Question Answering on Linked Data: Challenges and Future Directions

Saeedeh Shekarpour
Knoesis Research Center, USA
saeedeh@knoesis.org

Dennis Lukovnikov
University of Bonn, Germany
lukovnikov@outlook.com

Ashwini Jaya Kumar
Fraunhofer IAIS, Germany
ashwinijk18@gmail.com

Kemele M. Endris
University of Bonn, Germany
endris@cs.uni-bonn.de

Kuldeep Singh
Fraunhofer IAIS, Germany
kskuldeepvit@gmail.com

Harsh Thakkar
University of Bonn, Germany
thakkar@cs.uni-bonn.de

Christoph Lange
University of Bonn / Fraunhofer IAIS, Germany
langec@cs.uni-bonn.de

ABSTRACT

Question Answering (QA) systems are becoming the inspiring model for the future of search engines. While recently, underlying datasets for QA systems have been promoted from unstructured datasets to structured datasets with highly semantic-enriched metadata, but still question answering systems involve serious challenges which cause to be far beyond desired expectations. In this paper, we raise the challenges for building a Question Answering (QA) system especially with the focus of employing structured data (i.e. knowledge graph). This paper provide an exhaustive insight of the known challenges, so far. Thus, it helps researchers to easily spot open rooms for the future research agenda.

Keywords

Question Answering System, Research Challenge, Speech Interface, Query Understanding, Data Quality, Distributed and Heterogenous datasets, Interoperability of Components.

1. INTRODUCTION

Web of Data is enormously growing (currently more than 84 billion triples[1]). This data contains both structured and unstructured data. Still, taking advantage of this rapidly growing data is challenging. Traditional information retrieval approaches based on keyword search are user-friendly but can not exploit the internal structure of data due to their bag-of-words semantic. For searching information on the Data Web we need similar user friendly approaches i.e. keyword-based interfaces, which lie on the internal structure of the data.

Question Answering is a specialized form of information retrieval. A Question Answering system retrieves exact answers to questions posed in natural language by user. While recently, underlying datasets for QA systems have been promoted from unstructured datasets to structured datasets with highly semantic-enriched metadata, but still question answering systems involve serious challenges which cause to be far beyond desired expectations.

Question Answering systems consists of elements which independently can be studied and developed. These elements consists of (1) input interface for obtaining query, (2) understanding, interpreting, disambiguating and parsing query, (3) issues related to the employed datasets such as heterogeneity, quality and indexing and (4) interoperability issue for interacting different components. In the following, we elaborately discuss challenges related to each element and possibly future directions which can be considered. We close with the conclusion and future plan.

2. CHALLENGES

In this section we present question answering challenges from four different aspects namely, (i) Speech-based interface challenge, (ii) query understanding, interpreting, disambiguating and parsing challenges, (iii) data-oriented challenges (iv) interoperability of QA components challenge.

2.1 Speech Interface

Interfacing speech to QA systems has become a focus of research for a long time. But, the main focus of research effort so far has been spent on interfacing speech to IR-based QA systems.[41, 43, ?], and much less on interfacing speech input to KG based QA systems. Typical state-of-the-art IR approaches integrate a speech recognition (SR) unit directly with the QA system. An effort beyond merely interfacing the two units is required to enable natural conversation in question answering system for both IR and KG methods.

An SR system mainly consists of an acoustic model and a

[1]observed on 14 October 2015 at http://stats.lod2.eu/

ACM ISBN 978-1-4503-2138-9.
DOI: 10.1145/1235

language model, where the main objective is to decode what is uttered by the user. In contrast, a general IR based QA system comprises question processing (to extract the query from the input and to determine the answer type), passage retrieval, document retrieval, passage extraction, and finally answer selection depending on the relatedness of the named entities found to the question keyword. The accuracy of recognizing spoken words has a vital influence on the success of the whole QA process. For example, if 'Indore' (a city in India) is recognized as 'in door' (articulation style and duration of utterance is the key difference), then the whole QA process is altered. The city name, which may determine the answer type, is not recognized by the QA system. This can be avoided if there is a rich dataset to train the recognizer; however, it is not possible to have acoustic signals for an open domain. Hence speech recognizers are usually built for a specified domain. The same applies to QA systems: developing an open-domain QA is a challenge.

