect_Vector=intersect_Vector+S _{1-i} ; //match
	i++; flag = true; }
10	else flag = false;
11	}
12	j++;}
13	sim_score2=abs(count_intersect_vector)/sqrt(summation(abs(length(Vec1)))+abs(length(Vec2))))
14	Vec2 _x = sim_score2;
15	x++;}
16	Sorting Vec2 as descending;
17	Select top three Vec2;

Fig.4 Answer Determination Algorithm

5 Evaluation and Conclusion

There are two evaluations of the proposed research, the question type identification and the answer determination. The evaluation of the question type identification performance is based on the precision and the recall with the testing question corpus of 180 questions. The results of precision and recall are evaluated by three expert judgments with max win voting. The average precision and the average recall of the

question type identification by the proposed machine learning is 0.918 and 0.863 respectively (see Table1). SVM yields the better precision and recall results than NB whereas some features are dependency.

Table1 Show Precision and Recall of Question Type Identification by NB and SVM

180 Web Board Questions	NB		SVM	
	Precision	Recall	Precision	Recall
Why-Q	0.864	0.877	0.923	0.907
How-Q	0.891	0.865	0.932	0.922
Request-Diagnosis	0.952	0.819	0.960	0.941
Other	0.885	0.755	0.936	0.818

And, the evaluation of the answer determination performance of the proposed method is based on the percent correctness of the answer based on the knowledge source as the answer source. The results of answer correctness are evaluated by three expert judgments with max win voting. We drew a random sample of 120 correct questions from the question type identification for the answer evaluation. The answer correctness for Why-Q, How-Q, and Request-Diagnosis questions at the top 3 ranked similarity scores is 83%. The correctness of answer determination results should be increased if the symptom-EDU vectors of the same disease or the same cause-symptom vector pair are generalized before determining the proposed similarity scores. Therefore, the web board QA system of this research provides the explanation knowledge or reasoning knowledge for solving problems and also for the feasibility of self-study.

Acknowledgement

This is supported by Thai Research Fund. The medical-care knowledge and the pharmacology knowledge applied in this research are provided by Puangthong Kraipiboon, a clinician of Division of Medical Oncology, Department of Medicine, Ramathibodi Hospital, and Uraivan Janviriyasopak, a pharmacist of RexPharmacy, respectively.

References

1. Pechsiri ,C., Moolwat, O., Piriyaikul, R.: Symptom -Treatment Relation Extraction from Web- Documents for Construct Know-How Map. KICSS' 2013 proceedings (2013).
2. Pechsiri, C. Piriyaikul, R.: Explanation Knowledge Graph Construction through Causality Extraction from Texts. Journal of Computer Science and Technology, vol 25, no.5, pp.1055-1070 (2010).

3. Carlson, L., Marcu, D., Okurowski, M.E.: Building a Discourse-Tagged Corpus in the Framework of Rhetorical Structure Theory. *Current and New Directions in Discourse and Dialogue*, vol. 22, pp. 85-112 (2003).
4. Girju, R.: Automatic detection of causal relations for question answering. *Proc. 41st annual meeting of the assoc. for computational linguistics, workshop on multilingual summarization and question answering-Machine learning and beyond*, Japan (2003).
5. Schwitter, R., Rinaldi, F., Clematide, S.: The Importance of How-Questions in Technical Domains. In: *Proc. TALN-04, Workshop Question – Réponse*, Fez, Morocco (2004).
6. Verberne, S., Boves, L., Coppen, P-A., Oostdijk, N.: Discourse-based answering of why-questions. *Traitement Automatique des Langues*, 47, 2 (2007).
7. Baral, C., Vo, N.H., Liang, S.: Answering Why and How questions with respect to a frame-based knowledge base: a preliminary report. In: *Proc. ICLP 2012*, Hungary (2012).
8. Oh, J-H., Torisawa, K., Hashimoto, C., Sano, M., Saeger, S.D., Ohtake, K.: Why-Question Answering using Intra- and Inter-Sentential Causal Relations. In: *Proc. of the 51st Annual Meeting of the Association for Computational Linguistics*, Bulgaria (2013).
9. Sudprasert, S., Kawtrakul, A.: Thai Word Segmentation based on Global and Local Unsupervised Learning. *NCSEC'2003 Proceedings* (2003)
10. Chanlekha, H., Kawtrakul, A.: Thai Named Entity Extraction by incorporating Maximum Entropy Model with Simple Heuristic Information. *IJCNLP' 2004 proceedings* (2004)
11. Chareonsuk, J ., Sukvakree, T., Kawtrakul, A.: Elementary Discourse unit Segmentation for Thai using Discourse Cue and Syntactic Information. *NCSEC 2005 proceedings* (2005)
12. Mitchell, T.M.: *Machine Learning*. The McGraw-Hill Companies Inc. and MIT-Press, Singapore (1997)
13. Vapnik, V.N.: *The nature of statistical learning theory*. Springer, USA (1995).
14. Biggins, S., Mohammed, S., Oakley, S.: University Of Sheffield: Two Approaches to Semantic Text Similarity. In: *Proc. First Joint Conference on Lexical and Computational Semantics*, Montre´al, Canada (2012).

Front End System for Personal Health Record with Data Migration Facility from Printed Information

Atsuo Yoshitaka¹, Shinobu Chujyou², and Hiroshi Kato³

¹ School of Information Science
Japan Advanced Institute of Science and Technology
1-1 Asahidai, Nomi, Ishikawa, 923-1292 Japan

² goowa inc.
3-4 Unetanaka, Kanazawa, Ishikawa 920-0343 Japan

³ Life Care on Demand
3-13-5 Asahimachi, Kanazawa, Ishikawa 920-0941 Japan

Abstract. Countries where the society is aging face to issues on helping the advancement of health of citizens. PHR (Personal Health Record) systems are expected to be supplemental effect for improving health care environment in addition to the improvement of medical care environment. One of the major issues in PHR environment is that only a very limited data is distributed from hospital or medical institution to citizens as electric data; most of the data is still distributed by paper-printed document. Since it is estimated to take years to offer fully electric data distribution online, one of the practical solutions for this issue is to migrate and utilize PHR data from printed document in more efficient way. In this paper, we propose a PHR system which enables to migrate PHR data of printed document into the database by photography. It diminishes the burdens of an end-user in the process of data migration. In addition to this feature, we designed the user interface suitable for touch panel for more intuitive data manipulation. As a result of preliminary user test, we got positive feedbacks from testing subjects which show the effectiveness of the proposed system.

Keywords. Personal health record, data migration by photography, OCR, adaptive dictionary configuration

1 Introduction

Average life expectancy is high in Japan, and the importance of health management is recognized in accordance with aging society. Because of the growing ratio of aged person, Japanese government recognizes that policies for improving healthy life expectancy are becoming more important. These include not only the development of medical care environment but also that of the environment which enables effective use of medical data for both of medical personnel and citizen. The

development of health-care environment toward both of the above-mentioned two directions is expected to improve the quality of life especially for aged persons.

For the next step for matured aged society, Japanese government has put more focus on developing infrastructure for the citizens to maintain health by their effort. In accordance with this context, various aspects of EHR (Electric Health Record)[1] or PHR (Personal Health Record) systems have been studied.

Currently, sharing of electric medical data among hospitals or clinics is not established yet. One of the reasons is that the widely accepted standard of medical data format is still in progress. HL7 CDA [5] is one of the examples of such standard. Current status of medical data sharing is so-called “intra-sharing”, and “inter-sharing” is expected as the next step to improve medical service. Therefore, EHR or PHR system is considered to be a breakthrough for the above-mentioned issue. The EHR infrastructure includes the idea of data sharing among hospitals, clinics, medical inspection institutions, local governments, and citizen. Practical experiments of PHR system are carried out, and some of them are under the cooperation with local government. When we mention PHR system, we put more emphasis is on how the citizen takes care of his/her own health record by his/her own effort.

In this paper, we focus on PHR systems, since our motivation is based on how we can offer more effective platform for PHR data management. PHR systems[2][3][4] consolidates health information such as height, weight, BMI (Body Mass Index), blood type, blood pressure, hepatic function indices, uric protein, medication, history of disease, and so on. Currently, such health information is managed independently in each hospital/clinic or institution for medical examination, and therefore such data is distributed as many places as a person had been seen. The purpose of the development of a PHR system is to offer an environment for citizens to consolidate distributed information, and contribute to watch his/her health condition by his/her own effort. A PHR system makes it possible to refer to all his/her personal health information anytime and anywhere it needed: it is expected to contribute to diminish unintentional duplication of basic medical check.

There are studies and demonstration experiments on PHR system in Japan. One of the examples of PHR system is called Lico which is implemented by “Life care on demand” project. The Lico is a web-based PHR service and is aimed at maintaining his/her PHR by his/her own effort. One of the functions of Lico is to maintain the information of annual health checkup. There are approximately 40 items to enter manually with keyboard. As a result of operating the Life care on demand project, it is revealed that better interface in migrating printed data is awaited.

One of the possible solutions for this issue is to exploit an image scanner to capture a printed data of health check-up, and to apply OCR (Optical Character Recognition) processing for data migration. A general OCR engine reads out characters in each row, without recognizing logical structure of the table. Therefore, simple installation of a general image scanner with OCR engine will not be a promising solution. Alternative solution for this issue is to utilize a camera device to capture a form. For minimizing additional cost for equipping devices which may be devoted only for PHR data migration and for employing intuitive user interface, we think using a tablet PC is one of the promising solutions. Our scenario is as follows. A user takes photos of a

printed report of health checkup with the built-in camera of a tablet PC in which a PHR front-end system is installed. Then the captured images are sent to PHR front-end system for OCR processing to migrate printed PHR data.

One of the major issues in data migration from printed form with OCR is how we assure acceptable accuracy in OCR process. The accuracy of character recognition is not always satisfactory considering the burden for checking and correcting errors. This issue is more serious in the scenario of migrating printed data by photography, because the images shot with a hand-held tablet may be blurred, which results to deteriorate the accuracy of character recognition. In addition to this, we need to consider how we cope with geometric normality issue because of the flexibility of condition in photography.

In this paper, we describe a front-end system which enables a user to migrate PHR data by taking photography of printed document. We discuss the issue of improving the accuracy of character recognition by OCR engine for PHR data migration. Though this is considered to be a transient technique, however, we believe this is significant in order to offer better PHR platform at earlier convenience.

2 A PHR System Lico

2.1 System and Interface Design in Lico

As introduced in section 1, there are several demonstrational experiments for PHR systems. An example of them is led by a local government of Uchinada town in Ishikawa prefecture, Japan, which is called “Lico”. The Lico system is in operation which is implemented by the project called “Life care on demand” [6]. Lico offers a web based PHR service, which enables a user to enter PHR data by filling values into a form. Lico enables a registered user to store PHR data such as a report of annual health checkup, and to view stored PHR data.

An example of the report of annual health checkup is shown in Figure 1. Note that the form is actually used for distributing the report to each person who had a medical checkup, but the data printed is for demonstration. The report includes the result of the measurement of height, weight, BMI (body mass index), abdominal circumference, glucose, protein, and occult blood in urine, and so on. The number of items to be measured depends on the age, gender, and/or category of labor in case of an employee in a company. Figure 2 shows an example of screen snapshot when a user enters PHR data using Lico system. As mentioned, current implementation of Lico is a web-based system with simple data migration method. Lico data migration interface is designed using text boxes, and a user need to enter the values on the printed report of annual checkup one by one with keyboard and mouse. In this example, a user is prompted to enter the date of health checkup, height, weight, abdominal circumference, and body fat percentage.

Though the numbers of items to be measured during an annual health checkup depends on age, gender, and company/institution (i.e., employer or health insurance system) which offers an annual medical check, there are approximately 40 items to enter. It is not difficult to imagine that entering all of such data into the form as shown

in Fig. 2 is time-consuming, stressful, and error-prone for the user. Note that all the reports of annual health checkup are distributed as a printed form, but no electric data is distributed.



Fig. 1. An Example of the Report of Annual Health Checkup

健康データ

測定したデータを健康手帳に入力します。

測定値を入力したい項目セットをメニューから選択し、健康データの測定日を設定します。

測定値を入力

ユーザー項目セット

測定データを登録します。

測定日を選択して、測定データを入力してください。

※ 指定した日付に測定データが存在している場合、測定データの上書き確認画面が表示されます。

測定日: 2010 年 (9月) 18日

コメント:

身長: cm

体重: kg

腹囲: cm

BMI (BMI): 0.00

体脂肪率: %

保存する

キャンセル

Fig. 2. User Interface for Entering PHR data in Lico

2.2 Data Migration Issues in Lico PHR System

Up to now, there are over 600 registered users of Lico system. As introduced in the previous section, current implementation of PHR data migration interface of Lico system is simple, text-box based user interface. A user needs to enter the values of measured items one by one into the fields manually. According to the experience of

operating Lico system and feedback from the Lico user, insufficient operability of data migration is revealed in the current system. That is, the task of migrating printed data by filling the values into tens of fields in a Web page one by one discourages the user to keep using the system. One reason is that the task of typing all the data manually is time-consuming, and confirming if there is a miss-typing is stressful. Another reason relates to the unfamiliarity of specific terms such as MCV, MCH, GOT, or GPT, all of which are the indices on blood quality or soundness of liver. Since it is difficult for most of the users to understand the meaning of such values, it is stressful for them to enter such values without errors because it is not obvious whether a certain value is erroneous or not.

Since electric data distribution from the hospitals/clinics or medical institutions to end users is not established, one of the most practical solutions is to migrate PHR data printed on a paper into electric data by OCR. Two kinds of devices are considered to capture printed data; one is an image scanner and another is a digital camera. We think the image scanner is rather special and less versatile device, and the device such as a camera module in tablet PC or smart phone is more general. Moreover, general users are more familiar with using a camera than an image scanner. Since most tablet PCs and smart phones equip camera modules for taking photos, we chose a tablet PC as a device for capturing forms of annual health checkup. Another advantage of assuming a tablet PC as the device for PHR front-end, we can provide a user with intuitive user interface with direct manipulation on LCD display with touch panel.

As far as we experimented, the accuracy of character recognition by OCR processing is not always perfect, in spite of the maturity of character recognition technique. Therefore, when we design a PHR data migration facility by OCR, what we need to take into consideration is whether the accuracy of OCR is acceptable or not, in the sense that the amount of errors that need to be corrected manually by the user. One reason of imperfect accuracy in OCR is in the insufficient resolution of source image or bleeding and/or blur of character printed on a paper. Another is the case where an image is captured with blur, especially in a case where an image is captured by holding the camera device by hands. In addition to the above mentioned issues, there is also an issue where a character is misclassified into the different character whose visual appearance is quite similar to the character to be recognized. One example is the case where a single “l” (the lower case of the alphabet “L”) is misrecognized as a single “1” (the letter which corresponds to the number “one”). This problem is inevitable if the OCR engine does not hold a dictionary of the letters specific to the printed fonts.

Moreover, the fundamental problem in applying general purpose OCR engine is difficulty in proper recognition of structured data such as a table. If a table which includes items with null values is processed, output of recognition will be a sequence of characters in horizontal order, which may lose the correspondence between item name and value and/or the logical structure of the table.

3 PHR Data Management Front-end

3.1 System Organization

Figure 3 illustrates the system organization of the PHR system which we propose in this paper. The PHR system consists of a PHR front-end and PHR manager. The PHR front-end is implemented as application software on a tablet PC. Form capturing module assists a user to capture printed form by showing guide frames. Captured image is then transferred to OCR processor so as to carry out OCR processing. Each item name is associated with the value which is indicated in the same row, and 'null' case is also detected when no character is found. At the final process of data migration, PHR data received from the PHR front-end is stored into PHR database by PHR server. When a user accesses his/her own PHR data, it is transferred to the PHR front-end via PHR server. It is then shown on the display of PHR front-end. In the succeeding sections, we will describe the detail of form capturing assistance and the method for improving OCR accuracy which we call adaptive dictionary configuration.

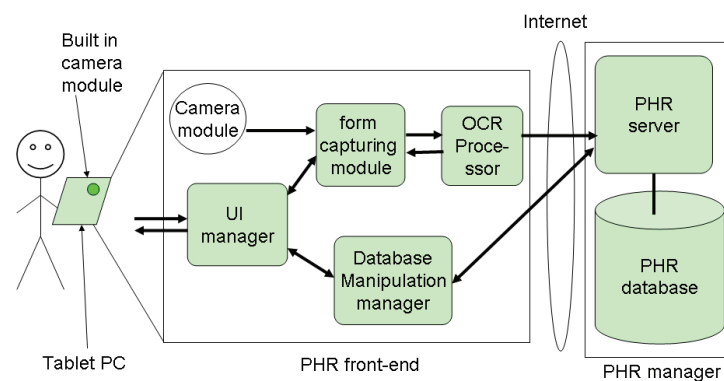


Fig. 3. The Organization of the PHR Front-end

3.2 Form Capture Assistance in Data Migration by Photography

Data migration by photography is the idea of migrating printed information into a system as electric data by OCR process, where the input image is obtained by photography. If we allow a user much freedom in shooting a form, many kinds of pre-processing such as geometric normalization are mandatory. In addition to this, there appear cases where image captured by photography is not usable because of low resolution or blur that is not negligible for realistic OCR processing. Therefore, proper shooting assistance is mandatory which guides a user to capture images properly. The detail of form capture module and OCR processor is shown in Figure 4. Prior to the data migration, the user is prompted to choose the institution which conducted his/her annual checkup. After that, guide frames are superimposed on the image captured by the build-in camera module, and is shown on the display. The

image is updated frame by frame, and all the user has to do is to adjust the position of the tablet that he/she holds in hands so that the borders of each table component overlap with guide frames, and finally touch a shutter icon to capture. A snapshot in the process of data migration by photography is shown in Figure 5. Figure 5 is an example when a user tries to adjust the position of his tablet so as to migrate the table of “anthropometric”, “urine test”, and “circulatory organ test”.

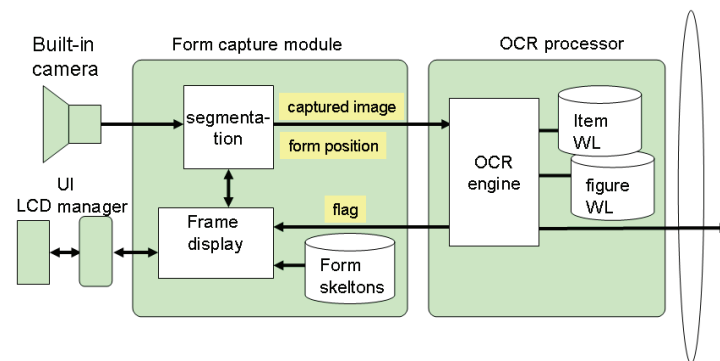


Fig. 4. The Organization of Capture Module and OCR Processor



Fig. 5. Displaying Table Frames for Capturing Printed Report.

3.3 OCR Processing with Adaptive Dictionary Configuration

Our preliminary experiment showed that OCR processing for images captured by handheld tablet with improper image results more recognition errors with standard dictionary, compared with OCR process with image scanner. Figure 6 shows an

example of OCR processing where an image is captured by a handheld tablet. The entire image is 1223x1631 pixels, and we clipped two columns as shown in red and green rectangle. The result of OCR process with standard Japanese dictionary is shown in the rectangles on the right. Red characters in the figure correspond to recognition error. Capturing condition in the sense of image quality such as resolution or blur largely affects the accuracy of character recognition.

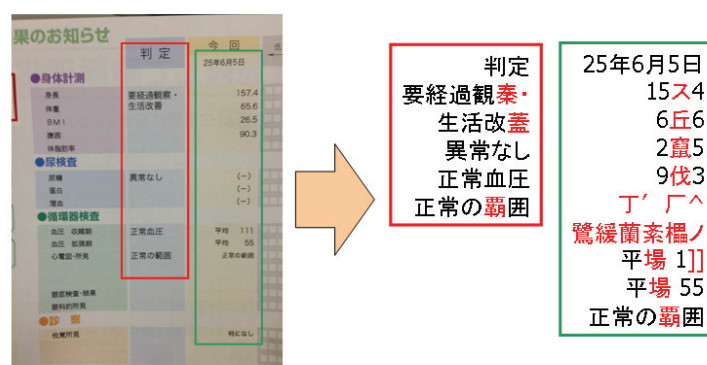


Fig. 6. An example of OCR processing

In order to improve the accuracy of OCR processing, adaptive dictionary configuration is carried out in OCR processor as shown in Figure 4. In case of recognizing characters in an annual health checkup report, we can expect a word or figure within a limited number of instances. For example, the possible words appeared in a report of annual health check-up only include item names such as 'height', 'weight', or 'BMI', but no other words. The number of possible word is quite limited compared with the case of general documents.

Adaptive dictionary configuration is based on this observation. Since the number of kanji characters (i.e., Chinese letters used in Japan) designated for everyday use is more than 2000, excluding Chinese letters that are not used in annual health check-up report from OCR dictionary contributes to improve the accuracy. Figure 7 illustrates how adaptive dictionary configuration is performed in OCR process. Candidate set of character, i.e., Chinese letter, Japanese character, and/or figures is associated with each sub-region in the guide frames as shown in Figure 8. In this Figure, the left side part for the area of anthropometric is associated with OCR region 1, word/value dictionary1, and item WL1 (white list 1) which define possible words/range of characters and values, respectively.

After an image is captured, it is segmented into sub-regions by following the guide frames, i.e., the spatial structure of the table. Then, each sub-image is sent to OCR processor with the meta-data as explained above. After that, the OCR engine switches the dictionary in accordance with the sub-region so as to converge to one of the characters that is supposed to appear in the region.

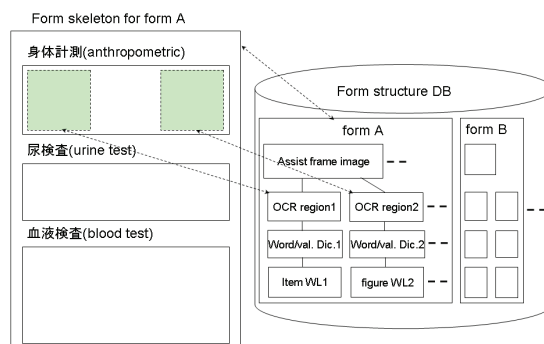


Fig. 7. Adaptive Dictionary Configuration for OCR processing

3.4 Data Confirmation and Correction by Display Synchronization

The adaptive dictionary configuration described in the previous section contributes to improve the OCR accuracy, however, it does not always make a user free from confirmation and correction of error. In other words, confirmation and correction process is mandatory step before registering the PHR data into the PHR database.



Fig. 8. Confirming Migrated PHR Data with Display Synchronization

Since the current implementation of Lico offers a simple, Web-based migration interface, the confirmation process impose a user a large amount of eye movement between a field displayed on the computer screen and that corresponds to it on the printed form. Keeping correspondence between them is a stressful and error-prone task because of the long distance of eye movement.

Our PHR front-end takes this issue into consideration, and designed the user interface as shown in Figure 8. There are two examples of aligning the captured image and the result of OCR processing; one is aligning them horizontally, and another is vertically. When a user drags either the captured image or the table that shows the result of recognition, both of them is scrolled synchronously on the display so as to minimize burden on keeping recognizing the correspondence between an item in the image and that in the resultant table.

4 Experimental Evaluation

4.1 Evaluation OCR Accuracy Improvement

The result of an experiment to evaluate the effectiveness of adaptive dictionary configuration is shown in Table 1. We used Tesseract-OCR [7] as OCR engine, since it performs the most precisely among OCR engines available. The standard dictionary configuration for recognizing Japanese (including Hiragana, Katakana, Kanji) and figure is carried out by specifying an option as '-l jpn'. Another standard option of '-l eng' is preferable in case for recognizing figures, simply because there are less candidates of characters to be recognized. We regard the results obtained by these configurations as baseline. In adaptive dictionary configuration, we configured dictionary so that the OCR engine refers to only the characters that are expected to be appeared in item name in case of item name recognition, and configured it so that it refers to only the figure (i.e., number from 0 to 9) and decimal in case of recognizing the region of numeric values. The result of experiment is shown in Table 1. Fractions beside the percentage denote the number of correctly recognized characters out of characters being tested.

Table 1. Comparison of Accuracy

	Standard config.	Adaptive config.
Item name	52% (456/876)	79% (692/876)
figures	82% (570/695)	91% (630/695)

4.2 User Evaluation

We conducted usability test for the visitors of technology exposition, which is called e-messe 2014 held at Kanazawa city in Ishikawa prefecture. We asked visitors who came to our booth to use our prototype system as well as current Lico system. The total number of subjects is 68, which consists of 51 male subjects and 14 female subjects. To avoid order effect, we let 33 subjects to use proposed system first, and let 31 subjects to use current Lico system first. We collected questionnaires from all these visitors, where the profile of the subjects with regard to age and proficiency or familiarity of IT devices is summarized in Table 2 and Table 3, respectively.

Table 2. Age Distribution of Subjects

~10	20-29	30-39	40-49	50-59	60-69	70~	Total
4	19	14	14	4	12	1	68
5.9%	27.9%	20.6%	20.6%	5.9%	17.6%	1.5%	

Table 3. Proficiency/familiarity to computers and communication device.

Device	N/A	upto once/week	2-3times /week	everyday	total
Desktop PC	17	6	3	40	66
Note/laptop PC	5	6	13	44	68
Tablet PC	43	5	4	15	67
Mobile phone	30	0	4	34	68
Smart phone	20	1	1	46	68

As the result of calculating summary math, about 78% of subjects evaluated that the proposed system is more useful, and 4.7% of subjects evaluated the current implementation of Lico system is easier to use as shown in Table. 4. Concerning the comparison between manual data migration by entering every value into a form and data migration by photography, about 93% of subjects evaluated data migration by photography is easier to use, whereas 3% of the subject still recognize the current, form based data migration is easier to use.

Table 4. Comparison of usability between two systems.

proposed	Current Lico	No difference	undecided	total
3	50	3	8	64
4.7%	78.1%	4.7%	12.5%	

Concerning the evaluation by acceptance of the accuracy of character recognition which is presented in section 4.1, about 67% subjects evaluated the OCR accuracy is acceptable, and only 9% of subject evaluated error correction is not negligible, as shown in Table 5. (Note that two subjects did not answer this inquiry.)

Table 5. Acceptance ratio of subject with regard to OCR accuracy.

acceptable	More errors than acceptable	undecided	total
44	6	16	66
66.7%	9.1%	24.2%	

5 Conclusion and Future Work

In this paper, we described a PHR system for end users to maintain their PHR data by their own effort. Proposed front-end system assumes to use tablet PC with camera module. Data migration by photography, which enables a user to migrate printed data by taking photos, is employed. It enables a user to migrate printed data more easily, since guide frames are displayed so that the user can understand which part of the form to take intuitively. Adaptive dictionary configuration improves OCR accuracy even for images captured by handheld camera device. Data confirmation and correction by synchronized display offers less stressful confirmation task before registering the PHR data into the database. According to the result of evaluation of proposed system, though it is still a prototype system, about 78% of subject rated that the proposed system is easier to use, and 93% of subjects evaluated that the data migration by photography is more useful than form based manual data migration.

Acknowledgment

The authors would like to express appreciation to Ms. Y. Nakada at the Industrial Collaboration Promotion Center, JAIST, for their management of this project. We would also like to acknowledge Mr. Y. Nishikawa, Mr. S. Urabe, and Mr. H. Yamashina for their effort of developing the system. This work is supported by the Strategic Information and Communications R&D Promotion Programme (SCOPE) supervised by the Ministry of Internal Affairs and Communications, Japan.

References

1. M. Tsiknakis, D. Katehakis, S.C. Orphanoudakis, "A health information infrastructure enabling secure access to the life-long multimedia electronic health record," CARS 2004 - Computer Assisted Radiology and Surgery, Proceedings of the 18th International Congress and Exhibition, Volume 1268, pp. 289–294, 2004.
2. Dean F. Sittig, "Personal health records on the internet: a snapshot of the pioneers at the end of the 20th Century", International Journal of Medical Informatics, Vol. 65, Issue 1, pp. 1–6, 2002.
3. J. S. Kahn, V. Aulakh and A. Bosworth, "What It Takes: Characteristics Of The Ideal Personal Health Record," Health Affairs, 28, no.2, pp. 369-376, 2009.
4. S. R Reti, H. J Feldman, S. E Ross, et al., "Improving personal health records for patient-centered care, "Journal American Medical Informatics Association, Vol. 17, pp. 192-195, 2010..
5. R. H Dolin, L. Alschuler, et al., "The HL7 Clinical Document Architecture," Journal American Medical Informatics Association, Vol. 8, pp. 552-569, 2001.
6. <https://www.l-cod.com/>
7. tesseract-ocr, <https://code.google.com/p/tesseract-ocr/>

Effect of Variant Factors on Multi-Document Summarization

[†]Nongnuch Ketui and [‡]Thanaruk Theeramunkong

[†]Faculty of Science and Agricultural Technology,
Rajamangala University of Technology Lanna, Nan and

[‡]School of Information, Computer and Communication Technology,
Sirindhorn International Institute of Technology, Thammasat University
`nongnuchketui@rmutl.ac.th`,
`thanaruk@siit.tu.ac.th`

Abstract. Multi-Document Summarization is challenge task for NLP researchers specially in Thai. In this paper, we study effect of variant factors i.e., weighting factors and unit-selection factors on Thai news summarization. To evaluate the proposed summarization by using fifty sets of Thai news articles with their manually constructed reference summary. Four ROUGE summarization evaluations, the results show that the iterative weighting gets higher performance of traditional TF-IDF, the node weight inclusion, query relevance, centroid-based selection, and unit redundancy consideration can help improving summary quality, and our summarization method is superior to the baselines.

Keywords: variant factor, multi-document summarization, iterative weighting

1 Introduction

Automatic summarization is the challenge task to natural language processing (NLP) community. A system of multi-document summarization may cover single-document summarization issues such as considering the relevant information, extracting the important content from a set of documents, and eliminating the redundant. In the past, multi-document summarization as a clustering problem was solved. The centroid-based techniques to generate a composite sentence from each cluster were proposed by [1, 11, 15]. They used features to modify sentence in each cluster while feature vectors were represented by a set of single words weighted by some weighting system such as TF-IDF. Carbonell and Goldstein [2] combined query relevance with information novelty as the topic and made a major contribution to topic-driven summarization by introducing the maximal marginal relevance (MMR) measure. Mani [9] presented an information extraction framework for summarization as well as a graph-based method to find similarities and dissimilarities in pairs of documents. Moreover, Mihalcea [13] introduced a new formula for a traditional graph-based ranking so-called

2 N.Ketui and T.Theeramunkong

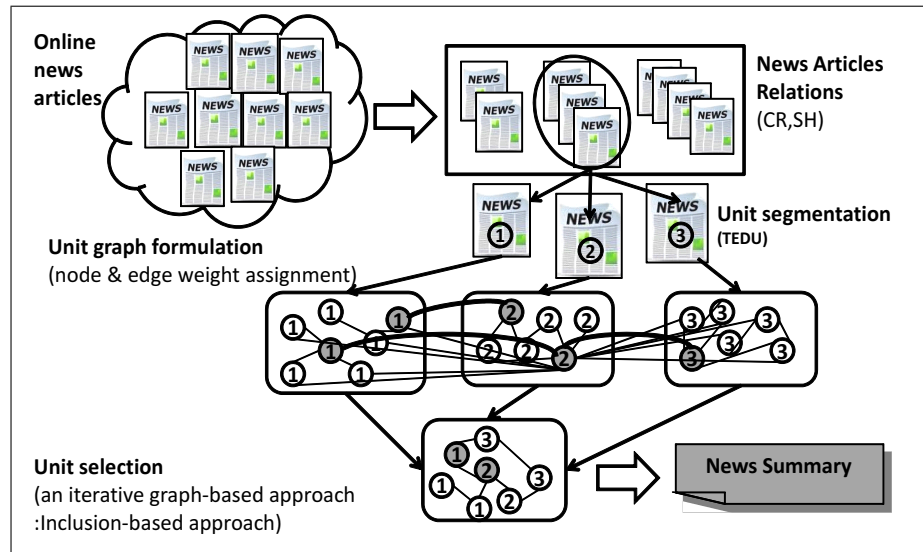


Fig. 1. The concept of Multi-Document Summarization.

TexRank that takes into account edge weights when computing the score associated with a vertex in the graph. The sentence scoring function was known as the PageRank algorithm [14].

In this work, we propose three processes of multi-document summarization. A Thai running text is segmented into Thai elementary discourse units (TEDUs) [5] as units in unit segmentation process. A graph-based approach is used to construct the graph unit while each node weight is assigned by the weighting factors (either TF-IDF or an iterative weighting) and the edge weight is calculated by the similarity measure. The inclusion-based approach using the unit-selection factors i.e., a PageRank consideration, node weight, query relevance, centroid-based selection, and redundancy removal is applied to unit selection process for improving the multi-document summarization. After selecting important non-redundant units, a summary is generated with factors of temporal factors and ordinal relations in original documents. To evaluate our method, a number of experiments are conducted using fifty sets of Thai news articles, the model summaries of which are given. Four measures of ROUGE-1, ROUGE-2, ROUGE-S, and ROUGE-SU4 are used as performance metrics.

In the rest, Section 2 describes the process of multi-document summarization. Variant factors are defined in Section 3. Experimental setup is shown in Section 4. Experimental results are discussed in Section 5. Finally, conclusion and future work are given in Section 6.

2 Multi-Document Summarization

This section presents our multi-document summarization model which is composed of three processes; 1) unit segmentation, 2) unit graph formulation, and 3) unit selection as shown in Figure 1. In the unit segmentation, a Thai running text is segmented into a sequence of tractable units and tagged with part-of-speech (POSS) and named entities (NEs). TEDUs are proposed for segmenting a Thai running text into units. In the unit graph formulation, a graph of units is constructed by conceptualizing a unit as a weighted node in a graph and a relationship of two nodes is formulated as a weighted link between the nodes. The weight of a node or a link is determined by considering its importance, i.e., its contribution in the graph. In the unit selection, a number of important nodes and links are selected by considering importance level of nodes or links, together with redundancy among units (nodes), and focusing on the node weight recalculation.

2.1 Unit Segmentation

Since Thai language has no sentence boundary (or even word boundary), it is difficult to define a unit for summarization. The most straightforward approach for segmenting blocks of Thai documents are paragraphs, as done [3]. However, using paragraphs as units for Thai document summarization is inflexible since a paragraph may contain heterogeneous contents and such paragraph-based approach does not allow us to exclude some unimportant parts in a paragraph. In this work, one alternative for unit segmentation in Thai language is to apply Thai Elementary Discourse Units (TEDUs) and Common Phrases (COMPs) [5] and segment a Thai running text into TEDUs+COMPs by using context-free grammar rules with a bottom-up chart parser with the longest matching technique [8] to detect TEDUs in a text, together with their structures.

2.2 Unit Graph Formulation

After a running text is split into tractable unit; TEDI+COMP, we need a model to determine importance of units and select the most suitable ones for a summary. Towards this, a graph model is proposed to express units and their relations extracted from multiple documents targeted for summarization. In our approach, a node in a graph corresponds to a unit while a link in a graph expresses connections between two units. Two subprocesses are implemented; (1) Node Weight Calculation and (2) Link Weight Calculation. For the node weight calculation process, we assign the node weight either TF-IDF (Term Frequency times Inverse Document Frequency) or the iterative weighting [4]. For the link weight calculation, it is worth investigating relations between two units (nodes). A link weight (relation strength) between two nodes (units) describes how much two units are identical or related. Although there have been several possibilities on definition of relations between units, this work simply uses a common method, namely cosine similarity. The cosine similarity ranges from 0 to 1 and the highest

4 N.Ketui and T.Theeramunkong

value indicates high similarity, implying that two units are duplicates or highly related.

2.3 Unit Selection

Given the unit graph derived from the process described in the previous subsection, unit selection is a task to select a set of suitable units for constructing a summary. In the past, several works [2] applied a straightforward method to select units based on their weights. In this work, we utilize a variant of the inclusion-based summarization method proposed in [4]. In the algorithm, the number of units to be selected is set. The most important nodes are repeatedly added one-by-one into the summary-band deleted from a set of unselected units until the number of selected units reaches the predefined compression rate. After the node addition, the weight of each unselected unit is recalculated. When the number of selected units satisfies the predefined compression rate, the algorithm returns the graph of summary. In the detail, the variant of the inclusion-based summarization factors are discussed in next section.

3 Variant Factors

In this section, we classify the variant factors into two groups; (1) weighting factors and (2) unit-selection factors as shown the detail below.

3.1 Weighting Factors

In node weight calculation, two node weighting factors are investigated below.

TF-IDF: As the most naive method, it is possible to apply TF-IDF (Term Frequency times Inverse Document Frequency) to weight words in a document and then weight a unit by calculating the summation of weights of all words in that unit.

Iterative Weighting: As a more sophisticated weighting method, we have proposed a so-called iterative weighting [4] to obtain more accurate weights of units by considering importance of documents, units, and words and reflect them when we weigh units.

3.2 Unit-Selection Factors

The multiple criteria of the inclusion-based summarization factors are discussed below.

PageRank consideration: A traditional graph-based ranking model applies the sentence scoring function for finding the relevance information [14] as shown in Eq. 1. Let $PR(s_i)$ is the ranking score of unit. s_j represents the

unit which relates to the considered unit s_i while s_k is the neighbour of s_j . Here, the constant d equals 0.001.

$$PR(s_i) = (1 - d) + \left(d \times \sum_{s_j \in Neighbour(s_i)} \frac{sim(v_{s_j}, v_{s_i})}{\sum_{s_k \in Neighbour(s_j)} sim(v_{s_j}, v_{s_k})} \right) \quad (1)$$

Node weighting: A node weight is assigned by either TF-IDF or the iterative weighting. It is possible to the high weight has a important information.

Query relevance: A unit combines query relevance with information as the title of a set of related news articles. The overlapped content between the unit and query may improve the text summarization since the query guides to find the main content.

Centroid-based selection: A preferable unselected unit is a node with high similarity to all other unselected units. As well as, the unit close to the centroid of the unselected units should be included in the summary. In contrast of the centroid-based selection, the unit far from the centroid of the unselected units should be considered to include in the summary.

Redundancy removal: Two units with an identical content or a highly similar content should not be selected simultaneously. In order words, it is reasonable to eliminate content redundancy in order to have a good short summary.

Reflecting the first concepts on weighting units, the unit selection can be formulated as follows. Here, the original weighting of a unit is modified to reflect the linkage page factor, query relevance factor, centroid-related factor, and redundancy-related factor as shown in Eq.(2). The best unit (\hat{u}) can be selected by maximizing the value in the equation, where the five terms indicates the PageRank consideration, its node weight, the query relevance, the centroid-based selection and the redundancy removal factors.

$$\hat{u} = \left(\lambda \times (PR(s_i))^\beta \times (W(s_i))^\phi \times ((\varepsilon + sim(v_q, v_{s_i})))^\varphi \right. \\ \left. \times \left[\left(1 - \frac{\sum_{k \neq 1, v \neq v_{s_i}}^{sim(v, v_{s_i})}}{|V|-1} \right)^\delta \times \left(\frac{\sum_{k \neq 1, v \neq v_{s_i}}^{sim(v, v_{s_i})}}{|V|-1} \right)^{1-\delta} \right]^\alpha \right) \\ - \left((1 - \lambda) \max_{s_j \in R} (sim(v_{s_j}, v_{s_i})) \right)^\omega \quad (2)$$

In the past, TF-IDF was usually used to express such importance levels of words by using term (word) frequency and inverse document frequency. As an alternative mentioned in [4], iterative weighting may be used. Such weight helps selecting suitable units for summarization. In Eq. 2, the constant (λ and ε) are 0.003 since the current unit weight may equal to 1. For multi-criteria terms, the first term expresses the original weight $W(s_i)$ which is assigned by either TF-IDF or iterative weighting ($\beta = 1$). The second term displays the PageRank consideration $PR(s_i)$ ($\phi = 1$) while the query relevance indicates in the third term ($\varphi = 1$). Let v_q is the word vector of query and v_{s_i} is the vector of considered unit. The fourth term explores the centroid/dicentroid/non-consideration of the

6 N.Ketui and T.Theeramunkong

unselected units. If δ equals 0 that we focus on the centroid-based selection, the discentroid-based selection can not consider. Let $|V|$ is the total number of unselected units. While α is 0, both the centroid and discentroid-based selection will not work. The fifth term represents the level of content redundancy ($\omega = 1$). In the extreme case, if the unit has similar content with any in the set of selected units, the maximum is one and the term will become 0.

4 Experimental Setup

4.1 Datasets

This work utilizes the THAI-NEST corpus developed in [18] which comprises 10,000 news articles in seven categories; crimes (CR), sports (SP), foreign affairs (FO), politics (PO), entertainment (EN), economics (EC), and education (ED), gathered from seventeen on-line news sources Later, a method for discovering document relations in [6] is applied to find relation between news documents and to group highly related news documents into a data set for summarization. While most previous works focused on finding document relations judged to be either relevant or non-relevant, this work classified documents into three main types of relations; (1) ‘completely related’ (CR), and (2) ‘somehow related’ (SH), and (3) ‘unrelated’ (UR). In this work, we randomly selected 50 sets of related documents with CR and SH relations for testing our proposed graph-based summarization approach. Given each set of related documents, the documents were tagged with POSs/NEs by Thai E-Class [19].

Later a Thai running text with POSs and NEs tagging is segmented into TEDUs, COMPs, and TEDU-LPs by using 446 context-free grammar rules (CFG rules) with a bottom-up chart parser [5] with the longest matching technique [8] to detect TEDUs in a text, together with their structures. Here, the CFG rules are built based on the syntactic categories defined by a Thai E-Class [19]. We applied three groups of CFG rules; 342 rules for TEDUs, 95 rules for COMPs, and 9 rules for TEDU-LPs. To form a reference summary (i.e., model summary) for evaluation, we have asked a number of Thai language experts in the Faculty of Liberal Art, Thammasat University to manually summarize the prepared set of news articles as gold standard for evaluating system results. The summarizers were instructed to construct an abstractive-based summary with the size of 20-100 words for each of the fifty datasets. The summaries contain main contents in the original documents. Some discourse markers are added to connect clauses. These reference summaries are used for evaluating a summary obtained from the system.

4.2 Experimental Settings

To examine effect of weighing factor and unit-selection factors on text summarization, we have conducted two experiments using 50 sets of related news documents, containing 23,781 words. The first experiment targets to compare

performance of top 20 ranked combination methods of the weighting factors and the five unit-selection factors. Moreover, the performances of twenty combinations are summarized and ranked to clarify which factor combination is optimal. The type of unit is based on TEDUs+COMPs. The two weighting factors we consider are (1) TF-IDF ('TF'), or (2) the iterative weighting ('IT') for assigning to the node weights. The five unit-selection factors we explore are (1) the PageRank consideration ('PR'), (2) the node weighting ('NW'), (3) the query relevance ('QR'), (3) the centroid-based selection ('CS'), and (4) the redundancy removal ('RR'). For the compression rate, we examine ten rates of 0.1 to 1.0. Here, the compression rate is defined as the ratio of the number of units in a summary to the number of units in the original documents. Moreover, we compare our summarization methods with a basic extractive summary called the first n character which the summary is composed of the ranked units by position in the original document, a traditional graph-based ranking model called the PageRank algorithm [14] and a query-based model called the maximal marginal relevance (MMR) measure [2]. The second experiment aims to investigate summarization performance according to TEDUs+COMPs unit type, weighing factor, and five unit-selection factors, and ten compression rates.

4.3 Evaluation Method

To evaluate a summary output from a system, we use the reference summaries as described in Section 4.1. A reference summary is an abstractive-based summary (with consideration of semantics; i.e., what, where, who, whom, why, and how) constructed by requesting Thai language experts to manually summarize a set of related news articles into 20-100 words. In this work, we utilize a standard metric, called ROUGE [7], to evaluate a system's summarization result by comparing it with its reference summary. Among various types of ROUGE, this work uses ROUGE-1 (unigram-based co-occurrence statistics), ROUGE-2 (bigram-based co-occurrence statistics), ROUGE-S (skip-bigram-based co-occurrence statistics), and ROUGE-SU4 (skip-bigram plus unigram-based co-occurrence statistics). Originally developed by NIST, the ROUGE we used is a new variant of ROUGES that considers precision ($R-1_P$, $R-2_P$, $R-S_P$, $R-SU4_P$), recall ($R-1_R$, $R-2_R$, $R-S_R$, $R-SU4_R$), and F-score ($R-1_F$, $R-2_F$, $R-S_F$, $R-SU4_F$).

5 Experimental Results

5.1 Effect of Variant Factors

Table 5.1 shows the effect of weighting factors and unit-selection factors to the summarization methods. $R-1_F$, $R-2_F$, $R-S_F$, $R-SU4_F$, and the average of four $ROUGE_F$ values of top-20 summarization methods are illustrated. Our methods are compared with the first n character, PageRank algorithm, and MMR method. For the weighting factors, the performance of iterative weighting ('IT') (the top-9 summarization methods) is higher than the standard TF-IDF weighting ('IF'). Since 'IT' weighting considers the important documents, units, and

8 N.Ketui and T.Theeramunkong

Table 1. Classification accuracy of the 20 best weighting factors and unit-selection factors

Methods	Weighting Factors	Unit-Selection Factors					R-1 _F	R-2 _F	R-S _F	R-SU4 _F	AVG
		PR	NW	QR	CS	RR					
MDS1	IT	N	Y	Y	C	Y	0.2946	0.1473	0.1159	0.2137	0.1929
MDS2	IT	N	Y	N	C	Y	0.2924	0.1477	0.1155	0.2126	0.1921
MDS3	IT	N	Y	Y	N	Y	0.2910	0.1457	0.1146	0.2111	0.1906
MDS4	IT	Y	Y	Y	N	Y	0.2913	0.1449	0.1136	0.2105	0.1901
MDS5	IT	Y	Y	Y	D	Y	0.2912	0.1448	0.1135	0.2105	0.1900
MDS6	IT	N	Y	Y	D	Y	0.2903	0.1452	0.1137	0.2104	0.1899
MDS7	IT	Y	Y	Y	C	Y	0.2909	0.1449	0.1131	0.2105	0.1899
MDS8	IT	N	Y	N	D	Y	0.2891	0.1459	0.1129	0.2092	0.1893
MDS9	IT	N	Y	N	N	Y	0.2889	0.1456	0.1127	0.2088	0.1890
MDS10	TF	N	N	N	C	Y	0.2879	0.1433	0.1122	0.2103	0.1884
MDS11	IT	N	N	Y	C	Y	0.2882	0.1417	0.1124	0.2107	0.1882
MDS12	IT	Y	N	Y	N	Y	0.2874	0.1421	0.1128	0.2105	0.1882
MDS13	IT	Y	Y	N	C	Y	0.2880	0.1438	0.1120	0.2086	0.1881
MDS14	IT	Y	N	Y	D	Y	0.2863	0.1422	0.1123	0.2099	0.1877
MDS15	IT	Y	N	Y	C	Y	0.2868	0.1404	0.1119	0.2108	0.1875
MDS16	TF	N	N	Y	C	Y	0.2880	0.1410	0.1086	0.2081	0.1864
MDS17	TF	N	Y	N	D	Y	0.2800	0.1451	0.1133	0.2051	0.1859
MDS18	TF	N	Y	N	N	Y	0.2800	0.1451	0.1133	0.2050	0.1858
MDS19	IT	Y	Y	N	D	Y	0.2866	0.1409	0.1095	0.2057	0.1857
MDS20	IT	Y	Y	N	N	Y	0.2865	0.1408	0.1095	0.2056	0.1856
FNC	-	-	-	-	-	-	0.2802	0.1438	0.1150	0.2098	0.1872
PR	TF	Y	N	N	N	N	0.2450	0.1211	0.0968	0.1931	0.1640
MMR	TF	N	N	Y	N	Y	0.2767	0.1382	0.1098	0.2021	0.1817

words in a set of related document, TF-IDF only focuses on the word occurrence in documents. For the unit-selection factors, the summarization does not need the PageRank factor ('PR') since the extractive summarization considers the importance information (not focus on the linkage unit). While the top-10 summarization methods include the node weight factor ('NW'), they achieve the high R-1_F, R-2_F, R-S_F, and R-SU4_F values. For the query relevance factor ('QR'), it can be improve the performance of summarization methods (the MDS1-MDS7 except MDS2) because the query as title of a set of news article represents the main content. To consider the centroid-based selection factor ('CS'), even though the MDS1 method having this factor achieves the highest average of four ROUGEs performance, the dis-centroid selection and non-consideration have occurred in the top-10 summarization methods. In the contrast of the previous factors, the redundancy removal ('RR') absolutely improves the performance of summarization methods since the elimination of duplicated information should be considered in multi-document summarization.

To conclude the effect of unit-selection factors, four factors (the node weight, query relevance, centroid-based selection, and redundancy removal) can affect to get the performance of summarization method highly. Comparing the perfor-

mance of summarization methods and the baselines, the average of four ROUGE values of the MDS1-MDS14 methods are higher than the first n character ('FNC') while PageRank ('PR') and MMR get the lowest performance.

Table 2. ROUGE performance of text summarization methods and compression rates.

ROUGE Method		Compression Rate									
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
R-1	MDS1	0.3075	0.3368	0.3361	0.3351	0.3216	0.3042	0.2813	0.2608	0.2406	0.2218
	FNC	0.2532	0.3105	0.3206	0.3263	0.3189	0.2941	0.2711	0.2504	0.2351	0.2218
	PR	0.2366	0.2594	0.2641	0.2595	0.2553	0.2516	0.2439	0.2316	0.2259	0.2218
	MMR	0.2186	0.2820	0.2993	0.3261	0.3216	0.3094	0.2839	0.2615	0.2428	0.2218
R-2	MDS1	0.1316	0.1499	0.1550	0.1596	0.1584	0.1567	0.1514	0.1452	0.1371	0.1283
	FNC	0.1165	0.1500	0.1529	0.1623	0.1632	0.1528	0.1455	0.1360	0.1306	0.1283
	PR	0.0846	0.1019	0.1188	0.1234	0.1299	0.1343	0.1332	0.1279	0.1288	0.1283
	MMR	0.0987	0.1152	0.1227	0.1523	0.1603	0.1637	0.1551	0.1464	0.1388	0.1283
R-S	MDS1	0.1047	0.1164	0.1210	0.1259	0.1237	0.1228	0.1189	0.1138	0.1090	0.1030
	FNC	0.0929	0.1195	0.1250	0.1297	0.1307	0.1201	0.1153	0.1088	0.1045	0.1030
	PR	0.0684	0.0794	0.0961	0.0995	0.1026	0.1077	0.1069	0.1014	0.1025	0.1030
	MMR	0.0778	0.0908	0.0966	0.1222	0.1276	0.1292	0.1227	0.1167	0.1119	0.1030
R-SU4	MDS1	0.1929	0.2180	0.2237	0.2314	0.2265	0.2231	0.2150	0.2077	0.2009	0.1977
	FNC	0.1652	0.2120	0.2221	0.2294	0.2278	0.2211	0.2152	0.2069	0.2006	0.1977
	PR	0.1452	0.1753	0.1910	0.2024	0.2052	0.2093	0.2069	0.1989	0.1986	0.1977
	MMR	0.1348	0.1792	0.1945	0.2272	0.2295	0.2282	0.2181	0.2093	0.2027	0.1977

5.2 Performance of Text Summarization Methods

In this experiment, we investigate the performance of summarization methods, and ten compression rates (0.1-1.0). R-1_F, R-2_F, R-S_F, and R-SU4_F values of the best summarization method ('MDS1') and three baselines i.e., the first n character ('FNC'), the PageRank algorithm ('PR'), and the query-based approach ('MMR') are shown in Table 2. Here, the bold font shows the highest performance of ROUGE value. For R-1_F value, TEDUs+COMPs with MDS1 gets the highest R-1_F performance at the compression rate of 0.2 (0.3368). Both FNC and MMR are the best performance at the compression rate of 0.4 (0.3263 and 0.3261, respectively). PR is the lowest performance at the compression rate of 0.3 (0.2641). For R-2_F value, even though R-2_F value of MDS1 (0.1596) is lower than FNC and MMR (0.1632 and 0.1637, respectively), our method gains the high R-2_F performance at the lower compression rate (at the compression rate 0.4). Normally, the good summary is less than 50% of original text. As well as R-2_F value, TEDUs+COMPs with FNC achieves the highest performance of R-S_F at the compression rate 0.5 (0.1307) while R-S_F value of MDS1 is higher than the baseline (PR) at the compression rate 0.4. For R-SU4_F value, TEDUs+COMPs with MDS1 is superior to other methods with R-SU4_F performance of 0.2314 (at the compression rate 0.4). Both MMR and FNC performance are similar (0.2295 and 0.2294, respectively) at the compression rate 0.4-0.5 while PR achieves the lowest performance of R-SU4_F at the compression rate 0.6 (0.2093).

We conclude that MDS1 obtains the best performance of R-1_F and R-SU4_F at the compression rate 0.2 and 0.4, respectively while R-2_F and R-S_F values of

10 N.Ketui and T.Theeramunkong

MDS1 are lower than MMR and FNC. Another baseline ('PR') gets the lowest performance of all ROUGE values.

6 Conclusion and Future Work

This paper introduced three processes of multi-document summarization, i.e., unit segmentation, unit graph formulation, and unit selection for summarization. We applied unit type as Thai elementary discourse units (TEDUs+COMPs). These units are represented as nodes and their relationships are formed as links among units with weights. In this work, we investigated the effect of weighting factors; TF-IDF and the iterative weighting and considered five unit-selection factors; the PageRank consideration, the node weighting, the query relevance, the centroid-based selection, and the redundancy removal. Using fifty sets of Thai news articles, the results showed that TEDUs+COMPs with the MDS1 method (composed of iterative weighting, the PageRank non-consideration, the node weight inclusion, the query relevance, the highest-weight priority with centroid preference, and the redundancy removal) yielded the best performance in $R-1_F$, $R-S_F$, and $R-SU4_F$ evaluations while the MDS2 method (having the same factors as MDS1 except the query relevance) achieved the highest $R-2_F$ value. In future works, we will analyze the relation between TEDUs in order to form a more suitable set of combined TEDU with consideration of semantic. Finally, it is worth exploring our approach on a larger dataset.

Acknowledgments.

This work was supported by the National Research University Project of Thailand Office of Higher Education Commission and Rajamangala University of Technology Lanna Nan. We would like to thank to all members at KINDML laboratory at Sirindhorn International Institute of Technology for fruitful discussion.

References

1. Barzilay, R.–McKeown, K. R.–Elhadad, M.: Information fusion in the context of multi-document summarization. In: Proc. of the 37th Annual Meeting of the ACL, 1999, pp. 550–557.
2. Carbonell, J. and Goldstein, J.: The use of MMR, diversity-based reranking for reordering documents and producing summaries. Research and Development in Information Retrieval, 1998, pp. 335–336.
3. Jaruskulchai, C.–Kruengkrai, C.: A practical text summarizer by paragraph extraction for Thai. In: Proc. of the sixth international workshop on Information retrieval with Asian languages (AsianIR '03), Vol. 11, 2003, pp. 9–16.
4. Ketui, N.–Theeramunkong, T.: Inclusion-based and exclusion-based approaches in graph-based multiple news summarization. In: Knowledge, Information and Creativity Support Systems, LNCS 6746, 2010, pp. 91–102.

5. Ketui, N.–Theeramunkong, T.–Onsuwan, C.: Thai Elementary Discourse Unit Analysis and Syntactic-based Segmentation, *InformationAn International Interdisciplinary Journal (INFORMATION-TOKYO)*, Vol. 16(10B), 2013, pp. 7423–7436.
6. Kittiphattanabawon, N.–Theeramunkong, T.–Nantajeewarawat, E.: News Relation Discovery Based on Association Rule Mining with Combining Factors. *IEICE Transactions*, Vol. 94-D(3), 2011, pp. 404–415.
7. Lin, C.-Y.: Rouge: A package for automatic evaluation of summaries. In: *Proc. of ACL Workshop on Text Summarization*, 2004, pp. 74–81.
8. Maier, D.: The complexity of some problems on subsequences and supersequences. *Journal of Association for Computing Machinery*, Vol. 25(2), 1978, pp. 322–336.
9. Mani, I.: Multi-document summarization by graph search and matching. In: *Proc. of the Fifteenth National Conference on Artificial Intelligence (AAAI-97)*, 1997, pp. 622–628.
10. Mann W. C.–Thompson S. A.: Rhetorical Structure Theory: Toward a functional theory of text organization, *Text*, Vol. 8(3), 1988, pp. 243–281.
11. McKeown, K.–Klavans, J.–Hatzivassiloglou, V.–Barzilay, R.–Eskin, E.: Towards multidocument summarization by reformulation: Progress and prospects. *AAAI/IAAI*, 1999, pp. 453–460.
12. Meknavin, S.–Charoenpornasawat, P.–Kijirikul, B.: Feature-based Thai word segmentation. In: *Proc. of the Natural Language Processing Pacific Rim Symposium (NLP RS '97)*, 1997.
13. Mihalcea, R.: Graph-based ranking algorithms for sentence extraction, applied to text summarization. In *Proc. of the ACL 2004 on Interactive poster and demonstration sessions, ACLDEMO '04*, Association for Computational Linguistics, Stroudsburg, PA, USA, 2004.
14. Page, L.–Brin, S.–Motwani, R.–Winograd, T.: The PageRank Citation Ranking: Bringing Order to the Web. Technical report, Stanford University, 1998.
15. Radev, D. R.–Jing, H.–Budzikowska, M.: Centroid-based summarization of multiple documents: Sentence extraction, utility-based evaluation, and user studies. In: *Proc. of NAACL-ANLP 2000 Workshop on Automatic summarization*, 2000, pp. 21–30.
16. Suwanno, N.–Suzuki, Y.–Yamazaki, H.: Extracting Thai compound nouns for paragraph extraction in Thai text. In: *Proc. of IEEE International Conference on Natural Language Processing and Knowledge Engineering (IEEE NLP-KE '05)*, 2005, pp. 657–662.
17. Thangthai, A.–Jaruskulchai, C.: Impact parameter on LSA performance for Thai text summarization. In: *Proc. of the 43rd Kasetsart University Annual Conference: Veterinary Medicine, Science (Vichakarn '43)*, 2004, pp. 331–339.
18. Theeramunkong, T.–Boriboon, M.–Haruechaiyasak, C.–Kittiphattanabawon, N.–Kosawat, K.–Onsuwan, C.–Siriwat, I.–Suwanapong, T.–N.Tongtep: Thai-nest: A framework for Thai named entity tagging specification and tools. In: *Proc. of the 2nd International Conference on Corpus Linguistics (CILC '10)*, University of A Coruna, Spain, 2010, pp. 895–908.
19. Tongtep, N.–Theeramunkong, T.: Multi-stage automatic NE and POS annotation using pattern-based and statistical-based techniques for Thai corpus construction. *IEICE Transaction on Information and Systems*, Vol. E96-D(10), 2013, pp. 2245–2256.

A Model of Character Affinity for Agent-Based Story Generation

Gonzalo Méndez, Pablo Gervás, and Carlos León

Facultad de Informática, Universidad Complutense de Madrid, Madrid, Spain,
 {gmendez,pgervas,cleon}@ucm.es,
 WWW home page: <http://nil.fdi.ucm.es>

Abstract. One of the aspects that is used to keep the reader's interest in a story is the network of relationships among the characters that take part in that story. We can model the relationship between two characters using their mutual affinities, which allow us to define which interactions are possible between two characters. In this paper we present a model to represent characters' affinities and we describe how we have implemented this model using a multi-agent system that is used to generate stories. We also present the result of one experiment to measure the evolution of the affinities between two characters throughout a story.

Keywords: computational creativity, narrative, story generation, multi-agent simulation, character affinities

1 Introduction

In the book *The Thirty-Six Dramatic Situations* [11] Polti explores the assertion made by Gozzi (author of *Turandot*) saying that there can only be thirty-six tragic situations. Polti analyses what these thirty-six situations are, their variations, and what characters are involved. At the end of the book, he begins his conclusions by saying that, to obtain the nuances of the situations, the first thing he did was to “*enumerate the ties of friendship or kinship between the characters*”. A century before that, Goethe had already proposed his theory (maybe just metaphorical) of *elective affinities* [15] to depict human relations, specially marriages, and he showed how affinities between characters can be represented by a topological chart.

Even in modern tv shows which expand for several seasons, one of the aspects that create more engagement with spectators are the relationships that exist between characters and the way in which they evolve from the beginning of the first season to end of the last one.

Through all of these examples, we can see that the affinity between characters is an important factor to take into account when generating stories, and one that can help us to maintain the necessary narrative tension to keep the reader interested in the story.

In the following sections, we present a model of character affinities and the way in which we have implemented it using a multi-agent system that is used to generate stories based on the relationships between characters.

2 Related Work

The first story telling system for which there is a record is the Novel Writer system developed by Sheldon Klein [4]. Novel Writer created murder stories within the context of a weekend party. It relied on a micro-simulation model where the behaviour of individual characters and events were governed by probabilistic rules that progressively changed the state of the simulated world. The particular murderer and victim depended on the character traits specified as input. The motives arise as a function of the events during the course of the story. The set of rules is highly constraining, and allows for the construction of only one very specific type of story. Personality characteristics are explicitly represented but marked as “not to be described in output”.

TALESPIN [9], a system which told stories about the lives of simple woodland creatures, was based on planning: to create a story, a character is given a goal, and then the plan is developed to solve the goal. TALESPIN introduces character goals as triggers for action. Actions are no longer set off directly by satisfaction of their conditions. The system allows the possibility of having more than one problem-solving character in the story. The validity of a story is established in terms of: existence of a problem, degree of difficulty in solving the problem, and nature or level of problem solved.

Lebowitz’s UNIVERSE [6] modelled the generation of scripts for a succession of TV soap opera episodes (a large cast of characters play out multiple, simultaneous, overlapping stories that never end). UNIVERSE is the first storytelling system to devote special attention to the creation of characters. It is aimed at exploring extended story generation, a continuing serial rather than a story with a beginning and an end. Plot fragments provide narrative methods that achieve goals, but the goals considered here are not character goals, but author goals. This is intended to allow the system to lead characters into undertaking actions that they would not have chosen to do as independent agents (to make the story interesting, usually by giving rise to melodramatic conflicts).

MEXICA [10] is a computer model designed to study the creative process in writing in terms of the cycle of engagement and reflection [12]. MEXICA was a pioneer in that it takes into account emotional links and tensions between the characters as means for driving and evaluating ongoing stories. The reflection phase revises the plot so far, mainly checking it for coherence, novelty and interest. The checks for novelty and interest involve comparing the plot so far with that of previous stories. The check for coherence is only carried out over the final version of the story, and it involves inserting into the text actions that convey explicitly either character goals or tensions between the characters that are necessary to understand the story.

The Virtual Storyteller [14] introduces a multi-agent approach to story creation where a director agent is introduced to look after plot. Each agent has its own knowledge base and rules to govern its behaviour. In particular, the director agent has basic knowledge about plot structure and exercises control over agent’s actions in one of three ways: environmental (introduce new characters and object), motivational (giving characters specific goals), and proscriptive (dis-

allowing a character's intended action). The director has no prescriptive control (it cannot force characters to perform specific actions).

Comme il Faut (CiF) [8] is a knowledge-based system that models the interplay between social norm, social interactions, character desires, and cultural background. The underlying model of social interaction covers a range of aspects, from cultural static knowledge relevant to social interaction to fleeting desires of characters, with models for intervening factors like social exchanges, microtheories for significant concepts (such as friendship), and set of rules capturing likely behaviours of characters when faced with particular social circumstances.

Stella [7] performs story generation by traversing a conceptual space of partial world states based on narrative aspects. World states are generated as the result of non-deterministic interaction between characters and their environment. This generation is narrative agnostic, and an additional level built on top of the world evolution chooses the most promising ones in terms of their narrative features. Stella makes use of objective curves representing these features and selects world states whose characteristics match the ones represented by these curves. Stella is aligned to the current approach in the sense that simulation is also the base for generation. Stella, however, does not address characters' interactions as a key feature in the creative process.

3 A Model of Character Affinity

When running intelligent agents in simulations, and specially when they are in the form of intelligent virtual agents within virtual environments, some authors report the impossibility to run but a few of them at the same time [3, 2], since all the artificial intelligence involved in making them intelligent implies a high computational cost. One of our concerns when designing the current model has been for it to be as light-weight and cost-effective as possible, so its combination with other artefacts to create intelligent characters with personality traits, emotions and complex decision making maintains a low computational cost.

One of the most relevant research works on the subject is *Thespian* [13], the social behaviour framework used in [3]. In this work, the authors describe the use of an affinity factor to model social interaction which affects how characters can behave with each other. In this case, affinity is affected by other factors, such as social obligations and characters goals.

The first approach we used was to model affinity as a set of symbolic values that would be subsequently used to reason about the character's actions. The advantage of this approach is that it is easier to understand and reason about what is happening in the simulation. However, it is more difficult to operate with these values, a certain semantic has to be added to the code to understand how these values change and, on the long run, symbolic reasoning tends to be slow when combined with other processes.

Therefore, we have opted for a numeric representation that allows us to use common arithmetic operators to modify the degree of affinity between characters. The main drawback of this approach is that it is more difficult to calibrate

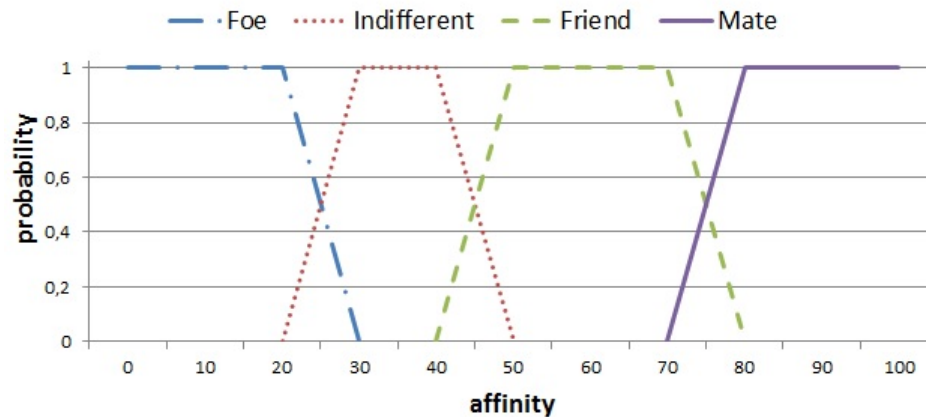


Fig. 1. Model of affinity

the model and interpret what is happening in the simulation. To reduce this drawback, we have opted for a representation similar to the fuzzy concepts proposed in [16], as shown in Fig. 1, an approach that has already been used by other authors to model cognitive architectures [2, 1].

We have modelled four levels of affinity according to four different kinds of affinity: foe (no affinity), indifferent (slight affinity), friend (medium affinity) and mate (high affinity). These four levels of affinity overlap on their limits, which allows for relationships not to change constantly when moving around the limits of two different levels. Therefore, the change from *indifferent* to *friend*, takes place when the affinity value is 70, and changing from *friend* to *indifferent* is done with an affinity value of 50.

An additional aspect of affinity is that it is not symmetrical. Given two characters, their mutual affinity is likely to have different values and it may even be situated in different levels, with the exception of mates: *character A* considers *character B* as its mate only if *character B* considers *character A* as its mate, too. However, if they are not mates, *character A* may think *character B* is a friend, while *character B* may think *character A* is a foe.

There are two ways in which the affinity value can change. The first one is by lack of interaction, in which case the affinity value moves towards the indifferent level. The second one is through interactions among characters, that obey a few simple rules. There is a set of interactions that is appropriate for each affinity level, so when dealing with a friend a character may only propose to carry out friend actions, but not mate actions. In addition, characters ignore proposals that do not correspond to their perceived affinity level, and receiving such proposals may penalise the affinity with the character proposing them. The exception to this rule are foes, who carry out what they intend to do irrespective of what the other character may want. When receiving a proposal, a character may decide to either accept or reject it. If the proposal is accepted, both characters increase

their mutual affinity. If it is rejected, the proposer will penalise its affinity with the receiver. Actions for the same level of affinity have different impact on it. For example, a romantic dinner has a higher effect on affinity than watching tv together. Similarly, the negative effect of rejecting an invitation is opposite to the positive effect of accepting it.

4 Implementation of the Model

The described model has been implemented by means of a multi-agent system developed using JADE¹.

The main objectives of the implementation were: to test the model apart from other factors such as the environment in which the story takes place or the personality traits and emotional state of the characters, which cause them to make different decisions in the same situation; and to implement the model as independent as possible from the domain of the story, so it can be easily used to generate different kinds of stories.

The system consists of two types of agents: a Director Agent, which is in charge of setting up the execution environment and creating the characters; and Character Agents, one for each character of the story, which are the ones that interact to generate the story. In the current case, the story consists of a set of interactions that make the affinity between characters change accordingly.

The information that the Director Agent needs to set up the execution environment is written in a text file that contains the number and names of the characters that have to be created. Subsequently, the information needed to configure each character is also included in a text file that currently contains the name and gender of the character, the name and affinity with its mate, a list of friend names and affinities and a list of foe names and affinities.

Each Character Agent is endowed with three different behaviours: one to interact with its mate, another one to interact with its friends and the last one to interact with its foes. Each behaviour is independent from the others, and they can all be added, blocked and removed dynamically to keep the system as lightweight as possible. These behaviours run the interaction protocols needed to implement and test the affinity model. The information needed by these behaviours, mainly the actions that characters can perform when executing them, along with the degree in which these actions affect the affinity between characters, is also stored in text files, so it is easy to add and remove actions and modify their influence on the affinity without having to change the code and recompile the system. This also means that, as far as the affinity model is concerned, actions have no semantic apart from their influence on the affinity value.

In Fig. 2 we can see how the MateBehaviour works. When a character receives a message from its mate, it checks whether it is a proposal to do something together or not. If it is, it may accept it, in which case it increases its affinity with its mate, or decline. In both cases, the decision is notified to its mate. When

¹ <http://jade.tilab.com>

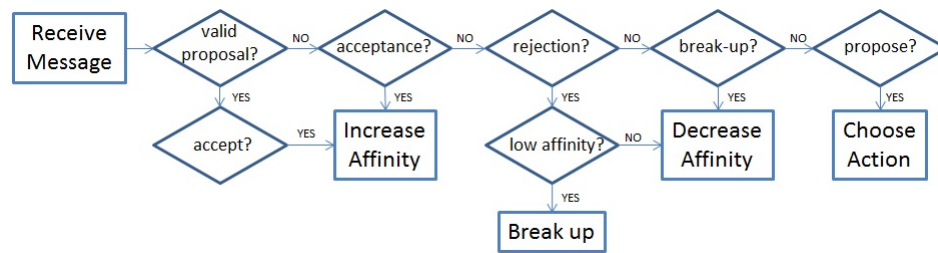


Fig. 2. Interaction protocol for mate characters

a character receives an acceptance, it increases its affinity with its mate, whereas if it is a rejection, it checks its affinity with its mate and, if it is already below a given threshold, it will decide to break up with its mate. When an agent receives a break-up notification, it decreases its affinity with its mate and decides whether to make it its friend or its foe. If none of this has happened, the character may then decide to propose its mate to do some activity together.

In the points where the characters should make a decision, such as whether to accept a proposal to do something or not, a random probabilistic decision has been made in order to be able to test the implementation of the affinity model by itself, without the interference of other processes. Thus, for example, the probability of accepting a proposal of the character's mate has been empirically established in 0.6. The reason for this value in our experiments is that it is high enough for couples to remain fairly stable, but it is low enough to keep things happening, so that stories don't turn boring.

Running the implemented system with 15 characters (8 females and 7 males forming 7 couples), we have chosen the couple formed by two of them, Betty and Clark, to show the evolution of their affinity over time, as shown in Fig. 3. The image shows how their affinity varies between 80 and 100 over the execution, but both affinities evolve separately (although they are not completely independent). In general, affinity increases at a lower speed than it decreases. This is due to two causes: the heavier impact that rejections have on the affinity than acceptances; and the fact that, if no other action is taken, affinity slowly fades over time, which affects the overall decreasing speed. The most remarkable fact that can be appreciated in the image is how, at the end of the execution, the affinity between both characters falls dramatically due to the final break-up of the couple. This break-up is caused by Clark's rejections of several of Betty's proposals, which in turn causes Betty's decision to put an end to the relationship, once the affinity level has gone below the threshold of mate affinity.

5 Conclusions

In the previous sections we have described a model to express characters relations based on their affinity, and we have shown how this model has been implemented using a multi-agent system to that generates character-based stories.

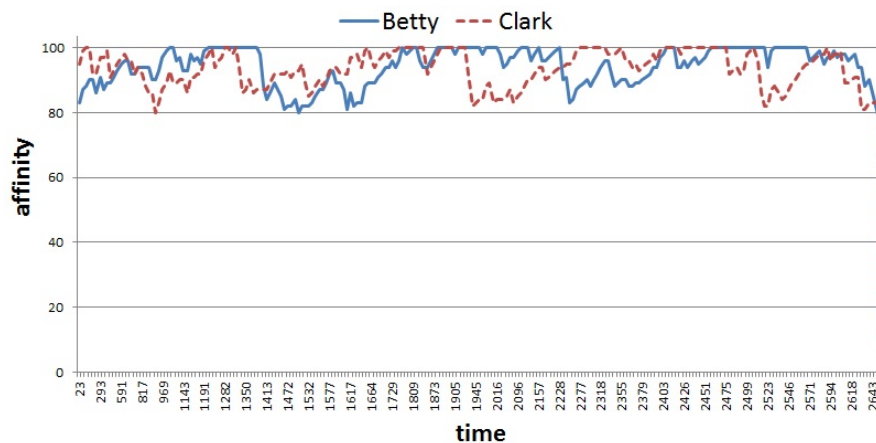


Fig. 3. Evolution of the affinity between characters Betty and Clark

We have run the implemented system with up to 15 characters and the results show that the possible interactions are rich enough to generate a high variety of stories. In addition, it is possible to change most of the information needed to generate the stories through configuration files, which makes it easy to produce new stories with different situations almost effortlessly.

In particular, we have seen that the model can be configured in such a way that it keeps relationships stable, but it allows enough flexibility so that unexpected events can happen to make the plot more interesting.

6 Future Work

The model can be further improved. More relations can be included in the system and a more refined selection of them can be tried and evaluated. The results on how the selection of features affects the complexity and the number of generated stories can shed light on what the set of relevant aspects of affinity are.

There is still much work to be done in order to generate stories that are not only based on character relationships. We will start by integrating the work described in this paper with the generator presented in [5], which will allow us to situate characters within a map and a context, giving us the chance to generate interactions only when proximity makes them possible.

We intend to endow characters with personality traits and emotions, in order to complement the affinity model and give characters the possibility to make decisions in a more cognitive way. We plan to use an approach similar to the one described in [2] to model emotions, so that it can be easily integrated with the present model and the implementation can maintain a low computational cost.

Finally, since we are capable of generating a high variety of different stories, we need to develop a mechanism to evaluate these stories in order to discard

those that lack interest and to refine the generation mechanism so that less non-interesting stories are generated.

Acknowledgements

This paper has been partially supported by the project WHIM 611560 funded by the European Commission, Framework Program 7, the ICT theme, and the Future Emerging Technologies FET program.

References

1. M. S. El-Nasr, J. Yen, and T. R. Ioerger. Flame - fuzzy logic adaptive model of emotions. *Autonomous Agents and Multi-agent systems*, 3(3):219–257, 2000.
2. R. Imbert and A. de Antonio. An emotional architecture for virtual characters. In *International Conference on Virtual Storytelling*, pages 63–72, 2005.
3. W. L. Johnson and A. Valente. Tactical language and culture training systems: Using artificial intelligence to teach foreign languages and cultures. In *AAAI*, pages 1632–1639, 2008.
4. S. Klein, J. F. Aeschliman, D. Balsiger, S. L. Converse, C. Court, M. Foster, R. Lao, J. D. Oakley, and J. Smith. Automatic novel writing: A status report. Technical Report 186, Computer Science Dept., The University of Wisconsin, December 1973.
5. I. M. Laclaustra, J. L. Ledesma, G. Méndez, and P. Gervás. Kill the dragon and rescue the princess: Designing a plan-based multi-agent story generator. In *5th International Conference on Computational Creativity*, Ljubljana, Slovenia, 2014.
6. M. Lebowitz. Storytelling as Planning and Learning. *Poetics*, 14:483–502, 1985.
7. C. León and P. Gervás. Creativity in story generation from the ground up: Non-deterministic simulation driven by narrative. In *5th International Conference on Computational Creativity, ICCO 2014*, Ljubljana, Slovenia, 06/2014 2014.
8. J. McCoy, M. Treanor, B. Samuel, A. A. Reed, M. Mateas, and N. Wardrip-Fruin. Social story worlds with *comme il faut*. *IEEE Trans. Comput. Intellig. and AI in Games*, 6(2):97–112, 2014.
9. J. R. Meehan. Tale-spin, an interactive program that writes stories. In *Proc. of the Fifth International Joint Conference on Artificial Intelligence*, pages 91–98, 1977.
10. R. Pérez y Pérez. *MEXICA: A Computer Model of Creativity in Writing*. PhD thesis, The University of Sussex, 1999.
11. G. Polti. *The Thirty-Six Dramatic Situations*. 1917.
12. M. Sharples. *How We Write*. Routledge, 1999.
13. M. Si, S. C. Marsella, and D. V. Pynadath. Thespian: Modeling socially normative behavior in a decision-theoretic framework. In *Intelligent Virtual Agents*, volume 4133 of *Lecture Notes in Computer Science*, pages 369–382. Springer, 2006.
14. M. Theune, E. Faas, A. Nijholt, and D. Heylen. The virtual storyteller: Story creation by intelligent agents. In *Proceedings of the Technologies for Interactive Digital Storytelling and Entertainment (TIDSE) Conference*, pages 204–215, 2003.
15. J. W. von Goethe. *Elective Affinities / Kindred by Choice*. 1809.
16. L. A. Zadeh. A computational approach to fuzzy quantifiers in natural languages. *Computers & Mathematics with Applications*, 9(1):149–184, 1983.

Vicarious: A flexible framework for the creative use of sensed biodata

Paul Tennent, Joe Marshall, Brendan Walker, Paul Harter, and
Steve Benford

Mixed Reality Laboratory, Department of Computer Science, University of
Nottingham, Nottingham, UK, NG8 1BB
{paul.tennent, joe.marshall, brendan.walker, steve.benford}@nottingham.ac.uk
paul@cleverplugs.com

Abstract. In this paper we discuss *vicarious*, a flexible, extensible, distributed framework for capturing, processing, visualising, recording and generally *handling* sensed biodata. We outline six specific creative needs: heterogeneity of sensor inputs, liveness, high quality video, synchronisation, scalability, usability. Next, by outlining the architecture and features of *vicarious*, we show how the system meets each of those expectations. We then provide three examples of creative experiences developed using *vicarious*. The paper thus contributes the tool *vicarious* as support for creative experience design, as well as concrete examples of creative, artistic or performative applications of the system.

1 Introduction

In recent years there has been a proliferation of computing systems that make use of biosensing and biodata. Traditionally these data have primarily been used for study as part of the process of capturing and analysing a user's experience. More recently, we have seen the emergence of so called 'wellbeing' systems to the commercial market with peripherals such as fitbit¹ and Nike+² becoming a common sight in fitness circles. In academia we have seen a move towards systems which make active and creative use of so called biodata (or physiological data) either for control or to in some way adapt an experience. These take many forms: for example there are several examples of biodata-driven games: Brainball [8], Perping [16], and others e.g. [11, 14] to cover but a few. Other, less game-oriented but similarly playful systems such as themepark-style rides like the broncomatic [10] and breathless [1], or more artistically oriented experiences such as 'ere be dragons [4] or Donnarumma's muscle-movement based music generation [5] are amongst a host of biodata-themed experiences punctuating the HCI literature.

One common feature of these systems is that to construct them requires the use of one or more biofeedback sensor, and the application or development of

¹ <http://www.fitbit.com/>

² <https://secure-nikeplus.nike.com/plus/>

2 Tennent et al.

software to make use of the data from that sensor. Choosing appropriate sensors when designing an experience turns out to be something of a challenge. The range of available kit is extensive and diverse; in most cases one is confronted with a staggering array of choices. Should my sensors be able to transmit live or simply record the data to local memory? What format might that data be in: EDF(+), CSV, XML or some proprietary standard? What about the quality of the equipment and data? Do I need consumer or medical grade monitoring? The next challenge relates to how that equipment may be used in coordination. What if we want one sensor from one company and another from somewhere else? Later in the paper we will discuss this more, but this should serve to demonstrate the myriad possibilities for which we have to account when choosing equipment.

After several years of designing experiences which require a new system built essentially from the ground up each time, or the integration of several different systems, it was concluded it was necessary to construct a framework through which heterogeneous sensors might be integrated. This framework developed into the eponymous *vicarious*, and in the rest of the paper we will explore the system itself, then present a series of examples where we were able to successfully apply the system in creative or performative experiences.

2 Motivation

HCI and computer science in general has a history of creating general purpose data visualisation tools. Many of these focus on post-experimental data analysis e.g. [3,6,9,12] Virtually none of them however support any kind of live data display. Of course there are commercial systems which do exactly that; *Biotrace*³ is one such example, however these tools tend to be developed to work with a particular system: Mind Media's *nexus* family in the case of Biotrace. Similarly Affectiva's *Q Live* and CamNTEch's *Actiwave Suite*⁴ provide comprehensive analysis tools as long as one has recorded the data using their *Q* and *actiwave* products. There is nothing intrinsically wrong with this, except that when creating experiences that use a fusion of different sensors a designer loses the ability to perform data integration. A combination of proprietary data formats and closed-source software make expansion challenging at best and impossible at worst.

There are then three specific challenges relating to the sensor data: liveness, biodata-specific tools and heterogeneity of input sensors - and no currently available tools meet all three of these challenges.

Changing focus slightly, let us consider video. synchronisation of video and data is challenging. Indeed synchronisation of video and video is challenging, and this is reflected in the relative paucity of support for it in many of these tools, despite the incredibly rich context it can offer to the data. Even when video is included in such systems it is rarely of a quality suitable for broadcast. And yet as a creative medium, video is incredibly familiar and often serves a key role in

³ <http://www.humankarigar.com/biotrace.htm>

⁴ <http://www.camntech.com/>

the delivery of experiences involving biodata - as we will see in several of the examples later in this paper.

Next we have questions of scale. The vast majority of these tools are designed to handle a single, or at most a handful of participants for a short duration experiment. What do we do if we want to scale a system to handle biodata from a larger group or over a larger period of time? What if we wish to broadcast our experience to an audience of millions. In an era of multi-screen, cross media design, the local area network may represent a creative prison.

One difficulty creative designers often face when building experiences that make use of biodata is a simple lack of knowledge about what to do with the data and how to get it into a quantifiable state. Signal processing is something of an arcane art to those without relevant training - and as creative experience developers, that state is reasonable to assume for at least a significant portion of potential users. Most data wrangling systems provide a plethora of choice in how one might wish to process the data, but making those choices meaningful is often challenging.

All this has led us to develop a set of key challenges for building a biodata handling system that is flexible, useful and usable. These are as follows:

- *Heterogeneity of sensors and data sources*: Can a system be designed to handle inputs from a whole range of possible sensors, and indeed a wider corpus of data sources?
- *Liveness and pre-recorded data*: Can we design a system able to usefully display, or make use of in some other way, live data, as well as allow for the playback and review of data recorded either within our system or from some other source or collection of sources, such as non-connected sensors?
- *Handling of broadcast quality video*: Can a system be designed in such a way as to support very high quality video suitable for use in broadcast media and can the video and biodata be usefully and attractively combined?
- *Effective data synchronisation*: If we have several videos and several feeds of data can these be effectively synchronised to allow for accurate analysis, and account for different data rates, framerates and dropouts?
- *Scalability*: Can we design a system that is able to scale up to handle lots of data sources over extended periods of time?
- *Usability*: Is it possible to create a system that can be essentially picked up and used by creative designers who may know little about the complexities of biodata signals, and can those signals be visualised in such a way as to be understandable to lay persons when broadcast?

In the remainder of this paper we will describe the vicarious system with reference to these questions, before giving some examples of experiences or performances delivered using vicarious as a backbone.

3 Architecture

Vicarious has been developed in such a way as to allow it to be both robust and scalable. It is built on the premise that each part of the system should be an

4 Tennent et al.

autonomous, atomic unit and that those units should communicate in a system of publish and subscribe via network protocols. This supports robustness by allowing any single component to fail and be restarted, or be added, removed, modified or moved without interfering with the rest of the system and allows complex work to be distributed across several devices if necessary. From a users' point of view this is virtually transparent, but from a design point of view, as we will see, it allows for a very flexible system. Each program, henceforth referred to as a component is typically one of six types: *producers*, *translators*, *processors*, *emphconsumers*, *gateways* and *datastores*.

These components are used to construct chains whereby data ultimately gets from producer to consumer. Once data has passed into the system all the data is transformed to a consistent format and thus any component can (generally) subscribe to and consume data from any other. This allows for very large ranges of configurations from relatively small numbers of components.

Another significant benefit of this approach is platform and language independence. The vast majority of vicarious is written in python and thus roughly platform agnostic. However, Because each component is atomic and thus not dependant on any other, it can potentially be written in any language that supports network sockets. In practice this has been implemented on multiple occasions to communicate with e.g. MaxMSP in *The Experiment Live* (see section 4.1). This inbuilt architectural flexibility in particular allows for creatives to construct visualisations etc. using whichever platforms and tools they are comfortable with and still have access to vicarious' data capturing and processing resources.

3.1 Communications

At its heart, Vicarious is a publish and subscribe system where each component publishes zero or more streams of data and any other component can subscribe to and consume zero or more of those streams. Streams can be consumed by as many subscribers as desired, allowing the construction of some fairly complex chains. In most cases a component publishes only one stream, however if multiple streams are required a single port listens for numerical queries which define which stream is being requested, then set up and service that logical connection as required. Streaming is handled over TCP/IP and consists of a series of newline-separated strings. In practice, the vast majority of what is sent through the system is numerical data, but there are some cases where one might wish to send more complex data. By using strings as our base medium we create implicit support for json encoding - however, the onus is then on the consumer to expect the data in that format.

Treating all our data as network ready streams supports both robustness and scalability. Individual components can be run on any system on the same subnet, allowing process heavy components to have dedicated CPU time; the whole system can be very effectively multithreaded on one machine letting the CPU sort out priority in most cases; and it supports the graceful stopping and starting of data at any point in the system. If we wish to leave the local subnet, one of the *gateways* discussed in section 3.7 can be applied.

3.2 Producers

We use the term producer to refer to any data source. In most cases this will be a single sensor such as a heart rate monitor, or sensor hardware platform such as the Nexus. It might however be something which programmatically generates data such as our person simulator (below). Much of the time producers are third party hardware/software and as such the produced data are likely to be in any one of a myriad of often proprietary formats. This leads us the need for *translators*. Some producers are also instances of *gateways* see section 3.7.

The Person Simulator is a test harness, used to generate a facsimile of one person's physiological data, as if it were being sensed. It includes electrodermal activity, an electrocardiogram and a respiration trace. It can generate data well beyond normal human ranges, a feature which has been creatively co-opted as we shall see in the examples section. While EDA, ECG and Respiration are by no means a complete set of signals, they have proved a reasonable analogue for the types of sensors typically applied in such systems, and we have used equivalent third party software to do a similar job for electroencephalography.

3.3 Translators

A *translator* serves as a bridge between the data format and communication medium used by any one of a vast array of possible third party producers and that used internally by vicarious. It is the job of a translator to serve streams of data in a standardized, simple form, (usually floats, but occasionally json objects) to any other component which may the consume that stream or pass it further down the chain. For vicarious to succeed as intended, there must be an expandable pool of translators, and they must be easily creatable. Any time one wishes to make use of a sensor or system that lacks a translator one has to be able to write a new one, and so a template is provided and the complexity required is dictated by the data format and connection method of the producer. Translators have been created for several popular sensor systems including *nexus*, *Q* and *emotiv* to name but a few. The open source framework encourages the sharing of additional translators as they are created.

3.4 Processors

By *processor* we mean any component that subscribes to one or more streams and publishes one or more streams, usually performing some kind of signal processing on the incoming data. Processors will be of one of four types:

- Linear: one stream in, one stream out.
- Combinatorial: Multiple streams in, one stream out.
- Multiplexing: One stream in, multiple streams out.
- Multilinear: Multiple Streams in, multiple streams out.

6 Tennent et al.

The vast majority of familiar signal processors are linear: bandpass filters, smoothing, rate of change etc. Similarly more specific data oriented ones such as ECG to heart rate or RSP to breathing rate are also typically linear. The processor group is also the place for threshold counters, concordancers, and a myriad of other possible processing tasks.

As with the translators, there is an ever increasing set of processor modules and a very simple template for creating new ones. This means that with only a little programming skill it is straightforward to create new processors to make project-specific calculations, combine incoming streams, or any one of countless other operations without the need for learning new and arcane scripting languages. We hope we will be able to encourage the community to upload any created processors for the benefit of others, but we will moderate these for one key practical reason: usability. One of our key motivating factors in this development was that everything should be at least usable without field specific knowledge (medical or signal processing). Thus every processor in vicarious includes a clear description of what inputs it requires, what outputs it produces, what it does, and what it is intended for. A user browsing for appropriate processors or wishing to build a system to use a particular type of data may search the database of processors to find exactly what is required and understand exactly what it does rather than having to use trial and error to get the necessary results. Of course this is not an issue for users who are experts in signal processing, but the target user group for vicarious is technical creative designers, so that knowledge cannot be assumed.

3.5 Consumers

Consumers, in this context refer to any component which subscribes to a stream and does not publish. Most consumers in our example systems are ultimately contained within the *visualisation suite*, but it is possible to create standalone programs that consume data from vicarious. In many cases these will be *gateways* to some other system.

3.6 Datastores

Datastores are special cases of consumers which write incoming data streams into databases. They can then be used to replay those streams as live - taking the role of producers. Incoming data is timestamped to allow for playback synchronisation and post-hoc analysis.

3.7 Gateways

A *gateway* is a special case of either a *translator* or a *consumer*. They allow vicarious to pass data into other systems and formats. For example, there exists a gateway to the massive-scale messaging system pubnub⁵ which buffers up

⁵ <http://www.pubnub.com>

some amount of some collection of streams, encodes them as json and delivers them to a pubnub channel. This can then be either handled by some external resource, or an incoming pubnub-vicarious gateway can be used to consume the stream from another internet connected computer and feed it into another vicarious setup. This allows us to sense at a distance, and is particularly desirable for massively mobile systems. Similarly, there is a gateway that serves as a websocket-javascript API for doing web based visualisation of streams of sensed data, and another that delivers accessible streams into the Arduino sketch environment for more physical computing-based experiences.

3.8 Component Management

The distributed nature of the vicarious framework makes for an uncomfortable number of separate processes. Starting each of these manually in a console window or similar, is hardly ideal, so vicarious includes a component manager, that reads an XML specification for a particular system and starts and manages all the necessary processes from within one local application (per device). A drag and drop visual-programming approach to building these XML files is currently under development to further simplify this process.

3.9 Visualisation Suite

Thus far we have examined the *backbone* of the vicarious framework - looking at tools for collecting, processing and storing data, but it is also necessary to provide tools for interacting with that data. While the range of creative uses to which sensed biodata might be put are far without our scope as technical designers to produce, we have created a suite through which incoming data can be visualised and combined with video, and which is in itself widely flexible and extensible. The system provides line charts, bar charts (in various flavours), heatmaps, alpha-shapes, videos and a plethora of other basic visualisation tools, which can be combined to create more complex visualisations.

The nature of the system is such that creatives can of course make use of the data outside vicarious using either *consumers* in their own system or vicarious' *gateways*, but the *visualisation suite*, which is written in OpenGL and has been designed to allow new visualisation components to be quickly added. The system is based on the idea of *active textures*, each of which constitutes one or more consumers (either of video or of data), and which can be laid out with suitable alpha-blending on the screen to create quite complex visualisation systems. Each of the example systems shown below (all of which were creatively directed by artists or performers) make use of the visualisation suite as the principle delivery method for data.

4 Examples of successful applications

In this section we give three examples of experiences which have made use of vicarious to deliver creative and interesting uses of biodata. The first is a live

8 Tennent et al.

performance, the second is a marketing campaign and the last is a live game at a music festival run by an environmental NGO. The diversity of these experience scratches the surface of what vicarious is capable of supporting. In the discussion section we will outline some future directions.

4.1 The Experiment Live



Fig. 1. The experiment live as presented on a cinema screen. Six cameras with one selected for main display at any given time, four ‘ghost hunters’ performing a seance with biodata displayed for ECG, EDA and RSP.

The experiment live was a live event run on Halloween and delivered on the screen of a local arts cinema. It centred around group of four ‘ghost hunters’ who, with their biodata being displayed to all, were exploring the ‘haunted’ basement of a local tearoom, to a soundtrack of music generated from their biosignals. A story was told in which the data-recorders became inactive and were then ‘possessed’ (unbeknownst to the ghost hunters themselves) after being removed. The experience is more fully documented in [15]. In system terms, each ‘ghost hunter’ was equipped with a nexus measuring ECG, EDA and RSP, each of which were passed through a processor to generate heart rate, excitement, and breathing rate respectively, and combined in another processor to create a ‘fear factor’. The streams were consumed by the visualisation suite along with video data from a series of static and roaming cameras. Similarly, the music - generated in MaxMSP consumed the streams straight out of vicarious. All this was delivered live to the cinema, then when it was time for the ‘possession’ the nexus data streams were swapped for the *person simulator*.

4.2 Juke: Built to Thrill

Unlike the experiment live, *Juke: Built to Thrill* was not exclusively a live performance. Instead it was part of an advertising campaign for the Nissan Juke in coordination with the advertising company TBWA, with a series of events and short films made for web consumption. There were four events: a pre-recorded stunt film to introduce the campaign, that required the data to be 'staged' post-hoc then overlaid. A skydiving simulator running live at the Goodwood festival of speed which showed live 'thrill data' processed from heart rate and EDA, and an adventure trip in which four participants had their 'thrill' data recorded to create a montage of their experience including an on screen comparison of their thrill data.



