Foreword

Discovering Knowledge through Highly Interactive Information Based Systems*

The new Internet era has increased a production of digital data. The mankind had an easy way to the knowledge access never before, but at the same time the rapidly increasing rate of new data, the ease of duplication and transmission of these data across the Net, the new available channels for information dissemination, the large amounts of historical data, questionable quality of the existing data and so on are issues for information overload that causes more difficult to make decision using the right data. Soft-computing techniques for decision support systems and business intelligent systems present pretty interesting and necessary solutions for data management and supporting decision-making processes, but the last step at the decision chain is usually supported by a human agent that has to process the system outcomes in form of reports or visualizations. These kinds of information representations are not enough to make decisions because of behind them could be hidden information patterns that are not obvious for automatic data processing and humans must interact with these data representation in order to discover knowledge. According to this, the current special issue is devoted to present nine experiences that combine visualization and visual analytics techniques, data mining methods, intelligent recommendation agents, user centered evaluation and usability patterns, etc. in interactive systems as a key issue for knowledge discovering in advanced and emerging information systems.

INTRODUCTION

In current digital information age, a critical issue is how we process the knowledge behind the huge amount of information that appears every day through every aspect of life (news, social media, TV, radio, email, blogs, papers and so on). Everywhere you look, the quantity of information in the world is soaring. At the beginning of 21st century there was the estimation that the world produced one to two exabytes (billion gigabytes) of unique information every year [1], three years after the same authors updated their study and

* This work partially supported by the Spanish Educational Ministry project "oiPLE: Open Integrated Personal Learning Environment" (reference number TIN2010-21695-C02).

they estimated that this quantity had grown between 3 to 5 exabytes per year [2]. The Economist (http://www.economist.com/node/15557443) estimates that mankind created 150 exabytes of data in 2005 and 1,200 exabytes in 2010. That means lots of data in form of text, numbers, images, sounds, etc. that are deemed important by humans for different purposes.

Obviously, this digital revolution has several and great advantages involving knowledge access and dissemination due to in knowledge economy information becomes the most valuable commodity [3], especially when an open access strategy or philosophy is presented [4], but it also goes with problems regarding the capacity to process, understand and take advantage from this huge world production and accumulation of information conformed as digital or electronic documents, which is referred as the Alvin Toffler's popularized term "information overload" [5], notion that just appeared in a 1964 Bertrand Gross' book [6]. The information overload may be defined as the difficulty a person can have understanding and issue and making decisions that can be caused by the presence of too much information [7].

There are different issues that cause information overload in the digital context, including the rapidly increasing rate of new data, the ease of duplication and transmission of these data across Internet, the new available channels for information dissemination, the large amounts of historical data, questionable quality of the existing data and so on [8, 9]. This information overload affects both individuals, distracting them in the context of the work environments [10] or confusing them in their personal decisions [11, 12], and especially companies where a bad strategy in the information management (that means a bad knowledge management strategy) would lead to poorer decision making [13-15].

Technological advances can help reduce information overload and decision making process. Soft-computing based techniques such as business intelligent systems [16], data mining strategies [17], visualization techniques [18], recommender systems [19], semantic data representations [20], etc. may support better and more efficiently and effectively the decision making processes. However, decision process is a conscious and human process [21], thus the soft-computing based decision support systems and business intelligent solutions are only tools in the hands of a manager that must accomplish the organizational mission.

Visualization techniques have the ability to convey information at the high bandwidth of the human perceptual system, facilitating recognition of patterns in the information space and supporting navigation in large collections [22]. However, the challenge is not only knowledge representation in

information system but highly interactive environments that offer a deep data interaction that allows non trivial knowledge discovering and decision making by users.

The combination of the above mentioned computing techniques with visual analytics techniques, user centered evaluation and usability patterns, etc. in interactive systems is a key issue for knowledge discovering in current and future information systems.

Visual analytics means a process for providing insight into vast amounts of data that are stored by heterogeneous data sources. It iteratively collects information, preprocesses data, carries out statistical analysis, performs data mining, and uses machine learning, knowledge representation, user interaction, visual representations, human cognition [23], perception, exploration and the human abilities for decision making [24] and it has been successfully applied in different areas such as Bioinformatics [25], Geography [26], Paleoceanography [27], Medicine [28], Databases [29], Software Development [30], Ontology Engineering [31], eLearning [32] or Pharmaceutical Industry [33].

