



JIS special issue preface

Richard Chbeir¹ · Nicolas Spyratos²

Received: 9 October 2019 / Revised: 9 October 2019 / Accepted: 10 October 2019 /
Published online: 11 November 2019
© Springer Science+Business Media, LLC, part of Springer Nature 2019

1 Introduction

The recent rapid growth and use of connected objects leads to the emergence of virtual environments composed of multiple and independent entities such as individuals, organizations, services, software and applications sharing one or several missions and focusing on the interactions and inter-relationships among them. These digital ecosystems exhibit self-organizing environments where the underlying resources mainly comprehend data management, innovative services, and computational collective intelligence. Due to the multidisciplinary nature of digital ecosystems and their characteristics, it is highly complex to study and design them. This also leads to a poor understanding as to how managing resources will empower digital ecosystems to be innovative and value-creating. The application of Information Technologies has the potential to enable the understanding of how entities request resources and ultimately interact to create benefits and added-values, impacting business practices and knowledge. This context introduces many new challenges from different theoretical and practical points of view. This special issue aims to assess the current status and technologies, as well as to outline the major challenges and future perspectives, related to the computational management of digital ecosystems. It includes 7 papers that have been selected after a very tight peer review, in which each paper has been reviewed by three reviewers. Several topics are addressed in the special issue: Digital ecosystems, Information extraction, Data management and governance.

2 Structure

In the first paper of this special issue, Fernando Ferri, Arianna D’Ulizia, and Patrizia Grifoni propose “A grammar inference approach for language self-adaptation and evolution in

✉ Richard Chbeir
rchbeir@acm.org

Nicolas Spyratos
nicolas.spyratos@lri.fr

¹ University Pau and Pays Adour, Anglet, France

² University of Paris-South, Orsay, France

digital ecosystems”. Here, effective socialization processes have been investigated for both “biotic” (human) and “abiotic” (virtual) entities also within digital ecosystems in the perspective of common and self-adaptive languages. In this paper, the authors propose an approach for socialization, language self-adaptation, and evolution that enables an effective communicative interaction among digital entities acting in a digital ecosystem. The proposed method relies on an adaptable and extensible grammatical formalism, named Digital Ecosystem Grammar (DEG). The proposed grammar enables digital entities to interpret the messages sent by other entities by using interaction, learning and evolution actions. Moreover, a grammar learning algorithm is applied to provide the self-adaptation mechanisms that allow the digital environment to adapt the interaction language according to new messages. The approach is suitable to support the characteristics of self-adaptation, context-awareness, evolvability, and semanticity of a digital ecosystem language.

The second paper is titled “A Framework for Aspect based Sentiment Analysis on Turkish Informal Texts” authored by Pinar Karagoz, Batuhan Kama, Murat Ozturk, I. Hakki Toroslu, and Deniz Canturk. It addresses the problem of online opinion sharing on various topics (such as consumer products, events or news) where users express different opinions on different features or aspects of the topics. It is possible that in a single post, a user may have a positive opinion on one aspect and a negative opinion on another aspect at the same time. Sentiment analysis methods applied on a whole text cannot capture such details, rather an overall sentiment score is generated. In aspect-based sentiment analysis, it is aimed to extract the opinions expressed for each aspect separately. Generally, a two-phase approach is used. The first phase is aspect extraction, which is detection of words that correspond to the aspects or features of the main topic or the target product. Once aspects are available, the next phase is to match aspects with correct sentiment words in the text. In this paper, the authors study and present a framework for the aspect-based sentiment analysis problem on Turkish informal texts. Particularly, they present improvements for aspect extraction as an unsupervised method, and for the second phase, they investigate enhancements for two cases: extracting implicit aspects and detecting sentiment words whose polarity depends on the aspect. They also present a tool developed to realize the proposed algorithms and to present the analysis results. In the experiments, the analysis is conducted on a collection of Turkish informal texts from an online products forum.

In the third paper titled “Crowd Sourced Semantic Enrichment (CroSSE) for Knowledge Driven Querying of Digital Resources”, Giacomo Cavallo, Francesco Di Mauro, Paolo Pasteris, Maria Luisa Sapino, and K. Selcuk Candan aim to contextual knowledge. In essence, most information sources provide factual, objective knowledge, but they fail to capture personalized contextual knowledge which could be used to enrich the available factual data and contribute to their interpretation, in the context of the knowledge of the user who queries the system. This would require a knowledge framework which can accommodate both objective data and semantic enrichments that capture user provided knowledge associated to the factual data in the database. Unfortunately, most conventional DBMSs lack the flexibility necessary (a) to prevent the data and metadata evolve quickly with changing application requirements and (b) to capture user-provided and/or crowdsourced data and knowledge for more effective decision support. In this paper, the authors present CrowdSourced Semantic Enrichment (CroSSE) knowledge framework which allows traditional databases and semantic enrichment modules to coexist. CroSSE provides a novel Semantically Enriched SQL (SESQL) language to enrich SQL queries with information from a knowledge base containing semantic annotations. The authors describe CroSSE and SESQL with examples taken from our SmartGround EU project. Safia Brinis, Caetano Traina Jr., and Agma J. M. Traina propose “Hollow-tree: a metric access method for data with missing values” in the fourth

paper. Similarity search is fundamental to store and retrieve large volumes of complex data required by many real-world applications.