Fig. 2. The visualisation of biodata in the Juke: Built to thrill campaign. Right: as rendered in vicarious, Left: as applied to the film in production

All three cases used the same visualisation (see figure 2) which was developed as a new active texture for the visualisation suite. In the case of the stunt film, this was driven with 'acted' biodata - a series of clips were made attempting to match a scripted 'bio-story' of the experience. These were then added into the final film by an external production team, using Adobe After Effects. At Goodwood, the data was captured live using a vilistus blood volume pulse sensor and an EDA sensor. This was processed and displayed live at the event using the visualisation suite, then these visualisations were sent to the same video production team and were again added to the final film in post. The adventure trip participants had their ECG and EDA data recorded using activwave cardios and Q sensors. The data was not monitored live, but downloaded each evening and processed, inspected and visualised for inclusion in the final video⁶.

4.3 Man vs Turtle

Man versus turtle was an interactive experience designed to help an environmental NGO called Medasset raise its public profile and create awareness regarding

⁶ The videos can be viewed at <https://www.youtube.com/playlist?list=PL2C59049F28D8B580>

10 Tennent et al.

its aims. It consisted of an interactive biodata driven game which then led to a reflective experience in the form of a short, but carefully staged interview and was developed to create conversation time between the players (general public) and the NGO representatives. Practically this was delivered live as a public two-player game in a marquee at the Plissken music festival in Athens. Each player of the game had their EDA captured using a nexus, this was then processed to rate of change and the differing values were used in the visualisation suite to drive an alpha-blended ‘tug-o-war’ between two videos: one of serene turtles, and another of a beach music festival.



Fig. 3. Players of the Man vs Turtle game at the Plissken music festival in Athens

The game itself was a spin on the classic relax to win mechanic established by brainball [8], wherein one player, representing the turtles needed to relax, and the other player, representing the humans, needed to get excited. This was supposed to be representative of the differing goals of turtles and humans and the relative success of each player served to drive the discussion after the game.

5 Discussion

Here we revisit the design challenges outlined in the motivation section with reference to the design of the system as discussed in the previous sections. In each case we outline how vicarious attempts to meet the challenge.

Heterogeneity of sensors and data sources: Vicarious uses an extensible system of translators and gateways to be able to talk to potentially anything that exposes a stream of data.

Liveness and pre-recorded data: The system allows for the monitoring, recording and playback of data and videos, and includes import and export for a variety of standard data formats such as CSV and EDF+.

Handling of broadcast quality video: As long as the hardware it is running on has the necessary support, e.g. Blackmagic *Decklinks* or similar, vicarious' visualisation suite can support multiple full HD video feeds and blend these with

complex visualisations. As we saw with Juke: Built to thrill - these visualisations can also be exported to be combined with video in post production using more familiar tools and workflows.

Effective data synchronisation: Vicarious' playback system uses absolute timestamps for every data point and video frame to ensure that everything remains perfectly in sync all the time when playing. This leads to generally excellent synchronisation, but is limited by the latency of the producer's communication medium as the timestamps are applied once the data enters the vicarious framework.

Scalability: The distributed nature of vicarious makes it scalable up to a point. Gateways to systems such as pubnub dramatically increase the scalability. Generally, vicarious has been designed with a broadcast view of the world, that is, data are generated by a small number of people and consumed by a larger audience. It remains to be seen whether the reverse would be feasible within this framework - and this stands as a remaining research question.

Usability: Vicarious as it stands is usable up to a point. Several areas of its interface - including a visual system generator for the backbone are currently under development. However, it still requires a significant degree of technical skill to operate and extend. The gateways mitigate this to some extent - delivering the data into familiar formats for visualisation, and the visualisation suite itself may be suitable for many applications, as demonstrated in the examples section, but for now, Vicarious remains something of a technical undertaking. Our local observations suggest that these techno-creative experiences are often composed by teams of people - with a designer/performer/artist at the helm and a supporting technical advisor or developer assisting them. For a team such as this then, vicarious is indeed a usable system, but perhaps not universally so.

6 Conclusion

Generally vicarious has achieved what we set out to do. We have created a system for ourselves and our artistic collaborators that allows us to rapidly develop and deploy interesting and creative experiences based on biodata. We are now looking beyond biodata as a source to other data types such as environmental sensors, and looking at other forms of delivery such as web based data explorers or second screens to live broadcasts. Much remains to be done, but vicarious may just prove to be an effective backbone with which to do it.

References

1. Steve Benford, Chris Greenhalgh, Gabriella Giannachi, Brendan Walker, Joe Marshall, and Tom Rodden. Uncomfortable interactions. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12*, page 2005, New York, New York, USA, May 2012. ACM Press.
2. H Brugman and A Russel. Annotating multi-media / multi-modal resources with ELAN. In *LREC2004*, pages 2065–2068, 2004.

12 Tennent et al.

3. Andy Crabtree, Steve Benford, Chris Greenhalgh, Paul Tennent, Matthew Chalmers, and Barry Brown. Supporting ethnographic studies of ubiquitous computing in the wild. In *Proceedings of the 6th ACM conference on Designing Interactive systems - DIS '06*, page 60, New York, New York, USA, June 2006. ACM Press.
4. Stephen Boyd Davis, Magnus Moar, Rachel Jacobs, Matt Watkins, Chris Riddoch, and Karl Cooke. ere be dragons: heartfelt gaming. *Digital Creativity*, 17(3):157–162, 2006.
5. Marco Donnarumma. Music for Flesh II: informing interactive music performance with the viscosity of the body system. In *New Interfaces for Musical Expression Conference (NIME)*, Ann Arbor, Michigan, USA, 2012.
6. Adam Fouse, Nadir Weibel, Edwin Hutchins, and James D. Hollan. ChronoViz: a system for supporting navigation of time-coded data. pages 299–299–304–304, May 2011.
7. Gabriella Giannachi, Duncan Rowland, Steve Benford, Jonathan Foster, Matt Adams, and Alan Chamberlain. Blast Theory's Rider Spoke , its Documentation and the Making of its Replay Archive. *Contemporary Theatre Review*, 20(3):353–367, August 2010.
8. Sara Ilstedt Hjelm and C Browall. Brainball - using brain activity for cool competition. In *Proc. NordiCHI*, Stockholm, Sweden, 2000.
9. Michael Kipp. Anvil - A Generic Annotation Tool for Multimodal Dialogue. In *Proceedings of the 7th European Conference on Speech Communication and Technology (Eurospeech)*, pages 1367–1370, Aalborg, September 2001.
10. Joe Marshall, Paul Harter, Jo Longhurst, Brendan Walker, Steve Benford, George Tomlinson, Stefan Rennick Egglestone, Stuart Reeves, Patrick Brundell, Paul Tennent, and Jo Cranwell. The gas mask. In *Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems - CHI EA '11*, page 127, New York, New York, USA, May 2011. ACM Press.
11. Soh Masuko and Junichi Hoshino. A fitness game reflecting heart rate. In *Proceedings of the 2006 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology*, ACE '06, New York, NY, USA, 2006. ACM.
12. A. Morrison, P. Tennent, and M. Chalmers. Coordinated Visualisation of Video and System Log Data. In *Fourth International Conference on Coordinated & Multiple Views in Exploratory Visualization (CMV'06)*, pages 91–102. IEEE, July 2006.
13. Greg Ross, Alistair Morrison, and Matthew Chalmers. Visualisation Techniques for Users and Designers of Layout Algorithms. In *IV '05: Proceedings of the Ninth International Conference on Information Visualisation (IV'05)*, pages 579–586, Washington, DC, USA, 2005. IEEE Computer Society.
14. Tadeusz Stach, T. C. Nicholas Graham, Jeffrey Yim, and Ryan E. Rhodes. Heart rate control of exercise video games. pages 125–132, May 2009.
15. Paul Tennent, Stuart Reeves, Steve Benford, Brendan Walker, Joe Marshall, Patrick Brundell, Rupert Meese, and Paul Harter. The machine in the ghost. In *Proceedings of the 2012 ACM annual conference extended abstracts on Human Factors in Computing Systems Extended Abstracts - CHI EA '12*, page 91, New York, New York, USA, May 2012. ACM Press.
16. Paul Tennent, Duncan Rowland, Joe Marshall, Stefan Rennick-Egglestone, A Harrison, Z Jaime, B Walker, and Steve Benford. Breathalising games: understanding the potential of breath control in game interfaces. In *Proceedings of the 8th International Conference on Advances in Computer Entertainment Technology*, page 58. ACM, 2011.

Rough k nearest neighbours for classification in the case of missing input data

Robert K. Nowicki¹, Bartosz A. Nowak^{1,2} and Marcin Woźniak³

¹ Institute of Computational Intelligence, Czestochowa University of Technology,
Al. Armii Krajowej 36, 42-200 Czestochowa, Poland

{robert.nowicki,bartosz.nowak}@iisi.pcz.pl
<http://www.iisi.pcz.pl>

² Department of Mathematical Methods in Computer Science, University
of Warmia and Mazury, ul. Słoneczna 54, 10-710 Olsztyn, Poland

³ Institute of Mathematics, Silesian University of Technology,
Kasubaska 23, 44-101 Gliwice, Poland
Marcin.Wozniak@polsl.pl

Abstract. Algorithm k -nn is often used for classification, but the majority of used distance meters are not designed to work with missing values. In the most appliances, this issue is solved using marginalization. Unfortunately, this approach can cause that even a large part of possessed data may be wasted. Also, after filling some of the missing values with valid data, the system may give completely different answer. Therefore in the paper a new algorithm is proposed, which in case of lacks in the sample analyses whole domain of possible values for corresponding attributes. Proposed system using k -nn algorithm gives rough-specific answer, which states if the test sample may or must belong to the certain set of classes. The system has important feature, that after filling some part of missing values it indicates the same set of possible classes or its subsets.

Keywords: rough sets, k -nn, missing values.

1 Introduction

The problem of incomplete input data is inherent in many real applications of decision-making systems. In the industry, some information could be unavailable due to e.g. measuring instrument failure or temporary exceeding the measure range in some part of monitoring process. In medical diagnosis procedures some tests are omitted because of the patient state, unacceptable cost, lack of reagents or rejected by the community because of beliefs. Moreover, it could be deemed unnecessary by a doctor. The decision support system cannot remain idle in such cases. Generally, regardless of applied methodology (neural networks, fuzzy systems, k -nn classifiers, svm systems etc.) there are two general methods to process data with missing values — marginalization and imputation — as well as its hybrids and modifications.

Methods that belong to the first group boil down to temporary reduce the dimensionality of considered space to the features of known values. Thus, some elements of

the system are just turned off. Moreover, sometimes the elimination of whole incomplete samples is also treated as the marginalization. However, it is eventually accepted only in developing time.

When we would like to use imputation, the unknown values are replaced by estimated ones. The palette of available methods is generally unlimited. The most primitive ones are confined to insertion of random, average or most common values. More sophisticated ones apply EM (Expectation Maximization) or k nearest neighbour algorithms, neural networks, and various fuzzy systems [5, 8, 21, 23]. Promising results are obtained by multiple imputation and interval imputation. If we know the probability density distribution, we can use the Bayesian solution [6, 9, 19, 20]. Then if we know the possibility distribution, we can use fuzzy imputation.

A specific approach to the problem comes from the rough set theory [17]. An object can be classified to a positive region of the class (i.e. the object certainly belongs to the class), to a negative region of the class (i.e. the object certainly do not belong to the class) or to a boundary region of the class (i.e. it is not possible to determine if the object belongs to the class or not). Membership to these regions depends on the quality of object description. If this description is good enough, the object belongs either to the positive or negative regions. If the description is too weak, then the object belongs to the boundary region. In the rough set theory [17] as well as in the theory of evidence [24], we do not use the individual elements of the consideration but some granules [18]. The granules contain elements which are indistinguishable basing on knowledge that we dispose. Thus, the size and the shape of granules depend on the used (known) knowledge about the elements. Hence, the many hybrid approaches apply rough sets together with other methods mentioned above.

In this paper, we focus on the hybrid system merging the rough set theory with one of the, probably, most popular classification method, i.e. k nearest neighbours [4]. The solutions based on this both ideas are already present in the literature. Let us give some examples. In [2] in order to improve the performance of k -nn classifier, the specific (tolerant) rough relation has been applied to build the similarity function. The quite popular idea is to apply rough set theory to reduction of dimensionality in nearest neighbour classifier. We can find various realizations of it in [10–12]. The rough sets and nearest neighbour algorithm are combined also with fuzzy sets (logic) and genetic (evolutionary) algorithms, see [13, 22, 26, 27] or [29].

The aim of solution presented in this paper is different from cited above. We propose to rebuild the k -nn classifier to be able to work with incomplete input data. However, we do not use either the imputation, or marginalization. In general, the idea could be described as follow. The new solution is built according to rough set theory [17]. When all input features have known and available values, the classifier works as elemental k -nn system. In the case, when some value is missing the all possible output of classifier are estimated. When only one decision is possible, the feature with missing value could be considered negligible and the classified sample is assigned to lower approximation (positive region) of classes indicated by the classifier. When more than one decision is possible, the classified sample is assigned to upper approximation of each class indicated by the classifier. Obviously, it is not possible to test all features with missing values. In presented solution we use the intervals and specific procedure shown below.

2 Proposed algorithm

The proposed algorithm can operate with interval-type value of attributes and missing values. In our approach lacks are replaced by interval $[v_i^{min}, v_i^{max}]$, where v_i^{min} , v_i^{max} are minimal and maximal values of the attribute v_i .

The first part of k -nn algorithm requires to compute distance from current sample $(x^{(t)})$ to each sample in the reference set. In our approach distance similar to Manhattan metric is used:

$$\bar{d}_s = \bar{d}(x^{(t)}, x_s) = \sum_{i=1 \dots n} \begin{cases} |\bar{v}_i^{(t)} - \bar{v}_{s,i}| & \text{if } v_{s,i} \in P \wedge v_{t,i} \in P \\ \max \left(|\bar{v}_i^{(t)} - v_i^{max}|, |\bar{v}_{t,i} - v_i^{min}| \right) & \text{if } v_{s,i} \in G \wedge v_{t,i} \in P \\ \max \left(|v_i^{max} - \bar{v}_{s,i}|, |v_i^{min} - \bar{v}_{s,i}| \right) & \text{if } v_{s,i} \in P \wedge v_{t,i} \in G \\ v_i^{max} - v_i^{min} & \text{if } v_{s,i} \in G \wedge v_{t,i} \in G, \end{cases} \quad (1)$$

$$\underline{d}_s = \underline{d}(x^{(t)}, x_s) = \sum_{\substack{i=1 \dots n \\ v_{s,i} \in P \wedge v_{t,i} \in P}} |\bar{v}_i^{(t)} - \bar{v}_{s,i}|, \quad (2)$$

where $\bar{v}_i^{(t)}$ is a value of i -th attribute in the test sample, $\bar{v}_{s,i}$ is a value of i -th attribute of s -th reference sample, P is set of attributes with available values, G is set of attributes with unavailable values, $[\underline{d}_s, \bar{d}_s]$ is interval-type distance of test sample $x^{(t)}$ to reference sample x_s .

Let the vector \mathbf{d}^{unique} contains all sorted and unique values of $\bar{d}_s, \underline{d}_s, s = 1 \dots M$, where M is a number of reference samples. Algorithm uses matrices $\bar{\Psi}^{m \times |d^{unique}|}$, $\underline{\Psi}^{m \times |d^{unique}|}$. Values $\bar{\Psi}_{j,c}, \underline{\Psi}_{j,c}$ define how many reference samples of the class ω_j are in the distance not greater than d_c^{unique} using \underline{d} for $\bar{\Psi}_{j,c}$ or \bar{d} for $\underline{\Psi}_{j,c}$, where

$$\bar{\Psi}_{j,c} = \sum_{\substack{s=1 \dots M: \\ \underline{d}_s \leq d_c^{unique} \\ \wedge x_s \in \omega_j}} 1, \quad (3)$$

$$\underline{\Psi}_{j,c} = \sum_{\substack{s=1 \dots M: \\ \bar{d}_s \leq d_c^{unique} \\ \wedge x_s \in \omega_j}} 1. \quad (4)$$

Let vectors $\bar{\psi}, \underline{\psi}$ contain decimal values. Values $\bar{\psi}, \underline{\psi}$ are equal to number of samples, which are in the distance not greater than d_c^{unique} using \underline{d} for $\bar{\psi}_{j,c}$ or \bar{d} for $\underline{\psi}_{j,c}$ we have:

$$\bar{\psi}_c = \sum_{\substack{s=1 \dots M: \\ \underline{d}_s \leq d_c^{unique}}} 1, \quad (5)$$

$$\underline{\psi}_c = \sum_{\substack{s=1 \dots M: \\ \bar{d}_s \leq d_c^{unique}}} 1. \quad (6)$$

4 Robert K. Nowicki, Bartosz A. Nowak, Marcin Woźniak

It is always true, that $\bar{\psi}_c \geq \underline{\psi}_c$. When all distances are singletons then $\bar{\psi}_c = \underline{\psi}_c$ and the system operates the same as k -nn.

For convenience of the users proposed method uses also symbols:

$$\bar{\Psi}_{\sim j,c} = \sum_{\substack{s=1\dots M: \\ d_s \leq d_c^{unique} \\ \wedge x_s \notin \omega_j}} 1 = \bar{\psi}_c - \bar{\Psi}_{j,c}, \quad (7)$$

$$\underline{\Psi}_{\sim j,c} = \sum_{\substack{s=1\dots M: \\ d_s \leq d_c^{unique}, \\ \wedge x_s \notin \omega_j}} 1 = \underline{\psi}_c - \underline{\Psi}_{j,c}. \quad (8)$$

Let c_{min} equals index of the first element in $\bar{\psi}$, which value is not lower than parameter k , and c_{max} equals index of the first element in $\underline{\psi}$ not lower than k . Because always $c_{min} \leq c_{max}$, we can write:

$$c_{min} = \arg \min_c (\bar{\psi}_c \geq k), \quad (9)$$

$$c_{max} = \arg \min_c (\underline{\psi}_c \geq k). \quad (10)$$

Proposed algorithm examines only distances $\{d_c^{unique} : c_{min} \leq c \leq c_{max}\}$. Because only for these distances the main idea of k -nn algorithm is fulfilled: exactly k -nn nearest reference samples (or more in case of a tie) are considered (in our case for classification).

Let $\Psi_{1,c}^{pot}, \Psi_{2,c}^{pot} \dots \Psi_{m,c}^{pot}$ equal some potential number of samples that belong to an appropriate class and are in the distance not greater than d_c^{unique} . Authors assumed that for each distance $\{d_c^{unique} : c_{min} \leq c \leq c_{max}\}$ all possible combinations of the values $\{\Psi_{j,c}^{pot} : 1 \leq j \leq m\}$ meet k -nn conditions if $\forall j = 1 \dots m, \underline{\Psi}_{j,c} \leq \Psi_{j,c}^{pot} \leq \bar{\Psi}_{j,c}$ and $\sum_{j=1\dots m} \Psi_{j,c}^{pot} \geq k$.

The test sample $(x^{(t)})$ may belong to an upper approximation of some class $(x^{(t)} \in P^* \omega_j)$ only if for any possible combination of $\{\Psi_{j,c}^{pot} : 1 \leq j \leq m\}$ the most of sample elements belong to that class (ω_j):

$$x^{(t)} \in P^* \omega_j \Leftrightarrow \left(\begin{array}{l} \exists \Psi_{1,c}^{pot}, \Psi_{2,c}^{pot} \dots \Psi_{m,c}^{pot} : \\ (\underline{\Psi}_{1,c} \leq \Psi_{1,c}^{pot} \leq \bar{\Psi}_{1,c}) \wedge \\ (\underline{\Psi}_{2,c} \leq \Psi_{2,c}^{pot} \leq \bar{\Psi}_{2,c}) \wedge \\ \dots \\ (\underline{\Psi}_{m,c} \leq \Psi_{m,c}^{pot} \leq \bar{\Psi}_{m,c}) \wedge \\ (\sum_{j'=1\dots m} \Psi_{j',c}^{pot} \geq k) \end{array} \right) \left(\begin{array}{l} \forall j' : 1 \leq j' \leq m \\ \wedge j' \neq j \end{array} \right) (\Psi_{j,c}^{pot} > \Psi_{j',c}^{pot}). \quad (11)$$

Similarly, test sample $(x^{(t)})$ must belong to a lower approximation of some class $(x^{(t)} \in P_* \omega_j)$ only if for all possible combination of $\{\Psi_{j,c}^{pot} : 1 \leq j \leq m\}$ the most of

sample elements belong to that class (ω_j),

$$x^{(t)} \in P_* \omega_j \Leftrightarrow \left(\begin{array}{l} \forall \Psi_{1,c}^{pot}, \Psi_{2,c}^{pot} \dots \Psi_{m,c}^{pot} : \\ (\underline{\Psi}_{1,c} \leq \Psi_{1,c}^{pot} \leq \overline{\Psi}_{1,c}) \wedge \\ (\underline{\Psi}_{2,c} \leq \Psi_{2,c}^{pot} \leq \overline{\Psi}_{2,c}) \wedge \\ \dots \\ (\underline{\Psi}_{m,c} \leq \Psi_{m,c}^{pot} \leq \overline{\Psi}_{m,c}) \wedge \\ (\sum_{j'=1 \dots m} \Psi_{j',c}^{pot} \geq k) \end{array} \right) \quad (12)$$

$$\left(\begin{array}{l} \forall j' : 1 \leq j' \leq m \\ \wedge j' \neq j \end{array} \right) \left(\Psi_{j,c}^{pot} > \Psi_{j',c}^{pot} \right).$$

In practice checking these two formulas — (11), (12) — may be difficult and time consuming. Therefore in the paper, the following classification rules are proposed.

To state that **the sample belongs to upper approximation of class** ω_j the following two steps should be processed using the all distances $c : c_{min} \leq c \leq c_{max}$:

- If for some c , $(\overline{\Psi}_{j,c}) + (\underline{\Psi}_{\sim j,c}) \geq k$ then this sample belongs to upper approximation of ω_j , if for all other classes $\omega_{j'}$, $\underline{\Psi}_{j',c} < \overline{\Psi}_{j,c}$.
- If for some c , $(\overline{\Psi}_{j,c}) + (\underline{\Psi}_{\sim j,c}) < k$, then additional conditions must be checked. The first one is the same as before: for all other classes $\omega_{j'}$, $\underline{\Psi}_{j',c} < \overline{\Psi}_{j,c}$. Because the main condition of k -nn is not necessarily fulfilled ($\overline{\Psi}_{j,c} + \underline{\Psi}_{\sim j,c} < k$), then more samples (than $\underline{\Psi}_{j',c}$) must be chosen from other classes until k samples will be selected, but number of chosen samples from other class cannot exceed $\overline{\Psi}_{j',c}$ and must be lower than $\overline{\Psi}_{j,c}$.

It may be presented in following, easy to implement formula:

$$x^{(t)} \in P^* \omega_j \Leftrightarrow \left(\begin{array}{l} (\exists c : c_{min} \leq c \leq c_{max}) \\ \left(\forall_{j': 1 \leq j' \leq m} \left(\overline{\Psi}_{j,c} > \underline{\Psi}_{j',c} \right) \right) \wedge \\ \left((\overline{\Psi}_{j,c} + \underline{\Psi}_{\sim j,c} < k) \Rightarrow \sum_{j': 1 \leq j' \leq m} (\min\{\overline{\Psi}_{j,c} - 1, \overline{\Psi}_{j',c}\}) \geq k - \overline{\Psi}_{j,c} \right) \end{array} \right). \quad (13)$$

To state that **the sample belongs to lower approximation of class** ω_j , the following two steps should be processed, taking into account each distance $c : c_{min} \leq c \leq c_{max}$:

- When $\underline{\Psi}_{j,c} + \overline{\Psi}_{\sim j,c} \geq k$ is fulfilled, and if $\underline{\Psi}_{j,c}$ is greater than $\overline{\Psi}_{j',c}$ for all other classes.
- When $\underline{\Psi}_{j,c} + \overline{\Psi}_{\sim j,c} < k$ the number of chosen samples for current class must be increased to $k - \overline{\Psi}_{\sim j,c}$. After that operation algorithm checks if the new number of chosen samples for current class is greater than $\overline{\Psi}_{j',c}$ for all other classes.

6 Robert K. Nowicki, Bartosz A. Nowak, Marcin Woźniak

It may be presented in following, easy to implement formula:

$$\begin{aligned}
 x^{(t)} \in P_* \omega_j \Leftrightarrow & \\
 (\forall c : c_{min} \leq c \leq c_{max}) & \\
 \left((\underline{\Psi}_{j,c} + \overline{\Psi}_{\sim j,c} \geq k) \Rightarrow \left(\bigwedge_{j': 1 \leq j' \leq m, j' \neq j} (\underline{\Psi}_{j,c} > \overline{\Psi}_{j',c}) \right) \right) \wedge & \quad (14) \\
 \left((\underline{\Psi}_{j,c} + \overline{\Psi}_{\sim j,c} < k) \Rightarrow \left(\bigwedge_{j': 1 \leq j' \leq m, j' \neq j} (k - \overline{\Psi}_{\sim j,c} > \overline{\Psi}_{j',c}) \right) \right), &
 \end{aligned}$$

that after simplification it is equal to

$$\begin{aligned}
 x^{(t)} \in P_* \omega_j \Leftrightarrow & (\forall c : c_{min} \leq c \leq c_{max}) \\
 \left(\bigwedge_{j': 1 \leq j' \leq m, j' \neq j} (\max\{\underline{\Psi}_{j,c}, k - \overline{\Psi}_{\sim j,c}\} > \overline{\Psi}_{j',c}) \right). & \quad (15)
 \end{aligned}$$

2.1 Performed simulations

Proposed algorithm was tested using databases (Tab. 1) from well-known UCI Repository [3]. All the research results were concluded according to 10-fold cross validation [1]. The main parameter of k -nn algorithm was set empirically. Authors used Manhattan distance instead of Euclidean, because it gave better results.

Databases In the paper following datasets were used:

- **Glass Identification (GI)** – samples were prepared by B. German from Home Office Forensic Science Service in Aldermaston. This database was composed for construction of the system to classify the type of glass for forensic purpose. The database contains 214 samples described by 9 attributes, which are non-negative real numbers. The number of classes was changed from original 7 to 2. The first one contains 163 samples and concerns window-type of glass, while the second one describes other type of glass.
- **Ionosphere** – samples were gathered by Space Physics Group from Johns Hopkins University. This database may be used for classification type in ionosphere analyzing radar [25]. It contains 351 samples, each described by 34 real numbers. The sample belongs to one of two classes of 225 or 126 items.
- **Iris** – data was prepared by R.A. Fisher [7]. This database is probably the most popular for benchmarking of classification algorithms. It contains 150 samples, each one is described by 4 attributes of real numbers. They represent characteristics of one of three types (classes) of iris flowers, containing 50 items each.
- **Wine Recognition (WR)** – database contains results of chemical analysis of 178 samples of wine, which can be divided into three classes. Every sample is described by 13 attributes.
- **Breast Cancer Wisconsin (BCW) Dataset** – samples were prepared by William H. Wolberg from University of Wisconsin Hospitals. The samples may be used to build classifier which recognizes type of breast cancer (malignant or not) [28]. The

Rough k nearest neighbours for classification in the case of missing input data 7

database contains 699 samples described by 9 attributes, which are positive decimals. Each sample belongs to one of two classes, containing 458 and 241 samples. Value of one of the attributes is missing for 16 samples.

Table 1. Properties of used databases from UCI repository

database name	#samples (M)	#attributes (n)	#classes (m)
Glass Identification (GI)	214	9	2
Ionosphere	351	34	2
Iris	150	4	3
Wine Recognition (WR)	178	13	3
Breast Cancer Wisconsin (BCW)	699	9	2

Evaluation measures Proposed algorithm indicates membership of sample to upper and lower approximations, what requires more complex evaluation model that used in case typical classification. The system was tested using 4 different measures:

1. The first criterion counts how often test sample belongs to the lower approximation of an appropriate class:

$$CCR_* = \frac{1}{M^{(t)}} \sum_{s=1}^{M^{(t)}} \sum_{j=1}^m \begin{cases} 1 & \text{if } x_s^{(t)} \in P_*\omega_j \wedge \omega(x_s^{(t)}) = \omega_j \\ 0 & \text{otherwise} \end{cases}, \quad (16)$$

where $M^{(t)}$ is a number of test samples, $x_s^{(t)}$ is s -th test sample, $x_s^{(t)} \in P_*\omega_j$ denotes that sample $x_s^{(t)}$ has been classified to the lower approximation of class ω_j and $\omega(x_s^{(t)}) = \omega_j$ — that the proper class of sample $x_s^{(t)}$ is class ω_j .

2. This criterion computes how often test sample belongs to the lower approximation, but improper class:

$$ICR_* = \frac{1}{M^{(t)}} \sum_{s=1}^{M^{(t)}} \sum_{j=1}^m \begin{cases} 1 & \text{if } x_s^{(t)} \in P_*\omega_j \wedge \omega(x_s^{(t)}) \neq \omega_j \\ 0 & \text{otherwise} \end{cases}, \quad (17)$$

where $\omega(x_s^{(t)}) \neq \omega_j$ denotes that the proper class of sample $x_s^{(t)}$ is other than class ω_j .

3. The third measure defines how often test sample belongs to upper approximation of appropriate class:

$$CCR^* = \frac{1}{M^{(t)}} \sum_{s=1}^{M^{(t)}} \sum_{j=1}^m \begin{cases} 1 & \text{if } x_s^{(t)} \in P^*\omega_j \wedge \omega(x_s^{(t)}) = \omega_j \\ 0 & \text{otherwise} \end{cases}, \quad (18)$$

8 Robert K. Nowicki, Bartosz A. Nowak, Marcin Woźniak

where $x_s^{(t)} \in P^*\omega_j$ denotes that sample $x_s^{(t)}$ has been classified to the upper approximation of class ω_j

4. The last criterion measures how often test sample belongs to the upper approximation of improper class:

$$ICR^* = \frac{1}{(m-1) \cdot M^{(t)}} \sum_{s=1}^{M^{(t)}} \sum_{j=1}^m \begin{cases} 1 & \text{if } x_s^{(t)} \in P^*\omega_j \wedge \omega(x_s^{(t)}) \neq \omega_j \\ 0 & \text{otherwise} \end{cases} \quad (19)$$

It is worth to note that in the denominator of criterion (19) occurs part $(m-1)$. This is due to the sample may belongs to upper approximations of more than one class.

Details of evaluations Proposed algorithm was tested using mentioned databases and measures. The k parameter was different for each dataset and was empirically defined: 11 for GI, 9 for Ionosphere, 5 for Iris, 11 for WR and 13 for BCW. For evaluation purposes, values of specified attributes in test samples were removed (considered as missing). The algorithm was tested in two different scenarios: in the first one (Tables: 2, 5, 7, 9) each attribute was erased separately, and in the second scenario (Tables: 3, 4, 6, 8, 10) set of "removed" attributes was growing until the algorithm gave completely useless answer ($CCR_* = 0, ICR^* = 100\%$). In case of no missing values, proposed

Table 2. Results of proposed algorithm for GI database and 8 available attributes

list of available attributes	results [%]			
	CCR_*	ICR_*	CCR^*	ICR^*
$v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_8$	65.8	0.5	99.5	32.7
$v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_9$	20.6	0.0	100.0	78.1
$v_1, v_2, v_3, v_4, v_5, v_6, v_8, v_9$	58.8	0.0	100.0	40.8
$v_1, v_2, v_3, v_4, v_5, v_7, v_8, v_9$	30.8	0.0	100.0	68.2
$v_1, v_2, v_3, v_4, v_6, v_7, v_8, v_9$	76.2	0.5	99.5	23.4
$v_1, v_2, v_3, v_5, v_6, v_7, v_8, v_9$	62.7	0.0	100.0	35.5
$v_1, v_2, v_4, v_5, v_6, v_7, v_8, v_9$	4.7	0.0	100.0	94.4
$v_1, v_3, v_4, v_5, v_6, v_7, v_8, v_9$	71.4	0.0	100.0	27.1
$v_2, v_3, v_4, v_5, v_6, v_7, v_8, v_9$	64.9	0.5	99.5	34.7

algorithm returns the same answer as typical k -nn method, because $c_{min} = c_{max}$ and identical samples used for voting as in standard k -nn algorithm. Usually, the level of ICR_* was very low in cost of low CCR_* , conversely CCR^* was very high in cost of high ICR^* . In the first scenario, when values of single attribute were removed, there were significant differences (Tables: 2, 7, 9 and particularly 5) in efficiency between various cases.

Rough k nearest neighbours for classification in the case of missing input data 9

Table 3. Results of proposed algorithm for GI database and chosen set of available attributes

list of available attributes	results [%]			
	CCR_*	ICR_*	CCR^*	ICR^*
$v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_8, v_9$	92.0	8.0	92.0	8.0
$v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_9$	20.6	0.0	100.0	78.1
$v_1, v_3, v_4, v_5, v_6, v_7, v_9$	0.0	0.0	100.0	100.0

Table 4. Results of proposed algorithm for Ionosphere database and chosen set of available attributes

list of removed attributes	results [%]			
	CCR_*	ICR_*	CCR^*	ICR^*
-	86.9	13.1	86.9	13.1
v_{29}	78.7	7.7	92.0	21.3
v_{18}, v_{29}	70.4	4.5	95.5	28.5
v_{10}, v_{18}, v_{29}	47.1	0.8	99.2	49.2
$v_{10}, v_{18}, v_{29}, v_{33}$	22.8	0.0	100.0	76.6
$v_8, v_{10}, v_{18}, v_{29}, v_{33}$	13.7	0.0	100.0	86.3
$v_8, v_{10}, v_{15}, v_{18}, v_{29}, v_{33}$	8.8	0.0	100.0	90.3
$v_7, v_8, v_{10}, v_{15}, v_{18}, v_{29}, v_{33}$	2.0	0.0	100.0	98.0
$v_7, v_8, v_{10}, v_{15}, v_{18}, v_{27}, v_{29}, v_{33}$	0.0	0.0	100.0	100.0

3 Conclusions

Proposed algorithm may be considered as generalization of k -nn method. The improvement over k -nn allows to operate with interval-type of attribute value and as consequent interval-type of the distance. This can be also applied in the case of missing values. Then the range of input features variation should be known. The algorithm processes the input interval and calculates the lower and upper distances between samples. Thus, it can take into account all possible but unknown input states. It is inspired directly by Pawlak rough sets theory.

The experimental studies confirm the analytical consideration. The proposed classifier, alike as the rough-neuro-fuzzy classifiers presented in [14–16] does not give incorrect answers as the result of missing values. If the classification is proper for complete input vector, it is also proper in case of incomplete data. If the available information is insufficient to make certain decision, the proposed classifier indicates the additional classes corresponding to possible values of feature with missing values. It expressed in this way, the uncertainty in the decision.

Acknowledgments The project was funded by the National Science Centre under decision number DEC-2012/05/B/ST6/03620.

10 Robert K. Nowicki, Bartosz A. Nowak, Marcin Woźniak

Table 5. Results of proposed algorithm for Iris database and 3 available attributes

list of available attributes	results [%]			
	CCR_*	ICR_*	CCR^*	ICR^*
v_1, v_2, v_3	0.0	0.0	100.0	61.0
v_1, v_2, v_4	0.0	0.0	100.0	56.3
v_1, v_3, v_4	26.7	0.0	100.0	40.0
v_2, v_3, v_4	26.7	0.0	100.0	36.3

Table 6. Results of proposed algorithm for Iris database and chosen set of available attributes

list of available attributes	results [%]			
	CCR_*	ICR_*	CCR^*	ICR^*
v_1, v_2, v_3, v_4	94.7	5.3	94.7	2.7
v_2, v_3, v_4	26.7	0.0	100.0	36.3
v_3, v_4	0.0	0.0	100.0	90.0
v_3	0.0	0.0	100.0	100.0

References

1. S. Arlot and A. Celisse. A survey of cross-validation procedures for model selection. *Statistics Surveys*, 4:40–79, 2010.
2. Y. Bao, X. Du, and N. Ishii. Improving performance of the k-nearest neighbor classifier by tolerant rough sets. In *Proc. of the Third International Symposium on Cooperative Database Systems for Advanced Applications*, pages 167–171, 2001.
3. collective work. Uci machine learning repository. <http://archive.ics.uci.edu/ml/datasets.html>.
4. T. Cover and P. Hart. Nearest neighbor pattern classification. *IEEE Trans. on Information Theory*, 13(1):21–27, January 1967.
5. K. Cpałka and L. Rutkowski. Flexible takagi-sugeno fuzzy systems. In *Proc. IEEE International Joint Conference on Neural Networks (IJCNN)*, volume 3, pages 1764–1769, 2005.
6. R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. A Wiley–Interscience Publication, John Wiley & Sons, INC., New York/Chichester/Weinheim/Brisbane/Singapore/Toronto, 2001.
7. R. Fisher. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7:179–188, 1936.
8. M. Gabryel, M. Korytkowski, R. Scherer, and L. Rutkowski. Object detection by simple fuzzy classifiers generated by boosting. In L. Rutkowski, M. Korytkowski, R. Scherer, R. Tadeusiewicz, L. Zadeh, and J. Zurada, editors, *Lecture Notes in Computer Science*, volume 7894, pages 540–547. Springer Berlin Heidelberg, 2013.
9. W. Greblicki and L. Rutkowski. Density-free Bayes risk consistency of nonparametric pattern recognition procedures. *Proceedings of the IEEE*, 69(4):482–483, April 1981.
10. M. He and Y.-P. Du. Research on attribute reduction using rough neighborhood model. In *Proc. International Seminar on Business and Information Management (ISBIM) 2008*, volume 1, pages 268–270, Dec 2008.
11. N. Ishii, I. Torii, Y. Bao, and H. Tanaka. Modified reduct: Nearest neighbor classification. In *Proc. IEEE/ACIS 11th International Conference on Computer and Information Science (ICIS)*, pages 310–315, May 2012.

Rough k nearest neighbours for classification in the case of missing input data

11

Table 7. Results of proposed algorithm for WR database and 12 available attributes

list of available attributes	results [%]			
	CCR_*	ICR_*	CCR^*	ICR^*
$v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_8, v_9, v_{10}, v_{11}, v_{12}$	42.7	0.0	100.0	30.3
$v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_8, v_9, v_{10}, v_{11}, v_{13}$	54.4	0.0	100.0	24.8
$v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_8, v_9, v_{10}, v_{12}, v_{13}$	58.3	0.0	99.4	19.4
$v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_8, v_9, v_{11}, v_{12}, v_{13}$	50.0	0.0	100.0	25.8
$v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_8, v_{10}, v_{11}, v_{12}, v_{13}$	60.6	0.0	100.0	18.8
$v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_9, v_{10}, v_{11}, v_{12}, v_{13}$	65.1	0.0	100.0	16.9
$v_1, v_2, v_3, v_4, v_5, v_6, v_8, v_9, v_{10}, v_{11}, v_{12}, v_{13}$	56.6	0.0	100.0	22.2
$v_1, v_2, v_3, v_4, v_5, v_7, v_8, v_9, v_{10}, v_{11}, v_{12}, v_{13}$	64.1	0.0	100.0	17.7
$v_1, v_2, v_3, v_4, v_6, v_7, v_8, v_9, v_{10}, v_{11}, v_{12}, v_{13}$	55.1	0.0	100.0	21.4
$v_1, v_2, v_3, v_5, v_6, v_7, v_8, v_9, v_{10}, v_{11}, v_{12}, v_{13}$	66.3	0.0	100.0	15.2
$v_1, v_2, v_4, v_5, v_6, v_7, v_8, v_9, v_{10}, v_{11}, v_{12}, v_{13}$	68.0	0.0	100.0	15.7
$v_1, v_3, v_4, v_5, v_6, v_7, v_8, v_9, v_{10}, v_{11}, v_{12}, v_{13}$	52.8	0.0	100.0	23.3
$v_2, v_3, v_4, v_5, v_6, v_7, v_8, v_9, v_{10}, v_{11}, v_{12}, v_{13}$	49.5	0.0	100.0	24.4

Table 8. Results of proposed algorithm for WR database and chosen set of available attributes

list of available attributes	results [%]			
	CCR_*	ICR_*	CCR^*	ICR^*
$v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_8, v_9, v_{10}, v_{11}, v_{12}, v_{13}$	97.7	2.3	97.7	1.1
$v_1, v_2, v_3, v_5, v_6, v_7, v_8, v_9, v_{10}, v_{11}, v_{12}, v_{13}$	66.3	0.0	100.0	15.2
$v_1, v_3, v_5, v_6, v_7, v_8, v_9, v_{10}, v_{11}, v_{12}, v_{13}$	6.1	0.0	100.0	61.7
$v_1, v_3, v_5, v_6, v_7, v_9, v_{10}, v_{11}, v_{12}, v_{13}$	0.0	0.0	100.0	96.9
$v_1, v_3, v_6, v_7, v_9, v_{10}, v_{11}, v_{12}, v_{13}$	0.0	0.0	100.0	100.0

12. N. Ishii, I. Torii, Y. Bao, and H. Tanaka. Mapping of nearest neighbor for classification. In *Proc. IEEE/ACIS 12th International Conference on Computer and Information Science (ICIS)*, pages 121–126, June 2013.
13. J. Keller, M. Gray, and J. Givens. A fuzzy k-nearest neighbor algorithm. *IEEE Trans. on Systems, Man and Cybernetics*, SMC-15(4):580–585, July 1985.
14. R. Nowicki. On combining neuro-fuzzy architectures with the rough set theory to solve classification problems with incomplete data. *IEEE Trans. on Knowledge and Data Engineering*, 20(9):1239–1253, Sept. 2008.
15. R. Nowicki. Rough-neuro-fuzzy structures for classification with missing data. *IEEE Trans. on Systems, Man, and Cybernetics—Part B: Cybernetics*, 39(6):1334–1347, 2009.
16. R. Nowicki. On classification with missing data using rough-neuro-fuzzy systems. *International Journal of Applied Mathematics and Computer Science*, 20(1):55–67, 2010.
17. Z. Pawlak. *Rough Sets: Theoretical Aspects of Reasoning About Data*. Kluwer, Dordrecht, 1991.
18. W. Pedrycz and A. Bargiela. Granular clustering: A granular signature of data. *IEEE Trans. on Systems, Man, and Cybernetics—Part B: Cybernetics*, 32(2):212–224, Apr. 2002.
19. L. Rutkowski. On Bayes risk consistent pattern recognition procedures in a quasi-stationary environment. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI-4(1):84–87, Jan 1982.

12 Robert K. Nowicki, Bartosz A. Nowak, Marcin Woźniak

Table 9. Results of proposed algorithm for BCW Dataset and 8 available attributes

list of available attributes	results [%]			
	CCR_*	ICR_*	CCR^*	ICR^*
$v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_8$	75.5	0.1	99.9	23.3
$v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_9$	72.4	0.0	99.9	25.3
$v_1, v_2, v_3, v_4, v_5, v_6, v_8, v_9$	80.8	0.4	99.4	18.5
$v_1, v_2, v_3, v_4, v_5, v_7, v_8, v_9$	68.8	0.3	99.6	30.5
$v_1, v_2, v_3, v_4, v_6, v_7, v_8, v_9$	80.5	0.1	99.9	18.8
$v_1, v_2, v_3, v_5, v_6, v_7, v_8, v_9$	75.7	0.3	99.7	23.8
$v_1, v_2, v_4, v_5, v_6, v_7, v_8, v_9$	78.5	0.1	99.7	20.2
$v_1, v_3, v_4, v_5, v_6, v_7, v_8, v_9$	76.4	0.1	99.9	22.0
$v_2, v_3, v_4, v_5, v_6, v_7, v_8, v_9$	78.5	0.6	99.4	20.3

Table 10. Results of proposed algorithm for BCW Dataset and chosen sets of available attributes

list of available attributes	results [%]			
	CCR_*	ICR_*	CCR^*	ICR^*
$v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_8, v_9$	95.6	2.9	97.0	4.0
$v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_9$	72.4	0.0	99.9	25.3
$v_1, v_2, v_3, v_4, v_5, v_6, v_7$	5.3	0.0	100.0	94.4
$v_1, v_3, v_4, v_5, v_6, v_7$	0.0	0.0	100.0	100.0

20. L. Rutkowski. Adaptive probabilistic neural networks for pattern classification in time-varying environment. *IEEE Trans. on Neural Networks*, 15(4):811–827, July 2004.
21. L. Rutkowski and K. Cpałka. Compromise approach to neuro-fuzzy systems. In P. Sincak, J. Vascak, V. Kvasnicka, and J. Pospichal, editors, *Intelligent Technologies - Theory and Applications*, volume 76, pages 85–90. IOS Press, 2002.
22. M. Sarkar. Fuzzy-rough nearest neighbors algorithm. In *Proc. IEEE International Conference on Systems, Man, and Cybernetics*, volume 5, pages 3556–3561 vol.5, 2000.
23. R. Scherer. Neuro-fuzzy systems with relation matrix. In *Lecture Notes in Artificial Intelligence*, volume 6113, pages 210–216. 2010.
24. G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, New York, 1976.
25. V. Sigillito, S. Wing, L. Hutton, and K. Baker. Classification of radar returns from the ionosphere using neural networks. *Johns Hopkins APL Technical Digest*, pages 262–266, 1989.
26. N. Verbiest, C. Cornelis, and R. Jensen. Fuzzy rough positive region based nearest neighbour classification. In *Proc. IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pages 1–7, June 2012.
27. T. Villmann, F. Schleif, and B. Hammer. Fuzzy labeled soft nearest neighbor classification with relevance learning. In *Proc. Fourth International Conference on Machine Learning and Applications*, pages 11–15, Dec 2005.
28. W. Wolberg and O. Mangasarian. Multisurface Method of Pattern Separation for Medical Diagnosis Applied to Breast Cytology. *Proc. of the National Academy of Sciences, U.S.A.*, 87:9193–9196, 1990.
29. R. Yager. Using fuzzy methods to model nearest neighbor rules. *IEEE Trans. on Systems, Man, and Cybernetics, Part B: Cybernetics*, 32(4):512–525, Aug 2002.

A novel approach to position traffic in NoSQL database systems by the use of firefly algorithm

Marcin Woźniak¹, Marcin Gabryel², Robert K. Nowicki² and Bartosz A. Nowak^{2,3}.

¹ Institute of Mathematics, Silesian University of Technology,
Kaszubska 23, 44-101 Gliwice, Poland
`Marcin.Wozniak@polsl.pl`

² Institute of Computational Intelligence, Czestochowa University of Technology,
Al. Armii Krajowej 36, 42-200 Czestochowa, Poland
`robert.nowicki,bartosz.nowak@iisi.pcz.pl`

³ Department of Mathematical Methods in Computer Science, University
of Warmia and Mazury, ul. Słoneczna 54, 10-710 Olsztyn, Poland

Abstract. In this paper, application of computational intelligence methods in positioning and optimization of traffic in NoSQL database system modeled with exponentially distributed service and vacation is discussed. Positioning of the system modeled with independent 2-order hyper exponential input stream of packets and exponential service time distribution is solved using firefly algorithm. Different scenarios of examined system operation are presented and analyzed.

Keywords: firefly algorithm, queueing system, NoSQL database positioning

1 Introduction

In modern computer systems more and more data is collected. The servers operate on large data sets. Service must be efficient and fast enough to work with many clients at the time. In NoSQL systems we use efficient and dedicated solutions to increase performance, please see [16]. Dedicated sorting algorithms help to organize large data sets and response to clients as fast as possible, please see examples in [28] and [29]. However it is not easy to position these systems for processing large amounts of data and response to many clients in the service. Extended research on this situations are presented in [25] and [26], where efficient methods of indexing and sorting in large NoSQL systems are discussed. Here will be discussed another important aspect of optimal service in large NoSQL systems.

Traffic in the network and therefore efficient organization of the service are other aspects that increase efficiency and quality of service (QoS). Let us now think of the network, where NoSQL database server is serving various clients. In common service many clients send requests at the time. The server collects them and proceeds actions. After processing knowledge in database server responses to

2 M. Woźniak, M. Gabryel, R. K. Nowicki and B. A. Nowak

the requests, but this goes according the income queue. Earlier requests shall be answered first and so on. The problem is to position this system for most efficient operation. One shall define optimal service, vacation and income parameters. In this paper NoSQL database system is positioned for most efficient service and lowest possible cost of work by the use of evolutionary computation (EC) methods.

For this type of object, various methods of modeling and simulation can be applied. One of them is analysis of provided stochastic characteristics. These are to describe operation of: IP routers, audio or video cards with buffer of packets waiting for displaying and various types of database servers with queue of requests from remote clients. In each of these objects, operation model is defined for applied queueing system (QS), see [23].

Let us take situation in network with NoSQL database server. Various clients may want to find information and send requests to the server. Server finds the information and sends it to the clients but according to arrivals. The server answers, however there is limited number of requests to be served at the time. Therefore, it gathers all the requests and put them in a queue. According to the order in this queue all the requests are serviced. Let us say, that the queue has defined capacity. The service in computer network is done with database, where dedicated QS can be applied to optimize operation cost and increase QoS. The situation can be modeled for $T_{service}$, T_{income} and $T_{vacation}$, which describe average time of service for each request, average time of incoming requests and average time in which the system takes break (for backup, conservation and etc.), respectively. All inter-arrival times, service times and server vacations are totally independent random variables, where the symbols used in QS description in time t are:

- τ_1 — the first busy period of the system (starting at $t = 0$);
- δ_1 — the first idle time of the system (consisting of the first vacation time v_1 and the first server standby time q_1);
- $h(\tau_1)$ — the number of packets completely served during τ_1 ;
- $X(t)$ — the number of packets present in the system at time t .

In this paper is discussed simulation and positioning of NoSQL database traffic modeled with dedicated QS, where at the time there is only one request arrival or response departure. This situation can be modeled with dedicated QS, which helps to position server operation, increases QoS and minimize costs of work. However this is non-trivial problem and many interesting papers are devoted to different positioned object. Classical cost structure is considered in [20]. While in [7] and [13] are presented most important aspects of positioning and cost optimization using QS. Various queueing models for applied server types are investigated in [2], [3], [18] and [17]. Please see also [19] and [21] for a review of important results on modeling and positioning.

In this paper are applied results of the research on similar QS. In [8] and [10] is discussed representation for joint transform of first busy period, first idle time and number of packets completely served during first busy period in $GI/G/1$ -type system with batch arrivals and exponential single server vacation. Further

characteristics of the system with single vacations, Poisson arrivals, generally distributed service times and infinite buffers can be found in [9] and [11]. In [12] non-stationary behavior of waiting time distribution in a finite-buffer queue with single server vacations is investigated. All these research results can help to model and position traffic in NoSQL system types discussed in [28], [16] and [15]. Let us see the model of QS for NoSQL database traffic.

2 Queueing model

Similar attempt to model and position QS by dividing observation horizon on relatively short time intervals in which arrival process is recurrent with different distributions in different interval times is presented in [1], [22] and [24] or [23]. To model NoSQL server operation was used a finite-buffer $H_2/M/1/N$ -type QS, which best model situation in this type of server traffic, for more details please see [4] and [14]. Here is presented only a brief description, just to help in understanding NoSQL positioning and simulation problem using Definition ??, a full analytical model is presented in details in [23].

In modeled service are used inter-arrival times of 2-order, hyper exponentially distributed random variables with distribution function:

$$F(t) = p_1(1 - e^{-\lambda_1 t}) + p_2(1 - e^{-\lambda_2 t}), \quad t > 0, \quad (1)$$

where $\lambda_i > 0$ for $i = 1, 2$ and $p_1, p_2 \geq 0$. Distribution of inter-arrival times is mixture of two exponential distributions with parameters λ_1 and λ_2 , which are being "chosen" with probabilities p_1 and p_2 respectively. In the system, there are $(N - 1)$ places in queue and one for packet in the service. System starts working at $t = 0$ with at least one packet present, see schematic situation in Fig. ?. After busy period the server begins compulsory single vacation in the service which is modeled with 2-order hyper exponentially distributed random variable with distribution function

$$V(t) = q_1(1 - e^{-\alpha_1 t}) + q_2(1 - e^{-\alpha_2 t}), \quad t > 0. \quad (2)$$

Interpretation of parameters α_i , $i = 1, 2$ and q_1, q_2 is similar to that for λ_i , $i = 1, 2$ and p_1 and p_2 . If at the end of vacation there is no packet present in the system, the server is on standby and waits for first arrival to start service. If there is at least one packet waiting for service in the buffer at the end of vacation, the service process starts immediately and new busy period begins.

2.1 Cost optimization problem

Functions $F(\cdot)$ and $V(\cdot)$ are implemented to investigate some various aspects of operation of the examined NoSQL system, where inter-arrival times and vacation period have distribution functions defined in (1) and (2), respectively. In the research find optimal set of parameters λ_i , p_i , μ and α_i is found. The set defines possible service situations, for which optimal cost of work is found. Therefore $r_n(c_1)$ is defined to describe minimal amount of resources to perform all operations.

4 M. Woźniak, M. Gabryel, R. K. Nowicki and B. A. Nowak

Definition 1. *Optimal cost of work in modeled QS is minimal amount of money, energy or any other resources that the system may need to perform all operations.*

Define in Definition 1 optimal cost of work has a form of equation to use it in optimization process, which is

$$r_n(c_1) = \frac{Q_n(c_1)}{\mathbf{E}_n(c_1)} = \frac{r(\tau_1)\mathbf{E}_n\tau_1 + r(\delta_1)\mathbf{E}_n\delta_1}{\mathbf{E}_n\tau_1 + \mathbf{E}_n\delta_1}, \quad (3)$$

where the notations are: $r(\tau_1)$ —fixed unit operation costs during busy period τ_1 , $r(\delta_1)$ —fixed unit operation costs during idle time δ_1 , $\mathbf{E}_n\tau_1$ —mean of busy period τ_1 and $\mathbf{E}_n\delta_1$ —mean of idle time δ_1 on condition that system starts with n packets present. The explicit formula with detailed information and description for conditional joint characteristic functions of τ_1 , δ_1 and $h(\tau_1)$ is presented in [23] (see also [22] and [24]). Here let us briefly discuss modeling of applied QS. General equation to calculate this values is

$$\begin{aligned} B_n(s, \varrho, z) &= \mathbf{E}\{e^{-s\tau_1 - \varrho\delta_1} z^{h(\tau_1)} \mid X(0) = n\} = \\ &= \frac{D(s, \varrho, z) - G(s, \varrho, z)}{H(s, z)} R_{n-1}(s, z) + \sum_{k=2}^n R_{n-k}(s, z) \Psi_k(s, \varrho, z), \quad 2 \leq n \leq N. \end{aligned} \quad (4)$$

where $s \geq 0$, $\varrho \geq 0$ and $|z| \leq 1$, $n \geq 1$. Moreover $a_n(s, z)$, $\Psi_n(s, \varrho, z)$, $D(s, \varrho, z)$, $G(s, \varrho, z)$ and $H(s, z)$ are defined in (5), (6), (9), (10) and (11) respectively.

$$a_n(s, z) = \int_0^\infty \frac{(z\mu t)^n}{n!} e^{-(\mu+s)t} dF(t), \quad n \geq 0, \quad (5)$$

$$\begin{aligned} \Psi_n(s, \varrho, z) &= -\frac{(z\mu)^n}{(n-1)!} \left[\int_0^\infty dF(t) \int_0^t x^{n-1} e^{-(\mu+s)x} \right. \\ &\quad \times \left. \left(e^{-\varrho(t-x)} V(t-x) + \int_{t-x}^\infty e^{-\varrho y} dV(y) \right) dx \right]. \end{aligned} \quad (6)$$

Moreover, sequence $(a_n(s, z))$ in (5) helps to recursively define

$$\begin{aligned} R_0(s, z) &= 0, \quad R_1(s, z) = a_0^{-1}(s, z), \\ R_{n+1}(s, z) &= R_1(s, z)(R_n(s, z) - \sum_{k=0}^n a_{k+1}(s, z) R_{n-k}(s, z)). \end{aligned} \quad (7)$$

With introduced following function

$$f(s) = \int_0^\infty e^{-st} dF(t), \quad s > 0 \quad (8)$$

we finally have components of $r_n(c_1)$ operation costs defined in (3):

$$D(s, \varrho, z) = \sum_{k=1}^{N-1} a_k(s, z) \sum_{i=2}^{N-k+1} R_{N-k+1-i}(s, z) \Psi_i(s, \varrho, z), \quad (9)$$

$$G(s, \varrho, z) = \Psi_N(s, \varrho, z) + (1 - f(\mu + s)) \sum_{k=2}^N R_{N-k}(s, z) \Psi_k(s, \varrho, z) \quad (10)$$

and

$$H(s, z) = (1 - f(\mu + s)) R_{N-1}(s, z) - \sum_{k=1}^{N-1} a_k(s, z) R_{N-k}(s, z). \quad (11)$$

Indeed, since:

$$\mathbf{E}_n e^{-s\tau_1} = \mathbf{E}\{e^{-s\tau_1} \mid X(0) = n\} = B_n(s, 0, 1), \quad (12)$$

then

$$\mathbf{E}_n \tau_1 = -\frac{\partial}{\partial s} B_n(s, 0, 1) \Big|_{s=0} \quad (13)$$

and

$$\mathbf{E}_n \delta_1 = -\frac{\partial}{\partial \varrho} B_n(0, \varrho, 1) \Big|_{\varrho=0}. \quad (14)$$

The QS model presented above was solved using Wolfram Mathematica 9.0 to prepare analytical form of total cost equation in (3). As You see QS has complicated mathematical model, which is not easy to solve even using analytical methods. Therefore application of EC method to position presented NoSQL system for best QoS at lowest cost is discussed.

3 Dedicated simplified firefly algorithm

Firefly algorithm (FA) is mapping behavior of flying insects that we all can see in the summer. FA maps searching for best partner in optimization process. Simplified method uses some characteristics of fireflies to search for optimum in examined object's solution space. As the research show, FA is efficient in various applications. In [27] some FA is applied to help in 2D images processing. In [5] FA is applied to minimum cross entropy threshold selection, while in [6] is described it's efficiency in image compression. It is also efficient in continuous optimization (see [30]) or multi-modal optimization (see [31]). Let us now see dedicated mathematical model.

Firefly of a given species can be described by several biological traits:

- Specific way of flashing,
- Specific way of moving,

6 M. Woźniak, M. Gabryel, R. K. Nowicki and B. A. Nowak

- Specific perception of other individuals.

These features are modeled to map them in specific numerical values. In this way we translate natural characteristics in mathematical model used to develop EC algorithm. Thus, in implementation of simplified firefly algorithm (SFA) are:

- γ –light absorption factor in given circumstances,
- μ –factor for random motion of each individual,
- β –factor for attractiveness of firefly species.

These features allow to implement behavior of different species or conditions of natural environment. Let us now see description of innate qualities:

- All fireflies are unisex, therefore one individual can be attracted to any other regardless of gender.
- Attractiveness is proportional to brightness. Thus, for every two fireflies less clear flashing one will move toward brighter one.
- Attractiveness is proportional to brightness and decreases with increasing distance between individuals.
- If there is no clearer and more visible firefly within the range, then each one will move randomly.

Movement of individuals is conditioned distance to other individuals surrounding it. Firefly will go to most attractive one measuring intensity of flicker. Thus, using mathematical assumptions, individuals will be characterized by suitable metric. Therefore we can model natural identification of individuals and their attraction to each other depending on the distance separating them. In nature, fireflies that are closer, not only see better, but also seem to be more attractive to the others. Using these features calculations remap natural behavior.

Distance between any two fireflies i and j situated at points \mathbf{x}_i and \mathbf{x}_j in object's solutions space can be defined as Cartesian metric

$$r_{ij}^t = \|\mathbf{x}_i^t - \mathbf{x}_j^t\| = \sqrt{\sum_{k=1}^n (x_{i,k}^t - x_{j,k}^t)^2}, \quad (15)$$

where the notations in t iteration are: \mathbf{x}_i^t ; \mathbf{x}_j^t –points in R^n space (n is dimension of the object's fitness function), $x_{i,k}^t$; $x_{j,k}^t$ – k -th components of the spatial coordinates \mathbf{x}_i and \mathbf{x}_j that describe each firefly position in the space.

Attractiveness of firefly i to firefly j decreases with increasing distance. It is proportional to intensity of light seen by surrounding individuals defined as

$$\beta_{ij}(r_{ij}^t) = \beta \cdot e^{-\gamma \cdot (r_{ij}^t)^2}, \quad (16)$$

where the notations in t iteration are: $\beta_{ij}(r_{ij}^t)$ –attractiveness of firefly i to firefly j , r_{ij}^t –distance between firefly i and firefly j defined in (15), γ –light absorption factor mapping natural conditions, β –attractiveness factor.

Finally, firefly i motions toward more attractive (clearer flashing) individual j using information about other individuals in the population denotes formula

$$\mathbf{x}_i^{t+1} = \mathbf{x}_i^t + (\mathbf{x}_j^t - \mathbf{x}_i^t) \cdot \beta_{ij}(r_{ij}^t) + \mu \cdot e_i, \quad (17)$$

where the notations in t iteration are: \mathbf{x}_i^t ; \mathbf{x}_j^t —points in R^n space (n is dimension of the object's model), $\beta_{ij}(r_{ij}^t)$ —attractiveness of firefly i to firefly j modeled in (16), r_{ij}^t —distance between fireflies i and j modeled in (15), γ —light absorption factor mapping natural conditions, μ —factor mapping natural random motion of individuals, e_i —randomized vector changing position of fireflies. Implementation of dedicated FA is presented in algorithm 1.

Start,

Define all coefficients: γ —light absorption factor, β —attractiveness factor, μ —natural random motion, number of *fireflies* and *generation*—number of iterations in the algorithm,

Define fitness function for the algorithm: $r_n(c_1)$ in (3),

Create at random initial population P of *fireflies* in QS solution space, $t = 0$,

while $t \leq \text{generation}$ **do**

 Calculate distance between individuals in population P using (15),

 Calculate attractiveness for individuals in population P using (16),

 Evaluate individuals in population P using (3),

 Create population O : move individuals towards closest and most attractive individual using (17),

 Evaluate individuals in population O using (3),

 Replace δ worst individuals from P with δ best individuals from O and the rest of individuals take at random,

 Next generation $t = t + 1$,

end

Values from population P with best fitness are solution,

Stop.

Algorithm 1: Dedicated FA to optimize defined QS of NoSQL database

4 Optimal NoSQL database system positioning

Research results help to position the system for lowest response time and optimize service cost considered in variants: under-load, critical load and overload. SFA simulations were performed for $r(\tau_1) = 0.5$ and $r(\delta_1) = 0.5$. It means that according to Definition 1, modeled system service and vacation uses 0.5 energy unit each, which costs differ depending on the local energy billing rate. These values may be changed, what makes presented model flexible and easily applicable to any similar system. Presented results are averaged values of 100 samplings, 80 fireflies in 20 generations with $\beta = 0.3$, $\gamma = 0.3$, $\mu = 0.25$, $\delta = 25\%$, where:

8 M. Woźniak, M. Gabryel, R. K. Nowicki and B. A. Nowak

- Average service time: $T_{service} = \frac{1}{\mu}$,
- Average time between packages income into the system: $T_{income} = \frac{p_1}{\lambda_1} + \frac{p_2}{\lambda_2}$,
- Average vacation time: $T_{vacation} = \frac{q_1}{\alpha_1} + \frac{q_2}{\alpha_2}$,
- Examined system size: $N = \text{buffer size} + 1$.

Scenario 1.

SFA was performed to find set of parameters for lowest cost of work and best QoS. In Table 1 are optimum values that affect NoSQL system operation.

Table 1. Parameters μ , λ_i , α_i , p_i , q_i for $i = 1, 2$ and lowest value of (3).

λ_1	λ_2	α_1	α_2	p_1	p_2	q_1	q_2
2.80	2.60	1.20	0.44	1.51	1.0	5.90	4.90
μ	0.51	$r_n(c_1)$		0.32			
$T_{service}$		T_{income}		$T_{vacation}$			
[sec]	1.96	0.92		16.05			

SFA was also arranged to position NoSQL in various scenarios. Hence system parameters must be set in peculiar way to optimize operation cost. In each scenario is discussed possible situation of traffic.

Scenario 2.

NoSQL $T_{service} = 2[sec]$, what means that request service takes about $2[sec]$. This is similar to situation in small on-line shops or customer services. Research results are shown in Table 2.

Table 2. Parameters μ , λ_i , α_i , p_i , q_i for $i = 1, 2$ and lowest value of (3).

λ_1	λ_2	α_1	α_2	p_1	p_2	q_1	q_2
1.93	3.21	0.81	0.9	81	1	2.11	10.20
μ	0.5	$r_n(c_1)$			0.36		
$T_{service}$		T_{income}		$T_{vacation}$			
[sec]	2.00	42.28		13.94			

Scenario 3.

NoSQL $T_{service} = 0.5[sec]$. This situation represents NoSQL business service with heavy traffic and efficient server machine. Research results with system positioning are shown in Table 3.

Using SFA we can also position NoSQL system for given T_{income} . This will correspond to peculiar incoming situations.

Scenario 4.

NoSQL T_{income} was given as $2[sec]$, what means that requests are incoming to the server once in 2 seconds, similarly to scenario 2. Research results are shown in Table 4.

A novel approach to position traffic in NoSQL database systems ...

9

Table 3. Parameters μ , λ_i , α_i , p_i , q_i for $i = 1, 2$ and lowest value of (3).

λ_1	λ_2	α_1	α_2	p_1	p_2	q_1	q_2
45.12	24.17	128.01	1.30	1.70	1.41	67.16	15.65
μ	2.00	$r_n(c_1)$		0.28			
$T_{service}$		T_{income}		$T_{vacation}$			
[sec]	0.5	0.10		12.56			

Table 4. Parameters μ , λ_i , α_i , p_i , q_i for $i = 1, 2$ and lowest value of (3).

λ_1	λ_2	α_1	α_2	p_1	p_2	q_1	q_2
3.83	4.01	0.7	1.24	6.57	2.02	11.01	6.45
μ	0.3	$r_n(c_1)$		0.39			
$T_{service}$		T_{income}		$T_{vacation}$			
[sec]	3.33	2.03		20.93			

Scenario 5.

NoSQL T_{income} was given as 0.5[sec], what means that requests are incoming to the server twice in every second. This situation is describing an extensively used database, like these of business purpose. Research results are shown in Table 5.

Table 5. Parameters μ , λ_i , α_i , p_i , q_i for $i = 1, 2$ and lowest value of (3).

λ_1	λ_2	α_1	α_2	p_1	p_2	q_1	q_2
28.13	26.56	0.92	0.54	14.13	1.40	6.20	4.90
μ	0.92	$r_n(c_1)$		0.30			
$T_{service}$		T_{income}		$T_{vacation}$			
[sec]	1.09	0.50		15.38			

5 Conclusions

EC methods are useful in simulation or positioning of various types of objects. They help to collect representative samples, which can be used by AI decision support systems. CI helps to simulate complicated objects and because of free design, calculations are possible even in discontinuous spaces. Conducted experiments confirm efficiency in simulation of examined object in many possible scenarios representing common situations in reality.

Positioned model was simulated in situations with predefined time of service or income. It reflects situations when traffic is common or heavy and system must serve many requests. Further work should be carried out to reduce time consuming operations, tentatively by using some knowledge prior to generate

10 M. Woźniak, M. Gabryel, R. K. Nowicki and B. A. Nowak

initial population. Moreover, modeled system could be non-stationary and parameters could change during work due to e.g. wear (expenditure) of elements (battery, voltage, etc.) or environment changes (temperature, air composition etc.). So, in future research it is important to take into account this aspects by e.g. fuzzyfication of the system parameters.

Acknowledgments The project was funded by the National Science Centre under decision number DEC-2012/05/B/ST6/03620.

References

1. M. Gabryel, R. K. Nowicki, M. Woźniak, and W. M. Kempa. Genetic cost optimization of the $GI/M/1/N$ finite-buffer queue with a single vacation policy. *Lecture Notes in Artificial Intelligence - ICAISC'2013*, (7895):12–23, 2013.
2. U. C. Gupta, A. D. Banik, and S. Pathak. Complete analysis of $MAP/G/1/N$ queue with single (multiple) vacation(s) under limited service discipline. *Journal of Applied Mathematics and Stochastic Analysis*, (3):353–373, 2005.
3. U. C. Gupta and K. Sikdar. Computing queue length distributions in $MAP/G/1/N$ queue under single and multiple vacation. *Applied Mathematics and Computation*, 2(174):1498–1525, 2006.
4. D. Hongwei, Z. Dongfeng, and Z. Yifan. Performance analysis of wireless sensor networks of serial transmission mode with vacation on fire prevention. *ICCET'10 IEEE CPS*, pages 153–155, 2010.
5. M.H. Horng. Multilevel minimum cross entropy threshold selection based on the firefly algorithm. *Expert Systems with Applications*, 38(12):14805–14811, 2011.
6. M.H. Horng. Vector quantization using the firefly algorithm for image compression. *Expert Systems with Applications*, 39(1):1078–1091, 2012.
7. O. Kella. Optimal control of the vacation scheme in an $M/G/1$ queue. *Operations Research Journal*, 4(38):724–728, 1990.
8. W. M. Kempa. $GI/G/1/\infty$ batch arrival queuing system with a single exponential vacation. *Mathematical Methods of Operations Research*, 1(69):81–97, 2009.
9. W. M. Kempa. Some new results for departure process in the $M^X/G/1$ queuing system with a single vacation and exhaustive service. *Stochastic Analysis and Applications*, 1(28):26–43, 2009.
10. W. M. Kempa. Characteristics of vacation cycle in the batch arrival queuing system with single vacations and exhaustive service. *International Journal of Applied Mathematics*, 4(23):747–758, 2010.
11. W. M. Kempa. On departure process in the batch arrival queue with single vacation and setup time. *Annales UMCS Informatica*, 1(10):93–102, 2010.
12. W. M. Kempa. The virtual waiting time in a finite-buffer queue with a single vacation policy. *Lecture Notes in Computer Science*, (7314):47–60, 2012.
13. R.E. Lillo. Optimal operating policy for an $M/G/1$ exhaustive server-vacation model. *Methodology and Computing in Applied Probability*, 2(2):153–167, 2000.
14. V. Mancuso and S. Alouf. Analysis of power saving with continuous connectivity. *Computer Networks*, 56(10):2481–2493, 2012.
15. Z. Marszałek, D. Połap, and M. Woźniak. On preprocessing large data sets by the use of triple merge sort algorithm. In *Proceedings of International Conference on Advances in Information Processing and Communication Technologies*, pages 65–72, Santa Barbara, California, USA, 2014. The IRED, Seek Digital Library.

16. Z. Marszałek and M. Woźniak. On possible organizing nosql database systems. *International Journal of Information Science and Intelligent System*, 2(2):51–59, 2013.
17. Z. Niu, T. Shu, and Y. Takahashi. A vacation queue with setup and close-down times and batch markovian arrival processes. *Performance Evaluation Journal*, 3(54):225–248, 2003.
18. Z. Niu and Y. Takahashi. A finite-capacity queue with exhaustive vacation/close-down/setup times and markovian arrival processes. *Queueing Systems*, 1(31):1–23, 1999.
19. H. Takagi. *Queueing Analysis, vol. 1: Vacation and Priority Systems, vol. 2. Finite Systems*. North-Holland, Amsterdam, 1993.
20. J. Teghem. Control of the service process in a queueing system. *European Journal of Operations Research*, 1(23):141–158, 1986.
21. N. Tian and Z. G. Zhang. *Vacation queueing models. Theory and applications*. Springer - Verlag, Berlin, Heidelberg, 2006.
22. M. Woźniak. On applying cuckoo search algorithm to positioning GI/M/1/N finite-buffer queue with a single vacation policy. In *Mechican International Conference on Artificial Intelligence - MICAI'2013 Proceedings*, pages 59–64. IEEE, 2013.
23. M. Woźniak, W. M. Kempa, M. Gabryel, and R. K. Nowicki. A finite-buffer queue with single vacation policy - analytical study with evolutionary positioning. *International Journal of Applied Mathematics and Computer Science*, 24(4):accepted–in press, 2014.
24. M. Woźniak, W. M. Kempa, M. Gabryel, R. K. Nowicki, and Z. Shao. On applying evolutionary computation methods to optimization of vacation cycle costs in finite-buffer queue. *Lecture Notes in Artificial Intelligence - ICAISC'2014*, (8467):480–491, 2014.
25. M. Woźniak and Z. Marszałek. *Selected Algorithms for Sorting Large Data Sets*. Silesian University of Technology Press, Gliwice, Poland, 2013.
26. M. Woźniak and Z. Marszałek. *Extended Algorithms for Sorting Large Data Sets*. Silesian University of Technology Press, Gliwice, Poland, 2014.
27. M. Woźniak and Z. Marszałek. An idea to apply firefly algorithm in 2D image key-points search. *Communications in Computer and Information Science - ICIST'2014*, 465:312–323, 2014.
28. M. Woźniak, Z. Marszałek, Gabryel M., and Nowicki R. K. On quick sort algorithm performance for large data sets. In A.M.J. Skulimowski, editor, *Looking into the Future of Creativity and Decision Support Systems - KICSS'2013*, pages 647–656. Progress & Business Publishers, Cracow, Poland, 2013.
29. M. Woźniak, Z. Marszałek, Gabryel M., and Nowicki R. K. Triple heap sort algorithm for large data sets. In A.M.J. Skulimowski, editor, *Looking into the Future of Creativity and Decision Support Systems - KICSS'2013*, pages 657–665. Progress & Business Publishers, Cracow, Poland, 2013.
30. X.S. Yang, Z.H. Cui, R.B. Xiao, A.H. Gandomi, and M. Karamanoglu. *Swarm Intelligence and Bio-inspired Computation: Theory and Applications*. Elsevier, Waltham, USA, 2013.
31. X.S. Yang and S. Deb. Cuckoo search via lévy flights. In *NaBIC'2009 Proceedings*, pages 210–214, 2009.

An Assessment on the Length of Hospital Stay through Artificial Neural Networks

Vasco Abelha¹, Henrique Vicente², José Machado³, and José Neves^{3,*}

¹ Departamento de Informática, Universidade do Minho, Braga, Portugal
vascoabelha91@gmail.com

² Departamento de Química, Centro de Química de Évora, Escola de Ciências e Tecnologia,
Universidade de Évora, Évora, Portugal

hvicente@uevora.pt
³ CCTC/Departamento de Informática, Universidade do Minho, Braga, Portugal
{jmac, jneves}@di.uminho.pt

* Corresponding author: phone: +351-934201337; fax: +351-253604471;
e-mail: jneves@di.uminho.pt

Abstract. The attitude of stashing costs and preventing people of any kind of intervention on the valuation of the problem referred to above, may undermine their belief on present society values and on their way of living. Cutting funds from Education to Health is at best delaying the inevitable *crash* that is foreshadowed. Indeed, regarding people, a major concern may be described as jeopardizing their health condition, i.e., providing healthcare is a very sensitive issue and prunes to drastic changes in short spaces of time. Factors like age, sex, and context – house conditions, daily lives – should also be central when deciding how long a specific patient should remain in a hospital. In no way, ought this be decided by economic circumstances alone. To fulfill this goal, a Logic Programming based approach is used for knowledge representation and reasoning, letting the modeling of the universe of discourse in terms of defective data, information and knowledge. Artificial Neural Networks are enforced as the computational framework, allowing one to predict how long a patient should remain in a hospital.

Keywords: Healthcare, Length of Hospital Stay, Logic Programming, Knowledge Representation and Reasoning, Artificial Neuronal Networks.

1 Introduction

On the one hand, and in light with recent events, our world is going through troubled social and economic times. In order to succeed and live in this era, our leaders are cutting any kind of outgoing expenses, which means reducing costs, namely in education and healthcare. In terms of health, this can be described as cutting out staff members, closing healthcare facilities, or even minimizing a patient's length of hospital stay. Such an attitude may end up in the worsening of a patient health condition or, in extreme cases, may lead to death. On the other hand, the hospital should be regarded

as a place where patients are treated to their illnesses and express and validate their health states. Indeed, a patient length of hospital stay must be decided against some features, namely among those that are mentioned below.

- Age and Sex;
- Comorbidity;
- Context;
- Hospital Records;
- Treatments;
- Patient History; and
- Exams.

Therefore, the variables of the universe of discourse are organized into two clusters, namely:

- Patient Source; and
- Hospital Source.

Patient Source will include age, sex, history, context and comorbidity, i.e., the subjective information that is retrieved by a dialog between the individual and the physician, whose accuracy must be judged and evaluated. Indeed, it can be compromised by factors such as memory, unawareness or fraud, just to name a few. The patient can easily forget, depending on his/her condition, details or even events that may be of great help to the physician. On top of that, they can also be lying and trying to deceive the clinic due to regular factors, i.e., its *context*.

One has a reliable and objective Source of Information, which is provided by the hospital. It includes vital signs and measurements, such as weight; findings from physical examination; results from laboratory and other diagnostics already completed, and Hospital Diaries. The two groups of information, Subjective – Patient – and Objective – Hospital – were inspired by a method of documentation employed by the healthcare arena – *S.O.A.P. (Subjective, Objective, Analyses and Planning)*. The aspects previously enumerated, are more than necessary, as presented in a wide-range of literature or among experts, to complete a risk assessment [1,2,3], i.e., to determine, with a certain Degree of Confidence, if a patient should stay in health care or not. The most prominent aspects can be linked to Patient History, Comorbidity and Context. To shed some light, the latter refers to the patient daily lives. If they have no conditions to be “*set out in the open*” – home hygienic conditions, no guidance - they ought to stay in the hospital, because, if not, they will most likely deteriorate their health and be routed once again to the hospital in a near future and even increase their costs of stay.

It is better to do it right the first time and not having the past come back to bite. Nevertheless and in spite of these politics, this is exactly what happens most of the times. Patients tend to decrease their already short time span of admission in the hospital. The rest of aspects are easy to infer their relevance and why they should as well to be taken into consideration. Such variables as *current treatments, exams and daily*

hospital diaries have a correlation between them. Most of the time they share information or complement each other.

In order to explain the basics that are described below, namely the aspects needed to determine the length of stay in a hospital, which were clustered in 4 (four) groups displayed in Fig. 1. All the data relative to the patient such as age, sex, and comorbidity are linked to the Biological Factors. Patient History refers to the detailed patient medical past. The information here comes from the patient himself, or any past records that the health system has recorded regarding his previous admissions. Ambient Factors are simply about the context embracing the patient. Anything that surrounds him/her is significant to determine the hospital length of stay. Hospital Factors simply summarizes every exam, diagnosis he/she did at the healthcare facility. Indeed, after a briefly description of the relevance and meaning of the values and moments under inspection, their drawbacks lay on the incomplete and default information, noise and uncertainty to which they are attached. Our analysis aims to speculate and estimate the necessary length of stay in a hospital of a given patient, while reducing any mishap in their health. This is achieved by the use of a Logical Programming based approach to Knowledge Representation and Reasoning, complemented with a computational framework based on Artificial Neural Networks.

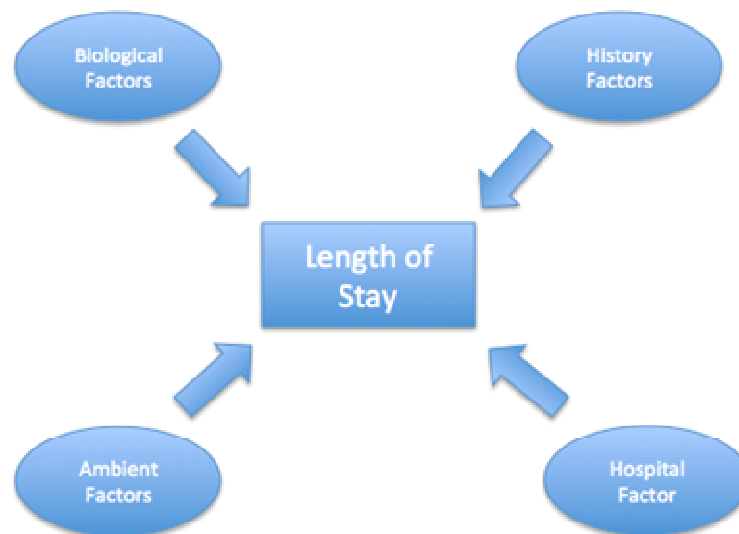


Fig. 1. Relevant aspects on the Length of Hospital Stay

2 Knowledge Representation and Reasoning

Many approaches for knowledge representations and reasoning have been proposed using a Logic Programming (LP) methodology to problem solving, namely in the area of Model Theory [4,5,6], and Proof Theory [7,8]. We follow the proof theoretical approach and an extension to the LP language, to knowledge representations and reasoning. An Extended Logic Program is a finite set of clauses in the form:

$$p \leftarrow p_1, \dots, p_n, \text{not } q_1, \dots, \text{not } q_m \quad (1)$$

$$?(p_1, \dots, p_n, \text{not } q_1, \dots, \text{not } q_m) \quad (n, m \geq 0) \quad (2)$$

where $?$ is a domain atom denoting falsity, the p_i , q_j , and p are classical ground literals, i.e., either positive atoms or atoms preceded by the classical negation sign \neg [8]. Under this representation formalism, every program is associated with a set of abducibles [4], [6] given here in the form of exceptions to the extensions of the predicates that make the program. Once again, LP emerged as an attractive formalism for knowledge representations and reasoning tasks, introducing an efficient search mechanism for problem solving.

Due to the growing need to offer user support in decision-making processes some studies have been presented related to the qualitative models and qualitative reasoning in Database Theory and in Artificial Intelligence research [9,10]. With respect to the problem of knowledge representation and reasoning in Logic Programming (LP), a measure of the *Quality-of-Information (QoI)* of such programs has been object of some work with promising results [11,12]. The *QoI* with respect to the extension of a predicate i will be given by a truth-value in the interval $[0,1]$, i.e., if the information is *known (positive)* or *false (negative)* the *QoI* for the extension of $predicate_i$ is 1. For situations where the information is unknown, the *QoI* is given by:

$$QoI_i = \lim_{N \rightarrow \infty} \frac{1}{N} = 0 \quad (N \gg 0) \quad (3)$$

where N denotes the cardinality of the set of terms or clauses of the extension of $predicate_i$ that stand for the incompleteness under consideration. For situations where the extension of $predicate_i$ is unknown but can be taken from a set of values, the *QoI* is given by:

$$QoI_i = 1/_{Card} \quad (4)$$

where $Card$ denotes the cardinality of the *abducibles* set for i , if the *abducibles* set is disjoint. If the *abducibles* set is not disjoint, the *QoI* is given by:

$$QoI_i = \frac{1}{C_1^{Card} + \dots + C_{Card}^{Card}} \quad (5)$$

where C_{Card}^{Card} is a card-combination subset, with $Card$ elements. The next element of the model to be considered is the relative importance that a predicate assigns to each of its

attributes under observation, i.e., w_i^k , which stands for the relevance of attribute k in the extension of $predicate_i$. It is also assumed that the weights of all the attribute predicates are normalized, i.e.:

$$\sum_{1 \leq k \leq n} w_i^k = 1, \forall_i \quad (6)$$

where \forall denotes the universal quantifier. It is now possible to define a predicate's scoring function $V_i(x)$ so that, for a value $x = (x_1, \dots, x_n)$, defined in terms of the attributes of $predicate_i$, one may have:

$$V_i(x) = \sum_{1 \leq k \leq n} w_i^k \times QoI_i(x)/n \quad (7)$$

It is now possible to engender the universe of discourse, according to the information given in the logic programs that endorse the information about the problem under consideration, according to productions of the type:

$$predicate_i(x_1, \dots, x_n) :: QoI \quad (8)$$

and evaluate the Degree of Confidence given by $DoC = V_i(x_1, \dots, x_n)/n$, which denotes one's confidence in a particular term of the extension of $predicate_i$. To be more general, let us suppose that the Universe of Discourse is described by the extension of the predicates:

$$f_1(\dots), f_2(\dots), \dots, f_n(\dots) \text{ where } (n \geq 0) \quad (9)$$

Assuming we have a clause that is mapped into a case, that clause has as argument all the attributes that make the case. The argument values may be of the type unknown or members of a set, may be in the scope of a given interval or may qualify a particular observation. Let us consider the following clause where the first argument value may fit into the interval [35,65] with a domain of [0,100], the value of the second argument is unknown, which is represented by the symbol \perp , with a domain that ranges in the interval [0,20], and the third argument stands for itself, with a domain that ranges in the interval [0,4]. Let us consider that the case data is given by the extension of predicate f_i , given in the form:

$$f_1: x_1, x_2, x_3 \rightarrow \{True, False\} \quad (10)$$

where “{“ and “}” is one's notation for sets. One may have:

```
{
   $\neg f_1(x_1, x_2, x_3) \leftarrow not f_1(x_1, x_2, x_3)$ 
   $f_1([35, 65], \perp, 1) :: 0.85$ 
   $\underbrace{[0, 100][0, 20][0, 4]}_{\text{attribute's domains for } x_1, x_2, x_3}$ 
  ...
}
```

Once the clauses or terms of the extension of the predicate are established, the next step is to transform all the arguments, of each clause, into continuous intervals. In this phase, it is essential to consider the domain of the arguments. As the second argument is unknown, its interval will cover all the possibilities of the domain. The third argument speaks for itself. 0.85 taken from the interval [0,1] denoting the *QoI* of the term $f_1([35, 65], \perp, 1)$. Therefore, one may have:

```
{
   $\neg f_1(x_1, x_2, x_3) \leftarrow not f_1(x_1, x_2, x_3)$ 
   $f_1(\underbrace{[35, 65], [0, 20], [1, 1]}_{\substack{\text{attribute's values ranges} \\ \text{for } x_1, x_2, x_3}}) :: 0.85$ 
   $\underbrace{[0, 100] [0, 20] [0, 4]}_{\text{attribute's domains for } x_1, x_2, x_3}$ 
  ...
}
```

Now, one is in position to calculate the *Degree of Confidence* for each attribute that makes the term's arguments (e.g. for attribute one it denotes one's confidence that the attribute under consideration fits into the interval [35,65]). Next, we set the boundaries of the arguments intervals to be fitted in the interval [0,1] according to the normalization procedure given in the form by $(Y - Y_{min}) / (Y_{max} - Y_{min})$, where the Y_s stand for themselves.

```
{
   $\neg f_1(x_1, x_2, x_3) \leftarrow not f_1(x_1, x_2, x_3)$ 
   $x_1 = \left[ \frac{35-0}{100-0}, \frac{65-0}{100-0} \right], x_2 = \left[ \frac{0-0}{20-0}, \frac{20-0}{20-0} \right], x_3 = \left[ \frac{1-0}{4-0}, \frac{1-0}{4-0} \right]$ 
   $f_1(\underbrace{[0.35, 0.65], [0, 1], [0.25, 0.25]}_{\substack{\text{attribute's values ranges for } x_1, x_2, x_3 \\ \text{once normalized}}}) :: 0.85$ 
   $\underbrace{[0, 1] [0, 1] [0, 1]}_{\substack{\text{attribute's domains for } x_1, x_2, x_3 \\ \text{once normalized}}}$ 
  ...
}
```

The *Degree of Confidence* (*DoC*) is evaluated using the equation $DoC = \sqrt{1 - \Delta l^2}$, as it is illustrated in Fig. 2. Here Δl stands for the length of the arguments intervals, once normalized.

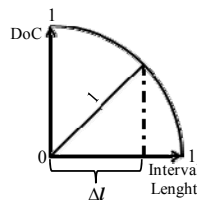


Fig. 2. Evaluation of the Degree of Confidence

```

{
   $\neg f_1(x_1, x_2, x_3) \leftarrow \text{not } f_1(x_1, x_2, x_3)$ 
   $f_{1_{\text{doc}}} \frac{(0.95, \quad 0, \quad 1)}{\text{attribute's confidence values for } x_1, x_2, x_3}} :: 0.85$ 
   $\frac{[0.35, 0.65][0, 1] [0.25, 0.25]}{\text{attribute's values ranges for } x_1, x_2, x_3 \text{ once normalized}}$ 
   $\frac{[0, 1] \quad [0, 1] \quad [0, 1]}{\text{attribute's domains for } x_1, x_2, x_3 \text{ once normalized}}$ 
  ...
}

```

3 A Case Study

In order to exemplify the applicability of our model, we will look at the relational database model, since it provides a basic framework that fits into our expectations [13], and is understood as the genesis of the LP approach to Knowledge Representation and Reasoning [8].

As a case study, consider a scenario where a relational database is given in terms of the extensions of the relations depicted in Fig. 3, which stands for a situation where one has to interpret the information and data about a patient with a given disease. Under this scenario some incomplete and/or default data is also available. For instance, relatively to *Ambient Factors* the presence/absence of underlying problems for case 1 is unknown, while in case 2, it varies from 5 to 7. The *Length of Stay* relation (Fig. 3) is populated according to various results of the different columns, namely *Biological Factors*, *Historical Factors*, *Hospital Factors* and *Ambient Factors*.

These are all the necessary variables to estimate the importance/risk of a patient remaining in Hospital Care, as well as their boundaries.

It should be reminded and noted that every single value is classified from level 0 to 10. The higher the level of stay energises the more important it is. As seen in the Fig. 3 every set has the equivalent result. These will and can be, later, calibrated by the medical staff according to their needs.

Then, one may abstract on the subject and describe the extension of the *length* predicate as:

length: $Bio_{logical\ Factor}, Hist_{orical\ Factor}, Hosp_{ital\ Factor}, Amb_{ient\ Factor} \rightarrow \{0,1\}$

where 0 (zero) and 1 (one) denote, respectively the situation of *not need* or *need* of keeping the patient in the health care facility. It is now possible to give the extension of the predicate *length*, in the form:

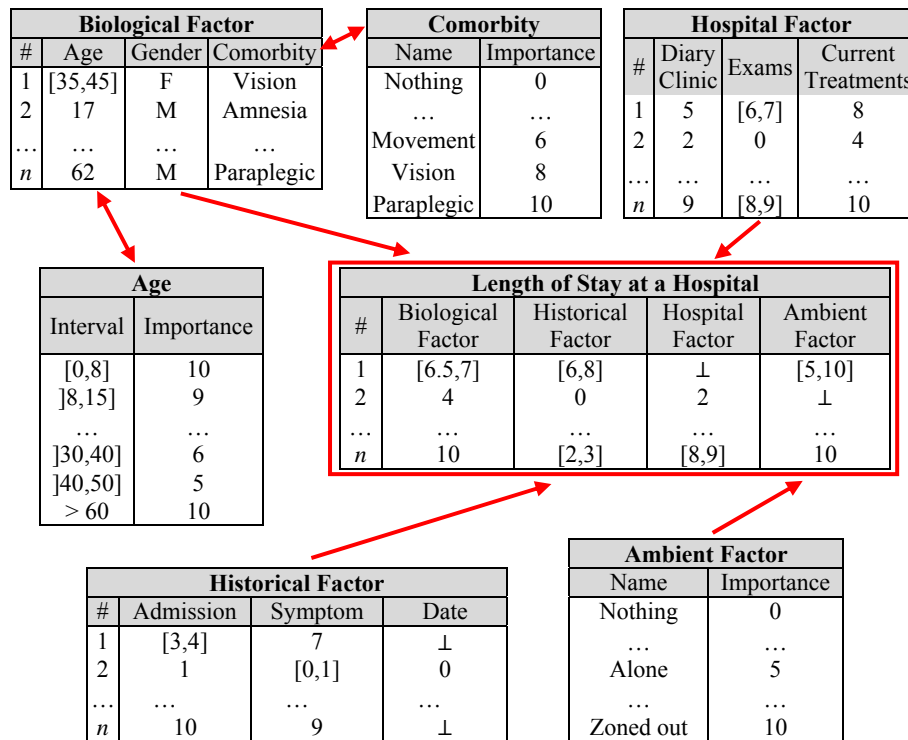


Fig. 3. A summary of the Relational Database model

```

{
  ¬length(Bio,Hist,Hosp,Amb) ← not length(Bio,Hist,Hosp,Amb)
  length(
    
$$\underbrace{([6.5,7], [6,8], \perp, [5,10])}_{\text{attribute's values}}$$

    
$$\underbrace{[0,10] [0,10] [0,10] [0,10]}_{\text{attribute's domains}}$$

  ) :: 1
  ...
}

```

In this program, the first clause denotes the closure of predicate *length*. The next clause corresponds to patient 2, taken from the extension of the *length* relation presented in Fig. 3. Patient 2 is a female, an age uncertain, somewhere between 35 to 45 years with problems in her vision, i.e., blind. Given that her age varies between 35 and 50, it can correspond to *Importance* 6 or *Importance* 5, according to *Age* Table. *Comorbidity* Table shows that vision corresponds to *Importance* 8. So doing an average of the respective values to her biological factors will give us a variable whose values range in the interval [6.5,7]. We keep doing these kinds of operations for all the aspects that are not unknown. If they are unknown we just assume the level can go from 1 to 10. To eliminate any doubt, case 2, refers to a patient that should no longer

remain in the Health Care facility, while case 1 is the opposite and ought to remain held for further examination and care. For further enlightenment, it is essential to explain that in the *Historical Factor Table*, *Admission*, *Symptom* and *Date* columns is relative to the severity of his past status. *Date* is the number of times he/she has been in the hospital. The latter, *Admission* and *Symptom* is referent to his/her health condition and how it evolved during the admission and subsequent period of time in the medical facility.

Moving on, the next step is to transform all the argument values into continuous intervals and then normalize the predicate's arguments to obtain the *Degree of Confidence* of the *length* predicate. One may have:

$$\begin{aligned}
 &\{ \\
 &\quad \neg \text{length}(\text{Bio}, \text{Hist}, \text{Hosp}, \text{Amb}) \leftarrow \text{not length}(\text{Bio}, \text{Hist}, \text{Hosp}, \text{Amb}) \\
 &\quad \text{length}_{\text{Doc}} \left(\underbrace{\begin{pmatrix} 0.99, & 0.98, & 0, & 0.87 \end{pmatrix}}_{\text{attribute's confidence values}} \right) :: 1 \\
 &\quad \quad \underbrace{\begin{pmatrix} [0.65, 0.7] & [0.7, 0.8] & [0, 1] & [0.5, 1] \end{pmatrix}}_{\text{attribute's values ranges once normalized}} \\
 &\quad \quad \underbrace{\begin{pmatrix} [0, 1] & [0, 1] & [0, 1] & [0, 1] \end{pmatrix}}_{\text{attribute's domains once normalized}} \\
 &\quad \dots \\
 &\}
 \end{aligned}$$

where its terms make the training and test sets of the Artificial Neural Network given in Fig. 4.

