

Semantics-Aware Approaches to Big Data Engineering

Alfredo Cuzzocrea¹ · Timos Sellis²

Published online: 16 May 2017
© Springer-Verlag Berlin Heidelberg 2017

Big Data are gaining the momentum in the database research community, by posing novel challenges to be deeply investigated by actual and future research efforts. Among these, *big data engineering* plays, without doubts, a dominant role and it is attracting the attention of larger and larger communities. Basically, big data engineering aims at extending decades of research results proposed and developed in the context of data engineering models and techniques for (very-large) databases and non-traditional data sources, such as graph data, XML data, OLAP data, to more probing application scenarios depicted by big data management and analytics environments. To this end, new paradigms and methodologies must take into account the well-known three Vs of big data processing: *volume, velocity, variety*.

Big data engineering can take great advantages from *semantics-aware methods*, which exploit knowledge kept in (big) data to better reasoning on data beyond the possibilities offered by more traditional data-instance-oriented approaches. Exploiting semantics of data is indeed possible to devise novel data management and optimization solutions that successfully make use of derived knowledge. From this evidence, several research issues arise, ranging from *big data indexing* to *big data design models*, from *big data querying* to *big stream data processing*, from *big data mining* to *big data analytics*, and so forth.

At the application-side, a plethora of systems manage and process big data on the basis of the semantics they express, trying to take their “big picture” beyond big data instances. A blueprint case study is clearly represented by emerging *social networks* that, from a side, found on massive big data repositories and, from another side, manage and process such data via fortunate *graph-like metaphors*. This is applied, for instance, in order to extract complex *behavioral models* from big data-populated social networks, with the final goal of supporting big data analytics for decision-making purposes.

In order to fulfill innovative requirements posed by the issue of integrating semantics approaches with big data engineering principles, this special issue on “*Semantics-Aware Approaches to Big Data Engineering*” of *Journal on Data Semantics* presents a selection of papers of the 13th *International Conference on Ontologies, DataBases, and Applications of Semantics* (ODBASE 2014), held in Aمانtea, Calabria, Italy, during October 27–31, 2014. ODBASE 2014 has attracted a large number of submissions, and, after a rigorous selection process over the accepted conference papers, only five papers have been invited for submission to the *Journal on Data Semantics* special issue on “*Semantics-Aware Approaches to Big Data Engineering*”. After two rigorous review rounds, only *three* papers have been accepted for final publication in the special issue.

The aim of the special issue is to offer an innovative, modern research perspective on the issue of semantics-aware big data engineering, with particular emphasis on models, methods and techniques, by highlighting recent top-quality contributions and results in this scientific context, and, at the same, stimulating further investigation in the reference field. In the following, we provide a summary of papers contained in the special issue.

The first paper, titled “*Constructing Event Processing Systems of Layered and Heterogeneous Events with SPARQL*”,

✉ Alfredo Cuzzocrea
alfredo.cuzzocrea@dia.units.it
Timos Sellis
tsellis@swin.edu.au

¹ DIA Department, University of Trieste and ICAR-CNR, Trieste, Italy

² CSSE Department, Swinburne University of Technology, Hawthorn, Australia

by Mikko Rinne and Esko Nuutila, focuses the attention on *complex event processing systems based on SPARQL query processing*. Authors recognize that SPARQL (*SPARQL Protocol and RDF Query Language*) has been originally developed to process queries over finite-length data sets encoded as RDF (*Resource Description Framework*) graphs. On the other hand, processing of infinite data streams can be enabled through continuous incremental evaluation of an incoming event stream. *SPARQL Update* provides tools for interconnecting queries, enabling event processing applications to be constructed out of multiple incrementally processed collaborating rules. These rule networks can perform event processing on heterogeneous event structures. Heterogeneous event support combined with the capability to synthesize new events enables the creation of layered event processing networks. Starting from the so-delineate big data processing scenario, authors review the different types of complex event processing building blocks presented in the literature and show their translations to SPARQL Update rules through examples, by supporting a modular and layered approach. The interconnected examples demonstrate the creation of an elaborate network for solving event processing tasks. The performance of the example event processing network is verified on the INSTANS platform.

The second paper, titled “*Detecting User Profiles in Collaborative Ontology Engineering using a User’s Interactions*”, by Sven Van Laere, Ronald Buyl, Marc Nyssen and Christophe Debruyne, moves the attention on *collaborative ontology engineering methods*, which are now very relevant for large-scale organizations that make extensive usage of big data. The targeted methods usually prescribe a set of processes, activities, stakeholders, and the roles each stakeholder plays in these activities. On the other hand, authors recognize that (1) the stakeholder community of each ontology engineering project is different, and (2) one can observe different types of user behavior. It may thus very well be that the prescribed set of stakeholder types and roles do not suffice. If one were able to identify these user behavior types, which are named as *user profiles*, one can compliment or revisit those predefined roles. For instance, those user profiles can be used to provide customized interfaces for optimizing activities in certain ontology engineering projects. In the paper, author presents a method for discovering *different* user profiles based on the interactions users have with each other on a collaborative ontology engineering environment. The proposed approach clusters users based on the types of interactions they perform, which are retrieved from data sets that

were annotated with an interaction ontology, which has been built on top of the SIOC ontology for online communities over the Web of data. Authors provide a demonstration of the proposed method via using the database of two instances of the GOSPL ontology engineering tool. The databases contain the interactions of two distinct ontology engineering projects involving, respectively, 42 and 36 users. For each data set, author discusses the findings by analyzing the different clusters. They finally prove that different user profiles are indeed discovered, indicating that the proposed approach is viable, though more experiments are needed to validate the results and to discover patterns across ontology engineering projects.

Finally, the third paper, titled “*Design Life-Cycle Driven Approach for Data Warehouse Systems Configurability*”, by Selma Khouri and Ladjel Bellatreche, considers the specific problem of *data warehouse systems configurability*. As authors recognize, many modern software systems are designed to be highly configurable. Configurability is the ability to build consistent systems from a common architecture through selecting and synthesizing provided design elements. Indeed, configurability offers high customizability and efficient reuse strategy, but it has not enjoyed the same popularity in data warehouse (DW) design, by comparing to other types of software systems. Starting from this limitation, the paper proposes a configurability-aware approach for DW design, which allows designers to specify requirements defining suitable design options oriented to generate a customized DW. To this end, three tasks are necessary: (1) a deep understanding of the DW design life cycle analyzed by reviewing its evolutions, (2) a formalization of each design phase, and (3) an identification of the interactions between phases. This analysis determines the approach that is proposed in the paper, which contains the following components containing: (1) the *configuration model* that tailors the DW system to meet designers’ requirements; (2) the *configuration process* that produces the corresponding DW configuration. A case study providing two DW configuration examples is finally proposed.

The editors would like to express their sincere gratitude to the Editor-In-Chief of *Journal on Data Semantics*, Prof. Esteban Zimanyi, for accepting their proposal of a special issue focused on Semantics-Aware Approaches to Big Data Engineering, and for assisting them whenever required. The editors would also like to thank all the reviewers who have contributed to substantial improvement in the quality of final papers.