Computing (2019) 101:1391–1395
https://doi.org/10.1007/s00607-019-00726-3

EDITORIAL

Model and data engineering for advanced data-intensive systems and applications

Yassine Ouhammou[1] · Ladjel Bellatreche[1] · Mirjana Ivanovic[2] · Alberto Abelló[3]

Published online: 23 July 2019

Nowadays, data are in the core of our daily lives and several business domains including Internet of Things, e-governance, social networks, Semantic Web, etc. Designing, building, managing, and evaluating advanced data-intensive systems and applications in the era of digitalization have become a crucial necessity for companies. Modeling efforts in terms of models, languages and methods have to be deployed to follow the development of reliable and efficient data-intensive systems and applications. These efforts allow dealing with interoperability, integration, disambiguation, orchestration, assessment issues. This development has to consider the presence of emerging hardware trends such as multi-core processors and high-speed and modern memory hierarchies such as solid-state drive.

Numerous workshops, conferences, and journals susceptible to publish findings related to our topics. Usually, they do not cover topics related to both data engineering and model engineering. This special issue has been associated with the 7th International Conference on Model and Data Engineering (MEDI 2017), that was held in Barcelona, Spain from 4 to 6 October 2017. Over the past years, MEDI has become one of the few international scientific events promoting the interaction and collaboration between the models and data research communities. MEDI 2017 received 70 papers form over 30 countries. The program committee finally selected 20 full papers

✉ Yassine Ouhammou
ouhammoy@ensma.fr

Ladjel Bellatreche
bellatreche@ensma.fr

Mirjana Ivanovic
mira@dmi.uns.ac.rs

Alberto Abelló
aabello@essi.upc.edu

1 LIAS, ISAE-ENSMA and University of Poitiers, Poitiers, France

2 Department of Mathematics and Informatics, Faculty of Sciences, University of Novi Sad, Novi Sad, Serbia

3 Universitat Politècnica de Catalunya, Barcelona, Spain

and 7 short papers. The accepted papers cover a number of broad research areas on both theoretical and practical aspects of new challenges related to systems assessment, advanced information systems, and mining complex databases.

To attract good papers, we manage our special issue for Computing Journal, Springer as follows: out of the 20 full papers accepted in MEDI 2017, only the authors of six papers related to the topics of our special issue were invited to extend their papers by at least 30% new content. Also, an open call for papers has been organized and attracted three papers covering the different topics of MEDI 2017. In total, our special issue got 9 papers from ten countries: Bosnia and Herzegovina, China, Cyprus, Greece, India, Morocco, Serbia, Slovenia, Spain and Sweden. After the second round of reviews, we finally accepted six papers distributed as follows: 5 extended papers from MEDI 2017 and one from the open call. Thus, the relative acceptance rate for the papers included in this special issue is competitive. We congratulate the authors who submitted articles to MEDI 2017 and our special issue.

These papers are authored by outstanding researchers in their respective fields, and tackle various issues from different concerns, interests and applications domains. Their topics include database evolution, the physical design of traditional and RDF databases, recommender systems, semantic web, data mining, machine learning, recommender systems. These topics concern several application domains like health field and automotive systems.

It is useful to note that most of these papers were results of funded projects by national agencies such as National Natural Science Foundation of China, Spanish Ministry of Economy and Competitiveness, Ministry of Education, Science and Technological Development of the Republic of Serbia and the Slovenian Research Agency.

The presentation of the six selected papers are performed as follows: we start with the five papers selected from MEDI 2017 and we end with the paper selected from the open call.

The first paper titled "Schema Evolution and Foreign Keys: a Study on Usage, Heartbeat of Change and Relationship of Foreign Keys to Table Activity", authored by Panos Vassiliadis, Michail-Romanos Kolozoff, Maria Zerva and Apostolos V. Zarras, deals with an interesting problem that concerns the reality reflected by database developers on how they use and change the schemas to satisfy their development requirements. More precisely, the authors focus on the evolution of foreign keys in the context of schema evolution for relational databases. Indeed, the paper tackles the behavior of a database from the point of view of adding tables and adding/deleting columns that may or may not have correct foreign keys so that they can be correctly joined with other tables in the database. The authors consider the schema histories of six free, open-source databases that contain foreign keys. The study observes different "cultures" for the handling of foreign keys they do not necessarily grow in sync with table growth. The initial version of this paper has been presented in MEDI 2017 by Panos Vassiliadis as a keynote paper.

The second paper titled "Feature selection based on community detection in feature correlation networks", authored by Milos Savic, Vladimir Kurbalija, Zoran Bosnic and Mirjana Ivanovic, presents an interesting filter-based method for feature selection which is an fundamental data preprocessing step in data mining and machine learning tasks, especially in the case of high dimensional data. The authors present a novel

feature selection method based on a complex weighted network structure describing the strongest correlations among features. A feature correlation network is a weighted graph where nodes and links represent respectively features and the strongest correlations among them. This graph structure is used by the proposed method that utilizes community detection techniques to identify cohesive groups of features in feature correlation networks. A subset of features exhibiting a strong association with the class variable is selected according to the identified community structure taking into account the size of feature communities and connections within them. The proposed method is experimentally evaluated on a high dimensional dataset containing signaling protein features related to the diagnosis of Alzheimers disease and compared against seven widely used classifiers that were trained without feature selection, with feature selection by four state-of-the-art methods provided by the WEKA tool, and with feature selection by four variants of our method determined by four different community detection techniques. The obtained results show that the proposed method improves the classification accuracy of several classification models while drastically reducing the dimensionality of the dataset.