With the evolution of neural network (NN) based methods for SR, the conventional approach to speech recognition has changed. NN based methods have surpassed the various pipeline stages in SR system construction. Usually the acoustic model and the language model were built as two independent units, but now a simplified neural network architecture [**?**, **?**, **?**] is used to obtain a transcript of the audio input. Following the same principle, it is possible to build a whole QA system with deep neural networks. A current research direction is towards exploring the interface of speech to knowledge graphs using deep neural networks.

2.2 Understanding Questions

In the case of QA over structured data, for example over a knowledge base such as Freebase [8], the question must be translated into a logical representation that conveys its meaning in terms of entities, relations, types as well as logical operators. This task of translating from NL to a logical form (semantic parsing) is characterized by the mismatch between natural language (NL) and knowledge base (KB). The semantic parsing problem can be divided into two parts: (1) determining KB constituents mentioned in the NL expression and (2) determining how these constituents should be arranged in a logical structure. The mismatch between NL and KB brings several problems. One problem is Entity Linking (EL), recognizing parts of NL input that refer to an entity (NER) and determining which named entities are meant by that part (disambiguation). An important challenge in EL is how to take into account the context of an entity mention in order to find the correct meaning. Another challenge is finding an optimal set of suitable candidates for a mention, where the *lexicon* plays an important role.

Another problem is relation detection and classification. Given an NL phrase, we want to determine which KB relation is implied by the phrase. Sometimes, the relation is explicitly denoted by a NL constituent, for example verb-mediated statements (e.g. "X married Y"), in which case a lexicon can help a lot to solve the problem. However, in general, a lexicon-based approach is not sufficient. Sometimes there are no relation-specific words in the sentence. Sometimes prepositions are used, for example "works by Repin" or "cars from Germany" and sometimes the semantics of the relations and the entities/types they connect are lexicalized as one, for example, "Russian chemists" or "Tolstoy plays". Such cases require context-based inference, taking into account the semantics of the entities that would be connected by the to-be-determined relation (which in turn is related to parsing).

Merely linking entities and recognizing the relations is not sufficient to produce a logical representation that can be used to query a data source. The remaining problem is to determine the overall logical structure of the NL input. This problem becomes difficult for longer, more complex sentences, where different linguistic phenomena, such as coordination and coreference, must be handled. Formal grammars, such as CCG [49], can help to parse NL input. CCG in particular is well-suited for semantic parsing because of its transparent interface between syntactic structure and underlying semantic form. One problem with grammar-based semantic parsers is their rigidity, which is not well-suited for incomplete input as often found in real-world QA scenarios. Some works have explored learning relaxed grammars [59] to handle such input.

The straightforward way of training semantic parsers requires training data consisting of NL sentences annotated with the corresponding logical representation, which are very cumbersome to obtain. Recent works have explored different ways to reduce the annotation effort in order to bypass this challenge. One proposed way is to train on question-answer pairs instead [7]. Another way is to automatically generate training data from the KBand/or from entity-linked corpora [39] (e.g. ClueWeb)Training with paraphrasing corpora [7]is another technique explored in several works to improve the range of expressions the system can cover.

Recently, impressive advances in different tasks in Artificial Intelligence have been achieved using deep learning techniques. Embedding-based language models, such as Word2Vec [34, 35] and GloVe [37], have helped to improve performance in many NLP tasks. One of the most interesting and the most promising future directions for semantic parsing and question answering is further exploration of deep learning techniques in their context.

Using deep learning to better understand questions can be done by using (possibly custom-trained) word (, word sense and entity) embeddings, which capture their syntactic and semantic properties, as features to improve more traditional workflows. However, a "deeper" approach would be to also devise new models that provide the machine with more freedom to figure out how to accomplish the task. An excellent and very recent example in NLP is the Dynamic Memory Network (DMN [29]), that does not use any manually engineered features or problem-tailored models, yet outperforms the state-of-the-art on all tested tasks, which are dissimilar enough to leave one impressed (POS tagging, coreference resolution, sentiment analysis and Question Answering on the bAbI dataset). The DMN is one of the works focussing on attention and memory in deep learning that enables learning to reason. We share the belief that the investigation and application of more advanced deep learning models (such as DMN and NTM [21]) could yield impressive results for different tasks in AI, including question answering.