Thus, visual analytics combines the advantages of machines with the strength of humans such as analysis, intuition, problem solving and visual perception. Therefore, the human is at the heart of visual analytics [34] and human computer interaction is a key component for supporting knowledge discovery.

According to this, the current special issue is devoted to present nine experiences that combine different techniques and strategies in interactive systems as a key issue for knowledge discovering in advanced and emerging information systems.

SPECIAL ISSUE CONTENTS

In this special issue, guest editors are glad to introduce nine innovative and interesting works that present different knowledge discovering techniques in highly interactive systems.

Madkour et al., in their paper "Context-aware service adaptation: An approach based on fuzzy sets and service composition", deal with a pervasive computing approach in which they propose a user-centric and service-centric mixed solution for ubiquitous context-aware systems. This way both users-related and services-related contextual requirements are taken into account in their service composition proposal.

González-Torres et al. contribution is entitled "How evolutionary visual

software analytics supports knowledge discovery". In this paper authors define evolutionary visual software analytics as a specialization of visual analytics, which is aimed at supporting software maintenance processes by aiding the understanding and comprehension of software evolution with the active participation of users. Specifically, in this paper they discuss the implementation of an architecture based on the evolutionary visual software analytics process and how it supports knowledge discovery during software maintenance tasks.

The paper entitled "Visualization of Document Retrieval using External Cluster Relationship", by Jee et al., proposes a new method for the real-time visualization of document retrieval results via clustering and also a two-level visualization algorithm that presents the cluster centers onto a two-dimensional space using multidimensional scaling in order to illustrate the relationships among different clusters, and displays individual documents at locations determined by the external cluster relationship in low dimensional space in order to allow the comparison of individual documents.

Jaddi and Abdullah, with their paper entitled "Nonlinear Great Deluge algorithm for rough set attribute reduction," deal with attribute reduction problem in information systems that is an important step of pre-processing in data mining. They propose a modification of Great Deluge algorithm for rough set attribute reduction wherein the "Level" is nonlinear. They use the standard datasets available in a UCI machine-learning repository to examine the proposed method.

The next paper in this special issue is entitled "Recommender System based on Product Taxonomy in E-commerce Sites," by Dr. Yong Soo Kim. In this paper the author describes a novel and fast recommender system for websites based on product taxonomy and user click patterns. The effectiveness of the proposed approach was assessed by applying the F1 metric to an experimental e-commerce website.

Dr. Kwanghoon Pio Kim, in his paper entitled "Discovering activity-performer affiliation knowledge on ICN-based workflow models," focuses on a special type of organizational social network knowledge acquired from deploying workflow technologies, which is dubbed activity-performer affiliation knowledge, describing the main implications of the activity-performer affiliation knowledge, and how much it is worth discovering them in workflow-driven organizations and enterprises producing massively parallel interactions and large-scaled operational data collections.

Man et al. contribution to this special issue, entitled "A text categorization method using extended vector space model by frequent term sets," regards

to natural language processing and information retrieval, introducing a new text categorization method using frequent term sets. The novelty of their proposal is related to the introduction of a constraint measure AD-Sup to extract discriminative features from frequent term sets for classification task.

The paper "Interactive Website Filter for Safe Web Browsing," authored by Jo et al., faces the browsing security issues and introduces an interactive website filter based on heuristics for detecting malicious websites. The filter incorporates user-interaction into discovering the true identity of the suspect websites lets our filter avoid false positives caused by automatic detection.

The final paper, by Heredia et al., in their paper entitled "Interactive knowledge asset management: Acquiring and disseminating tacit knowledge," defines and validates a solution to support organizational learning through the interactive evolution of organizational knowledge. Their approach includes mechanisms to acquire the experiences of individuals working in different teams and environments, combine them with the existing company knowledge assets, and disseminate the evolved knowledge to all the project teams.

Editors are very grateful to the reviewers who kindly agreed to referee the manuscripts in a timely manner and provided valuable feedback to the authors. Finally, guest editors would like to take this opportunity to thank authors who have contributed to this special issue.