4 Artificial Neural Networks

Several studies have demonstrated how Artificial Neural Networks (ANNs) could be successfully used to model data and capture complex relationships between inputs and outputs [14,15,16]. ANNs simulate the structure of the human brain being populated by multiple layers of neurons. As an example, let us consider the last case presented in Fig. 3, where one may have a situation in which a prolonged hospital stay is needed, which is given in the form:

$$\begin{aligned}
 &\{ \\
 &\quad \neg \text{length}(\text{Bio}, \text{Hist}, \text{Hosp}, \text{Amb}) \leftarrow \text{not length}(\text{Bio}, \text{Hist}, \text{Hosp}, \text{Amb}) \\
 &\quad \text{length} \left(\underbrace{\begin{pmatrix} 10, & [2,3], & [8,9], & 10 \end{pmatrix}}_{\text{attribute's values}} \right) :: 1 \\
 &\quad \quad \underbrace{\begin{pmatrix} [0,10] & [0,10] & [0,10] & [0,10] \end{pmatrix}}_{\text{attribute's domains}} \\
 &\}
 \end{aligned}$$

↓ 1st interaction: transition to continuous intervals

$$\begin{aligned}
& \text{length} \left(\frac{\underbrace{[10,10], [2,3], [8,9], [10,10]}_{\text{attribute's values ranges}}}{\underbrace{[0,10] \ [0,10] [0,10] [0,10]}_{\text{attribute's domains}}} \right) :: 1 \\
& \Downarrow \text{2nd interaction: normalization } \frac{Y - Y_{\min}}{Y_{\max} - Y_{\min}} \\
& \text{length} \left(\frac{\underbrace{[1,1], [0.2,0.3], [0.8,0.9], [1,1]}_{\text{attribute's values ranges once normalized}}}{\underbrace{[0,1] \ [0,1] \ [0,1] \ [0,1]}_{\text{attribute's domains}}} \right) :: 1 \\
& \Downarrow \text{DoC calculation: } DoC = \sqrt{1 - \Delta I^2} \\
& \text{length}_{DoC} \left(\frac{\underbrace{1, \ 0.99, \ 0.99, \ 1}_{\text{attribute's confidence values}}}{\frac{\underbrace{[1,1] [0.2,0.3] [0.8,0.9] [1,1]}_{\text{attribute's values ranges}}}{\underbrace{[0,1] \ [0,1] \ [0,1] \ [0,1]}_{\text{attribute's domains}}}} \right) :: 1 \\
& \}
\end{aligned}$$

In Fig. 4 it is shown how the normalized values of the interval boundaries and their *DoC* and *QoI* values work as inputs to the ANN. The output translates the necessity of increasing the length of stay on a Hospital of a patient, and *DoC* the confidence that one has on such a happening. In addition, it also contributes to build a database of study cases that may be used to train and test the ANNs. In this study were considered 300 samples (i.e. three hundred terms or clauses of the extension of predicate), of which 67% were used to train the ANN, and 33% were considered for testing (the data came from a main health care center in the north of Portugal). The back propagation algorithm was used in the learning process of the ANN. As the output function in the pre-processing layer it was used the identity one. In the others layers we used the sigmoid function.

5 Conclusions and Future Work

To understand if the increase of the length of hospital stay is mandatory for the well being of the patient is a hard and complex task, which needs to consider many different conditions with intricate relations among them. These characteristics put this problem into the area of problems that may be tackled by Artificial Intelligence (AI) based methodologies and techniques to problem solving. Despite that, little to no work has been done in that direction. This work presents the founding of a computa-

tional framework that uses powerful knowledge representation and reasoning techniques to set the structure of the information and the associate inference mechanisms. This representation is above everything else, very versatile and capable of covering every possible instance by considering incomplete, contradictory, and even unknown data.

The main contribution of this work is to be understood in terms of the evaluation of the *DoC*, and the possibility to address the issue of incomplete information and default data. Indeed, the new paradigm to knowledge representation and reasoning enables one to use the normalized values of the interval boundaries and their *DoC* values, as inputs to the ANN. The output translates the necessity of keeping a patient in a hospital order to improve their health and the degree of confidence that one has on such a happening. Future work may recommend that the same problem must be approached using other computational frameworks like Case Based Reasoning [17] or Particle Swarm [18], among others.

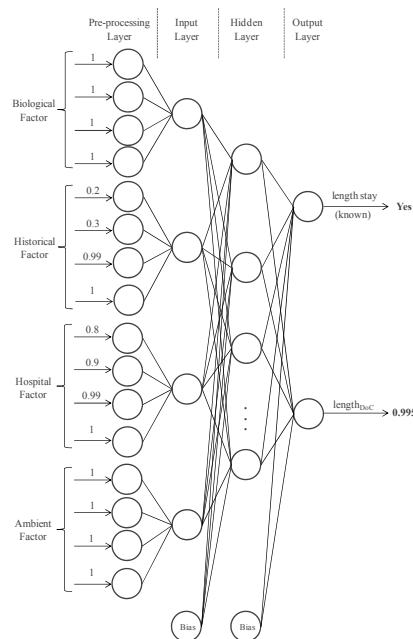


Fig. 4. The Artificial Neural Network topology

Acknowledgments. This work is funded by National Funds through the FCT – Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) within projects PEst-OE/EEI/UI0752/2014 and PEst-OE/QUI/UI0619/2012.

References

1. Fry, J.: General Practice: The Facts. Radcliffe Medical Press, Oxford (1993)

2. Pritchard, P.: *Manual of Primary Health Care: Its Nature and Organisation*. Oxford University Press, Oxford (1981)
3. World Health Organisation: *Primary Health Care*. World Health Organisation, Geneva (1978)
4. Kakas, A., Kowalski, R., Toni, F.: The role of abduction in logic programming. In: Gabbay, D., Hogger, C., Robinson, I. (eds.) *Handbook of Logic in Artificial Intelligence and Logic Programming*, vol. 5, pp. 235–324. Oxford University Press, Oxford (1998)
5. Gelfond, M., Lifschitz, V.: The stable model semantics for logic programming. In: Kowalski, R., Bowen, K. (eds.) *Logic Programming – Proceedings of the Fifth International Conference and Symposium*, pp. 1070–1080 (1988)
6. Pereira, L.M., Anh, H.T.: Evolution prospection. In: Nakamatsu, K. (ed.) *New Advances in Intelligent Decision Technologies – Results of the First KES International Symposium IDT 2009*, *Studies in Computational Intelligence*, vol. 199, pp. 51–64. Berlin: Springer (2009)
7. Neves, J., Machado, J., Analide, C., Abelha, A., Brito, L.: The halt condition in genetic programming. In: Neves, J., Santos, M.F., Machado, J. (eds) *Progress in Artificial Intelligence*, *Lecture Notes in Artificial Intelligence*, vol. 4874, pp. 160–169. Springer, Berlin (2007)
8. Neves, J.: A logic interpreter to handle time and negation in logic databases. In Muller, R.L., Pottmyer, J.J. (eds) *ACM’84 Proceedings of the 1984 annual conference of the ACM on The Fifth Generation Challenge*, pp. 50–54. Association for Computing Machinery, New York (1984)
9. Halpern, J.: *Reasoning about uncertainty*. MIT Press, Massachusetts (2005)
10. Kovalerchuck, B., Resconi, G.: Agent-based uncertainty logic network. In *Proceedings of the IEEE International Conference on Fuzzy Systems – FUZZ-IEEE 2010*, Barcelona, pp. 596–603 (2010)
11. Lucas, P.: Quality checking of medical guidelines through logical abduction. In Coenen, F., Preece, A., Mackintosh A. (eds) *Proceedings of AI-2003 (Research and Developments in Intelligent Systems XX)*, pp. 309–321. Springer, London (2003)
12. Machado, J., Abelha, A., Novais, P., Neves, J., Neves, J.: Quality of Service in healthcare units. *International Journal of Computer Aided Engineering and Technology* 2, 436–449 (2010)
13. Liu, Y., Sun, M.: Fuzzy optimization BP neural network model for pavement performance assessment. In: *2007 IEEE international conference on grey systems and intelligent services*, Nanjing, China, pp. 18–20 (2007)
14. Caldeira, A.T., Arteiro, J.M., Roseiro, J.C. Neves, J., Vicente, H.: An Artificial Intelligence Approach to *Bacillus amyloliquefaciens* CCMI 1051 Cultures: Application to the Production of Antifungal Compounds. *Bioresource Technology*. 102, 1496-1502 (2011)
15. Vicente, H., Dias, S., Fernandes, A., Abelha, A., Machado, J., Neves, J.: Prediction of the Quality of Public Water Supply using Artificial Neural Networks. *Journal of Water Supply: Research and Technology – AQUA* 61, 446–459 (2012)
16. Salvador, C., Martins, M.R., Vicente, H., Neves, J., Arteiro J.M., Caldeira, A.T.: Modeling Molecular and Inorganic Data of *Amanita ponderosa* Mushrooms using Artificial Neural Networks. *Agroforestry Systems* 87, 295–302 (2013)
17. Mendes, R., Kennedy, J., Neves, J.: The Fully Informed Particle Swarm: Simpler, Maybe Better. *IEEE Transactions on Evolutionary Computation* 8, 204–210 (2004)
18. Carneiro, D., Novais, P., Andrade, F., Zeleznikow, J., Neves, J.: Using Case-Based Reasoning and Principled Negotiation to provide decision support for dispute resolution. *Knowledge and Information Systems* 36, 789–826 (2013)

Bio-Inspired Hybrid Intelligent Method for Detecting Android Malware

Konstantinos Demertzis¹, Lazaros Iliadis²

^{1,2} Democritus University of Thrace, Department of Forestry & Management of the Environment & Natural Resources, 193 Pandazidou st., 68200 N Orestiada, Greece, kdemertz@fmenr.duth.gr¹, liliadis@fmenr.duth.gr²

Abstract. Today's smartphones are capable of doing much more than the previous generation of mobile phones. However this extended range of capabilities is coming together with some new security risks. Also, mobile platforms often contain small, insecure and less well controlled applications from various single developers. Due to the open usage model of the Android market, malicious applications cannot be avoided completely. Especially pirated applications or multimedia content in popular demand, targeting user groups with typically low awareness levels are predestined to spread too many devices before being identified as malware. Generally malware applications utilizing root exploits to escalate their privileges can inject code and place binaries outside applications storage locations. This paper proposes a novel approach, which uses minimum computational power and resources, to identify Android malware or malicious applications. It is a bio-inspired Hybrid Intelligent Method for Detecting Android Malware (HIMDAM). This approach performs classification by employing Extreme Learning Machines (ELM) in order to properly label malware applications. At the same time, Evolving Spiking Neural Networks (eSNNs) are used to increase the accuracy and generalization of the entire model.

Keywords: Security, Android Malware, Evolving Spiking Neural Networks, Extreme Learning Machines, Radial Basis Function Networks, Polynomial Neural Networks, Self-Organizing Maps, Multilayer Perceptron

1. Introduction

Lately, the share of smartphones in the sales of handheld mobile communication devices has drastically increased. Among them, the number of Android based smartphones is growing rapidly. They are increasingly used for security critical private and business applications, such as online banking or to access corporate networks. This makes them a very valuable target for an adversary. Until recently, the Android Operating System's security model has succeeded in preventing any significant attacks by malware. This can be attributed to a lack of attack vectors which could be used for self-spreading infections and low sophistication of malicious applications. However, emerging malware deploys advanced attacks on operating system components to assume full device control [1]. Malware are the most common infection method because the malicious code can be packaged and redistributed with

popular applications. In Android, each application has an associated *.apk* file which is the executable file format for this platform. Due to the open software installation nature of Android, users are allowed to install any executable file from any application store. This could be from the official Google Play store, or a third party site. This case of installing applications makes Android users vulnerable to malicious applications. Some of the most widely used solutions such as antivirus software are inadequate for use on smartphones as they consume too much CPU and memory and might result in rapid draining of the power source. In addition, most antivirus detection capabilities depend on the existence of an updated malware signature repository, therefore the antivirus users are not protected from zero-day malware.

This research effort aims in the development and application of an innovative, fast and accurate bio-inspired Hybrid Intelligent Method for Detecting Android Malware (HIMDAM). This is achieved by employing Extreme Learning Machines (ELMs) and Evolving Spiking Neural Networks (eSNNs). A RBF Kernel ELM has been employed for malware detection, which offers high learning speed, ease of implementation and minimal human intervention. Also, an eSNN model has been applied to increase the accuracy and generalization of the entire method. In fact, the bio-inspired model has shown better performance when compared to other ANN methods, such as Multilayer Perceptrons (MLP), Radial Basis Function ANN (RBF), Self-Organizing Maps (SOM), Group Methods of Data Handling (GMDH) and Polynomial ANN. A main advantage of HIMDAM is the fact that it reduces overhead and overall analysis time, by classifying malicious and benign applications with high accuracy.

1.1 Literature review

Significant work has been done in applying machine learning (ML) techniques, using features derived from both static [2][3][4] and dynamic [5] analysis to identify malicious Android applications [6]. Amongst early efforts towards Android applications security was the “*install-time policy security system*” developed by Enck et al. which considered risks associated with combinations of the app permissions [7]. From another perspective, some works focused in the runtime analysis [8],[9] whereas others have tried a static analysis of apps [10]. For instance, Chin et al. [4] used a 2-means clustering [11] of apps’ call activities, to detect Trojans. Fuchs et al. [12] used formal static analysis of byte codes [13] to form data flow-permission consistency as a constrained optimization problem. Barrera et al. [14] used app permissions in self-organizing maps (SOMs) to visualize app permission usage as a U-matrix [15]. Besides, their SOM component plane analysis allowed identification of the frequently jointly requested permissions. However, they did not relate categories and permissions. In [16], Tesauro et al. train ANN to detect boot sector viruses, based on byte string trigrams. Schultz et al. [17] compare three machine learning algorithms trained on three features: DLL and system calls made by the program, strings found in the program binary and a raw hexadecimal representation of the binary [18]. Kotler and Maloof [19] used a collection of 1971 benign and 1651 malicious executable files. N-grams were extracted and 500 features were selected using the Information Gain measure. The vector of n-gram features was binary, presenting the presence or absence of a feature in the file. In their experiment, they trained several classifiers:

IBK k-Nearest Neighbors (k-NN), a similarity-based classifier called the TFIDF classifier, Naïve Bayes, Support Vector Machines (SVM) and Decision Trees under the algorithm J48 [20]. The last three of these were also boosted. In the experiments, the four best-performing classifiers were Boosted J48, SVM, Boosted SVM and IBK [20]. Also, Cheng et al. [21] proposed the use of ELM methods to classify binary and multi-class network traffic for intrusion detection. The performance of ELM in both binary-class and multi-class scenarios are investigated and compared to SVM based classifiers. Joseph et al., [22] developed an autonomous host-dependent Intrusion Detection System (IDS) for identifying malicious sinking behavior. This system increases the detection accuracy by using cross-layer features to describe a routing behavior. Two ML approaches were exploited towards learning and adjustment to new kind of attack circumstances and network surroundings. ELMs and Fisher Discriminant Analysis (FDA) are utilized collectively to develop better accuracy and quicker speed of method.

2. Methodologies comprising the proposed hybrid approach

2.1 Extreme Learning Machines (ELM)

The extreme learning machine (ELM) as an emerging learning technique provides efficient unified solutions to generalized feed-forward networks including but not limited to (both single- and multi-hidden-layer) neural networks, radial basis function (RBF) networks, and kernel learning [23]. ELM theories show that hidden neurons are important but can be randomly generated, independent from applications and that ELMs have both universal approximation and classification capabilities. They also build a direct link between multiple theories namely: ridge regression, optimization, ANN generalization performance, linear system stability and matrix theory. Thus, they have strong potential as a viable alternative technique for large-scale computing and ML. Also ELMs, are biologically inspired, because hidden neurons can be randomly generated independent of training data and application environments, which has recently been confirmed with concrete biological evidences. ELM theories and algorithms argue that “random hidden neurons” capture the essence of some brain learning mechanism as well as the intuitive sense that the efficiency of brain learning need not rely on computing power of neurons. This may somehow hint at possible reasons why brain is more intelligent and effective than computers [24].

ELM works for the “generalized” Single-hidden Layer feedforward Networks (SLFNs) but the hidden layer (or called feature mapping) in ELM need not be tuned. Such SLFNs include but are not limited to SVMs, polynomial networks, RBFs and the conventional (both single-hidden-layer and multi-hidden-layer) feedforward ANN. Different from the tenet that all the hidden nodes in SLFNs need to be tuned, ELM learning theory shows that the hidden nodes/neurons of generalized feedforward networks needn't be tuned and these hidden nodes/neurons can be randomly generated [23]. All the hidden node parameters are independent from the target functions or the training datasets. ELMs conjecture that this randomness may be true to biological learning in animal brains. Although in theory, all the parameters of

ELMs can be analytically determined instead of being tuned, for the sake of efficiency in real applications, the output weights of ELMs may be determined in different ways (with or without iterations, with or without incremental implementations) [23]. According to ELM theory the hidden node/neuron parameters are not only independent of the training data but also of each other. Unlike conventional learning methods which must see the training data before generating the hidden node/neuron parameters, ELMs could randomly generate the hidden node/neuron parameters before seeing the training data. In addition, ELMs can handle non-differentiable activation functions, and do not have issues such as finding a suitable stopping criterion, learning rate, and learning epochs. ELMs have several advantages, ease of use, faster learning speed, higher generalization performance, suitable for many nonlinear activation function and kernel functions [23].

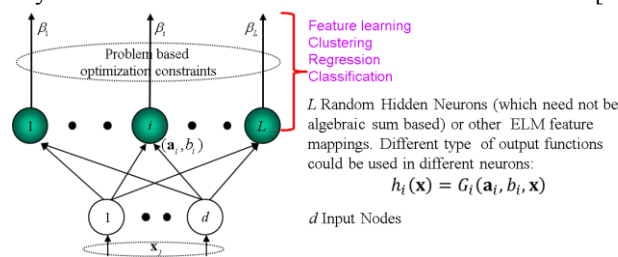


Fig 1. Extreme Learning Machine (ELM) [23]

2.2 Evolving Spiking Neural Networks (eSNNs)

The eSNNs are modular connectionist-based systems that evolve their structure and functionality in a continuous, self-organized, on-line, adaptive, interactive way from incoming information. These models use trains of spikes as internal information representation rather than continuous variables [25]. The eSNN developed and discussed herein is based in the “Thorpe” neural model [26]. This model intensifies the importance of the spikes taking place in an earlier moment, whereas the neural plasticity is used to monitor the learning algorithm by using one-pass learning. In order to classify real-valued data sets, each data sample, is mapped into a sequence of spikes using the Rank Order Population Encoding (ROPE) technique [27] [28]. The topology of the developed eSNN is strictly feed-forward, organized in several layers and weight modification occurs on the connections between the neurons of the existing layers.

The ROPE method is alternative to the conventional rate coding scheme (CRCS). It uses the order of firing neuron’s inputs to encode information. This allows the mapping of vectors of real-valued elements into a sequence of spikes. Neurons are organized into neuronal maps which share the same synaptic weights. Whenever the synaptic weight of a neuron is modified, the same modification is applied to the entire population of neurons within the map. Inhibition is also present between each neuronal map. If a neuron spikes, it inhibits all the neurons in the other maps with neighboring positions. This prevents all the neurons from learning the same pattern. When propagating new information, neuronal activity is initially reset to zero. Then, as the propagation goes on, each time one of their inputs fire, neurons are

progressively desensitized. This is making neuronal responses dependent upon the relative order of firing of the neuron's afferents [29] [30].

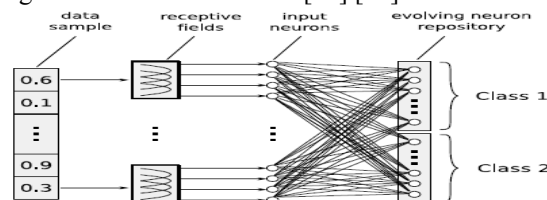


Fig. 2. The Evolving Spiking Neural Network (eSNN) architecture [29]

The aim of the one-pass learning method is to create a repository of trained output neurons during the presentation of training samples. After presenting a certain input sample to the network, the corresponding spike train is propagated through the eSNN which may result in the firing of certain output neurons. It is possible that no output neuron is activated and the network remains silent and the classification result is undetermined. If one or more output neurons have emitted a spike, the neuron with the shortest response time among all activated output neurons is determined. The label of this neuron is the classification result for the presented input [31].

3. Description of the proposed HIMDAM algorithm

The proposed herein, HIMDAM methodology uses an ELM classification approach to classify malware or benign applications with minimum computational power and time, combined with the eSNN method in order to detect and verify the malicious code. The general algorithm is described below:

Step 1: Train and test *datasets* are determined and normalized to the interval [-1,1]. The *datasets* are divided in 4 main sectors with “*permission*” feature. *Permission* is a security mechanism of mobile operating systems. For mobile phones any application executed under the device owner’s user ID would be able to access any other application’s data. The Android kernel assigns each application its own user ID on installation. To avoid the abuse of mobile phone functions, Android allows the user to effectively identify and manage mobile phone resources by setting permissions. If the application requires a certain function, the developer can announce permission. In the latest version of Android, there are a total of 130 permissions. To malware, some permissions are important and frequently needed, therefore they should be weighted. For example, the attacker needs permissions to transfer the stolen data to his account through the web, or to perform damaging behavior by sending out large number of SMS messages. The features involved can be divided in the sectors below:

1. *Battery + Permissions (5 features)*
2. *Binder + Permissions (18 features)*
3. *Memory + CPU + Permissions (10 features)*
4. *Network + Permissions (9 features)*

The *Hardware_Dataset* has been generated (16 features) including the most important variables from hardware related sectors (Battery, Memory, CPU, Network). On the other hand, the *All_Imp_Var_Dataset* (27 features) comprises of the most important

variables from all of the sectors (Battery, Memory, CPU, Network, Binder). To calculate the importance of variables we replace them with their mean values one by one and we measure the root mean squared error (RMSE) of the “new” model. Original model error is considered to have a zero percent impact on the RMSE and 100% impact is a case where all variables are replaced with their mean. The impact can easily exceed 100% when the variable in a model is multiplied by another one or it is squared. A small negative percentage can also happen if a variable is merely useless for the model.

In order to create a very fast and accurate prediction model with minimum requirements of hardware resources, we randomly check two sectors (e.g. Battery and Binder or Memory and Binder) every time with the ELM classifier. According to the ELM theory [32], the Gaussian Radial Basis Function kernel $K(u,v)=\exp(-\gamma||u-v||^2)$ is used. The hidden neurons are $k=20$, w_i are the assigned random input weights and b_i the biases, where $i=1,\dots,N$ and H is the hidden layer output matrix.

$$H = \begin{bmatrix} h(x_1) \\ \vdots \\ h(x_N) \end{bmatrix} = \begin{bmatrix} h_1(x_1) & \cdots & h_L(x_1) \\ \vdots & & \vdots \\ h_1(x_N) & \cdots & h_L(x_N) \end{bmatrix} \quad (1)$$

$h(x) = [h_1(x), \dots, h_L(x)]$ is the output (row) vector of the hidden layer with respect to the input x . Function $h(x)$ actually maps the data from the d -dimensional input space to the L -dimensional hidden-layer feature space (ELM feature space) H and thus, $h(x)$ is indeed a feature mapping. ELM aims to minimize the training error as well as the norm of the output weights as shown in equation 2:

$$\text{Minimize : } ||H\beta - T||^2 \text{ and } ||\beta|| \quad (2)$$

To minimize the norm of the output weights $||\beta||$ is actually to maximize the distance of the separating margins of the two different classes in the ELM feature space $2/||\beta||$. The calculation of the output weights β is done according to equation (3):

$$\beta \left(\frac{1}{c} + H^T H \right)^{-1} H^T T = \quad (3)$$

where c is a positive constant and T is obtained from the *Function Approximation of*

SLFNs with additive neurons with $t_i = [t_{i1}, t_{i2}, \dots, t_{im}]^T \in R^m$ and $T = \begin{bmatrix} t_1^T \\ \vdots \\ t_N^T \end{bmatrix}$.

It has been shown numerically in ELM theory [32] that the above solution has better generalization performance. More specifically, the reasoning of the new Malware detection algorithm that has been developed in this research is as seen below:

- *If both sectors' analysis with the ELM offer Negative results, no action is required and the next 2 sectors are examined.*
- *If the ELM analysis results in a Negative result for the one sector and positive for the other then:*
 - *If both sectors belong to the general Hardware field (eg. Network and Battery) then the Hardware_Dataset is reexamined.*
 - *If the result is Negative then we go further.*
 - *If the result is Positive then the whole Original Dataset with all 40 features is checked.*
- *If the ELM analysis of both sectors produces Positive results*
 - *Then the whole Original Dataset with all 40 features is checked.*

Step 2: The train and test datasets are determined and formed, related to n features. The required classes (malware and benign applications) that use the variable *Population Encoding* are imported. This variable controls the conversion of real-valued data samples into the corresponding time spikes. The encoding is performed with 20 Gaussian receptive fields per variable (Gaussian width parameter $\beta=1.5$). The data are normalized to the interval $[-1,1]$ and so the coverage of the Gaussians is determined by using i_{\min} and i_{\max} . For the normalization processing the following equation is used: $x_{i_{\text{norm}}} = 2 * \left(\frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \right) - 1, x \in \mathbb{R}$ (4)

The data is classified in two classes namely: **Class positive** which contains the benign results and **Class negative** which comprises of the malware ones. The eSNN is using modulation factor $m=0.9$, firing threshold ratio $c=0.7$ and similarity threshold $s=0.6$ in agreement with the vQEA algorithm [30] [31]. More precisely, let $A = \{a_1, a_2, a_3 \dots a_{m-1}, a_m\}$ be the ensemble of afferent neurons of neuron i and $W = \{w_{1,i}, w_{2,i}, w_{3,i} \dots w_{m-1,i}, w_{m,i}\}$ the weights of the m corresponding connections; let $\text{mod} \in [0,1]$ be an arbitrary modulation factor. The activation level of neuron i at time t is given by equation 5: $\text{Activation}(i,t) = \sum_{j \in [1,m]} \text{mod}^{\text{order}(a_j)} w_{j,i}$ (5) where $\text{order}(a_j)$ is the firing rank of neuron a_j in the ensemble A .

By convention, $\text{order}(a_j) = +8$ if a neuron a_j is not fired at time t , sets the corresponding term in the above sum to zero. This kind of desensitization function could correspond to a fast shunting inhibition mechanism. When a neuron reaches its threshold, it spikes and inhibits neurons at equivalent positions in the other maps so that only one neuron will respond at any location. Every spike triggers a time based Hebbian-like learning rule that adjusts the synaptic weights. Let t_e be the date of arrival of the Excitatory PostSynaptic Potential (EPSP) at synapse of weight W and t_a the date of discharge of the postsynaptic neuron.

If $t_e < t_a$ then $dW = a(1-W)e^{-|\Delta o|/\tau}$ else $dW = -aW e^{-|\Delta o|/\tau}$ (6). Δo is the difference between the date of the EPSP and the date of the neuronal discharge (expressed in term of order of arrival instead of time), a is a constant that controls the amount of synaptic potentiation and depression [27]. ROPE technique with receptive fields, allow the encoding of continuous values. Each input variable is encoded independently by a group of one-dimensional receptive fields (figure 1). For a variable n , an interval $[I_{\min}^n, I_{\max}^n]$ is defined. The Gaussian receptive field of neuron i is given by its center μ_i and width σ by equation 8. $\mu_i = I_{\min}^n + \frac{2i-3}{2} \frac{I_{\max}^n - I_{\min}^n}{M-2}$ (7) $\sigma = \frac{1}{\beta} \frac{I_{\max}^n - I_{\min}^n}{M-2}$ (8) where $1 \leq \beta \leq 2$ and the parameter β directly controls the width of each Gaussian receptive field. Figure 3 depicts an encoding example of a single variable.

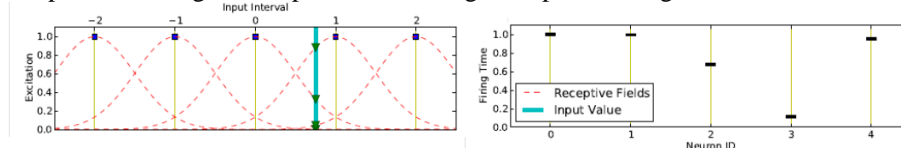


Fig. 3. Population encoding based on Gaussian receptive fields. Left Figure: Input Interval – Right Figure: Neuron ID

For an input value $v=0.75$ (thick straight line) the intersection points with each Gaussian is computed (triangles), which are in turn translated into spike time delays

(right figure) [30]. **Step 2:** The eSNN is trained with the *training* dataset vectors and the testing is performed with the *testing* vectors. The procedure of one pass learning is described in the following Algorithm 2 [29] [30].

Algorithm 2: Training an evolving Spiking Neural Network (eSNN) [30]

Require: m_l, s_l, c_l for a class label $l \in L$

- 1: initialize neuron repository $R_l = \{ \}$
 - 2: **for all** samples $X^{(i)}$ belonging to class l **do**
 - 3: $w_j^{(i)} \leftarrow (m_l)^{\text{order}(j)}, \forall j \mid j \text{ pre-synaptic neuron of } i$
 - 4: $u_{\max}^{(i)} \leftarrow \sum_j w_j^{(i)} (m_l)^{\text{order}(j)}$
 - 5: $\theta^{(i)} \leftarrow c_l u_{\max}^{(i)}$
 - 6: **if** $\min(d(w^{(i)}, w^{(k)})) < s_l, w^{(k)} \in R_l$ **then**
 - 7: $w^{(k)} \leftarrow \text{merge } w^{(i)} \text{ and } w^{(k)} \text{ according to Equation 7}$
 - 8: $\theta^{(k)} \leftarrow \text{merge } \theta^{(i)} \text{ and } \theta^{(k)} \text{ according to Equation 8}$
 - 9: **else**
 - 10: $R_l \leftarrow R_l \cup \{w^{(i)}\}$
 - 11: **end if**
 - 12: **end for**
-

For each training sample i with class label l which represent a benign software, a new output neuron is created and fully connected to the previous layer of neurons, resulting in a real-valued weight vector $w^{(i)}$ with $w_j^{(i)} \in R$ denoting the connection between the pre-synaptic neuron j and the created neuron i . In the next step, the input spikes are propagated through the network and the value of weight $w_j^{(i)}$ is computed according to the order of spike transmission through a synapse j : $w_j^{(i)} = (m_l)^{\text{order}(j)}, \forall j \mid j \text{ pre-synaptic neuron of } i$. Parameter m_l is the modulation factor of the Thorpe neural model. Differently labeled output neurons may have different modulation factors m_l . Function $\text{order}(j)$ represents the rank of the spike emitted by neuron j . The firing threshold $\theta^{(i)}$ of the created neuron i is defined as the fraction $c_l \in R, 0 < c_l < 1$, of the maximal possible potential $u_{\max}^{(i)}$: $\theta^{(i)} \leftarrow c_l u_{\max}^{(i)}$

$$(7) \quad u_{\max}^{(i)} \leftarrow \sum_j w_j^{(i)} (m_l)^{\text{order}(j)} \quad (9)$$

The fraction c_l is a parameter of the model and for each class label $l \in L$ a different fraction can be specified. The weight vector of the trained neuron is compared to the weights corresponding to neurons already stored in the repository. Two neurons are considered too “similar” if the minimal *Euclidean* distance between their weight vectors is smaller than a specified similarity threshold s_l (the eSNN object uses optimal similarity threshold $s=0.6$). All parameters of eSNN (modulation factor m_l , similarity threshold s_l , PSP fraction $c_l, l \in L$) included in this search space, were optimized according to the Versatile Quantum-inspired Evolutionary Algorithm (vQEA) [31]. Both the firing thresholds and the weight vectors were merged according to equations 10 and 11:

$$w_j^{(k)} \leftarrow \frac{w_j^{(i)} + N w_j^{(k)}}{1+N}, \forall j \mid j \text{ pre-synaptic neuron of } i \quad (10) \quad \theta^{(k)} \leftarrow \frac{\theta^{(i)} + N \theta^{(k)}}{1+N} \quad (11)$$

Integer N denotes the number of samples previously used to update neuron k . The merging is implemented as the (running) average of the connection weights, and the (running) average of the two firing thresholds. After merging, the trained neuron i is discarded and the next sample processed. If no other neuron in the repository is similar to the trained neuron i , the neuron i is added to the repository as a new output.

4. Data and Results

For this experiment, we used the free dataset provided by B. Amos [33]. The author developed a shell script to automatically analyze *.apk* Android application files by running them in available Android emulators. For each *.apk* file, the emulator simulates user interaction by randomly interacting with the application interface. This is done using the Android “adb-monkey” tool [34]. Based on inspection of the source code, we can conclude that each feature vector of the dataset is collected at 5 seconds’ intervals. The memory features were collected by observing the “proc” directory in the underlying Linux system of Android. The CPU information was collected by running the Linux “top” command. The Battery and Binder information was collected by using “intent” (Action listener) [35].

Table 1. Accuracy (ACC) Comparison between MLP, RBF, ELM, GMDH PNN, eSNN

	MLP		RBF		SOM		ELM		GMDH PNN		eSNN	
	ACC	TIME	ACC	TIME	ACC	TIME	ACC	TIME	ACC	TIME	ACC	TIME
All Features	95.38%	38.31	85.72%	0.34	87.34%	0.20	89.19%	0.17	90.50%	13.00	97.10%	20.22
Battery + Perm	73.58%	1.69	71.96%	0.19	64.28%	0.23	72.49%	0.17	71.30%	2.00	79.30%	1.22
Binder + Perm	93.35%	7.95	79.87%	0.08	72.82%	0.08	82.98%	0.05	84.80%	5.00	93.60%	8.71
Memory Cpu Perm	82.08%	4.84	79.65%	0.13	77.16%	0.23	80.08%	0.12	83.10%	4.00	84.90%	4.40
Network Perm	81.21%	3.02	69.83%	0.13	70.10%	0.19	71.27%	0.14	73.70%	3.00	80.10%	3.11
Important var. from all features	97.69%	14.59	83.76%	0.20	80.34%	0.20	89.47%	0.20	91.00%	8.00	98.20%	11.02
Important var. from Hardware	94.22%	6.69	88.25%	0.22	75.90%	0.17	88.00%	0.14	89.40%	6.00	94.80%	5.91

The original dataset has a total of 1153 data (feature vector) samples with 660 benign samples (classified as positive class) and 493 malicious samples (classified as negative class). It was divided randomly in two parts: 1) a training dataset containing 807 patterns (467 positive and 340 negative patterns) 2) a testing dataset containing 346 patterns (193 positive and 153 negative patterns). To identify the integrity of HIMDAM we have compared the ELM and eSNN classifiers with other neural network methods. The performance of both classifiers was evaluated on malware datasets. The results showed that the kernel based ELM has much faster learning speed (run thousands times faster than conventional methods) and the eSNN has much better generalization performance and more accurate and reliable classification results. The comparisons were performed on a dual boot PC with a P4 at 3.1 GHz CPU and 4GB RAM. For the eSNN classification, the *Linux Ubuntu 12.04 LTS* OS with *PyLab* (*NumPy*, *SciPy*, *Matplotlib* and *IPython*) was employed. The MLP, RBF and SOM tests were performed with the *Weka 3.7* [36], ELM with *Matlab 2013* and GMDH PNN with GMDH shell software [37]. The performance comparisons of NN

algorithms are shown in table 1. The confusion matrices for ELM and eSNN can be seen in table 2.

Table 2. Confusion matrices for ELM and eSNN algorithms

CONFUSION MATRICES FOR ELM			
All Features			
	Benign (Predicted)	Malware (Predicted)	Accuracy
Benign (Actual)	180	13	93.26%
Malware (Actual)	23	130	85.12%
Overall Accuracy			89.19%
Battery Permissions			
Benign (Actual)	149	44	77.20%
Malware (Actual)	50	103	67.78%
Overall Accuracy			72.49%
Binder Permissions			
Benign (Actual)	164	29	84.97%
Malware (Actual)	29	124	80.99%
Overall Accuracy			82.98%
Memory CPU Permissions			
Benign (Actual)	159	34	82.38%
Malware (Actual)	34	119	77.78%
Overall Accuracy			80.08%
Network Permissions			
Benign (Actual)	142	51	73.57%
Malware (Actual)	48	105	68.97%
Overall Accuracy			71.27%
Importance variables from all			
Benign (Actual)	180	13	93.26%
Malware (Actual)	22	131	85.68%
Overall Accuracy			89.47%
Importance variables from Hardware			
Benign (Actual)	175	23	90.67%
Malware (Actual)	23	130	85.33%
Overall Accuracy			88.00%

CONFUSION MATRICES FOR eSNN			
All Features			
	Benign (Predicted)	Malware (Predicted)	Accuracy
Benign (Actual)	190	3	98.40%
Malware (Actual)	7	146	95.80%
Overall Accuracy			97.10%
Battery Permissions			
Benign (Actual)	160	33	82.90%
Malware (Actual)	37	116	75.70%
Overall Accuracy			79.30%
Binder Permissions			
Benign (Actual)	185	8	95.80%
Malware (Actual)	13	140	91.40%
Overall Accuracy			93.60%
Memory CPU Permissions			
Benign (Actual)	168	25	87.00%
Malware (Actual)	26	127	82.80%
Overall Accuracy			84.90%
Network Permissions			
Benign (Actual)	161	32	83.40%
Malware (Actual)	36	117	76.80%
Overall Accuracy			80.10%
Importance variables from all			
Benign (Actual)	191	2	99.00%
Malware (Actual)	4	149	97.40%
Overall Accuracy			98.20%
Importance variables from Hardware			
Benign (Actual)	184	9	95.30%
Malware (Actual)	9	144	94.30%
Overall Accuracy			94.80%

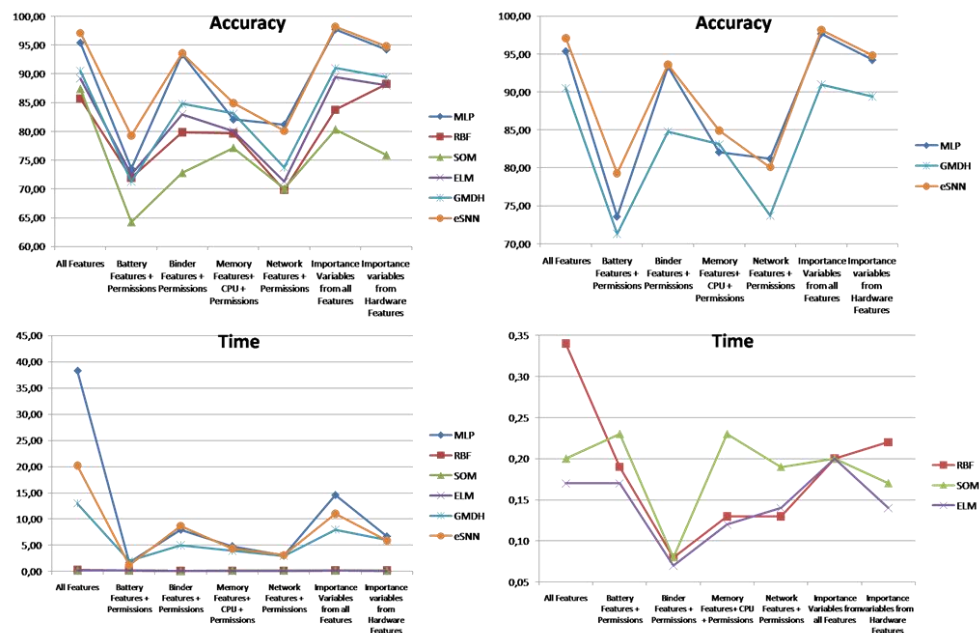


Fig. 4. Accuracy and time Comparison of MLP, RBF, ELM, GMDH PNN and eSNN

5. Discussion - Conclusions

An innovative bio-inspired Hybrid Intelligent Method for Detecting Android Malware (HIMDAM) has been introduced in this paper. It performs classification by using ELM (a very fast approach to properly label malicious executables) and eSNN for the detection of malware with high accuracy and generalization. An effort was made to achieve minimum computational power and resources. The classification performance of the ELM and the accuracy of the eSNN model were experimentally explored based on different datasets and reported promising results. Moreover the hybrid model detects the patterns and classifies them with high accuracy. In this way it adds a higher degree of integrity to the rest of the security infrastructure of Android Operating System. As a future direction, aiming to improve the efficiency of biologically realistic ANN for pattern recognition, it would be important to evaluate the eSNN model with ROC analysis and to perform feature minimization in order to achieve minimum processing time. Other coding schemes could be explored and compared on the same security task. Also what is really interesting is a scalability of ELM with other kernels in parallel and distributed computing in a real-time system. Finally the HIMDAM could be improved towards a better online learning with self-modified parameter values.

References

- [1] Fedler R., Banse C., Krauß Ch., Fusenig V.: Android OS Security: Risks and Limitations A Practical Evaluation, AISEC Technical Reports, AISEC-TR-2012-001, (2012).
- [2] Scandariato R., Walden J.: Predicting vulnerable classes in an android application, (2012).
- [3] Shabtai A., Fledel Y., Elovici Y.: Automated static code analysis for classifying android applications using machine learning, in CIS, Conference on. IEEE, 2010, pp. 329–333 (2010).
- [4] Chin E., Felt A., Greenwood K., Wagner D.: Analyzing inter-application communication in android, 9th conf. on Mobile systems, applications, and services. ACM, pp. 239–252 (2011).
- [5] Burguera I., Zurutuza U., Nadjm-Tehrani S.: Crowddroid: behavior-based malware detection system for android, 1st ACM workshop on on SPSM. ACM, pp. 15–26 (2011).
- [6] Glodek W., Harang R. R.: Permissions-based Detection and Analysis of Mobile Malware Using Random Decision Forests, IEEE Military Communications Conference, (2013).
- [7] Enck W., Ongtang M., McDaniel P.: On Lightweight Mobile Phone Application Certification, in: Proceedings of the 16th ACM Conf on Computer Security, CSS, (2009).
- [8] Lange M., Liebergeld S., Lackorzynski A., Peter M.: L4Android: A Generic Operating System Framework for Secure Smartphones, ACM Workshop on SPSM, (2011).
- [9] Portokalidis G., Homburg P., Anagnostakis K., Bos H.: Paranoid Android: Versatile Protection For Smartphones, 26th Annual Computer Security Applications Conference, (2010).
- [10] Ghorbanzadeh M., Chen Y., Zhongmin M., Clancy C. T., McGwier R.: A Neural Network Approach to Category Validation of Android Applications, Inter. Conference on Computing, Networking and Communications, Cognitive Computing and Networking Symposium (2013).
- [11] MacQueen J.: Some Methods for Classification and Analysis of Multivariate Observations, in Proceedings of the 5th Berkeley Symp. on Mathematical Statistics and Probability, (1967).
- [12] Fuchs A., Chaudhuri A., Foster J.: ScanDroid: Automated Security Certification of Android Applications, Technical report, University of Maryland, (2009).
- [13] WALA: <http://wala.sourceforge.net/wiki/index.php>.

- [14] Barrera D., Kayacik H., Oorshot P., Somayaji A.: A Methodology for Empirical Analysis of Permission-Based Security Models and its Application to Android, ACM, (2010).
- [15] Kohonen T.: Self-Organizing Networks, in Proceedings of the IEEE, (1990).
- [16] Tesauro G. J., Kephart J. O., Sorkin G. B.: Neural networks for computer virus recognition. *IEEE Expert*, 11(4):5–6, (1996).
- [17] Schultz M. G., Eskin E., Zadok E., Stolfo S. J.: Data mining methods for detection of new malicious executables, SP '01, pages 38–, Washington, DC, IEEE Computer Society (2001).
- [18] Sahs J., Khan L.: A Machine Learning Approach to Android Malware Detection, European Intelligence and Security Informatics Conference (2012).
- [19] Kolter J.Z., Maloof M.A.: Learning to detect malicious executables in the wild, International Conference on Knowledge Discovery and Data Mining, pp. 470–478 (2006).
- [20] Shabtai A., Fledel Y., Elovici Y.: Automated Static Code Analysis for Classifying Android Applications Using Machine Learning, IC Computational Intelligence and Security (2010).
- [21] Cheng C., Peng W. T., Huang G.-B.: Extreme Learning Machines for Intrusion Detection, WCCI IEEE World Congress on Computational Intelligence Brisbane, Australia (2012).
- [22] Joseph J.F.C., Lee B.-S., Das A., Seet B.-C.: Cross-Layer Detection of Sinking Behavior in Wireless Ad Hoc Networks Using ELM and FDA, *IEEE IJCA Vol 54– No.14*, (2012).
- [23] <http://extreme-learning-machines.org/>
- [24] Cambria E., Huang G.-B.: Extreme Learning Machines, *IEEE Intelligent Systems*, (2013).
- [25] Schliebs S., Kasabov N.: Evolving spiking neural network—a survey, *Evolving Systems*, June 2013, Volume 4, Issue 2, pp 87-98, Springer, (2013).
- [26] Thorpe S. J., Delorme A., Ruffin van Rullen: Spike-based strategies for rapid processing, *Neural Networks*, 14(6-7):715–725, (2001).
- [27] Delorme A., Perrinet L. & Thorpe S. J.: Networks of Integrate-and-Fire Neurons using Rank Order Coding B: Spike Timing Dependant Plasticity and Emergence of Orientation Selectivity, Published in *Neurocomputing*, 38-40(1-4), 539-545, (2000).
- [28] Thorpe S. J. and Gautrais J.: Rank order coding, In *CNS '97: 6th conf on Computational neuroscience: trends in research*, pages 113–118, Plenum Press (1998).
- [29] Kasabov, N.: *Evolving connectionist systems: Methods and Applications in Bioinformatics, Brain study and intelligent machines*, Springer Verlag, NY, (2002).
- [30] Wysoski S. G., Benuskova L., Kasabov N. K.: Adaptive learning procedure for a network of spiking neurons and visual pattern recognition. In *Advanced Concepts for Intelligent Vision Systems*, pages 1133–1142, Berlin/Heidelberg, Springer, (2006).
- [31] Schliebs S., Defoin-Platel M., Kasabov N.: Integrated feature and parameter optimization for an evolving spiking neural network, 5506, 2009, pages 1229 - 1236, Springer, (2009).
- [32] Huang G.-B.: An Insight into Extreme Learning Machines: Random Neurons, Random Features and Kernels, DOI 10.1007/s12559-014-9255-2, Springer (2014).
- [33] B. Amos. “Antimalware”. <https://github.com/VT-Magnum-Research/antimalware>, (2013).
- [34] Google. UI/Application Exerciser Monkey, <http://developer.android.com/tools/help/monkey.html>, (2013).
- [35] Alam M. S., Vuong S. T.: Random Forest Classification for Detecting Android Malware, *IEEE IC on Green Computing and Communications and Internet of Things* (2013).
- [36] <http://www.cs.waikato.ac.nz/ml/weka>
- [37] <http://www.gmdhshell.com/>

Predicting the impact of advertisements on web pages aesthetic impressions

Gianni Fenu, Gianfranco Fadda, and Lucio Davide Spano¹

University of Cagliari, Department of Mathematics and Computer Science
Via Ospedale 72, 09124, Cagliari, Italy
{fenu, gianfranco.fadda, davide.spano}@unica.it

Abstract. In this paper we study the impact of advertisements on a predictive model for web pages impressions, calculated according to a set of features on the rendered page images. We analysed the effects on two different predictors, the colourfulness and the visual complexity. We compared the prediction against ground-truth values, obtained through a user study. We conclude that the prediction model behaves correctly for the complexity, but it is not able to predict the increase on the colourfulness ratings.

Keywords: Online advertising, Predictive models, Visual features

1 Introduction

Users make their opinion on websites they visit for the first time after a few seconds [6]. This is a well known fact in the HCI community. However, we are also aware that commercial websites profit including advertisements, and their inclusion may change the first-look impression of a user. Even if they skip the content included in banners and text advertisements, their presence decreases the user performance while searching information or while interacting with the page, increasing the cognitive workload [2].

In this paper, we focus on predictive models for the aesthetic impression of web pages and we analyse the effects of advertisements on both the predicted and the perceived values. To achieve this goal, we exploited the prediction model proposed in [8]. We subsequently prepared a dataset including websites of different categories (e.g. news, e-commerce, business, etc.). For each website, we included in our experiment two different versions: one with and one without advertisements. In the experiment we compared the responses of the predictive model with the ground-truth results obtained through the user test.

Finally, we summarize the results and we discuss the possible development of this work.

2 Related Work

The user's opinion on an interface is influenced not only by its usability, but also by the aesthetic impression: the perceived usability is affected by the visual

appeal of the interface [5]. This is particularly important on the web, where it is really easy for users to switch between different sources. If a homepage does not impress a user, she will leave the page after a short time, without even evaluate the usability of the website [6].

Considering the fact that the first impression is built on the first 50-500 ms [6] of a webpage visit, different work focused on predicting such impression through image-related metrics. We can recall here the work by Zheng et al. [10], which shows how different low-level image statistics correlate with the user's ratings. In this paper, we exploit the work in [8] that has been validated with a large number of users. We provide a summary of their results in the next section.

We focus in particular on the advertisement impact on the aesthetic perception of a web page. Owens et al. in [7] studied the so-called "text advertisement blindness", which is the ability of users to ignore the advertisement content inside web pages, even for textual information and not only for banners. However, even if users are able to skip such contents, it has been demonstrated that their cognitive load increases when advertisements are included in a web page [2].

The visual properties that are correlated with the effectiveness of an advertisement include the spatial position and its visual appearance [1], which have an impact on with the overall aesthetic judgement of a website. In this work, we consider two different measures that may be affected by advertisements: the perceived colourfulness and the visual complexity.

3 Advertisement impact on rating prediction

In this paper, we exploit the predictive model described in [8], which considers two different factors for predicting the visual appeal: the *colourfulness* and the *visual complexity*. In order to obtain the ground-truth data on the users' perception, the authors selected a set of 450 websites and they proposed a randomly selected sample of 30 website to 242 volunteers, which rated the colourfulness and the visual complexity with a 9-point Likert scale (1 is the lower and 9 is the higher value).

These ground-truth values have been used in order to create a regression model for predicting the ratings according to different image metrics related to colors (e.g. percentage of pixels close to the sixteen W3C colors, the average pixel value in the HSV space, and the two definitions of colourfulness in [9] and [4]), quadtree decomposition [10] and space-based decomposition [3]). The two regression models are reported in Table 1. Reinecke et al. demonstrate that both complexity and colourfulness can be used for predicting the mean aesthetic user rating for a website. Additional details on both the models and the metrics are available in [8].

In this study, we analyse the impact of the advertisements on the prediction of both colourfulness and visual complexity of websites. We are interested in understanding if there is a statistically relevant difference between the rating of the website without advertisements and the version including them.

Colourfulness	<i>b</i>	<i>SE_b</i>	<i>p</i> >
Constant	-.68	.73	.05
Gray	2.45	.56	.001
White	2.74	.71	.001
Maroon	1.60	.71	.05
Green	1.85	.57	.01
Lime	-3.17	1.42	.05
Blue	-3.91	1.06	.001
Teal	1.01	.56	.05
Saturation	.01	.002	.01
Colorfulness [4]	.03	.003	.001
Image areas	.06	.008	.001
Quadtree leaves	3.74E-4	.000	.001
Text area	-1.48E-6	.000	.05
Non text area	1.86E-6	.000	.001

Complexity	<i>b</i>	<i>SE_b</i>	<i>p</i> >
Constant	.637	.179	.001
Text area	.000	.000	.001
Non text areas	.000	.000	.001
Number of leaves	.005	.002	.05
Number of text groups	.052	.008	.001
Number of image areas	.056	.012	.001
Colorfulness [9]	.011	.005	.05
Hue	.005	.002	.05

Table 1. Regression model for the perceived colourfulness ($r = .88$, $p < .001$, $R^2 = .78$, $R^2_{adj} = .77$) and visual complexity ($r = .80$, $p < .001$, $R^2 = .65$, $R^2_{adj} = .64$) in [8].

In order to measure this difference, we created a dataset of 50 websites belonging to different categories (news, weather, sports etc.). We selected the websites considering a list of the top 500 accessed websites¹. We did not consider websites without advertisements (e.g. institutional websites). We obtained a set of screenshots of the websites interfaces through a screen capture tool. After that, we manually removed the advertisements in the picture, replacing them with the website background color, pattern or image. Figure 1 shows two versions of the same website in the dataset with and without advertisements.

For each image in the dataset, we calculated the predicted values for the colourfulness and the visual complexity according to the regression models in Table 1. Then, we compared the ratings through a paired t-Test, in order to establish if the presence or the absence of advertisements introduced a significant difference on the predicted values.

For the colourfulness we registered a not significant difference between the values predicted for the websites with and without advertisements ($p = 0.77$). The 95% confidence interval for the difference is $\bar{x} = 0.18 \pm 1.25$ and the rating

¹ <http://www.alexa.com/topsites>



Fig. 1. Sample website with (top part) and without (bottom part) advertisements

is expressed in 1 to 9 Likert scale. Therefore, introducing the advertisements on websites does not produce effects on the predicted colourfulness.

Instead, for the visual complexity we registered a significant difference between the versions ($p = 10^{-7}$). Analysing the 95% confidence interval around the difference mean, we found that removing the advertisements we had a decrease of one point (out of 9) in the predicted complexity ($\bar{x} = 1.0 \pm 0.27$).

4 Advertisement impact on user ratings

In this section, we report on a user study we performed for establishing whether the effect of advertisements on the predicted ratings arises also for the perceived colourfulness and visual complexity.

In order to collect the ground-truth data for assessing the effects of the advertisements, we created a web application. The first presentation requires

the user to enter some demographic data, such as age, education level and the experience with web browsing. After providing this information, the application shows 40 screenshots of 20 different websites, presenting both the version with and without advertisements in a random order. The user rates the perceived colourfulness and complexity in a 1 to 9 Likert scale.

Fifty users participated to the test, 35 male and 15 female. They belonged to different age ranges: 2 are less than 15 years old, 10 between 16 and 25, 13 between 26 and 35, 14 between 36 and 45 and 11 are older than 45. The education level varied from middle school (4), high school (11), bachelor degree (17), master degree (16) and PhD (2).

The participants were recruited using social networks and emails. The test was performed using a remote setting. Once we collected the data, we performed a two way ANOVA for repeated measures analysis, in order to study the effect of the website version (with or without advertisements) on the colourfulness and the visual complexity ratings. In the analysis we consider two different factors that influence the rating outcome: the website and the version.

4.1 Colourfulness

We start from the analysis of the colourfulness ratings, which is summarised in Table 2. In this analysis we considered as factors the evaluated website (*Website*) and its configuration with or without advertisements (*Version*). We repeated the measures for each user all configurations for the two factors.

The Mauchly's test revealed that the sphericity assumption had been violated for the website factor ($\chi^2(189) = 6.33 \cdot 10^{-6}$, $p < 10^{-16}$) and for the interaction between the version and the website ($\chi^2(189) = 0.00136885$, $p = 3.2 \cdot 10^{-5}$). Therefore, we corrected the degrees of freedom applying the Greenhouse-Geisser estimation of sphericity, respectively with $\varepsilon_w = 0.3401$ and $\varepsilon_{v,w} = 0.61977$.

Table 2 shows an obvious significant effect of the website on the perceived colourfulness. It is more interesting to notice that both the version and the interaction between the version and the website have a significant impact on the users' ratings.

Even if the interaction is significant, the post-hoc analysis with the Tukey's HSD test revealed that the difference size is not meaningful for most values. The analysis is summarized in Figure 2: the upper part contains the differences for the pairs having the version with advertisements, while the lower parts shows

Colorfulness	SS	df(SS)	Error	df(Error)	F	p-value
Version	273	1	207.81	49	64.3842	10^{-10}
Website	1131	6.46779	2626.73	316.92171	21.1018	$< 10^{-16}$
Version * Website	121	11.77563	1057.22	577.00587	5.6271	10^{-9}

Table 2. ANOVA table for colourfulness ratings

the pairs without advertisements. As it is possible to see, most of them contain 0 inside the confidence interval.

With respect to the version factor, the Tukey's HSD test revealed that the ratings of websites with and without advertisements differ significantly ($p < 10^{-16}$). On average, the difference between the version with and without advertisement is 0.739, and the 95% confidence interval is [0.57339; 0.90461]. In a 1 to 9 Likert scale such difference is meaningful and, contrary to the predicted values, it demonstrates that users do perceive a colourfulness difference between the two versions.

This can be explained with the selective user's focus on a web page content: advertisements are perceived as an "added" content that does not really belong to the page, thus their presence always add colors to visual representation. The prediction instead is based on image features and, if the advertisements exploit roughly the same colors, their presence may not affect the measurement. Therefore, our suggestion is to remove the advertisements from the website image for predicting the colourfulness perception and to increase the value by 0.5 points.

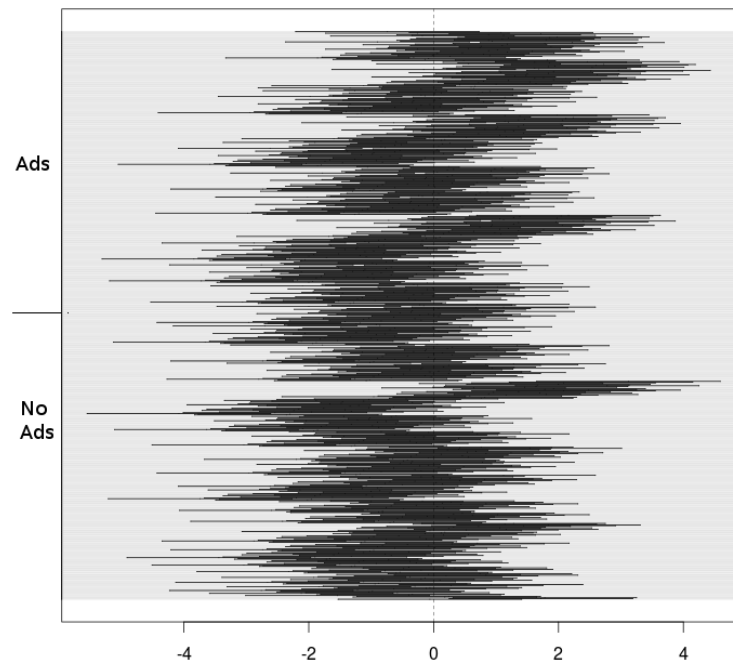


Fig. 2. 95% family-wise confidence level for colorfulness means comparison

4.2 Complexity

The analysis of the complexity ratings is summarized in Table 3, considering the same factors and conditions with respect to the previous experiment. The Mauchly's test revealed again the sphericity assumption had been violated for the website factor ($\chi^2(189) = 1.38 \cdot 10^{-4}$, $p < 10^{-16}$) and for the interaction between the version and the website ($\chi^2(189) = 9.93 \cdot 10^{-4}$, $p = 2.9 \cdot 10^{-6}$) and we corrected the degrees of freedom with Greenhouse-Geisser ($\varepsilon_w = 0.46016$, $\varepsilon_{v,w} = 0.60207$).

Table 3 reveals again the obvious significant effect of the website factor for the complexity rating. In addition, it reveals that the interaction between the website and the version is not significant, therefore our data can be represented with an additive model.

Complexity	SS	df(SS)	Error	df(Error)	F	p-value
Version	99	1	166.4	49	29.4154	10^{-6}
Website	361.6	8.74304	2932.2	428.40896	6.0434	10^{-14}
Version * Website	41.4	11.43933	1382.7	560.52717	1.4680	0.08873

Table 3. ANOVA table for complexity ratings

The most interesting part of this analysis is the significance of the version factor ($p < 10^{-6}$). In order to evaluate the difference in the ratings between the version with and without advertisements, we performed a post hoc Tukey's HSD test. It revealed that, on average, websites with advertisements are almost half point (0.447 in a 1 to 9 Likert scale) more complex than those without advertisements. The difference is contained in a $[0.26316; 0.63084]$ interval ($p = 2 \cdot 10^{-6}$).

Therefore, the study confirms that the predicted increase of complexity is perceived by users, even if it has been overestimated. In this case, we suggest to use website images that contain the advertisements in order to predict an upper bound for the perceived complexity of the page.

5 Conclusions and Future Work

In this paper, we analysed the impact of advertisements on predicting the user's rating of both complexity and colourfulness of web pages, which have been demonstrated correlated with the first aesthetic impression of the user. In order to perform such analysis, we selected the model in [8], which is able to predict the user's rating according to a set of metrics calculated on the website image. We found that the presence of advertisements does not change significantly the colourfulness prediction in the model discussed. Instead, it does have an impact

on the predicted complexity value: on average, websites with advertisements are one point more complex in a 1 to 9 Likert scale.

We conducted a user study for analyse the effects on real user ratings. We found a significant increase on the perception of both colourfulness and complexity ratings. While the latter is replicated also by the prediction model, the first one is not replicated in the prediction model. Therefore, we suggest to evaluate the colourfulness of the web page contents rather than considering also the advertisements for predicting the user's aesthetic judgement.

In the future, we aim to create a tool that, through the prediction based on the regression model, may help designers in selecting advertisements without increasing the perceived complexity. In addition, the tool may help them in comparing different positions and layouts for including advertisements in a web page.

Moreover, we plan to further investigate the relationship between colourfulness and advertisements. Our goal is to obtain a prediction model that is able to correctly evaluate the inclusion impact of contents that are not directly related with the page topic.

Acknowledgements

We gratefully acknowledge Sardinia Regional Government for the financial support (P.O.R. Sardegna F.S.E. Operational Programme of the Autonomous Region of Sardinia, European Social Fund 2007-2013 - Axis IV Human Resources, Objective 1.3, Line of Activity 1.3.1 "Avviso di chiamata per il finanziamento di Assegni di Ricerca").

References

1. Azimi, J., Zhang, R., Zhou, Y., Navalpakkam, V., Mao, J., Fern, X.: The impact of visual appearance on user response in online display advertising. In: Proceedings of the 21st International Conference Companion on World Wide Web. pp. 457–458. WWW '12 Companion, ACM, New York, NY, USA (2012), <http://doi.acm.org/10.1145/2187980.2188075>
2. Burke, M., Hornof, A., Nilsen, E., Gorman, N.: High-cost banner blindness: Ads increase perceived workload, hinder visual search, and are forgotten. *ACM Transactions on Computer-Human Interaction (TOCHI)* 12(4), 423–445 (2005)
3. Ha, J., Haralick, R., Phillips, I.: Recursive x-y cut using bounding boxes of connected components. In: Document Analysis and Recognition, 1995., Proceedings of the Third International Conference on. vol. 2, pp. 952–955 vol.2 (Aug 1995)
4. Hasler, D., Suesstrunk, S.E.: Measuring colorfulness in natural images. vol. 5007, pp. 87–95 (2003)
5. Lindgaard, G., Dudek, C., Sen, D., Sumegi, L., Noonan, P.: An exploration of relations between visual appeal, trustworthiness and perceived usability of homepages. *ACM Trans. Comput.-Hum. Interact.* 18(1), 1:1–1:30 (May 2011), <http://doi.acm.org/10.1145/1959022.1959023>

6. Lindgaard, G., Fernandes, G., Dudek, C., Brown, J.: Attention web designers: You have 50 milliseconds to make a good first impression! *Behaviour & Information Technology* 25(2), 115–126 (2006), <http://dx.doi.org/10.1080/01449290500330448>
7. Owens, J.W., Chaparro, B.S., Palmer, E.M.: Text advertising blindness: The new banner blindness? *J. Usability Studies* 6(3), 12:172–12:197 (May 2011), <http://dl.acm.org/citation.cfm?id=2007456.2007460>
8. Reinecke, K., Yeh, T., Miratrix, L., Mardiko, R., Zhao, Y., Liu, J., Gajos, K.Z.: Predicting users' first impressions of website aesthetics with a quantification of perceived visual complexity and colorfulness. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. pp. 2049–2058. CHI '13, ACM, New York, NY, USA (2013), <http://doi.acm.org/10.1145/2470654.2481281>
9. Yendrikhovskij, S., Blommaert, F., De Ridder, H.: Optimizing color reproduction of natural images. In: *Color and Imaging Conference*. vol. 1998, pp. 140–145. Society for Imaging Science and Technology (1998)
10. Zheng, X.S., Chakraborty, I., Lin, J.J.W., Rauschenberger, R.: Correlating low-level image statistics with users - rapid aesthetic and affective judgments of web pages. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. pp. 1–10. CHI '09, ACM, New York, NY, USA (2009), <http://doi.acm.org/10.1145/1518701.1518703>

A Process Model of Large-scale Open Idea Generation in Crowdsourcing Marketing

Toshihiko Yamakami

ACCESS, Marketing Strategy Office,
1-10-2 Nakase, Mihama-ku, Chiba-shi, JAPAN 261-0023
<http://www.access-company.com/>

Abstract. Advances in information and communication technology have impacted many businesses. Searching, comparison, and matching are domains where information technology has increased its influence. A variety of resources can be managed by technology in order to produce a deliverable that is impossible to achieve without information technology. Crowdsourcing is one such domain. Since marketing is a crucial part of any business, it is important to leverage crowdsourcing marketing in the era of social services. Large-scale open idea generation can play an important role in crowdsourcing marketing. The author discusses the requirements of large-scale open idea generation. Then, the author proposes a process view model of large-scale open idea generation. Finally, the author presents the design for large-scale open idea generation support system using the proposed process model.

1 Introduction

At the rise of the Internet, it was just an information highway. Therefore, there was little human aspect involved in the Internet in the early days. People were fascinated by the availability of global information it provided. As the capacity of communication was leveraged, the Internet started to cover more human aspects and a variety of roles in information consumption.

One of the landmarks was Web 2.0, in which a wide range of people were involved in the production of information thanks to the increased connectivity of the Internet. The massive number of participants constructed a new way of managing information and led to many new services such as Twitter, Facebook, Amazon (user reviews) and so on. This trend added a new spectrum to the information exchange on the Internet. In terms of games, massively multi-player online games became popular within decades. Early massively multi-player online games dated back to multi-user dungeon games in the 1980's.

This extension in terms of the number of participants further led to more complicated collaboration on the Internet. Crowdsourcing is one such extension. And Web 2.0 accelerated this trend. In 2005, Amazon released Amazon Mechanical Turk, which is a Web service to incorporate human-intelligent tasks. It

2 Yamakami

provides an API that splits tasks into small pieces that a human will perform. This is an on-demand, flexible, and scalable way to incorporate the human intelligence that is available on the Internet into a business process.

The author extends this crowdsourcing model to marketing. Crowdsourcing marketing is a relatively new concept. In order to support crowdsourcing marketing, the author discusses the design of the large-scale open idea generation. The author discusses the requirements of large-scale open idea generation support systems. Then, the author presents the design for a large-scale open idea generation support system.

2 Background

2.1 Purpose of Research

The aim of this research is to identify a design for a large-scale open idea generation support system.

2.2 Related Work

The research of crowdsourcing consists of four domains: (1) validity and techniques, (2) social and regional studies, (3) applications, and (4) adjacent works.

First, there is research on the validity of and techniques for crowdsourcing. Hipp et al. showed the validity of crowdsourcing in annotating human behaviors using Amazon Mechanical Turk [7]. Nauman discussed the validity of creating new, integrated data from openly available public data [8]. Barbier et al. discussed the mining of crowdsourcing in order to leverage crowdsourcing [2]. Wang et al. discussed the utilization of human power to facilitate joins in databases [14].

Second, there are regional characteristics and the social network analysis of crowdsourcing. Stanley et al. discussed rural crowdsourcing development in Namibia [10]. Ebdon discussed the network analysis of crowdsourcing [5].

Third, there are studies on the applications of crowdsourcing. Robson discussed crowdsourcing for citizen science that is performed by non-professional scientists [9]. Tan et al. presented BeFaced, a tile matching game in order to crowdsource the maintenance of quality in an image database [11]. Burnett et al. discussed crowd curation in music preferences [3]. Celino et al. discussed an Android app for collecting urban information with a game interface [4]. Vondrick et al. presented their 3-year study on crowdsourced video annotation [13]. Varriale et al. discussed the time-delay reporting system using a crowdsourced technique [12]. Absalom et al. discussed Wiki-based classification using crowdsourcing [1].

Fourth, there are adjacent research domains to crowdsourcing. Heimerl et al. discussed community sourcing using a specific population, an alternative to crowdsourcing [6].

Creative activity support for crowdsourcing is still a challenge and not well covered in the past literature.

The originality of this paper lies in its identification of a framework for massively multiuser idea generation as part of crowdsourcing marketing.

3 Method

The author's method consists of the following steps:

- Identifying asynchronous, massively multi-user idea generation process,
- Identifying requirements for each stage of the process, and
- Identifying functions for each stage of the process.

Crowdsourcing is a new terminology that has emerged with advances in information and communication technology. Definitions are listed in Table 1.

Table 1. Definitions.

Term	Definition
Crowdsourcing	A term that combines “crowd” and sourcing. Outsourcing to open general public on a massive scale.
Crowdsourcing marketing	A term that combines “crowd sourcing” and marketing. Outsourcing a marketing job in a crowd sourcing manner. Applying crowdsourcing to marketing activities.
Large-scale open idea generation	Idea generation with a number of people where the number of members is beyond each member's cognitive capacity for creative work.

Large-scale open idea generation provides a supporting role in crowdsourcing marketing. In this paper, the assumption in terms of number of members is beyond 20. For example, “large-scale open” means more than 20 members without interpersonal recognition and synchronization for idea generation. A number that exceeds 20 means that it is hard to recognize and anticipate all members' current activities and future reactions.

The two types of crowdsourcing marketing are depicted in Table 2. The former

Table 2. Two types of crowdsourcing marketing.

Term	Definition
Mechanical crowdsourcing marketing	Split marketing activities into a large number of simple tasks and delegate it to the general public. For example, sending a marketing email can be delegated to the general public. Each participant simply repeats sending the same email to his/her friends.
Creative crowdsourcing marketing	Delegate creative marketing activity to the general public. For example, constructing a creative marketing plan.

type of mechanical crowdsourcing marketing can be handled by common crowdsourcing frameworks. In this paper, the author deals with the latter, creative crowdsourcing marketing.

4 Yamakami

4 Function Design

4.1 Process Model

Large-scale open idea generation as a sub-activity of creative crowdsource marketing needs to address collaborative processes including synchronization and harmonization. The requirements of large-scale open idea generation are depicted in Table 3.

Table 3. Requirements of large-scale open idea generation.

Requirement	Description
Management participation	of Registration and sharing of participants.
Status sharing	Visualization of idea generation process.
Encouragement	Rewards for contributing ideas and creating social ties for facilitating creative processes.
Management creativity	of Idea-generation-specific management functions that facilitate massively parallel creative processes.

It is important to capture and facilitate the dynamic part of idea generation, in which the number of members plays an important role. All of these requirements: management, status sharing, encouragement, and management of creativity include the challenges of dynamism that derive from the large number of members. When there are social ties with strong interpersonal cognition, many issues in these requirements are implicitly solved by the social interaction. Large-scale open idea generation requires technology-augmented facilitation due to the lack of such social interaction-based facilitation.

A comparison of closed collaboration within size limits and open collaboration beyond certain size limits is depicted in Table 4.

Table 4. Comparison of closed collaboration and open collaboration beyond size limits.

Type	Description
Closed collaboration within size limits	Awareness, expectation, shared goals and norms are managed by social cognition and interaction.
Open collaboration beyond size limits	Awareness, expectation, shared goals and norms are difficult due to the cognitive capacity of each member. Openness of unknown members and unknown contributions provides both opportunities and challenges of collaboration.

This contrast between open and closed collaboration are universal; beyond the application of idea generation.

Table 5. Issues of large-scale open idea generation

Issue	Description
Massive multi-player collaboration	The complexity of collaboration explodes when the number of participants grows.
Anonymous collaboration	The complexity of collaboration grows when there are many anonymous members without much socialization among them.
Asynchronous idea generation	Usually, idea generation is done in a synchronous manner. Asynchronous idea generation exacerbates social conflicts, gaps in expectations, and other collaboration issues.
Gaps in expectations	Massively parallel collaboration needs to address gaps in expectation among members at each stage.
Social conflict resolution	When the number of participants grows, the possibility of social conflict also grows.
Convergence of massive number of ideas	Even when a small number of members is involved, the growth of the number of generated ideas adds complexity in the idea convergence process.

This open collaboration adds burdens in the case of idea generation. The issues of large-scale open idea generation are depicted in Table 5.

These issues are rooted by the above-mentioned dynamism that is beyond each member's cognitive capacity to understand other members' collaboration.

Even with these challenges, large-scale open idea generation has advantages over the conventional marketing activities. The advantages are depicted in Table 6.

Table 6. Advantages of crowdsourcing marketing

Advantage	Description
Time-free and space-free marketing	Participation is not constrained by time or place.
Massive participation	The number of participants is not constrained.
Meaning of collaboration	People feel we-ness in social activities. People feel meaning when they form something larger than themselves.
Surprise	Something beyond expectation provides surprise and joy during the collaboration process.

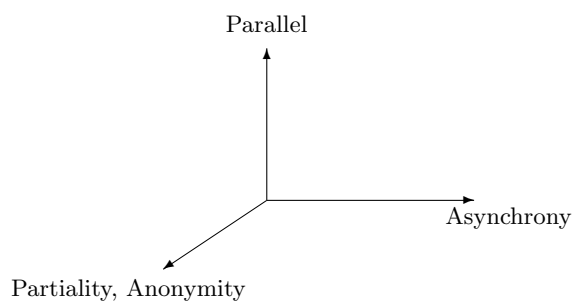
Three dimensions of collaborative characteristics are depicted in Fig. 1.

These dimensions represent the different types of impact from the challenge of open collaboration beyond size limits.

In order to design a large-scale open idea generation system, the author examines a process view model that represents the time-dimensional aspect of multiplayer idea generation.

A process view is depicted in Fig. 2. The time-dimensional view is impor-

6 Yamakami



Dimension	Description
Asynchrony	Crowdsourcing marketing facilitates asynchrony in marketing collaboration.
Parallel	Crowdsourcing marketing facilitates massive large-scale parallelism in collaboration.
Partiality, Anonymity	Crowdsourcing marketing facilitates partial collaboration with anonymity. It is difficult to collaborate in a non-full member manner.

Fig. 1. Dimensions of collaborative characteristics

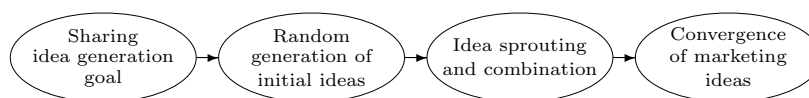


Fig. 2. Process view model of crowdsourcing marketing

tant because the time dimensional factor plays an important role when social cognition and interaction is not enough to manage the entire process of idea generation as it exists in large-scale open idea generation.

4.2 Functional Design

In order to provide a design framework, the author presents the functions of a large-scale open idea generation support system. The author adopts the stage-by-stage method for the functional design. The functions at each stage are depicted in Table 7.

Table 7. Functions of crowdsourcing marketing

Phase	Function	Description
Goal	Goal sharing	Setting a goal, time schedule, and links to social networks
	Sub goal sharing	Setting a sub goal (marketing channels, events, packaging)
Idea generation	Stimulus	Setting and showing a list of keywords, association words, figures, and episodes
	Linkage	Setting associations among generated ideas
	Storage	Storage, retrieval, and editing.
Idea sprouting	Combination	Setting a structure on a network of generated ideas.
	Recomposition	Setting a recomposition for a network of generated ideas.
Convergence	Filtering	Highlighting the sub-graph for a network of generated ideas.
	Structuring	Providing the structure for a network of generated ideas.

In order to facilitate these functions, it is important to facilitate the time-dimensional and emotional factors of open collaboration. A Four-stage model of emotion in crowdsourcing marketing is depicted in Fig. 3.

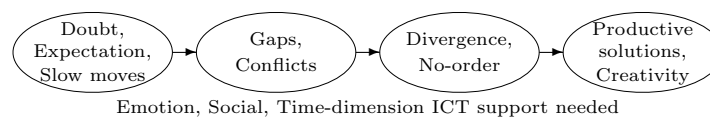


Fig. 3. Four-stage model of emotion in crowdsourcing marketing

8 Yamakami

This model of emotion provides the building blocks to deal with the time-dimensional emotional factors that are caused by large-scale open collaboration.

Above is the one-directional process model. A process model can be constructed with the dimensions of the number of generated ideas and the number of participants. A two-dimensional view model of crowdsource marketing is depicted in Fig. 4.

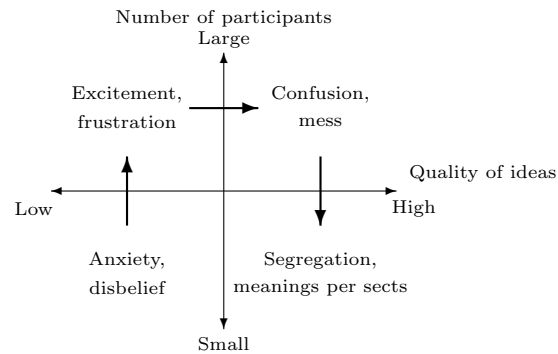


Fig. 4. Two-dimensional view model of large-scale open idea generation

In this view model, the arrows can be reversed depending upon the dynamic nature of idea generation and participation. A reverse clock wise movement from the bottom left corner to the bottom right corner is the case of successful large-scale idea generation.

In order to deal with dynamic contexts and social emotions at each stage, the features need to be designed stage-by-stage. The features that deal with emotion in creative crowdsource marketing are depicted in Table 8.

These features deal with the time dimensional phases in large-scale open collaboration.

5 Discussion

5.1 Advantages of the Proposed Method

Crowdsource marketing requires dealing with issues that are beyond the traditional closed synchronous collaboration. In order to cope with this, the author proposes a three-dimensional view model of the issues in crowdsource marketing.

The author presents the emotions associated with these three dimensions.

It is necessary to facilitate the time-dimensional aspects of crowdsource marketing in order to overcome the challenges of these three dimensions. The author

Table 8. Features that deal with emotion in creative crowdsourcing marketing

Phase	Function	Description
Goal	Goal sharing	Visualization of goals, diffusion support, real-time feedback of goal status.
	Sub goal sharing	Visualization of sub-goals, real-time update of micro-activities, timelines of social activities.
Idea generation	Stimulus	Idea generation supported by words, images, and associations. Encouragement of initiatives. Support for feelings of participation.
	Linkage	Dynamic linkages, association and social timelines.
Idea sprouting	Combination	Structuralization support, splitting a pile of ideas, coloring of ideas, network analysis of ideas, social encouragement.
	Recomposition	Visualization of recomposition process, encouragement of initiatives and leadership.
Convergence	Filtering	Tagging, clustering, semantic labeling.
	Structuring	Semantic construction support, achievement feeling support.

presents a process view model for facilitating the time-dimensional factors in crowdsourcing marketing.

The size of a group plays a critical role in collaboration in terms of awareness, social ties, and expectations of behavior. The proposed framework deals with aspects of open collaboration: (1) massively multi-player involvement, (2) partial involvement, (3) anonymous involvement, and (4) asynchronous collaboration.

The issues of size boundaries are depicted in Table 9.

Table 9. Issues of size boundaries.

Boundary	Issue
1–2	Token-passing, exclusive control, interpersonal communication, social power, norm, trust, . . .
2–3	Multiple interpersonal communications, combined social power, social emotion.
5–6	Boundary of real-time synchronous collaboration.
Around 20	Boundary of real-time awareness of activities.
Around 200	Boundary of conscious social network.

The author provides a process view model that covers the time-dimensional factors of large-scale open idea generation. This is universal for large-scale open collaboration beyond the application of idea generation. Open collaboration is fundamentally open-ended and full of asynchrony, conflicts, uncertainty, and

10 Yamakami

dynamism. The process-view-model based approach is useful for developing a systematic design framework for open collaboration systems.

5.2 Limitations

This paper deals with only the design framework of large-scale idea generation in the process of crowdsourcing marketing. The quantitative verification of each functional feature is beyond the scope of this paper.

The author did not perform any real-world evaluation of the proposed framework.

Experience, Internet literacy and cultures can impact the behavior of crowdsourcing. The quantitative analysis of these impacting factors is beyond the scope of this paper.

Implementation and deployment of a crowdsourcing marketing service remains for future studies.

The effectiveness of crowdsourcing marketing is not evaluated in this paper. The user satisfaction of participants in crowdsourcing marketing is not covered in this paper.

This paper does not cover initiatives, leadership, roles, social emotions, social conflicts, norms and other detailed social contexts.

6 Conclusion

Advances in information and communication technology have brought new waves of work processes. Crowdsourcing is one such new wave. Crowdsourcing started at the mechanical crowdsourcing and is shifting toward creative crowdsourcing such as crowdsourcing marketing.

Crowdsourcing marketing includes a wide spectrum of challenges to collaborative work support and creative work support. The author analyzes the issues and advantages of large-scale open idea generation, as a part of crowdsourcing marketing, from the viewpoint of massively parallel asynchronous creative marketing. Then, the author examines a stage view model of large-scale open idea generation. The author presents the functions and features at each stage from the viewpoint of a process view model.

Large-scale open idea generation inherits many issues from large-scale open collaboration. The process model provides the building blocks for real-world large-scale open idea generation support systems.

In aging societies, the utilization of senior know-how is important. Large-scale open idea generation can leverage senior know-how in terms of incubation and regional economic growth beyond space and time boundaries.

References

1. Absalom, A., Absalom, G., Hartmann, D.: A collaborative classification-based search engine for prior art and solution search: Durham zoo. In: Proceedings of

- the 2012 IEEE Sixth International Conference on Semantic Computing. pp. 87–92. ICSC '12, IEEE Computer Society, Washington, DC, USA (2012)
2. Barbier, G., Zafarani, R., Gao, H., Fung, G., Liu, H.: Maximizing benefits from crowdsourced data. *Comput. Math. Organ. Theory* 18(3), 257–279 (Sep 2012)
3. Burnett, D., Lochrie, M., Coulton, P.: "checkindj" using check-ins to crowdsourcing music preferences. In: *Proceeding of the 16th International Academic MindTrek Conference*. pp. 51–54. MindTrek '12, ACM, New York, NY, USA (2012)
4. Celino, I., Cerizza, D., Contessa, S., Corubolo, M., Dell'Aglio, D., Valle, E.D., Fumeo, S.: Urbanopoly – a social and location-based game with a purpose to crowdsourcing your urban data. In: *Proceedings of the 2012 ASE/IEEE International Conference on Social Computing and 2012 ASE/IEEE International Conference on Privacy, Security, Risk and Trust*. pp. 910–913. SOCIALCOM-PASSAT '12, IEEE Computer Society, Washington, DC, USA (2012)
5. Ebdem, M., Huynh, T.D., Moreau, L., Ramchurn, S., Roberts, S.: Network analysis on provenance graphs from a crowdsourcing application. In: *Proceedings of the 4th International Conference on Provenance and Annotation of Data and Processes*. pp. 168–182. IPAW'12, Springer-Verlag, Berlin, Heidelberg (2012)
6. Heimerl, K., Gawalt, B., Chen, K., Parikh, T., Hartmann, B.: Communitysourcing: Engaging local crowds to perform expert work via physical kiosks. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. pp. 1539–1548. CHI '12, ACM, New York, NY, USA (2012)
7. Hipp, J.A., Adlakha, D., Gernes, R., Kargol, A., Pless, R.: Do you see what i see: Crowdsourcing annotation of captured scenes. In: *Proceedings of the 4th International SenseCam & Pervasive Imaging Conference*. pp. 24–25. SenseCam '13, ACM, New York, NY, USA (2013)
8. Nauman, F.: Dr. crowdsourcing: Or how i learned to stop worrying and love web data. In: *Proceedings of the 2nd International Workshop on Business intelligenceE and the WEB*. pp. 2–2. BEWEB '11, ACM, New York, NY, USA (2011)
9. Robson, C.: Using Mobile Technology and Social Networking to Crowdsourcing Citizen Science. Ph.D. thesis, Berkeley, CA, USA (2012), aAI3555887
10. Stanley, C., Winschiers-Theophilus, H., Onwordi, M., Kapuire, G.K.: Rural communities crowdsourcing technology development: A namibian expedition. In: *Proceedings of the Sixth International Conference on Information and Communications Technologies and Development: Notes - Volume 2*. pp. 155–158. ICTD '13, ACM, New York, NY, USA (2013)
11. Tan, C.T., Sapkota, H., Rosser, D.: Befaced: A casual game to crowdsourcing facial expressions in the wild. In: *CHI '14 Extended Abstracts on Human Factors in Computing Systems*. pp. 491–494. CHI EA '14, ACM, New York, NY, USA (2014)
12. Varriale, R., Ma, S., Wolfson, O.: Vtis: A volunteered travelers information system. In: *Proceedings of the Sixth ACM SIGSPATIAL International Workshop on Computational Transportation Science*. pp. 13:13–13:18. IWCTS '13, ACM, New York, NY, USA (2013)
13. Vondrick, C., Patterson, D., Ramanan, D.: Efficiently scaling up crowdsourced video annotation. *Int. J. Comput. Vision* 101(1), 184–204 (Jan 2013)
14. Wang, J., Li, G., Kraska, T., Franklin, M.J., Feng, J.: Leveraging transitive relations for crowdsourced joins. In: *Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data*. pp. 229–240. SIGMOD '13, ACM, New York, NY, USA (2013)

OntOSN - Towards an Integrated Ontology for the Analysis of Online Social Networks

Richard Braun and Werner Esswein

Technische Universität Dresden
Chair of Wirtschaftsinformatik, esp. Systems Development
01062 Dresden, Germany
`{richard.braun,werner.esswein}@tu-dresden.de`

Abstract. The analysis of online social networks (OSN) like Facebook is an emerging field of research and of high relevance for businesses in order to derive new knowledge about customers. Generally, analyzing OSN requires the participation of different stakeholders, for example domain experts, data mining experts and IT engineers. Each of these groups operates with specific modeling approaches what can cause communication problems, redundancy or model discrepancies. We proclaim the design of an integrated ontology for OSN analysis in order to avoid these problems. Therefore, a set of requirements on the aimed ontology is evolved based on the literature. Afterwards, relevant ontologies are examined against these requirements and a need for an integrated ontology is identified. Finally, the OntOSN ontology is designed and presented.

Key words: Ontology Engineering, Domain-specific Ontology, Social User Ontology, Online Social Networks, OSN

1 Introduction and Motivation

Social media in general and Online Social Networks (OSN) in particular became substantial technologies of the daily life within the last years ([1], [2]). OSN provide new opportunities for businesses in the field of customer relationship management (CRM), social business intelligence and social media analytics ([3], [4], [5], [6]). On the other hand, OSN can cause several threats and risks for business like reputational damages, data leakages or IT-related risks [7]. Due to a power shift in OSN, businesses are in a more reactive position against customers and need to monitor both OSN users and their interactions associated with the business. From an academic point of view, there are several topics for investigation like the analysis of social influence structures [8]. It can be stated, that OSN are in the scope of both businesses and academia for a range of reasons ([9], [2]) and the multi-perspective analysis of OSN is an emerging field of research within the information systems discipline ([10], [9]).

Multi-perspective analysis of OSN describes the situation where experts from different domains are interested in gaining information from OSN or to conduct OSN data for their specific purposes. Therefore, purpose-specific analysis of OSN

is necessary [6]. [11] proclaims a framework in order to systematize analysis of social media: First, domain experts initially create a demand for specific data or information. Generally, analysis purposes are explicated as informal statements. Occasionally, conceptual domain models outline analysis purposes (e.g., [12], [13]), but these models are semiformal, high-level and neither detailed or implementation-oriented. In contrast to that, OSN data tracking describes in detail, what kind of data is available and accessible in OSN and what kind of technologies can be used for tracking and parsing the data. This area is managed by IT engineers in order to implement algorithms and queries on API level or code level. Therefore, social web ontologies are prevalent in order to present common OSN concepts in specific applications and to adapt them in dedicated tracking tools. The theoretical background for the fulfillment of specific analysis purposes comes from analytical methods which constitute as formal models or graph models (e.g., [14], [8], [15]). As stated, each area has distinct models for the description of the relevant context in OSN. The integration of different areas is challenging due to the heterogeneity of description languages and different levels of abstraction, formalization and specific domain concepts.

Although, OSN analysis is an emerging research topic, this problem is not explored so far. The following works at least touch the issue or state its relevance. [6] emphasizes the need for common technical frameworks, [10] and [4] see the need of integrated data models, [9] states the relevance of OSN conceptualization in general, [6] addresses problems in business integration, [16] criticizes the missing integration of social graph and social text data and both [17] and [18] see the need for a multi-perspective social user model. We suppose, that the insufficient consideration in research is not an evidence for missing relevance, but rather based on the lack in terms of OSN analysis frameworks, that leads to missing awareness of the issue. Based on the stated research gap, we aim to provide an ontology that integrates concepts from business-driven analysis purposes and the data tracking area (with respect to the space of this paper, the ontology design is limited to these areas). We argue, that an appropriate ontology has the potential to both close the gap between different modeling approaches and to facilitate the representation of knowledge regarding the analysis of OSN. Thus, a better understanding and a better analysis of OSN are promised.

Therefore, we first introduce some fundamentals from the field of ontology engineering and acquire requirements to the ontology from literature. These requirements are compared with relevant modeling approaches of the state of the art in order to prepare the design of the ontology (section 2). In section 3, the consolidated OSN ontology is presented and explained. The paper ends with a short summary and the consideration of further research aspects in section 4.

The research paper has a constructivist focus and refers to design science research by aiming the construction of an artifact [19]. Due to the novelty of the topic and the scope of the ontology design, we focus solely on the analysis and design phase and plan to evaluate the artifact in a separate research article [20].

2 Fundamentals, Requirements and State of the Art

2.1 Adapted Ontology Engineering Approach

An ontology can be defined as explicit, formal specification that conceptualizes a delimited discourse for a specific group of actors having a defined aim [21]. Essential elements of ontologies are concepts, properties, relations, axioms and relation types [22]. In contrast to conceptual models, ontologies do not strictly distinguish type level and instance level. Thus, they can define specific rules and integrity constraints between particular concepts [23]. Ontologies are primarily used for knowledge management and semantic web issues, but have also relevance for engineering tasks (what includes domain modeling, e.g., [24]). Several methodologies for the development of ontologies exist in the field of ontology engineering [25]. A prevalent ontology engineering method is the Methontology approach [26]. We have adapted this approach for the design of the OSN ontology: First, the aimed scope of the ontology is described by a general specification and the derivation of requirements in an argumentative manner. Therefore, general requirements from the OSN domain and exemplarily analysis purposes are described and consolidated within a requirements catalog in figure 1. Subsequently, existing approaches are examined against them in order to identify a need for integration, consolidation and extension. Afterwards, the final OSN ontology is consolidated and presented in section 3. The aimed ontology focuses the multi-perspective description of users and structures within OSN under special consideration of business-oriented analysis purposes. Thus, we aim to provide a comprehensive ontology containing several aspects of existing ontologies. The formality level of the aimed ontology is semi-formal using the OWL Lite language. The source of knowledge of the ontology is divided into the reuse and consolidation of existing ontologies and frameworks (as claimed in [26], p. 34) and the analysis of existing research works on purpose-specific analysis of OSN. For the second case, an informal text analysis of selected papers was conducted in order to identify relevant concepts and integrate them.

2.2 Requirements from User and Network Based Characteristics

Users are the central concept within OSN modeling. Although there are several ontologies for social user modeling, a consolidated approach that integrates different perspectives is still missing. For reasons of clarity, the consideration of users is divided into the aspects “identity”, “structure”, “activity” and “knowledge” [27]. The *identity* aspect describes who a user is and which characteristics the user has. Basic attributes are demographic data, interests, preferences or information about the current location (**R1**, **R2**, **R3**). Further attributes can be found in social user ontologies (e.g., [28], [29], [30], [31] or [32]). A single user can be associated with a specific role within OSN or within a specific group (**R6**). Examples for such roles are experts, opinion leaders, influencers or passive users ([33], [27]). A role always depends on the specific analytical view on the

OSN. For example, marketers want to identify opinion leaders or prospective customers, a security manager wants to identify users disclosing corporate data.

The aspect *structure* describes the relation between users ([27]; **R4**). Additionally, the relation of a user to some kind of a resource is understood in terms of the structural aspect. Relationships can have different reasons (e.g., real-world contact or business purpose [34]). The relationship between two users can be assessed regarding its distance or its strengths [35]. [2] emphasizes the need for a more detailed specification of relationship types (**R5**). Also, a relationship between users can be characterized by a specific influence level (positive, neutral, negative). [36] classifies the relationship between users and resources into the types “own”, “refers to” and “is trusted by” (**R9**). Social Network Analysis (SNA) comes from the field of sociology and analyzes OSN both from structural and activity-related perspectives (e.g., [15], [8]). The structural analysis refers to the analysis of the entire graph or network by generating measures like density, path length, clustering coefficients or the eigenvector centrality ([8], [37]). Also, the “classical” node-oriented centrality measures of [38] can be applied.

The aspect *activity* describes how active some user is [27]. Typical user activities are the exchange or publication of resources by posting any kind of statements (see detailed examples in [39]; **R7**, **R8**). Activities can be represented by activity graphs [8]. A basal activity between OSN users is communication. Communication is thereby significantly influenced by electronic word-of-mouth-effects (eWoM; [40], [41]). The adaption of a transmitted information object is influenced by a certain level of trustworthiness, the credibility of the sender, the expertise of the sender, the quality of the message and contextual involvement of the participated users or similarity between users ([40], [42]; **R9**, **R13**). The message has a neutral, positive or negative intention [43]. The aspect *knowledge* describes special know-how of a user regarding to analysis-specific topics (referring [27]; **R8**). Information objects (resources) are substantial for analytical reasons. Comments, ratings, favorite lists, events, opinions and profiles are very important analysis objects ([44]; **R10**, **R11**).