Deep neural networks are typically implemented as Recursive neural networks or recurrent neural networks (RNN). Convolutional Neural Networks (CNN), a special case of recursive NNs are well-explored for computer vision. Recursive NNs have also been applied for parsing. RNNs produce state-of-the-art results in speech processing as well as in NLP because of their natural vigor for processing variable-length sequences. They have been applied for machine translation (SMT) [13], language generation (NLG) [50], language modelling and more and are also fundamental for the success of the DMN and the NTM.

An additional interesting direction is the investigation of joint models for the subproblems involved in question interpretation (EL, coreference resolution, parsing, ...). Many tasks in NLP depend on each other to some degree, motivating the investigation of efficient approaches to make the decisions for those tasks jointly. For example, coreference resolution and EL can benefit from each other as entity information from a KB can serve as quite powerful features for coreference resolution and coreference resolution in turn can improve EL as it transfers KB features to phrases where anaphora refer to entities. Factor Graphs (and Markov Networks) are by nature very well-suited for explicit joint models (e.g. [47]). However, more implicit joint inference can also be achieved within a deep neural approach. For example, the DMN mentioned above needs to resolve anaphora in order to correctly answer certain questions. Based on the results of the DMN on that dataset, it can be assumed that it actually learned to perform coreference resolution where needed without any explicit model or a supervision signal for coreference resolution.

The concluding thought is that the further investigation of language [34, 35, 37, 24] and knowledge modelling [20, 55, 40, 48, 12] and powerful deep neural architectures with self-regulating abilities (attention, memory) as well as implicit or explicit joint models will continue to push the state of the art in QA. Well-designed deep neural architectures, given proper supervision and powerful input models, have the potential to learn to solve many different NLU problems with minimal customizations, eliminating the need for carefully engineered features, strict formalisms to process complex structures or pipelines arranging problem-tailored algorithms. We believe that this line of research on QA could be the next yellow brick [5] in the road to true AI, which has fascinated humanity since the ancient tales of Talos and Yan Shi's mechanical men.

2.3 Data challenges

2.3.1 Indexing heterogeneous data

A typical QA system is empirically only as good as the performance of its indexing module [17]. The performance of indexing serves as an upper bound to the overall output of the QA system, since it can process only as much data as is being presented/served to it from the indices. The precision and recall of the system may be good, but if all or most of the top relevant documents are not indexed in the system, the system performance suffers and hence does the end user.

Many researchers have compared effectiveness across a variety of indexing techniques. Their studies show improvement if multiple techniques were combined compared to any single individual indexing technique [38]. In the present scenario, information retrieval systems are carefully tailored and optimised to deliver highly accurate results for specific tasks. Over the years, efforts of developing such task specific systems have been diversified based on a variety of factors discussed in the following.

Based on the type of the data and the application setting, a wide range of indexing techniques are deployed. They can broadly be categorized into three categories based on the format and type of data indexed, namely: structured (e.g. RDF, SQL, etc.), semi-structured (e.g. HTML, XML, JSON, CSV, etc.) and/or unstructured data (e.g. text dumps). They are further distinguished by the type of technique they use for indexing and/or also by the type of queries that a particular technique can address. The different techniques inherently make use of a wide spectrum of underlying fundamental data structures in order to achieve the desirable result.

Most of the systems dealing with unstructured or semi-structured data make use of inverted indices and lists for indexing. For structured systems, a variety of data structures such as AVL trees, B-Trees, sparse indices, IR trees, etc., have been developed in the past decades. Many systems combine two or more data structures to maintain different indices for different data attributes. We present a short survey of indexing platforms and data structures used in a wide range of QA systems in table 1.

Table 1 is an excerpt from a table in our exhaustive survey of open QA systems[2] Our current work in progress is focusing on benchmarking different datasets such as Wikidata [54], DBpedia [2], and FreeBase [9] against a wide spectrum of indexing library platforms, such as Indri[3], Solr[4], ElasticSearch[5], Sphinx[6], Neo4j[7], Titan[8], Xapian[9], and Terrier[10].

