J Data Semant (2016) 5:1–2
DOI 10.1007/s13740-015-0054-z

EDITORIAL

Semantics Meets Big Data: Formal Models, Practical Issues, Novel Paradigms

Alfredo Cuzzocrea[1]

Published online: 19 October 2015

Big Data management is playing a major role within the database research community and is posing a plethora of challenges that are of interest for both the academic side and the industrial side of the community, still reminding that this exciting research topic has been initially born within the context of the main industrial Web players (e.g., *Google*, *Facebook*, etc.). Several research issues arise here, ranging from *representing Big Data* to *managing Big Data*, from *indexing Big Data* to *querying Big Data*, from *ranking Big Data* to *mining Big Data*, etc. As orthogonal to all the outlined research challenges, *Semantics of Big Data*, which mainly concerns with the issue of equipping Big Data with fruitable, actionable knowledge, is a first-class axiomatic problem to deal with.

Indeed, due to the enormous, large-scale size of Big Data, keeping knowledge in Big Data is really significant, due to the fact the latter one is the first step towards Big Data understanding. Several initiatives have been proposed to this end, but the research area is still in its embryonic phase; hence, a lot of work must be done in future. As an example, *Big Data Semantics* plays a leading role in next-generation *Cloud Computing architectures*, where heterogeneous platforms need to interact and interchange Big Data for a wide spectrum of modern applications, ranging from *e-life* to *e-science*, from *Web advertisement* to *social networks*, etc. As a consequence, *distributed approaches to Big Data Semantics* clearly appear to be a successful direction to be considered in next research efforts.

A special case where Big Data Semantics models and methodologies show their power is, without doubts, represented by the case of *Big Linked Data*, which are populating the Web more and more. In this specific application scenario, enriching Big Data with metadata and semantics has the direct advantage of achieving more powerful *analytics* on top of it, still simplifying data access and data integration procedures.

In order to fulfill innovative requirements posed by the issue of combining Semantics methodologies with Big Data requirements, this special issue on *Semantics Meets Big Data: Formal Models, Practical Issues, Novel Paradigms* of *Journal on Data Semantics* presents a selection of papers of the 3rd *International Conference on Model and Data Engineering* (MEDI 2013), held in Amantea, Calabria, Italy, during September 25–27, 2013. MEDI 2013 has attracted a large number of submissions, and after a rigorous selection process over the accepted conference papers only six papers have been invited for submission to the *Journal on Data Semantics* special issue on *Semantics Meets Big Data: Formal Models, Practical Issues, Novel Paradigms*. 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 and Big Data, 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 *Random Query Answering with the Crowd*, by Roberto De Virgilio and Antonio Maccioni, focuses the attention on *random data generators* that play an important role in computer science and engineering, due to the fact they aim at simulating reality in information and communication systems. Software random data gen-

✉ Alfredo Cuzzocrea
alfredo.cuzzocrea@dia.units.it

1 DIA Department, University of Trieste and ICAR-CNR, Via A. Valerio, 6, 34127 Trieste, TS, Italy

erators cannot be reliable enough for critical applications due to their intrinsic determinism, while hardware random data generators are difficult to integrate within applications and are not always affordable in all circumstances. Based on these considerations, authors present an approach that makes use of *entropic data sources* to compute the random data generation task. In particular, the proposed approach exploits the chaotic phenomena happening in the *crowd*, by extracting these phenomena from social networks since they reflect the behavior of the crowd. The resulting system, called RANDOMDB, has been implemented on top of a classical DBMS, and an experimental evaluation, where data coming from *Twitter*, *Facebook* and *Flickr* have been integrated, has been performed in order to show RANDOMDB's efficiency and flexibility over the competitor approaches. Experimental results show that these social networks are sources to generate reliable randomness, and RANDOMDB being a system that can be used for this task effectively and efficiently.

The second paper, titled *ROMULUS: The Repository of Ontologies for MULtiple USes populated with mediated foundational ontologies*, by Zubeida Casmod Khan and C. Maria Keet, moves the attention on *ontology-driven conceptual data modeling*, with particular regards to its usage for solving *interoperability issues among domain ontologies*. In this respect, authors correctly recognize that several foundational ontologies have been developed in recent years, and most of them are available in multiple versions, yet with only limited modules that could facilitate wider uptake of ontologies and ontology-based modeling. From this, the ontology interoperability problem gets worse again, and two peculiar research challenges arise: (1) increase of the need for a coordinated and structured comparison and elucidation of conceptual data modeling decisions; (2) requirement for suitable software infrastructures to address the previous need. In order to deal with these challenges, authors present the *Repository of Ontologies for MULtiple USes* (*ROMULUS*), which is the first online library of *machine-processable, modularized, aligned, and logic-based merged foundational ontologies*. In addition to the typical features of a model repository, ROMULUS also introduces a special *foundational ontology recommender* covering features of six foundational ontologies, tailor-made modules for easier reuse, and a catalog of mappable and non-mappable elements among the well-known *BFO*, *GFO* and *DOLCE* foundational ontologies.

Finally, the third paper, titled "*Producing Just Enough Documentation: An Optimization Approach*", by J. Andres Diaz-Pace, Christian Villavicencio, Silvia Schiaffino, Matas Nicoletti and Hernn Vzquez, considers the specific problem of *software architecture documentation production*. Authors start from noticing that software architectures are important assets in a *software development process*, which serves to share and discuss the main design concerns among the *project stakeholders*. The architecture must be properly documented (e.g., via a Wiki environment) in order to be effectively used by these stakeholders. However, the process of producing architecture documentation often fails to deliver contents that address the stakeholders' information needs. To address the problem, authors argue for *a novel knowledge management strategy* according to which: (1) architecture documentation is created *incrementally*, and (2) its contents are driven by a model of *stakeholder preferences*. In this paper, authors thus present an *information optimization approach* applied to the architecture documentation domain, derived from an existing documentation method called *Views and Beyond*. To this end, authors introduce the so-called *Next SAD Version Problem* (*NSVP*), and provide suitable tools capable of supporting assist architects in producing cost-effective documentation. Authors also perform a sensitivity analysis of the proposed optimization model and develop a robust formulation that takes into account uncertainty in the parameter estimations for NSVP instances, thus improving the outcomes of the introduced documentation assistant.

The Editor would like to express his sincere gratitude to the Editor-In-Chief of *Journal on Data Semantics*, Prof. Esteban Zimanyi, for accepting his proposal of a special issue focused on Semantics and Big Data, and for assisting him whenever required. The Editor would also like to thank all the reviewers who have worked within a tight schedule and whose detailed and constructive feedbacks to authors have contributed to substantial improvement in the quality of final papers.