2.3 Requirements from Selected Analysis Purposes

Monitoring OSN for social CRM [5] and risk management [7] are prominent application areas for purpose specific analysis of OSN. The following business use cases are conceivable in social CRM: Personalized offers and target advertising, opinion and trend analysis, monitoring of the communication between customers and identification of opinion leaders, influencers and experts. It is also necessary to identify complaints, suggestions, reviews, ratings, comments and product-related statements ([5], [3], [45], [9], [15]). From a business point of view, statements can have a priority and are associated to a specific business object [46]. It is necessary to match user and customer information and integrate customer data and opinions from social media to CRM systems ([13], [45]; **R14**). In contrast to these benefits, businesses also face several technical, socio-technical and organizational threats and risks in OSN like data leakage; identify theft, social engineering attacks and industrial espionage [7]. The main

risks are loss of control and reputational damage, which is mainly related to user interactions and transferred information objects [7]. Thus, it is also necessary to monitor business-related conversations from a risk perspective. Data and information leakage is an other severe risk that is primarily caused by privacy problems (e.g., [47]). From a business perspective, insufficient privacy settings of employees in OSN can facilitate social engineering attacks and data leakage (e.g., [48]). A deep understanding and description of the accessibility of data objects within OSN is necessary to reduce uncertainty and complexity and to enhance the understanding of the reach of personal data in OSN (**R15**).

2.4 Existing Ontologies in the Context of OSN

FOAF is an OWL based domain ontology for modeling both users and structures in the semantic web [28]. FOAF is primarily used for user-adaptive systems and for user aggregation issues. Core concepts are user basics, personal information, accounts, projects, groups and documents. **SIOC** is a RDF based domain ontology for the connection of discussion media such as blogs, forums or OSN [29]. Core concepts are posts, roles and items. **GUMO** is a generic ontology that focuses the generic description of users within their life cycle [31]. GUMO is mainly used for data exchange between user adaptive systems. Special concepts are emotional state, knowledge and privacy preferences. **URM** is an application-oriented ontology for cross-system personalization and user modeling. Essential concepts are preferences, relationships and tasks [49]. **U2M** is a RDF based ontology for comprehensive user modeling and was designed as integration of GUMO and UserML [50]. Important concepts are preferences, demographics and roles. **HiddenU** is less an ontology but rather a framework that aims to facilitate social nexus for the integration of scattered social data [32]. Thus, HiddenU has an aggregation and integration focus and provides more abstract concepts (e.g., competences, social relationships, preferences). **U-Sem** is a semantic web service for enriching and mining user data, which contains user modeling in U-Sem profiles that bases on the several ontologies [51]. Description groups within U-Sem are observations (activities, objects), user characteristics and domain knowledge. Social aspects are not considered. **SWUM** is a RDF based application-oriented ontology for the aggregation of user profiles in social media. Core concepts are user role, relationships, locations and demographics [52]. **OpenSocial** was developed as API standard for data exchange of OSN apps. Thus, it is very data-oriented providing concepts like activity, friends, meta data and messages [53].

In view of the variety of approaches, **Kapsammer et al. (2011)** proclaim a comprehensive and extensible reference model for the description of social user profiles. Their reference model is divided into the packages core, meta info (privacy, provenance, quality, context), resource (structure and behavior) and relation (universal, social; [18]). **SocIoS** is an application-oriented ontology for independent development of OSN apps that provides concepts like activities, events or reputations. The approach of **Bozzon et al. (2013)** focuses modeling of crowd sourcing scenarios and adapts user ontologies and content description

models. New concepts like confidence level, network roles and topical affinities are defined [17]. Further, some data-oriented reference models can be found in literature. For instance, in the context of SCRM systems [13] and risk analysis of OSN [12]. [16] evolves a formal model for social data analysis and formalize concepts of social graphs and social text.

2.5 Requirements Catalogue

Based on the analysis of the OSN domain, fifteen requirements were identified. First, identity related attributes like general personal data, context information and user preferences need to be addressed. Second, relations between single users and their respective type of relation need to be considered. Also network-specific measures are assigned to the structure group in order to describe a larger set of users. The role type enables the specification of user roles within a specific group or situation (e.g., opinion leader, product expert or critic). Third, the activity group encompasses the description of user activities and the description of social interaction. Fourth, the knowledge group stands for the ability of describing competences and for the representation of his reputation within the OSN, which influences his standing and his level of trust within the network. In the fifth place, the aimed ontology should describe information objects regarding their simple content, their meta data and their level of confidence or quality. The last group contains more analysis-specific features. Therefore, the ontology should provide concepts for the integration of users with existing customers of a business and for the annotation of relevant properties. For reasons of internal security and risk management, users should also be monitored in regard of their privacy settings and some kind of assessment of the business confidentiality of their statements.

The majority of ontologies have a technical and data-oriented focus and aim to support issues of data exchange and integration. There are deficits regarding the representation of social attributes, interaction features, activity types and respective roles. Also, there is a lack regarding modeling topical affinities, reliabilities and trust ([17], p. 17). These aspects might be relevant for the identification of opinion leaders ([15], [8]). Also, not all languages support the adequate description of information objects, especially regarding their meta data and confidence. Generally, a suitable integration of existing languages is missing and there is no universal data model or standard, that both contain structural, personal and social features! Consequently, there is also a lack of explicit considerations of business relevant features; for example in terms of a customer user integration (referring to [13]).

3 Ontology Engineering

By reviewing the stated approaches, it became obvious that none of the ontologies or reference models meet all requirements. Thus, we decided to specify a new integrated ontology called **OntOSN**. With regard to Methontology approach, we tried to reuse as much as possible concepts from existing ontologies

Approach:		FOAF	SIOC	GUMO	URM	HiddenU	U-Sem	SWUM	OpenSocial	Kapsammer et al. (2012)	Soclos	Bozzon et al. (2013)	Rosemann et al. (2012)	Braun & Esswein (2012)	Consequence for Design
Requirement:	Identity														
	R1 Personal Data	●	○	●	⊙	●	●	●	○	○	●	●	●	●	☑
	R2 Context	○	○	○	○	○	○	○	○	○	○	○	○	○	☑
	R3 Preferences	○	○	○	○	○	○	○	○	○	○	○	○	○	☑
	Structure														
	R4 Relation, Groups	●	○	○	⊙	●	○	○	●	○	●	●	●	○	☑
	R5 Relation Types	⊙	⊙	○	○	○	○	○	○	○	○	●	○	○	☑
	R6 Role Types	⊙	⊙	○	○	○	○	●	○	○	○	●	○	○	☑
	Activity														
	R7 Social Interaction	⊙	○	○	○	●	○	○	○	○	●	○	○	○	☑
	R8 Activity Types	○	○	○	○	○	○	○	○	○	○	○	○	○	☑
	Knowledge														
	R9 Competences	○	○	⊙	⊙	●	○	○	○	○	○	○	○	○	☑
	R10 Trust, Reputation	○	○	⊙	○	○	○	○	●	○	○	○	○	○	☑
	Information Objects														
	R11 Content	⊙	●	○	○	○	○	○	●	○	●	●	○	○	☑
	R12 Meta Data	○	●	○	○	○	○	○	●	○	○	○	○	○	☑
	R13 Confidence, Quality	○	○	○	○	○	○	○	○	○	○	●	○	○	☑
	Others														
	R14 Customer Integration	○	○	○	○	○	○	○	○	○	○	○	○	○	☑
	R15 Privacy, Visibility	○	○	○	●	●	○	○	○	○	○	○	○	○	☑
Specification		●	●	○	○	○	○	●	●	○	○	●	●	●	

Fig. 1. Requirements of the aimed OSN ontology in comparison with existing ontologies, frameworks and conceptual models

[26]. Therefore, each ontology was modeled graphically in order to identify relevant concepts, properties, relations and inheritance structures. Unfortunately, only a few ontologies are well-defined within OWL specifications or some kind of a meta-model (e.g., FOAF and SIOC). The majority of the approaches does not provide an integrated specification (e.g., GUMO, U2M). This circumstance makes it difficult to identify definable concepts from several approaches. Thus, it was necessary to evolve relevant concepts based on informal descriptions or frameworks within the stated articles. Based on all explicated models, the OntoSN ontology was consolidated by integrating existing concepts and adding new concepts. The ontology was designed by using the OWL Lite ontology language and we decided to explicate the ontology in diagrams in order to provide an appropriate overview. The legend of the notation can be found at the lower right corner of figure 2. For reasons of readability, the “range” and “domain” specifications of property relations as well as specifications of inheritances (“subClassOf” and “subPropertyOf”) are hidden in the respective diagrams. Concepts that were reused and adapted from existing ontologies are marked by the respective namespace in front of the concept name (FOAF, SIOC, U2M, Ka+11 (Kapsammer et al. 2011), Bo+13 (Bozzon et al. 2013), SWUM, SocIoS). Thus, an integration or mapping of the new ontology with existing ontologies remain possible. Besides, owned properties of concepts (especially “Person”, “Event”, “Rating” and “User Role”) are not stated with respect to the limited space of this paper. Particular property values can be found in the original ontologies.

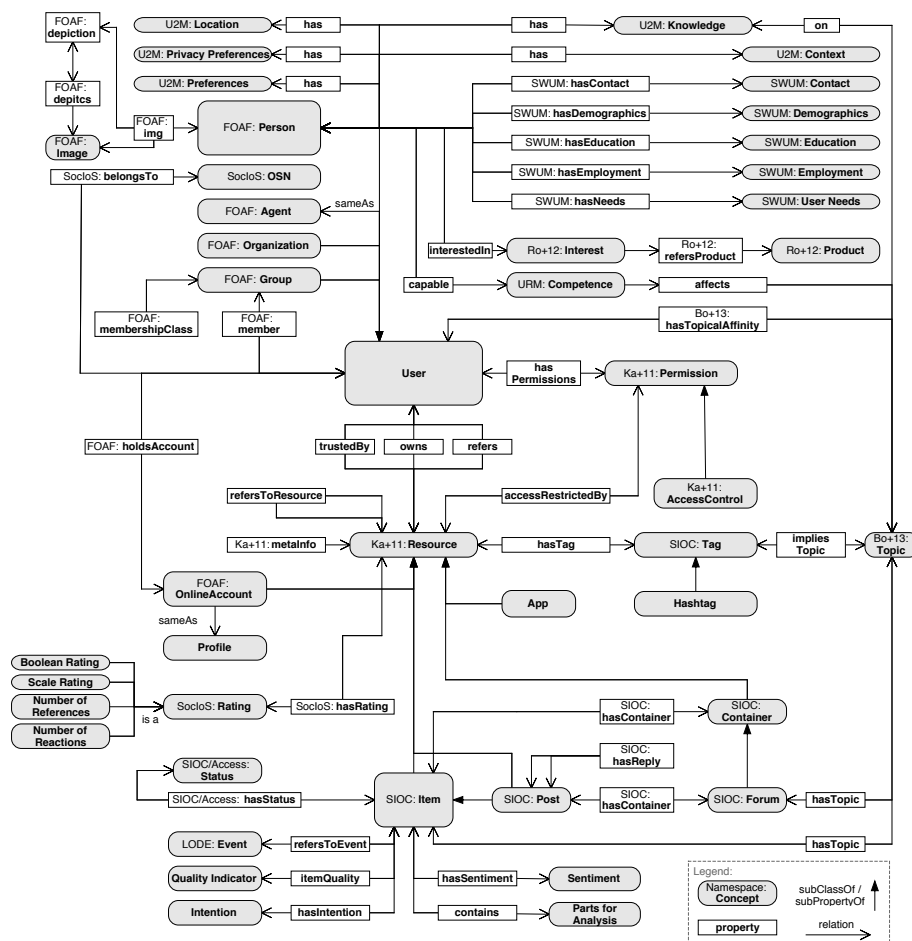


Fig. 2. Part of the OntOSN ontology representing essential user concepts and content description aspects

3.1 OntOSN: User and Information Objects

Figure 2 presents the part of OntOSN that depicts large parts of the user concepts and the information object concepts. Thus, the “node” elements of the abstract OSN structure are considered. The central concept of the OSN is the user concept covering all personal attributes and features of either a person (normal case), an organization or a group. Next to demographical attributes, a single user has a range of attributes like employment information, preferences, specific knowledge or competences, whereby the latter two affect a specific topic. A user always belongs to a specific OSN (e.g., Facebook) and holds an online account within this OSN. If a user maintains several accounts, the “holdsAccount” property enables an inference to the actual user.

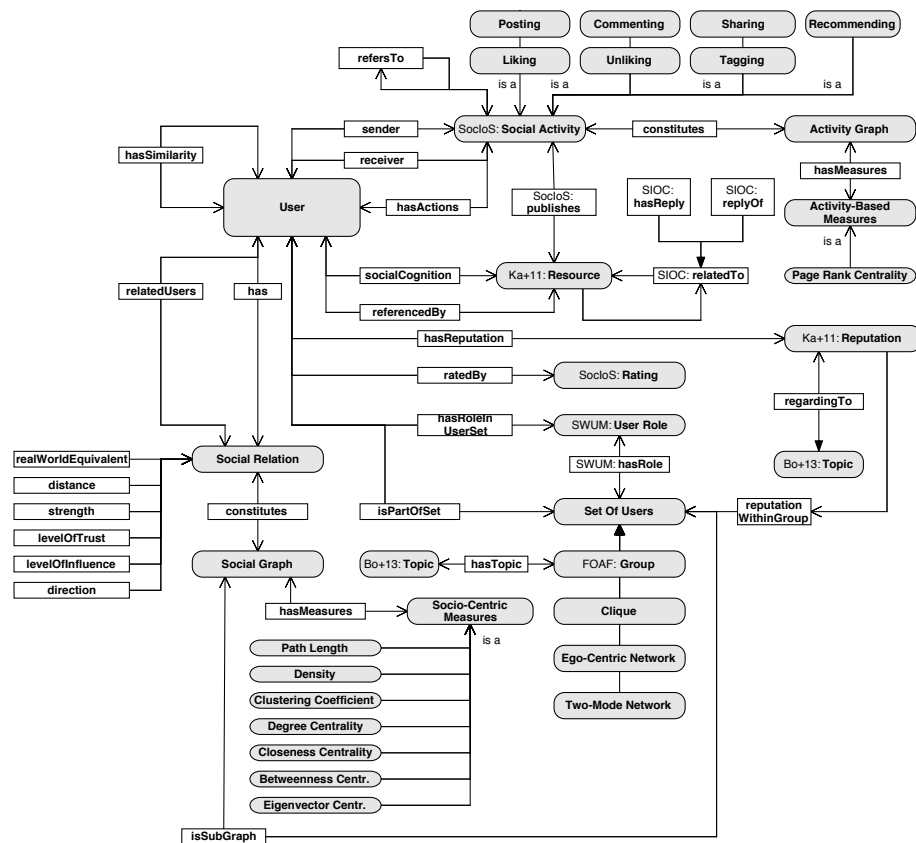


Fig. 3. Part of the OntOSN ontology representing structural aspects

The second very important concept within the analysis of OSN is the resource concept that stands for all OSN objects that can be created, altered, transferred, referenced or deleted by users. In particular, the resource types from SIOC (item, post, forum and container [29]) were adapted, whereby forums are understood both as structured discussions (e.g., within user groups) and as conversation between a set of users. Please note, that the item concept is the central point for analysis and stands for the “atomic” content element that can be analyzed regarding its contents. The other concepts (post, forum, container) are more abstract concepts lacking single items. Each resource can be semantically enriched by tags (keywords) in order to emphasize its topic. Further, a resource has some meta information (e.g., creating timestamp, language) and could refer to an other resource (e.g., as a link). The relation between users and resources is considered in several aspects: In accordance to [18], every user has a specific permission for a resource (e.g., full access, no access, read-only) that limits the scope of elements he can access within the OSN. In particular, a user can refer to specific resources and can be referenced by a resource (e.g., as link to the profile). If the

user is the creator or author of a resource, he owns it. Besides, the proclaimed ontology allows to specify a “trustedBy” relation between users and resources to give implicit statements on the trustworthiness of some content.

As stated above, the item concept is central for content-related analysis. With reference to the SIOC access module, an item may have a specific status (e.g., draft). Also, it can correspond to some kind of event. Events can be described by the LODE event description ontology. Further, an item generally has a specific sentiment, an intention and implies a certain level of quality (see [18]). Besides, the content of an item provides several elements (tokens) for analysis; such as product names, acronyms, negative words or locations (e.g., [46]).

3.2 OntOSN: Structural Aspects

Figure 3 depicts the part of OntOSN that focuses on the structural aspects of an OSN; namely social relations and interactions between users or users and resources. Thus, “edge” elements of the abstract OSN structure are considered. Generally, a user has several social relations to other users. Each relation can be specified by several properties like “realWorldEquivalent” (real relations), “distance” (number of intermediate nodes or users) or “strength” (e.g., [35]). Additionally, the level of trust between two users, the level of influence and the prevailing influence direction between two users can be modeled (see [42]). The set of social relations constitutes the social graph that can be measured by several socio-central measures [38], [37]. The similarity between two users (e.g., [40]) is covered by the “hasSimilarity” property.

Further, a user can perform social actions and interactions to or with other users. A social action (e.g., posting, liking, sharing) results in a resource (e.g., a public message), which could reference other users. Also, some social actions can refer to an earlier action. The set of social actions and interactions constitutes the activity graph that can be measured by activity-based measures like the page rank centrality index (see [8], [27]). Resources may also cause the social cognition of users by associating them with different information objects. The relation of a single user within a set of users is depicted in the middle of figure 3. Thereby, a “set of users” is understood as an abstract concept covering several group-related concepts like cliques, ego-centric networks or two-mode networks within the entire OSN. We used this construction for pointing out different roles a user can have within a OSN; depending a the related set of other users. Hence, users can also have different reputations; depending on the relevant group or topic.

3.3 OntOSN: Business Aspects

As stated in section 2.5, any kind of business relevant concepts are missing in existing ontologies since the approaches mainly focus technical aspects. With respect to the increasing pertinence of OSN for businesses, it is necessary to integrate business relevant concepts. We therefore decided to both consider beneficial aspects (customer integration) and risky aspects (e.g., avoid internal data

leakage). Figure 4 depicts the part of OntOSN that focuses on business aspects. It is possible to match some user (person) to a customer. The customer concept can be extended with some kind of information coming from a CRM system (e.g., customer history). Also, users can be related to other stakeholders in the context of a company (e.g., competitors or business partners). It might be necessary to monitor these users in the context of competitive intelligence, for example. In contrast to that, a user can also represent an employee of a company. The identification of employees is crucial for recognizing and avoiding internal data leakage (see [7]). Such internal data might be disclosed by explicit statements resulting in any kind of content items. With regard to both positive and negative effects of OSN for businesses, items can be identified as business relevant for analysis-specific reasons (concept “business relevant content”). An item can imply a specific business relevance type like a complaint, a question or some product ratings. That type has a specific priority (for reaction) and can be specified whether it affects the own company or a competitor.

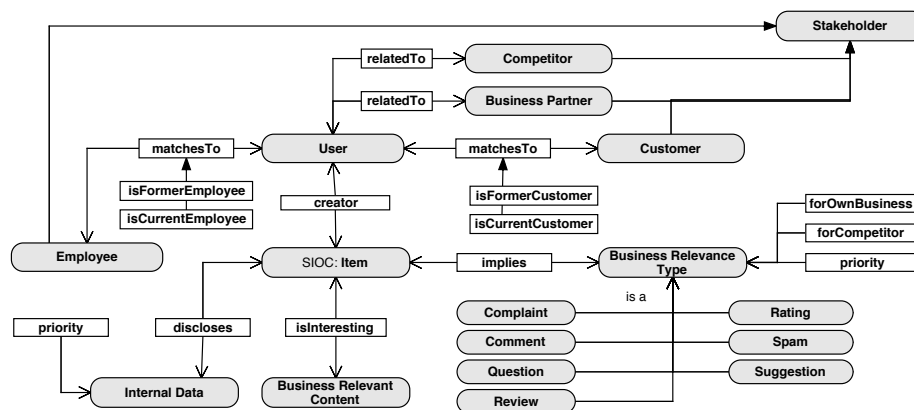


Fig. 4. Part of the OntOSN ontology representing concepts for business analysis

4 Conclusion and Further Research

This research paper designs an ontology for the analysis of OSN by integrating existing ontologies and adding new concepts and properties. The evolved ontology can be used for the analysis of OSN from different viewpoints and facilitates the analysis-specific design and development of OSN analysis tools due to the unified language base. Hence, it is possible to provide an integrated, comprehensive understanding of the emerging research field of OSN. Also, the ontology allows its extension due to the well-defined specification as OWL Lite ontology. In order to enable an integration with specific business purposes, the ontology explicitly considers concepts for the integration of customers. Thus, the contributions of this paper are as follows: (1) Development of a requirements set for

a both business-oriented and technical-oriented OSN ontology. (2) Examination of existing ontologies in regard to the requirements set. (3) Proposal for an integrated OSN ontology under special consideration of business relevant aspects.

Nevertheless, there are some points for further research and improvement. First, the evolved ontology needs to be published as RDF (XML) file in order to facilitate its application. Therefore, also owned properties of concepts need to be explicated. Second, the consideration of further views on OSN needs to be discussed. Especially, an integration with concepts from data mining is promising. Third, existing concepts of the OntOSN ontology can be enriched by specific vocabularies or further ontologies in order to enable an even better expressiveness of single concepts.

References

1. Ellison, N.B., et al.: Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication* **13**(1) (2007) 210–230
2. Heidemann, J., Klier, M., Probst, F.: Online social networks: A survey of a global phenomenon. *Computer Networks* (2012)
3. Culnan, M.J., McHugh, P.J., Zubillaga, J.I.: How large us companies can use twitter and other social media to gain business value. *MIS Quarterly Executive* **9**(4) (2010) 243–259
4. Dinter, B., Lorenz, A.: Social business intelligence: a literature review and research agenda. In: *Proceedings of the ICIS*. (2012)
5. Faase, R., Helms, R., Spruit, M.: Web 2.0 in the crm domain: defining social crm. *International Journal of Electronic Customer Relationship Management* **5**(1) (2011) 1–22
6. Stieglitz, S., Dang-Xuan, L., Bruns, A., Neuberger, C.: Social media analytics. *Business & Information Systems Engineering* **6**(2) (2014) 89–96
7. Braun, R., Esswein, W.: Towards a conceptualization of corporate risks in online social networks. In: *Proceedings of the 17th EDOC, IEEE* (2013) 267–274
8. Heidemann, J., Klier, M., Probst, F.: Identifying key users in online social networks: A pagerank based approach. In: *Proceedings of the ICIS*. (2010)
9. Richter, D., Riemer, K., vom Brocke, J.: Internet social networking - research state of the art and implications for enterprise 2.0. *Business & Information Systems Engineering* **3**(2) (2011) 89–101
10. Appleford, S., Bottum, J.R., Thatcher, J.B.: Understanding the social web: towards defining an interdisciplinary research agenda for information systems. *ACM SIGMIS Database* **45**(1) (2014) 29–37
11. Stieglitz, S., Dang-Xuan, L.: Emotions and information diffusion in social media - sentiment of microblogs and sharing behavior. *Journal of Management Information Systems* **29**(4) (2013) 217–248
12. Braun, R., Esswein, W.: Corporate risks in social networks – towards a risk management framework. In: *Proceedings of the 18th ACMIS*. (2012)
13. Rosemann, M., Eggert, M., Voigt, M., Beverungen, D.: Leveraging social network data for analytical crm strategies. In: *Proceedings of the ECIS*. (2012)
14. Gundecha, P., Liu, H.: Mining social media: a brief introduction. *Tutorials in Operations Research* **1**(4) (2012)

15. Bonchi, F., Castillo, C., Gionis, A., Jaimes, A.: Social network analysis and mining for business applications. *ACM Transactions on Intelligent Systems and Technology (TIST)* **2**(3) (2011) 22
16. Mukkamala, R.R., Vatrupu, R., Hussain, A.: Towards a formal model of social data. Technical report, IT-Universitetet i København (2013)
17. Bozzon, A., Fraternali, P., Galli, L., Karam, R.: Modeling crowdsourcing scenarios in socially-enabled human computation applications. *Journal on Data Semantics* (2013) 1–20
18. Kapsammer, E., Mitsch, S., Pröll, B., Schwinger, W., Wimmer, M., Wischenbart, M.: A first step towards a conceptual reference model for comparing social user profiles. In: *Proc. of the Int. Workshop on User Profile Data on the Social Semantic Web*. (2011)
19. Hevner, A.R., March, S.T., Park, J., Ram, S.: Design science in information systems research. *MIS quarterly* **28**(1) (2004) 75–105
20. Alturki, A., Gable, G.G., Bandara, W.: A design science research roadmap. In: *Service-Oriented Perspectives in Design Science Research*. Springer (2011) 107–123
21. Gruber, T.R.: A translation approach to portable ontology specifications. *Knowledge Acquisition* **5**(2) (1993) 199–220
22. Corcho, O., Fernández-López, M., Gómez-Pérez, A.: Methodologies, tools and languages for building ontologies. where is their meeting point? *Data & Knowledge Engineering* **46**(1) (2003) 41–64
23. Spyns, P., Meersman, R., Jarrar, M.: Data modelling versus ontology engineering. *ACM SIGMod Record* **31**(4) (2002) 12–17
24. Happel, H.J., Seedorf, S.: Applications of ontologies in software engineering. In: *Proc. of Workshop on Sematic Web Enabled Software Engineering*. (2006) 5–9
25. Sure, Y., Tempich, C., Vrandecic, D.: Ontology engineering methodologies. *Semantic Web Technologies: Trends and Research in Ontology-based Systems* (2006) 171–190
26. Fernández-López, M., Gómez-Pérez, A., Juristo, N.: *Methontology: from ontological art towards ontological engineering* (1997)
27. Probst, F., Grosswiele, L., Pflieger, R.: Who will lead and who will follow: Identifying influential users in online social networks. *Business & Information Systems Engineering* (2013) 1–15
28. Brickley, D., Miller, L.: Foaf vocabulary specification 0.99. <http://xmlns.com/foaf/spec/>, 2014. (2014)
29. Breslin, J.G., Decker, S., Harth, A., Bojars, U.: Sioc: an approach to connect web-based communities. *Int. Journal of Web Based Communities* **2**(2) (2006) 133–142
30. Tserpes, K., Papadakis, G., Kardara, M., Papaoikonomou, A., Aisopos, F., Sardis, E., Varvarigou, T.: Socios: A social media application ontology. In: *OTM 2012 Workshops*. (2012) 574–584
31. Heckmann, D., Schwartz, T., Brandherm, B., Schmitz, M., von Wilamowitz-Moellendorff, M.: Gumo—the general user model ontology. In: *User Modeling 2005*. Springer (2005) 428–432
32. Kappel, G., Schönböck, J., Wimmer, M., Kotsis, G., Kusel, A., Pröll, B., Retschitzegger, W., Schwinger, W., Wagner, R., Lechner, S.: The hidden u - a social nexus for privacy-assured personalisation brokerage. In: *Proceedings of the ICEIS*. (2010) 158–162
33. Braun, R., Esswein, W.: Towards a reference architecture model for virtual communities. *Proc. of the IADIS Int. Conference on Web-Based Communities* (2011) 48–56

14 Braun and Esswein

34. Mislove, A., Marcon, M., Gummadi, K.P., Druschel, P., Bhattacharjee, B.: Measurement and analysis of online social networks. In: *Proceedings of the 7th ACM SIGCOMM*, ACM (2007) 29–42
35. Granovetter, M.: Economic action and social structure: the problem of embeddedness. *American journal of sociology* (1985) 481–510
36. Le Malécot, E., Suzuki, M., Eto, M., Inoue, D., Nakao, K.: Online social network platforms: toward a model-backed security evaluation. In: *Proceedings of the 1st Workshop on Privacy and Security in Online Social Media*, ACM (2012) 3
37. Zhang, M.: Social network analysis: history, concepts, and research. In: *Handbook of social network technologies and applications*. Springer (2010) 3–21
38. Freeman, L.C.: Centrality in social networks conceptual clarification. *Social networks* **1**(3) (1979) 215–239
39. Berthon, P.R., Pitt, L.F., Plangger, K., Shapiro, D.: Marketing meets web 2.0, social media, and creative consumers: Implications for international marketing strategy. *Business Horizons* **55**(3) (2012) 261–271
40. Steffes, E.M., Burgee, L.E.: Social ties and online word of mouth. *Internet Research* **19**(1) (2009) 42–59
41. Cheung, C.M., Thadani, D.R.: The effectiveness of electronic word-of-mouth communication: a literature analysis. 23rd Bled eConference, Bled, Slovenia (2010)
42. Lis, B.: In ewom we trust. *Business & Information Systems Engineering* (2013) 1–12
43. Rafaeeli, S., Raban, D.R.: Information sharing online: a research challenge. *International Journal of Knowledge and Learning* **1**(1) (2005) 62–79
44. Costa, P.R., Souza, F.F., Times, V.C., Benevenuto, F.: Towards integrating online social networks and business intelligence. In: *IADIS International Conference on Web Based Communities and Social Media*. Volume 2012. (2012)
45. Sigala, M.: ecrm 2.0 applications and trends: The use and perceptions of greek tourism firms of social networks and intelligence. *Computers in Human Behavior* **27**(2) (2011) 655–661
46. Ajmera, J., Ahn, H.i., Nagarajan, M., Verma, A., Contractor, D., Dill, S., Denesuk, M.: A crm system for social media: challenges and experiences. In: *Proceedings of the 22nd WWW Conference*. (2013) 49–58
47. Huber, M., Mulazzani, M., Leithner, M., Schrittwieser, S., Wondracek, G., Weippl, E.: Social snapshots: Digital forensics for online social networks. In: *Proceedings of the 27th Annual Computer Security Applications Conference*, ACM (2011) 113–122
48. Al Hasib, A.: Threats of online social networks. *International Journal of Computer Science and Network Security* **9**(11) (2009) 288–93
49. Zhang, F., Song, Z., Zhang, H.: Web service based architecture and ontology based user model for cross-system personalization. In: *IEEE/WIC/ACM International Conference on Web Intelligence*. (2006) 849–852
50. Heckmann, D., Schwarzkopf, E., Mori, J., Dengler, D., Kröner, A.: The user model and context ontology gumo revisited for future web 2.0 extensions. *Contexts and Ontologies: Representation and Reasoning* (2007) 37–46
51. Abel, F., Celik, I., Hauff, C., Hollink, L., Houben, G.J.: U-sem: Semantic enrichment, user modeling and mining of usage data on the social web. *arXiv preprint arXiv:1104.0126* (2011)
52. Plumbaum, T., Wu, S., De Luca, E.W., Albayrak, S.: User modeling for the social semantic web. In: *SPIM*. (2011) 78–89
53. Häsel, M.: Opensocial: an enabler for social applications on the web. *Communications of the ACM* **54**(1) (2011) 139–144

Social Media and Communication Processes at Work: Evidence from China

Carol XJ Ou¹, Louie HM Wong², Robert M Davison² and Zhang Cheng³

¹Tilburg University, Tilburg, The Netherlands
carol.ou@uvt.nl

²City University of Hong Kong, Kowloon, Hong Kong
{louie.wong@my., isrobert@}cityu.edu.hk

³Fudan University, Shanghai, China
zhangche@fudan.edu.cn

Abstract. Web 2.0 applications, such as instant messengers and other social media platforms, are fast becoming ubiquitous at work, yet their impact on performance is poorly understood. We investigate these impacts in the Chinese workplace, analyzing data from 179 organizational employees. We find that vertical and horizontal communication contribute significantly to individual, team and organizational performance, with high levels of variance explained. We provide empirical evidence for how Web 2.0 enables organizations to reach out to collaborators and business partners, thereby boosting individual productivity, team collaboration and organizational agility. We discuss the implications of the research for scholars and practicing managers, also identifying future research directions.

1 Introduction

Social media applications are not only ubiquitous in our personal lives, but they are fast becoming common at work as well. A small number of researchers have started to extend investigations of social media beyond the social context into the workplace, assessing the impact that social media has on actual work practices. In this paper, we specifically examine the situation in China, surveying employees across a wide range of industries with respect to the impact of social media on their communication processes and the consequent performance outcomes.

Recent evidence [10], [11], [23], [24] suggests that social networking applications are widely adopted in the Chinese workplace, where they enhance a variety of work related processes, including collaboration, problem solving and information sharing. Not all companies are equally enthusiastic, however, and some do attempt to restrict employees' access to social media at work [9]. As a form of communication media, we find that social media are extensively used for a variety of communication purposes: inside the organization and beyond it; vertically (e.g. from senior to junior levels) and horizontally (e.g. between peers, or between a sales team and its customers).

In China, many Western social media platforms like Facebook, YouTube and Twitter are blocked by government censors and are hence very hard to access. In their place, a crop of new, domestic social media platforms has evolved, notably: RenRen

(a social media platform akin to Facebook), Youku (a video sharing site akin to YouTube) and Weibo (a microblog site similar to Twitter). Instant messengers are also very popular, including MSN (which has been terminated elsewhere in the world yet still runs in China), WeChat (popular globally, yet developed by Chinese firm TenCent), and QQ (an indigenous messenger similar to MSN, yet with additional functionality such as news feeding and voice messages).

The presence of an eclectic mix of social media options for Chinese workers, coupled with a culture that favours this form of informal communication [19], inspired us to ask: What is the Impact of Social Media Use on Work-Based Communication Processes and Outcomes in China?

To answer this question, we adapt the communicative ecology framework (CEF) [11, 12], creating a new theoretical model specific to the way social media enhances communication processes and outcomes. We test this model with data obtained through a survey of employees ($n=179$) in a variety of Chinese firms.

Following this introduction, we briefly review the relevant literature before proposing a theoretical model, explaining our research methods and presenting our findings. These are discussed in the context of practical and research implications before we conclude the paper.

2 Literature Review

The CEF [12] has been developed in the context of analysing the relationships between different social groups of people, the technologies that they leverage in their communication and the nature of their interactions. The idea that media and technology influence the way we communicate dates back at least as far as McLuhan [20], though more recent work [3], [12] has consolidated earlier arguments and led to the defining of the term communicative ecology as “the context in which communicative processes occur” [12, p.756]. The three layers of the framework, viz.: technological, social and discursive, are considered to be “intricately interwoven and mutually constitutive” [11]. As such, it is not practical to try to separate out the different technological media being used, the different social contexts in which communication takes place or even the discursive content – all are entangled together. As Tacchi et al. [30, p.17) note, communication requires “processes that involve a mix of media, organized in specific ways, through which people connect with their social networks”.

In parallel with this work on communicative ecology, social media (sometimes known as Web 2.0) exerts an increasing influence on organizational communication processes. A comprehensive review of the extensive literature on social media is beyond the scope of this paper. Much of the applied research focuses on social contexts, including social activism, disaster response and political campaigns [1], [18]. Research on organizational contexts is more limited, with a focus on microblogging and instant messengers.

Microblogging is noted for its potential to increase the speed and quality of communications, notably in situations characterised by urgency and low bandwidth [18]. It also increases transparency and accountability [15]. Business applications of mi-

croblogging include internal and external communication, information sharing, relationship marketing. Yammer, the corporate microblogging application, is used for team-task coordination [28] and e-marketing [16]. Studies of microblogging in China are very limited, but Zhang [33] noted that it is used in viral marketing.

Instant messengers (IM), such as MSN, QQ and G-Talk, have been extensively studied in the Western context for several years, with an extensive literature on the topic. IM users often signal their online status and thus their availability to be contacted [13]. Studies of IM in the Chinese workspace also suggest its popularity and value for enhancing communication [24]. Although IM tools enhance communication quality and interpersonal connectedness, they also increase work interruptions – a potential negative consequence [5], [23].

Although the literature is supportive of the value of social media for organizations, it is important to note that not all organizations are equally enthusiastic. Davison et al. [10] report that there are considerable corporate concerns, not least with the potential for time wasting and the loss of productivity. Social media applications, like any other technology, need to be embedded into an appropriate organizational culture that values and leverages them to organizational advantage [15]. From the theoretical point of view, how Web 2 technologies facilitate the organizational communication process and subsequently improve performance is under-investigated. These research gaps lead to the next section, in which we develop our theoretical framework.

3 Theoretical Framework

In adapting the Communicative Ecology Framework (CEF) [12], and its later instantiation as a high level model [11], we have focused on the primary communication functions where we see social media applications providing significant benefits. The CEF connects the communication media, people and the social environment, as well as the content of communication into an ecological system. In this system, the interaction between people in networks is facilitated by information and communication technology. This ecological metaphor and research perspective can be applied into either holistic (macro) or individual (micro) levels of analysis. Accordingly, we propose that a mediating role for the different types of communication, vertical and horizontal, is necessary for Web 2 technologies to exercise their positive influence on performance at the individual, team and organizational levels. In the next few paragraphs, we present arguments to justify this new theoretical model (see Figure 1) and the associated hypotheses.

3.1 The Use of Web 2.0 in Facilitating Organizational Communication

Following the explanations in the previous sections, we contend that the use of Web 2.0 technology at work supports many functions including: to stay informed; for information and knowledge sharing; to connect customers and colleagues; and for learning and training purposes [10, 23]. Meanwhile, organizational communication can be classified into two dimensions: vertical and horizontal [32]. The former refers

to the information disseminated from senior to junior colleagues, or from a management team down to operational level staff to establish and maintain corporate strategy and norms [21]; the latter includes peer communication for information and knowledge exchange, as well as social relationship building between colleagues and departments, and with external customers and suppliers [10].

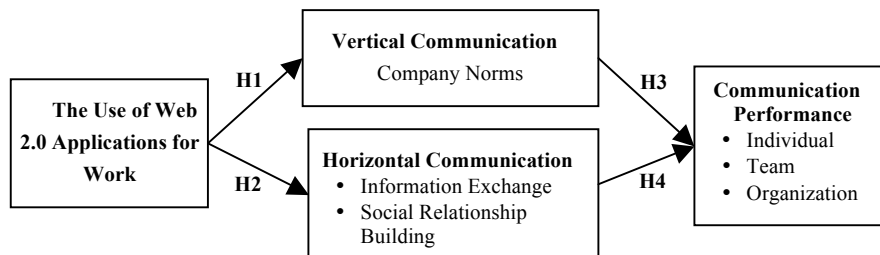


Fig. 1. The Proposed Theoretical Model

We argue that the use of Web 2.0 at work can facilitate both vertical communication to establish and maintain corporate norms, and horizontal communication for information exchange and social relationship building. From a top-down perspective, the use of web 2.0 (such as blogs, podcasts and prediction markets) can help disseminate the company strategy and deliver industry information to the whole company. Participatory Web 2.0 technologies facilitate vertical communications such as announcements, information about organizational activities and outcomes. With the frequent information dissemination and interaction with employees who engage with Web 2.0, the company identity can be established and emphasized as a result of the interaction. From a bottom-up perspective, Web 2.0 technology can provide a channel that aggregates opinions from individual employees who wish to contribute to strategic planning. Web 2.0 technology thus enhances the efficiency of information flow between the top and bottom of the organization. We thus propose that:

H1: The use of Web 2.0 technology at work can enhance vertical communication, specifically for establishing company norms, across an organization.

In the horizontal dimension of organizational communication, Web 2.0 technologies, typically wikis and knowledge forums, enable mass internal content creation and the broadcasting of solutions across internal departments, as well as different physical locations, as solutions are found to problems. The ease-of-use characteristics of Web 2.0 can also enable external parties to participate in product development, provide feedback and aid in customer support. Many companies, including General Electric, Procter & Gamble, Shell and Airbus, have found that Web 2.0 is valuable for “horizontal collaboration” since it “harnesses the power of collective intelligence to boost productivity, foster innovation and create enhanced value” [7]. The immediacy and the ease of use of Web 2.0 can help integrate customers, suppliers and employees into the communicative ecology by either providing a dialogue forum or facilitating the one-to-one interactive and informal communication that subsequently contributes to the social relationship building. Therefore, we propose that:

H2: The use of Web 2.0 technology at work can enhance horizontal communication, including information sharing and social relationship building, across an organization.

3.2 The Impacts of Communication on Improving Performance

Organizational communication is often described as being both vertical and horizontal [32]. As explained previously, vertical communication for establishing corporate norms requires both top-down information flows in the chain of command and bottom-up information gathering for decision making. The need for and the presence of vertical communication channels have long been considered important in an organization [21]. We argue that vertical communications can provide employees with important and updated information about their jobs, organizations, market and industrial environment. A positive communication climate and effective vertical communication are regarded as important means to strengthen employees' recognition and therefore identification with their organizations, which subsequently contributes to employees' productivity, a team's coherence and an organization's long-term financial performance [29]. More recently, scholars have argued that internal communication, of which vertical communication is one of the components, is the most "fundamental driver of business performance" [14, p.11]. Accordingly, we propose that:

H3: Vertical communication that is undertaken in order to support company norms contributes to individual performance (H3a), team performance (H3b) and organizational performance (H3c).

As explained previously, horizontal communication in this study includes the information exchange and social relationship building that takes place by means of communication among persons who have no hierarchical relationship. These people include: team members; employees from different departments; procurement managers and their suppliers; marketing & sales personnel and their customers. With the aid of Web 2.0 technology, open dialogues between different participants in various locations can be easily established to stimulate interactions between employees and external parties like suppliers and customers. These interactions are valuable because they provide relevant information and capture participants' insights and issues in a timely manner. The information shared inside an organization is critical for understanding, learning, clarifying ambiguity, conflict resolution and productive team work, while social relationships are established as interaction occurs. Researchers [10], [23] have found that effective communication, supported by Web 2.0 technologies, can facilitate the building of trust, a critical ingredient not only in work and social relationships, but also in successful organizations. Accordingly, we propose that:

H4: Horizontal communication, covering information exchange and social relationship building, contributes to individual performance (H4a), team performance (H4b) and organizational performance (H4c).

We integrate the above hypotheses into the theoretical research model (as shown in Figure 1) and explain the empirical validation of the research model in the following section.

4 Methodology

We used the survey method to verify the research model. This section briefly explains the development and validation of measures, and the data collection procedure.

4.1 Measures

We use existing measures from the literature to form the items used in this study. The independent variable, the use of Web 2.0 at work, is operationalized based on Ou and Davison's [23] measure of IM usage at work as well as Olasina's [22] measure of "the use of Web 2.0 tools and social networking sites by librarians". We extended their measures into the context of using Web 2.0 technology in organizational workplaces. Regarding vertical communication for establishing company norms, the measures are adapted based on Lewis et al.'s [17] items about maintaining corporate norms. We operationalize horizontal communication as a second-level construct that covers information exchange and social relationship building. Information exchange as the first-level construct is measured based on Phang et al.'s [26] items of knowledge seeking and contribution through online community systems. We adapted their measures into the context of information exchange in an organization in general. For social relationship building, we made use of Ou and Davison's [23] scale of social relationships. We separate performance into three distinct dependent variables that cover individual, team and organizational levels. Specifically, the items of individual and team performance were adapted based on the scales developed by Rice [27]. With respect to the organizational performance, we used the scale from Cousins and Lawson [8]. All the items were anchored on a scale from strongly disagree (1) to strongly agree (7). Table 1 provides a sample list of items and the complete list is available from the authors.

Table 1. Sample list of items

Construct	Item	Source
Use of Web 2.0 for Work [9 items]	How frequently do you use Web 2.0 for (e.g. Information seeking/sharing, Office communication, etc.) at work?	[23]
Vertical Communication [4 items]	e.g. My organization has identified/established organizational norms.	[17]
Social Relationship Building [6 items]	e.g. I frequently/regularly exchange information with my network members inside/outside the organization.	[26]
Information Exchange [4 items]	e.g. I have developed a good guanxi network with (my colleagues / external business partners / outside my organization)	[23]
Individual Performance [4 items]	e.g. I am / My team is (confident when undertaking my work / a productive worker / an effective decision maker). My work quality is high.	[27]
Team Performance [4 items]		
Organizational Performance [8 items]	e.g. Increased Market Share / Shareholder Return / Employee Satisfaction, Improved Time to Market / Cash Flows, etc.	[8]

4.2 Data Collection

As the items were established based on the existing literature in English, we followed the guidance of Van de Vijver and Leung [31] for translating items into Chinese. Accordingly, a professional translator, who has no relationship with our research project, was hired to translate the Chinese questionnaire back to English. During the consistency checking process, we found no semantic discrepancies between the retranslated questionnaire and the original English items. Furthermore, we also invited 10 working professionals in China to review and critique the measurement items in order to ensure their content validity before the actual data collection process. We collected survey data using Qualtrics, a professional survey website. In order to identify respondents, we used a customer panel database provided by a marketing research firm in China. In this database, we randomly selected 3000 panel members who were registered as employees. We sent out the online survey hyperlink to the 3000 panel members. We indicated that 30RMB would be provided as a token of appreciation for the survey participation on the first page of the online survey. Over a two month period we successfully reached 2873 email addresses. After removing those voluntary responses with more than 8% missing data and those with either too little or too much time spent in the online survey, we obtained 179 valid responses at the end, yielding a response rate of 6%. We examined non-response bias using the method described by Armstrong and Overton [3]. Specifically, we conducted a t-test of the demographic characteristics of respondents in the first two weeks and in the second two weeks did not significantly differ ($p>0.10$). Therefore these 179 data points formed the data set for subsequent statistical analysis. We summarize the demographic characteristics of these 179 survey respondents in Table 2.

Table 2. Demographic Characteristics (n=179)

Gender	%	Age range	%	Industry Type	%
Male	47	25 or below	58	Agriculture/forestry/fishery/mining	6
Female	53	26–35	31	Construction	8
		36–45	9	Culture, sports, and entertainment	4
		46 and above	2	Electric power, gas, and water production and supply	4
				Financial intermediation	7
Education level	%	Position	%	Manufacturing	7
High school or below	28	Non-Management Employee	57	Hotels and catering services	2
College	28	Manager	15	Information transfer, computer services, and software	18
Undergraduate	35	Senior or Executive Manager	7	International organization	3
Master or above	8	Undisclosed	21	Real estate	3
				Transport, storage and post	4
				Wholesale and retail trades	7
				Others	27

5 Data Analysis

5.1 The Measurement Model

We used SPSS and Smart Partial Least Squares (SPLS) to examine the constructs' validity and reliability. The factor analysis results first confirmed the convergent and discriminant validity of those reflectively measured first-order constructs. Specifically, all the factor loading scores on their expected factors are above 0.69. Furthermore, the own factor loading scores are higher than the cross-loading values. Meanwhile, the eigenvalues of the constructs are all higher than 1.0, with the communality scores above 0.70. As explained in the methodology section, we operationalize horizontal communication as a second-order construct formatively measured by information exchange and social relationship building. We used the factor scores of these two first-level constructs as the composite dimensions of horizontal communication in the SPLS analysis, following the method of handling second-order constructs in SPLS suggested by Petter et al. [25]. As shown in Figure 2, the loading of each composite dimension of horizontal communication was significant. In addition, we also examine construct reliability for all principal constructs with reflective measures by identifying the composite reliability scores, all of which are above 0.90 (as shown in Table 3), suggesting acceptable internal consistency. The square roots of the Average Variance Extracted (AVE) are all above 0.84, which is greater than all other cross correlations. This shows that all constructs capture more construct-related variance than error variance. Taken together, these results demonstrate adequate convergent and discriminant validity for all constructs used in this study.

Table 3. Descriptive Statistics, Correlation Matrix, and Average Variance Extracted.

Correlation Matrix	UOW2	CN	IE	SRD	IP	TP	OP
UOW2: The Use of Web 2.0	-						
CN: Company Norm	0.34	0.71					
IE: Information Exchange	0.30	0.43	0.84				
SRD: Social Relationship Building	0.40	0.60	0.61	0.89			
IP: Individual Performance	0.35	0.45	0.54	0.65	0.86		
TP: Team Performance	0.42	0.54	0.55	0.71	0.78	0.89	
OP: Org. Performance	0.27	0.44	0.39	0.57	0.59	0.73	0.79
Mean	-	4.97	4.35	4.97	5.12	5.17	5.24
STD	-	1.84	1.81	1.70	1.71	1.71	1.74
Composite Reliability	-	0.91	0.95	0.98	0.96	0.97	0.97
Cronbachs Alpha	-	0.86	0.93	0.97	0.94	0.96	0.96

Note: The Use of Web 2.0 was measured formatively. So its mean, standard deviation, average variance extracted (AVE), composite reliability and Cronbach's alphas are not relevant for measuring its construct validity. Diagonal elements above the "mean" row are the AVEs from their indicators. Off-diagonal elements above the "mean" row are correlations between constructs.

Considering that all the data were collected from a single source, we also tested common method bias. First, our principal components factor analysis indicated that each factor explains roughly equal variance. Second, our correction matrix shows that the highest inter-construct correlations are below 0.78, while common method bias is usually evidenced by extremely high ($r=.90$) inter-construct correlations [4]. These tests indicate that common method bias is not a major concern for the current study.

5.2 Analyzing the Research Model

After verifying the measurement model, we then examined the whole structural model in SPLS. The results in Figure 2 suggest the research model is largely supported by the data, except hypothesis H3a. Specifically, the use of Web 2.0 at work has significant influence on vertical communication ($b=0.34$, $p<0.01$) and horizontal communication ($b=0.41$, $p<0.01$) in organizations, with the explained variance of 12% and 17%, respectively, thus validating H1 and H2. Vertical communication is found to significantly impact team performance ($b=0.18$, $p<0.01$) and organizational performance ($b=0.17$, $0.05<p<0.10$), but not individual performance ($b=0.07$, $p>0.10$), therefore supporting H3b and H4b, but rejecting H3a. Our data indicate that horizontal communication has strong direct effects on individual performance ($b=0.64$, $p<0.01$), team performance ($b=0.62$, $p<0.01$) and organizational performance ($b=0.47$, $p<0.01$). Together with vertical communication, horizontal communication explain substantial amounts of variance of individual performance ($R^2=46\%$), team performance ($R^2=54\%$) and organizational performance ($R^2=34\%$). The R^2 scores for all dependent variables in this study, together with the factor loadings, yield an excellent goodness-of-fit for the whole research model [6].

In addition to the proposed research model, we also conducted an additional test for the direct path from the use of Web 2.0 at work to the three performance constructs. However, our analysis indicates that none of the three direct paths was significant. We discuss the implications of these findings in the following section.

6 Discussion, Implications, Future Research and Conclusions

This research has several key findings and implications. Building on the CEF [11], we have conceptualized the communication process as involving vertical and horizontal elements in organizations that can be enhanced with Web 2.0 applications. We also quantify both the benefits of using Web 2.0 and the subsequent communication processes for organizations. Our data suggest that vertical and horizontal communication contribute to individual, team and organizational performance significantly, with explained variances of 46%, 54% and 34% respectively. Such high scores testify to the substantial contribution of communication in the levels of operation at an organization, including individual tasks, teamwork and business processes.

Interestingly, our data analyses indicate that vertical communication contributed less than did horizontal communication to performance. This reflects how contemporary horizontal communication patterns, including information and knowledge exchange, as well as social relationship building, has become the cornerstone for organ-

izations to undertake tasks and run a business at different levels. We provide empirical evidence for how Web 2.0 enables organizations to reach out to collaborators and business partners, thereby boosting individual productivity, team collaboration and organizational agility. Our study also highlights the fit between Web 2.0 and the need for organizational horizontal communication in this era of “Knowledge, Information and Creativity”. This may directly contradict former organizational culture preferences for controlling and centralizing organizational information. We suggest that the instrumental value of Web 2.0 in harnessing and managing personal and public knowledge should not be ignored as it leads to higher levels of performance.

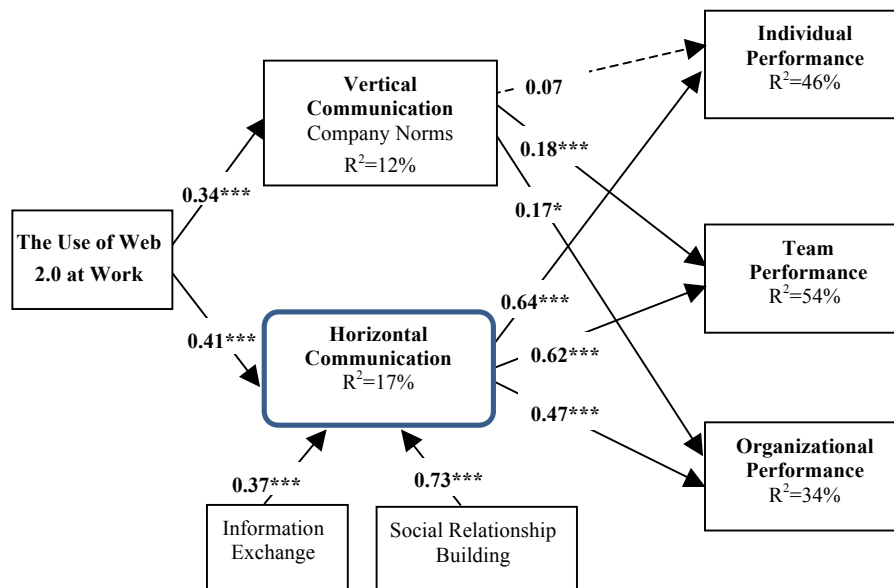


Fig 2. SPLS Results

Note: Square boxes represent the first-order factors. Bevels represent second order factor (Horizontal Communication) formatively measured by two first-order constructs (Information Exchange and Social Relationship Building). Paths in solid lines are significant links (*0.05<p<0.10; **0.01<p<0.05; ***p<0.01). Paths in dotted lines are insignificant links.

However, it is also worth noting that this is not an automatic process. Our mediation test of communication between the use of Web 2.0 and performance suggests that the direct impact of using Web 2.0 on all three levels of performance is not significant: Web 2.0 alone guarantees nothing. Nevertheless, encouraging employees to participate in Web 2.0 based interactions may help enhance team outcomes and thus stimulate organizations to consider how they can engage with employees directly in order to achieve these performance gains.

In addition to the above theoretical and practical implications, this research has several limitations that suggest future research opportunities. First, our data was all collected in China. It would be valuable to verify the research model in different countries, including local contextual characteristics as either independent variables or

moderators. Second, we collected subjective data from the survey participants. Objective performance data could usefully supplement the subjective data with additional measures. Finally, longitudinal data would permit the examination of the impacts of Web 2.0 on different tasks and projects over time.

In conclusion, social media is here to stay – in both personal and organizational contexts. We now need to find how to leverage it effectively so as to drive performance benefits – at the individual, team and organizational levels. This study represents a step in this direction, but we anticipate further investigations in different cultural contexts will identify new constructs. Ultimately, we need to engage in careful theorization of the relationships between constructs testing the theory across a wider range of populations.

7 References:

1. Acar, A., Muraki, Y.: Twitter for Crisis Communication: Lessons Learned from Japan's Tsunami Disaster. *Int. J. Web Based Comms.* 7(3), 392-402 (2011)
2. Altheide, D.L.: An Ecology of Communication: Toward a Mapping of the Effective Environment. *Sociol. Quart.* 35(4), 665-683 (1994)
3. Armstrong J.S., Overton, T.S.: Estimating Non-response Bias in Mail Surveys. *J. Marketing Res.* 14(3), 396-402 (1977)
4. Bagozzi, R.P., Yi, Y., Phillips, L.W.: Assessing Construct Validity in Organizational Research. *Admin. Sci. Quart.* 36(3), 421-458 (1991)
5. Chang, H.J., Ian, W.Z.: Instant Messaging and Interruption in Organizational Settings: A Social Presence's Perspective. *First Monday.* 17(3), <http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/3741/3180> (2012)
6. Chin, W.W.: Issues and Opinion on Structural Equation Modeling. *MIS Quart.* 22(1), vii-xvi (1998)
7. Chui, M., Miller, A., Roberts, R.P.: Six Ways to Make Web 2.0 Work. *McKinsey Quart.*, http://www.mckinsey.com/insights/business_technology/six_ways_to_make_web_20_work (2009)
8. Cousins, P.D., Lawson, B.: The Effect of Socialization Mechanisms and Performance Measurement on Supplier Integration in New Product Development. *Brit. J. Management.* 18(3), 311-326 (2007)
9. Davison, R.M., Ou, C.X.J.: Sharing Knowledge in Technology Deficient Environments: Individual Workarounds Amid Corporate Restrictions, 17th European Conference on Information Systems, Utrecht, The Netherlands, http://aisel.aisnet.org/ecis2013_cr/23 (2013)
10. Davison, R.M., Ou, C.X.J., Martinsons, M.G.: Information Technology to Support Informal Knowledge Sharing. *Inform. Sys. J.* 23(1), 89-109 (2013)
11. Davison, R.M., Ou, C.X.J., Martinsons, M.G., Zhao, Y., Hua, X.Q. and Du, R.: The Communicative Ecology of Web 2.0 @ Work: Social Networking in the Workspace, 73rd Academy of Management Conference, Orlando, USA, August 9-13 (2013)
12. Foth, M., Hearn, G.: Networked Individualism of Urban Residents: Discovering the Communicative Ecology in Inner-city Apartment Buildings. *Inform., Comm. and Soc.* 10, 749-772 (2007)

13. Garrett, R.K., Danziger, J.N.: IM = Interruption Management? Instant Messaging and Disruption in the Workplace. *J. Comp.-Mediated Comm.* 13(1), <http://jcmc.indiana.edu/vol13/issue1/garrett.html> (2007)
14. Gay, C., Mahoney, M., Graves, J.: *Best Practices in Employee Communication: A Study of Global Challenges and Approaches*. IABC Research Foundation, San Francisco (2005)
15. Günther, O., Krasnova, H., Riehle, D., Schöndienst, V.: Modeling Microblogging Adoption in the Enterprise, *Proceedings of the 15th AMCIS*, San Francisco, August 6-9 (2009)
16. Jansen, B.J., Zhang, M., Sobel, K., Chowdury, A.: Twitter Power: Tweets as Electronic Word of Mouth. *J. Am. Soc. Inform. Sci.* 60(11), 2169-2188 (2009)
17. Lewis, J., Wright, P.C., Geroy, G.D.: Managing Human Capital: The Study of a Self-managed Group Venturing into Digital Economy. *Management Decision*. 42(2), 205-228 (2004)
18. Li, J., Rao, H.R.: Twitter as a Rapid Response News Service: An Exploration in the Context of the 2008 Sichuan Earthquake. *Elec. J. Inform. Sys. Developing Countries*, 42(4), 1-22 (2010)
19. Martinsons, M.G., Westwood, R.I.: Management Information Systems in the Chinese Business Culture: An Explanatory Theory. *Inform. & Management*. 32(5), 215-228 (1997)
20. McLuhan, M.: *The Gutenberg Galaxy: The Making of Typographic Man*. Routledge and Kegan Paul, London (1962)
21. McPhee, R.D.: Vertical Communication Chains: Toward an Integrated Approach. *Management Comm. Quart.* 1(4), 455-493 (1988)
22. Olasina, G.: The Use of Web 2.0 Tools and Social Networking Sites by Librarians, Information Professionals, and Other Professionals in Workplaces in Nigeria. *PNLA Quart.* 75(3), 1-40 (2011)
23. Ou, C.X.J., Davison, R.M.: Interactive or Interruptive: Instant Messaging at Work. *Decision Support Systems*. 52(1), 61-72 (2011)
24. Ou, C.X.J., Davison, R.M., Zhong, X.P., Liang, Y.: Empowering Employees through Instant Messaging. *Inform. Tech. and People*. 23(2), 193-211 (2010)
25. Petter, S., Straub, D.W., Rai, A.: Specifying Formative Constructs in Information Systems Research. *MIS Quart.* 31(4), 623-656 (2007)
26. Phang, C.W., Kankanhalli, A., Sabherwal, R.: Usability and Sociability in Online Communities: A Comparative Study of Knowledge Seeking. *J. Assoc. Inform. Sys.* 10(10), 721-747 (2009)
27. Rice, R.E.: Task Analyzability, Use of New Media, and Effectiveness: A Multi-Site Exploration of Media Richness. *Org. Sci.* 3(4), 475-500 (1992)
28. Riemer, K., Richter, A., Behringer, M.: Enterprise Microblogging: Procrastination or Productive Use? *Bus. Inform. Sys. Eng.* 2, 391-394 (2010)
29. Smidts, A., Pruyn, A.T.H., van Riel, C.B.M.: The Impact of Employee Communication and Perceived External Prestige on Organizational Identification. *Acad. Management J.* 44(5), 1051-1062 (2001)
30. Tacchi, J., Slater, D., Hearn, G.: *Ethnographic Action Research: A User's Handbook*. UNESCO, New Delhi (2003)
31. Van de Vijver, F.J., Leung, K.: *Methods and Data Analysis for Cross-Cultural Research*. Newbury Park, Sage, CA (1997)
32. Williams, P.: *Employee Communication: The Comprehensive Manual for those who Communicate with Today's Employees (3rd Ed.)*. Lawrence Ragan, Chicago (2008)
33. Zhang, H.C.: *Microblog Marketing in China*. Master of Science Thesis in Media Technology, Royal Institute of Technology, Stockholm, Sweden (2011)

Educators as Serious Game Designers – A Model-Driven Visual Programming Framework

Niroshan Thillainathan, Jan Marco Leimeister

Information Systems, Kassel University, Pfannkuchstr. 1, 34121 Kassel
{thillainathan, leimeister}@uni-kassel.de

Abstract. Students use their own mobile devices like laptops, smartphones and tablets to work and study anywhere and anytime. Consequently, there is a high interest in using one's own mobile devices in the educational context. This presents a chance as well as a challenge for educators, who can now make use of mobile learning for supplying their students with tailored didactical content that is accessible anywhere and anytime. Therefore, in our work, we focus on presenting educational content using serious games. In order to allow educators without technical knowledge to create custom serious games, we propose a model-driven visual programming framework. In this paper, we present our framework, which consists of GLiSMo - a domain-specific modeling language - and the visual programming environment for serious games (VIPeR) that allows educators to use their didactical knowledge with a point and click graphical editor to create their own serious games.

Keywords: visual programming · domain-specific modeling language · model-driven development · serious game

1 Introduction

In recent years, mobile computing has been one of the fastest growing areas of the technology industry worldwide. With innovative technologies inside small and portable devices a new generation of applications in various environments is rising. The fact that mobile devices, such as smartphones, tablets and notebooks, are more affordable, accessible and easier to use than desktop computers, makes their use more appealing for the younger generation, who include them more and more into their everyday life. This has also been noted by the NMC Horizon Reports for Higher Education [1, p. 7].

Pearson and Harris Interactive have conducted a survey among 1206 college students in the U.S. Their results show that 91% of the students use their laptops, 72% their smartphones and 40% their tablets regularly. Out of these students, 96% have used a laptop, 66% have used a smartphone and 83% have used a tablet for educational work. From the learners' perspective, this shows that students are eager to use their mobile devices in the educational context [2]. From the educators' perspective, mobile computing enables educators to fulfill both the requirements noted by Johnson et al. [1] as well the learners' expectations and thus support the learning process [3].

This resulting educational paradigm of mobile learning is a special form of the established e-learning and has gained widespread attention as one way to enable students to learn anywhere and anytime. The specific type of mobile learning tool in our work are so-called serious games. Serious Games differ from traditional e-learning or mobile learning approaches as they combine education with gameplay, by presenting learning objectives and contents in a game-like context. By providing a game-like structure, the learners stay motivated and continue using the game, and, at the same time, they have didactical content conveyed to them. So, the more fun the game is, the more the student is motivated to play and learn and thus has a higher learning rate [4].

While serious games have been shown to hold great potential as a didactical tool, developing them for specific learning objectives involves several challenges for the educators who wish to apply them. They have the didactical knowledge about what and how they want to teach. They typically do not have the knowledge needed for the technical realization, though and need support for developing the serious game for diverse mobile devices. In this paper, we present our model-driven visual programming framework, which enables non-technical educators to create simple serious games based on the “point and click” graphical adventure genre. Therefore, we first present some serious game development approaches and describe our research methodology. Next, the model-driven visual programming framework and its three components are introduced. The remainder of this paper depicts an experimental modeling and implementation of a serious game, an evaluation of the framework followed by a conclusion and future work.

2 Foundation: Serious Game Development Approaches

Creating a serious game is nearly impossible for most educators due to lack of programming skills and existing serious games generally do not allow the adaption of the content to their own learning objectives. Furthermore, serious game development not only consists of programming the software, but also designing the game world and game logic, which also is not the expertise of most educators. Fig. 1 shows the process of serious game development by educators. As highlighted, the educator needs support during the game world development, game logic creation and software code generation. Therefore, educators have two options. The first option is to hand over the development process and providing the didactical content, to a professional game developer, which comes at high development cost [5]. We however argue for the second option: the educators develop serious games by themselves, supported during game development and the implementation process by a serious game development framework.

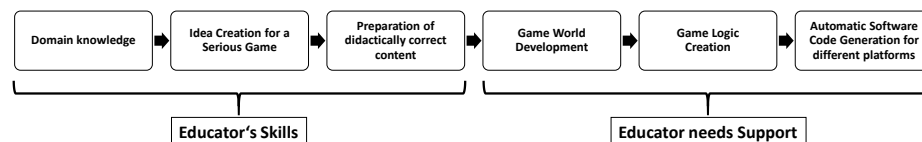


Fig. 1. Serious Game Development Process

Since it is not feasible to turn non-technical educators into programmers – learning a programming language is highly complex [6] – we consider the so-called visual programming approach as the way to support educators. Visual programming languages (VPL) have been shown to be easy to understand and help non-programmers create software [7]. A VPL implementation consists of a visual programming environment (VPE), which represents the user interface to create application models, a syntactic and semantic validator for these models, as well as a compiler to generate software code from them [8]. To support a wide variety of diverse mobile devices, we employ a model-driven development process as the underlying principle of the VPL. Model-Driven Development (MDD) is a methodology in software engineering, which combines techniques for automated generation of software code from formal models. To formalize the application structure, the behavior and its requirements, formal languages called domain-specific modeling languages (DSML) are designed and implemented for a specific domain [9]. The VPE is used to create game models and the game world, which are then transformed into different programming languages for specific platforms using the model-driven development approach.

In current literature, only very limited research concerning development tools for serious games can be found. For our own research, we have identified two tools that allow the development of customized serious games, Sealund's Serious Games Engines and Worldweaver's DX Studio. *Sealund's Serious Games Engines* [10] is suited to developing 2D board games, with multiple choice questions. The educator has the possibility of selecting the type of game and integrating a predefined number of questions. The main disadvantage, however, is that this tool does not allow modeling the logic of a serious game. It is particularly desirable to develop games with individual behavior instead of predefined logic. *DX Studio* [11] provides a development environment for 3D- and learning applications. It allows creating a game environment consisting of objects and characters within the world. This tool lacks the possibility of creating the logic and behavior of the game easily and intuitively. To do so, the developer must have knowledge in JavaScript programming.

The presented tools have several weaknesses and disadvantages. Either the existing approaches are lacking in functionality or are too complex to be used by non-technical domain experts. Therefore, this induces the need for a solution for the development of serious games, which allows educators to develop their own serious games, without technical or game design knowledge. We want to enable educators to develop serious games adapted to their needs by presenting an environment that is intuitive and easy to handle.

3 Methodological Approach

To provide the framework for educators to create working serious games for their students without having to write software code, we apply a design-oriented research approach. In design-oriented research, artifacts, i.e. artificial, man-made things, are specially crafted in order to fulfill certain purposes and goals given their functions and adaptability. A design-oriented approach allows researchers to assess the impact of their

artifacts and check whether they were able to achieve intended effects. Among the list of possible artifacts in IS research compiled by March & Smith [12] there are constructs, models, methods and instantiations. The framework we describe in this paper is comprised of constructs – the conceptualization used to describe the problem of creating mobile serious games by educators without technical background – as well as instantiations – the realization of artifacts in their target environment – in form of the concept and realization of our visual programming editor (VIPer), the modeling language for serious games (GLiSMo) and the model-driven development Toolchain.

In contrast to approaches found e.g. in the social sciences, which are well-defined and broadly accepted as rigorous, a corresponding universal approach for design-oriented research does not exist. However, there are notable similarities between phases of design-oriented research found in the key literature on design science (e.g. March and Smith [12], Takeda et al. [13]) that allow deriving a basic consensus on a basic structure for design-oriented research.

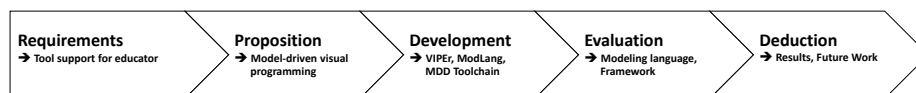


Fig. 2. Process steps followed in this research (adapted from Takeda et al. [13])

The first step in design-oriented research is to identify the problem and its context, as such research follows a problem-solving paradigm. Taking this as input, a scenario for a possible solution is designed. This includes both the definition of the goals for the artifacts as well as the desired effects of the artifact. In a third step, the artifact or artifacts are realized and put to use in the context they were designed for. Artifact evaluation in the fourth step is based on the changes that can be observed concerning the problem and its context. From this evaluation, conclusions are drawn and documented, e.g. as new input for another iteration of the process or as the basis to theorize. Fig. 2 shows the research process adapted from Takeda et al. [13], which we follow in this contribution.

4 A Framework for Model-Driven Visual Programming of Serious Games

As described in the introduction our aim is to present a solution for educators, which allows the development of mobile serious games without having knowledge in game development or game programming. Fig. 3 shows an overview of our framework, which supports the educator during development by offering a visual programming interface and creates the game for multiple different mobile computing platforms by employing a model-driven development approach. Our framework can be divided into the three abstract components visual programming, domain-specific modeling language and model-driven development Toolchain (MDD Toolchain).

In the visual programming component, we have the visual programming environment for serious games, which functions as the user interface for the educators. Here, the

educators are able to design and develop the serious game world in an intuitive manner. To create the logic and behavior of the game, we use a domain-specific modeling language, which mediates as an interface to connect the visual programming and the MDD Toolchain component. Within the visual programming component, the DSML is used to model the logic of the game. In the following step, the DSML is used for validating the created models in the MDD Toolchain. After the validation process the software code of the serious game will be generated for different platforms, like mobile devices, web browsers or desktop computers.

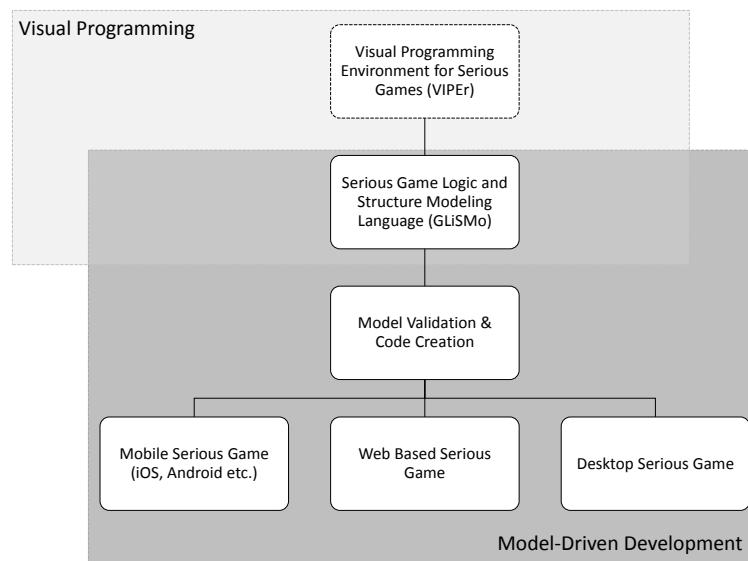


Fig. 3. Model-driven visual programming framework for serious games

In the subsequent sections we present the three major components of your framework, beginning with the core DSML, followed by the visual programming environment and MDD Toolchain.

4.1 Serious Game Modeling Language (GLiSMo)

In order to model a holistic serious game, a DSML tailored for serious games is needed. Therefore, we have created the Serious Game Logic and Structure Modeling Language (GLiSMo) as the underlying domain-specific modeling language, which allows designing both the structure as well as the logic of a serious game. The game structure describes the world the serious game is set in. It contains the layout of the game world, which characters and objects are included, and how the user/player interaction takes place. The logic characterizes the behavior of characters and objects in terms of how does the game react on specific actions performed by the player, or events occurring during game play. The logic also describes the assessment of player actions and the game adaption according to the adaption results. A detailed description of our domain-specific modeling language including the introduction of each language element can be

found in our previous publication [14]. Examples of GLiSMo models are presented in section 5.2.

4.2 Visual Programming Environment for Serious Games (VIPer)

While the Serious Game Modeling Language enables developers to describe a serious game in abstract terms, i.e. in form of a model, it does not ease serious game creation for educators without a technical background. In order to make model creation feasible for this target group, we have designed the Visual Programming Environment for Serious Games (VIPer) to provide educators with a point-and-click tool for model creation (see steps 4 & 5 in Fig. 1). As VIPer can be regarded as a frontend for creating valid and meaningful models for the game's structure and logic, the development of VIPer and GLiSMo are tightly linked.

As there already exists a plethora of development environments, we have chosen to extend an existing tool that is already in use rather than developing an integrated development environment (IDE) from scratch. For our purposes, we have chosen to create VIPer as a plugin for the cross-platform game engine Unity, an established tool for game creation that already includes a built-in integrated development environment (IDE). VIPer is especially made to develop point and click graphical adventure serious games in a simple manner. This game genre was chosen as the slow pace during game play, the possibility to study the environment and the fact that all tasks are based on problem solving are characteristics, which are relevant from a didactical point of view. These characteristics increase the learning success while playing [15]. We use GLiSMo as the underlying DSML with which the game logic and behavior can be created as intuitive as possible. To support this, we have taken good care while designing the clearly arranged and user-friendly UI. Our focus is on displaying only the required elements within the user interface to make sure that educators can use VIPer intuitively without frustration. At the same time, our intention is to relieve as much development work as possible from the educator. An example for this is that we have designed a room generator, which allows creating a given number of interconnected rooms automatically within the game world.

Our visual programming environment supports both game structure and logic development, as specified by GLiSMo. It is divided into game world design (room creation, placement of objects and characters) and game logic design (connection of individual tasks and actions in an editor), thus not only creating rooms and objects but also following the game logic happening directly in the player's view. By this, we want to make sure, only the most necessary information needs to be entered by educators.

As shown in Fig. 1, the educator needs support during the game world development, game model creation and software code generation. We have taken this as the basis to design the game development process in VIPer. Developing a serious game with VIPer needs only four simple steps. In the first step, the educator generates a basic structure of the level using the room generator. This is done by entering a level name, the desired number of rooms and by selecting a predefined setting (office, school, etc.). As an output of this process, a game world with the defined number of rooms will be automatically generated, having the associated doors in between the rooms. The second step

consists of placing all needed objects, characters and decorations into the game world. The educator has to insert default characters and objects via drag & drop. Furthermore, it is crucial to set one of the selected characters as the main player character, which will be followed by the main camera during gameplay. At this point, the structure of the serious game, thus the game world with its objects and characters, is built. These two steps cover the game world development from Fig. 1. To support the game model creation in step 3 the logic of the serious game is designed. As provided by GLiSMo the logic describes the behavior of the game by simply using the following elements: (i) actions, (ii) tasks and (iii) assessments. These elements are arranged by dragging & dropping into the visual programming interface. For example to create a conversation between two characters, the associated discussion element will be positioned into the visual programming interface, followed by entering the text of the conversation.

During this development process, the educators can test the game by using the play button. Finally, the game can be built by clicking on the build button, which starts the compiling process and ends with the final playable game. Both actions trigger the functions of the third component of our framework, the model-driven development Toolchain.

4.3 Model-driven Development Toolchain

In the introduction, we have described model-driven software development as a methodology in software engineering, which consists of techniques for automated generation of software code from formal models. These techniques include domain-specific modeling languages, such as GLiSMo in our case, and model validators and code generators. The model validator analyzes and checks the developed models for correctness, whereas the code generator outputs the equivalent software code. Changes made to the software will be done by updating the model [9].



Fig. 4. Process from a GLiSMo Model to the final software code

As presented in Fig. 3, our framework for serious games development contains a model-driven development Toolchain component, which consists of a model parser, model validator and a code generator. This component is connected with the visual programming component over GLiSMo as the interface. Fig. 4 gives an insight into the process within this MDD Toolchain, which is responsible for the transformation of GLiSMo models to the final serious game software code. The starting point in this process are GLiSMo models of the structure or logic of a serious game, which will be designed by the educators visually with VIPEr. These models will internally be parsed and as an output of this parsing process, a semantic model will be generated, which describes the meaning of the GLiSMo models. These semantic models will subsequently be validated and checked for correctness, which results in validated models.

Based on these models the associated software code will be generated for various platforms, i.e. mobile, web based and desktop devices.

The parsing algorithm as well as the validator are developed by us and are components of our serious game development framework. However, the code generator consists of two parts responsible for the two types of GLiSMo models. The first part generates the software code for the serious game world, which is mainly automatically done by Unity itself. The second part is responsible for the code generation of the game logic, which is programmed by us and is also part of our framework.

5 System Evaluation – Experimental Modeling and Implementation of a Serious Game

Our initial system evaluation of the proposed framework has been twofold. In an advanced modeling language evaluation, we conducted an expert review of the crucial modeling approach to ensure its feasibility before testing the whole framework with the guided realization of a serious game together with instructors and domain experts.

5.1 Advanced Modeling Language Evaluation

For a thorough evaluation focusing on the crucial serious game modeling language, we presented the current version of GLiSMo to four software engineering experts, from which two also are considered expert educators in higher education. During this step, we asked the evaluators to provide feedback on the modeling language, both from a technical standpoint, i.e. the technical feasibility of the modeling language as a basis for the visual programming environment as well as the model-driven development approach, as well as a game design standpoint, especially taking into account serious games for education. From the feedback we have received by observing their approach of analyzing the modeling language as well as the explicit comments from the experts, we could conclude that the current version of GLiSMo allows modeling the structure as well as the logic of the subset of serious games we aimed to support with the visual programming environment (VIPeR). While this would be a limitation for a comprehensive modeling language, this also indicates that no unnecessary features are included in the present form of GLiSMo.

5.2 Framework Evaluation

After the successful evaluation of the modeling languages feasibility for our purposes, we performed a guided realization of a serious game, called Shack City 3D, for apprentices in the field of sanitation, heating and cooling (SHaC) together with instructors and domain experts from that field. In this guided realization, we provided an initial idea for a serious game based on previous interviews with instructors in SHaC indicating the need for educational content covering real-life scenarios for apprentices (Step 1). We subsequently asked the instructors and domain experts to provide their didactical content, describe their goals for the game world as well as its behavior and model it in

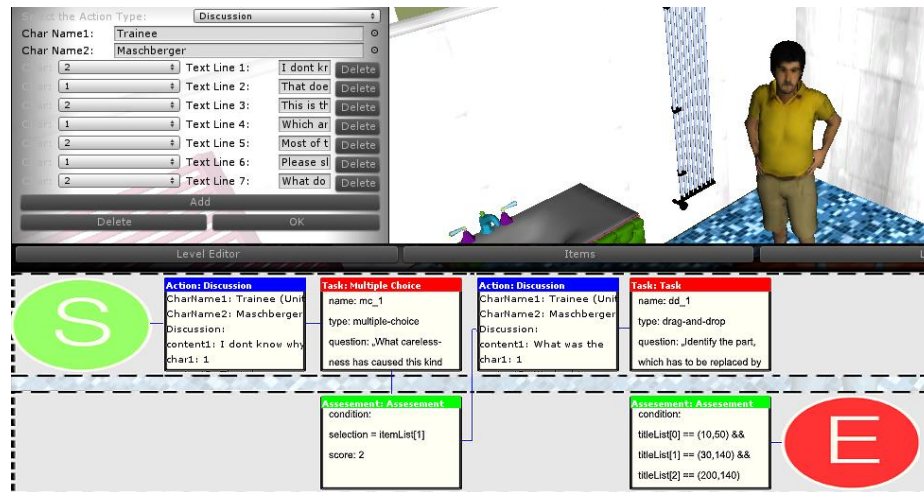


Fig. 5. Logic modeling interface in VIPer

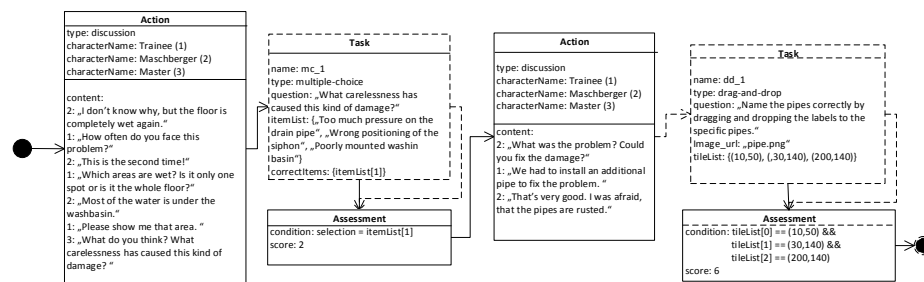


Fig. 6. Excerpt of the logic model of Shack City 3D

GLiSMo using VIPer (Step 2). The results were then compiled into a web-based game using the MDD Toolchain for instructors and domain experts to review (Step 3).

As the result of Step 1, the basic idea of the game is about the main character, a SHaC apprentice, who accompanies his foreman to different customers within a city to solve their problems with their SHaC. This is also supported by a story, which leads the player through the game. At each customer's place at first a dialogue between the apprentice and customer is shown, where the customer describes his problem in a novice manner, as in real life. Based on the information the apprentice has to find out solutions for the problem by solving different tasks. Possible types of tasks are multiple-choice questions, highlighting certain previously defined spots on objects and images, combination of certain objects or filling missing parts in images. After solving such subtasks there will be a conversation with the customer, which describes further problems. This will be alternated until all tasks are solved. At the end, the player gets detailed information and feedback about the solved tasks and learning progress. In Step 2, we derived

model elements from the overall game idea, combined them with more detailed descriptions of instructors and their didactical content. This was then used to model both the structure and logic in GLiSMo using VIPER. We performed multiple iterations of this modeling together with instructors, incorporating their expert feedback in each new iteration.

Fig. 6 shows the excerpt of the logic model in GLiSMo, which describes an interaction sequence with the customer “Maschberger”. This same interaction sequence is then modelled within VIPER by visually programming the logic as shown in Fig. 5.

The structure of the logic model in this game consists of an alternating sequence of conversations, tasks and assignments. To be precise, the conversation deals in this example with the problem context, where the customer describes what has to be repaired by the apprentice. With the support of an interesting background story, the player gets further details about the problem context. During these dialogues, the player will face tasks, which have to be solved by the player, followed by the assessment by giving feedback and scores. Among the world elements hence are the rooms, in which the conversations and tasks take place, as well as a player figure, representing the apprentice and non-player characters (NPC) for the customers providing the tasks. The resulting structure model in this case consists of following elements: (i) serious game root, (ii) different acts representing customer orders, (iii) scenes belonging to the acts and (iv) characters. In Fig. 7, an excerpt of the GLiSMo structure model of Shack City 3D is presented for one customer named “Maschberger”. The root element directly connects to the act, which contains two scenes. Based on the story of the game, the player can move to these different locations and solve tasks. Besides the character representing the apprentice played by the player, there is also a NPC acting as the instructor.

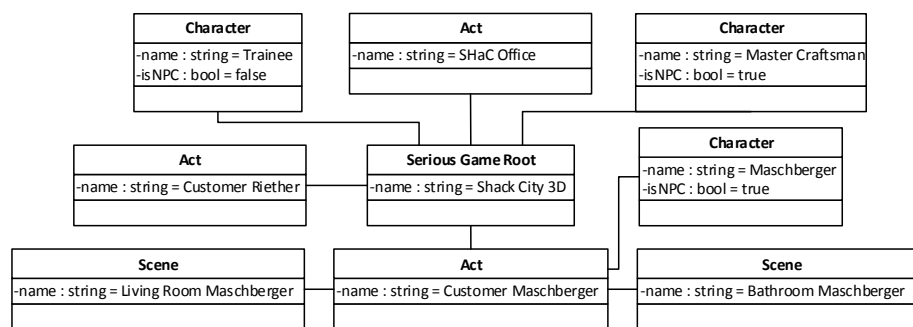


Fig. 7. Excerpt of the structure model of Shack City 3D

As the character elements can appear in all parts of the game, they connect to the serious game root element. However, the NPC representing the customer connects directly to the act as this character only appears in this act. In Step 3 of the guided evaluation, the first prototype of Shack City 3D, based on the models created in the step before, can be tested using the MDD Toolchain. As described by the logic model, a typical situation at a customer alternates between discussions and tasks. A multiple-choice task is given, which follows a discussion. The player has to select the correct answer, which then will be evaluated by the game based on the content of the associated assessment element

described in the logic model. First user tests with apprentices in the field of SHaC have been conducted together with their instructors, evaluating the didactical usefulness of the serious game. The apprentices have agreed that a serious game like this is a promising approach to motivate them for learning in casual settings. Furthermore, they have described the game as interesting and could imagine using such games during their studies as a supporting learning material. Their instructors and domain experts confirmed that the game built matches what they would like to provide for their apprentices when creating a serious game. They expect that the apprentices could benefit from using such serious games, by achieving better learning results and enjoying studying more at the same time.

6 Conclusion and Future Work

From the evaluation of our initial proof of concept, the idea of supporting educators to design and develop serious games through visual programming and model-driven development techniques appears to be a very promising approach. While educators have the domain knowledge and are able to create ideas for a serious game and prepare didactically correct learning content for them, they need support during the serious game design and development process, consisting of game world development, game model creation and software generation. We hence presented a framework that can be broken down into three abstract components: (i) visual programming, (ii) domain-specific modeling language and (iii) model-driven development Toolchain. Component (i) is realized in our visual programming environment VIPeR. VIPeR allows educators to apply their skills to come up with an idea and the content for a serious game and create a model of both via a drag-and-drop style interface and without the need for any technical knowledge. The modeling language we have designed for component (ii), GLiSMo, serves as interface and glues between the visual programming paradigm for the educator and the model-driven development to support a wide range of end devices. Finally, component (iii) takes the model representation of the user input, validates it and generates code for multiple mobile platforms from it. We have evaluated the feasibility of this component-based framework and our instantiation of tools for each component by designing a serious game called Shack City 3D. During this evaluation, our main focus lies on GLiSMo as the integral interfacing part between the two worlds of visual programming and model-driven development. The results from our evaluation show the feasibility and applicability of GLiSMo for modeling Shack City 3D and the ability to generate software code from GLiSMo models for different mobile devices.

A first evaluation has shown that GLiSMo is suitable for modeling the structure and logic of serious games. Furthermore, the prototype of Shack City 3D, the outcome of our first serious game created using our framework, has been considered as useful by apprentices and educators from the field of SHaC. Having established the feasibility of GLiSMo, the next evaluation will focus on the detailed evaluation of VIPeR as a visual development environment for educators, i.e. performing non-guided development experiments with instructors. Finally, we would also like to evaluate the serious games further, which are the final product of the educator's development process. This has to

be done both with the educator, i.e. to determine whether the serious game resembles what they anticipated to create, as well as with potential users in order to determine whether they reach the potential outlined by Brennecke and Schumann [4] as well as Pivec [16].

References

1. Johnson, L., et al., The NMC Horizon Report: 2013 Higher Education Edition. 2013.
2. Pearson, Pearson Student Mobile Device Survey 2013 National Report: College Students. 2013, Pearson.
3. Martin, S., J. Peire, and M. Castro. M2Learn: Towards a homogeneous vision of advanced mobile learning development. 2010. IEEE.
4. Brennecke, A. and H. Schumann. A General Framework for Digital Game-Based Training Systems. in Int. Conf. Game and Entertainment Technologies. 2009. Algarve.
5. Younis, B. and C. Loh. Integrating serious games in higher education programs. in Academic Colloquium. 2010. Ramallah.
6. Gomes, A.J., A.N. Santos, and A.J. Mendes, A study on students' behaviours and attitudes towards learning to program, in 17th ACM Conf. on Innovation and Technology in Computer Science Education. 2012: Haifa. p. 132-137.
7. Myers, B.A., Taxonomies of visual programming and program visualization. Journal of Visual Languages & Computing, 1990. 1(1): p. 97-123.
8. Kang, Z., D.Q. Zhang, and C. Jiannong, Design, construction, and application of a generic visual language generation environment. IEEE Trans. on Software Engineering, 2001. 27(4): p. 289-307.
9. Schmidt, D.C., Model-driven engineering. Computer, 2006. 39(2): p. 25.
10. Sealund. Sealund Serious Games - Serious Games Engines. Undated [cited 2013 16.08.2014]; Available from: <http://www.sealund.com/sgengines.php>.
11. Worldweaver. DX Studio. 2003 [cited 2013 19.08.2013]; Available from: <http://www.dxstudio.com>.
12. March, S.T. and G.F. Smith, Design and natural science research on information technology. Decis. Support Syst., 1995. 15(4): p. 251-266.
13. Takeda, H., et al., Modeling design processes. AI Mag., 1990. 11(4): p. 37-48.
14. Thillainathan, N., H. Hoffmann, and J.M. Leimeister. Shack City – A Serious Game for Apprentices in the Field of Sanitation, Heating and Cooling (SHaC). in Informatik 2013. 2013. Koblenz.
15. Van Eck, R., Building Artificially Intelligent Learning Games, in Intelligent Information Technologies: Concepts, Methodologies, Tools, and Applications. 2008, IGI Global. p. 793-825.
16. Pivec, M., Play and learn: potentials of game-based learning. British Journal of Educational Technology, 2007. 38(3): p. 387-393.

First year students' algorithmic skills in tertiary Computer Science education

Piroska Biró¹, Mária Csenoch¹, Kálmán Abari² and János Máth²

¹University of Debrecen, Faculty of Informatics, Debrecen, Hungary
{biro.piroska, csernoch.maria}@inf.unideb.hu

²University of Debrecen, Faculty of Arts and Humanities, Debrecen, Hungary
{abari.kalman, math.janos}@arts.unideb.hu

Abstract. Faculties of Informatics are facing the problem in Hungary that students starting their tertiary education in Computer Sciences do not have a satisfactory level of algorithmic skills, their knowledge seems superficial, and the dropout rate is extremely high in these courses. We have launched a project to test how students' algorithmic skills have been developed in their primary and secondary education, how students evaluate their knowledge. The test proved that an extremely high percentage of the students arrive at the Faculty of Informatics with underdeveloped algorithmic skills, with unreliable knowledge, and they do not consider the recently emerged, non-traditional environments as programming tools and facilities for developing algorithmic skills.

Keywords: algorithmic skills, first year students of Informatics, programming tools

1 Introduction

Over recent years Computer Sciences and Informatics (CSI) education has become accepted; in basic curricula digital competency has appeared alongside traditional competencies, and beyond that CSI has emerged as a separate school subject. However, in spite of the great expectations regarding this new science, students arriving into tertiary CSI education do not show great confidence in algorithmic skills, and consequently in digital competency. It seems that they mostly navigate the GUI (Graphic User Interface) unplanned, unaware of their activities, trying to solve computer related problems without algorithms.

A large number of studies have been published in the last two decades focusing on the methods adopted to teach programming, to develop algorithmic skills, and to measure students' knowledge [4,8,9,12]. To extend these studies we launched the TAaAS project (Testing Algorithmic and Application Skills) in the 2011/2012 academic year at the University of Debrecen, Hungary [2,3,4], in which 800 students were tested. The project focuses on the level of the students' algorithmic skills, their usage of terminology, and their problem solving abilities in traditional and non-traditional programming environments at the beginning of their tertiary studies.

adfa, p. 1, 2011.

© Springer-Verlag Berlin Heidelberg 2011

2 Methods

The three programming tasks compared in this study differ in terms of their environments, and consequently in the way they were presented to the students. Task 1 is designed to test the logical operators in a multilevel IF structure. Along with the source code a matrix is presented with nine pairs of inputs and the empty cells for the outputs, which are one of the four numbers from the algorithm.

In Task 2 three pseudo codes – Task 2.1, 2.2, and 2.3 – are given with background information on the environment. The results of the task are three natural language sentences to describe what the codes do. This task, beyond the decoding, requires the ability to “translate” the code of an artificial language to a natural language in the presented environment [4,8].

Task 3 is a multilevel spreadsheet function, an array formula. It tests how the students decode formulas, understand loops and logical operators in spreadsheet environment to Task 2, the solution should be given in a natural language sentence. The task is included to see how students would be able to handle programming task presented in a non-traditional programming environment.

Beyond the evaluation of the students’ results in the above mentioned three tasks, their self-assessment values in programming were recorded.

3 Hypotheses

The collected data and the statistical analyses of the TAaAS project can provide further information on the students’ background studies in CSI. The following hypotheses were checked with the analyses.

H1: The order of difficulty of the pseudo-codes is: Task 2.1, 2.2, 2.3, from the least to the most difficult, respectively.

H2: The students arrive in tertiary education with a low and unsatisfactory level of algorithmic skills. Their knowledge is haphazard and non-viable.

H3: The students’ uncertain knowledge of the subject can be differentiated: (1) a group of students with a very limited but quite convincing knowledge until they reach their limits, and (2) another group whose knowledge is somewhat arbitrary.

4 Results

4.1 Knowledge-based groups of students

Task 1 was found to be the most successful among the three tasks, so it was used to define clusters of students for the further analyses. The regrouping of the students was based on nine variables – the nine questions of Task 1 – and the four optional answers of this task – 0, 1, 2, 3. These values served as categorical data for the Two-step Cluster Analysis in SPSS. Based on these premises, 3+1 clearly distinguishable clusters were found.

Those students are in Cluster 1 who solved the task with an excellent result (99.7%, **Table 1**). In Cluster 2 there is an increase in the number of wrong answers, but this is still lower than the percentage of the right answers, which is 64.7%. The students in Cluster 3 answered those questions correctly which do not have 0 input. These students answered 0 if one of the inputs was 0. It seems that they found that the code works for inputs A and B, but with 0 input the only output was 0. This can be explained by the false interpretation of the algorithm, which does not provide an explicit output for all the 0 inputs.

The detailed results of Task 1 with the three recognized clusters are mapped in **Fig. 1** (left). The titles of the graphs show the order of the questions and the input values, the X-Y pairs of the algorithms to each question, while the columns the percentage of the students' correct answers in each cluster for the nine questions. The right answer to each question can easily be followed if we track the columns of Cluster 1 (black), since these students answered all the questions correctly. The gray and the white columns show the results of Clusters 2 and 3. As was mentioned in the previous section, the answers to Cluster 3 are 0s if one of the inputs is 0, $X=0$ and/or $Y=0$.