2.3.2 Data Quality Challenge

Recent advancements in the fields of Web of Data and Data Science have led to an outburst of standards related to structured data[11] such as RDF(a), Linked Data, schema.org, etc., to an increasing amount of such data, and to a wide range of tools to produce, manage and consume such data. To be available for ready consumption, especially in open question answering systems, any such data sources should meet a certain level of quality, e.g., defined by benchmarks. Quality can generally be defined as "fitness for use", but there are a lot of concrete factors that influence a dataset's fitness for use in question answering[12] settings and in specific application domains. Recently, a number of research activities have been concerned with automating the assessment of linked data quality. Debattista, who has developed one such tool (Luzzu [16]), provides an overview of other state-of-the-art tools [16], including one by Flemming [19],

[2]The full data collection can be found at https://docs.google.com/spreadsheets/d/1S_MfZKRLX2V3kjDhdTKHn2uYyx5hhnr22zym8iT45OE/edit#gi=388858238

[3]http://www.lemurproject.org/indri.php

[4]http://lucene.apache.org/solr/

[5]https://www.elastic.co/products/elasticsearch

[6]http://sphinxsearch.com/

[7]http://neo4j.com/

[8]http://thinkaurelius.github.io/titan/

[9]http://xapian.org/

[10]http://www.terrier.org/

[11]The amount not only of structured, but also of semi-structured and unstructured data available online is also steadily increasing; however, for the purpose of our work we assume that such data has first been translated to the RDF data model using standard tools, e.g. from the Linked Data Stack [3].

[12]In this section, we do not abbreviate "question answering" as "QA" to avoid confusion with "quality assessment".

System	Data structure used	Platform used for indexing
SWSE/YARS2 [23]	Sparse, Inverted Indices for RDF quads	Lucene
Sindice [36]	Inverted Index and On-disk persistent storage	Solr
Sina [45]	Bitmap index on RDF quads (in total 5 indices are maintained: 2 full RDF quad indices, 3 partial RDF quad indices)	OpenLink Virtuoso
HAWK [53]	*N/A	*N/A
TBSL [52]	Inverted Index	Solr
Ephyra [44]	Inverted index	Lemur-Indri
PowerAqua [32]	Inverted index	Lucene (two indices are prepared taxonomically)
AquaLog [31]	*N/A	GATE MIMIR possibly with Lucene
Sig.ma [51]	Inverted Index and On-disk persistent storage	Solr
QUADS [56]	Inverted index	Lucene
MAYA [27]	(key, value) pairs	Traditional index with RDBMS
ESTER [4]	Extended inverted index âĂŞ inverted index with scores for each word ; combines prefix search and join operations	Proprietary module
QAST [25]	Inverted index with term weighting ((Minimal Span Weighting))	Lucene
FREyA [15]	*N/A	Sesame/OWLIM (aka GraphDB))
QAKIS [11]	*N/A	*N/A
MEANS [6]	Inverted index	Terrier
Watson/DeepQA [26, 18]	Persistent disk caching	Watson Explorer Engine XML (VXML)

Table 1: Comparison of the indexing platforms and data structures used by different QA systems. *N/A = no data available

as well as Sieve [33], RDFUnit [28], TripleCheckMate [57], LinkQA [22], and LiQuate [42]. In this section, we summarise the concrete criteria by which the quality of linked data can be assessed, with a special focus on those criteria that are relevant to question answering.

In a comprehensive review of literature and systems, Zaveri et al. [58] have identified the dimensions of linked data quality and categorised them as follows:

