Contents lists available at ScienceDirect
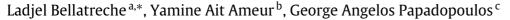
Future Generation Computer Systems

Editorial

Models and data engineering

Ladjel Bellatreche [a],[*], Yamine Ait Ameur [b], George Angelos Papadopoulos [c]

[a] LIAS/ISAE-ENSMA - Poitiers University, Poitiers, France
[b] IRIT/ENSEEIHT, Toulouse, France
[c] Department of Computer Science, University of Cyprus, Cyprus

ABSTRACT

Data and models are two well established communities that are continuously contributing in new challenges in different research domains including cyber-physical systems [1], cloud computing [2], service oriented applications, social networks [3], big data (with its five Vs characteristics: Volume, Variety, Velocity, Veracity and Value) [4], etc. The success story of data and models communities is mainly based on the availability of foundations relying on formal methods [5], modelling methods [6], storage systems and platforms [7], advanced optimization structures, benchmarking, scalability, etc. These foundations are usually associated with tools and commercial and academic systems.

The selected papers for this special issue address a variety of topics and concerns in models and data fields, including advanced databases, engagement systems, embedded and complex systems, etc.

© 2016 Published by Elsevier B.V.

1. Introduction

We welcome you in this special issue of Future Generation Computer Systems dedicated to the best papers presented at the fourth International Conference on Model and Data Engineering (MEDI) that was held in Larnaca, Cyprus, in September 2014. MEDI, initiated by researchers from Euro-Mediterranean countries, aims at promoting the creation of north-south scientific networks, projects and faculty/student exchanges as well as of other parts of the world. MEDI promotes the interaction and collaboration of research communities issued from the system modelling on the one hand and from the data modelling on the other hand. However, the explosion of the volume of data, models and processes within organizations has brought up new requirements and research issues, in particular to cope with problems related to advanced application design [8], complex systems [9], semantic modelling [10], formal methods [5], Business Intelligence [6], Big Data [4], user centric-applications [6], etc.

Each year, MEDI invites world known researchers from both model and data domains to give talks. For MEDI'2014 edition, two keynotes speakers were invited. Dominique Méry from the University of Lorraine, France, gave a talk entitled "Playing with State-Based Models for Designing Better Algorithms centered towards formal modelling". Also Mukesh Mohania, IBM Distinguished Engineer, Member IBM Academy of Technology, and ACM Distinguished Scientist, from IBM Australia gave a talk entitled "Some Issues in Modelling User Behaviour Data in Systems of Engagement", reporting the progress achieved within user data modelling.

The call for papers of MEDI'2014 attracted 64 submissions covering both model and data engineering activities. These papers range on a wide spectrum covering fundamental contributions, applications and tool developments and improvements. Each paper was reviewed by at least three reviewers and the programme committee accepted 16 long papers and 12 short papers leading to an attractive scientific programme. Out of the 16 full papers, we selected 8 papers to be invited for the special issue in the Future Generation Computer Systems. Due to the quality and timely presented topics of the keynote speakers, we invited them to extend their work to this special issue, which represents a total of 10 invited papers. After a second round of reviews, we finally accepted 6 papers.

We congratulate the authors of these 6 papers and thank all authors who submitted articles to MEDI 2014.

In general, research papers presented at MEDI 2014 covered the most recent and relevant topics in the areas of:

- Model Driven Engineering including transformation techniques;
- Ontology Engineering and its contributions in designing and exploiting trajectory data warehouses;

- Reuse and annotation of formal models to ease the development of complex systems and increase their quality;
- Big Data and Machine Learning techniques for user-centric applications;
- Performance and scalability of very large scale data and model stores.

2. Issue contents

Of the six papers in this issue, three are related to modeling activities and three are concerned with data management, including three cases studies related to transportation, eduction and marine mammals tracking applications, namely seal trajectories.

The first paper, "Playing with State-Based Models for Designing Better Algorithms", by Dominique Méry who gave a keynote at MEDI'2014 is modelling related [11]. It presents a novel view on the use of formal methods and the interest of formal model annotation in order to increase the quality of the design models. Quality is represented here by the capability to guarantee program and system properties. By properties, the paper intends static properties related to data and behavioral properties related to pre-conditions, post-conditions, invariants, and temporal properties. The author gives an interesting view of the state of the art related to the annotation of formal models of software and more generally of systems. The Event-B method together with its supporting tool Rodin (http://www.event-b.org/) is deployed to formally model the concepts introduced in this paper.