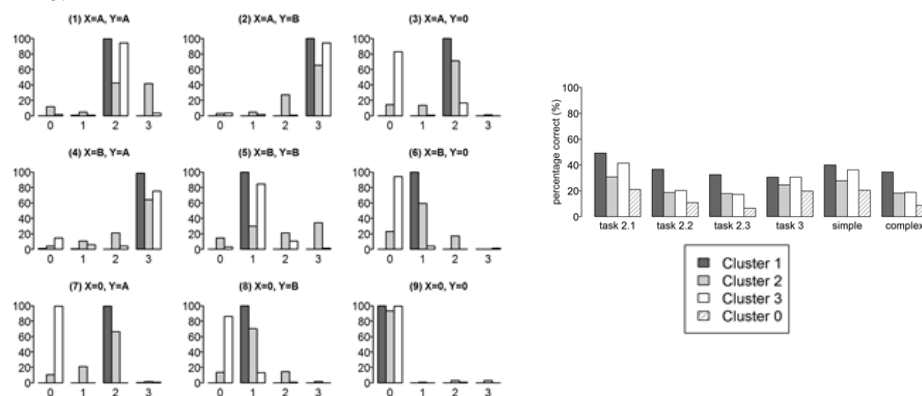


Fig. 1. The percentage of the answers of the 3 clusters in Task 1 (left), in Tasks 2 and 3 in the 3+1 clusters (right)

Table 1 presents the results of the three tasks of the 3+1 clusters. In Cluster 0 are students who did not attempt or complete all the nine answers in Task 1. It is clear from the data that Cluster 1 has the highest scores in all the three tasks, followed by Cluster 2 and 3, while Cluster 0 scored the lowest.

As was mentioned at the beginning of this section, the regrouping of the students was based on the results of Task 1; however, it is noteworthy that the results of Task 1 and the self-assessment values do not follow the same pattern. This can be explained by the major characteristic of the task: the output of the task is neither a source code nor a description of code, but numbers of four choices, and as such, is different in nature from traditional programming tasks. It will be obvious from the further analyses (section 4.3) that it is not Task 1 but 2 which fills the advisory role for the self-assessment values.

Table 1. The self-assessment values and the test results in the 3+1 clusters of Task 1

	Cluster 1	Cluster 2	Cluster 3	Cluster 0
N	213	134	122	331
self-assessment (%)	45.4	34.8	35.1	36.2
Task 1 (%)	99.7	64.7	53.7	0.0
Task 2 (%)	39.5	22.5	26.4	12.9
Task 3 (%)	30.7	24.6	30.6	19.8

4.2 Connections between the three tasks

Task 2 is divided into three subtasks, which are the three pseudo-codes. Task 3 is one multilevel function consisting of an AND connection and a counting of those objects which fulfill both of the requirements. Using the 3+1 clusters of Task 1 the results of Tasks 2 and 3 are presented in **Fig. 1** (right). It is clear from the percentage of the correct answers mapped in the graphs that the difficulty of the three pseudo-codes increases from Tasks 2.1 to 2.3 (percentage correct: 33.3%, 20.5%, 17.1%, respectively), which proves the H1 hypotheses, and Task 3 is as difficult as Task 2 in general (25.2%, 23.6%, respectively). Similarly to Task 1, the results of Cluster 1 (black) are the highest in these two tasks, Clusters 2 and 3 performed worse, and Cluster 0 has the lowest scores.

In the following the focus is on the connections between Task 1 and the other two tasks. In the comparison the differences in the results of the 3+1 clusters are examined. Four Kruskal-Wallis tests proved that there is a significant difference in the four clusters in terms of the percentage correct in Tasks 2.1, 2.2, 2.3 and 3 ($\chi^2_{2,1}=68.1$, $df_{2,1}=3$, $p_{2,1}<0.001$; $\chi^2_{2,2}=73.4$, $df_{2,2}=3$, $p_{2,2}<0.001$; $\chi^2_{2,3}=90.7$, $df_{2,3}=3$, $p_{2,3}<0.001$; $\chi^2_3=14.3$, $df_3=3$, $p_3=0.002$) compared to Task 1, which proves the relationship between Task 1 and the other four tasks – Tasks 2.1, 2.2, 2.3 and 3.

Regarding the complexity of the tasks, two artificial tasks were created: simple and complex. Tasks 2.1 and 3 were grouped as simple, while Tasks 2.2 and 2.3 as complex. The averages of the two artificial tasks were calculated from the averages of the composing tasks in each cluster (**Fig. 1**, right).

The results of Cluster 3 in Task simple and complex are the most noticeable. As was found in Task 1, the result of Cluster 3 is highly dependent on the complexity of the task, i.e., up to a certain degree of complexity the students are able to achieve high results; however, when the complexity of the task reaches their limits their results fall back. Similar results were found in Tasks simple and complex. The graphs in **Fig. 1** (right) clearly show that in Task simple the results of Cluster 1 are about as high as of Cluster 3, while Cluster 2 scored lower. There is a significant difference between the results of Clusters 2 and 3 in Task simple (Mann-Whitney test: $p=0.034$). However, this difference diminishes in Task complex ($p=0.784$); the results of Cluster 3 fall back to the level of Cluster 2, and both are significantly lower than the result of Cluster 1. In summary, in the simple task the result of Cluster 3 is similar to the result of Cluster 1 and the difference is not significant ($p=0.449$), while in the more complex task the results of Cluster 3 are significantly lower than of Cluster 1 ($p<0.001$).

4.3 Self-assessment

To decide which task(s) among the three would serve as guideline(s) for the self-evaluation in programming, the self-assessment values were compared to programming tasks. A strong positive correlation was only discovered between Tasks 2.1–2.3 and the self-assessment values ($r_{2.1}=0.50$, $df=788$, $p<0.001$; $r_{2.2}=0.42$, $df=788$, $p<0.001$; $r_{2.3}=0.41$, $df=788$, $p<0.001$), while there was only a weak positive correlation with Task 3 ($r_3=0.16$, $df=789$, $p<0.001$). This result proves that students only consider Task 2 as a programming task, which involves traditional pseudo codes, neither Task 1 nor 3. The programming self-assessment values do not seem able to measure the non-traditional programming knowledge.

The self-assessment analysis of the 3+1 clusters shows that Cluster 1 is different from the other clusters (Kruskal-Wallis test: $\chi^2=16.7$, $df=3$, $p<0.001$), since other comparisons of pairs did not prove any significant difference ($\chi^2=0.05$, $df=2$, $p=0.977$). Consequently, those who had the highest results in Task 1 had the highest self-assessment values.

The self-evaluation overestimates the test results both in Tasks 2 and 3 in all the clusters (**Table 1**). The measure of the overestimation was calculated, and it was examined whether or not the overestimation was different in the 3+1 clusters. Based on the results of the Kruskal-Wallis probe, no significant difference was found between the clusters testing Task 3. However, there is a significant difference in the case of Task 2 ($\chi^2=38.9$, $df=3$, $p<0.001$). Following the Kruskal-Wallis probe, the multiple comparison test proved that Cluster 0 is significantly different from the other three clusters. In this group the self-assessment values are 23.3% percent higher than the average results of Task 2, which is in accordance with the well-known Dunning–Kruger effect, and seems valid in connection with computer related activities, too [6]. These values in Clusters 1–3 are 5.9%, 12.3% and 8.7%, respectively. However, no significant differences were found between the other pairs.

5 Summary

The analysis revealed that the students' algorithmic skills have not developed in their previous education to a level which would serve as firm background knowledge for advanced level CSI studies (H2). A group of students were found in our analyses who up to a certain point are quite convincing in their knowledge. However, they are not able to solve problems which require knowledge beyond their limits; instead they try to find escape routes. The other group with uncertain knowledge proved rather uncertain both in their knowledge and their self-assessment (H3). Based on previous studies and reports, we can assume that the students of this latter group are socialized in an environment where the first level of mastery, the awareness of concept [6], in other words the construction of a mental model of a computer, is skipped and they only focus on the second level, which is usage with surface-approach metacognitive methods [2].

Being aware of these facts tertiary education has at least two things to consider. If students start their studies in Computer Sciences with underdeveloped algorithmic

skills they have to be offered extracurricular activities to develop these skills before starting “serious” studies in programming. If colleges and universities ignore the students’ actual knowledge, the dropout rate will not be reduced. On the other hand, and more importantly, teacher education in CSI has to be reconsidered and courses have to be offered which focus on the development of the algorithmic skills of students in primary and secondary education.

6 Acknowledgment

The research was supported by the TÁMOP-4.2.2.C- 11/1/KONV-2012-0001 project. The project has been supported by the European Union, co-financed by the European Social Fund. The research was supported partly by the Hungarian Scientific Research Fund under Grant No. OTKA K-105262.

7 References

1. Biggs, J. B., and Collis, K. E.: *Evaluating the Quality of Learning: The SOLO Taxonomy*. New York, Academic Press (1982).
2. Biró, P., Csernoch, M.: Deep and surface structural metacognitive abilities of the first year students of Informatics. 4th IEEE International Conference on Cognitive Informatics, Proceedings, Budapest, pp. 521–526 (2013).
3. Biró, P., Csernoch, M.: Programming skills of the first year students of Informatics. XXIII. International Conference on Computer Science, EMT, in Hungarian, pp. 154–159 (2013).
4. Clear, T., Whalley, J., Lister, R., Carbone, A., Hu, M., Sheard, J., Simon, B., Thompson, E.: Reliably Classifying Novice Programmer Exam Responses using the SOLO Taxonomy. 21st Annual Conference of the National Advisory Committee on Computing Qualifications 2008, Auckland, New Zealand. Samuel Mann and Mike Lopez (Eds). (2008).
5. Csernoch, M., Biró, P.: Teachers’ Assessment and Students’ Self-Assessment on The Students’ Spreadsheet Knowledge. EDULEARN13, Barcelona, Spain. Edited by L. Gómez Chova, A. López Martínez, I. Candel Torres. IATED. pp. 949–956 (2013).
6. Computer Science Curricula 2013: Curriculum Guidelines for Undergraduate Degree Programs in Computer Science. December 20, 2013. ACM & IEEE Computer Society (2013).
7. Kruger, J., Dunning, D.: Unskilled and Unaware of It: How Difficulties in Recognizing One’s Own Incompetence Lead to Inflated Self-Assessments. *Journal of Personality and Social Psychology* 77 (6): 1121–34 (1999).
8. Lister, R., Simon, B., Thompson, E., Whalley, J. L. & Prasad, C.: Not seeing the forest for the trees: novice programmers and the SOLO taxonomy, in *Proceedings of the 11th annual SIGCSE*, New York, NY, USA, pp. 118–122 (2006).
9. Mayer, R. E.: The Psychology of How Novices Learn Computer Programming. *ACM Comput. Surv.*, vol. 13 (1), pp. 121–141 (1981).
10. Schraw, G., Horn, C., Thorndike-Christ, T., Bruning, R.: Academic Goal Orientations and Student Classroom Achievement. *Contemporary Edu. Psycho.*, 20(3), pp. 359–368 (1995).
11. Sheard, J., Carbone, A., Lister, R., Simon, B., Thompson, E., Whalley, J. L.: Going SOLO to assess novice programmers, *SIGCSE Bull.*, vol 40, 3, pp. 209–213 (2008).
12. Soloway, E.: Should we teach students to program? *Communications of the ACM*. October 1993/Vol.36, No.10, pp. 21–24 (1993).

StalonaLe@rning: A Social Networking Educational System for Teacher Training

Pitsikalis Stavros¹ and Lasica Ilona-Elefteyja²

General Secretariat for Lifelong Learning, Greece

pitsikalis@gsae.edu.gr

Foundation for Youth and Lifelong Learning, Greece

e_ilona@outlook.com

Abstract. The purpose of this paper is the development, documentation and best use of a social networking tool to create a Web-based Social Networking Educational System aimed at the implementation of a training program designed along the lines of instructional design and deploying diverse teaching models available in relevant literature. For trainees to experience a real-life problem, an educational scenario has been devised, constituting the central theme of the training program throughout which participants are encouraged to receive training and act with different roles. The results derived from the evaluation of the training program are, finally, presented and discussed.

Keywords: Social Networking, ICT Use, Teacher Training

1 INTRODUCTION

In this paper, the utilization of Social Networking Applications in support of Teacher Training is proposed. The aim is the creation of a Web-based Social Networking Educational System (WSNES) that will serve as the basis of a learning community supported by teaching methods. Through the WSNEs in case (<http://learn2learn.gr/social/>), the sharing of digital educational content is encouraged and the promotion of constructive, collaborative and exploratory learning is attempted by providing new technological tools and by enabling sharing and reuse of resources. The actual challenge is the implementation and application of the educational activities in a manner that will prove useful in real-life learning process scenarios. Specifically, for the trainees to gain experience in real-life problem-solving, an educational scenario has been designed.

It is ultimately an approach towards creating attractive educational practices within the context of a social network. Trainees' development of creativity constitutes a central goal. The trainees function as 'Executives' within the WSNEs shouldering the responsibility of the Advertising and Promotion Divisions. The proposed scenario, addressing 2 classes of 16 trainees (part of a wider training program on the use and exploitation of Digital Tools and Systems towards improving Promotional Services) is titled: "Training Executives on YouTube to Promote their Services". Time allotted

adfa, p. 1, 2011.

© Springer-Verlag Berlin Heidelberg 2011

to the training program is 17 didactic hours (6days- 2weeks). The educational problem is related to the trainees' capacity for profoundly comprehending the technologies they observe through a specific process. The trainees develop skills relevant to the knowledge acquired, cultivate critical/systemic thinking and collaboration skills so as to use them as stepping stones once creating appealing educational environments.

2 DESIGN OF THE EDUCATIONAL APPROACH

The aim of the proposed pedagogical approach is to encourage trainees' active participation in the training program instead of merely attending it. The central objective is to focus on learning through setting problems, exploring possible solutions, developing projects, coming up with ideas and eventually presenting them.

2.1 TRAINEES' CHARACTERISTICS AND NEEDS

The trainees assume the role of business/organizations executives (responsible for the promotion and advertising divisions). Furthermore, they display adequate knowledge and skills in Introductory IT Concepts (Computer Use, Office Software, Internet and Communications). Some of the adult trainees' characteristics that were taken into consideration involve [1] [2]: (1) not always disclosing their real intentions for joining the training program (may have joined to make new acquaintances, to become members of a team and, to a smaller extent, to gain knowledge), (2) wishing to make social relationships (feel at ease), (3) appreciating the trainer's genuine interest, (4) needing reassurance and encouragement they can succeed in whatever they wish to do or learn, (5) needing to feel the satisfaction deriving from achieving goals e.tc. A key component of success for training programs is the implementation methodology and the emphasis placed on the participants' effective needs' analysis. The constituent parts of the methodology followed within this context (appropriately customized to fit the needs of the training program in question) comprise five stages composing the Systemic Approach to Training (SAT) (see Figure 1):

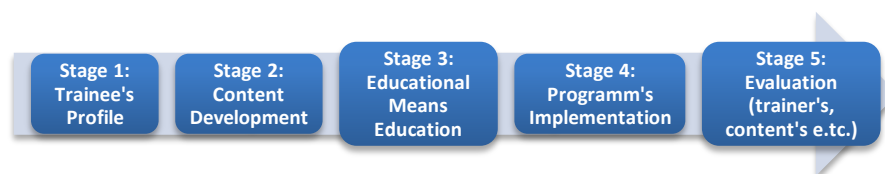


Fig. 1. Implementation Methodology of the Training Program

As regards the trainees' learning needs, YouTube may have an added value for the strategic design of promotional practices. The trainees feel the need to explore, discover and realize new technologies' actual dimension through their practical use. At the same time, they are in need of actively participating in the educational process and learning of how to act in a constructive manner as members of a team.

2.2 TASK & LEARNING OBJECTIVES ANALYSIS

With reference to the training program as a whole, these are the details concerning the implementation and nature of its activities: (1) Primary Information: Delimit 'accidental' correlations between two elements. Learning primarily takes place through memorization, (2) Concepts: The organized knowledge a person possesses in terms of the critical features of persons, objects, events, ideas, statements and procedures, (3) Principles-Generalizations: Causal or regulatory relationships linking concepts and serving the purpose of mentally simplifying the complexity of reality while assisting the prediction of facts and situations, (4) Procedures: While primary information, concepts and generalizations refer to the information and understanding an individual displays of the world, procedures allude to an individual's actions within society, (5) Interpersonal Relationships: Interaction skills fall under this category, (6) Attitudes - Values: Internal, well-established predispositions of an emotional and evaluative nature that affect a person's relationship with things and social surroundings.

By deploying the abovementioned activities the general purpose of the training program is identified as: increasing trainees' understanding of the YouTube functions and features, enhancing trainees' recognition of its added value within the context of promotion and enabling trainees to recognize and plan quality videos with a promotional orientation. A crucial axis for the achievement of the general aim is the WSNES displaying the necessary functions and educational content to support trainees. To clearly define the content of the program 'functionalization' of the general purpose is being conducted, formulating specific objectives. According to Bloom's taxonomy, such objectives fall into three domains: cognitive domain (identify, understand, perceive, interpret e.tc.), skills (create, use, leverage, make, become familiar e.tc.) and affective domain (will, wish, grasp, recognize, appreciate, realize e.tc.) [3].

2.3 TRAINING PROGRAM DESIGN AND TEACHING STRATEGIES

No technological or communication system can enhance learning effectiveness in its own unless it is accompanied or is part of a pedagogic logic and serves a structured pedagogic process. The quality of distance educational programs could be evaluated through: (a) overall program design (serving general principles of education, taking into account particularities of distance teaching), (b) the fact that the objectives and expected outcomes of learning should be closely associated, (c) the program's management and implementation (ensuring that trainees will achieve their learning goals), (d) encouraging trainees to take control of their learning, (e) trainees' assessment.

The program takes place in four distinct phases containing individual modules as illustrated in Figure 2. The modules have been configured in accordance with the learner-centered learning-related strategy [4] which advocates, among others, the following: (a) identification of prerequisites, (b) intimacy; relevant concepts are presented prior to unknown ones, (c) graded difficulty; the easy parts are taught first progressively moving to more complicated ones, (d) motivation; at first, trainees are presented with highly motivating issues. For each of the program's modules a series of activities based on carefully selected instructional strategies has been provided.

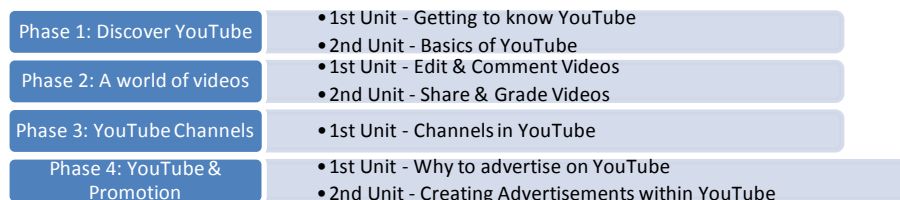


Fig. 2. Training Program Phases

The nature of the teaching content, the trainees' characteristics and the program's objectives necessitate the selective and combined use of four teaching strategies, namely: (a) Pre-instructional Strategy (can cater for the need of comprehending concepts and generalizations involved in each module while strengthening trainees' cognitive structures [5]), (b) Effective Teaching Strategy (it is aimed at formulating and automating processes that ensure quick and accurate promotion channels' creation by the trainees. Trainees receive individual practice in the new skill while the trainer provides guidance and assistance), (c) 5W1H Strategy (is designed to split a problem into individual sub-problems, addresses the development of problem-solving skills and consists of an organized series of questions inquiry focused (What, Where, When, Why, Who, How) [6]. It is employed within the training program to support trainees in developing original ideas as regards the use of YouTube in promotional practices) and (d) Webquests (series of guided inquiry activities in which learners undertake the responsibility of solving a problem using the internet as the primary source of information. Within this training program, the WebQuest strategy has been employed, as studies show that it has the potential to develop online search skills, comprehension of concepts but also to enhance trainees' motivation and collaboration [7]). Within this context, the themes taught need to be approached through group work.

Such strategies are primarily founded on a theory of cognitive development [8] which emphasizes that cognitive processes can be modeled in the social world before internalized by the individual. The selected learning theory and teaching strategies are supported within the proposed WSNES, with Web 2.0 technologies that reinforce cooperation and active participation in learning [9].

3 EDUCATIONAL CONTENT DEVELOPMENT

With reference to the WSNES' structure, an appealing environment has been created to enhance participants' positive attitude and mood; at the same time, being user friendly the WSNES allows novice users to easily navigate themselves throughout and actively participate in existing educational activities by means of merely possessing basic computer and internet use skills [9]. The tools employed for the development of the teaching materials or for reusing existing educational content can be hierarchically divided in the categories to follow: (1) Central Online Social Networking Training System (Oxwall), (2) Supportive Material for Phases/Modules (Online tools for interactive books, animation, comics, presentations, games and videos crea-

tion), (3) Educational Activities' Material (file sharing applications, WebQuests, video recording and sharing tools), and (4) Educational Content (either designed within the context of the educational program in case or has been the result of relevant Internet resources). In Figure 3 parts of the educational content available are depicted.



Fig. 3. Sample of Educational Content

4 TRAINING PROGRAM EVALUATION RESULTS

To facilitate the process of both qualitative and quantitative result extraction two (2) online questionnaires, one aimed at assessing the use of the StalonaLe@rning and one for assessing the training program have been designed. In addition, (a) the participants' views added to the forum within the context of the tasks assigned and their face-to-face discussions, (b) the functionality, aesthetics and content of the final videos created and (c) the trainees' participation in the activities (quiz, games, deliverables e.tc.) were also taken into consideration.

With regard to the first questionnaire, all trainees stated that their overall impression of the WSNES has been ranging from "very good" to "outstanding", while none of them had ever used a similar educational system. 80% claimed that the educational content available in digitized audiovisual form is interesting and the diversity of content (combination of notes, tasks, examinations, discussions), as opposed to the static lecture notes' presentation is more intriguing. The ease of communication among the participants and the distribution of educational content and ideas that encouraged a fun atmosphere, have been reported as the WSNES' strong point. Another feature of the WSNES that has been reported by the trainees involves information management and familiarization with diverse technologies. Finally, 60% of the trainees expressed their intention to integrate corresponding WSNESs into their own teaching programs.

In terms of the second questionnaire, all trainees felt that the training program was "very good" to "outstanding" and would definitely recommend it to their colleagues. 75% of the trainees mentioned that their expectations of the program were fully met and that the teaching method employed was "very good" to "outstanding". Also, 80% claimed that all activities met the program's objectives but also argued that the program's duration could be redefined.

5 CONCLUSIONS

The educational environment in which the training program and learning take place has been designed and implemented in compliance with the Instructional Design and Educational Approach principles presented in the above sections. Within this context: (a) the flexibility of applying various teaching methods depending on the learning conditions and the trainees' cognitive demands is ensured, (b) the trainees are dynamically involved in the teaching process having opportunities for communication with other trainees and with the instructor, enjoying a no longer passive role, (c) direct interaction and collaboration between the trainer and the trainees and the trainees themselves are ensured, (d) the trainees are getting engaged with the educational content and activities, taking over and completing various tasks, (e) they develop analytic, generalization, synthesis and application of knowledge skills, critical thinking and interest display in scientific matters.

The present educational program which has been implemented to assist trainees in-to acting as actual managers completing specific projects so as to evaluate and correspond to the demands of real life situations has been welcomed by the trainees who realized the need for developing good practices to be shared and the opportunity for getting involved in their own educational process.

Finally, the trainees were particularly interested in the interactive and digitized educational content, completed all activities and actively participated in the community created. It needs to be mentioned that the trainees got impressed by the design and the selection of a WSNES; concurrently, they reported that the balanced use of corresponding systems could have been exploited towards attracting trainees' interest and in support of traditional educational systems as well.

References

1. Smith, C., Hofer, J.: The characteristics and concerns of adult basic education teachers (NCSALL Rep. No. 26). Boston: National Center for the Study of Adult Learning and Literacy (2003)
2. Tight, M.: Education for adults, Volume 1: adult learning and education (2012).
3. Anderson, L.W., Krathwohl, D.R.: A taxonomy for learning, teaching, and assessment. A revision of Bloom's taxonomy of educational outcomes. New York: Longman (2001)
4. Morrison, R. G., Ross, M. S., Kemp, E. J.: Designing effective instruction (3rd edition), New York: John-Wiley & Sons (2001)
5. Joyce, R.B., Weil, M., Calhoun, E.: Models of teaching, 7th Edition. Upper Saddle River, N.J.; Prentice Hall (2008)
6. Jang, S., Woo, W.: 5W1H: Unified user-centric context. In: 7th International Conference on Ubiquitous Computing, Springer-Verlag Berlin (2005)
7. Gulbahar, Y., Madran, R.O., Kalelioglu, F.: Development & Evaluation of an Interactive WebQuest Environment: "Web Macerasi". Ed. Tech. & Soc., vol.13 no.3, pp.139-150 (2010)
8. Vygotsky, L.: Thought and language, Cambridge, MA: MIT Press (1986)
9. Jothi, P.S., Neelamalar, M., Prasad, R.S.: Analysis of social networking sites: A study on effective communication strategy in developing brand communication. J. of Med. and Com. St., vol.3 no.7, pp. 234-242 (2011)

Virtual Environment for Creative and Collaborative Learning

Anna Bilyatdinova, Andrey Karsakov, Alexey Bezgodov, Alexey Dukhanov

ITMO University, Saint Petersburg, Russia

a.bilyatdinova@gmail.com, kapc3d@gmail.com,
demiurghg@gmail.com, dukhanov@niuitmo.ru

Abstract. New guidelines for Russian technical universities and creative industries require competitive professionals and we need to apply novel educational methods to make their training efficient. This tendency calls for new tools to be adopted and implemented in the learning purposes. Program curriculum is based on three levels of desirable exit qualifications students need to develop - Hard Skills, Soft Skills and Professional competencies in a specialized field. Novel not only in content but also in form Elective courses in Scientific Visualization, and Virtual Reality of Double Degree Master's Program in Computational Science (ITMO University, Russian Federation and University of Amsterdam (UvA), The Netherlands) incorporate 3D anatomical atlas of human body, which became a prototype for the Virtual Learning Environment (VLE). In this paper, we describe architecture of the VLE and our experience of the introduction of VLE in the master program courses and master theses preparation. We use our flexible VLE to acquire knowledge, develop collaborative skills, and monitor knowledge acquisition within the same unified environment using different methods.

Keywords: Virtual Reality, Virtual Environment, Curriculum Design, Learning Competence.

1 Introduction and related works

Interconnection of the fields of study and application fields implies cultivation of multi-domain Hard and Professional Skills, which is reflected in growing popularity of interdisciplinary master's programs [1]. However, apart from audio, video and social media communication tools for learning and teaching purposes the lack of the multi- and interdisciplinary educational methods still exists. This situation is observed in most CIS countries due to rigid traditional approaches to training.

Diverse virtual reality environments are gaining their place in education and learning processes for the last 15 years. An Educational Virtual Environment (EVE) or Virtual Learning Environment (VLE) can be defined as a virtual environment that is based on a certain pedagogical model, and incorporates or implies one or more didactic objectives, provides users with experiences they would otherwise not be able to experience in the physical world and redounds specific learning outcomes [2]. In Dil-

adfa, p. 1, 2011.

© Springer-Verlag Berlin Heidelberg 2011

lenbourg's [3] and Mikropoulos' [4] papers common and unique features, which clearly determine VLE were described and currently there are many research works about effectiveness and possibilities of VLEs [5–11].

Real application fields of VLEs are wide enough and they may vary depending on the learning activities, i.e. virtual teams [12], and learning outcomes [13] that can be gained using different environments. Quite a large number of existing works dedicated to the research of implementation of the existing solutions in the educational process, i.e. game-like environment Second Life [14–16]. Others use their own solutions [17–19], that are also based on the gaming activities.

Separately it is possible to allocate the VLEs that are used to train encyclopedic knowledge, such as anatomic atlases used to provide access information in conjunction with a three-dimensional representation of the object of study. In this case such object of study is a human body 3d model. Implementation architecture of anatomical atlases is rather diverse starting from web-based applications accessible through web browser [20–22] with the advantage that the developers can always monitor the relevance of knowledge and up to the applications that fully store and run directly on client's computer [23–25]. Key issue of using such types of environments is that all of them are mainly locked to datasets integrated by developers without abilities for users to amend these datasets or data.

Scientific Visualization Team of eScience Research Institute of ITMO University was organized in 2007, equipped with the state-of-the art facilities and during these years successfully fulfilled several ambitious projects in scientific visualization domain, e.g. Problem Solving Environment for Flood Protection Barrier of Saint Petersburg, original 3D engine with realistic sea waves, ship motion and sun system simulation, and a 3d anatomic atlas of normal human body [26, 27]. The latter project has become a prototype for the VLE implemented in the syllabi of several elective courses of the master's program in Computational Science "Supercomputer technologies in Interdisciplinary Research". Below we describe one of our projects of 2-days group work with the elements of competition. Simultaneously students had a chance to test the key functions of technical (optimization of graphic engine to create complex models, introduction of Virtual Reality Tools such as dual-head stereo and 3D Vision, support of human-computer interaction - Microsoft Kinect and multi-touch support), and functional component (efficient structure of graphical user interface, content management system) of VLE developed by their lecturers.

2 Virtual Environment infrastructure

Virtual Environment based on client-server architecture. Client System consists of two parts - Content Management System (CMS) and User Shell (US). A server part of the VE is an ordinary web-server deployed on a remote PC. It works on Apache or IIS with installed PHP interpreter (version 5.2 or higher) and all basic modules that are required for normal work. In addition, the system requires a database server with MySQL 5.1.7 or higher. CMS and the data storage are implemented using PHP language and MySQL.

CMS is a simple web application and is accessible through the web browser interface. Left side of the page has several drop-down sections with the lists of options to select. Right side contains editable fields, which purpose and content depend on selected option in the left part of the page. CMS has “Edit content”, “Statistics and reports”, “Users” and “System setup” expandable groups of options. “Edit content” section contains a list of the elements of a hierarchical tree of 3D models loaded into the VE for editing the content. “Statistics and reports” section contains statistical information about the use of the environment. “Users” section contains options to manage the user’s rights. “System setup” contains global CMS and VLE parameters.

Differentiation system allows to us restrict access to the different CMS sections depending on the user types:

- Students have access only to US application (view-only rights).
- Advanced students in addition have access to the “Edit content” section of CMS.
- Lecturers have additional access to “Users” section, could create new users and define their type, except ‘System Administrator’ type.
- System administrator has complete access to US and CMS.

US is a main visualization application with graphical user interface. It is based on self-developed Fusion Engine platform, which is mainly implemented using C#. It runs under Windows 7 or Windows 8 and uses Direct3D 11 technology for visualization. In VLE students can open 3D models, and either learn hierarchical connection of its components in spatial representation, or study supplementary materials. Optionally they can edit it (both description of the objects and information about it). In fact, we have a tool for collaborative learning, tool for individual encyclopedic knowledge acquisition and knowledge control. More than that, our environment is not tied to a specific subject matter, which gives a wide range of opportunities to apply our learning environment in any field of knowledge, where the subject matter can be visualized in a spatial or temporal-spatial manner (animated objects, processes in or with the objects of study).

3 Implementation of VLE in the Learning Process

A global transition from post-industrial to knowledge economy raised new challenges for design and content of the courses in engineering education. So in 2012, ITMO University and University of Amsterdam (UvA) launched a double-degree Master's program in Computational Science. The program of 120 ECTS designed by the mutual effort of Russian and Dutch specialists lasts two years (60 ECTS in Russia and 60 ECTS in The Netherlands). The list of courses is based on three levels of desirable exit qualifications (competencies) - Soft Skills, Hard Skills and Professional Competencies. The goal to develop competencies, skills and abilities of students and not just to impose knowledge is a prerequisite for modernization and is supposed to bring the outcomes of educational process in line with international standards, in particular, Bologna Process.

At the level of design of the courses and following delivery of these courses, a lot of attention was given to their format and technical support to keep up with the standards of the UvA. Effective management and structuring of knowledge and information share could help to stimulate the shift from traditional approach to teaching to student-centered and technology-enhanced learning to solve interdisciplinary tasks.

We conjecture that in our VLE researchers, teachers and students could work together aiming to acquire and create knowledge and enhance the ability to self-tuition in accordance with the principles of knowledge society. In general, inside this VLE students can plan, design, produce and apply new entities as a part of innovative practice forming research skills.

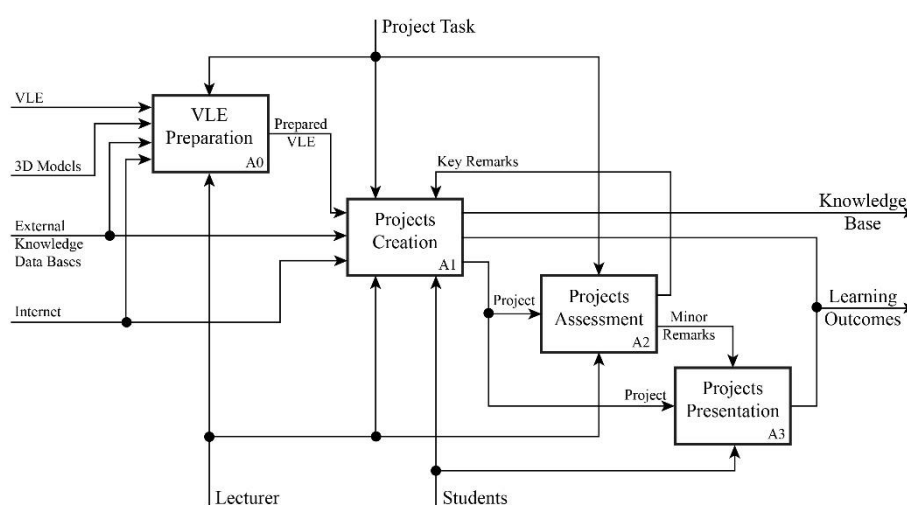


Fig. 1. Scheme of VLE implementation cycle

Our test group project was carried out during 2 days (5 hours during the first day, 5 hours on the next day including group presentations) aiming mainly to run VLE in real life and receive feedback on VLE operation or, to be more precise, how useful it would be to implement it in teaching courses (for example, Scientific Visualization and Virtual Reality, etc.) and group projects in the chosen specialization in the second semester of the master's program. We divided it into three principal tasks: to teach students not only to retrieve already available information (from lectures, www), but to generate knowledge, then define methods of research and assessment of learning outcomes and finally to identify the possible impact of the implementation of VL.

Optionally, at the stage of opinion poll, we offered our students to reflect upon new fields of VLE implementation and influence of the proposed approach.

At the planning stage we implied that students will go from application of already received knowledge (discussion of the given task) to analysis (estimate resources and define the ways to solve the task) and synthesis (develop a working plan in a group) to evaluation (appraise, compare and assess their peers). To follow that logic we have chosen the following instructional techniques: group work (to improve collaborative

and communicative skills), competition of the groups (to motivate students to win), presentations (students choose content to develop creativity), and self-assessment (to evaluate peers).

While setting the task for the groups, we intentionally did not provide the detailed specification of the problem to solve. Only general criteria to assess the result of the group work were specified aiming at the development of creative thinking. 15 master students were divided into 5 groups of 3 people and 4 out of 5 groups had a female member. Each group was given a system of a human body to explore (e.g. cerebrum, cardiovascular system, digestive system, muscular system, skeleton), find the information about it, define the most important from the point of view of the group parts of the system, to learn to use the functions of the VLE (such as different shading parameters, main constructive representation of the model).

For 2 days students worked in a computer class equipped with all necessary software and internet access. The task was given in the first day and followed by 4 hours of work, the next day 2 hours of work were followed by final group presentations. Each group had to prepare texts and pictures to illustrate their chosen hierarchy and to present their results during the session in computer class and in Scientific Visualization Center, which provides completed graphical view.

Participants were offered a questionnaire to assess the experience in general consisting of four parts: Program and Task Design, Competition of Group Presentations, Experience and Software feedback. The quality of presentations was estimated by both students themselves and four experts in computer simulation, physics, cloud computing and education. Total result of the group consisted of the average score of the parameters (80%) and evaluation of the presentation skills of the group (20%). All participants had to rate all five presentations by giving a rating number between 1 and 10 (1 = lowest, 10 = highest) and write down their opinion about such type of work, suggest possible implementation of given VLE, and moreover provide Software User's feedback (bug reports, functional improvements, global improvement ideas, etc.).

At the final presentation, each group had to convince the listener that their choice of hierarchy within the system is correct, which in turn requires proper reasoning (communicative skills) and carefully selected visual aids shown with the help of VLE (creativity). The presentations were estimated due to the following criteria: View, Shading, Adequacy of Display, General Visual Quality, Illustrations, Texts and Adequacy of Hierarchy of the system.

The final scores of the groups give a lot of food for thoughts in terms of development of Hard and Soft Skills of the students. Two groups were very close to each other in the average number of parameter (9.06 out of 10), but the 20% for presentational skills were crucial when the decision to choose a winner was taken.

During the course of the group project, it was interesting to note that one of the challenges our master students face most often is the lack of collaborative skills since in this type of work do not resemble their day-to-day interaction. Members of a certain group with a strong female leader faced serious problems when planning their work. However, a heated discussion of the plan of work, division of roles and time management issues, which even demanded the involvement of the experts, brought outstandingly good results. Element of competition also had a curious effect on those

members of the group who did not successfully interact before the group work task. While working with VLE and learning to use it, members of one team built fruitful communication and performed well together. At the same time, some of the groups with sound technical background (e.g. high average marks) could not show the proper use of VLE functionality and their final product raised many questions and mixed impression in general.

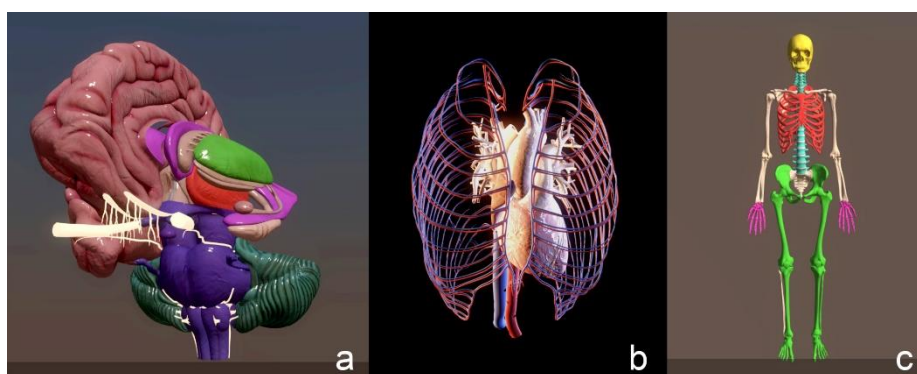


Fig. 2. Samples of group work's results in VE

On the figure 2, you can see the results of the students group work. The first two (figure 4, a and b) groups presented rather curious views of their systems, but the skeleton view (figure 4, c) did not go further than plain demonstration of general view with a frontal projection and color division of general functional groups. Therefore, the new trend in Russian graduate education to put emphasis on Soft Skills (such as common culture, communicative skills, self-criticism and creativity) is not groundless. In the final analysis, we might conclude that VLE could be used as a tool to form creative and collaborative skills and students respond positively to the presented format. Furthermore, we conclude that VLE could be used at best in combination with other forms of study, e.g. individual and group projects, division of roles by the teacher or by students themselves inside the groups.

One of the advantages of implemented VLE cycle, described in this paper is that it is not restricted to a subject field and could serve as a basis of multidisciplinary learning and research. In addition, it is could be deployed with different cycle duration (from a separate 'module style' group project lasting several day up to a part of a full course). Moreover, educators who took part in this group work project cannot but notice that students of our master's program with STEM background are more likely to adopt the new creative ways of tackling a problem when they are puts in interaction with both their peers and VLE.

4 Conclusion and future work

Our experience of the introduction of VLE in the master research project reaffirms that the effectiveness of VLE use depends on the combination of objective (task design and description, implementation of various teaching and learning strategies, software precision) and subjective conditions (participants' feedback in different fields of knowledge, ability to extract and analyze information, organize and plan group and individual work). 35, 7% out of the students who took part in the test group project indicated in the questionnaire that they have not gained so much from group work, VLE and presentations of the groups. At the same time when offered to describe the experience by writing down a couple of words, we came up with rather positive feedback. Moreover, most of the students took seriously the bug report and gave detailed description of the errors they ran into.

Implemented VLE cycle illustrates both acquisition of theoretical encyclopedic knowledge and more creative work to develop Soft skills. VLE could facilitate team-based learning and intensify formation of Soft and Professional Skills.

Possible application scenarios and development of existing VLE and/or development of a separate model of VLE corresponding to desired learning outcomes of the course could still be a subject of future research, since all VLE functionality is not fully discovered yet, and the only way to learn more is trying.

5 Acknowledgements

This paper is supported by Russian Scientific Foundation, grant #14-11-00823. The research is performed in Advanced Computing Lab (ITMO University), which is created in the frame of 220 Decree of Russian Government, contract #11.G34.31.0019.

6 References

1. Dukhanov, A. V., Krzhizhanovskaya, V. V., Bilyatdinova, A., Boukhanovsky, A. V., Sloot, P.M. a.: Double-degree Master's Program in Computational Science: Experiences of ITMO University and University of Amsterdam. *Procedia Comput. Sci.* 29, 1433–1445 (2014).
2. Mikropoulos, T. a., Natsis, A.: Educational virtual environments: A ten-year review of empirical research (1999–2009). *Comput. Educ.* 56, 769–780 (2011).
3. Dillenbourg, P., Schneider, D.K., Synteta, P.: Virtual learning environments. *Proc. 3rd Hell. Conf. "Information Commun. Technol. Educ.* 3–18 (2002).
4. Mikropoulos, T., Bellou, J.: The unique features of educational virtual environments. *Proc. e-society.* 122–128 (2006).
5. Simkova, M., Stepanek, J.: Effective Use of Virtual Learning Environment and LMS. *Procedia - Soc. Behav. Sci.* 83, 497–500 (2013).
6. Semradova, I., Hubackova, S.: Learning Strategies and the Possibilities of Virtual Learning Environment. *Procedia - Soc. Behav. Sci.* 83, 313–317 (2013).

7. Semradova, I., Hubackova, S.: Virtual Learning Environment and the Development of Communicative Competences. *Procedia - Soc. Behav. Sci.* 89, 450–453 (2013).
8. Loureiro, A., Bettencourt, T.: The Use of Virtual Environments as an Extended Classroom – A Case Study with Adult Learners in Tertiary Education. *Procedia Technol.* 13, 97–106 (2014).
9. Grenfell, J.: Immersive Interfaces for Art Education Teaching and Learning in Virtual and Real World Learning Environments. *Procedia - Soc. Behav. Sci.* 93, 1198–1211 (2013).
10. Dalgarno, B., Lee, M.: Exploring the relationship between afforded learning tasks and learning benefits in 3D virtual learning environments. ... *Aust. Soc. Comput. Learn.* (2012).
11. Dalgarno, B.: The potential of 3D virtual learning environments: A constructivist analysis. *Electron. J. Instr. Sci.* 1–19 (2002).
12. Bosch-Sijtsema, P.M., Haapamäki, J.: Perceived enablers of 3D virtual environments for virtual team learning and innovation. *Comput. Human Behav.* (2014).
13. Lorenzo, G., Pomares, J., Lledó, A.: Inclusion of immersive virtual learning environments and visual control systems to support the learning of students with Asperger syndrome. *Comput. Educ.* 62, 88–101 (2013).
14. Chau, M., Wong, A., Wang, M., Lai, S.: Using 3D virtual environments to facilitate students in constructivist learning. *Decis. Support* 56, 115–121 (2013).
15. Hsu, L.: Web 3D simulation-based application in tourism education: A case study with Second Life. ... *Hosp. Leis. Sport Tour. Educ.* 11, 113–124 (2012).
16. Land, S. Van Der, Schouten, A.: Lost in space? Cognitive fit and cognitive load in 3D virtual environments. *Comput. Hum.* 29, 1054–1064 (2013).
17. Berns, A., Gonzalez-Pardo, A., Camacho, D.: Game-like language learning in 3-D virtual environments. *Comput. Educ.* 60, 210–220 (2013).
18. Parsons, S., Leonard, A., Mitchell, P.: Virtual environments for social skills training: comments from two adolescents with autistic spectrum disorder. *Comput. Educ.* 47, 186–206 (2006).
19. Schmidt, M., Laffey, J., Schmidt, C.: Developing methods for understanding social behavior in a 3D virtual learning environment. *Comput. Human Behav.* 28, 405–413 (2012).
20. John, N.: The impact of Web3D technologies on medical education and training. *Comput. Educ.* 49, 19–31 (2007).
21. Brenton, H., Hernandez, J., Bello, F.: Using multimedia and Web3D to enhance anatomy teaching. *Comput.* 49, 32–53 (2007).
22. Lu, J., Pan, Z., Lin, H., Zhang, M., Shi, J.: Virtual learning environment for medical education based on VRML and VTK. *Comput. Graph.* 29, 283–288 (2005).
23. DeLaurier, A., Burton, N.: The Mouse Limb Anatomy Atlas: an interactive 3D tool for studying embryonic limb patterning. *BMC Dev.* 8, 83 (2008).
24. Hamrol, A., Górski, F., Grajewski, D., Zawadzki, P.: Virtual 3D Atlas of a Human Body–Development of an Educational Medical Software Application. *Procedia Comput. Sci.* 25, 302–314 (2013).
25. Hacker, S., Handels, H.: A framework for representation and visualization of 3D shape variability of organs in an interactive anatomical atlas. *Methods Inf Med.* 48, 272–281 (2009).
26. Bezgodov, A., Esin, D.: Complex Network Modeling for Maritime Search and Rescue Operations. *Procedia Comput. Sci.* 29, 2325–2335 (2014).
27. Zagarskikh, A., Karsakov, A., Tchurov, T.: The Framework for Problem Solving Environments in Urban Science. *Procedia Comput. Sci.* 29, 2483–2495 (2014).

A semantic system to support teachers to create differentiated digital learning resources

Françoise Greffier, Federico Tajariol

Objets et Usages Numériques, ELLIADD Laboratory, FR-EDUC
University of Franche-Comté
{francoise.greffier, federico.tajariol}@univ-fcomte.fr

Abstract. This paper presents the design of a semantic system to index digital learning resources according to the metadata describing their cognitive features. By “cognitive feature” we mean the cognitive activities (e.g., reading, listening, body interactions, etc.) associated with the form of presentation (e.g., text, audio, image, etc.). The semantic system includes a parser, which detects the semiotic components of a resource, and two ontologies that formally describe the cognitive styles and the semiotic descriptors, as well their association.

Keywords. Differentiated instruction, Cognitive styles, Semantic indexation, Digital learning resources.

1 Introduction

Information and Communication Technologies for Education enable teachers to carry out different types of Digital Learning Resources (DLR). A DLR could be a text explaining a concept, a practical work to train students, an interactive simulation, an exercise to evaluate their level of knowledge, etc., available on websites and on repositories of learning objects¹. Each DLR is the creative outcome of a teacher, who acts a pedagogical designer. In fact, even if a teacher does not create learning resources from scratch, (s)he can reuse and adapt preexisting DLR in order to fit learners’ pedagogical goals, cognitive abilities and learning preferences.

In regard to this personalized learning issue, a teacher runs into difficulties because of the little availability of details about the cognitive features of a DLR. By “cognitive feature” we mean the cognitive activities (e.g., reading, listening, body interactions, etc.) associated with the specific forms of presentation (e.g., audio, image, text, etc.) of a DLR. So, for example, a learner who is asked to read a text uses different cognitive abilities and skills than when (s)he observes a diagram illustrating the same content, and maybe (s)he better understands through the former cognitive activity than by the latter.

¹ For instance, AMSER <https://amser.org/>, MERLOT <http://www.merlot.org/merlot/index.htm>

The main goal of our research question is to help teachers to create differentiated DLR fitting student's learning abilities. To tackle this problem, we propose an automatic semantic system to index the cognitive features of a DLR.

This paper is built as follows. In the first section we explain the scientific problem and we suggest our approach to index the cognitive features of a DLR. In the second section, we define our set of descriptors concerning the cognitive features inserted in a DLR. In the third section, we present the architecture of our system we plan to develop.

2 Design differentiated digital learning resources: a creativity problem

Personalized learning means to adapt the pedagogical contents and activities to the learner's cognitive styles [1] [2], as well as to design differentiated forms of a same pedagogical content [3]. Thus, when a teacher is in charge of the design of a DLR, (s)he also has to reason on which structure and form of presentation could be in adequacy with (her)his learners' specific cognitive styles [4] and, therefore, the most efficient for their learning. For example, a philosophy teacher must prepare a lesson on modal logic intended for students in visual arts: (her)his first intention could only be to present logic symbols and formalisms, whereas those students would be more effective when they processed images and narrative forms, as if they owned a particular way of information processing. This research tackles the issue of design differentiated DLR: how could we help teachers to create differentiated DLR that would enable students to be more effective in their learning process?

Amongst works on cognitive and learning styles [5] [6], and the complex links between these two constructs [7], we suggest a different theoretical approach, the Multiple Intelligence Theory [4]. According to the MI theory, the intelligence of a person is a set of abilities and skills (s)he develops to solve problems and process information. In MI theory, there are between seven and nine sets of abilities and skills (intelligences), seldom applied separately, especially when a human being achieves a complex task in a precise context. These sets would be relatively autonomous to one another: intrapersonal, interpersonal, kinesthetic, linguistic, logical-mathematical, musical, visuo-spatial. For example, a person who has a visuo-spatial intelligence is able to better think through visual elements (e.g., images, graphs, cards, colors, etc.) and structures (e.g., patterns, diagrams, etc.) rather than speeches (linguistic intelligence), symbols (logical-mathematical), etc. Two main reasons trigger our theoretical choice. First, previous works have successfully applied the MI theory in the design of Technology Enhanced Learning environments [8]. Secondly, as we found in our previous works [9], the modularity of the MI theory let us formalize the association between a form of presentation (textual, audio, etc.), its semiotic components (words, sentence, images, etc.) and the cognitive activities required (reading, manipulating, etc). In fact, we built a grid of analysis linking the semiotic components of a DLR (sentences, keywords, etc) to the multiple intelligences (table 1). For instance, a DLR whose contents are presented by keywords, short sentences

and diagrams, generally expresses the logical-mathematical intelligence, whereas a DLR containing images (fixed or animated) and videos requires a visuo-spatial intelligence.

Semiotics components	Gardner's intelligences
Sentences describing and explaining a concept	Linguistic
Keywords, short sentences, diagrams	Logical-Mathematical
Sounds, music	Musical
Fixed images, animated images, videos	Visuo-Spatial
Interactive animations in which the learner must perform some gestures (e.g. click and drop)	Kinesthetic
Linguistic expressions showing the personal feeling of the author of the learning resource	Interpersonal
Linguistic expressions engaging the learner to think	Intrapersonal

Table 1. - Grid analysis linking Gardner's theory to the semiotic components of a DLR

3 A Semantic System to Index the Digital Learning Resources

Our SEMantic System to Index Digital learning resources (SESID) analyzes the semiotic components of a DLR and produces the metadata describing cognitive features of a DLR. The SESID enables the automatic indexation of DLR, because teachers usually do not fill several and high complex descriptors, such as Learning Object Metadata (LOM). The SESID consists of three components.

i) *Morphological analyzer*. A parser detects the semiotic components of each DLR by means of the grid analysis we presented above, and calculates their number. Its output is recorded in a database and also in an external spreadsheet file. We developed the parser under Java (1.7.0_11), and we used the Apache POI library (<http://poi.apache.org/>) to retrieve the semiotic components inside the .ppt, .pptx, .doc and .docx files, as well the itext library (<http://itextpdf.com/>) for the .pdf files. For example, for each slide of a .ppt file, the Apache POI creates a set of shapes. Then, by means of a set of algorithms we developed, the parser distinguishes between animated and static images (an icon, a draw or a scheme).

ii) *Domain Ontologies*. Considering the connections between semantic web and ontologies [10] [11], we created two domain ontologies [25]. First, in the *Semiotic Descriptors* ontology (SD) we formally specified the semiotic components of a DLR and their relations. The SD ontology (fig. 1) formally describes that, for instance, an icon is a fixed image.

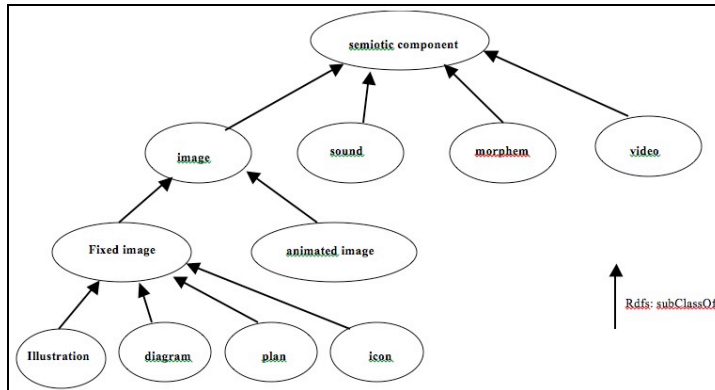


Figure 1 - Domain ontology of semiotic descriptors

Second, in the *Cognitive Style* ontology (CS) we formally described the Gardner's Multiple Intelligence theory [4] and we described the seven ways of presenting the learning contents of a DLR (fig. 2).

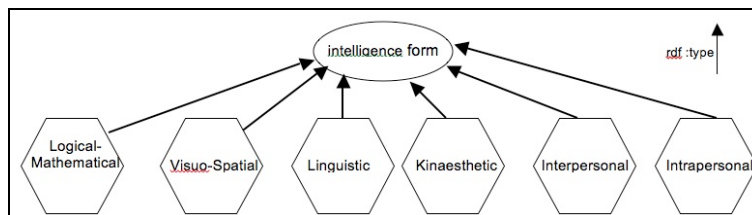


Figure 2 - Ontology of Multiple Intelligence

iii). *Application Ontology*. We designed two application ontologies to build both ontological classes and the relations between them. First, the *Linking Ontology* (LO), we built to describe formally the association between semiotic components and forms of intelligence according to the Gardner's approach (see Table 1). For instance, when the parser detects and counts the semiotic descriptors "morphemes", the Linking Ontology enables to associate the morpheme to the Linguistic form of intelligence. Second, we created the *Counting Ontology* (CO) to classify each form of intelligence associated to a DLR according to the number of semiotic descriptors that compose each DLR.

3.1 Justification of our choice

The function of a parser is necessary. In fact, in spite of some ontology-based image system [28] and some metadata granularity broad (e.g., the standard LOM allows the description of « image, fixed image and animated image »), we observed that the DLRs are seldom indexed in general, and in particular in the learning

repositories we have worked on for this research². The parser uses some API (Java) such as *itext* or *doc.office*, which includes some *getters* (ex: *getTitle*, *getSlides...*). The “file header” of each file gives us also useful information to distinguish a fixed from an animated image.

First, an ontology enables the description of complex interrelationships between the concepts [12] [13] and remains independent by the processing of the inference engine [14]. So, for instance, the Cognitive Style ontology could be reusable in other software, like an application monitoring when the user (teacher or student) tries to select the DLRs in adequacy with (his) her cognitive style [15] [16]. Our purpose is to establish the inter-ontology relations between the SID metadata and other existing ontologies and standards [17], as the Dublin Core [18] [19], the Learning Object Metadata (LOM) [20], the Standard Content Sharable Object Model Reference (SCORM). So, we aim to establish the links between our CS ontology and the Dublin Core metadata, which index digital resources from the title, the author, the date of edition, etc. Concerning the LOM [21] and MLR (Metadata for Learning Resources) [22] [23], we are going to study how integrating the metadata of the SESIID in LOM number of category 5 and 8 [24]. Category 5, entitled “Educational”, includes the pedagogical characteristics of a resource (type of resource, role of the user, context of use...), whereas category 8, called “annotation”, lets user add annotations on the pedagogical use of the DLR. About SCORM, in the fourth release we estimate to ingrate the metadata of SID with the section “system of content description”. We now introduce the two ontologies.

4 Conclusions

In this paper, we presented an original proposal to index DLRs according to their cognitive style. We choose to formally define Multiple Intelligences theory as ontology. We described the architecture of a system called SESID to index DLR, consisting of two ontologies and a parser. The SESID detects the cognitive style conveyed by a DLR, letting teacher select some parts of a DLR, combine them with others and create differentiated learning resources in order to fit learners’ multiple intelligences.

References

1. Brusilovsky, P., A. Kobsa, W. Nejdl, (eds): *The Adaptive Web: Methods and Strategies of Web Personalization*. Springer Verlag, Heidelberg, (2007).
2. Lefevre, M., Cordier, A., Jean-Daubias, S., Guin, N.: *A Teacher-dedicated Tool Supporting Personalization of Activities*. In: *ED-MEDIA 2009-World Conference on Educational Multimedia, Hypermedia & Telecommunications*, pp. 1136- 1141, (2009).
3. Türker, M. A., Görgün, İ., Conlan, O.: *The Challenge of Content Creation to facilitate Personalized eLearning Experiences*. *Int. J. E-Lear.*, vol. 5 (1), 11-17, (2006).

² SILO, <https://science-info-lycee.fr/>

4. Gardner, H.: *Frames of mind: The theory of multiple*. Basic Books, New York, (2004).
5. Petty, G., and Haltman, E.: Learning style and brain hemisphericity of technical institute students. *J. Stud. Tech. Careers* 13, 79–91, (1991).
6. Riding, R.: On the nature of cognitive style. *Educ. Psychol.* 17(1 & 2), 29–49, (1997).
7. Zhang, L., Sternberg, R. J.: A Threefold Model of Intellectual Styles. *Ed. Psy. Rev.*, vol. 17 (1), (2005). DOI: 10.1007/s10648-005-1635-4
8. Tracey, M.W.: Design and development research: a model validation case. *Ed. Tech. Res. Dev* 57, 553–571, (2009). DOI 10.1007/s11423-007-9075-0
9. Greffier, F., & Tajariol, F.: Les intelligences multiples: une approche pour analyser les ressources pédagogiques numériques. In: *Proceedings of Environnements Informatiques pour l'Apprentissage Humain*, Toulouse, France (2013).
10. Berners-Lee, T., Hendler, J., & Lassila, O.: The Semantic Web, *Scientific American*, May (2001). <http://www.sciam.com/2001/0501issue/0501berniers-lee.html>
11. Sowa, J. F.: *Knowledge Representation. Logical, Philosophical and Computational Foundations*. Brooks/Cole, CA, (2000).
12. Doan, B.-L. ; Bourda, Y. ; Bennacer, N.: Using OWL to describe pedagogical resources. In: *Proc. Conference on Advanced Learning Technologies*, pp. 916 – 917, (2004).
13. Bennacer, N. , Bourda, Y. Doan, B.-L.: Formalizing for querying learning objects using OWL. In: *Proc. Conference on Advanced Learning Technologies*, pp.321 – 325, (2004).
14. Gruber, T. R.: The Role of Common Ontology in Achieving Sharable, Reusable Knowledge Bases. In: *Proceedings of the 2nd Int. Conf. Principles of Knowledge Representation and Reasoning, (KR & R-91)*, pp.601-602, Morgan Kaufmann, CA, (1991).
15. Boyle, T.: Layered learning design: Towards an integration of learning design and learning object perspectives, *Computers & Education Volume* 54, Issue 3, 661–668, (2010)
16. Sicilia, M.A., Lytras, M., Sánchez-Alonso, S., García-Barriocanal, E.: Modeling instructional-design theories with ontologies: Using methods to check, generate and search learning designs *Comp. in Human Beh., Volume* 27, Issue 4, 1389–1398, (2011).
17. Zheng, Li, Yintao, L., Wang, J., Yang, F.: Multiple Standards Compatible Learning Resource Management. In: *International Conference on Advanced Learning Technologies*, pp.657-661, (2008).
18. Weibel, S.: The Dublin Core: A Simple Content Description Model for Electronic Resources, *Bull. of the Am. Soc. for Infor. Sci. and Tech.*, 24, 1, pp. 9–11, (1997)
19. Dublin Core Metadata Initiative. <http://www.dublincore.org/>.
20. IEEE,; LOM. Draft standard for learning object metadata, (2002). http://ltsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf
21. Neven, F., Duval, E.; Reusable Learning Objects: a survey of LOM-based Repositories. In: *Proceedings of the 10th ACM Conference on Multimedia*, pp.291-294, (2002).
22. MLR – Metadata for Learning Resources, ISO/IEC JTC1 SC36 19788.
23. LOM to MLR element mapping, ISO/IEC JTC1 SC36 WG4 N0188, (2007).
24. Nilsson, M., Palmer, M., Brase, J.: The LOM RDF binding - principles and implementation, 3rd ARIADNE conference, Belgium, (2003).
25. Guarino, N. (ed.): *Formal ontology and information systems*. Proceedings of the 1st International Conference. IOS PRESS, Amsterdam, (1998).
26. Champin, P.A., Prié, Y., Mille, A.: MUSETTE: a framework for knowledge capture from experience. 12^{ème} Atelier de Raisonnement à Partir de Cas, (2004).
27. Settouti, L.S., Prié, Y., Marty, J.C., Mille, A.: A Trace-Based System for Technology-Enhanced Learning Systems Personalisation. *ICALT*, pp.93-97, (2009).
28. Hyvönen, E., Styrman, A., Saarela, S.: Ontology-Based Image Retrieval. In: *Towards the Semantic Web and Web Services, Proceedings of XML Finland Conference*, 15-27, (2002).

The Significance of ‘Ba’ for the Successful Formation of Autonomous Personal Knowledge Management Systems

Ulrich Schmitt

University of Stellenbosch Business School, South Africa
PO Box 802, Gaborone, Botswana

schmitt@knowcations.org

Abstract. “Just as computer science underwent a revolution in the 1980s with the wide-spread use of personal computers, it is possible that Knowledge Management (KM) will in the twenty-first century experience a decentralizing revolution that gives more power and autonomy to individuals and self-organized groups.” Levy’s scenario - nearly seven decades after Vannevar Bush’s still unfulfilled vision of the Memex - stresses the dire need to provide overdue support tools for knowledge workers in the rising Creative Class and Knowledge Societies.

With a prototype system addressing the issues about to be converted into a viable Personal Knowledge Management System (PKMS), the author consolidates recent activities and publications and visualizes and presents a comprehensive Meta-PKMS-Concept covering the main environmental, design, development, deployment, and capacity building concerns.

The paper also portrays its substantial interrelationship with Nonaka’s concept of ‘Ba’ and its impact on the novel PKMS approach aiming to aid team-work, life-long-learning, resourcefulness, and creativity of knowledge workers throughout their academic and professional life and as contributors and beneficiaries of organizational performance.

Keywords: Personal Knowledge Management Systems. Knowledge Worker. Memes. Memex. SECI Model. Ba. Meta-Concept. Knowcations.

1 Pioneering Concepts for Autonomous Personal KM Systems

A current paper-under-review contributes to a theme where the original purposes of established Knowledge Management (KM) Models and Theories are re-examined to determine their value for the next KM generation. After qualifying the Personal KM (PKM) concept and prototype to be introduced as a ‘Next KM Generation’ system, their functionalities are reverse engineered in order to crystallize those KM Models and Theories which – in terms of the PKM system – have passed the test of time.

Forty established KM constructs have been identified as the ingredients for the ongoing integrating, adapting, broadening, deepening, repurposing, or innovating activi-

ties of the PKM system (PKMS) design. Over the past two years, a series of papers and presentations [17-35] have accompanied its progressive development which has been transparently visualized using the Information-Space Model [19, 21, 29, 32, 34].

However, such a comprehensive novel approach also has to be forward engineered and based on a meta-concept which is the focus of this paper. The role of this overarching meta-concept (figure 2) is to integrate the established and re-purposed with the new and innovative KM constructs for meeting the needs encountered and for planning ahead for the productive deployment of the final system.

The PKM concept fully supports the scenarios recently put forward by Levy (section 1.4) but also exemplifies the successful fusion of Bush's seventy year old vision of the Memex (1945), Simon's forty-five year old call for Attention Management (1971), and Nonaka's (2000) fifteen year old notion of Personal Autonomy.

1.1 Vannevar Bush's Vision of the 'Memex' (1945)

In 1945, convinced that "the world has arrived at an age of cheap complex devices of great reliability", President Truman's Director of Scientific Research imagined the 'Memex'. It resembles a hypothetical sort of mechanized private file/desk/library-device, acting as an enlarged intimate supplement to one's memory, and enables an individual to store and recall, study and share the "inherited knowledge of the ages".

It facilitates the addition of personal records, communications, annotations, and contributions and the building of non-fading trails of one's individual interest through the maze of materials available; all of which are easily accessible and sharable with the Memexes of acquaintances. Bush based his vision on the observations of a steadily "growing mountain of research" and an "increased evidence that we are being bogged down" as specialization extends further in the name of progress. He regarded our methods of transmitting and reviewing the results of research "to be generations old" and "totally inadequate for their purpose". Bush foresaw the appearance of new forms of encyclopedias with an extensive mesh of associative multi-disciplinary trails already built-in as well as a new profession of trail blazers who find delight in the task of establishing useful trails through the enormous mass of the common record.

A 'Memex' would provide the means to make intellectual excursions more enjoyable and an individual would "reacquire the privilege of forgetting the manifold things he does not need to have immediately at hand, with some assurance that he can find them again if they prove important" and could enjoy more time for learning or knowledge development. As an added benefit of the trails captured, "the inheritance from the master becomes, not only his additions to the world's record, but includes for his disciples the entire scaffolding by which they were erected" [2].

Unfortunately, a Memex-like device has not yet emerged; in 'Still building the Memex', Davies concludes: "Personal knowledge management [PKM] is [still] a real and pressing problem. [...] Yet it does not appear that Vannevar Bush's dream has yet been fully realized on a wide scale" [6]. Although there are many PKM tools available, "they are not integrated with each other", and the "currently available PKM systems can provide only a partial support to knowledge workers" [14].

1.2 Simon's Notion of 'Attention Management' (1971)

In 'Designing Organizations in an Information-rich World', Simon pointed out that the "wealth of information creates a poverty of attention and [with it] a need to allocate that attention efficiently among the overabundance of information sources that might consume it". Thus, "it is not enough to know how much it costs to produce and transmit information; we must also know how much it costs, in terms of scarce attention, to receive it. [...] In a knowledge-rich world, progress does not lie in the direction of reading information faster, writing it faster, and storing more of it. Progress lies in the direction of extracting and exploiting the patterns of the world – its redundancy – so that far less information needs to be read, written, or stored" [36].

Since then, accelerating technological progress and economic pressures are bringing about organizational, commercial, social and legal innovations with profound impacts on the way we work and live, including the change from information scarcity (few sources/channels, high associated costs) to a never before experienced ever-increasing attention-consuming information abundance. Accordingly, Simon's notion is more relevant than ever. As Kahle observed: "While today we have many powerful applications for locating vast amounts of digital information, we lack effective tools for selecting, structuring, personalizing, and making sense of the digital resources available to us" [10].

1.3 Nonaka's Theory of Organizational Dynamic Knowledge Creation (2000)

Nonaka's theory of organizational dynamic knowledge creation represents one of the most widely cited theories in knowledge management. Its SECI Model (socialization, externalization, combination, and internalization) addresses a major objective of Organizational KM (OKM) by aiming to make the tacit knowledge (gained only experientially and difficult to articulate, explain, share - as opposed to formal or explicit knowledge) of knowledge workers explicit, so it can be measured, captured, stored, protected, shared and further utilized in a 'spiral' of knowledge creation for the benefit of the organization and its stakeholders and independent of the availability of the persons concerned [13].

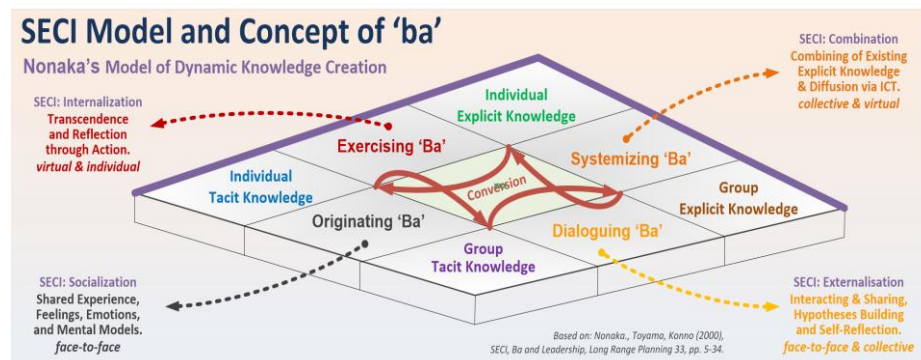


Figure 1: Nonaka's SECI Model and the Space Concept of 'ba' [13]

In the further development of his model, Nonaka defines Knowledge Creation as “a continuous, self-transcending process through which one transcends the boundary of the old self into a new self by acquiring a new context, a new view of the world, and new knowledge” and - in order to guide this journey - introduces the concept of ‘ba’ (figure 1) as a shared context or place (physical or virtual) in which knowledge is shared, created, interpreted and utilized.

To provide adequate organizational leadership for building, connecting, and energizing such dynamic knowledge-creating environments and their expansion (plurality of ‘ba’), Nonaka furnishes management with some recommendations which share a common ground with the PKMS concepts and systems to be introduced, such as to:

- Foster and utilize one’s personal knowledge-related proficiencies to keep track of one’s individual knowledge assets, and utilize them when they are needed.
- Continuously ‘map’ personal knowledge assets by taking account of any new stock dynamically created from one’s existing knowledge assets.
- Intentionally stimulate ‘ba’ by employing adequate physical or virtual space, or mental space such as distinctive goals or a knowledge vision.
- Choose the right mix of collaborators and promote their interaction.
- Preserve the conditions for an enabling environment, such as autonomy¹, creative chaos, redundancy, requisite variety, and love, care, trust and commitment.

1.4 Levy’s Autonomous PKM Capacities and Creative Conversations (2011)

In 2011, Levy envisaged: “just as computer science underwent a revolution in the 1980s with the widespread use of personal computers, it is possible that Knowledge Management (KM) will in the twenty-first century experience a decentralizing revolution that gives more power and autonomy to individuals and self-organized groups”.

His scenario is based on a future of decentralized autonomous PKM capacities, networked in continuous feedback loops. Nourished by the creative conversation of many individuals’ PKM devices, the systems are expected to assume an elementary role that enables “the emergence of the distributed processes of collective intelligence, which in turn feed them”. Accordingly, “one of the most important functions of teaching [...] will therefore be to encourage in students the sustainable growth of autonomous capacities in PKM” [11].

Levy’s notions are in line with Wiig’s view that the “root objective of PKM is the desire to make citizens highly knowledgeable” in order to “function competently and effectively in their daily lives, as part of the workforce and as public citizens” [37]. The thoughts mark a departure from the - until then - narrow individualistic confinement PKM had been placed in [3]. In limiting its scope, it has been labelled as sophisticated career and life management with a core focus on personal enquiry [15] or as a means to improve some skills or capabilities of individuals, negating its importance relating to group member performance, new technologies or business processes [5].

¹ “Autonomous individuals and groups in knowledge-creating organizations set their task boundaries for themselves in pursuit of the ultimate goal expressed by the organization” [13].

2 Knowcations – An Autonomous PKM Prototype System

Due to technological and economic forces, the world of work is changing rapidly² and the “responsibility for self-development and lifelong learning is [said to be] now in the hands of the individual, who increasingly controls the development of his/her career and destiny. [...] In the world of the modern knowledge worker, it has become necessary for individuals to maintain, develop and market their skills to give them any chance of competitive advantage in the job market in both the short and long term” [15] and of participating productively as professionals in the SECI activities of the organizational world realities.

Hence, autonomy is a key virtue in both the world of work as well as the world of education, it increases the initiative and commitment of an individual for finding valuable information, for utilizing existing knowledge and generating new and unexpected knowledge. But, as sovereign professionals move from one project or responsibility to another, they will want to take their version of a knowledge management system with them. Accordingly, an effective PKMS has to support the notion that “knowledge and skills of a knowledge worker are portable and mobile. Unlike manual workers, they have numerous options on where, how, and for whom they will put their knowledge to work” [16].

In following the pioneering concepts presented, the PKM System-in-progress addresses these needs and aims to provide knowledge workers with the overdue support for aiding life-long-learning, resourcefulness, creativity, and team-work throughout their social, academic, and professional life and as contributors and beneficiaries of organizational and societal performance. The artificial name ‘Knowcations’ crafted for the system is made up of KNOW as a reference to knowledge and know-how and CATIONS as an intended association to the locations or spaces as well as to the vocations or abilities which are so vital to further our expertise and careers.

The ‘Knowcations’ idea originated during the author’s PhD studies in the early 90s. Since then, the resulting prototype has been continuously expanded and used personally for career support as a management consultant, scholar, professor, and academic manager. With the development platforms and cloud-based services available now, an innovation opportunity has presented itself for converting and advancing the prototype into a viable PKM system across multiple platforms.

² Work has suffered from a process of fragmentation which will continue to accelerate. Gratton analyzes its implications and highlights the slipping control over constant interruptions, the loss of time for real concentration, and less learning by observation and reflection [9]. With specializations and domain-specific knowledge on the rise, the identification of people has also shifted from their company to their occupation and profession, and “the vertical hierarchy and traditional career ladder have been replaced by sideways career moves between companies, and a horizontal labor market” [7]. Moreover, the 2013 study ‘Future of Employment’ analyzed 702 occupations and estimated about 47% of total US employment to be still at risk due to recent technological breakthroughs able to turn previously non-routine tasks into well-defined problems susceptible to computerization [8]. Similar predictions have been put forward for Europe by Bowles [1].

3 The Benefits of a Comprehensive System Concept and Design

The complexities involved in realizing *Bush's* functionalities and in addressing *Simon's* attention-related concerns as well as the challenges³ posed by *Nonaka's* plea for autonomy and *Levy's* push for creative conversations and capacity development call for an all-embracing high-level system concept and design. Its aim is to mobilize and motivate the relevant audience and instill in them a sustainable commitment to interact with and/or promote the PKMS technologies for keeping personal human capitals à-jour (see Mostert's Six Levels of Appreciation [12] presented in [23, 25]).

Fortuitously, what might have appeared initially as difficult to reconcile or conflicting (e.g. objectives, philosophies, methodologies) has found its way into an overarching meta-concept which even provided the system's name 'Knowcations' with a further lease of life as an acronym as well as a heuristic method for similar conceptual design challenges. The present state of this meta-concept has been visualized in a clock-like fashion in figure 2 (numbers 1 to 12 used refer to position on a watch dial).

Table 1: Legend for right-hand (1-6) and left-hand (7-12) side of figure 2 (* = in-work)

Topics	Legend	With particular focus on:	Paper
1	K	Knowledge & Human Development	Four Challenges
2	N	Nescience & Ignorance Matrix	- Acquisition/Preservation
3	O	Occupation & Capacity Development	- Capacity Development
4	W	Workflows & Cycles in Information-Space	- Collaborative
5	C	Concepts & Circuits in the Meme Ideosphere	- Conceptual Level
6	A	Autonomous Devices, Barriers & Provisions	Transformation
7	T	Technology & Systems Design	• Combination
8	I	Interactions & Creative Conversations	• Externalization
9	O	Outcomes & Value Chains	• Socialization
10	N	Systems Thinking & Creative Solutions	• Internalization
11	S	Individualization versus Institutionalization	Meta-Level, Meta-'Ba'
12	!		Integration of PKM/OKM

³ The novel PKMS approach entails a departure from the current heavyweight, prohibitive, centralized, top-down, institutional developments in order to strengthen individual sovereignty by employing grass roots, bottoms-up, lightweight, affordable, personal applications. The needs have been summarized in form of Five Vital Provisions [23]:

- a. Digital personal and personalized knowledge is always in possession and at the personal disposal of its owner or eligible co-worker, residing on personal hardware and/or personalized cloud-databases.
- b. Contents are kept in a standardized, consistent, transparent, flexible, and secure format for easy retrieval, expansion, sharing, pooling, re-use and authoring, or migration.
- c. Information and functionalities can continually be used without disruption independent of changing one's social, educational, professional, or technological environment.
- d. Collaboration capabilities have to be mutually beneficial to facilitate consolidated team and enterprise actions that convert individual into organizational performances.
- e. The PKMS design and its complex operations are based on a concept, functionalities, and interventions which are clearly understood and are painlessly applied in practice.

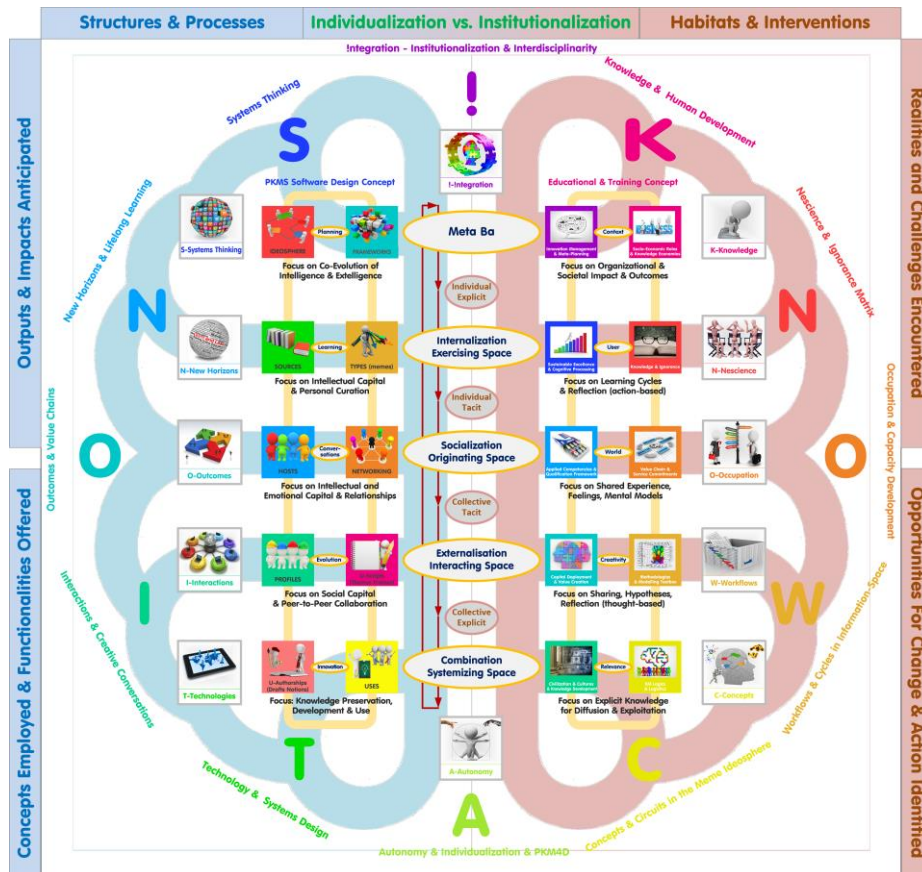


Figure 2: Visualizing the 'KNOWCATIONS!' overarching Meta-Concept

The 'KNOWCATIONS!' Meta-Concept visualizes closely associated aspects:

- The twelve topics (clock-wise from 1 to 12 o'clock) deal - in the appropriate sequence - with the realities and challenges encountered, the opportunities for changes and actions identified, the concepts employed and functionalities offered, and the outputs and impacts anticipated (legends right and left of figure 2).
- Thus, the right-hand side (1-5) represents the external system-independent environment and the left-hand side (7-11) the central features and objectives of the PKM system technicalities. Furthermore, circumstances and barriers preventing and provisions enabling the transition from the current inadequate state (right) to the recommended scenario (left) are given (6) and the importance of integrating personal and organizational KM is emphasized (12).
- Due to the closely related issues addressed by the topics (challenges on the right versus solutions on the left), topics can be paired (1&11, 2&10, 3&9, 4&8, and 5&7) as indicated by the bridging ellipses between the right and the left hemisphere.
- The system analysis and design phase covering these topics has resulted in the development of the PKMS prototype (depicted as a separate dark yellow rounded rec-

tangle in the middle of the left hemisphere) whose main entities are shown as icons placed appropriately on the bridges connecting them to their most relevant topics.

- Similarly, key content to be addressed in the accompanying roll-out and educational context have been exemplified and positioned (depicted as a further separate dark yellow rounded rectangle in the middle of the right hemisphere).
- Accordingly and in a reversal of the initial left-right analysis-design direction, a right-left development-deployment path is shown from the system to the environment.
- The five bridges with their distinctive environment-design-development-deployment-links are passing through a further distinctive section each (five larger ellipses in the center section) which themselves are interconnected and represent Nonaka's SECI model and 'Ba' concept, individually and in their plurality.

The Meta-Concept depicted provides a good example of how ongoing analytical and design activities can inform the setting of a blueprint which then, in turn, informs the further integration and advancement of all system components. As a consequence:

- Terminology, colors, acronyms, and icons are used consistently across system layers and dimensions resulting in a unified and coherent user familiarization/experience.
- Prior publications have been individually pitched at each of the 12 topics (as indicated in table 1). Their meme-based⁴ representations have been integrated in the PKMS knowledge bases and form part of the test data for system validation. They are providing not only assistance and help functionalities for the user in the deployment phase but also case study examples and knowledge structures for further authorship.
- This closely integrated content is also setting the stage for a PKM/OKM lecturing concept serving higher education and professional training; it enables a delivery based on a cohesive, tried and tested framework rather than on the fragmented collections of KM notions which typify too many Knowledge Management course outlines. To paraphrase Bush: "As an added benefit of the trails captured, the [content provided by the developer] becomes not only his additions to the world's record, but entails for his [customers] the entire scaffolding by which they [have been] erected" [2].
- The visual representations and mappings in the Information-Space offer not only a concise system overview for users but also provide a road map together with - as Bush put it - "an extensive mesh of associative multi-disciplinary trails already built-in of alternative pathways" [2]. These structures will conveniently accommodate the establishment and navigation of eLearning courses, planned after the completion of the face-to-face course design.

⁴ The novel distributive facet of the PKMS has a dual dimension. The first, from centralized organizational OKM systems to decentralized personalized PKM devices, has already been alluded to. The second, from document-centric repositories to meme-based knowledge bases is portrayed in the second conference paper [18]. Rather than whole documents, what is referred to and what is stored at the same time, is smaller and more distinct, a basic building block of knowledge in the eyes of the beholder (a meme). The information-structure captured should be perfectly understandable alone by itself without piggybacking irrelevant or potentially redundant information. In this quasi-atomic state it can be repeatedly utilized (and traced back) at any later stage in combination with other memes to form memplexes or knowledge assets, defined as "nonphysical claims to future value or benefits" [4].

4 The Concept of ‘Ba’ as the Underpinning of ‘Knowcations’

At the heart of the Meta-Concept in figure 2 is the SECI-model which also has been depicted in the ‘I-Space’ model alluded to [29]. Although the analysis and design activities have not been based initially on Nonaka’s concepts, this center stage has become apparent due to the closeness and correlation with the forthcoming findings and design choices. The four ‘ba’ and their plurality emphasize the shared contexts or places in which the PKMS philosophy is implemented, understood, followed, shared, utilized, and further developed as well as the practice of repositing, referencing, recalling, retrieving, reusing, repurposing, relating, and reflecting.

While the ‘Exercising Ba’ addresses the development of the PKMS user’s intellectual, social, and emotional capital via study and desktop research, the ‘Originating Ba’ is present in the social interaction during field research and teamwork activities. The ‘Systemizing Ba’ surfaces as Human-Computer Interface of the PKMS and its cloud-based networking endeavors, and the ‘Dialoguing Ba’ is represented during the dissemination of the system outputs and related feedbacks and, of course, by the Creative Conversations referred to by Levy.

The ‘Ba’ spaces are equally relevant at on-line and off-line PKMS usage times⁵ as well as during user training and education. They provide an important linkage between personal and organizational KM philosophies (as referred to in section 1.3), but also apply at the grass-roots level of repurposing and combining relevant memes to form knowledge assets:

Memes “can either be encoded in durable vectors (e.g. storage devices, books, great art, major myths, or artefacts) spreading almost unchanged for millennia, or they succeed in competing for a host’s limited attention span to be memorized (internalization) until they are forgotten, codified (externalization), or spread by the spoken word to other hosts’ brains (socialization) with the potential to mutate into new variants. To gain an advantage in competing for attention and survival, it pays to form symbiotic relationships (combination) with other memes (memeplexes) to mutually support each other’s fitness and to replicate together” [18].

Hence, in fostering the ‘individualization versus institutionalization’ integration, Nonaka’s notion of Experiential, Routine, Conceptual, and Systemic Knowledge Assets as “firm-specific resources that are indispensable to create value” [13] is equally important in the PKMS context and its authorship capabilities. To add a hands-on perspective and example to the conceptual meme-based knowledge creation cycles, a recent ‘demonstration’ paper has utilized the prototype for the paper’s development which, in turn, has described the iterative process steps involved in its creation [25].

⁵ In contrast to the various knowledge models published, the PKMS activities have been differentiated into eight on-line-system-supporting ‘PROJECTS’ tasks (**p**rojecting, **r**eferencing, **o**riginating, **j**ustifying, **e**videncing, **c**ontexting, **t**opicing, and **s**cripting) and eight off-line external ‘ALFRESCO’ tasks (**a**dministering, **l**ocating, **f**eed-backing, **r**eflecting, **e**ngaging, **s**ourcing, **c**rafting, and **o**bserving). The system, however, also provides functionalities to plan, manage, document, track, and evaluate the external tasks.

5 The Road ahead

The PKMS concept and prototype-in-progress add a further level of detail to the JAIST Nanatsudaki Model with its sequence of seven waterfalls or seven creative spirals which describes academic/individually-oriented as well as practice/group-oriented knowledge creation processes [38] and which has been aligned to the PKMS's information-space cycles in a forthcoming post-proceedings paper [17].

After finalizing the prototype test phase, an estimated 12 months migration to a viable PKM application is envisaged based on a production environment for cross-platform mobile, web and cloud applications. In parallel, the series of recent conference sessions⁶ and publications will be concluded in line with the meta-concept:

- *Meta Ba*: The Meta-Concept described is firmly rooted in two prior publications. The very first paper identifies four challenges for individuals managing their knowledge at acquisition/preservation, collaborative, capacity development, and conceptual level and looks at approaches to address them [35] and another which positions PKM systems in the context of human development [27]. This paper described the Meta-Concept in detail and will be followed by an exploration of PKM's impact on social, cultural, economic, and environmental sustainability.
- *Exercising Space*: Two papers introduce an extended ignorance matrix and discuss the role of PKM systems in making citizens highly knowledgeable [31, 26], followed by the 'demonstration' paper mentioned [25]. It is planned to publish suggestions for a KM course structure and e-learning system in support of the educational needs Levy highlighted; a respective three-hour conference tutorial has already been scheduled.
- *Originating Space*: Two papers and a presentation look at the workplace and careers of knowledge workers in the business, educational, and developmental context [30, 28, 22]. They just have been concluded by presenting a framework of twelve PKM for Development (PKM4D) criteria (aligned to Maslow's Extended Hierarchy of Needs) which explored the close relationship between PKM and e-Skills [20].
- *Interacting Space*: Two papers and a presentation demonstrate the close proximity of the PKM concept and functionalities with renowned KM models and established organizational KM practices [33, 29, 24]. A follow-up will explore the benefits of PKMS as a "Structural Intellectual Capital Asset" [37] for teamwork and leadership.
- *Systemizing Space*: The integration of the models and workflows in the context of the meme-based PKMS concept and its information-space have been presented [32, 18, 19]. A paper-in-progress adds a further perspective by contemplating how the combination's innovative potential is able to invigorate digital scholarship, individual and institutional curation, and the traceability of knowledge in a new configuration of citation systems. A more technical paper will compare the PKMS prototype to other related recent system development activities as well as semantic technologies.

⁶ In acknowledging the trans-disciplinarity of the PKM notion, the papers and presentations received feedback from and addressed a wide scope of conference themes, including Knowledge Management and Knowledge Technologies, Management and Social Sciences, Higher Education and Human Resource Development, Innovation and Creativity Support Systems, e-Skills and e-Learning, as well as Organizational Learning and Future Studies.

References

1. Bowles, J.: The Computerisation of European Jobs. Bruegel, Retrieved from www.bruegel.org/nc/blog/detail/article/1394-the-computerisation-of-european-jobs/ (2014)
2. Bush, V.: As we may think. The Atlantic Monthly, Issue 176.1, pp. 101-108. (1945)
3. Cheong, R.K., Tsui, E.: From Skills and Competencies to Outcome-based Collaborative Work. Knowledge and Process Management, 18(3), pp. 175-193 (2011)
4. Dalkir, K.: Knowledge Management in Theory and Practice, Butterworth-Heinemann (2005)
5. Davenport, T. H.: Personal Knowledge Management and Knowledge Worker Capabilities. In: Pauleen, D. J., Gorman, G. E. (eds) Personal KM. Gower, pp. 167-188 (2011)
6. Davies, S.: Still Building the Memex. Communications of the ACM, 53(2), pp. 80-88 (2011)
7. Florida, R.: The Rise of the Creative Class – Revisited. Basic Books (2012)
8. Frey, C. B., Osborne, M. A.: The Future of Employment. Oxford Martin (2013)
9. Gratton, L.: The Shift - The Future of Work is Already Here. HarperCollins UK (2011)
10. Kahle, D.: Designing Open Educational Technology. In: Iiyoshi, T., Vijay Kumar, M. S. (eds.) Opening up Education. MIT Press (2009)
11. Levy, P.: The Semantic Sphere 1. Wiley. pp. 115-116 (2011)
12. Mostert, M.: Systemic Leadership Learning, Knowres Publishing (2013).
13. Nonaka, I., Toyama, R., Konno, N.: SECI, Ba and Leadership: a Unified Model of Dynamic Knowledge Creation. Long Range Planning 33, pp. 5-34 (2000)
14. Osis, K., Gaidnsenkis, J.: Modular personal knowledge management system and mobile technology cross-platform solution towards learning environment support. Annual Int. Conference on Virtual and Augmented Reality in Education (VARE). pp. 114-124 (2011)
15. Pauleen, D.J., Gorman, G.E.: The Nature and Value of Personal Knowledge Management. In: Pauleen, D. J., Gorman, G. E. (eds) Personal KM. Gower, pp. 1-16 (2011)
16. Rosenstein, B.: Living in More Than One World. Berrett-Koehler Publishers. p. 53 (2009)
17. Schmitt, U.: Knowcations - Positioning a Meme and Cloud-based 2nd Generation Personal Knowledge Management System. Forthcoming Post-Proceedings of the KICSS'2013 Conference. Springer Series: Advances in Intelligent Systems and Computing, 2194-5357 (2014)
18. Schmitt, U.: Significance of Memes for the Successful Formation of Autonomous PKMS. 9th International Conference on Knowledge, Information and Creativity Support Systems (KICSS), Nov 06-08, 2014, Limassol, Cyprus. (2014)
19. Schmitt, U.: Who needs Personal Knowledge Management anyway and what for? Poster Presentation at the eSkills for Knowledge Production and Innovation Conference and e-Summit of the South African iKamva National e-Skills Institute (iNeSI) and the Informing Science Institute (ISI), Nov 17-21, 2014, Cape Town, South Africa. (2014).
20. Schmitt, U.: Making Sense of e-Skills at the Dawn of a New Personal Knowledge Management Paradigm. Proceedings of the e-Skills for Knowledge Production and Innovation Conference, November 17-21, 2014, Cape Town, South Africa. (2014)
21. Schmitt, U.: Proposing a Next Generation of Knowledge Management Systems for Creative Collaborations in Support of Individuals and Institutions. In Proceedings of the 6th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K), Oct 21-24, 2014, Rome, Italy. (2014).
22. Schmitt, U., Butchart, B.: Making Personal Knowledge Management Part and Parcel of Higher Education Programme and Services Portfolios. In Journal of the World Universities Forum. Vol. 6(4) pp. 87-103, <http://wuj.cgpublisher.com/product/pub.173/prod.392> (2014)
23. Schmitt, U.: Overcoming the Seven Barriers to Innovating Personal Knowledge Management Systems. In Proceedings of the International Forum on Knowledge Asset Dynamics IFKAD, June 11-13, 2014, Matera, Italy. 978-88-96687-04-8 (2014)

24. Schmitt, U.: From Circuits of Knowledge to Circuits of Personal Knowledge Management Concepts (Presentation). Paper presented at the Conference on Organizational Learning, Knowledge and Capabilities OLKC. April 22-24, 2014. Oslo, Norway. (2014)
25. Schmitt, U.: How this Paper has been created by leveraging a Personal Knowledge Management System. Proceedings of the 8th Int. Conference on Higher Education ICHE, March 16-18, 2014, Tel Aviv, Israel. <http://www.intconfhighered.org/Webpage%20Schmitt.pdf> (2014)
26. Schmitt, U.: The Role of Personal Knowledge Management Systems in making Citizens highly knowledgeable. In Proceedings of the 8th International Technology, Education and Development Conference INTED, March 10-12, 2014, Valencia, Spain. 978-84-616-8412-0, <http://library.iated.org/view/SCHMITT2014ROL> (2014)
27. Schmitt, U.: Personal Knowledge Management Devices - The next Co-evolutionary Driver of Human Development?! In Proceedings of the International Conference on Education and Social Sciences INTCESS14, February 03-05, 2014, Istanbul, Turkey. 978-605-64453-0-9, <http://www.ocerint.org/cd/abstracts/a384.pdf> (2014)
28. Schmitt, U.: Leveraging Personal Knowledge Management Systems for Business and Development. Paper presented at the 2nd Conference Africa Academy of Management (AFAM), Jan 9-11, 2014, Gaborone, Botswana. (2014)
29. Schmitt, U.: Knowcations - Conceptualizing a Personal Second Generation Knowledge Management System. In: Skulimowski, A. M. J. (ed). Looking into the Future of Creativity and Decision Support Systems. Proceedings of the 8th Conference on Knowledge, Information and Creativity Support Systems KICSS, Nov 07-09, 2013, Krakow, Poland. 978-83-912831-6-5. <http://da.xmachina.gr/Proceedings/KICSS2013/files/papers/paper5.pdf> (2013)
30. Schmitt, U.: Furnishing Knowledge Workers with the Career Tools they so badly need. In International HR Development Conference HRDC. October, 17-18, 2013, Mauritius. 1694-2140, <http://www.hrd-conference.com/images/site/Papers/Conference%20Proceedings%20-%20Ulrich%20Schmitt.pdf> (2013)
31. Schmitt, U.: Managing Personal Knowledge to make a Difference. In Proceedings of the 27th British Academy of Management Conference BAM, Liverpool, Sep 10-12, 2013. 978-0-9549608-6-5 (2013)
32. Schmitt, U.: Managing Personal Knowledge for the Creative Class Economy (Poster). World-Future 2013 - Exploring the Next Horizon, July 19-21, 2013, Chicago, U.S.A. (2013)
33. Schmitt, U.: Innovating Personal Knowledge Creation and Exploitation. In Proceedings of the 2nd Global Innovation and Knowledge Academy GIKA, Jul 9-11, 2013, Valencia, Spain. 978-1-909507-26-5 (2013)
34. Schmitt, U.: Knowcations - A Meme-based Personal Knowledge Management System-in-Progress. In Proceedings of the 8th International Conference on e-Learning ICEL, ACPI, June 27-28, 2013, Cape Town, South Africa. 978-1-909507-26-5 (2013)
35. Schmitt, U.: Knowcations - The Quest for a Personal Knowledge Management Solution. In 12th International Conference on Knowledge Management and Knowledge Technologies i-Know, Sep 05-07 2012, Graz, Austria. Copyright 2012 ACM 978-1-4503-1242-4/12/09. Graz: ACM. <http://dl.acm.org/citation.cfm?id=2362469>. (2012)
36. Simon, H. A.: Designing Organizations for an Information-Rich World. In: M. Greenberger (Ed.) Computers, Communication, and the Public Interest. Baltimore, Johns Hopkins Press (1971)
37. Wiig, K.M.: The Importance of Personal Knowledge Management in the Knowledge Society. In: Pauleen, D. J., Gorman, G. E. (eds) Personal Knowledge Management. Gower, pp. 229-262 (2011)
38. Wierzbicki, A. P. & Nakamori, Y.: Creative Environments: Issues of Creativity Support for the Knowledge Civilization Age. Springer Publishing Company. (2007)

The Significance of Memes for the Successful Formation of Autonomous Personal Knowledge Management Systems

Ulrich Schmitt

University of Stellenbosch Business School
PO Box 802, Gaborone, Botswana

schmitt@knowcations.org

Abstract. Recent papers based on a prototype-in-progress focused on advancing Personal Knowledge Management (PKM) as the next generation of Knowledge Management Systems (KMS). Based on assumptions of autonomous capacities engaged in creative conversation, the personal devices are supposed to enable the emergence of the distributed processes of collective extelligence and intelligence, which in turn feed them. This shift from centralized organizational KMS to decentralized personalized KMS devices is expected to give more power and autonomy to individuals and self-organized groups.