A useful mechanism for such concept is the query-by-similarity. Based on their topological properties, metric similarity functions can be used to index sets of data which can be queried effectively and efficiently by the so-called metric access methods. However, data produced by various application domains and the varying data types handled often lead to missing data, hence, they do not follow the metric similarity requirements. As a consequence, missing data cause distortions in the index structure and yield bias in the query answer. In this paper, the authors propose a novel access method aimed at successfully retrieving data with missing attribute values. It employs new strategies for indexing and searching data elements, capable of handling the missing data issues when the cause is not known. The indexing strategy is based on a family of distance functions that allow measuring the distance between elements with missing values, along with a set of policies able to organize the elements in the index without causing distortions to its internal structure. The searching strategy employs fractal dimension property of the data to achieve accurate query answer while considering data with missing values part of the response. Results from experiments performed on a variety of real and synthetic data sets showed that, while other metric access methods deteriorate with small amounts of missing values, the Hollow-tree maintains a remarkable performance with almost 100% of precision and recall for range queries and more than 90% for k-nearest neighbor queries, for up to 40% of missing values.

In the fifth paper, “Skyline-based Dissimilarity of Images” is proposed by Nikolaos Georgiadis, Eleftherios Tiakas, Yannis Manolopoulos, and Apostolos N. Papadopoulos. In this paper, the aim is to capture the intrinsic dissimilarities of image descriptors in large image collections, that is to detect dissimilar (i.e., diverse) images without defining an explicit similarity or distance function. To this end, the authors adopt skyline query processing techniques for large image databases, based on their high-dimensional descriptor vectors. The novelty of the proposed methodology lies in the use of skyline techniques empowered by state-of-the-art hashing schemes to enable effective data partitioning and indexing in secondary memory, towards supporting large image databases. The proposed approach is evaluated experimentally by using three real-world image datasets. Performance evaluation results demonstrate that images lying on the skyline have largely different characteristics, which depend on the type of the descriptor. Thus, these skyline items can be used as seeds to apply clustering in large image databases. In addition, the authors observe that skyline processing using hash-based indexing structures is significantly faster than index-free skyline computation and also more efficient than skyline computation with hierarchical indexing structures. Based on these results, the proposed approach is both efficient (regarding runtime) and effective (with respect to image diversity) and therefore can be used as a base for more complex data mining tasks such as clustering.

“Safe Disassociation of Set-Valued Datasets” is proposed as the sixth paper by Nancy Awad, Bechara Al Bouna, Jean-Francois Couchot, and Laurent Philippe. Here, the authors address the notion of disassociation introduced by Terrovitis as a bucketization based anonymization technique that divides a set-valued dataset into several clusters to hide the link between individuals and their complete set of items. It increases the utility of the anonymized dataset, but on the other side, it raises many privacy concerns, one in particular, is when the items are tightly coupled to form what is called, a cover problem. Here, the authors present safe disassociation, a technique that relies on partial-suppression, to overcome the aforementioned privacy breach encountered when disassociating set-valued datasets. Safe disassociation allows the km-anonymity privacy constraint to be extended to a bucketized dataset and copes with the cover problem. The authors describe an algorithm

that achieves the safe disassociation and provide a set of experiments to demonstrate its efficiency.

The last paper of this special issue is titled “Business Information Architecture for successful project implementation based on Sentiment Analysis in the tourist sector” written by Gianpierre Zapata, Javier Murga, Carlos Raymundo, Francisco Dominguez, Javier M. Moguerza, and Jose Maria Alvarez. It addresses a practical problem related to the failure of IT projects in specialized small and medium-sized companies due to the poor control in the gap between the business and its vision. In other words, acquired goods are not being sold, a scenario which is very common in tourism retail companies. These companies buy a number of travel packages from big companies and due to lack of demand, these packages expire and become an expense, rather than an investment. To solve this problem, the authors propose to detect the problems that limit a company by re-engineering the processes, enabling the implementation of a business architecture based on sentiment analysis, allowing small and medium-sized tourism enterprises (SMEs) to make better decisions and analyze the information that most possess, without exploitation pre-knowledge. In addition, a case study was conducted using a real company, comparing data before and after using the proposed model in order to validate feasibility of the applied model.

3 Conclusion

We hope this special issue motivates researchers to take the next step beyond building models to implement, evaluate, compare and extend proposed approaches. Many people worked long and hard to help this edition become a reality. We gratefully acknowledge and sincerely thank all the editorial board members and reviewers for their timely and insightful valuable comments and evaluations of the manuscripts that greatly improved the quality of the final versions. Of course, we offer thanks to all the authors for their contribution and cooperation. Finally, we express our thanks to the editor of JIIS for his support and trust in us.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.