The third paper titled "Bulk-loading and Bulk-Insertion Algorithms for xBR+-trees in Solid State Drives", authored by George Roumelis, Athanasios Fevgas, Michael Vassilakopoulos, Antonio Corral, Panayiotis Bozanis and Yannis Manolopoulos, deals with the problem of inserting a large batch of data to a new or existing xBR+-tree, by taking advantage of the special features of Solid State Drives (SSDs). To the best of our knowledge, it is the pioneer paper dealing with this problem. An xBR+-tree is a balanced, disk-resident, Quadtree-based index for point data, which is very efficient for processing spatial queries. Bulk loading/insertion of new data is a very important operation in many applications, where updates are processed in a batch form. An interesting state-of-art of the existing studies related to bulk loading/insertion techniques and flash efficient indexes is given. Intensive experiments, using several large real and artificial datasets, are conducted to compare the proposed algorithms against non-SSD specific counterparts, to analyze their performance (Inputs/Outputs and execution time) and to study the characteristics of the resulting trees. Interesting research directions related to spatial indexing and query processing on SSDs are also discussed.

The fourth paper titled "Data Aggregation Processes: A Survey, A Taxonomy, and Design Guidelines", authored by Simin Cai, Barbara Gallina, Dag Nystrm and Cristina Seceleanu, is a survey paper that investigates the characteristics of DAP (data aggregation processes) across a variety of applications, with a particular focus on the real-time properties. The survey has inspired a taxonomy of DAP called DAGGTAX, which presents the common and variable characteristics as features. The taxonomy provides a comprehensive view of data aggregation processes for the designers. In addition, a set of design constraints and heuristics have been proposed, which can reduce the design space and guide the realization of the selected features, so that the timing constraints can be satisfied.

The fifth paper entitled "The role of collaborative tagging and ontologies in emerging semantic of web resources", authored by Sara Qassimi and El Hassan Abdelwahed proposes an approach that aims at using collaborative tagging methods and ontologies to better describe the semantic of Web resources to increase their usage, exploitation, and reuse. More concretely, this work shows the important role of linking tagging and

ontologies to annotate Web resources for a particular context. The proposed approach is described in terms of two main phases. The first phase aims to automatically extract keywords that describe the main topic of a web resource. An extension of the classifier-based approach is proposed to extract content based on main keywords and to identify a matching set of terms contained in a given ontology. The second phase aims to retrieve relevant folksonomy tags that can better summarize the content of a web resource. The ontology in this work can be viewed as a referential of keywords and relationships that may exist among them. Comprehensive experiments were conducted to validate the proposal using Medical Subject Heading as the controlled vocabulary.

The sixth paper titled "Indexing Temporal RDF Graph", authored by Li Yan, Ping Zhao, and Zongmin Ma, proposes a new approach allowing indexing large temporal RDF (Resource Description Framework) datasets to speed up their exploitation by the means of complex queries. The authors start by motivating the importance of capturing the entire histories of RDF data and the necessity of having efficient data structures allowing querying this data. Instead of naively representing temporal RDF model by a set of temporal triples, the authors propose a graph structure representation. On top of this structure, a temporal RDF index is built. An algorithm for prefixing paths, and building the B-tree indexes are largely discussed. This index is used to materialize extracted paths and consequently, it contributes to speeding up the lookup of suffixes and prefixes which are considered as important aspects of temporal queries. Experiments were conducted to compare the proposed temporal RDF graph against the nave indexing approach. An interesting discussion on the experimentation preparation, in terms of used datasets and their adaptation to the context of the study, has to be highlighted. Datasets of Lehigh University Benchmark and DBpedia are used and enriched by intervals generated randomly and added to the initial triples. The obtained results show the effectiveness and efficiency of the proposed indexes for temporal RDF graphs for various types of queries: single triple-pattern, path query, star, and interval search.

We hope readers will find the content of this special issue interesting and that it will inspire them to look further into the challenges raised by advanced Data-Intensive Systems and Applications. 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 reviewers who ensured the high quality of this special issue. Last but not least, we would like to thank Professor Schahram Dustdar, Editor-In-Chief of Computing Journal, for accepting our proposal of a special issue, supporting for a long time our Model and Data Engineering Conference, and for assisting us whenever required. We would like to thank Linda Xavier and Christine Kamper for their help and support. The complete International Program Committee of this special issue is listed below.

1 International program committee

- Idir Ait Sadoune, CentraleSupelec, France.
- Jesus M. Almendros-Jimenez, University of Almeria, Spain.
- Youness Bazhar, ASML, Netherlands.
- Djamal Benslimane, LIRIS, Lyon, France.