A second novel feature proposed is the substitution of document-centric repositories with meme-based knowledge bases. While a parallel conference paper has considered the significance of Nonaka's concept of 'Ba' for the overarching meta-concept of the PKM system-in-progress, this contribution concentrates on the particular aspect of memes and introduces their relevant features and potential allowing for information-rich, multi-dimensional information structures and trails as well as for more elaborate dissemination concepts and citation systems.

Keywords: Personal Knowledge Management Systems. Knowledge Worker. Memes. Memetics. Digital Scholarship. Citation Systems. Knowcations.

1 The Inadequacy of Today's Reference and Citation Systems

Traceability forms the backbone of quality management in modern manufacturing and stands for the ability to trace the history, application or location of an entity. "For discrete manufacturers, this means you can track and trace each component that comprises your product - from suppliers and manufacturers through assembly and final delivery to customers by creating an as-built genealogy. And for food and other process manufacturers, it's the ability to trace each ingredient of a product from 'farm to fork' through the creation of a batch genealogy" [14].

Unfortunately, for knowledge seekers and producers such an enabling state of management did not emerge with the advances and widespread affordability of Information and Communication Technologies (ICT). Instead, we are faced with a never before experienced ever-increasing attention-consuming information abundance.

It is still the academic paper-based citation system which dominates today's knowledge tracking and it has been - since the 17th century - the basis for the reputation economy in science. It "allows scientists to build on the earlier work without having to repeat that work [and] the citation both credits the original discoverer, and provides a link in a chain of evidence" [15].

In motivating his vision of the 'Memex',¹ seven decades ago, Bush already called these methods of transmitting and reviewing our research results "to be generations old" and "totally inadequate for their purpose" [4]. In order to take advantage of today's online realities, Nielsen calls for a "New Era of Networked Science" and urges removing barriers that prevent potential contributors from engaging in a wider sharing and faster diffusion of their ideas, sources, data, work-in-progress, preprints, and/or code for the benefit of more rapid iterative improvement [15]. Several of these counterproductive barriers have been identified, including incompatible formats and structures, fragmentations and inconsistencies, redundancies and discouraging services as well as the prohibitive institutional approaches and centralized developments controlled by large corporations interested in captured audiences [24] and in "knowledge being shut up in silos and balkanized within small closed communities" [13]. Missing, broken, or pretentious web links or references add further to the sorry state of inefficiencies, but so do the current reference and citation systems in general.

"Any correct reference indicates a discrete source (e.g. book, article, web site) with a page number or access date added only occasionally. It represents a granularity which might have been adequate in the stable context of paper-based worlds, but is far from sufficient in a volatile digital world where contents referenced are constantly altered or erased. It also is far too unrefined for the potential capabilities of digital Personal Knowledge Management Systems (PKMS).

A PKMS is expected to serve its master over a lifetime of educational and professional careers and is supposed to constantly evolve in the process. What we have to refer to and what we need to store at the same time, has to be smaller and more distinct, a basic building block of knowledge in the eyes of the beholder. Captured best in a quasi-atomic state, the information-structure should be perfectly understandable alone by itself but be able to be used at any later time in combination with other building blocks stored without piggybacking irrelevant or potentially redundant information" [24]. This basic building block of knowledge is also referred to as a meme.

¹ In 1945, Vannevar Bush (then President Truman's Director of Scientific Research) imagined the 'Memex', a hypothetical sort of mechanized private file/desk/library-device. It is supposed to act as an enlarged intimate supplement to one's memory, and enables an individual to store, recall, study, and share the "inherited knowledge of the ages". The 'Memex' would have facilitated the addition of personal records, communications, annotations, and contributions, but, above all, the recording of non-fading trails of one's individual interests through the maze of materials available - all easily accessible and sharable with the 'Memexes' of acquaintances. In the seventy years since Bush's famous article [4], a 'Memex' has never been realized [6, 11, 17] but it is one of the inspirations for the PKM system proposed [21].

2 It is a Meme's World

Memes were originally described by Dawkins as units of cultural transmission or imitation [7]. They are (cognitive) information-structures [1] that evolve over time through a Darwinian process of variation, selection and transmission [5]. Able to self-replicate by utilizing mental storage in human hosts, they influence their hosts' behavior to promote further replication [1]. From the meme's-eye view, every human is a machine for making more memes, a vehicle for propagation, an opportunity for replication and a resource to compete for [2]. But, memes exist only virtually and have no intentions of their own; they are merely information pieces in a feedback loop with their longevity being determined by their environment [5].

To live on, a meme has to be able to survive in the medium it occupies and the medium itself has to survive [1]. They can either be encoded in durable vectors (e.g. storage devices, books, great art, major myths, or artefacts) spreading almost unchanged for millennia [1], or they succeed in competing for a host's limited attention span to be memorized (internalization*) until they are forgotten, codified (externalization*) or spread by the spoken word to other hosts' brains (socialization*) with the potential to mutate into new variants (figure 1). To gain an advantage in competing for attention and survival, it pays to form symbiotic relationships (combination*) with other memes (memplexes) [9] to mutually support each other's fitness and to replicate together (*-markings refer to comparable SECI Model stages [16, 21]).

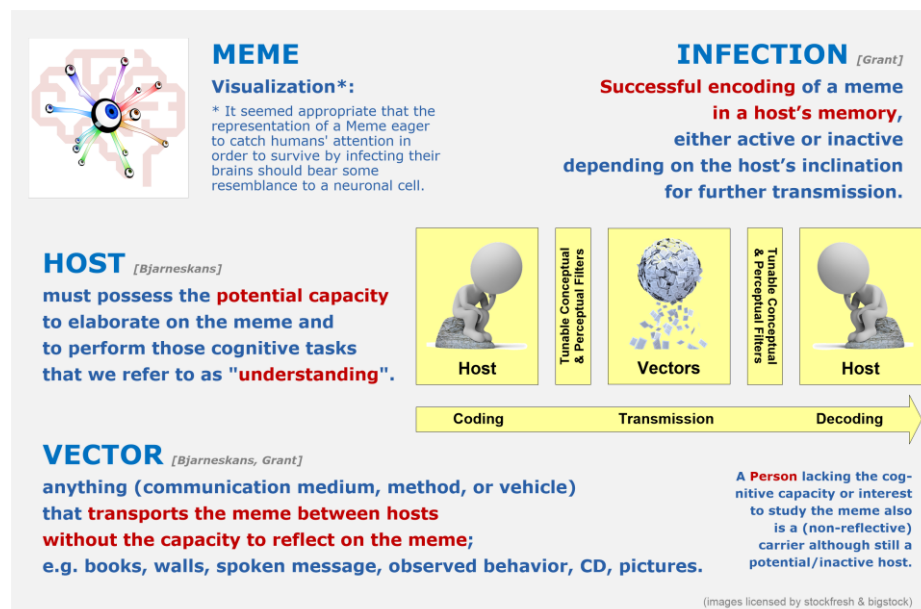


Figure 1: The Transmissions of Memes between Hosts via Vectors [25, 1, 9]

Hughes adds the point that human evolution and memes' endurances have not only thrived on big brain memory and communication technology with a high degree of accuracy but also on peoples' insatiable urge to use these technologies for the purposes intended [10]. A recent paper magnifies this perspective and argues that the progress of human development is primarily due to successive technological co-evolutions. At each stage, civilization runs into constraints which could only be overcome with the emergence of a further powerful co-evolution [27].

Today's constraint in need to be dealt with is clearly the over-abundance of information. Instead, the 'breakthrough' technology under way, termed the 'Industrial Internet' or 'Internet of Things', will exploit the opportunities of cloud-based platforms and applications by incorporating networked sensors and software into goods and machines resulting in the self-organizational capability of complex value chains. As a consequence, ever more information will be generated - further defeating the very attention² our cognitive capabilities are able to master [27].

3 The Ideosphere as the Memes' Habitat of Operation

Memetics studies ideas and concepts viewed as 'living' organisms, capable of reproduction and evolution in an 'Ideosphere' [18], an "invisible but intelligible, metaphysical sphere of ideas and ideation" where we engage in the creation of our world. "This means that the substance of the world is idea, which forms, reforms, and transforms itself via the conversations of humankind, synergetically organizing itself as an evolutionary, multidimensional network [with technology just an artefact of idea]. The problem, however, is that the majority of 'humanity remains the consumer of ideas without being the producer'" [12]. Hence, Kimura calls for an ideospheric transformation set off by a synergetic phenomenon that emerges "when individuals in sufficient numbers become authentic, independent thinkers, that is, originators of ideas, producers of dialogues, and contributors to the network of conversations that comprises the world" [12]. Fittingly, Levy stresses the "need for a personal discipline for collection, filtering and creative connection (among data, among people, and between people and data flows)" and the need for education "to encourage in students the sustainable growth of autonomous capacities in Personal Knowledge Management" [13].

But, if memes and their inbuilt ideas are able to flourish in a virtual 'Ideosphere' as their habitat of operation, PKM Systems aiming at supporting individual capacity and repertoire for innovation, sharing and collaboration are well advised to utilize the very same space and resources and to form a digital counterpart of this 'Ideosphere'.

In this respect, the three-dimensional Information-Space Model [3] which has been used to visualize the PKMS's entities, workflows, and its integrated KM models [28, 22, 19] also represents this 'ideosphere', and its memes and memeplexes are displayed in their amalgamated states as the PKMS user's knowledge assets and his/her intellectual, social, and emotional capital [20].

² Simon already noted over forty years ago, "the wealth of information is creating a poverty of attention and with it a need to allocate that attention efficiently among the overabundance of information sources that might consume it" [30].

4 The Modes of Operandi of a Meme-based PKMS

A PKMS is expected to serve its user unceasingly over life-long-learning periods of educational, professional, social and private activity and experience. Hence, any meme captured has to be enabled to further evolve during learning processes as well as to form the symbiotic relationships with other relevant memes as alluded to.

Hence, a meme-based PKMS focuses on the relevant textual, visual, audio, or video memes a document or message contains, conserves them with their relevant frames of references (e.g. origins, titles, formats, licenses), embedded in a more-dimensional classification system for subsequent easy retrieval, and as a pure, pre-edited, re-purposed, and/or already re-combined version according to one's individual preferences and objectives. By capturing and referencing these basic information units, the PKMS allows recalling, sequencing and combining already stored memes with one's own new meme creations for integration in any type of authoring and sharing activity one would like to pursue. As a result, the user obtains the means to retain and build upon knowledge acquired in order to sustain personal growth and facilitate productive collaborations between fellow learners and/or professional acquaintances.

Thus, known or newly imagined memes are combined during authorship or sense-making sessions into increasingly complex memplexes or knowledge assets (e.g. articles, presentations, or scripts). As Distin points out: "In recombination, existing memes are appropriately recombined in new situations, creating new ways of thought and novel effects, perhaps as the result of previously recessive memes' 'effects' being revealed in the reshuffle". As a result of mutations, "copies will not always be exact, and the idea or skill in question may change in some way en route" [8].

Dawkins [7] points out three qualities of a meme to enhance its fitness in order to maintain a continued presence in future generations³: Fecundity, Longevity, and Copying Fidelity. All three features are profiting extensively from the secure, convenient, and standardized storage in the PKMS's knowledge bases and from the creative conversations between networked autonomous PKMS devices [13].

What this means for creative authorship has been exemplified in a recent 'demonstration' paper which added a hands-on perspective to the conceptual Information-Space cycles by utilizing the PKMS prototype for the paper's development; it describes the iterative process steps involved in its very own creation (25). However, while the resulting published paper version adopts a book-like linear format, its internal virtual structure in the knowledge base remains multi-dimensional, information-rich, and its memes are ready to be re-used, rephrased, re-purposed, and shared - with their trails in place to be visited and revised. As envisaged in the 'Memex', "the inheritances from the master become, not only his additions to the world's record, but includes for his disciples the entire scaffolding by which they were erected" [4].

³ Fecundity asserts that the more copies of itself a meme can encode in vectors, the higher its fitness. Longevity refers to a meme's replication rate which has to be appropriate to its environment and depends on cost-benefit ratios. Copying Fidelity rests on the accuracy with which a meme is copied resulting in either the integrity of information, sensible variations, or erroneous mutations. [7]

5 Conclusions for the Road ahead

The novel meme-based Personal KMS approach incorporates a number of appealing outcomes which will be further explored in upcoming and planned papers:

- It provides solutions to the challenges faced by knowledge workers identified in the initial paper [29]. They just have been reviewed against a newly devised framework of twelve PKM for Development (PKM4D) criteria (aligned to Maslow's Extended Hierarchy of Needs) in the context of e-Skills [20].
- It offers opportunities in the educational context [23] to be further elaborated on in the context of digital scholarship, mentorship, individual and institutional curation.
- It further presents synergies with life-long-learning in form of face-to-face or e-learning coursework. With all papers captured in their meme-based representations and their information-rich references to established KM models, theories, and practices, it is planned to publish suggestions for a KM course structure in support of the PKM concept; a respective three-hour conference tutorial has already been accepted.
- A paper-in-progress follows up on the issue of knowledge traceability raised and contemplates the potential impact of interlinked and shared autonomous meme-based knowledge repositories for the transformation of citation systems contributing to the "New Era of Networked Science" as advocated by Nielsen [15].
- With the PKMS's envisaged role of making citizens highly knowledgeable [26], a paper-in-progress utilizes the PKM4D framework for an inquiry of the PKM's impact on professions and society (social), autonomy and collaboration (cultural), extelligence and leadership (economic), and personalized habitats (environmental).

References

1. Bjarneskans, H., Grønnevik, B., Sandberg, A.: The lifecycle of memes. <http://www.aleph.se/Trans/Cultural/Memetics/memecycle.html> (1999)
2. Blackmore, S. J.: The power of memes. *Scientific American*, Vol. 283(4), pp. 52-61. (2000)
3. Boiso, M.: Exploring the information space: a strategic perspective on information systems. Working Paper Series WP04-003. University of Pennsylvania. (2004)
4. Bush, V.: As we may think. *The Atlantic Monthly*, Issue 176.1, pp. 101-108. (1945)
5. Collis, J.: Introducing Memetics. <http://meme.sourceforge.net/docs/memetics.php>. (2003)
6. Davies, S.: Still Building the Memex. *Communications of the ACM*, 53(2), pp. 80-88 (2011)
7. Dawkins, R.: *The Selfish Gene*. Paw Prints (1976)
8. Distin, K.: *The Selfish Meme*. Cambridge University Press. (2005)
9. Grant, G., Sandberg, A. and McFadzean, D.: Memetic. <http://www.lucifer.com/virus/memlex.html>. (1999)
10. Hughes, J.: *On the Origin of Tepees*. Free Press. (2011)
11. Kahle, D.: Designing Open Educational Technology. In: Iiyoshi, T., Vijay Kumar, M. S. (eds.) *Opening up Education*. MIT Press (2009)
12. Kimura, Y. G.: Kosmic Alignment - A Principle of Global Unity. *Kosmos Journal*. http://www.via-visioninaction.org/via-li/articles/Kosmic_Alignment.pdf. (2005)
13. Levy, P.: *The Semantic Sphere I*. Wiley. pp. 115-116 (2011)

14. Motorola: Quick Reference Guide: Traceability for Manufacturing. Motorola Inc. (2011)
15. Nielsen, M.: Reinventing Discovery - The New Era of Networked Science. Princeton University Press. (2011)
16. Nonaka, I., Takeuchi, H.: The Knowledge-Creating Company. Oxford University Press. pp. 71-2, 89 (1995)
17. Osis, K., Gainspenkis, J.: Modular personal knowledge management system and mobile technology cross-platform solution towards learning environment support. Proceedings of the Annual International Conference on Virtual and Augmented Reality in Education (VARE). pp. 114-124 (2011)
18. Sandberg, A.: Memetics. <http://www.aleph.se/Trans/Cultural/Memetics>. (2000)
19. Schmitt, U.: Knowcations - Positioning a Meme and Cloud-based 2nd Generation Personal Knowledge Management System. Forthcoming Post-Proceedings of the KICSS'2013 Conference. Springer Series: Advances in Intelligent Systems and Computing, 2194-5357 (2014o)
20. Schmitt, U.: Making Sense of e-Skills at the Dawn of a New Personal Knowledge Management Paradigm. Proceedings of the e-Skills for Knowledge Production and Innovation Conference, November 17-21, 2014, Cape Town, South Africa. (2014k)
21. Schmitt, U.: The Significance of 'Ba' for the Successful Formation of Autonomous Personal Knowledge Management Systems. 9th International Conference on Knowledge, Information and Creativity Support Systems (KICSS), Nov 06-08, 2014, Limassol, Cyprus. (2014i).
22. Schmitt, U.: Proposing a Next Generation of Knowledge Management Systems for Creative Collaborations in Support of Individuals and Institutions. In Proceedings of the 6th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K), Oct 21-24, 2014, Rome, Italy. (2014h).
23. Schmitt, U., Butchart, B.: Making Personal Knowledge Management Part and Parcel of Higher Education Programme and Services Portfolios. In Journal of the World Universities Forum. Vol. 6(4) pp. 87-103, 1835-2030 (2014g)
24. Schmitt, U.: Overcoming the Seven Barriers to Innovating Personal Knowledge Management Systems. In Proceedings of the International Forum on Knowledge Asset Dynamics IFKAD, June 11-13, 2014, Matera, Italy. 978-88-96687-04-8 (2014f)
25. Schmitt, U.: How this Paper has been created by leveraging a Personal Knowledge Management System. In Proceedings of the 8th International Conference on Higher Education ICHE, March 16-18, 2014, Tel Aviv, Israel (2014d)
26. Schmitt, U.: The Role of Personal Knowledge Management Systems in making Citizens highly knowledgeable. In Proceedings of the 8th Int. Technology, Education and Development Conference INTED, March 10-12, 2014, Valencia, Spain. 978-84-616-8412-0 (2014c)
27. Schmitt, U.: Personal Knowledge Management Devices - the next Co-evolutionary Driver of Human Development?! In Proceedings of the Int. Conference on Education and Social Sciences INTCESS14, February 03-05, 2014, Istanbul, Turkey. 978-605-64453-0-9 (2014b)
28. Schmitt, U.: Knowcations - Conceptualizing a Personal Second Generation Knowledge Management System. In: Skulimowski, A. M. J. (ed). Looking into the Future of Creativity and Decision Support Systems. Proceedings of the 8th Conference on Knowledge, Information and Creativity Support Systems KICSS, Nov 07-09, 2013, Krakow, Poland. (2013g)
29. Schmitt, U.: Knowcations - The Quest for a Personal Knowledge Management Solution. In 12th International Conference on Knowledge Management and Knowledge Technologies i-Know, Sep 05-07 2012, Graz, Austria. Copyright 2012 ACM 978-1-4503-1242-4/12/09. Graz: ACM. <http://dl.acm.org/citation.cfm?id=2362469>. (2012)
30. Simon, H. A.: Designing Organizations for an Information-Rich World. In: M. Greenberger (Ed.) Computers, Communication & Public Interest. Baltimore, Johns Hopkins Press. (1971)

Computer creativity in games — how much knowledge is needed?

David C. Moffat and Paul Hanson

Department of Computing, Glasgow Caledonian University, UK.
email: D.C.Moffat@gcu.ac.uk

Abstract. In seeking to make computer systems that can be creative, there are several problems, including the definition of the concept of creativity, understanding how to design creative algorithms, encoding the domain knowledge needed for them, and evaluating the creative products of those algorithms.

We make a case that games are in some ways a good domain in which to research and develop creative algorithms, in particular the design of computer programs that can play games in ways that human players would consider to be creative.

As an example, we show the play of our program for a falling-blocks puzzle game, that uses simple search to play to a fairly good standard, and which a human player may well consider to be creative at times. In that case, we conclude that at least in some non-trivial domains, computational creativity might be achieved without the typically heavy requirement for a large or sophisticated knowledge-base, or even machine learning.

1 Creativity — concept and requirements

1.1 Definition of creativity

In the literature for creativity (and computational creativity), if a product or idea is generally said to be the result of a creative thought process then it must be both *novel*, and *valuable* (e.g. [1], [2]). What counts as valuable will of course depend on the domain; and the novelty of any idea should somehow be calculated by comparison with previous ideas in the domain. In many or most domains these are vague, informal and dependent on the intuition of domain experts. For example, this is the case for all the arts, clearly; and for practically all the so-called “creative industries.” Indeed, we always start with human intuition as our only guide to what is creative, until we can develop more formal measures.

Some authors include further conditions alongside novelty and value. Notably, Boden says that a creative idea should generally be in some sense *surprising* [3]. This could mean that the idea is not obvious even to domain experts, and that they would not easily have thought of it themselves. It could also mean that the creative idea might seem elegant, once it is heard. In the more striking cases, it might then even seem obvious; but that is only with the benefit of hindsight.

While there remains debate about the definition of creativity, we shall take it to mean those three conditions, of *novelty*, *value* and *surprise*, with *elegance* an optional fourth. As well as the matter of definition, in the field of computational creativity (CC),

there are the difficult matter of designing *algorithms* that could be capable of showing something like a human level of creativity; and the vexatious need to *evaluate* the creative products, even without any clear criteria that can be formulated.

1.2 Requirement for KA (knowledge acquisition)

As well as designing algorithms for CC, a prior problem is to provide them with any necessary knowledge to work with. When we aim to make programs that independently contribute to a field of human endeavour, they also need to have all or most of the knowledge that humans already discovered, in that field. This is a formidable requirement, in the general case requiring us to formalise a suitable knowledge representation scheme and then encode in it all relevant human knowledge. This venerable old problem of AI and expert systems is still with us today in most domains of real-world expertise. Even if we had arrived at a fully competent knowledge representation language for all domains, the task of encoding any single domain into it would be extremely intimidating. It is often referred to as the “knowledge acquisition (KA) bottleneck.”

Instead of encoding knowledge into a system (or as well as doing that), it is possible to make the system learn for itself, from its own experience in the domain. This approach of machine learning, however, is no nearer a general solution than the more direct approach of encoding. There being as yet no general solution to the KA-bottleneck, we have to choose our domain for creativity carefully, so that we can feasibly solve the KA issue for the chosen domain at least.

1.3 Requirement for evaluation

Another reason to choose the creative domain carefully is to enable us to evaluate the ideas or other creative products as reliably and as uncontroversially as possible. The evaluation should also be quick, and formalisable, especially in CC fields, to incorporate them into the creative generation algorithms.

In general, however, it is not easy to find a consensus on what ideas in any given domain are truly *novel*, *valuable*, and *surprising* (not to mention *elegant*). Nor can we usually agree on a suitable formalisation of these concepts, which are still an active research area.

Therefore we generally fall back on naive human judgement, possibly with several judges, at least to begin with, in order to conduct at least some kind of evaluation of creative programs. This is the approach taken in this paper.

2 Gameplay as a domain for computational creativity

Here we intend for game playing programs to be creative, just in the way they play their games. There are other ways to apply creative programs to the domain of games, such as by helping to design them (so called “procedural content generation”); but that is not the focus of this paper.

The attractions of gameplay as a testbed for CC (computational creativity) research are that it offers at least partial solutions to most of the above problems.

Firstly, games are typically easy to formalise briefly and clearly, because they are designed to have simple and exact rules. As games are often highly abstracted from their original worldly metaphor (e.g. real armies at war, in the case of chess), they do not require mountains of common sense or other domain knowledge to be encoded. Many games thus avoid or alleviate the *KA-bottleneck problem*.

Secondly, the playing of games has been a testbed application of AI since the beginning, so we have a depth of knowledge about how to write algorithms for this task. This largely solves the *algorithm problem*.

Thirdly, each game situation (or position) is potentially very similar to many that have occurred before in other games; and yet the smallest difference might have big consequences, so that there is rich novelty available, even against any database of games played in the past. Given such a database, it should be feasible to program a variety of possible estimates of *novelty*. Games typically award points or have other ways to quantify play performance, which can conveniently be used to assess *value*. Taken together, these two criteria of creativity are thus fairly measurable, which goes a long way to solving the *evaluation problem*.

Finally, the criterion of *surprise* or even *elegance* are less easy to formalise. We shall therefore use the game score as a basis for value, and rely on human intuitive judgement for the criteria of novelty, surprise or elegance.

3 PuyoPuyo as a game for CC research

The subject of our study is a simple arcade game called PuyoPuyo. First we implement a competent player for it. We then illustrate its performance that could be said to be creative gameplay, and end with some of its potential advantages as a testbed for CC research.

3.1 The game *Puyo-Puyo*

PuyoPuyo is played at the same time by two opposing players, where each player has a game board of 6x13 spaces. However, our program only plays a simpler, single-player version, so we here leave out the features of the game which are relevant when there are two players.

Pairs of different coloured puyo are dropped onto the player's board, which must then be positioned to avoid filling up the game board. As in *Tetris*, if and when the board fills up entirely to the top, the player loses [4]. Spaces on the board can be cleared by popping four or more puyo of the same colour. When this happens, the player's score increases. After a chain of Puyo is cleared in this way, any other puyo above them fall into their spaces (again, similar to *Tetris*); and if that makes a new chain (of four or more puyo of another colour), then it is also cleared away, and so on until no more puyo can be cleared. Then the next pair of puyo enter the top of the board for the player to drop.

In the example a blue and green puyo pair is dropped. The blue one completes a chain, which then disappears to let the higher puyos fall. A green chain is thereby completed, which also disappears. This causes a blue chain and then a red chain to go

as well, leaving only five puyos at the end of the move. Because four chains have gone in a single move, it has got a high score (much more than the four chains would have got separately). To build up the potential for such “cascades” of chains is thus the strategic aim of the game.



Fig. 1. Screenshots of *PuyoPuyo*. A green/blue pair falls, to make a chain of four blues, which clear away; then the five greens go; then the next four blues; and finally four reds.

While the aim may be clear, it is unclear how to achieve it. In other games like *Tetris*, one can postulate general rules or heuristics about good ways to play the game. Players soon learn to fit blocks in patterns and wait or hope for blocks to fall that exactly fit into the gaps, so that they can clear lines. But in *PuyoPuyo* there is not such an obvious strategic method. We thus fall back on brute search methods, which are not guided by expert heuristics. It is for these reasons that MC algorithms are interesting to investigate.

3.2 Monte Carlo Search (MCS) for *PuyoPuyo*

Monte Carlo (MC) simulations are predominantly used as a substitute to heuristic evaluation functions in game AI — the advantage being that a value in a search tree can be determined without the need for complex, time consuming expert knowledge based functions. In order to capitalize on the advantages of MC simulations and to improve the search capabilities of it, the concept of MC evolved into what is known as the Monte Carlo Tree Search (MCTS).

The MCTS is a form of best-first search which uses the results of multiple MC simulations to determine which path to take [5] This form of search was used by Bouzy [6] who combined a standard MC search, with a shallow and selective global tree search. His ComputerGo programs Indigo and Olga play the game pretty well [6].

In this study we investigate the possible value of a simpler MC algorithm, as follows, which shows how even a small and simple algorithm can play with some apparent creativity.

Because the game *PuyoPuyo* has a large branching factor it becomes computationally difficult to exhaustively search the tree more than a few layers deep. Each new

pair of *puyo* can be rotated four ways (similar to *Tetris*), and dropped in any of the six columns. Less the two positions where one of the *puyo* would be off the edge of the board, this means that the top level of the search tree branches 22 ways (i.e. $4 \times 6 - 2$). In this study the simple MCS algorithm only searches the first layer exhaustively, by visiting each of the 22 children in turn, in round-robin fashion. None of the nodes are favoured either; in contrast to what happens with many MCTS algorithms, where more attractive looking nodes are visited more often under the assumption that the best solutions may be found under those ones. These two simplifications of the more complex and advanced MCTS family of algorithms make this MCS algorithm much shorter to implement in code.

After some experimentation, we found that a depth-limit of only three was enough to give a good performance with real-time play on a modest PC (with Intel's Celeron two-core processor). This happens to be the most authentic depth-limit, as it corresponds to being aware of the next two blocks that will come after the current one. As a human player of PuyoPuyo can see the next two blocks queueing up at the side of the board (similar to *Tetris* again), this means the AI player is playing under the same conditions as the human player. More detail about the algorithm itself can be found in our earlier paper [7].

4 Performance of the PuyoPuyo player

The MCS algorithm described plays PuyoPuyo pretty well. An example of one of its more surprising and possibly creative moves, see Figure 2.

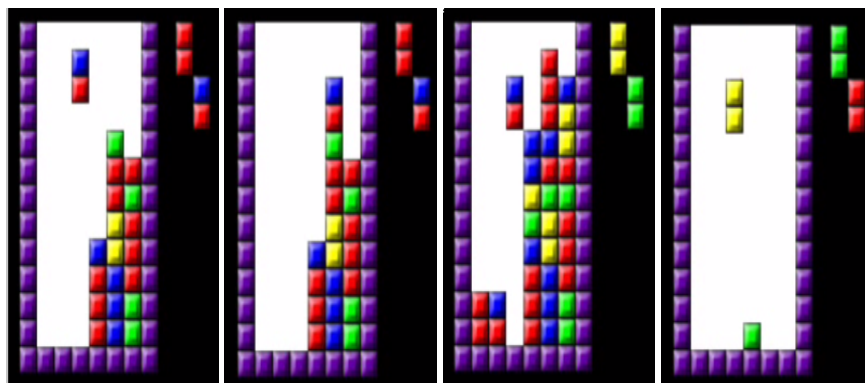


Fig. 2. Program playing *PuyoPuyo*. The program chooses a sequence of less obvious moves than seem to risk losing, but builds a good position on the left of the board, and finally exploits it to clear the whole board.

In the figure we see that a good move would be to drop the red/blue pair from where it is, as that would clear first the reds, and then more chains after that, to get a good score. But the program instead moves the pair to the side, and then moves the next few pairs to the sides as well; but putting more blues and reds to the left. This endangers

the play as the blocks nearly reach the top; but then the last pair shown (the blue/red) is dropped right into the whole, and nearly the entire board is cleared in one go.

This piece of play seems elegant, even to the two authors who are novice PuyoPuyo players; as well as apparently risky, which makes the final clearance more satisfying. As a high score is achieved by these moves, and a more obvious and still good move was foregone to make this play, it seems to be a good candidate for a creative play.

Although it may appear that the program is planning a long way ahead, it is in fact only searching to depth three. The reason that it appears to be conceiving of the game in terms of strategies like “building a strong position to the left of the board” and tactics like “aim to plug the gap with same coloured puyo” is probably due to our own human tendency to conceive the game in those terms. The program itself has no such complexities within it, however. It only has knowledge of the rules of the game, and it only looks three moves ahead. Anything else it appears to be doing must therefore be an artefact, or a result of the way we perceive the game.

5 Conclusion

Having identified some key ways in which KA problems are presenting some major road-blocks to CC research, we argued for the role of games (and in particular game-play) as a suitable domain.

We presented a simple game, *PuyoPuyo*, that can be understood by anybody who knows *Tetris*, but that does not yet have strategies formulated for how to play it well. Therefore a correspondingly simple monte-carlo search algorithm was encoded to play it, which had no special game strategy or other heuristic knowledge put into it. The algorithm did not learn either; and yet it was able to achieve a creditable performance. The example play of the game showed that the program is capable of some apparently novel, strong and even surprising play. In other words its play appears to be creative to some extent.

The main significance of the work as we see it, is that for the chosen game and algorithm, a modestly creative performance was achieved without any special knowledge, and without even any machine learning.

References

1. Csikszentmihalyi, M.: Creativity: flow and the Psychology of Discovery and Invention. HarperPerennial, New York (1997)
2. Boden, M.A.: The creative mind. (1998)
3. Boden, M.A.: Creativity and artificial intelligence. *Artificial Intelligence* **103** (1998) 347–356
4. Team, S.: Puyopuyo (2009)
5. Chaslot, G.M.J.B., Winands, M.H.M., van den Herik, H.J., Uiterwijk, J.W.H.M., Bouzy, B.: Progressive Strategies for Monte-Carlo Tree Search. *New Math. Nat. Comput.* **4**(3) (2008) 343–357
6. Bouzy, B.: Associating Shallow and Global Tree Search with Monte Carlo for 9x9 Go. In: *Proc. Int. Conf. Comput. and Games*, LNCS 3846, Ramat-Gan, Israel (2004) 67–80
7. Hanson, P., Moffat, D.C.: Monte carlo search for a real-time arcade game, puyo-puyo. In: *Proceedings of the Symposium on AI Games, AISB 2014 Convention*, at Goldsmiths, University of London (2014)

When Engineering and Design Students Collaborate: The Case of Website Development Workshop

Meira Levy¹, Yaron Shlomi¹, Yuval Etzioni²

¹Department of Industrial Engineering and Management,

²Department of Textile Design,
Shenkar College of Engineering and Design,
12 Anna Frank Street, Ramat-Gan 52526, Israel
{lmeira, yshlomi, yetzioni}@shenkar.ac.il

Abstract. Designers are often characterized by their creativity, whereas engineers are often characterized by their analytical thinking style. Both capabilities are vital in the new economic era, particularly for innovation. Thus, the collaboration between designers and engineers is crucial and should be fostered by academia and industry. Our preliminary research examined perceptions and interactions among 14 students in engineering, art and design programs during a website development workshop. The research indicates that students enjoyed the experience and found it educational and socially engaging. The data collected through questionnaires and observations were qualitatively analyzed, revealing that students learned each other's jargon, conducted discussions that addressed both technological and design concerns and considered various opinions leading to joint decision making. This study sheds light on the potential of multidisciplinary teams in development processes in general and in website development processes in particular.

Keywords: multidisciplinary collaboration, website development, qualitative research.

1 Introduction

In the new economic era multidisciplinary knowledge and capabilities are required for gaining competitive advantages [1,2]. Previous research addressed the need to enable effective collaboration among people from different disciplines, and found that successful multidisciplinary encounters depend on tailoring the selection of a theme, participants, and location to the encounter's particular objectives [3]. In particular, universities are preparing their students for a vital need of industry and society; i.e., to bridge the gap between producers and consumers of technology. To this end, universities are enriching their curricula with relevant themes from social sciences, humanities, cultural and management studies. Moreover, students from various disciplines need to understand the financial, business, environmental economic and social constraints in which businesses operate [4]. For this aim, multidisciplinary workshops are organized where participants are familiarized with one another's profession and learn to appreciate dissimilar viewpoints [5]. Although the literature suggests that multidis-

ciplinary programs are beneficial for broadening the students' perspectives, there are scant reports that describe the interdisciplinary educational perspectives and the interactions that occur among their participants.

In this paper we report on a multidisciplinary workshop aimed at website development. Workshop participants were students from engineering, design and art departments. During the workshop they gained technical knowledge related to a platform for web design, as well as knowledge related to digital culture and images. Understanding digital culture is vital to the development of a website as it encompasses the social implications of the knowledge and images presented in the site. Our preliminary research aimed at understanding the students' perceptions and interactions during the workshop for drawing inferences on how multidisciplinary teams might work collaboratively. Our paper is organized as follows: we present the importance and challenge of collaboration in the design process and elaborate on digital culture. We then present our empirical study and suggest some conclusions and directions for future research.

2 The Importance and Challenge of Collaborative Design

Designers are often characterized as creators, empathizers, pattern recognizers and meaning makers [1]. They are trained to deal with fuzzy and unpredictable situations and projects that require personal perspective, visual presentations, emotional involvement and almost no quantification [6]. Conversely, engineers are often characterized as more logical and linear thinkers with computer-related capabilities. They study science, accounting, finance, and related analytical subjects. There is a growing awareness of the need to create multifunctional teams that include designers and engineers in new product development process. The multifunctional teams are viewed as a strategic tool aimed at fostering innovation and gaining competitive advantage [2,7].

Working together, designers and engineers can change the way organizations are operating. Their collaboration will transform the organizations from the Information Age economy to the Conceptual Age economy characterized by inventive, empathic, and big picture capabilities [1]. Moreover, integrating the capabilities of sequential, logical and analytical thinking associated with engineers with the nonlinear, intuitive, and holistic thinking associated with designers [1] is essential to innovation. However, enabling the collaboration of engineers and designers is challenging. People from different professional backgrounds often have different value systems leading to misunderstanding and conflict [2]. Some of these differences can be traced to the training (i.e., the academic curriculum) that these professionals received (ibid).

3 Digital Culture

Digital culture exhibits the intersection of technology, knowledge, and culture in the digital age. The digital art object or image is a realization of human experience and hence considered an expression of human life in the field of aesthetics. Manovich [8] claims that the "New Media" refers to a new approach to culture representations in-

cluding an art object or an image as means of communication. The new media technology does not create new images but it brings new representations and manipulations capabilities with enormous presentation opportunities in private and social spaces (ibid). Specifically, manipulated digital images yield less detailed representations of reality than those obtained with the naked eye. Thus, researchers predict that watching digital images desensitizes our awareness to missing information [9]. At the same time new digital media tools and software can produce or edit digital images that may replace reality with virtual visibility, and can reduce human ability to fulfill human dialogue and relationships at the reality sphere [10].

As we turn to digital culture our visual experience is both expressing and challenging every aspect of our life. We are facing changes in our relationships with human body, objects and materials, our attitudes to written text and toward local and global time and space (ibid).

4 Empirical Study

4.1 Method and Settings

The main objective of our study was to explore the students' perceptions, expectations and interactions that occur during a design-engineering workshop. The students were third year students at Shenkar College of Engineering, Design and Art in Israel. Lecturers from the engineering, design and art departments formed teams for Shenkar's annual "Jam Week". Each team of lecturers prepared a four-day workshop in which students from the different departments worked together on projects guided by the workshop organizers. The first and last authors led the workshop discussed in this article, entitled "The Internet Sphere and Our Lives Sphere, Reflection and Alteration" with the aim of studying the digital culture and web technology principles and developing a website according to the students' interests. During the workshop students learnt, besides the web design platform, what digital culture is, and in particular what digital image means for people, and how artists develop it [11]. Understanding the digital culture concept was fundamental and allowed design discussions beyond technological issues.

The students learnt the Wix¹ Platform, which is a commercial free platform for creating websites. The four days of the workshop were mostly devoted to the teams' independent work on their projects, with few breaks for short lectures given by a representative from WIX and by the two workshop organizers. The WIX representative also served as a tutor during two days of the workshop, while the lecturers advised the teams as needed. The workshop participants were 14 students (7 from design and art and 7 from engineering) who were divided to four multidisciplinary teams.

The workshop started with a social game in which students reported on three subjects of interest to them. Then we used a string to create a physical network among students who reported on related topics. Finally, students were assigned to teams

¹ <http://www.wix.com/>

based on common interests, with the constraint that each team include at least one student from engineering and one from design or art.

Data were collected through open-ended questionnaires (see Appendix 1,2) and through observations of the teams during their work. Students completed one questionnaire at the beginning of the workshop, and a second one at the workshop's conclusion. The gathered data were qualitatively analyzed [12].

4.2 Analysis and Findings

The analysis of the pre-workshop questionnaire revealed that all of the students were enthusiastic to learn a new web building platform that can serve them in the future. Eight students used one of the words "innovation"/"creativity"/"thinking out of the box"/"inspiring" in their list of associations. The words "expectations"/"excitement"/"experience"/"curiosity"/"enjoyment"/"pleasant" were found in 11 questionnaires. Ten students mentioned that they want to get to know other workshop participants, especially those from a different discipline, for gaining other perspectives.

These findings suggest that students are eager to meet colleagues from various disciplines, for extending their perceptions and learning experience.

The data that were gathered during the workshop focused on conversations among the teams' members. Qualitative analysis of these exchanges revealed that students, at the beginning of the workshop, tended to comment on issues that were more related to their discipline. For example, a design student talked about the site's logo, and an engineering student commented "color is your field", while pointing at the design student. Engineering students focused on the functionality of the site or handling the work: "we need to reduce clicks"/"lets upload the content, we won't be able to finish", while design students focused on the design or the general site theme: "we need to add affects, like something is burning"/"the site should be more clean"/"The main page is the most important one, we must invest in the design".

As the workshop continued, we heard engineering students adopt the design role and terminology: "the site gives the feeling of wholeness"/"What is your design concept?"/ "we need a digital image that expresses the hard work of the artist therefore we need a close-up picture". At the same time, design students adopted the more engineering perspectives: "we need to present a work process"/"we need to add a button to stop the music".

These findings suggest that although the workshop lasted only four days students got involved with issues that initially belonged to their colleagues from the counter discipline. The general atmosphere among the team members was that while the engineering or design student was considered the professional in her discipline, there was mutual respect to hear opinions from everyone in the team.

The analysis of the post-workshop questionnaires revealed that the students found the workshop pleasant, instructive, and socially beneficial. Twelve students mentioned "collaboration"/"team work" and "mutual work" as terms describing their thoughts at the end of the workshop. They mentioned the collaborative atmosphere and decision making processes where each student could express thoughts and preferences.

5 Conclusion and Future Research

We report on a multidisciplinary workshop in which students from the engineering, design and art departments collaboratively developed websites according to their interests while being engaged in technological and cultural enrichment activities. Our view is that the interactions among the workshop participants exemplify collaboration among engineering and design students. Moreover, the study revealed that students were enthused to participate in such experience and found it educational and enjoyable. The interactions enabled mutual learning of each discipline's terminology while respecting various opinions when arriving at joint decisions regarding the characteristics of their websites. The multidisciplinary discussions, encompassing both technological and design perspectives, enriched the class and created a creative atmosphere that fostered creation of the websites.

While this case study was successful, the generalization of this preliminary research may be limited. However, it sheds light on the potential of multidisciplinary teams in development processes in general and in web design process in particular. Engineers and designers rely on a variety of design terminologies to define and model their problems and alternative solutions. Further, we assume that these professionals may not use the same design modeling tools (e.g., UML, wireframe). More generally, the two disciplines might be unfamiliar with each other's goals, perspectives, and tools. This is important because each of the disciplines relies on these elements to present its design solution (e.g., website).

Future research will continue to explore differences between the goals, perspectives and modeling tools of the two disciplines. In the long term, this exploration will guide us in creating a modeling tool that is accessible and beneficial to both disciplines. This tool will facilitate collaboration within multidisciplinary teams.

6 References

1. Pink, D. H.: *A Whole New Mind: Why Right-Brainers Will Rule the Future*. New York, NY: Riverhead Books (2005)
2. Stamm, V. B.: *Managing Innovation, Design and Creativity*, Second edition, John Wiley & Sons Ltd. (2008)
3. Bridle, H., Vrieling, A., Cardillo, M.: Preparing for an interdisciplinary future: A perspective from early-career researchers. *Futures*, 53, 22-32 (2013)
4. Parashar, K.M., & Parashar, R.: Innovations and Curriculum Development for Engineering Education and Research in India. *Procedia – Social and Behavioral Sciences*, 56, 685-690 (2012)
5. Taneli, Y. Yurtkuran, S. Kirli, G.: A Multidisciplinary Design Exercise: Myndos Excavation Site. *Procedia – Social and Behavioral Sciences*, 106, 120-129. (2013)
6. Managers and Designers: Two Tribes at War? In Oakley, M. (ed.), *Design Management: A Handbook of Issues and Methods*. Oxford: Blackwell. (1990)
7. Rigby, D.K., Gruver, K., & Allen, J. : *Innovation in Turbulent Times*. Harvard Business Review. (2009). Retrieved March 6, 2014. <http://hbr.org/2009/06/innovation-in-turbulent-times/ar/1>
8. Manovich, L.: *Avant-garde as Software*. <http://www.manovich.net/> (1999)

9. Rushkoff, D.: Program Or Be Programmed: Ten Commands for a Digital Age. Berkeley: Soft Skull Press.
http://share.pdfonline.com/6d37ec8261714ac18920d77697ad83b3/Program_or_Be_Programmed.htm
10. Verilio, P.: Speed and Information: Cyberspace Alarm
<http://www.ctheory.net/articles.aspx?id=72>
11. Sagmeister, S.. Everything I do always comes back to me
<http://www.sagmeisterwalsh.com/>
12. Strauss, A., Corbin, J.: Basics of qualitative research grounded theory procedures and techniques, Sage Publications, Inc (1990)

Appendix 1: Workshop Entry Questionnaire

Welcome to the workshop “The Internet Sphere and Our Lives Sphere, Reflection and Alteration”. The workshop staff is interested in documenting your attitude to the workshop and to topics related to it. We would like to ask you to answer some questions at the beginning of the workshop. Thank you for your cooperation.

1. Name (optional): _____
2. Email: _____
3. Academic program: _____
4. Why did you choose this workshop?
5. What are your expectations and goals from the workshop?
6. List 10 words describing associations, perceptions, emotions, etc. that you can think of at the beginning of the workshop.
7. Describe your knowledge and experience as it relates to the workshop agenda as presented so far.
8. During the workshop you will meet and work with students from different academic programs. What are your expectations and goals from meeting and working together?
9. Do you have any professional work experience? If so, please elaborate.

Appendix 2: Workshop Summary Questionnaire

Towards the end of the workshop, you are requested to give the workshop organizers a feedback about the workshop activities and personal experiences you went through as individual and as a member of a joint team of engineering and design. Thank you for your cooperation.

1. Name (optional): _____
2. Email: _____
3. Academic program: _____
4. Did the activities in the workshop match your expectations and goals from the workshop?
5. Describe your experience working with colleagues from other academic programs
6. List 10 words describing associations, perceptions, emotions, etc. relating to the end of the workshop
7. What new subjects have you learned in the workshop?
8. Describe how decisions were made in the team, and whether there was a priority in decision-making related to the academic program (design/engineering). (in issues of choosing the site idea, site design, site content, technological issues, work documentation, team handling issues)
9. Do you have any recommendations for a similar workshop in the future?

Knowledge-Managing Organizational Change Effectors, Affectors and Modulators in an Oil and Gas Sector Environment

Anthony Ayoola

Dean, School of Business & Management, American University of Ras Al Khaimah,
Ras Al Khaimah, United Arab Emirates (Anthony.Ayoola@aurak.ae)

Abstract. Knowledge Management techniques (KMT) have been used effectively in a number of business domains, for dealing with complex management issues. A key area of interest is the use of KMT for managing organizational change in an effective manner. Developmental, transformational and transitional change types are generally influenced by a range of inter – dependent factors. These factors can be broadly categorized as change effectors, affectors and modulators.

This paper investigates the use of Knowledge Management techniques for modulating salient change effectors, affectors and modulators influential in the oil and gas sector, within a Middle-Eastern context.

Research findings from a recent multi-company case study of oil and gas sector organizations in the Gulf region are presented based on a conceptual framework predicated on crucial change effectors, affectors and modulators. Knowledge Management modulated change factors investigated in the study include worker empowerment, autonomy, participation, communication, culture, normative responsibility, training/re-skilling and reward systems.

Keywords: Knowledge Management, Interaction Circles, Knowledge Diffusion, Knowledge Externalization, Sustainable Change

1 Introduction and Background

Companies in the oil and gas sector have undergone significant change in recent years, driven by technological growth, business competition, market-place issues and public expectations. Technology is continually advancing and non-progressive

companies find their organizational set-up outdated and inefficient. To counter this, oil and gas sector firms embark on frequent bouts of change with attendant developmental, and management issues.

This paper provides a systematic framework for characterizing developmental, transformational and transitional change types and determining how these are influenced by a range of HRM – dependent factors. The causal HRM-related factors are systematically categorized as change effectors, affectors and modulators. This paper will establish which effectors, affectors and modulators exert the most significant levels of influence in the oil and gas sector; within a Middle Eastern context.

Studies by Stace and Dunphy [1] map the change driven formats to various change types. So, for example, they recommend the mapping of unilaterally driven change to the transformational change type, primarily for reasons of convenience and the much shorter timescales.

For transitional or developmental transitions, multilaterally driven (participative) change is recommended, since there is greater scope for worker involvement, and longer timescales. Knowledge-Managed HRM (KM-HRM) spans several people-management activities and is applied at multiple levels within the organization. When properly practiced, KM-HRM provides a framework for holistic people management, serving to select, train, motivate and performance manage the company's workforce. KM-HRM has significant strategic and operational implications. The work we have carried out systematically clarifies specific KM-HRM practices as being effectors, affectors or modulators.

2 Knowledge-Managing change

Standard change management schemes [2] are typically linear, with feedback loops often used to regulate and fine-tune the process. It is possible to integrate Knowledge-Management formalism within standard change management. We term this approach KMCM (Knowledge-Managed Change Management). KMCM manages tacit and explicit change-related knowledge sources within the organization. KMCM is predicated on key Knowledge Management (KM) phases which are in turn based on Socialization, Externalization, Combination (Group-learning and Dissemination) and Internalization KM activities. Approaches of this nature are termed SECI-based Systemic Models [3]–[7]. We have used SECI-based models to formulate a more relevant and robust framework to underpin the proposed KMCM technique. The derived scheme is shown in Fig. 1 –

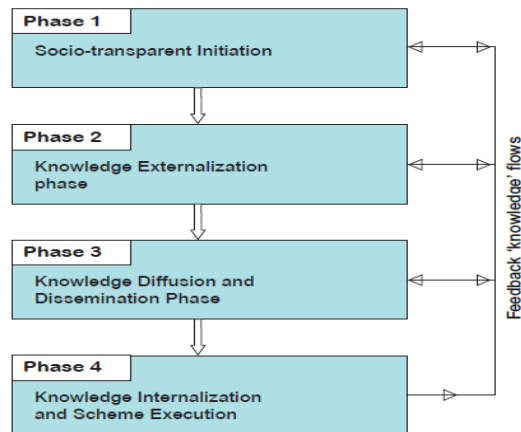


Fig. 1. Multi-phased KMCM workflow

3 HRM Effectors, Affectors and Modulators

The KMCM work carried out systematically clarifies specific HRM practices as being effectors, affectors or modulators. Effector HRM practices seek to implement people management policies and processes. This includes recruitment activities, disciplinary processes, job design and performance appraisal. Effector HRM is driven by targets and performance levels. Affector HRM practices are concerned with building commitment, loyalty and trust within the workforce. Affectors seek to enhance the psychological contract between the company and its employees. Affector HRM provides the caring and supportive face to HRM. It also strives to elicit participation where this is in alignment with the organizational ethos or culture. Affector HRM roles include providing mechanisms and policies for benefits, dispute resolution, grievance processes etc. Affector HRM seeks to minimize worker discontent, and match employee expectations to what the company is able to offer.

Modulator HRM practices provide the necessary control mechanisms for people management. This could include organizing training for re-skilling or up-skilling, and the restructuring of worker reward systems and procedures. Modulator HRM is able to modulate or elevate states induced by effectors or affectors e.g. changing skills levels of newly recruitment employees through training or improving levels of worker motivation through well designed reward systems. In other words, Modulators can sometimes be viewed as secondary transitioning agents, whilst Effector and Affector HRM practices are primary influences. Having specified concepts pertaining to KM, change management and HRM, we can proceed to integrate the three strands within relevant parts of the KMCM workflow.

4 Methodology of the Study

A multi-company comparative case study of companies in the Middle Eastern Oil and Gas sector was carried out over several months [6]. The study was based on a critical realism research epistemology. A mixed inductive- deductive research approach was used. Data collection utilized a mix of primary and secondary data collection instruments. The primary data was collected using self-completed questionnaires (SCQ) and semi-structured interviews (SSI). Secondary data was used for additional triangulation. The time-frame was essentially longitudinal, with cross-sectional elements. The selected companies had been involved in significant change processes, and stratified samples of HR managers, line managers, training managers and regular employees were used. The work identifies the key HRM effectors, affectors and modulators that influence change management within the organizations, as a prelude to adopting and implementing a formal KMCM scheme for managing future change.

5 Study Findings

Impact of Communications & Training (Effector & Modulator HRM domains)

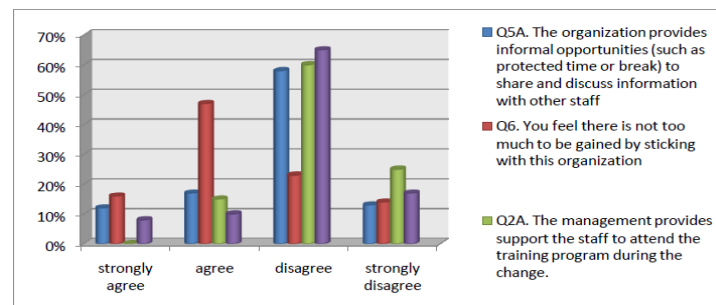


Fig. 2. Impact of Communications & Training

Impact of Level of Commitment to the Company (Affector HRM domain)

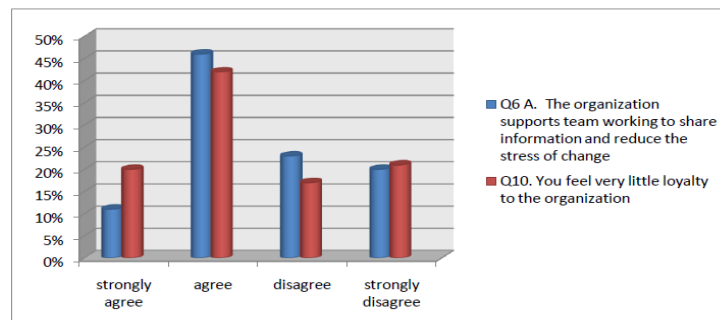


Fig. 3. Impact of Level of Commitment to the Company

Impact of Determinants of Resistance to Change (Affector HRM domain)

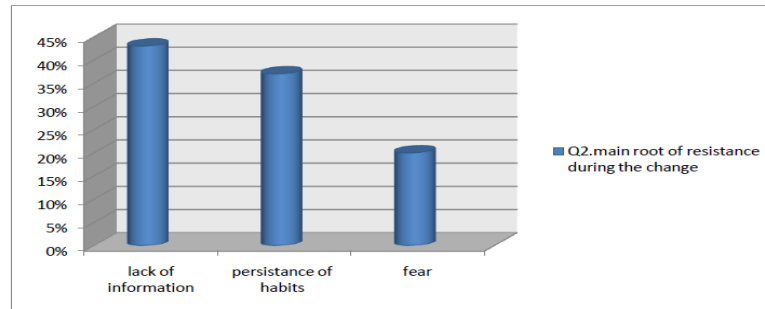


Fig. 4. Impact of Determinants of Resistance to Change

Impact of Worker Empowerment & Participation (Affector HRM domain)

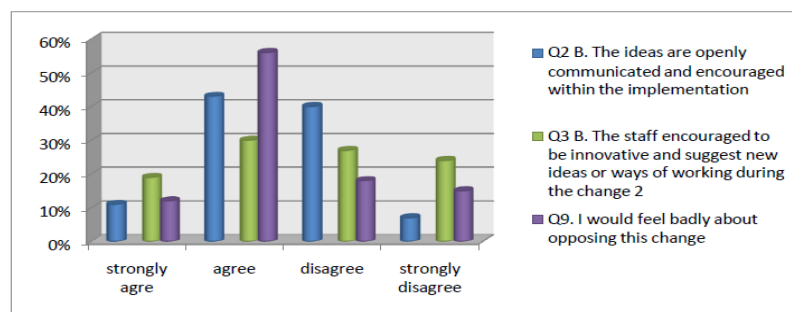


Fig. 5. Impact of Worker Empowerment & Participation

Impact of Level of Consultation about the Change (Affector HRM domain)

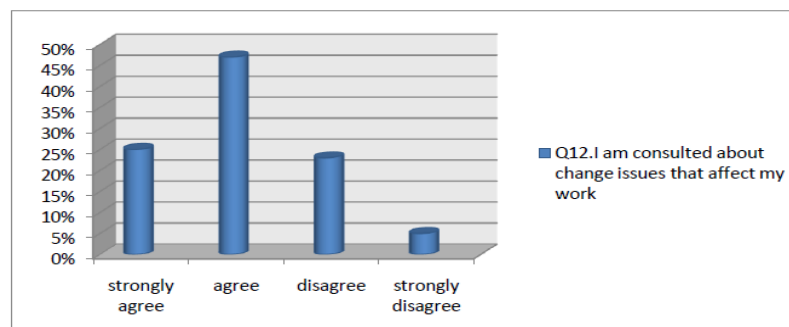


Fig. 6. Impact of Level of Consultation about the Change

Impact of Rewards & Performance Appraisal (Affecter, Effector & Modulator domains)

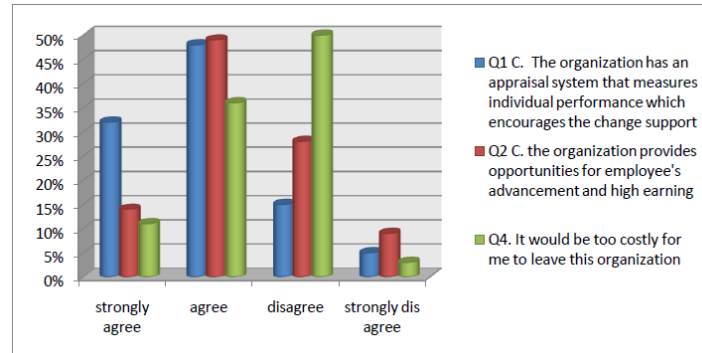


Fig. 7. Impact of Employee Rewards & Performance Appraisal

6 Comments and Conclusions

Varying levels of Effector, Affecter and Modulator HRM practices were used by the companies during the period of change, with the qualitative responses and perceptions shown in the charts. In a majority of cases, the changes were judged to be difficult, and the overall conclusion was that more judicious use of the prevailing HRM Effector, Affecter and Modulator practices, within a KMCM scheme, would have resulted in more efficient and effective change processes.

References

- [1] Dunphy, D. and Stace, D. (1993) *The Strategic Management of Corporate Change*, Human relations, Vol 46 (8) pp.905-20. D. Kelleher and S. Levene, *Knowledge Management: A Guide to Good Practice*, British Standards Institute: London, 2001.
- [2] SA. Carlsson, "Knowledge managing and knowledge management systems in inter-organisational networks," *Knowledge and Process Management*, vol. 10, no. 3, pp. 194-206, 2003.
- [3] T. Knight and T. Howes, *Knowledge Management: A Blueprint for Delivery*, Butterworth Heinemann: Oxford, 2003.
- [4] PJ. Lane and M. Lubatkin, "Relative absorptive capacity and inter-organizational learning," *Strategic Management Journal*, vol. 19, no. 5, pp. 461-477, 1998.
- [5] S. Newell, M. Robertson, H. Scarbrough and J. Swan, *Managing Knowledge Work and Innovation*. London: Palgrave Macmillan, 2009.
- [6] H. Rezaeian (2010), MBA Dissertation, University of Bolton
- [7] H. Scarbrough, *Knowledge Management: A Literature Review*, Issues in People Management Series, Institute of Personnel and Development: London, 1999.

From Computational Creativity Metrics to the Principal Components of Human Creativity

Pythagoras Karampiperis, Antonis Koukourikos, George Panagopoulos

Computational Systems & Human Mind Research Unit
Institute of Informatics & Telecommunications
National Center for Scientific Research “Demokritos”

Abstract. Within the field of Computational Creativity, significant effort has been devoted towards identifying variegating aspects of the creative process and constructing appropriate metrics for determining the degree that an artefact exhibits creativity with respect to these aspects. However, in the effort to determine if an artefact is creative by human standards, it is also important to examine the perception of creativity by humans and to which extent this perception can be formalized and applied on the evaluation of creative works. In this paper, we investigate how the human perception for creativity exhibited in text artefacts can be correlated by the usage of appropriate formulations of computational creativity metrics. To this end, we propose a model for transitioning from traditional metrics to a space that adheres to the principal components of human creativity and reflects the way that human approach the assessment of the creative process.

Keywords: Computational Creativity; Creativity Measurement; Creativity in Texts; Cognitive Models.

1 Introduction

Human creativity is a multifaceted, vague concept, combining undisclosed or paradoxical characteristics. As a general notion, creativity adheres to the ability to move beyond traditional and established patterns and associations, by transforming them to new ideas and concepts or using them in innovative, unprecedented contexts and settings [1]. In general, the creativity of a person can be divided qualitatively by taking into account its origin in psychometric or cognitive aspects of their thinking process [2]. On the other hand, machines can mimic human creativity, or provide the necessary stimuli for encouraging and promoting the production of creative ideas and artefacts, it is not straightforward to assess the exhibited creativity by using automated techniques. Rather, most efforts have been focused on analyzing creativity on different aspects and producing different metrics, based on the nature of the examined artefacts.

In the present work, we propose text-based metrics for the core aspects of creativity as the latter are determined in the relevant literature and examine their conformance with the human perception of what constitutes a creative artefact. We proceed to identify the

deviations between these two perspectives (computational metrics and human judgment) and propose a model for transforming the automatic measures to a space that more accurately reflects the human opinion.

2 Understanding the Human Perception of Creativity

2.1 Metrics for Computational Creativity over Textual Content

The association of creativity metrics with quantifiable results derived from the textual data is the critical step for automating the evaluation process. The formalization of creativity metrics for textual content is a complex task, and the related work is focuses on very specific characteristics of the examined content in order to model creativity.

Zhu et al. [1] propose a machine-learning approach based on features derived from computer science and psychology perspectives. Other works focus on concrete linguistic and morphological characteristics of the text, e.g. analogies [3].

In our previous work [4], we presented a formulization of a set of Computational Creativity Metrics for Novelty, Surprise, Rarity, and Recreational Effort, over textual artefacts. In this paper, we extend this work in order to correlate these Computational Creativity Metrics with the perception of creativity by humans.

2.2 Correlation of Computational Creativity Metrics with the Human Perception of Creativity

In order to assess the adherence of the proposed metric formulization with the human perception for creativity, we organized and conducted an experimental session based on storytelling activities. For the execution of the experiment, we employed forty (40) human participants, split in ten (10) teams of four (4) members each. All teams were asked to construct a story, on a specified premise, the survival of a village's habitants under a ravaging snow storm. The stories were created incrementally, with twenty (20) fragments produced for each story.

Following the completion of the stories, the teams were organized in two groups, each consisting of five teams. Without any interaction between the groups, each team was called to rate the stories of the remaining four teams belonging to their group, using a rank-based 4-star scale (i.e. the best story received 4 stars, the second-best story received 3 stars etc.). In this way, we obtained a ranked list of the five stories in each group. The goal of our experiment was to determine if, using the ranked lists of one of the test groups and a formalized representation of the computational creativity metrics, we can identify their correlation and examine if the distribution of values for the metrics follow the pattern of human judgment. To this end, we define a constrained optimization problem over functions of the aforementioned metrics, which is described below.

2.3 Extracting a Model for the Human Perception of Creativity.

Each artefact (story) S_n is characterized (via the execution of the computational creativity metrics presented in the previous section) by a set of 4 independent properties $g^{S_n} = (g_1^{S_n}, g_2^{S_n}, g_3^{S_n}, g_4^{S_n})$ where g_1 stands for “Novelty”, g_2 for “Surprise”, g_3 for “Rarity” and g_4 for “Recreational Effort”. We define as partial creativity function (PCF) related to artefact property g_k a function that indicates how important is a specific value of the property g_k when calculating the creativity of an artefact S_n . This function is defined by the following formula:

$$PCF_{g_k}(g_k^{S_n}) = w_{g_k} * \left(\frac{c_{g_k} * (1 - d_{g_k})}{e^{(a_{g_k} * g_k^{S_n} + b_{g_k})^2}} + \frac{d_{g_k}}{2} \right), \quad (1)$$

where $g_k^{S_n} \in [0, 2]$ is the value of property g_k for the artefact S_n , and $0 \leq a_{g_k} \leq 5$, $-4 \leq b_{g_k} \leq 4$, $0 \leq c_{g_k} \leq 1$, $0 \leq d_{g_k} \leq 2$ are parameters that define the form of the partial creativity function, whereas $0 \leq w_{g_k} \leq 1$ represents the weight of property g_k in the calculation of the overall creativity. The calculation of the above parameters for all g_k properties lead to the calculation of the complete creativity function (CCF), as the aggregation of the partial creativity functions, as follows: $CCF(g^{S_n}) = \frac{1}{4} * \sum_{k=1}^4 PCF_{g_k}(g_k^{S_n})$

If CCF_{S_1} is the complete creativity of an artefact S_1 and CCF_{S_2} is the complete creativity of an artefact S_2 , then the following properties generally hold for the complete creativity function:

$$CCF_{S_1} > CCF_{S_2} \Leftrightarrow (S_1)P(S_2),$$

$$CCF_{S_1} = CCF_{S_2} \Leftrightarrow (S_1)I(S_2),$$

where P is a strict preference relation and I is an indifference relation, as perceived by humans when evaluating the creativity of these artefacts.

Given a preference ranking of a reference set of artefacts, we define the creativity differences $\Delta = (\Delta_1, \Delta_2, \dots, \Delta_{q-1})$, where q is the number of artefacts in the reference set and $\Delta_i = CCF_{S_i} - CCF_{S_{i+1}} \geq 0$ is the creativity difference between two subsequent artefacts in the ranked set.

We then define an error parameter E for each creativity difference:

$$\Delta_i = CCF_{S_i} - CCF_{S_{i+1}} + E_i \geq 0.$$

We can then solve the following constrained optimisation problem:

$$\text{Minimise } \sum_{i=1}^{q-1} (E_i)^2 \text{ s.t. } \begin{cases} \Delta_i \geq 0, \text{ if } (S_i)P(S_{i+1}) \\ \Delta_i = 0, \text{ if } (S_i)I(S_{i+1}) \end{cases}$$

This optimisation problem leads to the calculation of the partial creativity function parameters (a , b , c , d and w) for each property g_k .

Regarding the impact of the various metrics in the computation of the overall creativity, we observed that Novelty is generally considered a particularly positive attribute creativity-wise for the stories, its partial creativity (PC) increasing as its value increases. In contrast, the remaining metrics reached their maximum partial creativity at a certain

value, after which their partial creativity started to decrease, indicating that e.g. recreational effort greater than a certain point is not perceived as a direct indication of creativity. Hence, the obtained results indicate that, while the proposed computational creativity metrics are correlated with the perception of humans for creativity, this correlation is not direct for all metrics. The following section discusses on the implications of these observations and details our approach for using the proposed metrics towards building a dimensional plane that more accurately reflects the human perspective for creativity.

3 Transferring Computational Creativity Metrics to the Human Perspective

As stated, the four computational creativity metrics discussed, provide a good estimation for the respective creativity aspects exhibited by textual artefacts. However, in the process of using such metrics to approach the human notion for creativity, the derived results indicated some deviations between this formalization and the way humans think. In broad terms, we observe the following two characteristics that should be taken into account when trying to model the human perspective:

- Humans prefer to think *monotonically*, perceiving the value of a metric / dimension as analogous to the “quality” of the examined artefact in that dimension;
- Humans prefer to think *orthogonally*, perceiving each of the features as a dimension independent from the rest.

The first step towards identifying the adherence of our metrics with the human perspective is to examine the orthogonality of the proposed metrics formulation. To this end, we ran an experiment for calculating the four basic computational creativity metrics on two datasets derived from distinct and distant domains, and determined whether the four metrics are orthogonal.

The first dataset comprised transcriptions of European Parliament Proceedings [5]. Given the described formulation of computational creativity metrics, we consider as a “story” the proceedings of a distinct Parliament session and as a fragment the speech of an individual MP within the examined session. The second dataset was derived from a literary work, *Stories from Northern Myths*, by E.K. Baker, available via the Project Gutenberg collection. In this case, the story is a book chapter and the story fragment is a paragraph within the chapter. In total, we examined 50 distinct parliament sessions from the Europarl dataset and 40 chapters from the storybook.

Based on the obtained results, we calculated the correlation between the four computational creativity metrics. Table 1 (left) provides the correlation values between the four metrics. It is evident that the computational creativity metrics by themselves are not orthogonal. Hence, in order to better approximate the human perception for creativity, we propose the following abstraction for modelling the examined aspects of creativity to a space more closely resembling human thinking:

	N.	S.	R.	R.E.	Type	I.	At.	
N.	1.000	0.134	0.123	-0.407	Formal Verbal Transcripts	1.0000	2.986E-07	I.
S.	0.134	1.000	0.265	-0.432		2.986E-07	1.0000	At.
R.	0.123	0.265	1.000	-0.335				
R.E.	-0.407	-0.432	-0.335	1.000				
N.	1.000	-0.642	0.104	-0.108	Literary Work	1.0000	1.436E-07	I.
S.	-0.642	1.000	0.074	-0.025		1.436E-07	1.0000	At.
R.	0.104	0.074	1.000	-0.039				
R.E.	-0.108	-0.025	-0.039	1.00000				

Table 1. Correlation of Computational Metrics (left) and Creativity Dimensions (right)

-*Innovation*, that is, the tendency to produce ideas and artefacts that are disassociated with the other elements on a given context.

-*Atypicality*, that is, the tendency to deviate from the norm without actually breaking through.

Innovation, by its nature, is a perspective which closely associated with the Novelty computational metric. To this end, the magnitude of Innovation for a textual artefact is equal with its value for the novelty metric. On the other hand, we consider Atypicality as a combination of the Surprise, Rarity and Recreational Effort metrics, each bearing a different weight towards determining Atypicality. These two axes also provide a rough conceptualization of the two major qualitative aspects of creative work: whether the said work is *visionary*, i.e. it provides a groundbreaking approach on a given field; and whether it is *constructive*, i.e. it uses in a novel way established techniques and ideas in order to produce a high-quality artefact.

As indicated by the experiment described in Section 2.2, Innovation has an analogous and close to monotonic association with the human judgment for creativity. Therefore, and in order to satisfy our second requirement (orthogonality), we consider Innovation as the strictly defined dimension of our space and seek for the formulation of Atypicality that results to a dimension orthogonal to Innovation.

More specifically, let Atypicality of a text t be the normalized weighted sum of its Surprise, Rarity, and Recreational Effort:

$$A(t) = \frac{w_s \text{Sur}(t) + w_r \text{Rar}(t) + w_e \text{Eff}(t)}{w_s + w_r + w_e}, \text{ with the weights } w_s, w_r, w_e \in [-1, 1].$$

We aim to find the weight values that constitute Atypicality orthogonal to Innovation, i.e. those weight values for which $\text{Correl}(\text{Innovation}, \text{Atypicality}) = 0$. We thus define the following optimization problem:

Minimise $\sum_{i=1}^n (\text{Correl}(\text{Innovation}_i, \text{Atypicality}_i))^2$ s.t. $w_s, w_r, w_e \in [-1, 1]$, where n is the number of the combined datasets.

Although the search space of the optimization problem above is highly non-linear, solving this problem is straightforward. The optimum weight values in our case are:

$$(w_s, w_r, w_e) = (0.13951, 0.10154, 0.06905)$$

Table 1 (right) presents the correlations between innovation and atypicality in the two datasets for the found optimum weight values. The resulting model defines two orthogonal axes, Innovation and Atypicality, which define the space for measuring and characterizing the observed creativity, as an Euclidean vector, the length of which indicates the quantitative aspect of the creativity exhibited by the artefact, while its direction indicates the tendency for either Innovation (visionary creativity) or Atypicality (constructive creativity). Using this model for the evaluation of the storyset of the initial experiment, and taking into account the vector length for each story, we obtained the same ranking as the one produced by the human evaluation. This is a strong indication that the proposed model accurately reflects human judgement, while also pertaining to core principles of the human perception of creativity.

4 Conclusions

Understanding the elements of the Creative Process, is a challenging research issue, which combines research results from several research fields, like neuroscience, psychology and philosophy. Within the field of Computational Creativity, there have been significant results for the identification of various aspects of the creative process and the construction of appropriate metrics for these aspects. However, in the effort to determine if an artefact is creative by human standards, it is also important to examine the perception of creativity by humans and to which extend this perception can be formalized and applied on the evaluation of creative works.

The work described in the present manuscript showcases our findings towards transitioning from computational creativity metrics associating specific attributes of text artefacts with creativity aspects to a creativity calculation model that better reflects the human perception of creativity.

5 REFERENCES

1. Zhu, X., Xu, Z., Khot, T.: How creative is your writing? A linguistic creativity measure from computer science and cognitive psychology perspectives. In: Workshop on Computational Approaches to Linguistic Creativity, ACL, 87-93 (2009)
2. Boden, M.: The creative mind, Routledge, London, 2nd Edition (2004)
3. Chiru, C-G.: Creativity detection in texts. In: The 8th International Conference on Internet and Web Applications and Services (ICIW2013), Rome, Italy, 174-180 (2013)
4. P. Karampiperis, A. Koukourikos, E. Koliopoulou, "Towards Machines for Measuring Creativity: The Use of Computational Tools in Storytelling Activities", in Proc of the 14th IEEE International Conference on Advanced Learning Technologies (ICALT2014) Advanced Technologies for Supporting Open Access to Formal and Informal Learning, Athens, Greece, July 7-9, 2014
5. Koehn, P.: Europarl: A Parallel Corpus for Statistical Machine Translation. MT Summit (2005)

The Approach to Extension of the CLAVIRE Cloud Platform for the Researchers' Collaboration

A.V. Dukhanov¹, E.V. Bolgova¹, A.A. Bezgodov¹, A.V. Boukhanovsky¹

¹ITMO University, Saint Petersburg, Russian Federation

dukhanov@niuimo.ru, {katerina.bolgova, demiurghg}@gmail.com,
avb_mail@mail.ru

Abstract. This paper describes an approach to extend the CLAVIRE platform for sharing traditional scientific documents, data sources, executable services and 3D-visualization means for collaborations and peer review. This approach includes the development of an intellectual editor for scientific package integration into CLAVIRE; C++ based graphical library and composite application with interaction between package and visualization tools. This means facilitating the process of scientific results sharing and presentation in a vivid and interactive manner. The considered approach was developed within the principles of research object ontologies and with the use of the workflow-centric research object approach and method of interactive workflow. The examination of the approach was performed by evolving young scientists from cities located across Russia.

Keywords: scientific collaboration, scientific result sharing, scientific package, cloud service, CLAVIRE cloud platform, research object

1 Introduction

Modern scientific society is characterized by many changes in methods, approach, and technologies. Modern researchers should respond quickly to these changes and offer rapidly novel scientific results. Contemporary scientific society should not only be provided with traditional scientific databases complemented by 3D pictures, data sources and social networks like ResearchGate, but they should also be provided with platforms, with easy-to-master interfaces, for sharing the executable scientific results (which require original computational resources) in the form of cloud service. They should also be provided with instruments for easy development of visualization tools, which support desktop graphical adapters, top systems of 3D-Stereo visualization.

Contemporary researcher have a wide choice of means to present his results in interactive and vivid manner. However, part of them has limited functionality and another part require skills and abilities in programming area, which may distract scientist from the main study. We have following questions under consideration.

Q1. How to present (share) scientist results in vivid and interactive manner for collaboration and peer review with the use of cloud computing technologies and modern concepts including Research Object?

Q2. How to simplify scientists' efforts (including programming costs reduction) to share his results in manner mentioned in the first question?

Q3. How to create a scientific collaboration environment in the trend of modern technological achievements in cloud computing and results presentation including 3D Visualization?

This paper offers approach to extend the CLAVIRE platform for sharing traditional scientific documents, data sources, executable services and 3D-visualization means for collaborations and peer review. This approach includes development intellectual editor for scientific package integration into CLAVIRE; C++ based graphical library and composite application with interaction between package and visualization tool.

2 Related Works

A modern researcher (or researcher's team) has a wide choice of abilities to present (share) his scientific results. He can disseminate results in the traditional form of a scientific paper in journals, or conference proceedings to use special instruments, plug-ins and other computer services for sharing data arrays, visualizations, scientific workflows (eScience applications), multimedia content and other digital materials (Q1).

2.1 Collaboration Approaches Based on Scientific Papers

Today scientists can collaborate with each other on scientific papers with use of various tools, platforms and social networks, e.g. Google Docs for Researchers¹, Mendeley², Research Gate³ and Life Science Network⁴. These resources allow not only to share published papers but also to provide drafts and manuscripts for collaborative works and peer review (and significantly increase a number of qualified reviewers [1]). The peer review within these resources may be carried out much faster than the traditional peer review procedure before acceptance of a paper in a journal or conference proceedings instead [2]. The "Content innovation" built in the "Article of the Future" project [3] provides authors with digital tools "...to disseminate their research in its full digital richness in Elsevier articles"⁵. These tools allow a scientific paper to be published in Elsevier's journals with interactive 3D visualizations, data sources, interactive plots, etc.

2.2 Use Traditional Models of Cloud Services

The services mentioned above allow authors to make theirs papers more vivid and interactive. However, researchers are limited by feasibilities of these services and

¹ <http://gigaom.com/2014/07/23/macmillan-invests-in-google-docs-for-researchers-firm-writelatex/>

² <http://www.mendeley.com/>

³ <http://www.researchgate.net/>

⁴ <http://www.lifescience.net/>

⁵ <http://www.elsevier.com/about/content-innovation>

sometimes do not have the ability to present all features of his scientific product (application packages, script or project of scientific environment, etc.). Use cloud computing approaches within models of SaaS, PaaS and IaaS [4] allows researchers to present functions of their original tools. He can provide remote desktop for a launch application package, make a remote interface or create another original decision for sharing functionality of his scientific tools and middleware, e.g. DataTurbine, OPeN Dap [5]. The first case is limited by the level of confidence between partners. Other cases may require large efforts and time spending. These efforts may be needed in the case when a distributed computational infrastructure is required. The growing amount of interdisciplinary problems, which require use of two and more regular and special application packages working in original computational resources, was the crucial factor to implement Scientific Workflow Paradigm.

2.3 Scientific Workflow Paradigm

Currently the scientific workflow paradigm (SWfP) is the fundamental instrument for scientific collaboration [6], [7]. SWfP enables composition and execution one, two and more application packages and tools in a distributed computational environment. On the one hand workflow (WF) is "... a high-level specification of a set of tasks and the dependencies between them that must be satisfied in order to accomplish a specific goal" [8]. On the other hand the WF can be presented as a computer program written in high level programming language. At the present time scientists have a wide choice of ways and tools to workflows representation (including WF languages and graphical instruments – GUI [9] to develop WF as Directed Acyclic Graph (DAG)) as well as a wide choice of Scientific Workflow Management Systems (SWfMS): Taverna, Kepler, Pegasus, etc. In particular the Taverna has instruments to share WFs realized within the "MyExperiment" project [8]. One of the key properties of SWfMS is WF execution scheduling, including tasks to resources mapping. There are lot of algorithms for task scheduling, which covers most properties and features of WFs (set of WFs) and distributed computational resources. For example, the set of scheduling algorithms for the multiple deadline-constrained workflows [10] was developed for heterogeneous computing systems with time windows. These algorithms may be useful for WFs shared for not only scientific collaboration, but also educational purposes, where the student has ability to run one or more real packages when required computational resources are idle.

The concept of collaborative scientific workflows [11] considers required features of collaborative WFs for "...enablement of scientific collaboration" including the life cycle of scientific collaboration. The research workflow platforms approach described in [12] offers to unite most of the useful features, mentioned above, in one digital solution. This work gives one successful result – platform eSciDoc developed for the aims of Max Planck Society. The authors of the paper note "Building and maintaining completely independent platforms for each domains... may not be sustainable". The concept of Research Object is encouraged to overcome this problem.

2.4 The Research Object Concept

Research Object (RO) may be defined as a container of scientific resources [13], which are received during all stages of the research life cycle. This container may include executable packages and WFs, data sources, program documentation, short descriptions, scientific papers, tools for collaborative work, etc. In this aspect the principles of RO ontologies are offered in [14]:

- P1. Preserving Data and Methods
- P2. Overcoming Obfuscation Through Annotation
- P3. Treating the Research Object as a Container
- P4. Citations and Credit in Research Objects
- P5. Treating Research Objects as Software

These five principles unite all the features of previous achievements mentioned above, standardize descriptions of the RO components and present RO as a whole structure. The paper [14] considers Research Object Digital Library (RODL) – a set of tools for “...collect, manages and preserves aggregations of workflows and related objects and annotations...”. The RO-Enabled “MyExperiment”⁶ is an extension of the virtual research environment mentioned above made on the base of workflow-centric RO.

Summarizing the above, we can see that the modern researcher has a big choice of tools, environments and platforms to facilitate his research and share related results in different forms for collaboration and peer review. These tools based on widely recognized standards of computer-aided objects descriptions, hence they are easy to master by scientists, who have a basic knowledge of ontologies, domain specific languages and contiguous areas of IT. Nevertheless, a researcher without the required knowledge and skills in appropriate areas of IT need to appeal to colleagues for preparing RO or learn required technologies and will be distracted from the main study.

3 The CLAVIRE Platform as a Means to Facilitate Sharing of Workflow-Centric Scientific Results

The CLAVIRE platform was developed by ITMO University. It based on the concept of iPSE (Intelligent Problem Solving Environment). This concept allows us “...to hide the technical details of a used infrastructure: it is possible to interact with the system with the use of domain specific languages which are automatically translated into the composite applications performed using the services available within eScience infrastructure.” [15].

The CLAVIRE lets the scientific team create eScience infrastructure based on data sources, application packages installed in distributed heterogeneous high-performance environment and solving complex issues, including multidisciplinary issues. The iPSE concept, implemented in the CLAVIRE, allows us to describe the computational process of problem solving in the form of *abstract* workflow, which are translated into instructions of executing application packages presented as cloud (remote) services

⁶ <http://alpha.myexperiment.org/>

(concrete WF). An abstract WF described in the special domain specific language called “EasyFlow”. It is a very easy language, which defines WF as a sequence of simple steps (Q2). These steps contain definitions of input data and execution mode (*standard, urgent*, etc.) of application packages integrated into CLAVIRE. In order to present an installed application package as a cloud service (integrated into CLAVIRE), it is enough to describe it in the special language called “Easy package” [15]. The description contains statements of the input and output parameters, instruction of execution in the form of a command line. The package’s owner may make compositions of input and output parameters of cloud service different from input and output parameters of the application package itself. All data conversion instructions are described in the form of lambda expressions or a more complex form of object, which are presented in the Ruby language. The EasyPackage supports all structures of this language. Therefore, an owner of the package has many opportunities to present one application package in different forms depending on the case study statement and/or subject area. These opportunities allow to develop a joint-free composite application and do not require a researcher to digress from the main study. The simplified grammars of the EasyFlow and the EasyPackage were presented in [15].

The CLAVIRE platform contains several user interfaces (UI) to develop and execute workflows and present it as a cloud service (P5) on the base the model AaaS. This model allows us to create and use composite applications (CA), which consists of interacting cloud services for solving complex tasks: data source access, computer simulation, results processing and analysis [16]. The CLAVIRE provides data preservation (inputs and outputs of WFs and interim data) and the means for it transfer, converting, upload, download and visualization (P1). When the user describe the application package on the EasyPackage, he can type information about his future service (partly annotate it – P2).

There are four main UIs of the CLAVIRE (the only web-browser is required), which were described in [15], [17] and [18].

The WF-based UI is the most close to principles of the RO ontologies, because WF project may include WF scripts with commands to application packages launching, input, output and interim data of WF and other data, documents in any format (including documentations, descriptions and scientific papers). This interface provides users with means for creating WFs, working with data and media, visualizing results (pictures and file in the “Cube” format), sharing a certain WF or the whole project for collaboration, and other reasons.

In addition, the VSO UI is close to the principle of RO ontologies, since VSOs and VSO projects refer to data and application packages as methods, well represents real process in the form entities, which are related each other in defined manner.

The CLAVIRE contains a Web 2.0-based component called “Site,” which allows users to make sites with required media and embed containers with the first, second, and fourth CLAVIRE UIs considered above. The web site’s developer can specify a project, which will be loaded when the user opens the appropriate web page and calls the embedded CLAVIRE UI. This instrument is useful for not only annotating executable scientific results but also creating remote educational resources with embedded Virtual Learning Labs [16].

In addition, the CLAVIRE supports the Interactive WF (IWF) method [19]: one or more services based on application packages communicate with each other when they are active (running).

Since a researcher, before sharing his executable scientific result, has to integrate his application package, he may be faced with the problem of learning the EasyPackage language. If CLAVIRE UIs do not require program activities, the package integration asks to develop text script with statements.

In addition, the CLAVIRE does not have means for 3D visualization (including stereo mode), which support large sized data (more than 10 Mb) and modern human interface devices, e.g. motion capture system Microsoft Kinect. This means should support simple graphical hardware for common desktops and top visualization system like the Center of Scientific Visualization and Virtual Reality (CSVVR) built in the Advanced Computer Lab (ACL), which was founded in 2011 in the ITMO University within Russian Government Resolution #220⁷. Since not all researchers have direct access to top visualization systems, the technological process is performed with the use of distributed eScience infrastructure without virtual reality system up to the stage of interpretation.

The next three sections are organized as follows. The next section describes Web-based intellectual editor, which facilitate the integration an application package into CLAVIRE. The section 5 deal with C++ based graphical library for the rapid development tools of 3D visualization. Section 6 presents the information about development composite application on the base IWF method (interaction between scientific package and visualization tool).

4 Web-based Intellectual Editor for the Integration Application Packages into the CLAVIRE

The Web-based Intellectual Editor called “Package Manager” (PM) was developed to facilitate the process of application packages integration into the CLAVIRE platform (Q2). The JavaScript-framework, HTML and CSS-tables was used for the PM implementation. For data transfer and requests the JSON text interchange format is used.

The PM provides package owner with interface, which offers the user to input packages general information (the package authors, the organization, license type, etc.), its inputs and outputs (Fig. 1).

In addition, the PM allows user to insert parameters, which are not input parameters of the application package. For each this parameter user can specify the expression in the Ruby’s notation, which is depended on existing parameters defined in the package description. This option allows owner to present his package in the desired form. If it is necessary, PM gives ability to define a validation expression for any input parameter, e.g. for foolproof or protection of the legitimate user.

⁷ The ACL was founded in the collaboration with the research group «Computational Science» at the University of Amsterdam led by prof. Peter M.A. Sloot. Please, see more by the link <http://www.p220.ru/en/home/projects/item/297-11-g34-31-0019/297-11-g34-31-0019>

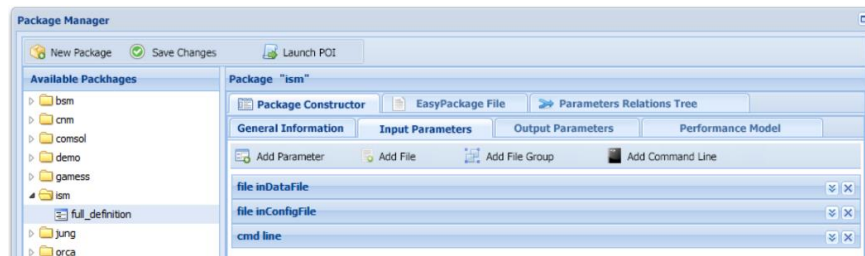


Fig. 1. The Example of the Package Manager interface

The command line is specified in the same form as in the usual case except the expression “{0}” puts execution file instead. This option does the package description independent on location on or more instances of the package installed in the distributed computational environment.

The ready package description PM generates into the text file with the code in EasyPackage language. The package owner can add Ruby’s classes for inputs and outputs conversion or make other changes. To check the prepared package description he can launch the POI from PM directly.

The PM rids the researcher to code the description script (Q2), hence save his time. The integrated into the CLAVIRE package become a cloud service, and the researcher can share it for all or selected users. If the source package was modified (e.g. for the improvement of some features), the owner can make the required changes in the package description so that input parameters of the relevant cloud service will be unaffected. This option makes workflows built on the packages of different owners stable from changes of source packages.

5 C++ based Library for the Rapid Development of Visualization Tools

The C++ based library for the rapid development visualization tools offers the template called “base_demo,” which can be changed to build the custom domain-specific visualization tool. This template contains the four abstract methods, which provide users with abilities to build prepared 3D dynamic graphical scenes:

1. *init* – the method that reads all required data and performs preliminary computations (e.g. calculation of the rotation matrix);
2. *update* – the method that performs a fast (no longer than 0,01 sec.) update of the visualization data (e.g. data renewal, addition/change/deletion of visualization objects);
3. *draw3d* – the method that calls the OpenGL functions to form the 3D scene (this method requires that all demanded OpenGL matrixes were calculated.);
4. *draw2d* – the method that calls the OpenGL functions to form 2D scene pictures: 2D diagrams, graphics, texts, etc. (this method requires that all demanded OpenGL matrixes were calculated.).

The template “base_demo” allows developers to connect the library Kinect for Windows SDK и Kinect for Windows Developer Toolkit, hence providing operators with the ability to control scenes through hand gestures (Q1). This library supports the detection of the following hand flaps:

- two hands round flap away from the body – distancing the scene;
- two hands round flap towards the body – the scene approaching;
- one hand flap in parallel the body – rotation of the scene.

The C++ base library considered above supports the mode of stereoscope visualization of different type (anaglyph, polarization, etc.). The special glasses are required for the immersion in a virtual environment. Fig. 2 shows the operator interacting with the one of visualization tools in the CSVVR of the ITMO University.

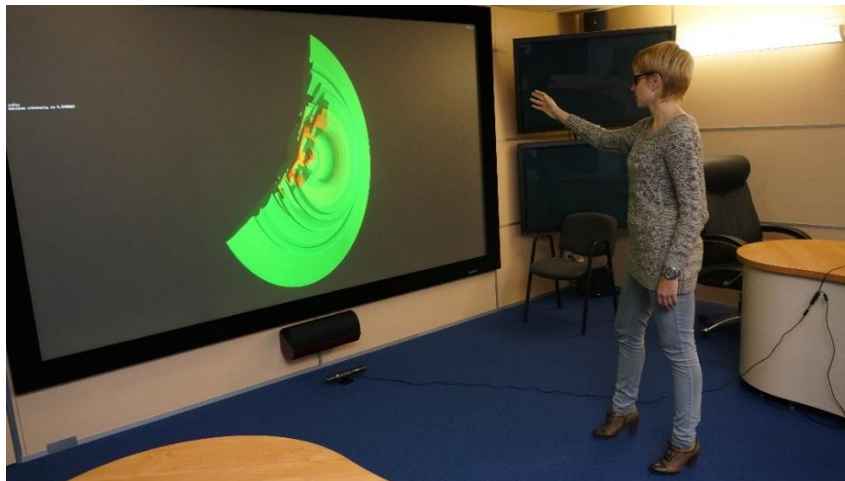


Fig. 2. The interaction between operator and visualization tool through MS Kinect.

The released visualization tool based on template “base_demo” may be run both on a desktop with the discrete video adapter and on a top system of visualization and virtual reality. These abilities allows the researcher to use the developed visualization tool at job or install and share it in the organization, which has top visualization system.

6 Developing the Composite Application for the Interactive visualization

The composite application (WF) is developing through the WF-based UI or VSO interfaces and contains a least two steps: *run_package* and *run_vis*: which calls integrated application packages: scientific package (e.g. simulation package) and visualization tool. This composite application is implemented with the use of the IWF method: interaction between one of scientific packages and visualization tool (Q1).

Fig. 3 shows the interaction of packages with the use of the ZMQ protocol. At first, the instance of the scientific package is launched as a translator of a data by request. After that (without waiting for sign off scientific paper), the IWF calls the instance of the visualization tool. The last receives data from the scientific package required data, shows the visualization picture, and waits for the operator's commands. The commands of the operator are performed and sent to the scientific package. If necessary, the scientific package performs the required calculations and sends data updates to the visualization tool. If the visualization tool receives the operator's command "Exit," then it sends the relevant command to the scientific package. In this case, both packages are closed.