- **Accessibility dimensions**: This category covers aspects related to retrieving and accessing data, which includes full or partial access and different technical means of access (e.g. the possibility to download a data dump vs. the availability of a SPARQL endpoint, i.e. a standardised query interface).
  - Availability is generally defined as the ease of access with which particular information is obtainable or rapidly retrievable for readily consumption. In a linked data context, availability can be referred to as the accessibility of a SPARQL endpoint or RDF dumps or dereferenceable URIs.
  - Interlinking is relevant as it refers to the data integration and interoperability. The output of interlinking is a *linkset*, i.e. a set of RDF triples linking subjects and recognised related objects.
  - Security denotes the degree to which a particular dataset is resistant to misuse or alteration without appropriate user access rights.
  - Verifiability, usually by an unbiased third party, addresses the authenticity and correctness of the dataset. Verifiability is typically enabled by provenance metadata.
- **Intrinsic dimensions**: This category covers aspects that are independent of the user's context, or the out of the application context âĂŞ such as accuracy and consistency.
  - Accuracy refers to the degree of a dataset correctly representing the captured real world facts and figures in the form of information with high precision.
  - Consistency refers to the independence from logical, formal or representational contradictions of a dataset with respect to others.
  - Completeness is referred to as the degree to which information in the dataset is complete or not missing. The dataset should have all the required objects or values for a given task in order to be considered as complete. Thus, arguing intuitively, completeness is one of the concrete metrics for linked data quality assessment.
- **Contextual dimensions**: This category is concerned with the context of the task being pursued.
  - Timeliness is concerned with the freshness of data over time or timeliness, i.e. the regularity of updates or merges and so on.
  - Understandability can be achieved by providing appropriate human readable annotations to a dataset and its entities, and by consistently following a certain regular expression as a pattern for forming entity URIs.
  - Trustworthiness is concerned with the reliability or trustworthiness of the data and its source.

- **Representational dimensions**: This category is concerned with the design and representation of the data and its schema. For instance, understandability and interpretability.
  - Interpretability refers to adhering to the standard practice of representing information using appropriate notations, symbols, units and languages.

Data quality dimensions in all of these categories can be relevant in question answering scenarios. Our next step is to identify more systematically what dimensions of data quality are specifically relevant in the typical application domains of question answering, or sufficient for determining a dataset's "fitness" for question answering. Having identified such dimensions, we have two goals: identifying datasets that are suitable for question answering at all, and then, for those datasets that are, identifying more specifically what quality problems they still suffer from. This leads to the question of what concrete quality metrics for the relevant quality dimensions can be computed on such datasets in a reasonable way, given, e.g., the expressiveness of their schemas, and, secondly, whether implementations are available to effectively and efficiently compute these metrics on the given datasets. Regarding implementation, we expect that the Luzzu linked data quality assessment framework provides a sufficient number of fully implemented quality metrics that are ready to use in a question answering setting, that further existing implementations of metrics in Luzzu can be specifically adapted to make them suitable for quality assessment related to question answering, and that, finally, Luzzu's flexible extensibility even enables us to implement new metrics that may be required. In summary, our near-future work will be concerned with defining a generally and flexibly applicable framework for automating the process of rigorously assessing the quality of linked datasets for question answering by identifying, formalising and implementing the required metrics.

For this purpose, we have so far identified 10 dimensions and 17 metrics, from [58] to be evaluated on LOD datasets such as DBpedia, Wikipedia and many more to follow. We plan to extend Luzzu over the next year and implement more open question answering related metrics. The results of our preliminary study on DBpedia and Wikidata are published online as a public spreadsheet.[13]. For this evaluation we have obtained four slices of both DBpedia and Wikidata, namely *Restaurants, Politicians, Films* and *Soccer players*. The scores of evaluation on all of the four slices are reported on the link mentioned above.

2.3.3 Distributed heterogeneous data

The decentralized architecture of the Web has produced a wealth of knowledge distributed across different data sources and different data types. Question answering systems consume different types of data: *structured*, *semi-structured* or *unstructured* data. Most question answering systems uses either of these types of data to answer user queries. Only few systems exploit the wealth of data on the Web by combining these types of data. Hybrid question answering systems are able to answer queries by combining both structured and unstructured types of data. HAWK [53], for instance, provides entity search for hybrid question answering using Linked Data and textual data. HAWK is able to achieve an F-measure of up to 0.68 on the QUALD-4 benchmark.

Most question answering systems today uses a single source to answer users question. It should rather be possible to answer questions imposed by a user by combining different interconnected sources. The challenges imposed by the distributed nature of the Web are, on the one hand, finding the right sources that can answer user query and, on the other hand, integrating partial answers found from different sources. Source selection is one of the challenges in federated question answering approaches. In [46], the authors presented an approach to construct a federated query from user supplied (natural language) questions using disambiguated resources.

