GUEST EDITORIAL PREFACE

Special Issue on Modeling for Data-Intensive Computing (MoDID'12)

Juan Trujillo, Department of Software and Computing Systems, University of Alicante, Alicante, Spain

Il-Yeol Song, College of Computing and Informatics, Drexel University, Philadelphia, PA, USA

David Gil, Department of Computing Technology and Data Processing, University of Alicante, Alicante, Spain

The aim of the International Workshop on Modeling for Data-Intensive Computing is to bring together researchers, developers and practitioners to discuss research issues and experience in modeling, developing and deploying systems to deal with Big Data. The first International Workshop on Modeling for Data-Intensive Computing (MoDID'12), held in Florence, Italy, October 15-18, 2012 is presented with the aim to attract papers on the latest and best proposals for modeling and managing Big Data in this new data-drive paradigm. Papers focusing on novel applications and using conceptual modeling approaches for any aspects of Big Data such as MapReduce, Hadoop and its ecosystems, big data analytics, social networking, security and privacy and data science approaches were highly welcomed. Therefore, the workshop has been an international forum for researchers and practitioners who are interested in the different facets related to the use of the conceptual modeling approaches for the development of next generation applications based on Big Data.

MoDIC'12 attracted papers from 9 different countries distributed all over the world: Brazil, China, Cuba, France, Poland, Spain, Sweden, Switzerland and USA, receiving 12 papers. The Program Committee selected only four papers, making an acceptance rate of 33.3%. Authors of best papers were invited to extend their papers and re-submit them for this special issue. These extended papers had two more rounds of reviews where reviewers performed thorough reviews paying special attention on the new material. In addition, an invited paper entitled "*A Distributed System for The Management of Fine-grained Provenance*" complete the papers included in this special issue. In summary, in this special issue, the first two papers are focused on Big Data: general issues and modeling approaches; then, the invited paper and the third one cover Ontologies and Conceptual Models. In the following, we summarize these selected papers:

The first paper, *"Modeling and Implementing Scientific Hypotheses"*, by Fabio Porto, Ana Maria de C. Moura, Bernardo Gonçalves, Ramon Costa, Hermano Lustosa, and Frederico Corrêa da Silva, relies on the fact that data needs to be structured and managed in a way that readily makes sense to scientists, so that relevant knowledge may be extracted to contribute to the scientific investigation process. The authors propose a conceptual model to assist scientists in representing concepts like phenomena and hypotheses, and also describe a means to store results of experiment simulations with their corresponding provenance metadata. A cardiovascular numerical simulation illustrates the applicability of the model and an initial implementation using SciDB is discussed.

The second paper, entitled *"A novel multidimensional approach to integrate Big Data in Business Intelligence"*, by Alejandro Maté, Hector Llorens, Elisa de Gregorio, Roberto Tardío, David Gil, and Rafa Muñoz-Terol, argues that the huge amount of information available and its heterogeneity have surpassed the capacity of current data management technologies. In this paper, authors present an approach aimed to enable OLAP queries over different, heterogeneous, data sources. This approach is based on a MapReduce paradigm, which integrates different formats into the recent RDF Data Cube format. The benefits of this approach are that it is capable of querying different sources of information, while maintaining at the same time, an integrated, comprehensive view of the data available. In addition, the approach is evaluated in detail by means of a case study.

The invited paper, "*A Distributed System for The Management of Fine-grained Provenance*", written by Salmin Sultana and Elisa Bertino, focuses on a conceptual model to represent the provenance of data objects at different abstraction layers. The paper introduces the requirements of a provenance model and presents the authors' approach, in which the main benefit is that it is a unified provenance infrastructure. They illustrate the utility of the model in real world data processing systems. In addition, the authors introduce a data provenance distributed middleware system composed of several different components and services that capture provenance according to that model and securely stores it in a central repository.

