

Editorial

Special issue on question answering for Linked Data

Axel-Cyrille Ngonga Ngomo ^{a,b,*} and Christina Unger ^c

^a *Semantic Data Management Unit, Institute for Applied Informatics, Hainstraße 11, 04109 Leipzig, Germany*

^b *Data Science Group, University of Paderborn, Warburgerstraße 100, 33098 Paderborn, Germany*

^c *Semantic Computing Group, University of Bielefeld, Universitätsstraße 25, 33615 Bielefeld, Germany*

Abstract. This editorial summarizes the content of the special issue of the Semantic Web Journal on question answering for linked data.

Keywords: Question answering, Linked Data, natural language processing

The Data Web is now a reality for a large number of experts. With more than 10,000 datasets published according to the Linked Data principles and more than 150 billion facts¹ [3], the Data Web is now also used in a large number of applications ranging from health-care [9] to urban data management [2]. Accessing data published according to the Linked Data principles is easy for experts fluent in SPARQL.² However, these experts are a minute fraction of the potential benefactors of the Linked Data Web. One of the means to improve the access to the Linked Data Web lies in the development of natural-language interfaces that can transform human languages or even controlled languages into a representation suitable for querying large knowledge bases [10]. The goal of this special issue on question answering was to continue the ongoing efforts initiated through the Question Answering on Linked Data challenge [10] and to gather some of the newest developments all around better natural language interfaces for Linked Data.

*This work was partly supported by a grant from the EU H2020 Framework Programme provided for the project HOBBIT (GA no. 688227).

¹<http://stats.lod2.eu>.

²<http://www.w3.org/TR/sparql11-query/>.

The papers accepted in this special issue all present innovative approaches to dealing with question answering on Linked Data. [8] addresses the problem of building a knowledge base of rules for question answering system. To this end, the authors introduce an intermediate representation for questions that can be used across languages. The approach is applied to Vietnamese with high accuracies. This paper presents an alternative approach to a large number of state-of-the-art systems, which focus on English and use the large number of NLP tools available for this particular language to generate question parses and corresponding answers to these questions. Therewith, it can potentially lead the way towards novel approaches for question answering.

The authors of [4] address the problem of finding approximate answers for SPARQL 1.1 queries. The provision of solutions for this problem is of central importance when faced with zero-result queries or queries with unsatisfactory results. The authors present a framework that allows generating relaxations incrementally, making the idea of relaxation theoretically amenable to interactive applications.

GFMed [7] shows how designing specialized question answering systems can lead to high-performance question answering for the biomedical domain. The

approach presented in this work relies on a controlled vocabulary, which allows generating SPARQL queries when coupled with a corresponding grammar. Once again, the idea of language-independence is tackled as the approach is evaluated on Romanian and English.

The authors of [5] address the same problem as the paper aforementioned but rely on a different approach. Here, the authors use natural-language processing techniques to generate abstractions of questions, which are converted into SPARQL query templates. The templates are then instantiated and executed. The approach is shown to perform well on benchmark data and suggests that the way of mapping languages is still a viable option for building question answering systems.

The paper [1] addresses the problem of information reconciliation for question answering. The distributed nature of the Linked Data Web is made use of to collect and integrate information necessary to answer questions. In particular, the authors use a framework based on argumentation theory for the reconciliation and are able to provide explanations for their results. The reconciliation approach is applied to DBpedia and used to create a dataset that subsumes all chapters and that can be used for better question answering. This paper displays how improve data quality can lead to better Semantic Web applications.

While the papers presented in this special issue present a significant advance over the state of the art, current surveys suggest [6] that there is still a long ahead before achieving highly accurate question answering on RDF data. Amongst the most important challenges lie the problem of multilinguality, which remains particularly hard to tackle for languages with only few linguistic resources. Achieving user-friendly runtimes on complex queries is also still ongoing and demands improved storage and indexing solutions for the Linked Data Web. Domain-specific questions (e.g., procedural, temporal, spatial and statistical questions) demand different types of processing as dedicated semantics are needed to replicate the model of the nat-

ural language used to formulate the query into a formal language such as SPARQL. More diverse and intelligent natural-language interfaces such as dialog and recommender systems for the Linked Data Web complete this non-exhaustive set of possible improvements for the future.

References

- [1] E. Cabrio, S. Villata and A. Palmero Aprosio, A RADAR for information reconciliation in question answering systems over linked data, *Semantic Web* **8**(4) (2017), 601–617.
- [2] S. Egami, T. Kawamura and A. Ohsuga, Building urban LOD for solving illegally parked bicycles in Tokyo, in: *The Semantic Web – ISWC 2016 – 15th International Semantic Web Conference, Proceedings, Part II*, Kobe, Japan, October 17–21, 2016, 2016, pp. 291–307.
- [3] I. Ermilov, J. Lehmann, M. Martin and S. Auer, LODStats: The data web census dataset, in: *The Semantic Web – ISWC 2016 – 15th International Semantic Web Conference, Proceedings, Part II*, Kobe, Japan, October 17–21, 2016, 2016, pp. 38–46.
- [4] R. Frosini, A. Cali, A. Poulouvasilis and P.T. Wood, Flexible query processing for SPARQL, *Semantic Web* **8**(4) (2017), 533–563.
- [5] T. Hamon, N. Grabar and F. Mougin, Querying biomedical linked data with natural language questions, *Semantic Web* **8**(4) (2017), 581–599.
- [6] K. Höffner, S. Walter, E. Marx, R. Usbeck, J. Lehmann, and A.-C. Ngonga Ngomo, Survey on challenges of question answering in the semantic web, *Semantic Web Journal* (2016), 1–26.
- [7] A. Marginean, Question answering over biomedical linked data with grammatical framework, *Semantic Web* **8**(4) (2017), 565–580.
- [8] D.Q. Nguyen, D.Q. Nguyen and S.B. Pham, Ripple down rules for question answering, *Semantic Web* **8**(4) (2017), 511–532.
- [9] R. Piro, Y. Nenov, B. Motik, I. Horrocks, P. Hendler, S. Kimberley and M. Rossman, Semantic technologies for data analysis in health care, in: *The Semantic Web – ISWC 2016 – 15th International Semantic Web Conference, Proceedings, Part II*, Kobe, Japan, October 17–21, 2016, 2016, pp. 400–417.
- [10] C. Unger, C. Forascu, V. Lopez, A.-C. Ngonga Ngomo, E. Cabrio, P. Cimiano and S. Walter, Question answering over linked data (QALD-4), in: *Working Notes for CLEF 2014 Conference*, 2014.