The second paper, "Modeling User Behavior Data in Systems of Engagement", by Oliver Bent, Prasenjit Dey, Komminist Weldemariam and Mukesh K. Mohania (who gave a keynote at MEDI'2014), proposes a framework for modelling user behaviour in the context of personalized education. The goal of the framework is to model student behavior information, captured from different activities that the student performs during the process of learning [12]. The framework first describes the characteristics of the user interaction, affective and cognitive data and maps them to Big Data characteristics (5Vs for volume, velocity, variety, veracity and value). The framework proposes then to analyse the user cognition and behavior data. The proposed system stores the user interaction data as JSON objects in IBM Cloudant and processes data using MapReduce functions. A large scale running pilot/experiment has been used to perform experiments to analyze user behavior for a class of students and their teachers.

The third paper, "Landmark Selection for Spectral Clustering based on Weighted PageRank", by Dimitrios Rafailidis, Eleni Constantinou and Yannis Manolopoulos, presents an accelerated spectral clustering method, using a landmark selection strategy. Based on the weighted *PageRank* algorithm [13], the most important nodes of the data affinity graph are selected as landmarks and are provided to a landmark spectral clustering technique to obtain scalable clustering. The paper gives the mathematical formulation of the problem and presents experiments with two benchmark face and shape image data sets, using several landmark selection strategies for spectral clustering that either ignores or considers the topological properties of the data in the affinity graph. The results show that the proposed method outperforms baseline and accelerated spectral clustering methods, in terms of computational cost and clustering accuracy, respectively.

The fourth paper, "A Model-Driven Approach for Developing a Model Repository: Methodology and Tool Support", by, Hamid Brahim, presents a model-based methodology to develop a repository of models, to promote reuse and to interconnect the process of models' specification and the system development with models are defined in different areas, in particular the embedded systems area [14]. One of the main characteristics of reuse is its ability to cut software development time and costs. Meta-modelling and model transformation techniques are widely used for the specification of the repository structure, interfaces and the stored modelling artifacts. Another important contribution of this paper is the development of a model-driven engineering tool-chain supporting the methodology and the deployment and operational use of the repository of models. A proof of concept through a use case from the railway domain is given to show the effectiveness of this approach.

The fifth paper, "Trajectory ontology inference considering domain and temporal dimensions. Application to marine mammals", by Rouaa Wannous, Jamal Malki, Alain Bouju and Cecile Vincent, proposes a modelling approach based on ontologies to build a trajectory ontology [15]. The ontology model is composed of a general trajectory domain model, a domain knowledge and a temporal domain model. The approach defines a set of reasoning rules used on the ontology model. The thematic rules are based on the domain trajectory activities and the temporal rules are based on Allen relationships. In order to enhance the reasoning mechanism, the approach defines a two-tier inference refinement on trajectory data, distinguishing two distinct operations primary and secondary inference operations. The primary filter is applied to the captured data to classify them into a set of interested places, called area-restricted search (ARSs). The latter computes the inference mechanism considering the ARS. The approach is implemented using the framework of *Oracle Semantic Data Store*.

The sixth paper, "Advantages of Complex SQL Types in Storing XML Documents", by Kai Schweinsberg and Lutz Wegner proposes a novel approach to solving the problem of mapping XML documents into a relational database using new SQL features [16]. The authors review the interaction between XML documents and SQL of three most popular commercial database management systems (DBMSs): Informix, DB2 and Oracle. An excellent survey on the current database technology in the studied area, which covers XML documents, in terms of both storage requirements and retrieval from databases is provided. A new mapping method (bijective) between XML documents and complex SQL types and an implementation prototype are also given. The rationale for the choice of *Informix* as the implementation vehicle is well documented. The prototype uses standard XML files and a performance comparison between the Informix prototype and an Oracle DBMS implementation is provided.

We hope readers will find the content of this special issue interesting and will inspire them to look further into the challenges that are still ahead before designing advanced data applications and complex software/systems. We would like to thank all the authors who submitted their papers to this special issue. In addition, we are grateful for the support of various reviews that ensured the high quality of this special issue. Last but not least, we would like to thank Professor Peter Sloot, Editor-In-Chief of the Future Generation Computer Systems Journal, Elsevier, for accepting our proposal of a special issue focused on models and data engineering and promoting MEDI conference, and for assisting us whenever required. We would like to thank very much Xu Hilda for his endless help and support. The complete International Program Committee of this special issue is listed next.