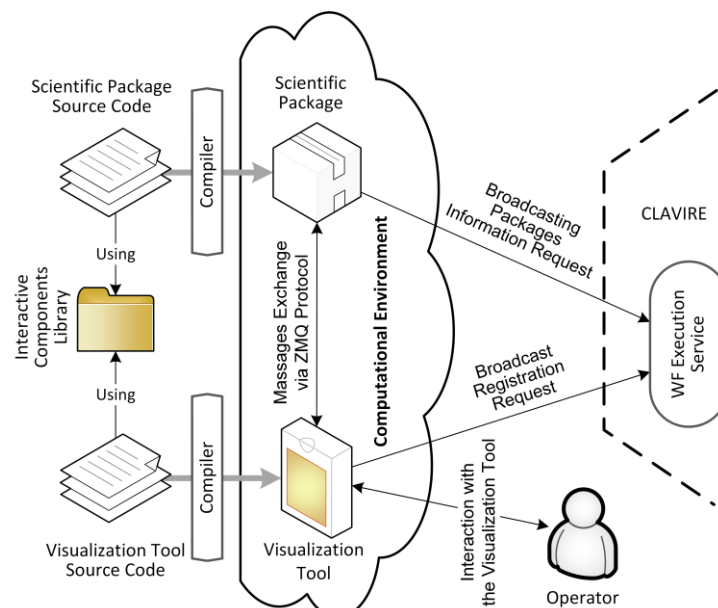


Fig. 3.

The IWF method allows the researcher to present scientific results in the modern visualization form with the use of top systems of virtual reality even scientific package and visualization tool work in different places.

7 The Examination of the Developed Approach and Tools by Evolving Russian Yung Scientists

The developed approach, CLAVIRE platform, PM, and C++ based visualization library were examined by an evolving group of 20 young scientists (from 23 to 33 years old), who are in the different cities of Russian Federation (including Nizhniy Novgorod,

Moscow, Ufa, Rostov-on-Don, and Novosibirsk). All young scientists were as a participants or winners in/of competition in computational science [20], but not more half of them had good skills in the area of programming.

We split this group into two equal subgroups (10 participants). Members of the first subgroup had to develop relevant to theirs scientific packages (which supported the mode of the command line) visualization tool and integrated both solutions into CLAVIRE. Members of the second subgroup had to integrate theirs scientific packages into CLAVIRE only. The first subgroup had six weeks for solving all tasks and the second subgroup – three weeks. In addition, all participants in the experiments were required to prepare a report with the scientific area and application package description, visualization concept, description of the package integration and of visualization tool.

Fig. 4 shows the process of application packages integration and using the visualization tool (Q3). In the case if young researcher decided to send us the scientific package, administrator of CLAVIRE installed it on our resources. Participant did have the rights to install the package for security reasons. If scientist did not send the package (three of members had packages which original requirements to computational resources), he shared his resource within CLAVIRE's requirements to launch it in the external re-source (gave information about location of the package and its input and output files, and order to launch the package).

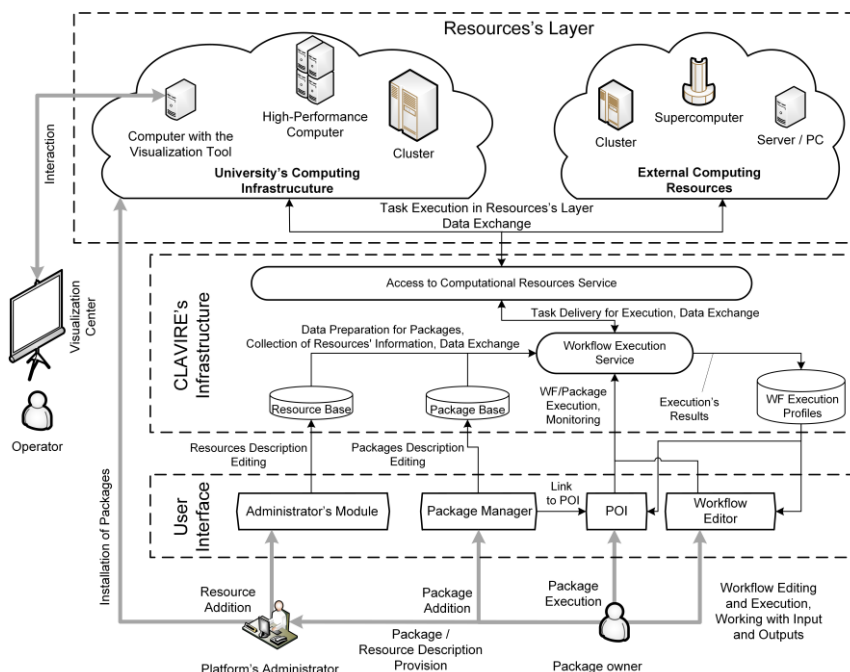


Fig. 4. The scheme of packages integration into CLAVIRE and using the visualization tool.

After that, all the participants described his packages via PM and checked these descriptions with the use of POI. Then almost all the members of the first group developed

visualization tools and sent them to the administrator, who installed them in our CSVVR. Moreover, they developed the composite application – IWF, which included calls of the scientific package and visualization tool. The members of the second group made a regular composite application with calls of one or more scientific application packages only.

All the participants prepared reports and included them into their WF project. After that, we checked the composite applications and visualization tools.

All the members of both groups successfully integrated their scientific packages into CLAVIRE. Nine of the 10 first-group members developed their visualization tools. Three of these tools (33,3%) show an excellent picture, four (44,5%) show a good picture, and two of them (22,2%) could present a picture of satisfactory quality. Therefore, almost all participants had mastered the C++ based library. One member did not develop his visualization tool since he was overcrowded during the experiment period. The participants spent from one to three weeks to integrate their packages via PM. Only two of them asked many questions; the other members mastered the PM without serious problems. The wide scatter of time length is explained from different levels of the members' workload.

Fig. 5 shows the results of four visualization tools, which represent calculations of packages built within the following research.

1. Active data storage processes visualization for the load balancing dynamical distribution in computational nodes of the cluster [21] (Fig 5, a);
2. The paraxial distribution simulation of ultrashort laser pulses [22] (Fig 5, b)
3. Subsurface hydromechanics: simulation of straining action during oil field development [23] (Fig 5, c);
4. Electromagnetic field simulation with the use of iteration solvers for the Maxwell's equations [24] (Fig 5, d).

Fig. 5 shows that all pictures presented on the simply flatness look informative and visually poor. This negative effect explains that the studied phenomena have powerful 3D-demencial features in stereo mode, which are overlapped and interfused in the 2D picture. This feature implicitly confirms the necessity of applying means of interactive 3D (stereo) visualization for complex processes and systems study.

All the participants used PM, where they described (including general information, authors and organization, inputs and outputs – P2 and partly P4) and prepared executable cloud service. They used POI for the running the prepared cloud service (P5) and WF-based UI for development the composite application (WF), adding data for the experiment and demonstration (including results in different form of presentation), which is stored in the WF-project (P1). The participants also added into the WF-project prepared within the experiment report and papers, related theirs research areas (partly P3). We do not claim for the full compliance with RO ontologies, but we showed, that these principles are useful for creating scientific collaboration environments based on the CLAVIRE platform and developed tools. Using them, scientists can develop and share easy and rapidly own scientific results in the forms, related modern achievements of eScience (including executable cloud services and 3D visualization tools).

The participants of the experiment did not use VSO interface, since they was solving problem in one scientific area and had on or two scientific packages only. The VSO concept and relevant interfaces are useful for heads on scientific group, who lead inter- and multidisciplinary projects. The application of the VSO-interfaces is one of future works topics.

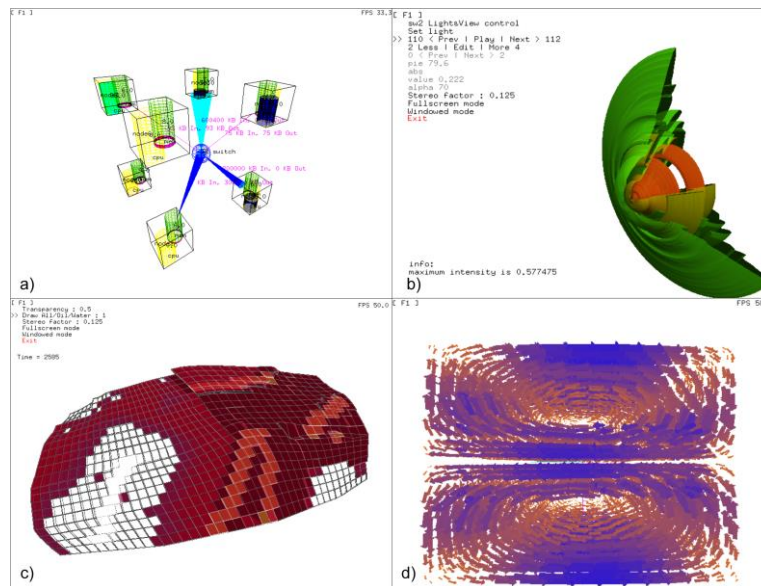


Fig. 5. Examples of pictures received with the use of developed visualization tools.

8 Conclusion and Future Works

Modern scientific society is characterized by many changes in methods, approach and technologies. Modern researcher should respond quickly these changes and offer rapidly novel scientific results in vivid and interactive manner.

In this paper, we developed an approach to extend the CLAVIRE platform for sharing traditional scientific documents, data sources, executable services, and 3D-visualization means for collaborations and peer review (Q1). We developed a Web-based intellectual editor, “Package Manager,” for rapid and easy integration of scientific packages into CLAVIRE (Q2). We created C++-based graphical library for developing visualization tools with a motion capture sensor (Microsoft Kinect) support. The IWF method is offered to develop a composite application with the interaction between the scientific package and visualization tool (Q1) via ZMQ protocol. As was shown above the approach, the CLAVIRE’s interfaces and developed tools are within five RO ontologies principles. Only the C++-based graphical library requires skills in programming, but this library significantly facilitates the development of visualization tools (Q2).

The approach mentioned above was examined via 20 young, evolving scientists from more than 10 cities located across the Russian Federation. All of the participants successfully integrated their scientific packages into CLAVIRE through PM (we created a prototype of the scientific environment (Q3), which contains 20 new cloud services). Seven of the ten participants developed visualization tools that present a good, quality 3D picture and a relevant composite application.

In the future, we are planning to develop a graphical platform for making a virtual environment on the base existing graphical model and high-realistic graphical engine [25]. This platform should allow scientists to create 3D scenes without programming efforts. In addition, we will perform experiments by evolving scientific teams, which work on inter- and multidisciplinary problems (with the use of VSO interfaces). Furthermore, we are planning to apply the developed means in the implementation of double-degree international educational programs [26]. Moreover, we intend to carry out the social research on the subject of users' opinions about the offered approach.

Acknowledgement

This paper is supported by Russian Scientific Foundation, grant #14-21-00137 "Supercomputer simulation of critical phenomena in complex social systems". The research is done in Advanced Computing Lab (ITMO University), which is opened in frame of 220 Decree of Russian Government, contract #11.G34.31.0019

References

1. Hardaway D (2005) Sharing Research in the 21st Century: borrowing a page from open source software. *Commun ACM* 48:125–128.
2. Meyer D (2014) Academic social network ResearchGate aids debunking of stem cell study. In: GIGAOM. <http://gigaom.com/2014/03/14/academic-social-network-researchgate-aids-debunking-of-stem-cell-study/>. Accessed 29 Jul 2014
3. IJsbrand A, Atzeni S, Koers H, Elena Z-S (2014) Bringing Digital Science Deep Inside the Scientific Article: the Elsevier Article of the Future Project. *Lib. Q.* 24:
4. Mell P, Grance T (2011) The NIST Definition of Cloud Computing Recommendations of the National Institute of Standards and Technology. Gaithersburg
5. Barseghian D, Altintas I, Jones MB, et al. (2010) Workflows and extensions to the Kepler scientific workflow system to support environmental sensor data access and analysis. *Ecol Inform* 5:42–50. doi: 10.1016/j.ecoinf.2009.08.008
6. Elmroth E, Hernández F, Tordsson J (2010) Three fundamental dimensions of scientific workflow interoperability: Model of computation, language, and execution environment. *Futur Gener Comput Syst* 26:245–256. doi: 10.1016/j.future.2009.08.011
7. Manuel S, Batista V, Dávila AMR, et al. (2008) OrthoSearch: A Scientific Workflow Approach to Detect Distant Homologies on Protozoans. *Proc. 2008 ACM Symp. Appl. Comput.* New York, New York, USA, pp 1282–1286
8. Deelman E, Gannon D, Shields M, Taylor I (2009) Workflows and e-Science: An overview of workflow system features and capabilities. *Futur Gener Comput Syst* 25:528–540. doi: 10.1016/j.future.2008.06.012

9. Oinn T, Addis M, Ferris J, et al. (2004) Taverna: a tool for the composition and enactment of bioinformatics workflows. *Bioinformatics* 20:3045–54. doi: 10.1093/bioinformatics/bth361
10. Bochenina K (2014) A Comparative Study of Scheduling Algorithms for the Multiple Deadline-constrained Workflows in Heterogeneous Computing Systems with Time Windows. *Procedia Comput Sci* 29:509–522. doi: 10.1016/j.procs.2014.05.046
11. Lu S, Zhang J (2011) Collaborative scientific workflows supporting collaborative science. *Int J Bus Process Integr* ... 15:39 – 47. doi: 10.1109/MIC.2011.87
12. Förstner K, Hagedorn G, Koltzenburg C, et al. (2011) Collaborative platforms for streamlining workflows in Open Science. *Nat Preced.* doi: 10.1038/npre.2011.6066.1
13. Bechhofer S, De Roure D, Gamble M, et al. (2010) Research Objects: Towards Exchange and Reuse of Digital Knowledge. *Nat Preced.* doi: 10.1038/npre.2010.4626
14. Belhajjame K, Zhao J, Garijo D (2014) The research object suite of ontologies: Sharing and exchanging research data and methods on the open web. *arXiv Prepr. arXiv* ...
15. V. Knyazkov K, V. Kovalchuk S, N. Tchurov T, et al. (2012) CLAVIRE: e-Science infrastructure for data-driven computing. *J Comput Sci* 3:504–510. doi: 10.1016/j.jocs.2012.08.006
16. Dukhanov A, Karpova M, Bochenina K (2014) Design Virtual Learning Labs for Courses in Computational Science with Use of Cloud Computing Technologies. *Procedia Comput Sci* 29:2472–2482. doi: 10.1016/j.procs.2014.05.231
17. Kovalchuk S V., Smirnov PA, Kosukhin SS, Boukhanovsky A V. (2012) Virtual Simulation Objects concept as a framework for system-level simulation. 2012 IEEE 8th Int. Conf. E-Science. IEEE, pp 1–8
18. Smirnov PA, Kovalchuk S V., Dukhanov A V. (2014) Domain Ontologies Integration for Virtual Modelling and Simulation Environments. *Procedia Comput Sci* 29:2507–2514. doi: 10.1016/j.procs.2014.05.234
19. Knyazkov K V., Nasonov D a., Tchurov TN, Boukhanovsky A V. (2013) Interactive Workflow-based Infrastructure for Urgent Computing. *Procedia Comput Sci* 18:2223–2232. doi: 10.1016/j.procs.2013.05.393
20. Sloot PMA, Boukhanovsky A V. (2012) Young Russian researchers take up challenges in the computational sciences. *J Comput Sci* 3:439–440. doi: 10.1016/j.jocs.2012.08.009
21. Tyutlyaeva E, Kurin E, Moskovsky A, Konuhov S (2012) Abstract: Using Active Storage Concept for Seismic Data Processing. 2012 SC Companion High Perform. Comput. Netw. Storage Anal. IEEE, pp 1389–1390
22. Suvorov E, Akhmedzhanov R, Fadeev D, et al. (2011) On the Peculiarities of THz Radiation Generation in a Laser Induced Plasmas. *J Infrared Millim Terahertz Waves* 32:1243 – 1252.
23. Biktimirov MR, Biryaltsev EV, Demidov DE, et al. (2012) INFORMATION INFRASTRUCTURE OF TATARSTAN: FROM «SENet-Tatarstan» TO «SEGrig-Tatarstan». *Russ. Natl. Supercomput. Forum*
24. Butyugin DS, Il'in VP, Petukhov AV (2009) Comparative Analysis of Approaches for High Frequency Electromagnetic Simulation. *Prog. Electromagn. Res. Symp. Proc. Moscow*, pp 1483 – 1487
25. Bezgodov A, Esin D, Karsakov A, et al. (2013) Graphic toolkit for virtual testbed creation: application for marine research and design. *Dyn Complex Syst century* 7:34.
26. Dukhanov A V., Krzhizhanovskaya V V., Bilyatdinova A, et al. (2014) Double-degree Master's Program in Computational Science: Experiences of ITMO University and University of Amsterdam. *Procedia Comput Sci* 29:1433–1445. doi: 10.1016/j.procs.2014.05.130

iDAF-drum: Supporting Everyday Practice of Drum by Adding an Unperceivable Factor

Kazushi Nishimoto, Akari Ikenoue ^{*}, and Masashi Unoki

Japan Advanced Institute of Science and Technology
knishi@jaist.ac.jp, identity0811@yahoo.co.jp, unoki@jaist.ac.jp
<http://www.jaist.ac.jp/ks/labs/knishi/>

Abstract. To achieve excellent drum performances, sufficient use of the extensor muscles of the wrists is important. However, it is actually very difficult and there have been no efficient methods and tools to train them. This paper proposes iDAF-drum, which is a novel training system of the extensor muscles in everyday drum practice. “iDAF” is an acronym of “insignificantly delayed auditory feedback” and usual people cannot perceive such a very slight delay. We found an interesting phenomenon that drummers raise the drumsticks higher than usual by inserting the unperceivable delay between impact and sound. By exploiting this phenomenon, iDAF-drum can efficiently train the drummers’ extensor muscles without giving them any unusual feeling. We demonstrate the efficiency of iDAF-drum based on user studies and discuss possibilities of exploiting unperceivable factors like iDAF for effectively bringing out and fostering buried creative abilities.

Keywords: Unperceivable factors, drum practice, delayed auditory feedback, illusory feelings.

1 Introduction

Good control of drumsticks is very important in drum performance. A drummer must not only drum at accurate tempo with adequate strength but also control the tone of the drum through the motion of the sticks. Wrist motion is the key to stick control. The flexor muscle, which contracts to cock the wrist, and the extensor muscle, which contracts to extend the wrist, govern the wrists’ motion. Balanced usage of both muscles allows the drummer to drum fast for a long period as well as to improve the tone of the drum[1–4].

However, it is generally difficult to master the technique of stick control where the extensor muscle is sufficiently used. Special heavy sticks have often been used for practicing the intentional use of the extensor muscle in traditional drum training methods. However, using such heavy sticks causes damage to wrists due to overload. A new method for training the dominant use of the extensor muscle has recently been developed [5]. In this method, a special drumming form that

^{*} Currently with Next Co., Ltd.

2 iDAF-drum: Supporting Everyday Practice of Drum

forces the drummer to use the extensor muscle is proposed. However, this form is quite different from normal ones. The drummer has to master the special form only for training the extensor muscle. Thus, this method is not so efficient because it requires extra training time.

This paper proposes a novel method exploiting an “insignificantly delayed auditory feedback (iDAF)” for training drum stick control and a support system for drumming practice named iDAF-drum [6]. Here, insignificantly delayed auditory feedback means a very short time delay between the impact of the stick with the drumhead and the emission of sound. Typically, humans cannot recognize the existence of iDAF: iDAF is unperceivable. Different from the conventional methods, drummers will not find any difference between drumming under our method and the conventional one. Nonetheless, it allows drummers to efficiently train in a way where the extensor muscle is sufficiently used.

2 Related Works

Several support and augmenting systems for drum performance have been studied so far. Jam-O-Drum [7] is a collaborative multimedia percussion system for performing interactive improvisations. Voice Drummer [8] is a system for inputting a percussion score by so-called “voice percussion.” Many such systems focusing on drumming performance have been proposed, created and studied. However, they did not support practice while training how to use the extensor muscle of the wrists.

In contrast to performance training, there have not been so many attempts to support practice of the drum. Iwami and Miura [9] studied a computer system to help drummers practice loop patterns of the drum. It visualizes situations of drumming such as fluctuation of timing and impact strength to allow drummers to self-check their performances. This system shows where mistakes are made, but it does not tell the drummer how to practice to correct them. Beatback [10] is a system for supporting rhythm practice. However, “practice” here means “exploration” or “creation” of novel rhythmic patterns. The system encourages such generating processes of rhythm by working as a virtual partner of musical performance. Thus, this system does not truly focus on correction of the wrong drumming form.

Tsuji and Nishitaka [11] developed a system to improve drumming form. By showing rhythm lapses, drum-form lapses, hand-stroke amplitude, and striking strength, it leads the drummer to correcting his/her wrong form. This objective of that study is quite similar to ours. However, the way to correct the wrong form is indirect: No concrete instructions are provided. In contrast, we attempt to directly correct wrong usage of the muscles.

Several patents to improve drumming form have been applied for. “Practice aid device for percussionists” [12] proposed a small spacer for correcting the way of holding the drumsticks. “Muscle control development system and kit therefor” [13] proposed drum pads made of elastic materials. By preparing multiple pads whose degree of rebounding are different and by drumming on those pads, users

can train in ways of stick control while hitting drums having different rebounding features. They aimed at improving drumming form, but they did not focus on training the extensor muscle.

“Drummer stick control up-stroke practice method and device” [14] proposed a special attachment that is mounted over a drumhead: A horizontal bar is set tens of centimeters above the drumhead. By intentionally hitting the bar by the up-stroke of the sticks just after hitting the drumhead, users can learn ways to intentionally raise the sticks after impact. The objective of this patent is similar to ours: It focuses on the up-stroke that requires the drummer to use the extensor muscle. However, it also imposes special and unusual ways of performance on the drummer. There is the risk of he/she adopting the bad habit of excessively raising the sticks.

Consequently, although various systems for supporting the practice of drumming performance have been proposed, created and studied, no system has focused on training the extensor muscle. Furthermore, there has been no attempt to exploit the effects of unperceivable factors like iDAF.

3 Proposed Method and System

3.1 Definition of iDAF and Method

Delayed auditory feedback (DAF) usually means a feedback of voice to its speaker with a 100~200 msec delay. It is well known that such a DAF prevents the speaker from smoothly speaking, since it leads to the phenomena of repeating syllables and stuttering [15].

If a DAF is applied to the performance of a musical instrument, behaviors in the performance change. In the case where a person repeatedly taps using his/her forefinger, the raising height of the forefinger tends to increase if the tapping sounds are delayed [16]. Therefore, by applying this result to drumming, we can expect that the raising height of the drumsticks will increase and that this will in turn provoke a motion that makes much greater use of the extensor muscle. However, such a large delay as 100~200 msec makes it difficult for people to play musical instruments, and they became unable to keep accurate rhythm [17]. As a result, it becomes practically impossible to practice drumming.

Insignificantly delayed auditory feedback (iDAF) is an auditory feedback with a very short delay that people normally cannot perceive. It is known that people can usually perceive a time lag between an event and its sound if it is longer than 20~30 msec [18, 19]. Therefore, we define iDAF, in this paper, as an auditory feedback with a delay not exceeding 30 msec.

We propose a novel method for drumming practice that exploits iDAF. It provides an unperceivable delay between the impact of the stick and the sound emission from the impact. If it can change a drummer's motion similar to the behavior change of the forefinger tapping with a long delay [16], the extensor muscle would come to be used much more without obstructing the drumming.

4 iDAF-drum: Supporting Everyday Practice of Drum

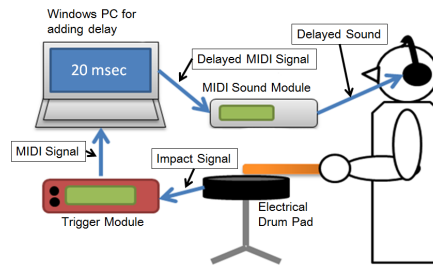


Fig. 1. System setup of iDAF drum

3.2 System Setup

Figure 1 illustrates the system setup of iDAF-drum. iDAF-drum consists of an electrical drum pad (YAMAHA TPS80S), a trigger module (YAMAHA DTX-PRESS), a MIDI sound module (Roland SD-50), a USB-MIDI interface (YAMAHA UX-16), and a Windows PC (Windows VISTA). An impact signal from the electrical drum pad is input to the trigger module and then converted to a MIDI signal. The MIDI signal is input to the Windows PC. After a given time (shorter than 30 msec) passes, the PC inputs the signal into the MIDI sound module. Finally, a hitting sound is emitted with an insignificant delay.

For convenience of data analysis, when the Windows PC receives the MIDI signal from the trigger module, the PC outputs a pulse signal from its serial port that is used for synchronizing with the myoelectric potential data of the extensor muscles (we call this pulse signal “synchronization signal” hereafter). We implemented the software runs on the PC for adding the delay while using C#. In order to achieve 1-msec-order resolution for adding the delay, we used the Windows API functions. All of the sounds of the performance and a metronome are output from a headphone. We assigned a snare drum tone as the performance sound.

4 Estimating Effects of iDAF on Drumming

To the best of the authors’ knowledge, there has been no study on estimating the effects of such a very slightly delayed auditory feedback as iDAF. Therefore, this section investigates the effects of iDAF on drumming performance. First, we confirm that iDAF produces no negative effects, and then we investigate whether iDAF changes the drumming behavior so that the drummer begins to use the extensor muscle to a much greater extent.

4.1 Experimental Procedure

We employed 12 subjects (Sex: 7 males and 5 females; Age: Average 20.67 y/o, STDV 3.94; Drum experience: Average 5.78 years, STDV 3.67, Max 13.0 years,

Min 0.5 years). We asked them to hit the electrical drum pad using sticks held by the right and left hands alternatively along with a metronome sound that ticks every 250 msec. We instructed them to hold the sticks in a matched-grip manner.

Before starting experimental drum performances, we measured the maximum voluntary contraction (MVC) of the radial extensor muscle of the wrist three times for each of the subjects. In the experiment, we asked each subject to conduct four performance sessions whose delay times changed in the order of 0, 20, 10 and 30 msec. The subjects were asked to perform for 90 sec in each session. How they synchronized with the metronome was individually different, e.g., some subjects attempted to set the timing of sound emission to the metronome sound while others attempted to set the timing of impact to the metronome. Therefore, to standardize the way of synchronization, we asked the subjects to wear an eye mask and to attempt to set the timing of the sound emission to the metronome as much as possible. No warm-up was permitted. During each session, the myoelectric potential data of the extensor muscle were measured using electromyography (TEAC Polymate AP1532). In addition, we recorded the performances using a high-speed video camera. After finishing each session, we asked the subject whether he/she felt or found anything unusual.

To increase the signal-to-noise ratio of the myoelectric potential data between the synchronization signal and 250 msec before that, we applied a signal-averaging method to the obtained data, and the root mean square (RMS) of the cleaned data was calculated. Then, the data were normalized using the MVC of each muscle so that the MVC value was set to 100 percent.

4.2 Results

The ANOVA for average IOI values of all subjects showed no significant main effect with or without delay ($F(3, 44) = 0.4, p < 0.754$). In the previous studies on DAF with 100~200 msec delay, an expansive speaking phenomenon was observed. However, such an expansive phenomenon does not arise under the iDAF condition. We also calculated the coefficient of variation (CV), which is usually used as an indicator of confusion caused by DAF. As a result, we could not find any significant difference between with and without insignificant delay ($F(3, 44) = 0.436, p < 0.728$). From these results, we can conclude that iDAF does not at all cause the confusion that arises in DAF. From the inquiry results obtained after each session, only one subject reported that he felt a slightly strange feeling when delay was given. However, he did not become unable to perform drumming and he could play as usual. None of the other subjects felt any difference or strangeness.

Figure 2 shows snapshots taken with the high-speed camera at one minute from the beginning of each session. The small red circle in each picture shows the head of the stick. The upper row shows the top reach point of the stick held by the right hand and the lower row shows that of the left hand. The delay times were 0, 10, 20 and 30 msec from the leftmost to the rightmost pictures. It can

6 iDAF-drum: Supporting Everyday Practice of Drum

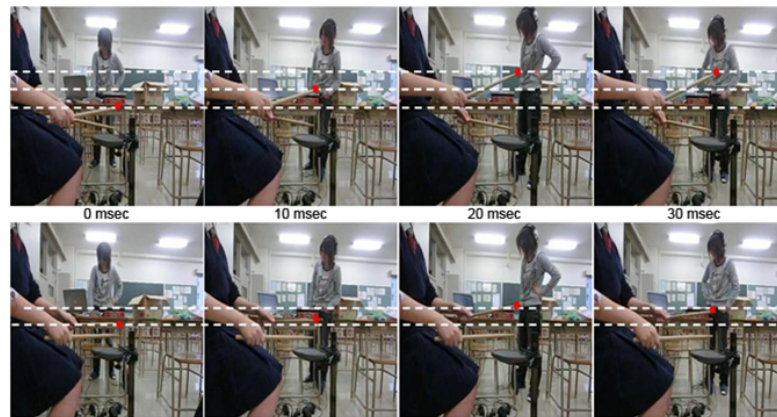


Fig. 2. Top reach points of the stick for 4 delay times. Upper row: Right hand, Lower row: Left hand

be seen that the more the delay time increases, the higher the stick is raised. This tendency was observed for most of the subjects.

Figures 3 and 4 show average electromyograms (EMGs) of the extensor muscle of the left and right wrists, respectively, for each delay time. The vertical line drawn at 250 msec shows the impact timing. Figure 3 shows that the peak value of EMG with no delay (blue line) is smaller than the other peak values with delays. In particular, the peak value with 10 msec delay (broken red line) is about 0.6% larger than that with no delay. In contrast, Figure 4 shows that the peak value of EMG with no delay is larger than those with delays. We calculated gross value of EMG for each delay and compared the gross value with no delay to the gross value with each delay by the t-test. As a result, the difference between the gross value with 10 msec delay and that with no delay for the left hand is marginally significant ($t(11) = 1.892, p < 0.085$). However, no significant difference could be found in all other combinations. As shown in Figure 2, the subjects raised the sticks higher under the iDAF conditions. Although this suggests that they used the extensor muscle much more, it was not supported by the electromyograms.

4.3 Discussion

From the results shown in Figure 2 and responses to inquiries after each session, we obtained very interesting findings: The motions of drumming performance changed due to iDAF and the subjects tended to raise the sticks much higher than usual (with no delay), although they were not aware of the delays. This suggests that iDAF can successfully induce the performer to sufficiently use the extensor muscle in drumming performance without perceiving differences. It remains unrevealed why humans react to such unperceivable auditory delays. This is a very interesting research issue, but it is out of the scope of this paper;

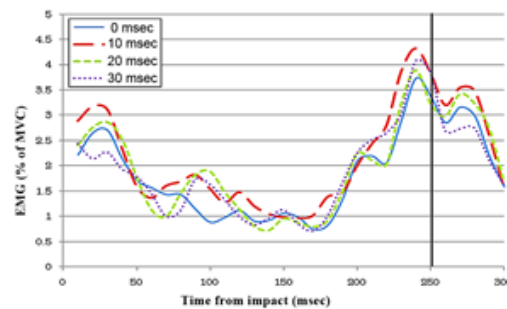


Fig. 3. Electromyograms of extensor muscle of left hand for each delay time

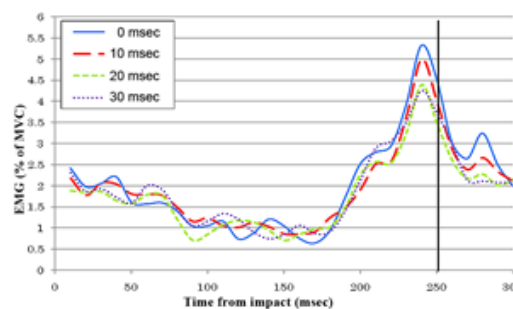


Fig. 4. Electromyograms of extensor muscle of right hand for each delay time

the objective of this paper is to create a useful support system for practicing drum performance. We only apply this phenomenon to the support system in this paper; elucidating its mechanism remains a future work.

Although iDAF may be able to provoke dominant use of the extensor muscle, the results shown in Figures 3 and 4 do not support the idea that the subjects came to use the extensor muscle more than usual when iDAF was given. Rather, although no significance was obtained, the longer the delays were, the less the extensor muscle was used for the right hand. For the left hand, the extensor muscle was mostly used with 10 msec delay, and then the longer the delays became, the less the extensor muscle was used.

It can be assumed that the reason why the sticks were raised higher when the delays were given is that the subjects unconsciously control the impact timing. However, to do so, they probably use not only their wrists but also their elbows.

It becomes impossible to effectively make subjects use their wrist extensor muscle if their elbows move. A possible method of making their elbows not move is to immobilize their upper arms. However, such a method imposes too great a load on the extensor muscle, and it can damage the extensor muscle in a similar way to the method using heavy sticks. Therefore, we think a better method is to make the impact-impact interval so short (namely, to make play faster) that the drummers need not move their elbow to raise the sticks higher

8 iDAF-drum: Supporting Everyday Practice of Drum

● **Single-stroke**

Perc. $\text{♩} = 120$

$\text{H} \frac{1}{2}$

R R R R R R R R

x 2 minutes

● **Change-up**

Perc. $\text{♩} = 120$

$\text{H} \frac{1}{2}$

R L R L

R LR LR LR L

Perc. $\text{♩} = 120$

$\text{H} \frac{1}{2}$

RLRLRLRLRL

RLRLRLRLRLRLRLRL

Perc. $\text{♩} = 120$

$\text{H} \frac{1}{2}$

RLRLRLRLRLRLRLRL

R LR LR LR L

x 3 minutes

● **16th notes drumroll**

Perc. $\text{♩} = 120$

$\text{H} \frac{1}{2}$

RL RL RL RL RLRL RLRL

x 3 minutes

Fig. 5. Test pieces

for adjusting the impact-impact interval. As a result, we can expect iDAF to effectively make them use the extensor muscle. Conversely, it becomes difficult for them to greatly exercise the extensor muscle. However, people usually cannot immediately master such ways of using the extensor muscle: Long-term training is required. Therefore, it is preferable, from a safety viewpoint, to make them use this muscle little by little.

5 Estimating Efficiency of iDAF-Drum

In the previous section, we showed the possibility that iDAF-drum leads the drummer to sufficiently use the extensor muscle. This section investigates whether the drummer can eventually master the correct drumming way of sufficiently using the extensor muscle by continually using iDAF-drum. If this can be achieved, the myoelectric potential of the extensor muscle without delay will increase along with the progress of training and finally become as strong as that with delay.

5.1 Experimental Procedure

In this experiment, we employed five subjects who are members of a brass band of a high school and are included in the 12 subjects of the experiment shown in the previous section. We asked them to perform a 10-minute practice session every day, which includes 2-minute single-stroke practices for the left and right hands, a 3-minute change-up practice, and a 3-minute drumroll of 16th notes (Figure 5). Before starting this session, each subject was allowed a warm-up performance.

We set the metronome to a 500 msec interval. In the previous section, we pointed out that it is preferable to set the performance speed relatively fast.

However, some subjects did not have so much experience. If we set the performance speed too fast, they might not have been able to perform the test pieces. Therefore, we examined maximum speed for each subject before the experiment and set the speed as fast as all of the subjects could perform the test pieces. Under this metronome interval, the speed when 8th notes were performed was the same as in the experiment conducted in the previous section, and it became faster when shorter notes like 16th notes were performed. Therefore, the overall performance speeds of this experiment are faster than those in the experiment of the previous section.

The system setup of iDAF-drum is the same as that used in the previous section. However, in this experiment, we used two sets of iDAF-drum.

Each subject practiced 10 minutes every day using the iDAF-drum with a 20 msec delay for twelve days. On the first day, the sixth day, and the last day, we measured the MVC of the extensor muscle of each subject three times, and then we asked each subject to continue drumming along with the metronome at 250 msec intervals for 1.5 minutes using iDAF-drum with no delay and with 20 msec delay. During these performances, we measured electromyogram of the extensor muscle of both arms together with synchronization signals.

We applied the signal-averaging method to the obtained data to reduce noise and calculated the root mean square (RMS) of the cleaned myoelectric potential data between the synchronization signal and 250 msec before that. Then the data were normalized using the MVC of each muscle so that the MVC value was set to 100 percent. In addition, for each muscle and for each delay time, we calculated the accumulated electromyogram data between 30 and 90 seconds from the beginning and the average of the gross amount of muscle activities of all subjects.

5.2 Results

Figures 6 and 7 show the averaged electromyograms of the extensor muscles of both hands of all subjects with no delay on the three measuring days. In the figures, vertical black lines drawn at 250 msec show the impact timing.

From these figures, we can see that the peak values became higher day by day. We calculated the averages of the myoelectric potential data between the synchronization signal and 250 msec before that. The ANOVA for the averaged myoelectric potential data showed a significant main effect of the measuring day for the left hand ($F(2, 12) = 4.815, p < 0.029$). However, no such significance was shown for the right hand ($F(2, 12) = 2.062, p < 0.170$).

A supervisor of the brass band of the high school pointed out that one of the subject's drumming sound changed during the experiment: The sound became sharp and clear.

5.3 Discussions

If the drummer learns how to use the extensor muscle while being forced to use it by iDAF-drum, we can expect that the extensor muscle will come to be

10 iDAF-drum: Supporting Everyday Practice of Drum

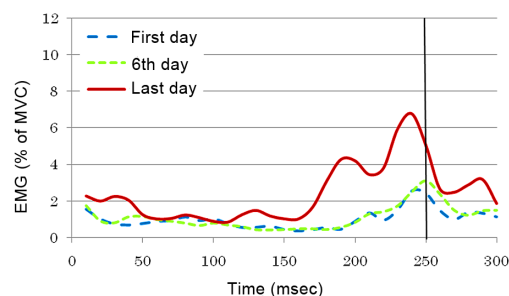


Fig. 6. Electromyogram of the extensor muscle of the left arm without delay

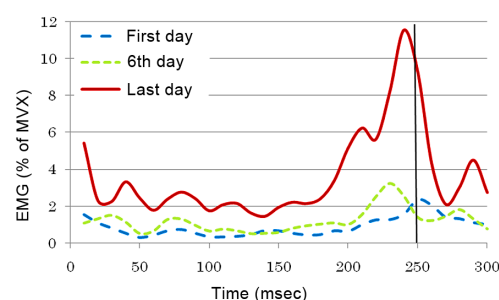


Fig. 7. Electromyogram of the extensor muscle of the right arm without delay

used even without delay day by day. From the results shown in Figures 6 and 7, although no significance could be obtained for the right hand, a day-by-day tendency for the extensor muscle to be used more without delay was observed.

We should conduct a control experiment by preparing a control group whose subjects practice using only a normal drum without delay. This time, unfortunately, we could not do so due to the lack of experimental equipment. However, most of the subjects we employed have practiced the drum for a long time (Average: 4.71 years, STDV: 3.29). If such experienced drummers can immediately master techniques for using the extensor muscle in only 12 days' practice, we can assume that they have already mastered the techniques. Therefore, it is unlikely that a control group would show changes within these 12 days. In contrast, we obtained some evident changes in this experiment. This fact supports the idea that iDAF-drum is effective for improving the use of the extensor muscle.

All of the subjects were right-handed, and they could use the right hand at will. From this, we can infer that they already had some skill in using the extensor muscle of the right hand. This is likely a reason why no significance was obtained for the right hand's myoelectric potential data (Figure 7). On the contrary, they could not use the left hand at will in the same manner as the right hand, and they did not have enough skill to use the extensor muscle of the left hand. As a result, the effect of training by using iDAF-drum clearly appeared.

This implication was supported by the result that significance was obtained for the left hand's myoelectric potential data (Figure 6).

6 Conclusions

In this paper, we proposed iDAF-drum to exploit the effects of insignificantly delayed auditory feedback in human behavior. Our purpose was to support drummers in mastering the sufficient use of the extensor muscle in controlling the drumsticks. Using iDAF-drum, we conducted user studies and investigated its efficiency. As a result, we found that the subjects came to raise the sticks higher than usual and that iDAF did not confuse the drummers in their performance. Furthermore, by continually using iDAF-drum, the myoelectric potential of the extensor muscle became stronger day by day. Consequently, we could confirm the usefulness of iDAF-drum.

In the case of iDAF-drum, the drummers' actions changed although they could not perceive the existence of the delay. We have applied iDAF to a keyboard such as a piano. Similar to the drum cases, performers could not perceive the delay. However, they reported that its keys were heavier when iDAF is inserted between key-touch and sound emission than those without delay. Thus, people involuntarily react to the unperceivable factors and they perceived them as a different feeling. These phenomena are very interesting. It suggests possibilities of the unperceivable factors that people's behaviors can be changed without being realized the existence of the factors by them and we can provide some different feelings without actually changing structure of an object (such as the keyboard) to provide the feelings. In future, we will further explore other cases that provide such illusory feelings and apply them for effectively bringing out and fostering buried creative abilities.

Acknowledgments

We thank the members of the light-music circle of Miyazaki University and the brass band of the high school associated with Kumamoto Gakuen University for their kind cooperation to our experiments. We also thank Dr. Kohei Matsumura of Ristumeikan University for his valuable advices on measuring electromyogram.

References

1. Fujii, S., Kudo, K., Shinya, M., Ohtsuki, T. and Oda, S.: Wrist Muscle Activity During Rapid Unimanual Tapping With a Drumstick in Drummers and Nondrummers, *Motor Control*, Vol. 13, pp. 237-250 (2009).
2. Osu, R., Franklin, D. W., Kato, H., Gomi, H., Domen, K., Yoshioka, T. and Kawato, M.: Shortand long-term changes in joint co-contraction associated with motor learning as revealed from surface EMG, *J. of Neurophysiology*, 88(2), 991-1004, 2002.

12 iDAF-drum: Supporting Everyday Practice of Drum

3. Fujii, S. and Moritani, T.: Rise rate and timing variability of surface electromyographic activity during rhythmic drumming movements in the world's fastest drummer, *Journal of Electromyography and Kinesiology*, Vol. 22, pp.60-66 (2012).
4. Fujii, S. and Moritani, T.: Spike shape analysis of surface electromyographic activity in wrist flexor and extensor muscles of the world's fastest drummer, *Neuroscience Letters*, Vol. 514, pp.185-188 (2012).
5. Nagano, Y., *Relax! Physical Drumming -An American way of stroke enables clear, fast and loud performance*, Ritto Music (2011). (in Japanese)
6. Ikenoue, A., Ogura, K., Unoki, M., and Nishimoto K., A supporting system to improve drumming form by applying insignificantly-delayed auditory feedback, *The Transactions of Human Interface Society*, Vol.15, No.1, pp.1-14 (2013). (in Japanese)
7. Blaine, T., and Perkis, T. The Jam-O-Drum Interactive Music System: A Study in Interaction Design, *Proc. The 3rd conference on Designing Interactive Systems: Processes, Practices, Methods, and techniques*, ACM Press (2000), 165-173.
8. Nakano, T., Goto, M., Ogata, J., and Hiraga, Y. Voice Drummer: A Music Notation Interface of Drum Sounds Using Voice Percussion Input, *Proc. the 18th Annual ACM Symposium on User Interface Software and Technology*, ACM Press (2005), 49-50.
9. Iwami, N., and Miura, M., A Support System for Basic Practice of Playing the Drums, *Proc. International Computer Music Conference*, pp.364-367 (2007).
10. Hawryshkewich, A., Pasquier, P., and Eigenfeldt, A., Beatback: A Real-time Interactive Percussion System for Rhythmic Practise and Exploration, *Proc. 2010 Conference on New Interfaces for Musical Expression*, pp.100-105 (2010).
11. Tsuji, Y., and Nishitaka, A., Development and evaluation of drum learning support system based on rhythm and drumming form, *Electronics and Communications in Japan (Part III: Fundamental Electronic Science)*, Vol. 89, Issue 9, pp. 11-21 (2006).
12. Anderson, R.E. Social impacts of computing: Codes of professional ethics. *Social Science Computing Review* 10, 2 (1992), 453-469.
13. Barke, S. J., Practice aid device for percussionists, United States Patent, US 7,473,836 B2 (Jan. 6, 2009).
14. Carter, D., Muscle control development system and kit therefor, United States Patent, US 5,744,737 (Apr. 28, 1998).
15. Bennett, H., Drummer stick control up-stroke practice method and device, United States Patent, US 2004/0025664 A1 (Feb. 12, 2004).
16. Lee, B. S., Effects of Delayed Speech Feedback, *Journal of the Acoustical Society of America*, Vol.22, Issue 6, pp.824-826 (1950).
17. Pfordresher, P. Q., and Bella, S. D., Delayed Auditory Feedback and Movement, *Journal of Experimental Psychology: Human Perception and Performance*, Vol.37, No.2, pp.566-579 (2011).
18. Pfordresher, P. Q., and Palmer, C., Effects of delayed auditory feedback on timing of music performance, *Psychological Research*, Vol.16, pp.71-79 (2002).
19. Cox, J., The minimum detectable delay of speech and music, *Proc. Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '84*, pp. 136-139 (1984).
20. Nishibori, Y., Tada, Y., and Sone, T., Study and Experiment of Recognition of the Delay in Musical Performance with Delay, *IPSJ SIG Notes*, Vol.2003-MUS-53, No.9, pp.37-42 (2003). (in Japanese)

Effects of random correlation on ANOVA and regression

Gergely Bencsik, Laszlo Bacsardi

University of West Hungary, Institute of Informatics and Economics, Sopron, Hungary
{bencsikg, bacsardi}@inf.nyme.hu

Abstract. The always increasing data volume can help to perform different analyses based on the available data using different analyzing methods. Since there are many outgoing research on different fields, analysis can be performed on bigger and bigger data sets and can be interpreted from different points of view. The entire process is controlled by the research methodology precisely. However, there are increasing number of contradictory results which follow the same methodology but interpret their results differently.

Our research focuses on how is possible to get different inconsistent results to a given question. The results are proofed by the methods mathematically and accepted by the experts, but we assume, that the decisions are even not valid, because the correlations originated from a random nature of the measured data. This random characteristics—named as random correlation—is unknown for the experts as well. But this phenomenon need to be handled to make correct decisions.

Keywords: decision support, random correlations, regression analyses

1 Introduction

The decision making processes are support by data analyses to predict the possible future effects. There are different mathematical, statistical and data mining techniques as well as experts, who accept and validate the results.

The standard research methodology is defined by many state-of-art publications [1,2]. Following these methodologies, different and contradictory results were born. In biology, the salt consumption is always generate opposite publications [4,5]. In sociology, there are arguments about data based analysis and since these results do not produce the real predictions, a new methodology was proposed [6]. Another example is based on questionnaires and scores related to internet addiction [7,8].

We assumed that these contradictory results could be originated from a phenomena named random correlations [RC]. This paper presents our work on RC and shows its effects on analysis of variance (ANOVA) and regression techniques.

2 Random correlation framework

The random correlation theory states that there can be connection between data rows randomly which could be misidentified as a real connection. The main idea is that data

rows as variables present the revealed, methodologically correct results, however, these variables are not connected really, and this unconnected property is hidden from the experts as well. Our framework contains three entities: (1) classes, (2) calculation processes and (3) parameters.

2.1 Three classes of random correlation

Class 1. The first class is related to the number of used methods. Different methods can be applied for a given problem set, and if we cannot find good results with one, then we choose another. This is a classical view for finding some kind of results. The number of chosen methods can raise with different input parameters range and seeking and removing outliers. On the other side, we do not [we cannot] define, when the data are not related to each other. When more and more methods are used with different circumstances, i.e., different parameters and error rates, then we could be not sure whether a real correlation was found or just a random one.

Class 2. The second class focuses on inconsistency and includes the cases when one or more methods produce opposite results. In the process of data analyses, we generally stop at the first method which we used to obtain the satisfying result and no further correlations are seek with further methods. Two or more methods give the same “correlated” result, which can be meant a real connection. The term “one or more methods” is used because even one method can be inconsistent with itself. For example, when it produces different types of results near given circumstances, i.e., sample size.

Class 3. The third group is based on number of data items. Data volume is growing fast nowadays. The classical approach is that if we have more data, we could create more precise results. However, it could lead to problem when a part of the data rows produce different results then the larger amount of that same data rows. For example, if we measure a data row from the start to the time t , and this part of the data row leads inconsistent result from another part which was measured from the start to time $t + k$. This is critical problem, because the time interval is not known for us. For this problem, the cross validation can be a solution. If all subsets of the data row for a given time period t are not perform the same result, only a random model could be found.

2.2 Computing and parameters

The core of the RC analyze is the calculation process. During this process, we cover the whole possibility space (Ω). For example, there is a questionnaire with 4 questions and each question can be answered with a number from 1 to 3. If 5 persons are asked, then the total possibility space is $3^{4*5} = 3,486 * 10^9$. We perform the analysis method in each case and notify the count of “correlated” and “non-correlated” subsets, from which rate R can be calculated as the quotient of the “correlated” frequency and $|\Omega|$. If this rate is not balanced or often changed near different parameter values, then the random correlation’s chance is high. For example, if R is very high, that means correlation can be found easily, because possibilities count on “correlated” judgment is high in Ω .

To describe all kind of data input structure, we need parameters, which are number of data columns (k) and number of data values (n). The third parameters is range (r),

which determines the possible values between a and b , where a and b are numbers. Because of r , Ω can be calculated, because we cannot produce it with infinite number line. However, the r is not a very strict condition because we measure some kind of value (> 0) but we cannot measure whatever (big) number in general. The number of methods (t) is necessary to define how many methods we use in the case of *class 1*. It follows, not all RC analyze session must contain all framework parameters.

To perform an analyses, we have defined the following steps:

1. Introducing the analyzed method's basic mathematical background;
2. Introducing which random correlation class contains the given case;
3. Define what is exactly understand under the method's random correlation;
4. Define and choose random correlation's parameters;
5. Calculations and proofing;
6. Validation with simulations.

3 Random correlation analyzes I: ANOVA

3.1 Applying the steps of our framework on ANOVA

The ANOVA is used to determine whether the groups' average are different or not. The null hypothesis H_0 state that the averages are equal statistically and the alternative hypothesis H_1 decline the equality statistically. ANOVA compare the outer, $SSB = \sum_{j=1}^k n_j * (\bar{x}_j - \bar{\bar{x}})^2$ and the inner, $SSW = \sum_{j=1}^k \sum_{i=1}^{n_j} (x_{i,j} - \bar{x}_j)^2$ variances of the data rows and that quotient are compared the critical F value at significance level α [Step 1].

As for the random correlations, ANOVA belongs into Class 2 and it has a specific place: both H_0 and H_1 can be meaningful [Step 2]. The "non-correlated" can be defined as the means are statistically similar [H_0], therefore the influencing variable has no effect on the subject. The "correlated" means that it has influence [H_1]. In this case, the random correlation means that the H_0 or the H_1 can took priority over against the other according to parameters defined later in Step 4. The seeking rate should be around 0.5 [Step 3]. It is allowed to get a little difference from that rate, but huge distortion is dangerous. The following parameters k , n and r are used in this case [Step 4]. To calculate all combination and perform ANOVA in each case [Step 5], a program was developed in .NET framework and written in C# language. Finally, we calculated the rates [Step 6].

ANOVA have three conditions: (1) sampling must be done randomly; (2) each sample size must follow the normal distribution and (3) the variances must be equaled statistically. We assumed that random sampling was passed. We used the D'Agostino-Pearson test for normality check and Bartlett test for variance test.

3.2 Reducing the total possibility space

By increasing k , n , and r , the Ω cannot be calculated in real time even with a fast computer. To make this calculation possible, the Finding Unique Sequences algorithm

(FUS) was developed to decrease the number of candidates. The FUS will be introduced focus on ANOVA, but it can be used in another cases as in Section 4.

The outer variance is the variance of group's means, but one group's mean is determined by data values, which are related to the inner variance. Therefore, we proceed from the inner variance: we have to calculate only one column total possibility space and repeat it k times. Note that is not enough to store the SSW only, because one SSW can belong to one m mean, but one mean can belong to several SSW . If we have $(i)[1, 2, 2]$ and $(ii)[1, 1, 3]$, the $SSW_{(i)} = 0.6666$ and $SSW_{(ii)} = 2.6666$, while the means are the same in both cases (1.6666). This leads to different F values. Therefore, we must store $mean - SSW$ tuples. This is the first level of decreasing.

At the second reducing level, we must be calculated only that combinations, which have different SSW . These possibilities can be achieved with repeated permutation technique. Besides, we need to store the frequency for each $mean - SSW$. The frequency of one tuple can be calculated with the repeated permutation

$$\frac{n!}{s_1! * s_2! * \dots * s_i!}, \quad (1)$$

where n is the number of elements, s_i is the number of repetitions. We produce all repeated combinations for one group, then we calculate each combination's $mean$, SSW and $frequency$. Based on these triples, we can calculate SSB , and F at the end. The calculated F frequency can be derived as

$$F_{i,k,n,a,b} = (\prod_i^k C(SSW_i)) * C(m_i), \quad (2)$$

where $C(SSW_i)$ is the count of the SSW frequency and $C(m_i)$ is the count of the given means combination. A given F is compared with the $F_{critical}$, and since the frequency of that comparison is known, we can define how many times this judgment occurs. At the end, R can be calculated.

3.3 Effects of random correlation on ANOVA

The first results are summarized in Table 1.

Table 1. ANOVA results with $r(1, 3)$

r(1,3)	k = 3	k = 4	k = 5	k = 6	k = 7	k = 8
n = 4	1	0.9991	0.9997	0.9999	0.9999	0.9999
n = 5	0.9877	0.9923	0.9946	0.9957	0.9968	0.9976
n = 6	0.9744	0.9788	0.9818	0.9847	0.9864	0.9876
n = 7	0.9699	0.9741	0.9768	0.9797	0.9817	0.9831
n = 8	0.9630	0.9661	0.9685	0.9707	0.9721	0.9736
n = 9	0.9563	0.9594	0.9611	0.9621	0.9635	0.9645
n = 10	0.9656	0.9695	0.9722	0.9741	0.9756	0.9768

The parameter values are changed in the further calculations. The results with different r , k , and n are shown in Table 2.

Table 2. ANOVA results with different r, k, n

r	k	n	R
(1,5)	3	4	0.9758
(1,5)	5	6	0.9594
(1,10)	3	4	0.9547

The rates are relatively high in the favor of H_0 . We can also discover that there are no trivial patterns in the tables. By increasing the k near to the given n and r , there are no trivial increasing or decreasing directions [horizontal case]. This is also true if n is increased [vertical case]. However, we showed a few example in Table 2, the $r(1,5)$ are calculated with the same internal of k and n as in the case of $r(1,3)$, and the same conclusion can be observed. Despite of FUS, we cannot even calculate all cases, since a large enough k, n [about 1000] and r [about 1-10 or 1-20] will result in a huge count of possibilities. This problem is already showed up in the case of $r(1,10)$ with increasing k and n . However, the high rate of H_0 is a big distortion.

4 Random correlation analyzes II: regression

In this section, we used our framework to analyze regression methods. Regressions are in the first class of random correlation. If we perform more and more regression techniques, then the chance of finding a correlation will be increased. There are many different regression methods, so we focused on linear, quadratic, exponential and logarithmic techniques. Therefore, $t = 4$. The k can be skipped, because we always have two columns [x and y coordinates]. We have two parameter r : r_1 determines the range of x values [$r_1(a_1, b_1)$], while r_2 stands for range of the y values [$r_2(a_2, b_2)$].

The earlier described FUS can be used only partly: the first level of reducing cannot be applied since the order of the coordinates is important. For example, the $x' = \{2, 1, 2\}$ and $y' = \{1, 3, 1\}$ do not give the same r^2 as $x = \{1, 2, 2\}$ and $y = \{1, 1, 3\}$. Therefore, all possibilities needs directly produced in the first phase. However, the second level of reducing can be used without any modification.

In each case, we perform all 4 types and seek the best fitting line or curve. If we found high r^2 with either of them, then we increase the count of “correlated” class. The acceptance level can be changed, we defined as $r^2 > 0.7$. Also in regression case, the simulation deal with applying conditions. We increased n from 5 to 10, changed r_1 and r_2 , used $t = 4$. The results are summarized in Table 3.

Table 3. Rates of r^2

$t = 4$	$r_1(1,5);r_2(1,3)$	$r_1(1,10);r_2(1,3)$	$r_1(1,3);r_2(1,5)$	$r_1(1,3);r_2(1,10)$
$n = 5$	0.2873	0.3071	0.3122	0.3288
$n = 6$	0.2092	0.2161	0.2530	0.3239
$n = 7$	0.1387	0.1379	0.2204	0.3102
$n = 8$	0.1142	0.1027	0.1947	0.3029
$n = 9$	0.1057	0.0796	0.1894	0.2927

We can conclude that the $r_1(1,3);r_2(1,10)$ case is very stable around 0.3. This is a significant results, since the “correlated” and “non-correlated” judgments have the same chance in each cases. On the other side, it is rightful assumption, that the theoretically rate cannot be around 0.5 in regression case, because 0.5 would mean that the “correlated” judgment is not more, like a simple coin fifty-fifty rate. To say “correlated”, the rate must be stricter. Therefore, the parameters related to rate 0.3 could also be suitable.

5 Conclusion

A new random correlation theory was presented in this paper. We introduced the classes of random correlations and defined its parameters. Getting mathematically proven correlations between data rows as variables do not mean, that these data are connected in real life since the analysis methods consider the data as numbers.

This work is not against the real existing correlations since we do not assume that real correlations are not exist. We just would like to avoid making false decisions based on false results because of this random characteristics. Scientists follow the research methodology, use the methods with conditions and perform transformation, filtering and outlier analysis if needed and they do not know that the received result could be just random, because the circumstances of random correlations are unknown for them.

Acknowledgement

This work was partially supported by the TAMOP-4.2.2.C-11/1/KONV-2012-0015 (Earth-system) project sponsored by the EU and European Social Foundation.

References

1. J. A. Khan: Research methodology. APH Publishing Corporation, New Delphi (2008), ISBN: 9788131301364
2. Petter Lake, Haakon Breien Benestad, Bjorn Reino Olsen: Research Methodology in the Medical and Biological Sciences. Academic Press, London (2007), ISBN: 9780123738745
3. Lee Hooper, Christopher Bartlett, George Davey Smith, Shah Ebrahim: Systematic review of long term effects of advice to reduce dietary salt in adults. British Medical Journal 325, 628–632 (2002), ISSN: 0959-8138
4. Steve Pljesa: The impact of Hypertension in Progression of Chronic Renal Failure. Bantao Journal, 1, 71-75 (2003), ISSN: 1312-2517
5. Mike Savage, Roger Burrows: The Coming Crisis of Empirical Sociology. Sociology, 41, 885-899 (2007)
6. Lawrence T. Lam, Zi-Wen Peng: Effect of Pathological Use of the Internet on Adolescent Mental Health. Archives of Pediatrics and Adolescent Medicine, 164, 164-174 (2010)
7. Cai-Xia Shen, Ru-De Liu, Dan Wang: Why are children attracted to the Internet? The role of need satisfaction perceived online and perceived in daily real life. Computers in Human Behavior, 29, 185–192 (2013)

A Proposed Clinical Prediction Rule Register Supporting Primary Care Decision Making

Atieh Zarabzadeh, Ronan McDonnell, Derek Corrigan, Rose Galvin, Susan M Smith, Tom Fahey

HRB Centre for Primary Care Research, Department of General Practice, Royal College of Surgeons in Ireland (RCSI), 123 St. Stephens Green, Dublin 2, Ireland
{atieh.zarabzadeh, rmcdonnell, derekcorrigan, rosegalvin, susansmith, tomfahey}@rcsi.ie

Abstract. Clinical Prediction Rules (CPRs) are developed and used as a means of leveraging evidence-based medicine and may be implemented as part of clinical decision support systems. In this paper functional and non-functional requirements of an electronic register for CPRs (eCPR) are described. Non-Functional requirements such as interoperability, security, integration, and consistency and functional requirements such as searching, maintenance of CPR records, computational functions, and data provenance are included in this paper. An eCPR architecture is proposed for use in the primary care setting. Subsequently the proposed eCPR is evaluated against the defined requirements.

Keywords: Clinical Prediction Rules (CPRs), Clinical Decision Support Systems (CDSSs), primary care

1 Introduction

Clinical Prediction Rules (CPRs) are tools that serve to assist clinicians to quantify the contribution of multiple predictors from a patient's history, physical examination or laboratory results to inform a diagnosis, prognosis or treatment response [1]. They offer clinicians one way of implementing evidence-based medicine, especially if incorporated into Clinical Decision Support Systems (CDSS) at the point of patient care [2]. CPRs should pass through three stages of development prior to clinical use: 1) derivation: predictor variables are identified to develop the CPR; 2) validation: this can be sub-divided into two stages – narrow validation, where the rule is tested for reliability and accuracy in a similar patient cohort or clinical setting and broad validation, where the rule is tested in a different patient cohort or clinical setting; and 3) impact analysis: the value of the rule may be tested in terms of patient outcomes, clinician behaviour or costs [3][4]. The majority of CPRs have been published without further validation, with relatively few undergoing impact analysis.

We have recently developed an international web based register of CPRs (eCPR) for use in primary care. The purpose of this register is to act as a resource for researchers and clinicians, where CPRs that are available at the point-of-care and rele-

vant to clinical practice are summarised and categorised in terms of derivation, validation and impact analysis. We have included 434 unique CPRs in the register, of which 238 (54.8%) have been validated at least once and 12 (2.8%) have undergone impact analysis [5]. Previous attempts at developing a register for CPRs include the integrated CPR (iCPR) for: 1) validated streptococcal pharyngitis, and 2) bacterial CPRs [6]. While iCPR is limited to two CPRs, the eCPR serves as a comprehensive repository for available CPRs. The outcome of the iCPR revealed that using iCPR, patients are significantly less likely to have antibiotic prescribed and significantly less likely to have rapid streptococcal tests ordered for them [7].

In this paper, the methods used to design and develop an electronic platform that contains the eCPR register are described.

2 eCPR Register Requirements

The non-functional and functional requirements for the eCPR register are the subject of this section. Non-functional requirements refer to the general system-level requirements whereas functional requirements refer to the functions and behaviours of the system.

2.1 Non-functional Requirements

In addition to the general non-functional requirements such as information quality and ease of use, additional non-functional requirements that must be considered when designing and implementing an eCPR register include the following:

Interoperability. System interoperability and semantic interoperability are two facets of interoperability. System interoperability refers to the ability to transfer easily understood and widely agreed data structures between different systems e.g. using standards such as Extensible Markup Language (XML) or Health Level-7 (HL7). Semantic interoperability goes beyond the ability of sharing data structures, elements and values [8]. It refers to the use of clinical terminologies or classifications to link predefined codes to data that associate it with defined clinical meanings so as there is no clinical ambiguity in relation to the data e.g. Systematized Nomenclature of Medicine (SNOMED), International Classification of Diseases (ICD10) and International Classification of Primary Care Second edition (ICPC2). Interoperability is required to facilitate sharing machine computable logic, knowledge discovery, and data federation between eCPR and other Electronic Health Record (EHR) systems and CDSS.

Security. Security, confidentiality and integrity of patient information are the vital requirements of any healthcare IT solution [9] including eCPR. Access authorisation control and other measures [9] should be put in place to ensure that eCPR is only accessible to authorised users. Secure network protocols and data encryption methods

to communicate with other EHR and CDSS is a requirement for eCPR to facilitate the highest level of security when exchanging patient information.

Integration. Integration with existing practices and workflow involves fusion of the new system to the existing processes without any major or significant perceived changes to the existing workflow. Seamless integration of eCPR with other EHR or CDSS already in use in primary care is a fundamental requirement if it is to be widely adopted and used.

Consistency. Consistency of data, data terminology and data representation methods should be coherent between eCPR and existing setting preferences. eCPR must adhere to the set of predefined guidelines and best practices.

2.2 Functional Requirements

Further to the general administration activities such as registering users, the functional requirements associated with the eCPR register include searching, maintenance of CPR records, computational functions, and data provenance.

Searching. Searching eCPR for the most relevant and recent CPR is the core functional requirement of this register. Users should be able to search by clinical domain, symptoms or a CPR name. The results should be sorted depending on the relevancy of the CPR to the searched terms and the users setting i.e. a CPR developed based on the population in a geographical jurisdiction may only be relevant to those located in that particular geographical jurisdiction.

Maintenance: Maintenance of eCPR involves updating existing CPR records when necessary and adding recently published CPRs. Authorised users should be able to update an existing CPR record on the eCPR system with any additional information e.g. CPR being validated in a different population. Upon the publication of a CPR, this CPR must be added to the eCPR through the system interface.

Computational function. A crucial eCPR function is its computational function by which a CPR score associated to the selected signs and symptoms is calculated. Based on the CPR algorithm, decision suggestions may be available.

Data provenance. In the context of eCPR, data provenance involves recording the input/output data, processes and events occurred on the system. The recorded data may be used to: 1) revert back any inappropriate or accidental changes made to the eCPR register, and 2) improving the system functionality by reviewing the trends of events and processes e.g. particular users viewing only a set of CPRs.

3 eCPR Register Architecture

Based on the requirements outlined in the previous section, an eCPR register has been designed and developed in the HRB Centre for Primary Care Research [10]. This eCPR consists of the following components as shown in Fig. 1:

- The core database that houses the CPR data manually extracted from existing literature [11]. This database is developed using MySQL and houses data on various components of a CPR such as paper metadata, the prediction variables, risk variables, decision criteria and other variables associated with a CPR.
- The query component acts as the communication channel between the engine, Application Program Interface (API) and computational components. The query component allows for search, create, read, update, and delete functions on the database depending on whether private or public API is making the call. This component is developed in PHP and incorporates MySQL to communicate with the database.
- The interface component is responsible for the representation of data to the users, i.e. User Interface (UI). The interface component communicates with the private API. The UI is developed in PHP, JQuery, HTML and CSS.
- Private and/or public API determines the functions, data and interface that should be available to various users depending on their roles. This two layered approach to API enhances the security of the system.
- Engine determines the appropriate CPR for the user based on basic data about the user e.g. location, setting, population characteristics etc. Java rule engine is deployed to serve as the engine component. Pattern matching algorithms are used as the basis of the engine component.
- Computational component is responsible for the score calculations and predictions. Score calculations and predictions are based on the algorithms suggested by each CPR.

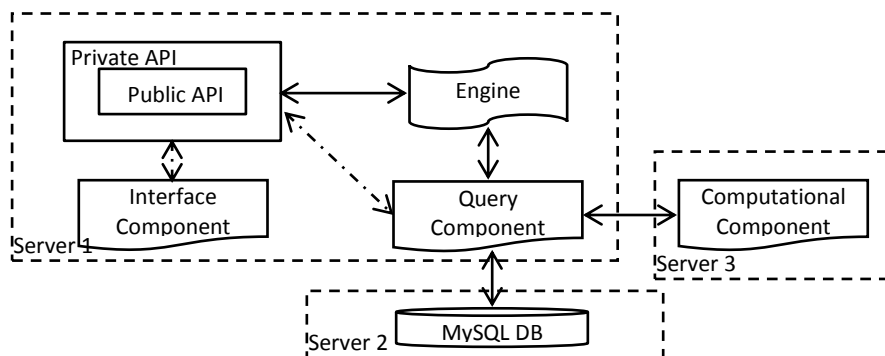


Fig. 1. eCPR register architecture

The query component, two APIs, interface component and engine run on one server (server 1), while the database and computational components run on two other separate servers (servers 2 and 3, respectively). This architecture allows for additional security, reliability, maintenance and speed of calculations. Server 1 facilitates APIs,

query and interface components. Server 2 only has the MySQL database running on it to ensure the separation of data and processors. The computational component is on a dedicated server to score computations without undermining other services by the eCPR register online application. With separation of components to three servers, the eCPR register is available for searching even if server 3 is not operational.

4 An eCPR Register Specifications

In response to the requirements defined, the eCPR register is developed based on the architecture described in the previous section. Table 1 demonstrates how the eCPR register has addressed each of the non-functional requirements. For each functional requirement outlined, a dedicated eCPR section is developed e.g. CPR score calculator section for computational function.

Table 1. Requirements and their associated eCPR features

Requirement	Associated eCPR Feature
Interoperability	eCPR facilitates data sharing using its private and public APIs in XML format to serve its system interoperability. For the purposes of semantic interoperability, the terminology used throughout the system is identical to that used by the original authors of the CPRs to prevent ambiguity along with their equivalent ICD10 and ICPC2 codes.
Security	Various components of the eCPR register are located on three different servers to ensure security of data. Although no patient information will be stored on the eCPR database, the two layered approach to APIs allows for additional security of the CPR data. User authentication measures are also put in place.
Integration	eCPR as an additional system to the existing practice is seamlessly integrated with the current workflow. This is done via its two APIs. At the primary care settings level, only private API is accessible. And outside the primary care settings only the public API is accessible.
Consistency	Using a Content Management Framework (CMF) the data presentation is consistent across the functional requirements. Also, using terminologies that the CPR authors use along with the list of various alternative terms for each CPR, it is anticipated that the data in the eCPR is consistent.

5 Future Work

Design and development of an eCPR system and how non-functional and functional requirements are addressed by its features are demonstrated in this paper. This eCPR system facilitates searching available CPRs and CPR score calculation. Previous attempts at developing an electronic repository for CPRs include iCPR that is limited to only two CPRs.

Further work is deemed necessary to evaluate the system and validate its usefulness and impact in practice. Evaluating the system involves assessing its functionality matching the user requirements for various use case scenarios and in various settings.

Validating its usefulness, impact and success involves a number of significant research projects dedicated to these.

Acknowledgments

This work was supported by the Health Research Board (HRB) of Ireland through the HRB Centre for Primary Care Research under Grant HRC/2007/1.

References

1. T. G. McGinn, G. H. Guyatt, P. C. Wyer, C. D. Naylor, I. G. Stiell, and W. S. Richardson, "Users' guides to the medical literature: Xxii: how to use articles about clinical decision rules," *JAMA*, vol. 284, no. 1, pp. 79–84, 2000.
2. D. Corrigan, L. Hederman, H. Khan, A. Taweel, O. Kostopoulou, and B. Delaney, "An ontology-driven approach to clinical evidence modelling implementing clinical prediction rules," in *E-Health Technologies and Improving Patient Safety: Exploring Organizational Factors, 2012*, IGI Global, 2013, pp. 257–284.
3. C. Keogh, E. Wallace, K. K. O'Brien, P. J. Murphy, C. Teljeur, B. McGrath, S. M. Smith, N. Doherty, B. D. Dimitrov, and T. Fahey, "Optimized retrieval of primary care clinical prediction rules from MEDLINE to establish a Web-based register," *J. Clin. Epidemiol.*, vol. 64, no. 8, pp. 848–60, Aug. 2011.
4. B. M. Reilly and A. T. Evans, "Translating Clinical Research into Clinical Practice: Impact of Using Prediction Rules To Make Decisions," *Ann. Intern. Med.*, vol. 144, no. 3, pp. 201–209, 2006.
5. C. Keogh, E. Wallace, K. K. O'Brien, R. Galvin, S. M. Smith, C. Lewis, A. Cummins, G. Cousins, B. D. Dimitrov, and T. Fahey, "Developing an International Register of Clinical Prediction Rules for Use in Primary Care: A Descriptive Analysis," *Ann. Fam. Med.*, vol. 12, no. 4, pp. 359–366, Jul. 2014.
6. D. Mann, J. Kannry, and D. Edonyabo, "Rationale, design, and implementation protocol of an electronic health record integrated clinical prediction rule (iCPR) randomized trial in primary care," *Implement Sci.*, vol. 6, no. 109, 2011.
7. T. G. McGinn, L. McCullagh, J. Kannry, M. Knaus, A. Sofianou, J. P. Wisnivesky, and D. M. Mann, "Efficacy of an evidence-based clinical decision support in primary care practices: A randomized clinical trial," *JAMA Intern. Med.*, vol. 173, no. 17, pp. 1584–1591, Sep. 2013.
8. D. Kalra and B. Blobel, "Semantic Interoperability of EHR Systems," in *Medical and Care Compunetics*, vol. 4, L. Bos and B. Blobel, Eds. 2007, pp. 231–245.
9. E. Smith and J. H. . Eloff, "Security in health-care information systems—current trends," *Int. J. Med. Inform.*, vol. 54, no. 1, pp. 39–54, Apr. 1999.
10. "HRB Centre for Primary Care Research." [Online]. Available: <http://hrbcentreprimarycare.ie/>. [Accessed: 29-Jul-2014].
11. C. Keogh, E. Wallace, K. K. O'Brien, P. J. Murphy, C. Teljeur, B. McGrath, S. M. Smith, N. Doherty, B. D. Dimitrov, and T. P. Fahey, "Optimized retrieval of primary care clinical prediction rules from MEDLINE to establish a Web-based register," *J Clin Epidemiol*, vol. 64, pp. 848–860, 2011.

Graph Visualization Performed by nVidia CUDA Platform

Ondrej Klapka, Antonin Slaby

University of Hradec Kralove, Faculty of Informatics and Management, Department of Informatics and Quantitative Methods, Hradec Kralove, the Czech Republic

{Ondrej.Klapka, Antotnin.Slaby}@uhk.cz

Abstract: Graph is way of structured data representation that is used for solving of many today's practical problems, e.g. bioinformatics, data mining or social networks. Big graphs, however, require high computing power, which current CPUs can't provide. But GPU is a high performance, extreme optimized, low cost platform. The article deals with Fruchterman-Reingold graph; how its layout algorithm can be parallelized for GPU using nVidia CUDA computing model. A comparison of execution time on CPU and GPU for processing of large randomly generated graphs is presented.

Keywords: CUDA • GPU computing • Graph layout • Parallelization

1 Introduction

In recent decades, importance of information and consequently the amount of information that is collected and processed has been dramatically increasing. One possibility to represent the information in a structured form suitable for further processing is expressing the information in the form of a graph. Graphs can be used in many application fields such as data mining, bioinformatics, social networks, etc. [1]

The amount of information in some applications raises the need to handle very large graphs, sometimes having hundreds of thousands to millions of nodes. Increasing size of graphs then requires more and more computing power for their processing. Computing power of current processors becomes insufficient for very large graphs and processing them by supercomputers is often not possible or is very expensive.

As a cost-effective solution to the above mentioned problem seems parallelization of calculations and utilization of the graphics processing unit (GPU) for calculations. GPU can process a large number of parallel tasks (called threads) simultaneously. [2]

The following text provides a brief description of the computing platform CUDA based on nVidia graphics chips and also presents a comparison of processing speed tested on randomly generated graphs of various sizes.

2 The Problem of Graph Visualization

Graphs are usually represented for processing by computers in some way enabling to execute this process easily. The most widely used ways include incidence matrix, adjacency matrix, edge list, a list of neighboring nodes, etc. [3] These forms of representation provide a structured and precise description of the graph, but at the same time these representations are confusing and therefore inappropriate to human perception.

On the other hand the best and the most intuitive way to express the graph for human is a graphic representation of nodes and edges. The graphical representation however is useful only if it meets certain basic aesthetic and cognitive criteria. There are not generally formulated and generally accepted criteria saying how a properly represented graph should look like. It is clear, however, that the graph with plenty of mutual edge crossings is not easily readable for people.

There can be found a number of ideas and linked algorithms to distribute graph nodes. Generally, the largest and most widely used group of algorithms for visualization graphs is called force-directed algorithms group. Force-directed algorithms are based on the idea of conception of the graph as a mechanical system in which nodes and edges are subjected to two kinds of forces: The nodes mutually repel each other and at the same time edges cause attraction of nodes connected to each other. [4]

Force-directed algorithms have quadratic computational complexity which arises from the need to calculate the repulsive forces interacting between all (every two) nodes. [5] Quadratic complexity can be reduced to the logarithmic complexity by different simplification heuristics (usually branch and bound based). [4, 5, 6]

Quadratic complexity of algorithms is generally acceptable for small graphs, comprising hundreds to thousands of nodes. For very large graphs it causes noticeable slowdown of the algorithm. Consequently even today's most powerful processors can not draw very large graphs in real time. CUDA platform is capable to provide high performance GPU computing and could enable real-time draw of graphs having hundreds to thousands time more nodes than those current processors are able to elaborate.

3 Description of the Computing Platform

Implementation of the non-graphic computation on the GPU enabled the arrival of Shader Model 4.0 and the birth of a unified architecture of separate GPU units in 2006. Since that time a very specific hardware, designed and optimized for applications and calculations in graphics, enables running not only graphics algorithms. GPU has become in such a way cost-effective, computationally very efficient architecture. [1] Modern GPU is a high-performance, parallel and programmable architecture. However, GPU programming, originally intended solely for graphic calculations, is unlike the programming for the processor very limited. [1, 7] Consequently effective use of this specific architecture requires at least a basic understanding of its principles.

3.1 CUDA Model

As to hardware, the GPU is a set stream multiprocessors (SM) and each multiprocessor consists of several processors. Each SM has a shared memory that is available for all processors in SM. Each processor executes during one cycle the same instruction but this instruction is provided over specific data that may differ. Communication between multiprocessors is organized through global device memory that is shared for all multiprocessors of the GPU. [1]

CUDA computing model is organized as a large number of parallel threads. The code that each thread executes is called a kernel function, and this code is the same for all threads. The threads are triggered off in groups, called blocks and the blocks are grouped in the grid. The thread and block scheduling on multiprocessors is supervised by GPU planner and so the programmer is completely shielded from the issue of planning. [1, 8]

Identification of individual threads in the program code of kernel function is enabled through the built-in variables. Thanks to the easy identification of individual threads the programmer can easily ensure that each thread elaborates different set of data even though all the threads execute the same code.

4 Fruchterman-Reingold algorithm using nVidia CUDA

In this contribution Fruchterman-Reingold algorithm from the family of force-directed algorithms was chosen for demonstration of work with nVidia CUDA for visualization of graphs. The main advantage of Fruchterman-Reingold algorithm is in its speed, achieved by using a very simplistic model of interacting forces in the graph. [10]

To make full use of the advantage of great potential of GPU computing it is necessary to transform the code which is traditionally executed sequentially on the CPU so that it could enable parallel execution.

The traditional code of one iteration of Fruchterman-Reingold algorithm for CPU is as follows: [9, 10]

```
For each pair of nodes v1 and v2 of the graph repeat:
    {d is a vector of distance between nodes}
    d := v1.position - v2.position;
    f_rep := - k * k / d; {repulsive force}

    {each node has two vectors: .position and .displacement}
    {Vector .displacement defines the distance by which the}
    {node will be transformed in given iteration}
    v1.displacement := v1.displacement + d/d * f_rep;

If nodes v1 and v2 are connected by the edge do:
    f_attr := d * d / k; {attractive force}
    v1.displacement := v1.displacement - d/d * f_attr;
```

```

    v2.displacement := v2.displacement - d/|d| * f_attr;
end
end

{limitation of displacement for the node v}
For each node v of the graph do:
    limitation := min(max_displacement, |v.displacement|);
    v.position := v.position + v.displacement /
|v.displacement| * limitation;
end

```

It is noticeable that the code that is in that version of the algorithm is repeated for each pair of nodes. Consequently it must necessarily consist of two cycles. In case of a parallel version of the code, it is possible to remove one cycle as it can be replaced by a large amount of threads. The modified algorithm will calculate for each thread forces acting between the nodes denoted by the index of the current thread and leading all the other nodes. Consequently the parallelized code of one iteration Fruchterman-Reingold algorithm is as follows:

```

{unique identification of the thread}
thread_id = (blockIdx * blockDim) + threadIdx;
{each thread will read one node according to its ID in
such a way one cycle in the algorithm is removed}
v1 = nodes_list[thread_id];

For all other nodes in the thread v2 from the array
nodes_list do
    d := v1.position - v2.position;
    f_rep := - k * k / d; {repulsive force}
    v1.displacement := v1. displacement + d/|d| * f_rep;

    If nodes v1 and v2 are connected by the edge do:
        f_attr := d * d / k; {attractive force}
        v1. displacement:= v1. displacement - d/|d| * f_attr;
        v2. displacement:= v2. displacement - d/|d| * f_attr;
    end

    limitation:= min(max_displacement, |v.displacement|);
    v1.position := v1.position + v. displacement / |v. displacement| * limitation;
end

```

5 Results

Layout of position of graph nodes was set by Fruchterman-Reingold algorithm. The algorithm was tested on CPU and GPU for different sizes of graphs and different density of edges in the graph. The comparison was carried out on CPU Intel Core 2

Duo T8100, running at 2100 Mhz and GPU nVidia GeForce 9300M G with a G86M core processor running at 400 Mhz.

Time needed for execution of 20 iterations for a random graph with an average node degree 5 and maximum node degree 20 was measured. The results of measurement are shown in the following Table 1.

Table 1. Time required for execution of the 20 iterations by the Fruchterman-Reingold algorithm of a randomly generated graph with different number of nodes.

Number of nodes	Number of edges	Execution time on CPU (ms)	Execution time on GPU (ms)
10	50	2	35
100	500	10	169
1 000	5 000	489	2 368
10 000	50 000	52 434	29 980
100 000	500 000	4 368 280	2 032 005

Fig. 1 shows a graph which expresses the dependence of execution time on the size of the graph. It can be seen, that the difference in execution time is significantly shorter by the GPU for graphs with more than 10 000 nodes. For this and bigger size of graph the calculation performed by the GPU is already several times faster than by the CPU. For small graphs, having tens to hundreds of nodes is but faster CPU. This is due to the fact that the calculation cannot be run in such a small amount of threads and performance GPU is not fully exploited.

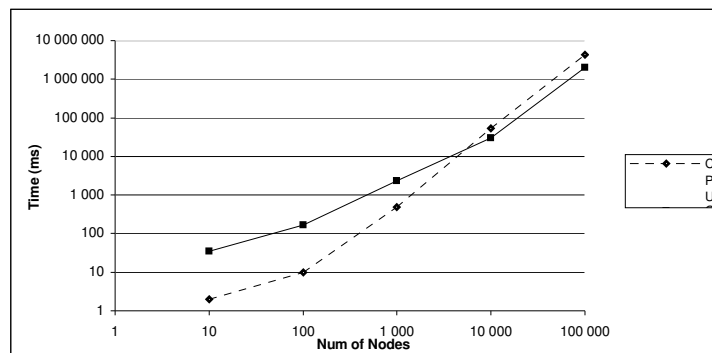


Fig. 1. The graph showing the dependence of calculation time on the number of nodes and edges of the graph. (Both the X axis and Y axis have logarithmic scale)

Second investigation was established to determine how the speed of the algorithm running on the CPU and GPU reflects the density of edges in the graph. Calculation and measurements were carried out for a graph with 5000 nodes, where in previous measurements had been found very similar processing times.

Graph, created on the basis of results is shown in Fig. 2. It can be deduced from Fig. 2 that increasing density of edges in the graph results into the slowing of execution on the GPU, while the CPU execution time still fluctuates around 10,000 ms. GPU is therefore more powerful than the CPU only for sparse graphs. For the graphs with a large number of edges is but faster CPU, which is mainly due to the fact that the logic for verifying whether the two nodes are connected by the edge may cause mutual blocking of some threads on GPU. Increasing density of edges in the graph then leads to more frequent blocking of the threads and subsequently to the fall of the performance of GPU.

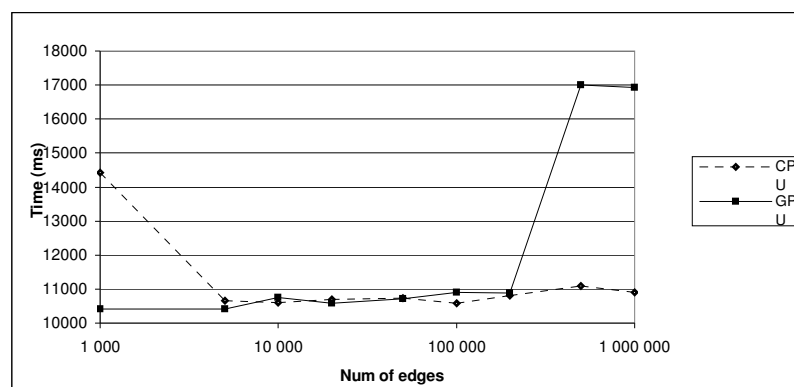


Fig. 2. Graph showing the dependence of the execution time for various densities of edges in the graph.

6 Conclusion

The article is the contribution to the problem of visualization of very large graphs. Parallelization of Fruchterman-Reingold algorithm code for automatic graph layout was largely tested. Testing was focused on comparing the differences in processing performance (execution time) of selected graphs by CPU and GPU.

Testing has shown that small size graphs having up to several hundreds of nodes can be processed faster by CPU than by GPU. Furthermore, it was verified that the GPU is due to the huge amount of running threads and high optimization for simple arithmetic operations capable for layout nodes of very large graphs faster than the CPU. For graphs having hundreds of thousands or more nodes can then be processing of the graph even several times faster. It was also found that the difference between the CPU and GPU in performance is especially noticeable for sparse graphs, with a large number of nodes and a reasonably small number of edges.

Acknowledgement

This work was supported by the project No. CZ.1.07/2.2.00/28.0327 Innovation and support of doctoral study program (INDOP), financed from EU and Czech Republic funds.

References

1. Pawan Harish, P. J. Narayanan: Accelerating large graph algorithms on the GPU using CUDA. Center for Visual Information Technology, International Institute of Information Technology Hyderabad, INDIA, <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.102.4206&rep=rep1&type=pdf>
2. John L. Hennessy, David A. Patterson: Computer Architecture: A Quantitative Approach. Morgan Kaufmann Publishers (2011)
3. Robert Vajdik: Reprezentace Grafu. Technická Univerzita Ostrava, Fakulta elektrotechniky a informatiky, Katedra informatiky, Ostrava (2009), http://home1.vsb.cz/~vaj049/AlgoritmyII/reprezentace_grafu.pdf
4. Yifan Hu: Efficient and High Quality Force-Directed Graph Drawing. Wolfram Research Inc, USA, http://yifanhu.net/PUB/graph_draw_small.pdf
5. Stephen G. Kobourov: Force-Directed Drawing Algorithms. University of Arizona, <http://cs.brown.edu/~rt/gdhandbook/chapters/force-directed.pdf>
6. Laurens van der Maaten: Barnes-Hut-SNE. Pattern Recognition and Bioinformatics Group, Delft University of Technology, Netherlands (2013), <http://arxiv.org/pdf/1301.3342v2.pdf>
7. Frank Dehne, Kumanan Yogaratnam: Exploring the Limits of GPUs With Parallel Graph Algorithms. School of Computer Science, Carleton University, Ottawa (2010)
8. Rafia Inam. An Introduction to GPGPU Programming CUDA Architecture. Mälardalen Real-Time Research Centre <http://www.diva-portal.org/smash/get/diva2:447977/FULLTEXT01.pdf>
9. Ondrej Klapka. Vizualni analyza dat: Vizualni analyza vlastnosti a vztahu dat. Hradec Králove: Univerzita Hradec Králove, Fakulta informatiky a managementu, Katedra informatiky a kvantitativních metod (2013).
10. Thomas M. J. Fruchterman, Edward M. Reingold. Graph Drawing by Force-directed Placement. University of Illinois, Department of Computer Science (1991), http://pdf.aminer.org/001/074/051/graph_drawing_by_force_directed_placement.pdf

Quantitative Evaluation of the Coffee Service Quality and Customer Value by Mining Twitter

Shu Takahashi¹, Ayumu Sugiyama², and Youji Kohda¹

¹ School of Knowledge Science,
Japan Advanced Institute of Science and Technology, Japan
{[takahashi.shu](mailto:takahashi.shu@jaist.ac.jp), [kohda](mailto:kohda@jaist.ac.jp)}@jaist.ac.jp

² Faculty of Humanities, Yamanashi Eiwa College, Japan
ayumu@yamanashi-eiwa.ac.jp

Abstract. Difference against the background of high technology is called commoditization. In this work, we focus on Japanese cafe service as a subject, and analyze the impressions of customers about the search, experience and credence qualities. The results of the analysis is suggested that the definition of the index for quantitative evaluation of search, experience and credence qualities, so the improvements of commoditization that leverage social data would be possible.

Keywords: Customer value, Natural language processing, Service value, Twitter

1 Introduction

In recent years, big data utilization in social network data (e.g. Facebook, Twitter, Flickr) has attracted much attention from researchers and research divisions of companies to analyze the customers market with high “objectivity” and high “completeness”[1][3][5]. In this study, we focus on the Twitter data from the viewpoint of customer’s value of products and services in the commoditization of service. Especially, we chose coffee service in Japan as an example of commoditization.

Specifically, demand for coffee beans has been expanding, and the methods of providing coffee diversify in the world coffee markets and service. In the decade from 2000, coffee bean consumption in importing countries has increased more than 8%. In addition, coffee beans consumption in exporting countries has increased by 48% or more³.

With the advent of convenience store coffee, competition in the coffee market has intensified in Japan. Total number of sales of convenience store coffee extends to 500 million cups in less than one year from the start of sales in 2012. Coffee consumption in Japan was 440 thousand tons in 2013[2]. This number increased 4.3% from the previous year by the hit of a convenience store coffee. Therefore, price competition occurs between Seattle-type coffee and convenience

³ International Coffee Organization - Historical data

stores coffee, it is a sign of commoditization. In addition, a new type of coffee shop called Third-wave appeared around the west coast of North America, and is spreading all over the world.

Controlling commoditization in coffee service is important for coffee market not only Japan, but also emerging countries. In this study, we explore the possibility to control the commoditization in coffee service. Therefore, we analyze the values of the customer against the Seattle-type coffee and convenience store coffee, to clarify the needs of the customer for coffee service.

2 Problem Statement

The type of coffee service is classified into three fields. The first one is convenient style coffee such kind of coffee has the authentic flavor and inexpensive price provided by convenience stores (CVS) and fast food shops (e.g. Mac cafe by McDonald's). The second type is a cafe service providing the relaxing place. Many number of Seattle-type coffee present such kind of service (e.g. Tully's Coffee, Starbucks Coffee). Among these, by providing the third place[6] nor at work or at home, the large chain store Starbucks have succeeded in to satisfy the needs of customers. The third field is Third-wave coffee. Third-wave coffee is only using single origin coffee beans that were harvested from seedlings of a single species, and made by drip.

In Japan, price competition occurs between Seattle-type coffee and convenience stores coffee. In this study, we analyze by three points "taste", "price" and "place", the values of the customers to the CVS and Starbucks. Using the tweets of customers obtained from the Twitter Streaming API data, to be analyzed by the following procedure.

1. Extract the Japanese tweets that are related to coffee service using the Twitter Streaming API.
2. Generate dictionary about "taste", "price" and "place" by the Japanese tweets extracted.
3. Map a feature space of "taste", "price" and "place" about CVS and Starbucks using the generated dictionary.
4. Adapt of the statistical methods to analyze tweet vectors in the feature space.

3 Data Collection and Feature Extraction

In order to extract the sense of values from the tweets of customers, we created a dictionary about sense of values (value dictionary) related to "taste", "price" and "place". Data for creating value dictionary is intended for 118,855 Japanese tweets including the words coffee and cafe in both English and Japanese(コーヒー, 珈琲, カフェ, coffee, cafe) which were acquired by Twitter Streaming API⁴

⁴ <https://dev.twitter.com/docs/api/streaming>

from March 10, 2014, to June 28, 2014. We perform morphological analysis using Mecab[4] for the collected tweets, and count the number of occurrences by fixing dictionary form for words. We extract the part of speech when it is a noun or adjective against the word been counted, and treat it as one word of value dictionary.

In the dictionary about “taste”, “price” and “place” have been produced, the top five of the words are shown in Table 1.

Table 1. The top 5 frequent words to define the 3 type of dictionary from the tweet data related to coffee tweet. (a): taste, (b): price, (c): place

(a)			(b)			(c)		
Word (JP)	Part	Meaning	Word (JP)	Part	Meaning	Word (JP)	Part	Meaning
美味しい	n.	delicious	円	n.	yen	雰囲気	n.	mood
濃い	adj.	strong	高い	adj.	expensive	いい	n.	niceness
熱い	n.	hot	得	n.	economical	好き	adj.	like
甘い	n.	sweet	価格	n.	price	店内	n.	interior
薄い	adj.	weak	安い	adj.	inexpensive	豊富	n.	wealth

Extract the tweets about CVS and Starbucks among 118,855 target tweets. Among the tweets, we defined as CVS that contains [コンビニ, convenience] , and Starbucks that contains [スターバックス, スタバ, starbucks]. Table 2 shows the counts of tweets related to CVS and Starbucks extracted. The target data

Table 2. Counts of tweets

Class	count
Starbucks	3612
CVS	1529
Total	5141

is 5,141 tweets out of 118,855 Japanese tweets about the coffee.

We represent the tweets of target data by a Vector Space Model[7] based on the value dictionary. As the N words presented in the value dictionary such as $V_1, V_2, V_3, \dots, V_N$ tweet T is expressed as tweet vector t such as the following.

$$t = \{(V_1, T), (V_2, T), (V_3, T), \dots, (V_N, T)\}$$

In the subsequent analysis, we analyze from tweet vectors, which are mapped to the feature space.

4 Analysis

If customers are focused on value related to “taste”, it can be expected that tweet vectors include value in “taste” dictionary. We compare value on “taste” of customers about CVS and Starbucks through the target tweets. Table 3 shows the number of tweet vectors containing the feature space that is generated from the dictionary.

Table 3. The number of tweets related to 3 kind of category. (a)price, (b)place, (c)taste.

(a)				(b)				(c)			
	F	T	Total		F	T	Total		F	T	Total
CVS	1280	249	1529	CVS	1344	185	1529	CVS	1216	313	1529
Starbucks	3017	595	3612	Starbucks	3327	285	3612	Starbucks	2722	890	3612
Total	4297	844	5141	Total	4671	470	5141	Total	3938	1204	5141

We test with difference of population proportion using Pearson’s χ^2 tests on the Table 3-(a). Result of Pearson’s χ^2 tests, there was no difference in the population proportion at level of significance 0.05. Because there is no difference in the population proportion of tweet related to taste, it can be said that consumers are focused in the same way as CVS and Starbucks on “taste”.

On the other hand, we find the difference in Pearson’s χ^2 tests on “price” and “place” at level of significance 0.05 shown in Table 3-(b),(c). These results indicate customers at CVS are focused on “price” compared with Starbucks. Moreover, customers at Starbucks are focused on “place” compared with CVS.

We analyze that the customers are focused on what kind of value on “place” when they buy CVS or Starbucks. Table 4 shows five words with a great difference at level of significance 0.05 in the population proportion, which corresponds to the value dictionary about place.

Table 4. Difference in the population proportion

Word	meaning	CVS	Starbucks
嬉しい	happy	2	55
高級	high-quality	3	48
大人	adult	2	41
若者	youth	2	25
オシャレ	fashionable	1	20

There are many positive words for Starbucks coffee, such as “happy”, “high-quality” and “fashionable”. It is evaluated that the place offering Starbucks is

a high quality compared to the CVS, and customers are satisfied. In addition, many words about Generation like “adult” and “youth” have appeared.