The third paper, "*Discovering and Analysing Ontological Models from Big RDF Data*", by Carlos R. Rivero, Inma Hernández, David Ruiz, and Rafael Corchuelo, starts with a clear problem statement: the existence of a gap between existing ontological models and RDF data in the context of the Web of Data and the Big Data management problem. The paper presents a technique to discover ontological models from raw RDF data that relies on a set of SPARQL 1.1 structural queries. This technique includes six steps to accomplish the generation of the ontological model. It is derived from these data and includes the types and properties, subtypes, domains and ranges of properties and subproperties. The experiments, with millions of triples, prove that the technique is suitable to deal with Big RDF Data. Moreover, authors argue that, as far as they know, this is the first technique to discover such ontological models in the context of RDF data and the Web of Data.

Finally, we would like to thank all the authors who revised and extended their papers for this special issue and the reviewers for their hard work in revising these extended papers twice and providing critical and useful comments that helped authors in improving their papers. Absolutely, all of them have contributed in having this special issue of a high quality. We hope the readers will enjoy reading this issue and find the content beneficial to their research.

Juan Trujillo
Il-Yeol Song
David Gil
Guest Editors
JDM

Juan Trujillo is a Full-time Professor at the Department of Software and Computing Systems in the University of Alicante (Spain) and the leader of the Lucentia Research Group. His main research topics include Business Intelligence applications, Big Data, Business Intelligence 2.0, data warehouses' development, OLAP, data mining, UML, MDA, data warehouses' security and quality. He has also participated in the official registration of different tools related to Data Warehouse modelling. He has advised 11 PhD students and published more than 150 papers in different highly impact conferences such as the ER, UML or CAiSE, and 38 papers in highly ranked international journals (JCR) such as the DKE, DSS, ISOFT, IS, or JDM. He has also been co-editor of nine special issues in different JCR journals (e.g. DKE). He has also been PC member of different events and JCR journals such as ER, CIKM, ICDE, DOLAP, DSS, JDM, or DKE, and PC Chair of DOLAP '05, DAWAK '05- '06, FP-UML '05- '09, MoDIC '12-13 and ER '13. Further information on his main research publications can be found on: http://www.informatik. uni-trier.de/~ley/db/indices/a-tree/t/Trujillo:Juan.html.

Il-Yeol Song is professor in the College of Computing and Informatics of Drexel University, Director of PhD Program in Information Studies in his college, and Deputy Director for NSF Research Center on Visual & Decision Informatics. He is also an affiliated professor of Computer Science Department of KAIST, Korea. He is an ACM Distinguished Scientist and an ER Fellow. His research interests include conceptual modeling, data warehousing & OLAP, business intelligence, CRM, object-oriented technologies, and healthcare applications. Song published over 190 peer-reviewed papers and co-edited 22 proceedings. He is a co-Editor-in-Chief of Journal of Computing Science and Engineering (JCSE) and is in an editorial board member of DKE, JDM, IJEBR, and JDFSL. He won the Best Paper Award in the IEEE CIBCB 2004. He won 14 research awards from competitions of annual Drexel Research Days. He also won four teaching awards from Drexel, including the most prestigious Lindback Distinguished Teaching Award. Dr. Song served as the Steering Committee chair of the ER conference between 2010 and 2012. He is also a steering committee member of DOLAP, BigComp, and ADFSL conferences. He served as a program/general chair of over 20 international conferences/workshops including DOLAP '98-13, CIKM '99, ER '03, FP-UML '06, DaWaK '07- '08,, DESRIST '09, CIKM '09, MoDiC '12 and MoBiD '13.

David Gil David Gil is an associated professor at the Department of Computing Technology and Data Processing at the University of Alicante, Spain. David received a PhD in Computer Science from the University of Alicante (Spain) in 2008. His research interests include Applications of Artificial Intelligence, data mining, data warehouses, multidimensional databases and Big data. He has published papers in high quality international conferences such as IJCNN, SAC, HEALTHINF, DCAI, SCAI, SAIS, etc. He has also published papers in highly cited international journals such as Expert Systems With Applications, Applied Soft Computing and Neural Networks. Gil has served as a Program Committee member of several conferences and workshops such as DAWAK, ARES and CAiSE. He is a reviewer of several journals such as Neurocomputing, Expert Systems and Soft Computing. He is also involved in the organization of several international workshops (MoDIC '12, MoBiD '13-15).