Answers may come from different sources which have different data quality and trust levels, ranking and fusion of data should be applied to select the best sources.

The amount of data to be used to answer users' queries should also be balanced with the response time.

2.4 Interoperability Challenge

The field of QA is so vast that the list of different QA systems can go long. Many Question Answering systems Based on specific domains have been developed. Domain-specific QA systems, for example [1] are limited to a specific knowledge, for example medicine. They are known as *closed domain* QA systems. However, when scope is limited to an explicit domain or ontology, there are less chances of ambiguity and high accuracy of answers. It is also difficult and costly to extend closed domain systems to a new domain or reusing it in implementing a new system.To overcome the limitations of closed domain QA systems, researchers have shifted their focus to *open domain* QA systems. FREyA [14], QAKiS [10], and PowerAqua [30] are few examples of open domain QA systems which use publicly available semantic knowledge for example DBpedia [2].

While many of these system achieved significant performance for special use cases, a shortage was observed in all of them. We figured out that the existing QA systems suffer from the following drawbacks: (1) potential of reusing its components is very weak, (2) extension of the components is problematic, and (3) interoperability between the employed components are not systematically defined. Interoperability of these tools and components is needed to enhance QA process which is still missing at the conceptual level and currently more focused on implementation details. Therefore, there is a need for a descriptive approach that define a conceptual view of QA systems. This approach must cover all needs of current QA systems and be abstracted from implementation details. Moreover it must be open such that it can be used in future QA systems. The generalized approach for architecture or ontology of a QA system and semantic search must focus to bring all state-of-the advancement of QA under a single umbrella. We envisioned that a generalized vocabulary for QA will be an abstraction level on top of all the existing QA approaches and will provide interoperability and exchangeability between them. This generalized vocabulary can be further used to integrate different components and web services within a QA system.

3. CONCLUSION AND FUTURE PLAN

[13] https://docs.google.com/spreadsheets/d/15E--hG0uQnyiYfU-OeNqj-dpQ1PzOijYGXq28XEG5cg/edit#gid=0

In this paper, we presented an exhaustive overview of all the open challenges being still controversial for developing a question answering system. The intuition is that Linked Data which provides advantages such as semantic metadata and interlinked dataset can influence all of the four major elements (i.e. interface, parsing, data and component interoperability) which play a key role in Question Answering systems. As our future research agenda, we are steering our research on all of the discussed issues with the focus of employing Linked Data technology to promote question answering capabilities.

Acknowledgments Parts of this work received funding from the European Union's Horizon 2020 research and innovation programm under the Marie Sklodowska-Curie grant agreement No. 642795 (WDAqua project). We thank Didier Cherix for his kind support while developing the approach.

4. ADDITIONAL AUTHORS

5. REFERENCES

[1] Asma Ben Abacha and Pierre Zweigenbaum. Medical question answering: translating medical questions into sparql queries. In *ACM International Health Informatics Symposium, IHI '12, Miami, FL, USA, January 28-30, 2012*, pages 41–50, 2012.

[2] Sören Auer, Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak, and Zachary G. Ives. Dbpedia: A nucleus for a web of open data. In *The Semantic Web, 6th International Semantic Web Conference, 2nd Asian Semantic Web Conference, ISWC 2007 + ASWC 2007, Busan, Korea, November 11-15, 2007.*, pages 722–735, 2007.

[3] Sören Auer, Jens Lehmann, and Axel-Cyrille Ngonga Ngomo. Introduction to linked data and its lifecycle on the web. In *Reasoning Web*, pages 1–75, 2011.

[4] Holger Bast, Alexandru Chitea, Fabian M Suchanek, and Ingmar Weber. ESTER : Efficient Search on Text , Entities , and Relations. *Search*, (2):671–678, 2007.

[5] L Frank Baum. *The wonderful wizard of Oz*. Oxford University Press, 2008.

[6] Asma Ben Abacha and Pierre Zweigenbaum. MEANS: A medical question-answering system combining NLP techniques and semantic Web technologies. *Information Processing & Management*, 51(5):570–594, 2015.