REFERENCES

1. P. Lyman and H. R. Varian, "How much information?" *Journal of Electronic Publishing*, Vol. 6, 2000, DOI: http://dx.doi.org/10.3998/3336451.0006.204.
2. P. Lyman and H. R. Varian, "How much information? 2003," University of California at Berkeley, 2003, http://www2.sims.berkeley.edu/research/projects/how-much-info-2003/.
3. P. Hemp, "Death by information overload," *Harvard Business Review*, Vol. 87, 2009, pp. 83-89.
4. F. J. García-Peñalvo, C. García de Figuerola and J. A. Merlo, "Open knowledge: Challenges and facts," *Online Information Review*, Vol. 34, 2010, pp. 520-539.
5. A. Toffler, *Future Shock*, Random House, New York, 1970.
6. B. M. Gross, *The Managing of Organizations: The Administrative Struggle*, Free Press of Glencoe, New York, 1964.
7. C. C. Yang, H. Chen and K. Hong, "Visualization of large category map for Internet browsing," *Decision Support Systems*, Vol. 35, 2003, pp. 89-102.

8. R. Swenson, *Margin: Restoring Emotional, Physical, Financial and Time Resources to Overloaded Lives*, NavPress, Colorado, 2004.
9. D. Allen and M. Shoard, "Spreading the load: Mobile information and communications technologies and their effect on information overload," *Information Research*, Vol. 10, 2005, http://informationr.net/ir/10-2/paper227.html.
10. C. Speier, J. S. Valacich and I. Vessey, "The influence of task interruption on individual decision making: An information overload perspective," *Decision Sciences*, Vol. 30, 1999, pp. 337-360.
11. J. Jacoby, D. E. Speller and C. K. Berning, "Brand choice behavior as a function of information load: Replication and extension," *Journal of Consumer Research*, Vol. 1, 1974, pp. 33-42.
12. A. Paul and H. K. Herbig, "The effect of information overload on the innovation choice process: Innovation overload," *Journal of Consumer Marketing*, Vol. 11, 1994, pp. 45-54.
13. W. K. Jackson, "Information overload and managerial roles: A naturalistic study of engineers," Doctoral dissertation, University of Texas at Austin, 2000.
14. I. Mulder, H. dePoot, C. Verwijs, R. Janssen and M. Bijlsma, "An information overload study: Using design methods for understanding," in *Proceedings of OZCHI*, 2006, pp. 245-252.
15. E. A. Carlevale, "Exploring the influence of information overload on middle management decision-making in organizations," UMI 3437432, ProQuest LLC, Ann Arbor, MI, 2010.
16. Z. Jourdan, R. K. Rainer and T. E. Marshall, "Business intelligence: An analysis of the literature," *Information Systems Management*, Vol. 25, 2008, pp. 121-131.
17. R. O. K. Rupnik, M. Kukar and M. Krisper, "Integrating data mining and decision support through data mining based decision support system," *Journal of Computer Information Systems*, Vol. 47, 2007, pp. 89-104.
18. D. A. Keim, F. Mansmann, J. Schneidewind and H. Ziegler, "Challenges in visual data analysis," in *Proceedings of the Conference on Information Visualization* (IV '06), IEEE Computer Society, Washington, DC, USA, 2006, pp. 9-16, DOI: http://dx.doi.org/10.1109/IV.2006.31.
19. F. J. Martín, J. Donaldson, A. Ashenfelter, M. Torrens and R. Hangartner, "The big promise of recommender systems," *AI Magazine*, Vol. 32, 2011, pp. 19-27.
20. F. J. García-Peñalvo, R. Colomo-Palacios, P. Soto-Acosta and E. Serradell-López, E. "SemSEDoc: Utilización de tecnologías semánticas en el aprovechamiento de los repositorios documentales de los proyectos de desarrollo de software," Information Research, Vol. 16, 2011, http://Infor-

