

A Note on Intelligent Exploration of Semantic Data

Editorial

Dhaval Kumar Thakker^a, Daniel Schwabe^b, Roberto García^c, Kouji Kozaki^d, Marco Brambilla^e, Vania Dimitrova^f

^a Faculty of Engineering and Informatics, University of Bradford, Bradford, UK

^b Department of Informatics, Pontifical Catholic University of Rio de Janeiro, Brazil

^c Computer Science and Engineering Department, Universitat de Lleida, Lleida, Spain

^d Department of Knowledge Science, Osaka University, Japan

^e Data Science Lab, Politecnico di Milano, Italy

^f Faculty of Engineering and Informatics, University of Leeds, Leeds, UK

Abstract. Welcome to this special issue of the Semantic Web (SWJ) journal. The special issue compiles three technical contributions that significantly advance the state-of-the-art in exploration of semantic data using semantic web techniques and technologies.

Keywords: Data graphs, Semantic data, data exploration, summarisation, exploratory search, learning

1. Introduction

Semantic data, including structured and unstructured data linked with knowledge models representing meaning, is becoming widely available in a diversity of contexts and usages (Zaveri, Rula et al. 2016). This can make significant contribution to the implementation and the broader utilisation of semantic web in human practice. Users with various backgrounds and expectations are becoming exposed to semantic data, and need to explore such data for different purposes (searching, browsing, summarising, making sense, taking decisions, or learning) (Shneiderman 1996) (Brunetti, García 2014) (Thakker, et al. 2016). Semantic data exploration is becoming a key activity in a range of application domains, such as government and public organisations (Hendler, Holm et al. , Höffner, Martin et al. 2016), education (Dietze, Yu et al. 2012), life science (Williams, Harland et al. 2012), cultural heritage (Hyvönen 2012), media and entertainment (Kobilarov, Scott et al. 2009) to name a few. Several novel interfaces and interaction means for exploration of

semantic data are being proposed, for example semantic data browsers, ontology/content visualisation and exploration environments, semantic wikis, intelligent agents. However, the efforts have not been consolidated to capitalise on experience in tackling key issues in exploration of semantic data, and to bring together the main themes of: (a) Human Factors: the utility and the added value of exploration approaches to humans; (b) Technology: new technological strands and how these strands impact the algorithms for interaction and exploration; (c) Application Domains: domains that have seen a change because of bringing in semantic data exploration approaches, and the uncharted application domains community has yet to address; and (d) Evaluation: sharing practices on evaluating exploration approaches including gold standards, benchmarks and methodologies.

The main goal of this Special Issue is to explore new directions and approaches in the aforementioned themes to leverage innovative research aimed at tackling exploration challenges while using semantic

data. We encouraged the submission of work with important theoretical and practical results, as well as case studies on existing use of technologies and techniques for semantic data exploration.

2. Summary of Contributions

In this section, we present the summary of the papers that were accepted for publication in this special issue.

It is common practice to query different linked datasets in order to get as much information as possible on distinct topics. These datasets can have a large number of complex vocabularies and data structures making it difficult to query them for any meaningful analysis. For this reason, there is a need for systems that can produce on-demand rewritings of queries. One of the challenges in query rewriting is the preservation of semantics due to heterogeneity of the vocabularies. Hence, the quality estimation of the produced rewriting becomes crucial. In the paper titled "Estimating query rewriting quality over LOD", Torre et al present a framework that, given a query written in the vocabulary the user is more familiar with, the system rewrites the query in terms of the vocabulary of a target dataset. Moreover, their system also informs about the quality of the rewritten query with two scores: firstly, a similarity factor which is based on the rewriting process itself, and secondly, a quality score offered by a predictive model. This model is constructed by a machine learning algorithm that learns from a set of queries and their intended (gold standard) rewritings. They have validated the feasibility of their framework in a real scenario.

In the paper titled "Quality Metrics for RDF Graph Summarization", Zneika et al. tackle the issue of graph summarization from the perspective of semantic data exploration. RDF Graph Summarization pertains to the process of extracting concise but meaningful summaries from RDF Knowledge Bases (KBs) representing as close as possible the actual contents of the KB both in terms of structure and data. The authors present a quality framework that works at two levels: the level of the ideal summary of the KB that could be provided by an expert user and the level of the instances contained by the KB. For the first level, they are computing how close the proposed summary is to the ideal solution (when this is available) by defining and computing its precision, recall and F-measure against the ideal solution. For the second level, they are computing if the existing instances are covered

(i.e. can be retrieved) and in what degree by the proposed summary. Authors have tested their quality framework against three well-known RDF Graph Summarization algorithms.

The third paper of the issue, titled "Using Knowledge Anchors to Facilitate User Exploration of Data Graphs" by Al-Tawil et al. focuses on the novel concept of knowledge expansion during data exploration. Authors introduce an exploration support mechanism underpinned by the subsumption theory of meaningful learning (Ausubel 1962), which postulates that new knowledge is grasped by starting from familiar concepts in the graph which serve as knowledge anchors from where links to new knowledge are made. A core component of their framework based on subsumption theory of meaningful learning is to the automatic identification of such knowledge anchors in a data graph (KADG). Authors present several metrics for identifying KADG which are evaluated against familiar concepts in human cognitive structures. These metrics and framework is evaluated in a task-driven experimental user study compared to free data graph exploration.

We hope that the readers will find the articles of this special issue to be informative and useful.

References

- Ausubel, D. P. (1962). "A Subsumption Theory of Meaningful Verbal Learning and Retention." The Journal of General Psychology 66(2): 213-224.
- Maria Brunetti, J., & García, R. (2014). "User-centered design and evaluation of overview components for semantic data exploration." Aslib Journal of Information Management 66(5): 519-536.
- Dietze, S., et al. (2012). Linked education: interlinking educational resources and the web of data. The 27th ACM Symposium On Applied Computing (SAC-2012), Special Track on Semantic Web and Applications. Trento, Italy.
- Hendler, J., Holm, J., Musialek, C., Thomas, G. "US government linked open data: semantic. data. gov." the IEEE Intelligent Systems, 27(3): 25-31.
- Höffner, K., Martin, M., Lehmann, J. (2016). "LinkedSpending: openspending becomes linked open data." Semantic Web 7(1): 95-104.

Hyvönen, E. (2012). "Publishing and Using Cultural Heritage Linked Data on the Semantic Web." Synthesis Lectures on the Semantic Web: Theory and Technology **2**(1): 1-159.

Kobilarov, G., et al. (2009). Media meets semantic web—how the bbc uses dbpedia and linked data to make connections. European semantic web conference, Springer, 723-737, Heraklion, Greece.

Shneiderman, B. (1996) "The eyes have it: a task by data type taxonomy for information visualizations." Proceedings of the IEEE Symposium on Visual Languages. pp. 336–343. Boulder, CO, USA .

Thakker, D., Yang-Turner, F., & Despotakis, D. (2016). User interaction with linked data: An exploratory search approach. International Journal of Distributed Systems and Technologies (IJ DST), **7**(1): 79-91.

William, A. J., Harland, L., Groth, P. T., Pettifier, S., Christine, C., Willighagen, E. L., Mons, B. (2012). Open PHACTS: Semantic interoperability for drug discovery. Drug Discovery Today, **17**(21-22), 1188-1198.

Zaveri, A., Rula, A., Maurino, A., Pietrobon, R., Lehmann, J., & Auer, S. (2016). Quality assessment for linked data: A survey. Semantic Web, **7**(1), 63-93.