Where is Linked Data
in Question Answering over Linked Data?

Tommaso Soru, Edgard Marx, André Valdestilhas,
Diego Moussallem, and Gustavo Publio

AKSW, University of Leipzig, Germany
{tsoru,marx,valdestilhas,moussallem,gpublio}@informatik.uni-leipzig.de

Abstract. We argue that "Question Answering with Knowledge Base" and "Question Answering over Linked Data" are currently two instances of the same problem, despite one explicitly declares to deal with Linked Data. We point out the lack of existing methods to evaluate question answering on datasets which exploit external links to the rest of the cloud or share common schema. To this end, we propose the creation of new evaluation settings to leverage the advantages of the Semantic Web to achieve AI-complete question answering.

1 Introduction

Question Answering with Knowledge Base (KBQA) parses a natural-language question and returns an appropriate answer that can be found in a Knowledge Base (KB). Currently, one of the most exciting scenarios for Question Answering (QA) is the Web of Data, a fast-growing distributed cloud of interlinked KBs which comprises more than 100 billions of edges [13]. Similarly, Question Answering over Linked Data (QALD) is a research field aimed at transforming utterances into SPARQL queries which can be executed towards the Linked Open Data (LOD) cloud [11]. QALD and KBQA are strictly related, as they both target the retrieval of answers from KBs. However, the current benchmarks and datasets available for evaluating QALD approaches are limited to an *unlinked* and *unstandardized* vision of the structured question answering task. In this paper, we point out the lack of existing methods to evaluate QA on datasets which exploit external links to the rest of the cloud. Moreover, we argue that several learning-based KBQA approaches may be very competitive in QALD challenges, as the current distinctions among their respective benchmarks are only in terms of underlying KBs. Instead, our plea is to let language experts do language and Web semantics experts do semantics. We propose the creation of new evaluation methods and settings to leverage the advantages of the Semantic Web (SW) to achieve AI-complete QA over the Web of Data [6].

2 State of the Art

2.1 Question Answering over Linked Data

The most popular datasets for QALD are collected in the homonym QALD benchmarks [11], which have been released through 9 challenge editions since 2011, the BIOASQ challenge for biomedical QA [15], and the more recent LC-QUAD [14].18 out of 29 datasets in the aforementioned benchmarks are *open-domain* (i.e., they target a knowledge graph such as DBpedia and Wikidata); 8 *domain-specific* datasets are from the biomedical domain, 3 describe music and 1 describes governmental data. QA systems are supposed to build queries aimed at retrieving information from the RDF dataset itself and, in 3 cases, associated textual resources such as abstracts [9]. As [7] reports, the QALD benchmark has been the evaluation subject of diverse systems, some of them based on query templates. Techniques adopted include graph search, Hidden Markov Models, structured perceptron and (only recently) machine learning, as well as string similarity, language taxonomies, and distributional semantics.

Even a 5-star LOD dataset may contain vocabulary which cannot be mapped to any other dataset in the cloud because of its uniqueness (e.g., a specific property of a protein might not exist in open-domain KBs). In order to achieve QA over the SW, we need to perform QA outside of a single KB by exploiting the advantages of the SW (e.g., external links, ontology alignments).

2.2 Question Answering with Knowledge Base

Within the SW community, KBQA is not as popular as QALD. Research in Semantic Parsing, defined as *"the task of converting a natural language utterance to a logical form"*, was often excluded from comparisons with QALD approaches for being irrelevant to RDF [9]. Datasets based on Freebase [4] such as WE-BQUESTIONSSP [3], SIMPLEQUESTIONS and COMPLEXQUESTIONS [5] are rarely used in the QALD community, despite their full compatibility with RDF standards. On the other hand, they are widely adopted in the KBQA community. Several works in the field of Computational Linguistics target both closed- and open-domain KBQA. Early on, [16] proposed a supervised method to translate questions into queries, which however required a lot of training data. Semantic parsing approaches were later introduced to address this problem by learning the queries out of question-answer pairs, as in [3]. *Neural Symbolic Machines* were devised for query induction using recurrent neural networks and reinforcement learning [10]. Very recently, [2] proposed an approach to automate the template generation and in [1], an architecture to reach never-ending learning from a small set of question-answer pairs. Most of current QALD approaches still struggle tackling the issues above, in addition to lexical gap and support for complex operators [9]. Given the task similarities, we expect the aforementioned KBQA approaches to achieve high scores on QALD.

3   Where is Linked Data?

We argue that the open-domain QALD benchmarks are to DBpedia and Wikidata what the KBQA benchmarks are to Freebase. Judging strictly from a practical viewpoint, the two areas do not differ in anything except in the underlying data. Hence the question in the title, *"Where is Linked Data in Question Answering over Linked Data"*?

Unfortunately, QALD benchmark datasets are *self-contained*, meaning that the desired information can be either found inside them or not found at all. Such scenario is different from the LOD cloud, where information about any real-world entity is usually spread across multiple datasets. As of today, to the best of our knowledge, a method to evaluate QA on datasets which exploits external links to the rest of the cloud does not yet exist. Let us introduce an example; say we have a source dataset reporting employee data including the city of their birth. The question *"which employees were born in the US?"* would need additional knowledge to be answered correctly (i.e., what "US" means and how they relate with the cities). We argue that this kind of QA problems can be addressed only by following the external links from the starting dataset; in this case, an employee can be born in `dbr:Monterey,_California` and – by virtue of the following DBpedia statement – be added to the result set.

```
dbr:Monterey,_California   dbo:country   dbr:United_States   .
```

One could argue that there is no difference between two or more interlinked KBs and their merge into one single KB. However, in a real-world scenario, Linked Datasets can be extremely different in size, structure, and format, as well as being subject of constant change. As up-to-dateness does not seem a concern for the KBQA community – since Freebase is now a defunct project containing obsolete data – the SW community needs to start considering this problem in its actual environment, i.e. the Web.