[7] Jonathan Berant and Percy Liang. Semantic parsing via paraphrasing. In *Proceedings of ACL*, volume 7, page 92, 2014.

[8] Kurt Bollacker, Colin Evans, Praveen Paritosh, Tim Sturge, and Jamie Taylor. Freebase: a collaboratively created graph database for structuring human knowledge. In *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*, pages 1247–1250. ACM, 2008.

[9] Kurt D. Bollacker, Robert P. Cook, and Patrick Tufts. Freebase: A shared database of structured general human knowledge. In *Proceedings of the Twenty-Second AAAI Conference on Artificial Intelligence, July 22-26, 2007, Vancouver, British Columbia, Canada*, pages 1962–1963, 2007.

[10] E. Cabrio, J. Cojan, A. P. Aprosio, B. Magnini, A. Lavelli, and F. Gandon. QAKiS: an open domain QA system based on relational patterns. In *Proc. of the ISWC 2012 Posters & Demonstrations*, 2012.

[11] Elena Cabrio, Julien Cojan, Alessio Palmero Aprosio, Bernardo Magnini, Alberto Lavelli, and Fabien Gandon. Qakis: an open domain qa system based on relational patterns. In *11th International Semantic Web Conference ISWC 2012*, page 9. Citeseer, 2012.

[12] Kai-Wei Chang, Wen-tau Yih, Bishan Yang, and Christopher Meek. Typed tensor decomposition of knowledge bases for relation extraction. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1568–1579, 2014.

[13] Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*, 2014.

[14] D. Damljanovic, M. Agatonovic, and H. Cunningham. Freya: An interactive way of querying linked data using natural language. In *ESWC Workshops*, 2011.

[15] Danica Damljanovic, Milan Agatonovic, and Hamish Cunningham. Freya: An interactive way of querying linked data using natural language. In *The Semantic Web: ESWC 2011 Workshops*, pages 125–138. Springer, 2012.

[16] Jeremy Debattista, Sören Auer, and Christoph Lange. Luzzu – a framework for linked data quality analysis. In *Semantic Computing*, 2016.

[17] Xin Dong and Alon Halevy. Indexing dataspaces. In *Proceedings of the 2007 ACM SIGMOD international conference on Management of data*, pages 43–54. ACM, 2007.

[18] David Ferrucci, Eric Brown, Jennifer Chu-Carroll, James Fan, David Gondek, Aditya A Kalyanpur, Adam Lally, J William Murdock, Eric Nyberg, John Prager, et al. Building watson: An overview of the deepqa project. *AI magazine*, 31(3):59–79, 2010.

[19] Annika Flemming. Quality Characteristics of Linked Data Publishing Datasources. http://sourceforge.net/apps/mediawiki/trdf/index.php?\\title=Quality_Criteria_for_Linked_Data_sources, 2010. [Online; accessed 13-February-2014].

[20] Alberto García-Durán, Antoine Bordes, and Nicolas Usunier. Effective blending of two and three-way interactions for modeling multi-relational data. In *Machine Learning and Knowledge Discovery in Databases*, pages 434–449. Springer, 2014.

[21] Alex Graves, Greg Wayne, and Ivo Danihelka. Neural turing machines. *arXiv preprint arXiv:1410.5401*, 2014.

[22] Christophe Guéret, Paul Groth, Claus Stadler, and Jens Lehmann. Assessing linked data mappings using network measures. In *The Semantic Web: Research and Applications*, pages 87–102. Springer, 2012.

[23] Andreas Harth, Jürgen Umbrich, Aidan Hogan, and Stefan Decker. YARS2: A federated repository for querying graph structured data from the Web. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 4825 LNCS:211–224, 2007.

[24] Ignacio Iacobacci, Mohammad Taher Pilehvar, and Roberto Navigli. SENSEMBED: Learning Sense Embeddings for Word and Relational Similarity. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics, Beijing, China*, 2015.

[25] Wittawat Jitkrittum, Choochart Haruechaiyasak, and Thanaruk Theeramunkong. QAST : Question Answering System for Thai Wikipedia. *Knowledge Creation Diffusion Utilization*, (August):11–14, 2009.