Above all, It is seen from generation words appeared from Starbucks tweet vectors that third place has been evaluated for not only equipment but also generations and people who gather there. In contrast, the place of CVS has not been evaluated, it is found that customers are not interested in the place for drinking CVS.

5 Discussion

In this study, we discuss the customer's value of coffee service by analyzing the huge tweet data set. We suggest the quantitative evaluation method of customer's value by using the morphological analysis of the tweets and generated value dictionary.

The characterizes of all service and product can be classified into three qualities: search qualities, experience qualities and credence qualities[9]. Search qualities, which can be called spec information, are relatively simple as in these three qualities, customer can search information and make a decision before the purchase literally. Experience qualities indicates a value that can not be evaluated unless try to purchase the products and services in practice. It is difficult to be evaluated and quantified without trying. Credence qualities, refers to a product that the specialization is high. In addition, this nature is very strong in such as legal services and medical care. Cause in these services, it is generally difficult to determine the value even examine and experience the service actually. There is no choice but to trust the contents of the service, it is called trust value.

With commoditization, the value of product and service fall into the search qualities[8]. Originally, as Japan is represented by “hospitality”, so it is good at imparting credence qualities and experience qualities. It is important to think how to scientifically handle experience qualities and credence qualities and how to incorporate product and service in order to avoid commoditization.

The value dictionary used in this analysis, it is assumed that the value for “price” is the search qualities, and the value for “place” is experience qualities. Therefore, the conceptual diagram of each field in the coffee service is shown in Fig 1. For the CVS, customers have a consumption with a focus on value related to “price” compared to the Starbucks. Accordingly, commoditization in the service area as “Convenience coffee” and “Seattle-type coffee” is because the values of the two regions are depressed in the search qualities. Therefore, it is considered that by making the service deployment focused on the searched value “place”, “Seattle-type coffee” will become possible to control commoditization.

Thus, SNS data have a useful value to analyze the service value of any market service and customer satisfaction in search qualities, experience qualities and credence qualities. Our suggested methodology can be applied to any service and product to improve the commoditization problem from the viewpoint of service science.

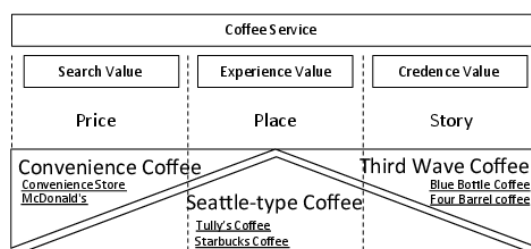


Fig. 1. Service value in coffee service

References

1. Anderson, C.: The end of theory. Wired magazine 16(7), 16–07 (2008)
2. Eric Pfanner, H.K.: 7-eleven gets sales boost from coffee drinkers in japan. Wall Street Journal (2014)
3. Godes, D., Mayzlin, D.: Using online conversations to study word-of-mouth communication. Marketing Science 23(4), 545–560 (2004)
4. Kudo, T., Yamamoto, K., Matsumoto, Y.: Applying conditional random fields to japanese morphological analysis. EMNLP 4, 230–237 (2004)
5. Liu, Y.: Word of mouth for movies: Its dynamics and impact on box office revenue. Journal of marketing 70(3), 74–89 (2006)
6. Oldenburg, R.: The great good place: Cafes, coffee shops, bookstores, bars, hair salons, and other hangouts at the heart of a community. Marlowe New York (1999)
7. Salton, G., McGill, M.J.: Introduction to modern information retrieval (1983)
8. Schmitt, B.: Experiential marketing. Journal of marketing management 15(1-3), 53–67 (1999)
9. Zeithaml, V.A.: How consumer evaluation processes differ between goods and services. Marketing of services 9(1), 25–32 (1981)

Using Wiki as a Collaboration Platform for Software Requirements and Design

Irit Hadar¹, Meira Levy², Yochai Ben-Chaim³, Eitan Farchi⁴

¹Information Systems Department, University of Haifa Haifa, Israel
hadari@is.haifa.ac.il

²Industrial Engineering and Management Department,
Shenkar College of Eng. and Design, Ramat-Gan, Israel
lmeira@shenkar.ac.il

³yochai.benchaim@gmail.com

⁴IBM Haifa Research Lab., Haifa, Israel
farchi@il.ibm.com

Abstract. Wiki is a collaboration platform allowing building a corpus of knowledge in interlinked web pages created and edited by different users. Wiki has been applied in different domains and usage contexts in business and education. This paper reports on an exploratory study conducted as a part of an ongoing research regarding the use of wiki in software development projects. The objective of this study was to capture the perceptions of users regarding wiki as a collaboration platform. Specifically, we aimed at understanding what hinders and what motivates users' contribution to documents constructed within the wiki environment. Based on data obtained via interviews with wiki users we found, among other things, that they do not perceive wiki as a stand-alone communication and collaboration tool, and tend to conduct off-line discussions prior to changes made in the wiki. We also identified several wiki features, which, if improved, may enhance wiki usage.

Keywords: Wiki, collaboration, software development, design, requirements, knowledge creation, knowledge capture.

1 Introduction

Wiki is a platform that allows different users to collaboratively build and share body of knowledge [1,2,3]. It has been applied in different domains and usage contexts including the most commonly known and used Wikipedia, educational wiki applications, and various business-related wiki-based collaboration systems [4,5].

In our ongoing research, we focus on wiki as a collaboration platform in software development, and specifically for developing requirements and design documents. Using wikis for improving software documentation [6], managing documentation of a software project [7], supporting collaborative requirements engineering [8] and as a basis for development environments [9] have been suggested in the past. When using wiki for design and requirements, overcoming cultural issues should be considered [10], otherwise these wikis become inactive after a short period of time [11].

In this study we explore this phenomenon by investigating the perceptions of wiki users in software projects using wiki as a collaboration platform. Furthermore, we aim to understand what motivates users to be more engaged in projects via wiki usage and to contribute to the corpus of knowledge built in the wiki environment.

The next section briefly describes the wiki platform and its existing and potential benefits as a collaboration tool. Next we present the settings of the exploratory study, followed by its preliminary results. We discuss and analyze our findings and, finally, we conclude and discuss future work.

2 Wiki Collaboration Platform

A wiki is a set of linked web pages created through incremental development by a group of collaborating users [2]. The first wiki was developed by Ward Cunningham in 1995, as the Portland Pattern Repository, to communicate specifications for software design. The term wiki (a Hawaiian term meaning “fast”) gives reference to the speed with which content can be created.

Wiki is defined as "a system that allows one or more people to build up a corpus of knowledge in a set of interlinked web pages, using a process of creating and editing pages" [1]. The wiki system is designed to increase collaboration and pool of ideas, as a centralized system designed to increase efficiency and reduce uncertainty [3]. Since the first wiki in 1995, wiki has been used in many contexts: in the public domain, most notably, the online encyclopedia Wikipedia project; in education applications [12] and in the business context [13]. Wiki use was found successful and increasing in the software development industry because it is an excellent means to collect asynchronous contributions from a group of distributed people in a centralized repository of textual artifacts [14]. Wikis are particularly valuable in distributed projects as global teams may use them to organize, track, and publish their work.

Using wiki as an intranet and content management system in a company holds benefits and challenges [15]. For example, a previous study of the use of wiki at IBM showed that it was simple for the creation of information artifacts and for access and use of shared information, resulting in better overview of a project's status [16].

3 The Empirical Study

3.1 Settings and Method

The objective of the exploratory study was to identify the perceptions of wiki users in software development projects regarding the wiki as a collaborative platform. The participants in the study were software developers from four different software development projects in IBM, exhibiting multiple case studies [17]. In these projects, the wiki platform was used mainly for the requirements analysis and design stages. The software development team members in these projects were globally distributed,

spread over three different continents, and from a wide range of managerial roles and seniority. Eleven local team-members participated in the current study.

The use of wiki for maintaining all project requirements and design documents was enforced in all projects participating in the study. The wiki was also used to monitor the project status and highlight the main requirements and design issues agreed upon. The wiki hosted informative documents such as requirements and design documents, and presentations. All the wiki pages were available to all the users for viewing and changing, according to the "wiki way" [2].

The use of the wiki incorporates automatic wiki-based characteristics:

- Data is visible to all, and always online and available.
- Changes are automatically visible to all as soon as they happen.
- Notifications are sent to stakeholders when changes are made.
- Wiki usage is logged and monitored (this was common knowledge to the users).

Our research method was based on the grounded theory approach [18] employing semi-structured interviews as a data collection tool. The interviews included questions regarding the usage of the wiki in general, specific usage behaviors, and about opinions regarding exiting wiki features, for the purpose of probing for parameters that may explain user behaviors and attitudes toward the wiki. The collected data underwent an inductive analysis, in which the transcripts of the interviews were divided to segments, followed by their analysis including open, axial and selective coding [18], in which categories emerged from the data. These categories will be presented next.

3.2 Findings

Tables 1 and 2 summarize the emergent categories regarding the usage and perceptions of wiki as a collaboration platform, and proposed improvements.

Main Category	Sub-category	Comments/Examples
General usage of wiki	Frequency: Several times a month	This was also validated by monitoring the users' actual behavior (log).
	Action: Mostly change existing material	Some users have never inserted new material.
Communicating via wiki	Perceived as a waste of time	"Wiki is an un-needed overhead" "Wiki is used in addition to other communication means, thus doubles the efforts."
	Prefer F2F or email discussions	"We discuss issues in 1:1 meetings or emails." "Using comments blocks the page; so using an off-line communication tools is better."
	Only few respond to wiki email notifications	Several interviewees indicated that they never respond to wiki notifications.
Attitude towards changing text within wiki	Only following an off-line discussion	"If a page is owned by someone, I would not change it without discussing with them first."
	Writing notes or highlight changes rather than erasing/changing.	"I would add but not erase or overwrite [...] I would mark with a different font [to indicate suggestions for change]."

Table 1. Categories of usage and perceptions of wiki as a collaboration platform

Our analysis was aimed at understanding the aspects of wiki that may hinder wiki usage. Therefore, the categories presented in Table 1 and 2 are focused specifically on these aspects. Moreover, these categories are not independent, and specifically, some categories in Table 1 (the current state) are directly linked to categories in Table 2 (possible improvements). For example, “Using comments blocks the page; so using an off-line communication tools is better” explains both – the current reluctance and the proposed improvement to enhance wiki as a communication tool by adding notes.

Main Category	Sub-category	Examples
Wiki user interface features to be improved	Personalized view	“Automated user-specific view: see what is meant for you first.”
	Editor	“Word processor is much more convenient.”
	Visual representations	“It’s easier to draw UML diagrams in a UML tool.”
	Attachment	“Using attachments is not comfortable.”
Wiki communication features to be improved	Notes writing	“More comfortable commenting is needed.”
	Reducing the overwhelming number of notification	“The notification mechanism should be personalized, so that notifications will be sent only to the specific users interested in the change made.” “Filter pages I receive notifications on.”
Accessibility	Should be included in single sign-on	“An additional login is not comfortable.”

Table 2. Required improvements for the wiki platform

3.3 Discussion

The most salient finding from the interviews is that the interviewees *do not perceive wiki as a stand-alone collaboration tool*, in which they are able to communicate with their peers, discuss the issues at hand and accordingly update the knowledge constructed in the wiki. This drives them to communicate in other means prior to making changes in the wiki environment, which in turn facilitates their perception of the wiki as an inconvenient and redundant overhead rather than an enabler of collaboration.

One reason for this perception may stem from the lack of understanding and embracing the “wiki way” [2]: no individual owns a wiki page, but rather all involved users have a mutual responsibility to develop it and keep it up to date. An additional obstacle is the technical limitations inherent in the specific wiki environment used in the investigated projects. In this environment, communicating with peers in order to discuss required changes is perceived as inconvenient and restricting, making the users reluctant to change text inserted by their peers without preceding off-line discussions. Once the discussion has been conducted and decisions have been made, documenting it in the wiki seems to be documentation-related overhead rather than a collaborative construction of knowledge. This, again, is not in line with the wiki way where knowledge is constructed in the wiki environment rather than merely documented.

Many advantages of the wiki are relevant in the settings of this research. For example, wiki is known to be beneficial in the workplace for groups requiring a collaborative medium, with relatively small number of participants who are geographically distributed, as is the case of software development projects [14] such as those participating in this research. Nevertheless, we found the above documented obstacles that prevent realizing the full potential of the benefits of wiki.

The findings of the current research may explain previously reported results regarding the lack of long-term continuity in wiki usage (e.g., [11]). Furthermore, our findings lay the foundations for possible future enhancements of the wiki platform in order to motivate its usage and augment its benefits. These enhancements can be implemented and then evaluated utilizing social network analysis for analyzing users' actual behavior, combined with qualitative perception analysis.

While this paper focuses on qualitative, interview-based data, our on-going research includes data collected via additional qualitative and quantitative methods, providing further findings and validation. Nevertheless, some limitations are to be taken into consideration. The study was conducted with the participation of team-members in four distributed software development projects in IBM. The specific wiki platform configuration used in these projects was selected, and its use was guided, by IBM personnel. As in any case study, the findings of this study reflect the setting of the specific cases investigated, and generalization of the conclusions should be done cautiously, taking the factors reported here into consideration. Additional case studies in various contexts are needed for further validation and generalization of the results.

4 Conclusion

Wiki is a collaboration platform that allows different users to collaboratively build and share body of knowledge. Wikis are used in various settings and contexts for supporting collaborative processes. This research examined the use of wiki in distributed software development projects, and specifically in the requirements analysis and design phases. Our aim was to explore the perceptions and attitudes of software developers using wiki in this context towards the collaborative nature of the wiki environment, and specifically to identify what hinders users from exhausting the full potential of the wiki as a collaborative platform.

Our findings suggest that the wiki environment as implemented in the IBM software development projects participating in this study is not perceived as a stand-alone communication and collaboration tool, and that its users tend to conduct off-line discussions prior to changes made within the wiki. We also identified several wiki features as well as cultural aspects, which, if improved, may enhance wiki usage and its benefits. In the future we plan to develop and implement enhancements for the wiki environment according to the findings of this study, and evaluate their contribution. This research is based on multiple case studies from a single organization. Generalization of the findings is limited, however possible, in cases where similar implementation of wiki is observed. Further research may examine additional cases of wiki usage in software development, to further validate and generalize the identified factors and stemming perceptions hindering the usage of wiki as a collaboration platform.

References

1. Franklin, T., Van Harmelen, M.: Web 2.0 for content for learning and teaching in higher education, Technical Report, Bristol: JISC (2007)
2. Leuf, B., Cunningham, W.: The wiki way: Quick collaboration on the web, Addison-Wesley, Boston, MA (2001)
3. Topper, C. M., Carley, K. M.: A Structural perspective on the emergence of 32 network organizations, *Journal of Mathematical Sociology*, 24, 1, 67-96 (1999)
4. Andersen, E.: Using wikis in a corporate context, in: Hohenstein, A. and Wilbers K. (Eds.), *Handbuch E-Learning*, Deutscher Wirtschaftsdienst, vol. 11, pp. 1-15 (2005)
5. Cole, M.: Using wiki technology to support student engagement: Lessons from the trenches. *Computers & Education*, 52, 1, 141-146 (2009)
6. Correia, F. F.: Extending and integrating wikis to improve software documentation, *The Fourth Workshop on Wikis for Software Engineering*, Porto, Portugal (2008)
7. Alonso, J. M. G., Olmeda, J. J. B, Rodriguez, J. M. M.: Documentation center – Simplifying the documentation of software projects. *The Fourth Workshop on Wikis for Software Engineering*, Porto, Portugal (2008)
8. Ferreira, D., Alberto R. S.: Wiki supported collaborative requirements engineering, *The Fourth Workshop on Wikis for Software Engineering*, Porto, Portugal (2008)
9. Luer, C.: A component-based, agile software development wiki, *The Third Workshop on Wikis for Software Engineering, WikiSym2007*, Montreal, Quebec, Canada (2007)
10. Biesack, D.: The culture of collaboration. *The Third Workshop on Wikis for Software Engineering, WikiSym2007*, October 21, Montreal, Quebec, Canada (2007)
11. Arazy, O., Croitoru, A., Jang, S.: The Life cycle of corporate wikis: An analysis of activity patterns, *The 19th Workshop on Information Technologies and Systems (WITS)*, December 14-15, Phoenix, AZ, USA, 163-168 (2009)
12. Ravid G.: Open large shared knowledge construction systems dominance: The Wikipedia social structure, *Academy of Management Annual Meeting* (2007)
13. Majchrzak, A., Wagner, C., Yates, D.: Corporate wiki users: Results of a survey, *The International Symposium on Wikis '06*, Aug. 21–23, Odense, Denmark, 99-104 (2006)
14. Louridas, P.: Using wikis in software development, *IEEE Software*, 23 (2), 88-91 (2006)
15. Trkman, M., Trkman, P.: A wiki as intranet: a critical analysis using the Delone and McLean model, *Online Information Review*, 33, 6, 1087-1102 (2009)
16. McCarty, E.: How IBM uses an intranet to connect a global audience, *Knowledge Management Review*, 11 (2), 28-33 (2008)
17. Yin, R. K.: *Case Study Research, Design and Methods*, 3rd ed, Newbury Park, Sage Publications (2003)
18. Strauss, A.L., Corbin, J.: *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. Sage Publications (1998)

Enhancing Software Architecture via a Knowledge Management and Collaboration Tool

Sofia Sherman¹, Irit Hadar¹, Meira Levy², Naomi Unkelos-Shpigel¹

¹Information Systems Department, University of Haifa, Haifa, Israel
{shermans, hadari, naomiu}@is.haifa.ac.il

²Industrial Engineering and Management Department,
Shenkar College of Eng. and Design, Ramat-Gan, Israel
lmeira@shenkar.ac.il

Abstract. Software architecture is an important part of software development, aiming at ensuring a high-quality product. Recent research has shown that collaboration and knowledge management are important parts of the architecture process, and have significant role in architecture design and review. In this paper we present a prototype for a tool we developed, as part of our ongoing research on the software architecture process, for supporting collaboration, communication and knowledge sharing during all steps of the architecture development process. This tool was developed based on the findings of a case-study research in a global, large software organization.

Keywords: Software architecture, knowledge management, collaboration.

1 Introduction

Over the last two decades, there has been increased focus on architecture within software development. Software architecture design is an essential stage in the software development process and has a significant impact on software quality. Literature in this field usually refers to technological aspects of architecture such as design methods, architectural notations, patterns, analysis, etc. Only few empirical studies focused on human and collaborative aspects of software architecture [e.g. 1, 3, 4].

During the software architecture process many obstacles arise, most of which stem from collaboration-related issues and lack of specific knowledge resources [2]. One of the main problems caused by insufficient collaboration between stakeholders is the misalignment between the architecture review and the architecture design processes, and between different architecture review process instances and their outcomes. Knowledge preservation, accessibility, and reuse are vital for good and efficient software architecture design and review. Using already created and proven knowledge reduces time and resources invested, and typically results in higher-quality outcomes.

This paper presents a prototype for a tool we developed, as part of our ongoing research on the software architecture process, for supporting different aspects of the

architecture process, including collaboration, communication, knowledge sharing, and review. This tool was developed based on the findings of a case-study research in a global, large software organization. A subset of these findings can be found in [2].

2 Motivation and Background

Several studies have shown the importance of collaboration and knowledge management during software development in general and during architecture design and review in particular [3, 4]. In this section we provide a brief overview of the topics we combine in our research.

Software architecture design consists of two steps: (1) choosing an overall strategy for the architecture; and, (2) specifying the individual components that make up the application [5]. These result in an architecture document that includes architecture views and design rationale. Architecture design includes extracting and understanding architecturally significant requirements, making choices, synthesizing a solution, exploring alternatives, validating them, and more [1]. Most definitions of software architecture in literature typically relate to technological aspect of architecture design, and do not emphasize the required resources, including tools and human stakeholders. Architects need to work and collaborate with other stakeholders in order to produce high-quality architecture. Insufficient collaboration and knowledge sharing between the stakeholders involved in the process leads to difficulties and obstacles towards successful design of a software architecture solution [2].

Software architecture review is a vital phase in the software architecture process, aiming at ensuring and validating the quality of a proposed architecture. The reviewers focus on identifying project problems before they become costly to fix; providing information to various stakeholders for better decision making; and improving product quality [6]. Literature review reveals a wide range of evaluation methods used in review process [e.g. 3], however, little attention has been given to the review process in general and we found no evidence for tool that supports this process [7].

Knowledge management (KM) KM plays a central role in architecture process, embodies knowledge and serves as a vehicle for communication among stakeholders. In particular, software architecture captures early design decisions and it is a transferable abstraction of the system, which enables further reuse in other software systems [8]. Thus, managing architecture knowledge is an important and inherent part of the software development processes. According to Maranzano et al. [6], KM should be embodied in architecture review for capturing best practices and socialize such practices across the organization. However, while their research fosters cross-organizational learning and transfer of lessons learned and best practices, and highlights the need to identify and share corporate assets and reusable components, it does not provide a framework for embedding the KM practices within the review process itself for facilitating collaboration and engagement among the review participants.

Collaboration between software architects for sharing architecture knowledge has been the focus of several research works. For example, Farenhorst et al. [9] proposed an architectural knowledge-sharing tool based on best practices of KM and character-

istics of architecture knowledge. However, we did not find any research focusing on the collaboration between software architects and architecture reviewers, or among architecture reviewers, towards or during review performance. The tool presented in this paper aims to bridge this gap by providing an environment and guidelines for embedding collaboration and KM practices throughout the architecture process.

3 Proposed Solution

3.1 Framework

The main concepts this solution is based on are collaboration, knowledge management and sharing, between and among different stakeholders involved in the architecture process. Here we address only the stakeholders who are directly involved in the process of developing architecture and approving it for the next stages. The framework for the conceptual solution model (Fig.1) presents a general view of the architecture process.

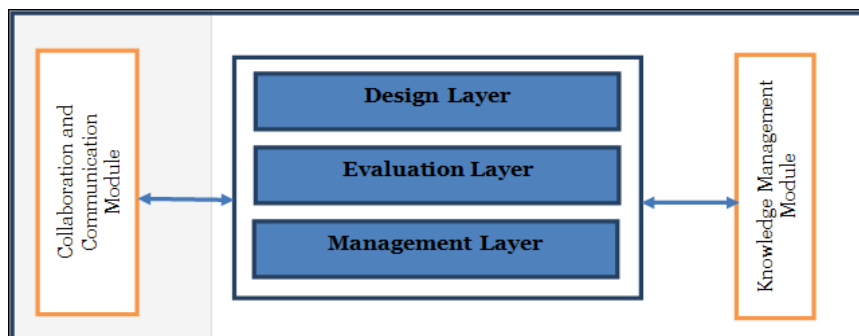


Fig. 1. A conceptual framework for architecture KM and collaboration.

We divided these stakeholders to three main layers: (1) Management – stakeholders who manage the architecture process including: project allocation, review allocation and architecture knowledge. (2) Design – software architects who are responsible for designing and documenting the architecture solution, preparing it for review, and improving the architecture solution according to the reviewers’ comments and instructions, if provided. (3) Evaluation – architects, who serve as architecture reviewers and operate in different levels of the architecture process, for example: peer review, manager review, or board review.

The proposed solution supports online collaboration and KM. Collaboration is enabled via the Collaboration and Communication module, which includes several types of collaboration means: forum, private messaging, and collaboration on documents under review. KM is enabled via the Knowledge Management module; each outcome produced during the architecture process via the collaboration module will be stored for future use in the KM module. These outcomes include discussions between reviewers and architects and among reviewers during the processes of the architecture

design, review preparation, and review discussions and decisions. This information will be made accessible for architects and reviewers: architects will be able to reuse this knowledge for future architecture development and learn about reviewers' requirements and comments; reviewers will be able to access previous reviews' outcomes in order to see, for example, previously raised issues regarding the same product, customer or architect. Additional information included in the KM module is theoretical architecture knowledge extracted from textbooks, and checklists formulated by managers and reviewers, based both on theory and practice. These checklists evolve within the tool over projects and time.

3.2 Tool

Based on the framework presented above, we developed a prototype for a tool that manages and supports the architecture process. The architecture design and review processes are supported by the proposed tool as described below.

Architecture knowledge management and collaboration: There are two types of knowledge available in the tool. *Formal knowledge* produced by the management and evaluation layers based on different sources: company requirements, architecture educational sources (e.g., textbooks), and previous review outcomes. This knowledge is presented to the architecture design layer as guidelines (Fig. 2.a) and is accessible directly in the guidelines or via the search engine. In the current version of the tool, we used [10] as the source for the quality attributes' description and as a basis for the first version of the checklist. *Informal knowledge* created while working with the tool. It allows relevant stakeholders to access documents from different projects, public conversations regarding projects, and relevant architecture artifacts. This knowledge is accessible in the tool via projects or the search engine (Fig. 2.b).

The collaboration within the system provides two important capabilities. First, it enables communication and joint decision making during architecture design and architecture review. Second, since all comments and discussions are documented in the system, this knowledge becomes available, not only to stakeholders of the current projects, but also to architects and reviewers in other projects, who may benefit from the lessons learned and knowledge accumulated over previous projects.

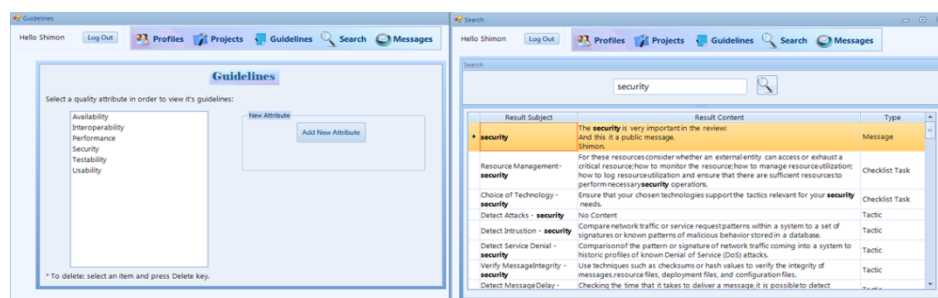
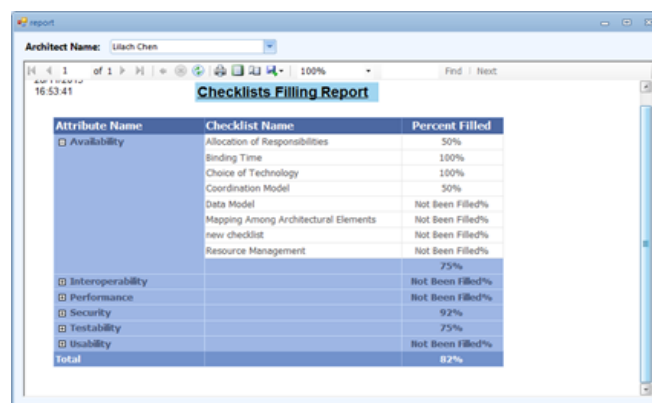


Fig. 2: Screen Shots

a. Formal knowledge content

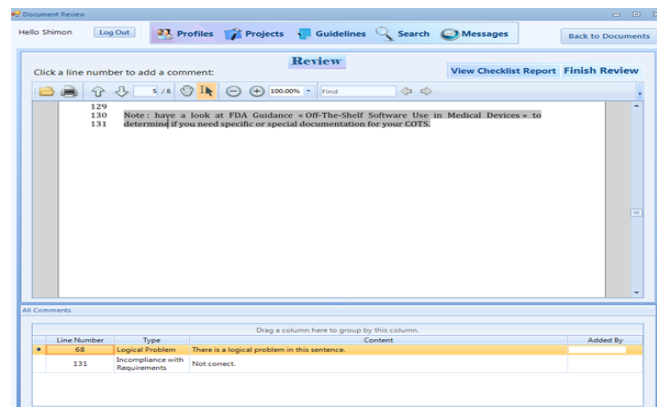
b. Knowledge search engine

The architecture design process: The main usage scenarios include: (1) Architects are allocated to projects by the management layer and design the architecture. During their work, architects can access formal knowledge in the guidelines section (Fig. 2.a) or informal knowledge in forums or comments that reside in other projects (Fig. 2.b). Architects can communicate with management and evaluation layer users, for questions and guidance. (2) During the architecture design, architects fill a checklist report (Fig. 3) regarding all aspects of the architecture under development. (3) When ready, architects submit the architecture document and the checklist report.



Attribute Name	Checklist Name	Percent Filled
Availability	Allocation of Responsibilities	50%
	Binding Time	100%
	Choice of Technology	100%
	Coordination Model	50%
	Data Model	Not Been Filled%
	Mapping Among Architectural Elements	Not Been Filled%
	new checklist	Not Been Filled%
	Resource Management	Not Been Filled%
		75%
Interoperability		Not Been Filled%
Performance		Not Been Filled%
Security		92%
Testability		75%
Usability		Not Been Filled%
Total		82%

Fig. 3: Screen Shot: Checklist report



Click a line number to add a comment:

Note: have a look at FDA Guidance « Off-The-Shelf Software Use in Medical Devices » to determine if you need specific or special documentation for your COTS.

Line Number	Type	Content	Added By
66	Logical Problem	There is a logical problem in this sentence.	
131	Incompliance with Requirements	Not correct.	

Fig. 4. Screen Shot: Review screen

The architecture review process: The main usage scenarios include: (1) Upon submission, the management layer representative allocates reviewers listed in the evaluation layer, according to their profiles. (2) While performing review, reviewers can to add comments within the architecture document and automatically follow comments in the document that were left by other reviewers (Fig. 4), as well as follow-up on changes made by the architect based on the review comments. During the

review, reviewers can communicate with each other using the Collaboration and Communication module, look for history of previous reviews of the same or other products, read comments, and participate in discussions among reviewers regarding other relevant software reviews. Using the system for discussions among reviewers, contributes to managing knowledge history by documenting and storing it in the database and making it accessible for future needs.

4 Conclusion and Future Work

Previous research has shown that insufficient collaboration and knowledge management hinders the quality of software architecture. The solution we propose supports these two aspects of the architecture process and provides a platform for managing architecture design and review with emphasis on collaboration between the stakeholders involved and knowledge creation and reuse.

In the next step of this research we will evaluate the proposed solution in collaboration with a global, large software organization in order to learn about its strengths and weaknesses from practitioners' point of view, and possibly conduct a feasibility check via a controlled usage of the tool in real industrial settings.

References

1. Kruchten, P.: What Do Software Architects Really Do? *The Journal of System and Software*, 81, 2413-2416 (2008)
2. Sherman, S., Hadar, I., Levy, M.: Enhancing Software Architecture Review Process via Knowledge Management, *Proceedings of the AMCIS*, Lima, August (2010)
3. Babar, M., Kitchenham, B., Zhu, L., Gorton, I., Jeffery, R.: An empirical study of groupware support for distributed software architecture evaluation process. *Journal of Systems and Software*, 79(7), 912-925 (2005)
4. Sherman, S., Hadar, I., Hadar, E., Harrison, J.: The Overall Value of Architecture Review in a Large Scale Software Organization. *Proceedings of the IWSSA co-located with CAISE'11* (2011)
5. Gorton, I.: *Essential Software Architecture*, Springer Berlin Heidelberg NY (2006)
6. Maranzano, J. F., Sandra A., Rozsypal, S. A., Zimmerman, G. H., Warnken, G. W., Wirth, P. E., Weiss, D. M.: *Architecture Reviews: Practice and Experience*. *IEEE Software*, March-April (2005)
7. Babar, M. A., Zhu, L., Jeffery, R.: A framework for classifying and comparing software architecture evaluation methods. *Australian Software Engineering Conference* (2004)
8. Babar, M. A., Dingsøyr, T., Lago, P., van Vliet, H.: *Software Architecture Knowledge Management - Theory and Practice*. Springer (2009)
9. Farenhorst, R., Lago, P., van Vliet, H.: EAGLE: Effective tool support for sharing architectural knowledge. *Int. J. Cooper. Inform. Syst.* 16(3/4), 413-437 (2007)
10. L. Bass, P. Clements, R. Kazman: "Software Architecture in Practice". 3rd ed., Addison-Wesley Professional (2013)

PRIORITIZATION OF THE REQUIREMENTS FOR THE PLATFORM TO BE USED FOR THE EMPLOYMENT OF QUALIFIED - DISADVANTAGED INDIVIDUALS

Ahmet Suerdem¹, Basar Oztaysi²

¹Istanbul Bilgi University, Department of Business Administration, Istanbul, Turkey
ahmet.suerdem@bilgi.edu.tr

²Istanbul Technical University, Department of Industrial Engineering, Istanbul, Turkey,
oztaysib@itu.edu.tr

Abstract: Collaboration of system designers and social researchers for the solution of social problems is increasingly becoming important. In this context, G@together project funded by JPI-Urban Europe is an interdisciplinary project aiming to build an e-platform facilitating the employment of disadvantaged overqualified people. The use of qualitative social data and the participation of end-users to the system design process provides a significant potential for the MIS research. In this study; requirements of different stakeholders concerning the platform are obtained through unstructured interviews. Subsequently, qualitative and quantitative analysis methods are used to structure the requirements list. Finally, Analytic Hierarchy Process, a commonly used multi-criteria decision making technique is used to prioritize the system requirements.

Keywords: Disadvantaged groups, employment, system design, Analytic Hierarchy Process

1 Introduction

Employment and social involvement of disadvantaged groups is a widely discussed area invoking the development of various projects in European Union and Turkey. G@together “Get together without barriers” is such a project which came into play through a reciprocal agreement between Austria and Turkey within the scope of JPI Urban Europe. The aim of the project is to reinforce the employment of qualified yet socially disadvantaged individuals, thus targeting at establishing an Internet-based online recruitment platform prototype.

When online recruitment systems first emerged in the mid-1990s, popular management press heralded these systems as a ‘recruiting revolution’ (Boydell, 2002). Despite this hype, however, some critics (e.g., Harris et al., 2003, Kehoe et al., 2005 and Stone et al., 2013) have noted that these systems are still not an

efficient alternative to traditional methods because of their unintended dysfunctional consequences. Hence, major shortcomings of these systems can be summarized as: being deprived of talent pool diversity, being uniform and not being flexible and customized. The major reason underlying these shortcomings is that the data collection required for the needs analyses performed when designing these platforms has been conducted suddenly and unexpectedly in accordance with the decisions of the company management and experts. (Palmer and Kaplan 2007). Such designed platforms mainly respond to the short-term overqualified personnel requirements of the major companies and ignore the talent pool diversity as well as the requirements of minor companies or NGOs with respect to employment (Maurer and Liu 2007). To make these platforms more efficient and effective to fulfill the requirements of various user groups, a more flexible, open and participative design is required. Establishing all stages of the design together with the stakeholders will prevent from the very beginning staying focused on an incorrect definition that may turn into an unattractive product for the users (Bhattacharya et al. 1998). A participative conceptualization process where all stakeholders produce ideas and turn them into concepts and definitions before specifying the properties of the system is highly important for the success of the final product. The previous problems regarding the incorrect concepts and definitions of the system may result in high cost and failures in the next stages of the product development process (Florén and Frishammar 2012). In particular, non-overlapping of the needs analyses for the product characteristics and the system architecture involving various components and interfaces will frequently lead to system failures for the complex systems including many various software modules, agents, data structures and interfaces (Wood and Jones 1995). A participative and process-spread needs analysis will yield flexible, integrated, efficient, effective and smart products for the end users. In this process, the system users should be identified not only as customers but also as the widest stakeholder spectrum that will be affected by or affect the completion of the system (Driessen and Hillebrand 2013). Therefore, the platform being designed aims at finding out the aspects of the present system not suitable for the use of the disadvantaged and establishing a system architecture based on the needs analyses involving the users. This study summarizes the studies regarding the manifestation and prioritization of the features that should be available in the respective platform. Within this framework, the subject of the article suggests a method that integrate the social sciences and system engineering studies in the needs analyses in order to ensure flexibility in the system design. The continuation of the study was arranged as follows: First of all, the methods used to establish the needs of the system users are summarized and the integrating Analytic Hierarchy Process method of the social research and system design that has been used to prioritize the above is introduced. The subsequent part describes the details of the application performed and gives a summary of the results. The final part discusses the results and implications.

2 Method

2.1 Main Steps of the study

Within this scope, a method consisting of five steps has been used to specify and prioritize the characteristics that an online platform should have to support the employment of qualified - disadvantaged individuals which is the content of our research. These steps may be briefly summarized as follows:

1. Determining the problems of the target group concerning employment: One-to-one interviews are performed with the target group members, employers and representative of the employment centers to describe the obstacles to employment of these individuals. The relationships between the obstacles and problems are introduced.
2. Determining the Good Practices: In this field, both the studies in the academic world and the good practices are reviewed. The properties of the system that may be used in the possible applications in Turkey are determined.
3. Determining the potential needs of the system through Brain Storming method: The characteristics that the aforementioned online platform should have are determined together with the relevant people.
4. Excluding the unnecessary ones among the possible system needs: The characteristics revealed in the previous step are combined and listed as the system needs. At this stage, the individuals who will realize the platform review the list as the decision-makers and the unnecessary or undesired needs for the system are specified before the initiation of the prioritization process.
5. Prioritization of the system needs: The residual system characteristics of the previous process are prioritized by using the Analytic Hierarchy Process. Following this step, the degree of importance of each system requirement is determined.

2.2 Analytic Hierarchy Process

Analytic Hierarchy Process (AHP) is a multi-criteria decision making method developed firstly by Thomas Saaty (1980). AHP method disintegrates the complex decision problems as per different perspectives and addresses them as a hierarchy. In this way, the complex problems are simplified into easier sub-problems. AHP method has an appropriate structure to make a decision by considering both the quantitative and qualitative factors.

The structure of AHP consists of a problem objective, the criteria to be used to decide for this objective and alternatives among which the selections will be made (Figure 1).

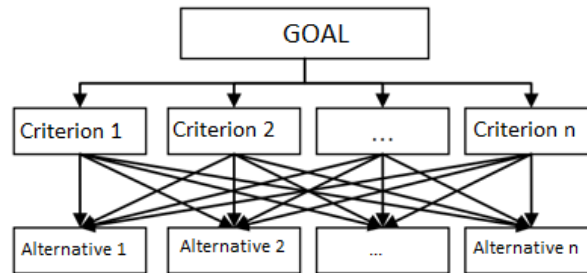


Fig. 1. Example of AHP hierarchy

The steps and processes contained within AHP method are listed as follows (Saaty, 1980; Özden, 2008):

Step 1: Identification of the Problem: The existing problems are identified in the first stage of AHP. Upon the identification of the problem, the objective to be achieved by the decision-maker(s) with AHP is determined.

Step 2: Identification of the Criteria: The characteristics that the alternative to be selected with respect to the problem should have are clarified. During the identification of the criteria, the opinions of the relevant people should be asked and a needs list should be created. The number of criteria may vary depending on the type of the problem.

Step 3: Identification of the Alternatives: At this stage, all the alternative options to be considered in order to achieve the objective during decision-making are determined.

Step 4: Establishing the Hierarchic Structure: The identified main objective is divided into several subgoals and a hierarchy as shown in Figure 1 is established. When establishing the hierarchy, the items at the same level are assumed as independent of each other.

Step 5: Performing Paired Comparisons: Based on the established hierarchy, the items at the same level are exposed to paired comparisons with respect to the item involved in the upper level. For example, regarding the structure given in Figure 1, all criteria are paired compared in terms of their effects on the objective. Afterwards, all alternatives are compared as per their success with respect to each criterion. At this stage, the paired comparisons are performed by the decision-maker(s). The evaluations are performed by using the relative significance levels (Table 1) developed by Saaty (1980).

Table 1. Relative significance level to be used in the paired comparisons

1-9 Scale	Definition
1	Equal Importance (E)
3	Modarate Importance (M)
5	Strong Importance (S)
7	Very Strong Importance (VS)
9	Extreme Importance (Ex)

Step 6: Measuring the weight of the criteria: The weight of the criteria is measured by using paired comparison matrix. At this stage, A paired comparison matrix has been obtained provided that the significance level between the criterion a_{ij} and criterion j is shown.

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \quad C = \begin{bmatrix} c_{11} & c_{12} & \dots & c_{1n} \\ c_{21} & c_{22} & \dots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \dots & c_{nn} \end{bmatrix}$$

For this reason, each paired comparison matrix is primarily normalized. The most commonly used method of normalization in practice is dividing each element of the column into the sum of the relevant column.

Then, each element of matrix A is divided into the sum of the relevant column.

$$b_i = \sum_{i=1}^n a_{i1} \quad c_{ij} = \frac{a_{ij}}{b_i}$$

As a result, matrix C with normalized values is obtained. The significance level of the criteria with each other is calculated with the arithmetic mean method by using the values in the matrix C.

Step 7: Consistency Analysis and Revision: After the paired comparison matrices are completed, the consistency of the evaluation is tested. The decision-makers are asked to revise their opinions for the paired comparison matrices with consistency problems.

3 Application and Findings

Within the scope of the project, qualified – disadvantaged group is defined as the "women who have interrupted their carriers because of marriage and child care and want to go back to the job market". In accordance with this definition, collaborative works have been conducted with "Yeniden Biz Platformu" (yenidenbiz.com) a women's NGO working supporting women who have interrupted in their careers.

We have conducted a literature research, to investigate the examples of worldwide good practices. Subsequently we have performed interviews between the employers and woman job seekers. Finally, we have specified the requirements of the target groups for the e-recruitment system to be developed.

The requirements for the system may be grouped under 7 titles. (i) Informing: The characteristics in this group are not directly related to employment; they just include information for the companies and people seeking jobs. (ii) Introduction services: They are the system characteristics that may be involved in the transmission of the people seeking for jobs to the companies, and the companies

and business opportunities to the people seeking for jobs. (iii) Filtering/Matching: They are the characteristics that support the employees to attain the business opportunities and the companies to reach the appropriate employees. (iv) Special Tests: The system characteristics to understand the competences, personal qualities and suitability of the people seeking for jobs for the job are involved in this group. (v) Ease of Use: It contains the characteristics intended for the ease of use of the system. (vi) Communication before the interview: Various characteristics to ensure communication between the candidates and the companies before the face-to-face interviews.

As a result of the interviews performed, it was concluded that the sixth group and certain requirements among the other groups were found unnecessary and excluded from the model. Paired comparison matrices were established for the remaining system requirement groups and system characteristics and the evaluations of the decision-makers regarding the priorities were summed up.

The example of paired comparison matrix for AHP is given in Table 2. The decision-makers were asked to compare the five requirement groups in terms of the platform. The diagonal of the matrix was selected to be equally-weighted. Each evaluation involved in the matrix should be read as the evaluation of the requirement group in the relevant line as per the characteristic group in the respective column. For example; the Introduction characteristics have been found "Slightly Important" with respect to the Informing characteristics.

Table 2. Paired comparison of the requirement groups as per the platform

	C1	C2	C3	C4	C5
C1	Eq		VS	S	M
C2	M	Eq	EX	VS	S
C3			Eq	S	M
C4				Eq	M
C5					Eq

The oral evaluations stated in Table 2 are converted into numeric values in the next step. The numeric values are given in Table 3. Unlike the previous table, the empty cells are completed as the reverse of the value symmetrical to the diagonal as per multiplication.

Table 3. Numeric values of the paired comparisons and the weights

	C1	C2	C3	C4	C5	Priorities
C1	1	0.33	7	5	3	0.260
C2	3	1	9	7	5	0.503
C3	0.33	0.2	1	5	3	0.134
C4	0.2	0.14	0.33	1	3	0.068
C5	0.14	0.11	0.2	0.33	1	0.035

The processes defined in the previous section are performed on this matrix and their significance levels are calculated. The significance levels are given in the last column of Table 3. Accordingly, "Introduction" is the most significant requirement group with 0.503 which is followed by informing and matching requirements.

The abovementioned paired comparison matrices were prepared for the similar and then each requirement groups, and thus the significance of each requirement in the relevant group was found.

The significance level of each requirement in the relevant group is stated in the Local Significance column in Table 5. The significance level of the requirement in the general platform is obtained by multiplying the significance level of the relevant group by the local significance. The significance level of the requirements as a whole is listed under the General Significance.

Table 4. Local and general significance level of the characteristics

	Local Significance	General Significance		Local Significance	General Significance
Informing	0.26		Matching	0.134	
R1.1		0.277	0.072	R3.1	0.197
R1.2		0.248	0.065	R3.2	0.153
R1.3		0.298	0.078	R3.4	0.309
R1.4		0.176	0.046	R3.5	0.341
Introduction	0.5		S. Tests	0.068	
R2.1		0.036	0.018	R4.1	0.264
R2.2		0.080	0.040	R4.2	0.076
R2.3		0.341	0.171	R4.3	0.263
R2.4		0.130	0.066	R4.4	0.397
R2.5		0.413	0.208	Ease of Use	0.035
				R5.1	0.250
				R5.2	0.750

R2.5 (multiple-choice data entry), R2.3 (anonymous identification of the applicant) and R1.3 (online training and courses) have been revealed as the most significant requirements among the characteristics the system should have as per the values in the table.

4 Implications

When the agencies involved in the employment field in Turkey are reviewed, it can be seen that people using İŞKUR and employment centers of the municipality generally do not have much qualifications. Kariyer.net and LinkedIn are frequently used for the preliminary and moderate management qualities, whereas private human resources firms and "head hunter" companies are utilized for the upper positions. All available systems head out with a general definition of disadvantaged or completely exclude the disadvantaged portions. Customized

systems appropriate for the specific needs and demands of the specific social groups will bring more effective solutions into the employment process. Accordingly, more flexible processes and systems are required for the employment of qualified yet disadvantaged ones. This study has been conducted to establish an instance for the use of social science methods together with the analytical methods for the design of such a platform.

5 References

1. Bhattacharya, S., V. Krishnan, and Mahajan V. (1998), "Capturing the Effect of Technology ... Sequence Design Methodology," IIE Transactions, 30, 933-945.
2. Boydell, M. (2002). Internet recruitment helps HR careers. Canadian HR Reporter, Feb 11 2002, Canada. Driessen P.H. and Hillebrand B. (2013), Integrating Multiple Stakeholder Issues in New Product Development: An Exploration, Journal of Product Innovation Management, Volume 30, Issue 2, pages 364–379
3. Florén and Frishammar, (2012) From preliminary ideas to corroborated product definitions: managing the front end of new product development. California Management Review. v54. 20-43.
4. Maurer, S. D., and Liu, Y. (2007). Developing effective e-recruiting websites: Insights for managers from marketers. Business Horizons, 50(4), 305-314.
5. Özden H.Ü. (2008) Primary School Selection using Analytic Hierarchy Method, Journal of Marmara University FEAS Volume XXIV, Issue 1
6. Palmer D. and Kaplan (2007) A framework for strategic innovation. Blending strategy and creative exploration to discover future business opportunities. www.1000ventures.com/business_guide/innovation_strategic_byip.html.
7. Saaty, T.L., 1980. "The Analytic Hierarchy Process." McGraw-Hill, New York
8. Wood D.J. and Jones R.E. (1993), stakeholder mismatching: a theoretical problem in empirical research on corporate social performance, International Journal of Organizational Analysis, Volume 3 issue 3.

Acknowledgement

This work is a part of Urban Europe Project entitled "Gettogether without Barriers" and is supported by the Scientific and Technological Research Council of Turkey (TÜBİTAK), Grant No: 113K027

Visualization of Comparison of Texts from Two Information Sources by Network Integration

Ryosuke Saga

Osaka Prefecture University, Graduate School of Engineering,
1-1 Gakuen-cho, Nakaku, Sakai, 559-8531, Japan
saga@cs.osakafu-u.ac.jp

Abstract. This paper describes a comparison visualization method based on network integration. A topic can be discussed by many people, thus resulting in various information sources with different viewpoints and perspectives. To comprehend a topic, network visualization has been proposed previously. However, it cannot compare claims and opinions. Hence, the current study investigates network visualization methods that individually compare texts from multiple information sources.

Keywords: Knowledge Visualization, Comparison Analysis, Co-occurrence graph, Graph Integration.

1 Introduction

One's impressions of things change depending on viewpoint. For instance, we often use the term "curious" to express a personal characteristic, but this characteristic often corresponds to "get tired easily". Viewpoints are also called positions, perspectives, and fields of view. For consistency, however, we use the term "viewpoint" in this paper.

Traditionally, many arguments are originate from a difference in viewpoints between two individuals. For example, competitors in court (the plaintiff and the defendant) discuss a claim. We can also consider the review system of papers to be a viewpoint-based discussion about originality and feasibility between the authors of a submission and the reviewers. With the progress of the Internet, two or more persons can argue back and forth. On review sites such Amazon.com and Tripadvisor, consumers evaluate items according to their opinions, and service providers such as hotels and producers reply to their customers. In anonymous bulletin board systems, many people submit opinions about a topic based on their viewpoints. Therefore, arguments and discussions are derived from the viewpoints of two or more persons.

Hence, outputs related to an object depend on viewpoint although facts about it are known. In other words, data are generated according to viewpoint (in this case, the viewpoint is known as a model or distribution). Viewpoints are extracted using several methods; for quantitative data, the methods of multivariable analysis and data mining are mainly used (Principal Component Analysis, DEA, etc.)[1][2][3].

To comprehend textual viewpoints, texts (bag of words) are mined through keyword and topic extraction. To extract important keywords and topics, methods of

clustering, PCA, probabilistic PCA and latent Dirichlet allocation are used [4][5]. With these keywords and topics, we can infer and understand their backgrounds and the viewpoints associated with them. However, these methods cannot directly determine the differences and gaps among the viewpoints of n persons. For example, we can analyze a topic in one document set (ex. A newspaper) using PCA, but we cannot comprehend the variation in the results of two document sets.

Given this background, my project aims to determine the difference between the viewpoints of n persons through network visualization [6]. In this study, $n = 2$, that is, I visualize the difference between the viewpoints of two individuals.

2 Network Visualization for Text

Information visualization techniques are very important in result comprehension. Thus, my research aims to apply network visualization in the clarification of viewpoints.

In this study, network visualization is based on graphical representation in mathematics. Hence, a network consists of vertices and edges, which also possess attributes. Vertex set is V ; edge set is E ; and a network G_i is presented as $G_i = G(V_i, E_i)$. V and E are also known as network elements, whereas V_i and E_i are elements of network G_i . Each vertex $v_i \in V_i$ and each edge $e_i \in E_i$ has n and m attributes, that is, $v_i = (v_{i1}, v_{i2}, \dots, v_{in})$, $e_i = (e_{i1}, e_{i2}, \dots, e_{im})$. The edge between two vertices A and B is also expressed as $e_{A \rightarrow B}$ (note that $e_{A \rightarrow B}$ is equal to $e_{B \rightarrow A}$ in an undirected network).

This network visualization has been applied in areas such as human relationships and social network analysis. In the network visualization of text data, a network is generated from a document set, which is derived from an information source. Each vertex is labelled with a keyword or a key phrase, and each edge often displays a co-occurrence coefficient. This research focuses on network visualization of text data within the domains.

The network visualization of text data involves themes such as network generation from data, graph layout, and graph representation. My research emphasizes the comparison of graphs obtained from several text datasets that correspond to viewpoint. If we analyze the differences in viewpoint, our network output must depict the viewpoint of each information source and verify the relationships of correspondent network elements through pairwise comparison.

However, network visualization contains many significant vertices for text data; hence, the analysis cost of comparing and recognizing the differences among networks is high. For instance, the vertices and edges of two networks may be similar in number, but they may be unique. However, two networks may also possess some similar vertices and edges. Thus, each vertex and edge must be verified between two networks, and the comparison cost increases when the network is large.

3 TEXT COMPARISON ON NETWORK VISUALIZATION

3.1 Principle Strategy

The principle strategy of the visualization of text comparison is detailed in the following steps. First, we pre-process the data from the information source to determine the attributes of network elements. This attributes are used for final representation. Subsequently, we extract network elements and construct networks based on each information source. To compare and integrate these networks, we merge network elements from two networks. Finally, the integrated and compared network is outputted.

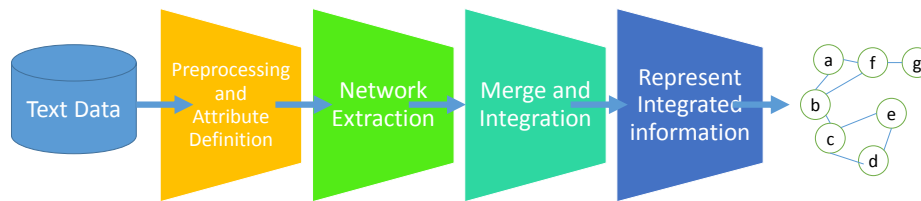


Fig. 1. Principle Strategy of integrated graph

3.2 Network Merging and Integration

We assume that the network derived from datasets D1 and D2 are $G_{D1}=G(V_{D1},E_{D1})$ and $G_{D2}=G(V_{D2},E_{D2})$ and that the integrated network between G_{D1} and G_{D2} is G_I . G_I is given by the integration function Integrate as follows:

$$G_I=G(V_I, V_E)=\text{Integrate}(G_{D1},G_{D2}) \quad (1)$$

In this study, the integration function matches the elements of two graphs and sets the method of integration. In general, we can add, subtract, multiply, and divide. When the integration function obtains the maximum elements of a certain attribute, each vertex and edge is applied as follows:

$$\text{Integrate}(G_{D1}, G_{D2}) = \begin{cases} \max(v_{D1}, v_{D2}) \\ \max(e_{D1}, e_{D2}) \end{cases} \quad (2)$$

An element may also be observed in one graph but not in another. For example, one graph has node “a”, but another does not. Thus, the formula above does not aggregate elements according to the given function set and does not compensate for limited elements. It can be compared with the integration function; for instance, the set function is union operation and the compensated graphs of G_{D1} and G_{D2} are denoted by $G'D1$ and $G'D2$. The vertices and edges are then merged using the following formula.

$$(G'_{D1}, G'_{D2}) = \text{Set}(G_{D1}, G_{D2}) = \begin{cases} V_{D1} \cup V_{D2} \\ E_{D1} \cup E_{D2} \end{cases} \quad (3)$$

Formula (1) is completely converted into

$$G_i = \text{Integrate}(\text{Set}(G_{D1}, G_{D2})) \quad (1)'$$

We suppose that a graph G_A has three vertices labelled “A”, “B”, and “C”, whereas graph G_B has “B”, “C”, and “D”. Furthermore, G_A has two edges $e_{A \rightarrow B}$ and $e_{B \rightarrow C}$ and G_B has three edges $e_{B \rightarrow C}$, $e_{B \rightarrow D}$, and $e_{C \rightarrow D}$. Then, the integrated network has four vertices and 4 edges shown in Fig 2.

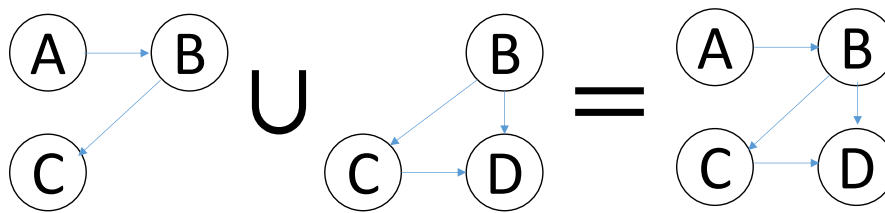


Fig. 2. Example of two graph by union operation

Using the integration function, we can determine much information. Assuming that multiple companies enter a single market, we can visualize their documents. Moreover, frequent and attractive keywords regarding this market can be extracted and summed up. We can also extract the gaps between companies by the differences in word frequency. Furthermore, we can determine emergent patterns as per the calculation of edge coefficients by quotient computation.

3.3 Integration Representation

In the final step, the integrated information is presented. Such information can be represented by variables such as color and figure. Color, in particular, is more useful.

4 CASE STUDY

Figure 3 shows a part of text comparison visualization of the topic “Beijing Olympic” between two newspapers in Japan. This figure is outputted according to the principle strategy by improving FACT-Graph[7]. Integration function is maximum function, which apply for the attribute of term frequency and document frequency, TF-IDF index. And this figure shows the difference between the two as shape and color (shape: the degree of document frequency and term frequency; red (blue) color elements: the specific elements for one or another newspapers about TF-IDF; otherwise elements; common elements between the two). We can check the difference between the two in the figure.

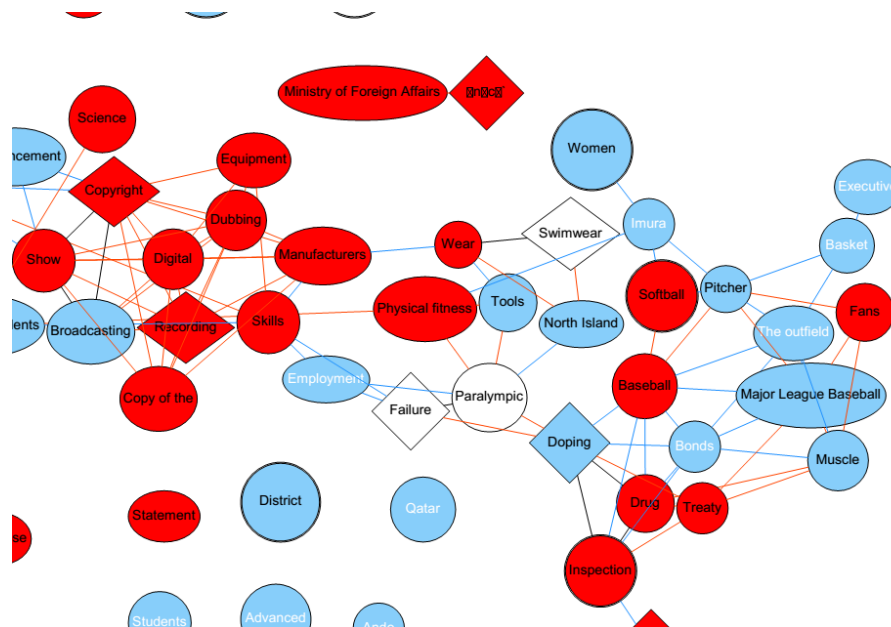


Fig. 3. Text comparison visualization between two newspapers

5. CONCLUSION

This paper described the comparison of text using network visualization. It also listed the principal strategy applied to compare text datasets.

REFERENCES

1. Manly, B. F. J., "Multivariate Statistical Methods A primer", CHAPMAN & HALL/CRC, pp. 93-106 (1986).
2. Han, J., Kamber, M., Pei, J., "Data Mining: Concepts and Techniques", Morgan Kaufmann (2011)
3. Aoki, S., Toyozumi, K., Tsuji, H., "Visualizing Method for Data Envelopment Analysis", SMC 2007, pp. 474-479 (2007)
4. Bishop, C. M., "Pattern Recognition and Machine Learning", Springer (2010)
5. Blei, D. M., Ng, A. Y., Jordan, M., "Latent Dirichlet Allocation", Journal of Machine Learning Research, Vol. 3, pp. 993-1022 (2003)
6. Ward, M., Grinstein, G., Keim, D., "Interactive Data Visualization: Foundations, Techniques, and Applications", A K Peters/CRC Press (2010)
7. Saga, R., Terachi, M., Tsuji, H., "FACT-Graph: Trend visualization by frequency and co-occurrence", Electronics and Communications in Japan, Vol. 95, No. 2, pp. 50-58 (2012)

Citizen Science in the Humanities: A Promise for Creativity

Milena Dobрева, Daniela Azzopardi

Library Information and Archive Sciences Department
Media and Knowledge Science Faculty
University of Malta, Msida 2280, Malta

`milena.dobрева@um.edu.mt, daniela.azzopardi@um.edu.mt`

Abstract. Citizen science is gaining popularity and becoming a new outlet for people who are not professionally trained to be researchers in order to contribute to a wide range of research. Bonney et al. (2009) suggested that citizen science projects differ in the type of involvement of citizens and pointed out that there are three types of projects, contributive, collaborative, and co-created. By their different nature they provide different opportunities for citizen scientists to participate in research, ranging from merely helping with trivial data collection tasks to formulating new research questions. Within the citizen research area, Humanities are still moderately present and tend to tap mainly into crowdsourcing activities which fall under the contributive project type. Our paper starts by outlining some of the key developments in citizen science; then we present some Humanities-related citizen science initiatives, and argue that citizen research in the Humanities is still under-developed compared to other domains of knowledge.

Keywords: citizen science models, crowdsourcing, motivation, activities

1 Introduction

In the Green Paper on Citizen Science commissioned by the EC [2], the titular term is defined as “general public engagement in scientific research activities when citizens actively contribute to science either with their intellectual effort or surrounding knowledge or with their tools and resources”. Although this term gained popularity recently to reflect on the engagement of unprofessional researchers and currently is associated with big groups of such contributors, it was normal that research in previous ages was normally done by people who had other professions. However today we make a clear distinction between those who chose research as their vocation and members of the general public who are willing to contribute some of their time to community or personal research. Thus despite the introductions of science as proper profession, citizen scientists are far from extinct. This is mostly evident in projects centred on the sciences, where citizens are led by professional researchers in studies which revolve mostly around observation and notation in multiple locations or across longer time spans.

adfa, p. 1, 2011.

© Springer-Verlag Berlin Heidelberg 2011

The longest running study utilizing citizen science is the Christmas Bird Count¹ which started in 1900 and this year concluded its 114th study. Volunteer birdwatchers are used to observe and note the amount of birds in an area outlined to them by the project leaders.

A similar long-continuing effort is the creation of a dictionary of Mediaeval Latin which took 101 years to complete. The effort produced seventeen volumes of work, the first of which was published in 1975 – with work on them launched as early as in 1913, and the last one published in 2013.² This project was only made possible with the help of crowdsourcing and the help of dedicated volunteers. The work will also be published online and made accessible for free thanks to a grant provided by the Packard Humanities Institute.³

The advancement of ICT, Internet and mobile technologies give new opportunities for bringing together groups of people to contribute to research. This resulted in a rapid growth of the citizen science initiatives around the globe, and subsequently in an increased body of academic publications discussing various aspects of it.

To illustrate the growing popularity of citizen science, we did a scholarly literature survey in two digital libraries (IEEE and ACM. Fig. 1 captures the change in the volumes of publications about citizen science retrieved from ACM and IEEE based on the year they were published.

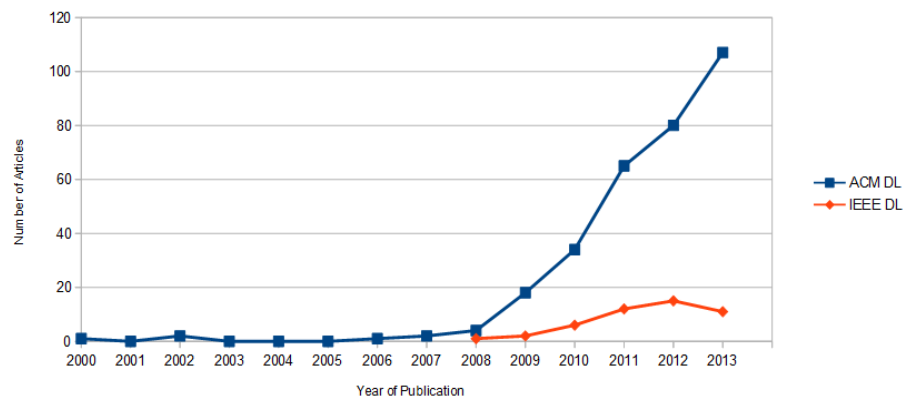


Fig. 1. ACM and IEEE DL articles containing mention of the term “Citizen Science”

There was also a substantial interest to the potential of citizen science in funding agencies who are the main source of funding of such projects. Wiggins and Crowston

¹ More information on this study can be found at <http://birds.audubon.org/christmas-bird-count>.

² <http://www.bbc.com/news/education-28952646>

³ http://www.oxfordtimes.co.uk/news/9070280.Latin_dictionary_is_a_lifetime_career

[7] who invited some 840 projects to respond to a survey about citizen science and as a result created 128 profiles of project, summarise the most popular funding sources of the projects as follows: federal and other grants – 68 projects; in-kind contributions – 31 projects, private donations – 23 projects, participant fees: 11 projects.

The potential of citizen science had been as well addressed in a number of EC-funded projects, e.g. **socientize** (<http://www.socientize.eu/>) which is working on a white paper on citizen science, and **Civic Epistemologies** (<http://www.civic-epistemologies.eu/>) which aims to develop a roadmap for citizen science use in the cultural heritage domain.

Since research is one of the most creative human activities, a still under-explored area is to what extent unprofessional researchers are involved in trivial repetitive tasks as opposed to creative activities related to research, and how citizen science could provide a creativity outlet for the members of the wide public. As an initial study related to this question we decided to explore how citizen science is used in the research in the Humanities.

2 Citizen science in the Humanities

2.1 Crowdsourcing Use in the Humanities

The application of Citizen Science in the field of Humanities has been less common than in the sciences, however one should not assume that they are not existent. Oomen and Aroyo [5] list several examples where amateur researchers and labourers contributed to process or gather data. They highlight six different types of crowdsourcing related to digital cultural heritage.

The first is **correction and transcription** where the citizen is granted access to a database of text, generally scanned manuscripts, and asked to transcribe or make correction to text which was already transcribed electronically via a computer programme. **Contextualization** happens when citizens submit data such as letters, photographs, stories or other materials in order to gather a meaningful context. Submitting data in databases with the aim of completing them or making them more sound is instead known as **complimenting collection**. **Classification** is the practice of tagging the data, or labelling it, in order to easily group similar data or locate relevant information in a short period of time. **Co-curation** seems to occur mostly with projects involving the aesthetic arts and allows the citizens to interact with institutions regarding selection activities for publication. Lastly there is **Crowdfunding** where the citizens gather together money and resources in order to support efforts initiated by others. Recently, Noordegraaf et al. suggested a model for crowdsourcing in the cultural heritage context which explores six pillars: institution, collection, goal, crowd, infrastructure, and evaluation [4].

2.2 Some Project Examples

An example of a recent humanities centred project based on data gathered via citizen science is the project “Letters of 1916”⁴. This website gathers letters to or from Irishmen submitted from people all around the world. People can also opt to translate submitted letters. This project has helped unveil various details surrounding the lifestyle of people back in those days, thus creating a new intimate perspective of the early twentieth century. One interesting feature of this project is that it develops a hybrid collection featuring letters belonging to the collections of cultural heritage institutions and letters belonging to personal archives.

The British Library also experiments in the crowdsourcing domain with its project “Georeferencing: help us place our digitized maps”⁵ where citizens are encouraged to help the British library identify their historic maps and their modern day location. Helpful users are cited and thanked. These examples as well as the aforementioned forms that citizen science can take show that these are mainly data-focused projects and that citizens are not really taking part in other stages of the research process, such as formulating research questions, choosing methodology and discussing the results.

2.3 Creativity, Citizens and Crowdsourcing in the Humanities

The range of activities to which unprofessional researchers contribute in citizen-science projects as suggested in [7] includes the following:

1. Define question
2. Gather information
3. Develop hypothesis
4. Design study
5. Data collection
6. Analyse sample
7. Analyse data
8. Interpret data
9. Draw conclusions
10. Disseminate results
11. Discuss results and ask new questions

Those activities assume different levels of creativity. The tasks of transcribing historical letters or providing geolocations would normally be considered to be quite trivial and are from the contributive type of citizen involvement as defined in [1]. Thus one research question for the future is how citizens involved in Humanities research could contribute to creative rather than trivial tasks? Furthermore, it is essential to understand what is the motivation of citizens to contribute to such projects. Some initial research on the motivation in citizen science projects in biodiversity had been done in [6] but studies in the Humanities-related citizen science initiatives are still lacking.

⁴ <http://dh.tcd.ie/letters1916/>

⁵ <http://www.bl.uk/maps/>

3 Discussion and future work

While citizen science grows in popularity in general, the majority of citizen science oriented projects take place in scientific areas. Fig. 2 captures the distribution of 47 citizen science projects across domains, which is based on research done by Franzoni and Sauermann [3].

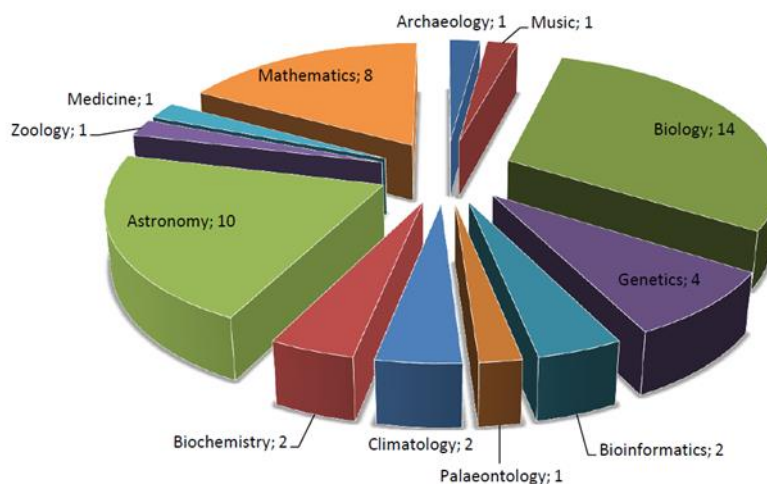


Fig. 2. Distribution of citizen science projects across disciplines based on [3]

Why citizen science is underutilised in the Humanities when it is a profitable and inexpensive means of expanding research? The currently ongoing project Civic Epistemologies is striving to develop a roadmap for the use of e-Infrastructures to aid in the inclusion of citizens in research related to cultural heritage and digital humanities. This project will use a mixed methods approach to understand the different demands and expectations of citizens and stakeholders (cultural institutions, infrastructure providers). It will combine focus groups to capture current opinions.

The project will also conduct three case studies. The first will explore the use of smartphone application based as a pocket guide for sightseeing in Coventry, United Kingdom. The second case study revolves around people who volunteer to take part in cultural activities as a way to get more involved in society and increase their sense of citizenship and love for the cultural arts. It will focus on their experience to get in touch with these cultural events, their methods, any difficulties they might have encountered and how important their cultural heritage is to them, amongst other things. The third case study revolves around location names in Ireland. As is the case with many rustic places, street names hold more than just a name to themselves, they carry with them a history, signalling important events which could have taken place in the nearby area or people who lived in the vicinity or who visited the place, amongst other things.

The combination of direct consultation with citizens through the focus groups which will be held in Malta, Sweden and the UK, and the in-depth case-will provide insights for the roadmap for e-infrastructures in order to better their integration of citizen scientists related to cultural heritage.

Another domain which can be explored further is related to the use of citizen science as an approach to use more intensively eInfrastructures offering access to cultural heritage content for educational purposes and enhancing skills of the citizens. This is an area addressed by some of the research activities of the Advanced Computing for Innovation (ACoMin) project.

With such current research we could expect that in the near future we will have better understanding how Humanities research could benefit more from citizens' contribution, and to what extent creative tasks constitute part of the motivation of citizen researchers.

Acknowledgements. The authors acknowledge the support of the EC-funded projects AComIn "Advanced Computing for Innovation", Grant 316087, funded by the FP7 Capacity Programme (Research Potential of Convergence Regions) and Civic Epistemologies (*Development of a Roadmap for Citizen Researchers in the age of Digital Culture*) funded under FP7 grant agreement 632694.

References

1. Bonne, R. et al. (2009) Public Participation in Scientific Research: Defining the Field and Assessing Its Potential for Informal Science Education. A CAISE Inquiry Group Report. Available: <http://files.eric.ed.gov/fulltext/ED519688.pdf>
2. European Commission; Green Paper on Citizen Science (2013). Available on: <http://www.socientize.eu/sites/default/files/Green%20Paper%20on%20Citizen%20Science%202013.pdf>
3. Franzoni, C., H.Sauermann (2014) Crowd Science: The Organization of Scientific Research in Open Collaborative Projects. *Research Policy*, 2014, Vol. 43 (1), pp. 1-20. Available: <http://ssrn.com/abstract=2167538>
4. Noordegraaf, J., Bartholomew A., A. Eveleigh (2014) Modeling Crowdsourcing for Cultural Heritage. *Museums and the Web 2014*, N. Proctor & R. Cherry (eds). Silver Spring, MD: Museums and the Web. Published February 17, 2014. Consulted September 10, 2014 <http://mw2014.museumsandtheweb.com/paper/modeling-crowdsourcing-for-cultural-heritage/>
5. Oomen, J., L. Aroyo (2011) Crowdsourcing in the Cultural Heritage Domain: Opportunities and Challenges. *Proceedings of the 5th International Conference on Communities and Technologies (C&T '11)*. ACM, New York, NY, USA, 138-149.
6. Rotman, D. et al. (2012) Dynamic changes in motivation in collaborative citizen-science projects. *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work (CSCW '12)*. ACM, New York, NY, USA, 217-226.
7. Wiggins, A., K. Crowston. (2012). Describing Public Participation in Scientific Research, iConference 2012 Toronto, Ontario, Canada. Available: <http://crowston.syr.edu/system/files/iConference2012.pdf>

Academic Achievement in Vocational Secondary Schools: Construction and Validation of a Test to assess potential Learning Performance

Metwaly Mabed¹ & Thomas Köhler²

¹ Faculty of Education, Suez University, Egypt, metwally.mabed@suezuniv.edu.eg

² Faculty of Education, TU Dresden, Germany, thomas.koehler@tu-dresden.de

Abstract.

The study intends to construct a valid academic achievement test for assessing students' learning performance in electrical engineering. Reviewing of experts, textbooks, and curriculum guides helped determining educational objectives. To assess learning performance a 76-item pilot test was generated. Cronbach's alpha, principle component analysis, varimax rotation, discrimination validity, item difficulty, and item discrimination indices were exploited to measure validity and reliability. Results showed test levels accounting 63.98% of total variance. Cronbach's alpha was high (.96) and correlational analyses reported that the test scores are valid indicators of student learning performance. Furthermore, the test provides an acceptable difficulty value ranging from .30 to .78, while the test introduced a sufficient discrimination value ranged from good to excellent items. Subsequently we eliminated 16 items from the pilot, leading to a final 60-item-version of the academic achievement test in electrical engineering.

Keywords: Academic achievement, Learning performance, Knowledge measurement, Test methodology, Vocational education

1 Introduction

1.1 Learning performance, academic achievement and knowledge

Learning performance links to the future career opportunities, especially, in vocational education field.¹ Therefore, it is not surprising that the academic achievement of students is a top concern of educators. Academic achievement is often used to refer to the knowledge obtained by student through a school program or curriculum. According to [2] academic achievement is defined as "*the competence of a person in relation to a domain of knowledge*" (p. 46) or the proficiency of students' performance in a certain course. From the given importance of this topic in educational context, it is necessary

¹ Authors thank the participating pupils for their active contribution. The Egyptian Government and the Free State of Saxony supported studies with grants.

to do the process of exploring, assessing, and evaluating academic achievement by a particular test. AERA, APA and NCME [3] pointed out that academic achievement tests “*are measures of academic knowledge and skills that a person acquired in formal and informal learning opportunities*” (p. 124). Other definition of academic achievement test can be found in the Dictionary of Education, which is defined as “*a test that measures the extent to which a person has acquired certain information or mastered certain skills, usually as a result of planned instruction or training*” [22]. Consequently, academic achievement tests tend to measure recent learning performance and are closely tied to particular subjects or courses [4].

1.2 Testing academic achievement

Testing academic achievement may serve three major functions [24]: a) measuring the effectiveness of learning performance at a particular point of time in an educational program; b) predicting the behavior of student across different situations; and c) assessing various psychological traits and characteristics. The information, which is collected by the achievement test, moreover, support the decisions concern the placement or diagnostic issues during instruction process. Furthermore, [23] claimed that there are wide range of reasons of using academic achievement test, some of these purposes are to: a) identify where the student falls along continuum of knowledge acquisition; b) classify students in groups according to their scores in the end of a course; c) determine the eligibility for particular educational programs; and d) measure the effectiveness of instruction process. Whether the achievement test is norm-referenced measure or criterion referenced measure, the vital issue that its results should represent accurately the evidences concerning student performance to enhance learning process as a central goal in any educational organization.

In vocational education, indeed, the precise specification of what to measure is poorly understood. The development of academic achievement test to assess students learning performance in electrical engineering is more beneficial for research and practice. Although many researchers are interested in measuring academic achievement ([5], [7], [8], [13], [19], [20], [21]), no existing measure was found to be adequate to deal with the electrical engineering content in this study. The necessity for the construction of such a test is related to the influence of achievement factor in many aspects of students’ professional skills in vocational secondary school. Therefore, the construction and development of an appropriate test to measure students’ achievement in electrical engineering will enhance the quality of learning outcomes.

2 Method

2.1 Aims, Procedures and Data Analysis

The aim of this study is twofold. The first purpose is to identify the learning performance that reflects students’ assimilation of the electrical engineering content. The sec-

ond purpose is to describe the construction of a reliable and valid an academic achievement test designed to measure students' learning performance. To achieve these goals, the researchers developed an academic achievement test in several phases by following the systematic approach and the procedures, which provided in [9], [10], [17], as well as in the light of the standards [3]. Statistical analysis was conducted using PASW (Predictive Analytics SoftWare) Version 18.0 for windows.

2.2 Development Phases

The construction of the Academic Achievement Test consisted of eight subsequent development phases as follows:

1. Clarifying the Purpose of the Academic Achievement Test [9];
2. Identifying the Educational Objectives [9], [6];
3. Performing Panel of Experts;
4. Developing Test Blueprint [9], [17], [3];
5. Determining and Generating Test Questions [17], [18];
6. Preparing Test Instructions [3];
7. Performing Panel of Experts [9];
8. Conducting the Pilot Study.

3 Findings

3.1 Test Validity

The researchers are keen to utilize a valid academic achievement test. Therefore, it was a central to check the validity of academic achievement test. However, test validity is defined as *"the process of collecting evidence to establish that the inferences, which are based on the test results, are appropriate"* ([1], p. 21). From this meaning, the validity term related to the scores, which are obtained from applying the test more than the test items itself. In the same stream, a complete description of the validity issue found in [3]: *"Validity refers to the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests. Validity is, therefore, the most fundamental consideration in developing and evaluating tests. The process of validation involves accumulating evidence to provide a sound scientific basis for the proposed score interpretations."* (p. 9). The process of assessing validity includes many actions such as asking test takers and experts about their subjective opinions on the content of the test items, as well as using factor analysis to evaluate the item internal structural [15]. In this study, the principle component's factor analysis with a varimax rotation procedure was employed to calculate the factor loadings and eigenvalues of the answers for the sample by using the statistical program PASW version 18.0. Two criteria were determined to achieve satisfactory validity of the test items. First, the factor loadings should be exceeded the minimum threshold of 0.5 as recommended [14]. Second, the items with multiple loadings should record higher load on their related level than the load value on other levels [12]. The results revealed that the factor loadings of

five-item were less than the cut-off of 0.5. Moreover, four-item were loaded greater onto another level than their corresponding level. In summary, nine items out of the 70 items were removed from the academic achievement test. Then, the factor analysis with varimax rotation was run a gain over the dataset for 61 questions. The results showed that the factor loading value of each item was at 0.5 or above and there was no cross loadings in between the test items. The first component referred to the application level and it had an eigenvalue of 20.89. The second component indicated to the knowledge level and it had eigenvalue of 9.86, while the third component referred to the analysis level and it had eigenvalue of 5.00. Finally, the fourth component pointed out to the comprehension level and it had eigenvalue of 3.22. Accordingly, all eigenvalues were more that the cut-off value of one. Moreover, it was yielded that the four levels accounted 63.98% of the total variance.

3.2 Test Reliability

Reliability as indicator was calculated during construction of the test. While test was defined as a scale describing student behavior in a specified domain, test reliability refers to the stability of this scale when the testing procedure is repeated on a population of individuals or groups [3]. Cronbach's alpha was higher than the restrictive criterion of 0.7 [11], with 0.966. Since the academic achievement test consisted of four levels (knowledge, comprehension, application, and analysis), it is not sufficient to compute the reliability for only the entire test when the score of sub-level will be used. According to standard 2.1 ([3], p. 31) *"For each total score, sub-score, or combination of scores that is to be interpreted, estimates of relevant reliabilities and standard errors of measurement or test information functions should be reported"*. Thus, the reliability for each level in the achievement test was also reported, as shows in Table 1. In regards to the knowledge level, the Cronbach's alpha was 0.967 for the 15-item and the item correlation values ranged from 0.69 to 0.88.

Table 1. Results of knowledge level reliability on the academic achievement test

Test level	# Items	Cronbach's alpha	Mean	Variance	SD
Knowledge	15	0.967	11.42	28.22	5.31
Comprehension	15	0.939	9.71	26.71	5.17
Application	22	0.958	14.68	56.43	7.51
Analysis	9	0.936	4.54	13.14	3.62

Concerning the comprehension level, the reliability coefficient alpha Cronbach was 0.939 for the 15-item and the item correlation values ranged from 0.51 to 0.77. In regards to the application level, an internal consistency analysis showed Alpha value of 0.958 for the 22-item and the item correlation values ranged from 0.52 to 0.86. Regarding to the analysis level, the Cronbach's alpha coefficient was 0.936 for the nine-item and the item correlation values ranged from 0.53 to 0.92. As a result, the items of levels

in the achievement test presented satisfaction reliability criteria.

3.3 Item Analysis

The level of items difficulty is considered further indicators that can be used to determine if the item is useful to be included in the academic achievement test or not. The difficulty level of an item refers to the proportion of students answering an item correctly. The item difficulty index, also known as the p-value, is calculated by dividing the number of students who answered the question correctly by the total number of students who answered the question [24]. Furthermore, the difficulty index can range from zero to one. While zero value indicates that no student answered the item correct, one value indicates that all students answered the item correctly. The first case points out that the item is very difficult, whereas the second result explains that the item is too easy. In the present study, the item is considered difficult when the difficulty value was less than 0.25. In addition, the item is considered to be easy when the difficulty value was greater than 0.80. The results illustrated that the test items provided acceptable difficulty values ranged from 0.30 to 0.78. On the other hand, the value item discrimination provides a suggestion of the degree to which an item correctly differentiates among the test takers on a certain domain [24]. Steps of calculating the discrimination index were described by [15]. First, students who have the highest and lowest overall test scores are sorted into two groups. The upper group which is made up of the 25%-33% who have the highest overall test scores, and the lower group which is made up of the bottom 25%-33% who have the lowest overall test scores. Subsequently, the p-value of each item for the upper and lower groups is computed. Finally, the item discrimination index is provided simply by subtract the p-values of the two groups, upper group and lower group. However, the item discrimination indices diverge from -1.00 to +1.00. The overall test scores of were arranged in the descending method. Subsequently, the highest 19 students (27% of the total number 69 of students) were included in the upper group, while the lowest 19 student (27% of the total number 69 of students) were selected in the lower group. The item discrimination indices were calculated. The items were categorized concerning their discrimination indices according to the criteria that recommended by Ebel (cited by [16]). The item with an index of discrimination of 0.40 and up is considered an excellent item, when the value ranged from 0.3 to 0.39 refers that it is a good item. Moreover, the item discrimination value of 0.2 to 0.29 is considered being acceptable, whereas the ratio from zero to 0.19 refers that test item should be revised. Moreover, the item with negative discrimination index value should be removed from the academic achievement test. Fortunately, the findings reported that all discrimination indices for the test items were positive values. Moreover, the results showed that the test items provided sufficient discrimination value ranged from the good to excellent item.

Based on the results of item analysis, 61 items was the number of test questions that were found to be valid to evaluate student achievement in electrical engineering. Conversely, the total number of the test items was 60 items. Therefore, one item was removed from the academic achievement test.

4 Discussion and Conclusions

With respect to the measurement of learning performance as means to assess academic achievement and thus the development of knowledge on an individual level empirical data is one of the key indicators. However, measurement is often difficult, appropriate tools need to be subject-specific and reliable. Due to that, the present study addresses the construction and validation of academic achievement test to measure student is learning performance in the electrical engineering. The development of the academic achievement test included a rigorous process of planning, creating, and testing. However, many of researchers explained that careful attention must be paid to the purpose of the test, because the usefulness of test interpretations depend on the agreement with which the purpose and the domain represented by the test have been explicated [3]. Therefore, the purpose, the educational content, and the target population of the academic achievement test were identified. This test aimed to assess student's learning performance in the electrical engineering for vocational secondary school in Egypt.

According to [9], one or more actions may be applied during developing an instrument such as content analysis, review of the literature, critical incidents, direct observations, expert judgment, and instruction objectives. The researchers were taken into account several activities to overcome of the problems, which might inherent in test. For that reason, the test items should be reflected the knowledge which be involved directly in the electrical engineering. Therefore, the content of electrical engineering was analyzed to identify the educational objectives. Experts review was employed to test face and content validity. The output of this stage was a 76-objective reflected five main topics in the electrical engineering. To increase the content validity of the test, the importance of each topic and the number of pages were determined during preparing the blueprint of the academic achievement test.

A 76-item was generated to assess student's learning performance in electrical engineering. The academic achievement test items covered four taxonomy levels, which are knowledge, comprehension, application, and analysis. Presenting a clear, simple and concise direction is integral part of well-constructed test items [17]. Therefore, the first draft of the academic achievement test and the directions for the test takers were given to panel of experts to judge the appropriateness of them. The pilot study was conducted to collect evidence that determine the validity and reliability issues. The characteristics of the pilot study sample were similar to the group, which the final test was intended to use with it. The results from the pilot study suggested that the academic achievement test demonstrated acceptable reliability across the test. The alpha coefficient for the whole test and all the sub-level was reasonable.

The academic achievement test showed evidence of discriminant and convergent validity. The Principle component's factor analysis as the extraction technique and a varimax rotation as the orthogonal rotation method procedure were executed by using the statistical program PASW. The results of factor analysis revealed that the application level explained the higher proportion of the variance, while the comprehension level accounted the lower percent of the variance. Moreover, the academic achievement test had a good item difficulty fall within the range of 0.30 to 0.78. Furthermore, it was also discovered that the discriminative index of the academic achievement test items

ranged from 0.41 to 0.79. Since the academic achievement test items were designed to reflect predetermined objectives, this variation in value to the items would be expected. In sum, the findings indicated that the 60-item test hold promise as a valid and reliable academic achievement test to measure of student's learning performance in electrical engineering. Finally yet importantly, the construction and development of the academic achievement test in the present study should be considered in the light of a few limitations. The academic achievement test is related to electrical engineering content in the vocational education school. Moreover, although the research sample of pilot study was appropriate for the correlation matrix and factor analysis, it was relatively small. Future research utilizing larger sample size is necessary to provide evidence regarding the test validity.

5 References

1. Adams, J. (2007). The language of assessment (2nd ed.). In M. McDonald (Ed.), *The nurse educator's guide to assessing learning outcomes* (pp. 9-26). Sudbury, MA: Jones and Bartlett.
2. Algarabel, S., & Dasi, C. (2001). The definition of achievement and the construction of tests for its measurement: A review of the main trends. *Psicologica*, 22(1), 43-66.
3. American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME) (1999). *The standards for educational and psychological testing*. Washington, DC: AERA.
4. Ariyo, A. O. (2007). Construction and validation of a general science aptitude test (GSAT) for Nigerian junior secondary school graduates. *Ilorin Journal of Education*, 27, 20-29.
5. Bayrak, B., & Bayram, H. (2010). The effect of computer aided teaching method on the students' academic achievement in the science and technology course. *Procedia - Social and Behavioral Sciences*, 9, 235-238.
6. Bloom, B. S. (Ed.). Englehard, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals, Handbook I, Cognitive domain*, New York: David McKay.
7. Carle, A., Jaffee, D., & Miller, D. (2009). Engaging college science students and changing academic achievement with technology: A quasi-experimental preliminary investigation. *Computers & Education*, 52(2), 376-380.
8. Choy, J. L., O'Grady, G., & Rotgans, J. I. (2012). Is the study process questionnaire (SPQ) a good predictor of academic achievement? Examining the mediating role of achievement-related classroom behaviours. *Instructional Science*, 40(1), 159-172.
9. Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Holt, Rinehart and Winston.
10. Downing, S. M. (2006). Twelve steps for effective test development. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of Test Development* (pp. 3-25). Mahwah, N.J.: Lawrence Erlbaum Associates.
11. Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18(1), 39-50.
12. Gefen, D., & Straub, D.W. (2005). A Practical guide to factorial validity using PLS-Graph: Tutorial and annotated example. *Communications of the Association for Information Systems*, 16(5), 91-109.
13. Haislett, J., & Hafer, A. A. (1990). Predicting success of engineering students during the freshman year. *Career Development Quarterly*, 39(1), 86-95.
14. Hulland, J. (1999). Use of partial least squares (PLS) in strategic management research: A review of four recent studies. *Strategic Management Journal*, 20(2), 195-204.

15. Kline, T. (2005). *Psychological testing: A practical approach to design and evaluation*. Thousand Oaks, California: Sage Publications.
16. Mitra, N. K., Nagaraja, H. S., Ponnudurai, G., & Judson, J. P. (2009). The levels of difficulty and discrimination indices in type a multiple choice questions of pre-clinical semester 1 multidisciplinary summative tests. *International E-Journal of Science, Medicine & Education*, 3(1), 2-7.
17. Osterlind, S. J. (1998). *Constructing test items: Multiple-choice, constructed-response, performance, and other formats* (2nd ed.). New York: Kluwer Academic Publishers.
18. Quellmalz, E., & Hoskyn, J. (1995). Classroom assessment of reasoning strategies. In G. D. Phye (Ed.), *Handbook of Classroom Assessment* (pp. 103-130). San Diego, CA: Academic Press.
19. Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, Schools, and Academic Achievement. *Econometrica*, 73, 417-458.
20. Romano, J., Wallace, T., Helmick, I., Carey, L., & Adkins, L. (2005). Study procrastination, achievement, and academic motivation in web-based and blended distance learning. *The Internet and Higher Education*, 8(4), 299-305.
21. Selçuk, G., Sahin, M., & Açıkgöz, K. (2011). The effects of learning strategy instruction on achievement, attitude, and achievement motivation in a physics course. *Research in Science Education*, 41(1), 39-62.
22. Shukla, R. (2005). *Dictionary of education*. New Delhi: APH Publishing.
23. Smith, D. R. (2001). Wechsler individual achievement test. In J. J. Andrews, D. H. Saklofske, & H. L. Janzen (Eds.), *Handbook of Psychoeducational Assessment: Ability, Achievement, and Behavior in Children* (pp. 169-193). San Diego: Academic Press.
24. Srivastav, G. N. P. (2000). *Management of teacher education: A handbook*. New Delhi: Concept Publishing Company.
25. Whiston, S. C. (2005). *Principles and applications of assessment in counseling* (2nd ed.). Belmont, CA: Thompson, Brooks Cole.

Committees and Reviewers

Conference Chairs

Susumu Kunifujii, JAIST, Japan
George Angelos Papadopoulos, University of Cyprus, Cyprus
Andrzej M.J. Skulimowski, P&BF and AGH, Poland

Local Organizing Chair

George Angelos Papadopoulos, University of Cyprus, Cyprus

Award Chairs

Susumu Kunifujii, JAIST, Japan
Thanaruk Theeramunkong, Thammasat University, Thailand

Invited Session Chairs

Tomoko Kajiyama, JAIST, Japan
Vincent Cheng-Siong Lee, Monash University, Australia
Takaya Yuizono, JAIST, Japan

Program Committee

Ateet Bhalla, Oriental Institute of Science and Technology, India
Chi-Hung Chi, CSIRO, Australia
João Clímaco, University of Coimbra, Portugal
Simon Colton, University of London, UK
Eugene Dubossarsky, Prescient, Australia
Mark Embrechts, RPI, USA
Ali Eydgahi, Eastern Michigan University, USA
John Garofalakis, University of Patras, Greece
Tessai Hayama, Kanazawa Institute of Technology, Japan
Hidehi Hayashi, Naruto University of Education, Japan
Hidehiko Hayashi, Naruto University of Education, Japan
Christopher Hinde, Loughborough University, UK
Josef Jablonsky, University of Economics Prague, Czech Republic
Ivan Jelinek, Czech Technical University, Czech
Janusz Kacprzyk, Systems Research Institute - PAS, Poland
Hideaki Kanai, JAIST, Japan
Takashi Kawaji, Ohtsuki City College, Japan
Takahiro Kawaji, Ohtsuki City College, Japan
Thomas Koehler, TU Dresden, Germany
Paul Kwan, University of New England, USA
Vincent Cs Lee, Monash University, Australia
Antoni Ligeza, AGH, Poland
Ahmad Lotfi, Nottingham Trent University, UK
Akinori Minaduk, Koshiro Prefecture University, Japan
Motoki Miura, Kyushu Institute of Technology, Japan
Kazunori Miyata, JAIST, Japan
David C. Moffat, Glasgow Caledonian University, UK
Anna Mura, University Pompeu Fabra, Spain
Toyohisa Nakada, Niigata University of International and Information Studies, Japan
Kazushi Nishimoto, JAIST, Japan
Maciej Nowak, University of Economics Prague, Czech Republic
Kok-Leong Ong, Deakin University, Australia
Francois Pachet, CSL Sony Paris, France
Robert Pearson
Przemyslaw Pukocz, AGH, Poland
Pawel Rotter, AGH, Poland
Jose L. Salmeron, University Pablo de Olavide, Spain
Jagannathan Sarangapani, Missouri University of Science and Technology, USA
Sheng Chai, Queen's University, Canada

Hsu-Shih Shih, Tamkang University, Taiwan
Mikifumi Shikida, JAIST, Japan
Marcin Skowron, Austrian Research Institute for Artificial Intelligence, Austria
Johan Suykens, K.U. Leuven, ESAT-SCD, Belgium
Ryszard Tadeusiewicz, AGH, Poland
I-Hsien Ting, National University of Kaohsiung, Taiwan
Brijesh Verma, Central Queensland University, Australia
Yongui Wang, Nanjing University, China
Michał Wozniak, Wrocław University of Technology, Poland
Fan Wu, National Chung-Cheng University, Taiwan
Takashi Yoshino, Wakayama University, Japan
Atsuo Yoshitaka, Japan Advanced Institute of Science and Technology, Japan
Takaya Yuizono, JAIST, Japan
John Zeleznikow, Victoria University, Australia
Songmao Zhang, Chinese Academy of Sciences, China
Constantin Zopounidis, Technical University of Crete, Greece

External Reviewers

Antonio Carmona, University Pablo de Olavide, Spain
Juan Manuel Berbel, University Pablo de Olavide, Spain
Jose Luis Barbero, University Pablo de Olavide, Spain