International Program Committee

- Yamine Ait Ameur, ENSEEIHT/IRIT, Toulouse, France
- Idir Ait Sadoune, LRI - CentraleSupélec, University of Paris Saclay, France
- Ladjel Bellatreche, ISAE-ENSMA, Poitiers, France
- Rafael Berlanga, Universitat Jaume I, Spain
- Azedine Boulmakoul, University of Hassan II Mohammedia, Morocco
- Sebastian Breß, German Research Center for Artificial Intelligence, Berlin, Germany

- Alain Crolotte, Teradata Corporation, USA
- Selma Khouri, ESI, Algiers, Algeria
- Yannis Manolopoulos, Aristotle University of Thessaloniki, Greece
- Atif Mashkoor, Software Competence Center Hagenberg, Austria
- Dominique Mery, LORIA, Nancy, France
- Anirban Mondal, Shiv Nadar University, India
- Carlos Ordonez, University of Houston, USA
- Yassine Ouhammou, ISAE-ENSMA, Poitiers, France
- Dimitris Sacharidis, Technische Universitat Wien, Vienna, Austria
- Guandong Xu, University of Technology Sydney, Australia

References

[1] H. Yoo, T. Shon, Challenges and research directions for heterogeneous cyber-physical system based on IEC 61850: Vulnerabilities, security requirements, and security architecture, Future Gener. Computer Systems 61 (2016) 128–136.

[2] Y. Jararweh, M. Al-Ayyoub, A. Darabseh, E. Benkhelifa, M.A. Vouk, A. Rindos, Software defined cloud: Survey, system and evaluation, Future Gener. Computer Systems 58 (2016) 56–74.

[3] H. Yahyaoui, Z. Maamar, E. Lim, P. Thiran, Towards a community-based, social network-driven framework for Web services management, Future Gener. Computer Systems 29 (6) (2013) 1363–1377.

[4] C. Hsu, Intelligent big data processing, Future Gener. Computer Systems 36 (2014) 16–18.

[5] H. Hassan, L.T. Yang, H. Zhang, M.D. Santambrogio, Special Issue on: Multicore and many-core architectures for future generation embedded systems, Future Gener. Computer Systems 56 (2016) 169–170.

[6] Y.A. Ameur, L. Bellatreche, G.A. Papadopoulos (Eds.), 4th International Conference on Model and Data Engineering -(MEDI), in: Lecture Notes in Computer Science, vol. 8748, Springer, 2014.

[7] J. Liu, D. Feng, Y. Hua, B. Peng, Z. Nie, Using provenance to efficiently improve metadata searching performance in storage systems, Future Gener. Computer Systems 50 (2015) 99–110.

[8] J. Gubbi, R. Buyya, S. Marusic, M. Palaniswami, Internet of Things (IoT): A vision, architectural elements, and future directions, Future Gener. Computer Systems 29 (7) (2013) 1645–1660.

[9] F. Zhu, Y. Yao, W. Tang, D. Chen, A high performance framework for modeling and simulation of large-scale complex systems, Future Gener. Computer Systems 51 (2015) 132–141.

[10] D. Gil, I. Song, Modeling and management of big data: Challenges and opportunities, Future Gener. Computer Systems 63 (2016) 96–99.

[11] D. Mery, Playing with state-based models for designing better algorithms, Future Gener. in Computer Systems 68 (2016) 445–455.

[12] O. Bent, P. Dey, K. Weldemariam, M.K. Mohania, Modeling user behavior data in systems of engagement, Future Generation Computer Systems 68 (2016) 456–464.

[13] D. Rafailidis, E. Constantinou, Y. Manolopoulos, Landmark selection for spectral clustering based on weighted pagerank, Future Gener. Computer Systems 68 (2016) 465–472.

[14] H. Brahim, A model-driven approach for developing a model repository: Methodology and Tool Support, Future Gener. Computer Systems 68 (2016) 473–490.

[15] R. Wannous, J. Malki, A. Bouju, C. Vincent, Trajectory ontology inference considering domain and temporal dimensions. Application to marine mammals, Future Gener. Computer Systems 68 (2016) 491–499.

[16] K. Schweinsberg, L. Wegner, Advantages of complex SQL types in storing XML documents, Future Gener. Computer Systems 68 (2016) 500–507.