Nevertheless, in order to scale to the size of the Web, systems must be prepared. It is known that scalability is still an issue for the majority of the QALD systems [7]. Recent works in KBQA managed to deal with billions of triples achieving a satisfiable waiting time for the end user [2]. With this respect, neural approaches are a promising alternative; despite being expensive during training, neural networks usually do not require as many resources for prediction.

Another point that would differentiate the two research areas is the variety of schemata used in the LOD cloud. Each dataset has its own vocabulary, which is adapted to the context, where some vocabularies use extremely specific or idiosyncratic terms. It is therefore not granted that an approach which can perform well on open-domain QA may also generalize well on other datasets. Term overlap is also frequent, leading to a scenario that is completely different than in KBQA.

4   What to do now?

In this section, we propose two settings of a hypothetical benchmark. Both settings could be generated semi-automatically without too much human effort, or at least no more than for a canonical QALD dataset. The availability of the numerous QALD benchmarks released so far reduces the risk factor consistently. After such new benchmarks are released, we expect the community to react and provide resolutions in the span of one year within the next QALD edition.

Setting 1. A domain-specific dataset containing links to (e.g.) DBpedia and Schema.org entities is given in full to the QA systems. Questions can be answered only if the DBpedia and Schema.org links are dereferenced. A snapshot of the used LOD cloud subset can be created for reproducibility and distributed through query-ready formats such as KBox [12] and HDT [8].

Setting 2. The second setting we propose exploits the fact that properties utilized in the LOD cloud are defined by widely-adopted *standardized* vocabularies. The task is to apply transfer learning over two or more datasets using one or more common upper ontologies (e.g., OWL, DCT, SKOS). Training data are only given on a dataset A, while questions target dataset B, where the different ontologies in A and B are aligned to an upper ontology.

5   Conclusion

We showed the necessity for the QALD community to tackle the homonym problem not as a single KB, from the perspective of the Web of Data. After such new benchmarks are released, we expect the community to react and provide resolutions in the span of one year within the next QALD edition. Our plea is to let language experts do language and Web semantics experts do semantics. While the former will keep addressing all problems related to human language, new challenges will arise for the SW.

Acknowledgments

References

1. Abdalghani Abujabal, Rishiraj Saha Roy, Mohamed Yahya, and Gerhard Weikum. Never-ending learning for open-domain question answering over knowledge bases. In *Proc. of the 2018 The Web Conference*, pages 1053–1062, 2018.
2. Abdalghani Abujabal, Mohamed Yahya, Mirek Riedewald, and Gerhard Weikum. Automated template generation for question answering over knowledge graphs. In *Proc. of the 26th Int. Conf. on World Wide Web*, pages 1191–1200, 2017.

3. Jonathan Berant, Andrew Chou, Roy Frostig, and Percy Liang. Semantic parsing on freebase from question-answer pairs. In *EMNLP*, 2013.

4. Kurt Bollacker, Colin Evans, Praveen Paritosh, Tim Sturge, and Jamie Taylor. Freebase: a collaboratively created graph database for structuring human knowledge. In *SIGMOD*, 2008.

5. Antoine Bordes, Nicolas Usunier, Sumit Chopra, and Jason Weston. Large-scale simple question answering with memory networks. 2015.

6. Dennis Diefenbach, Andreas Both, Kamal Singh, and Pierre Maret. Towards a question answering system over the semantic web. *arXiv preprint arXiv:1803.00832*, 2018.

7. Dennis Diefenbach, Vanessa Lopez, Kamal Singh, and Pierre Maret. Core techniques of question answering systems over knowledge bases: a survey. *KIS*, 2017.

8. Mario Arias Gallego, Javier D Fernández, Miguel A Martınez-Prieto, and Claudio Gutierrez. Hdt-it: Storing, sharing and visualizing huge rdf datasets. In *10th Int. Semantic Web Conference, Bonn, Germany, October*, pages 23–27, 2011.

9. Konrad Höffner, Sebastian Walter, Edgard Marx, Ricardo Usbeck, Jens Lehmann, and Axel-Cyrille Ngonga Ngomo. Survey on challenges of question answering in the semantic web. *Semantic Web*, 8(6):895–920, 2017.

10. Chen Liang, Jonathan Berant, Quoc Le, Kenneth D Forbus, and Ni Lao. Neural symbolic machines. 2016.

11. Vanessa Lopez, Christina Unger, Philipp Cimiano, and Enrico Motta. Evaluating question answering over linked data. *Journal of Web Semantics*, 21:3–13, 2013.

12. Edgard Marx, Ciro Baron, Tommaso Soru, and Sören Auer. Kbox transparently shifting query execution on knowledge graphs to the edge. In *ICSC*, 2017.

13. John P. McCrae, Andrejs Abele, Paul Buitelaar, Richard Cyganiak, Anja Jentzsch, and Vladimir Andryushechkin. The Linked Open Data Cloud, 2018.

14. Priyansh Trivedi, Gaurav Maheshwari, Mohnish Dubey, and Jens Lehmann. Lc-quad: A corpus for complex question answering over knowledge graphs. In *International Semantic Web Conference*, pages 210–218. Springer, 2017.

15. George Tsatsaronis, Michael Schroeder, Georgios Paliouras, Yannis Almirantis, et al. Bioasq: A challenge on large-scale biomedical semantic indexing and question answering. In *AAAI fall symposium*, 2012.

16. Mohamed Yahya, Klaus Berberich, Shady Elbassuoni, Maya Ramanath, Volker Tresp, and Gerhard Weikum. Natural language questions for the web of data. In *EMNLP*, pages 379–390. ACL, 2012.