mationR.net/ir/16-4/paper504.html.
21. A. Steward III, "Leader influences on the decision making process of executives level leaders in technology," Doctoral dissertation, University of Phoenix, Arizona, 2005.
22. J. Thomas, P. Cowley, O. Kuchar, L. Nowell, J. Thomson and P. C. Wong, "Discovering knowledge through visual analysis," *Journal of Universal Computer Science*, Vol. 7, 2001, pp. 517-529.
23. A. Drigas, L. Koukianakis and Y. Papagerasimou, "Towards an ICT-based psychology: E-psychology," *Computers in Human Behavior*, Vol. 27, pp. 1416-1423.
24. X. Llorá, K. Sastry, F. Alías, D. E., Goldberg and M. Welge, "Analyzing active interactive genetic algorithms using visual analytics," in *Proceedings of the 8th Annual Conference on Genetic and Evolutionary Computation*, ACM, 2006, pp. 1417-1418.
25. E. Baehrecke, N. Dang, K. Babaria and B. Shneiderman, "Visualization and analysis of microarray and gene ontology data with treemaps," *BMC Bioinformatics*, Vol. 1, 2004, pp. 12-14.
26. G. Andrienko, N. Andrienko, P. Jankowsk, D. Keim, M.-J. Kraak, A. MacEachren, and S. Wrobel, "Geovisual analytics for spatial decision support: Setting the research agenda," *International Journal of Geography Information Science*, Vol. 21, 2007, pp. 839-857.
27. R. Therón, "Visual analytics of paleoceanographic conditions," in *Proceedings of 2006 IEEE Symposium on Visual Analytics Science and Technology*, IEEE VAST, 2006, pp. 19-26.
28. C. Tominski, P. Schulze-Wollgast and H. Schumann, "Visual methods for analyzing human health data," in *Encyclopedia of Healthcare Information Systems, Information Science Reference, Hershey*, PA, USA, 2008, pp. 1357-1364.
29. B. Shneiderman, "Extreme visualization: Squeezing a billion records into a million pixels," in *Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data*, ACM, 2008, pp. 3-12.
30. A. González-Torres, F. J. García-Peñalvo and R. Therón, "Human-computer interaction in evolutionary visual software analytics," *Computers in Human Behavior*, In Press, http://dx.doi.org/10.1016/j.chb.2012.01.013.
31. F. J. García-Peñalvo, R. Colomo-Palacios, J. García and R. Therón, "Towards an ontology modeling tool. A validation in software engineering scenarios," *Expert Systems with Applications*, Vol. 39, 2012, pp. 11468-11478, DOI: http://dx.doi.org/10.1016/j.eswa.2012.04.009.
32. D. A. Gómez, R. Therón and F. J. García-Peñalvo, "Semantic spiral time-

line as a support for e-learning," *Journal of Universal Computer Science*, Vol. 15, 2009, pp. 1526-1545.
33. J. Saffer, V. Burnett, G. Chen and P. van der Spek, "Visual analytics in the pharmaceutical industry," *IEEE Computer Graphics and Applications*, Vol. 24, 2004, pp. 10-15.
34. A. Dix, M. Pohl and G. Ellis, "Perception and cognitive aspects," Mastering the information age solving problems with visual analytics, Eurographics Association, Goslar, Germany, 2010, pp. 109 130.

• Guest Editors •

Francisco José García-Peñalvo received the Ph.D. degree in Informatics, Salamanca University in 2000. He is working in Informatics Department of the Salamanca University, Spain and leads the researching group GRIAL (research GRoup in InterAction and eLearning). He is member of the Science Education Research Institute. His current research interests include human-computer interaction, software engineering, semantic web, visual analytics, services oriented systems and technology management applied to educational contexts, systems and institutions.

Ricardo Colomo-Palacios is an Associate Professor at the Computer Science Department of the Universidad Carlos III de Madrid. His research interests include applied research in Information Systems, software project management, people in software projects and social and Semantic Web. He received his PhD in Computer Science from the Universidad Politécnica of Madrid. He also holds a MBA from the Instituto de Empresa. He has been working as Software Engineer, Project Manager and Software Engineering Consultant in several companies including Spanish IT leader INDRA. He serves as Editorial Board Member and Associate Editor for several international journals and conferences and Editor in Chief of International Journal of Human Capital and Information Technology Professionals.

Jane Yung-Jen Hsu is a Professor of Computer Science and Information Engineering at National Taiwan University. Her research interests include intelligent multi-agent systems, data mining, service oriented computing and web technology. Prof. Hsu is on the editorial board of the International Journal of Service Oriented Computing and Applications (published by Springer). She has served on the editorial board of Intelligent Data Analysis - An International Journal (published by Elsevier and IOS Press) and the executive committee of the IEEE Technical Committee on E-Commerce. She is actively involved in many key international conferences as organizers and members of the program committee. She is a member of AAAI, IEEE, ACM, Phi Tau Phi, and has been an executive committee member of TAAI.