[26] Aditya Kalyanpur, Branimir K Boguraev, Siddharth Patwardhan, J William Murdock, Adam Lally, Chris Welty, John M Prager, Bonaventura Coppola, Achille Fokoue-Nkoutche, Lei Zhang, et al. Structured data and inference in deepqa. *IBM Journal of Research and Development*, 56(3.4):10–1, 2012.

[27] H. Kim, Kyungsun Kim, G.G. Lee, and Jungyun Seo. MAYA: A fast question-answering system based on a predictive answer indexer. *Proceedings of the workshop on Open-domain question answering-Volume 12*, page 8, 2001.

[28] Dimitris Kontokostas, Patrick Westphal, Sören Auer, Sebastian Hellmann, Jens Lehmann, Roland Cornelissen, and Amrapali Zaveri. Test-driven evaluation of linked data quality. In *Proceedings of the 23rd international conference on World Wide Web*, pages 747–758. ACM, 2014.

[29] Ankit Kumar, Ozan Irsoy, Jonathan Su, James Bradbury, Robert English, Brian Pierce, Peter Ondruska, Ishaan Gulrajani, and Richard Socher. Ask me anything: Dynamic memory networks for natural language processing. *arXiv preprint arXiv:1506.07285*, 2015.

[30] Vanessa Lopez, Miriam Fernández, Enrico Motta, and Nico Stieler. Poweraqua: Supporting users in querying and exploring the semantic web. *Semantic Web*, 3(3):249–265, 2011.

[31] Vanessa Lopez, Michele Pasin, and Enrico Motta. Aqualog: An ontology-portable question answering system for the semantic web. In *The Semantic Web: Research and Applications*, pages 546–562. Springer, 2005.

[32] Vanessa Lopez, Victoria Uren, Marta Reka Sabou, and Enrico Motta. Cross ontology query answering on the semantic web: an initial evaluation. In *Proceedings of the fifth international conference on Knowledge capture*, pages 17–24. ACM, 2009.

[33] Pablo N Mendes, Hannes Mühleisen, and Christian Bizer. Sieve: linked data quality assessment and fusion. In *Proceedings of the 2012 Joint EDBT/ICDT Workshops*, pages 116–123. ACM, 2012.

[34] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.

[35] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119, 2013.

[36] Eyal Oren, Renaud Delbru, Michele Catasta, and Richard Cyganiak. Sindice . com : A Document-oriented Lookup Index for Open Linked Data.

[37] Jeffrey Pennington, Richard Socher, and Christopher D Manning. Glove: Global vectors for word representation. *Proceedings of the Empiricial Methods in Natural Language Processing (EMNLP 2014)*, 12:1532–1543, 2014.

[38] TB Rajashekar and Bruce W Croft. Combining automatic and manual index representations in probabilistic retrieval. *Journal of the American society for information science*, 46(4):272–283, 1995.

[39] Siva Reddy, Mirella Lapata, and Mark Steedman. Large-scale semantic parsing without question-answer pairs. *Transactions of the Association for Computational Linguistics*, 2:377–392, 2014.

[40] Sebastian Riedel, Limin Yao, Andrew McCallum, and Benjamin M Marlin. Relation extraction with matrix factorization and universal schemas. 2013.

[41] Paolo Rosso, Lluis-F. Hurtado, Encarna Segarra, and Emilio Sanchis. On the voice-activated question answering. *IEEE Trans. on Systems, Man and Cybernetics-Part C: Applications and Reviews*, 42(1):75–85, Jan 2012.

[42] Edna Ruckhaus, Oriana Baldizán, and María-Esther Vidal. Analyzing linked data quality with liquate. In *On the Move to Meaningful Internet Systems: OTM 2013 Workshops*, pages 629–638. Springer, 2013.

[43] Emilio Sanchis, Davide Buscaldi, Sergio Grau, Lluis Hurtado, and David Griol. Spoken qa based on a passage retrieval engine. In *Spoken Language Technology Workshop, 2006. IEEE*, pages 62–65. IEEE, 2006.

[44] Nico Schlaefer, Jeongwoo Ko, Justin Betteridge, Manas A Pathak, Eric Nyberg, and Guido Sautter. Semantic extensions of the ephyra qa system for trec 2007. In *TREC*, 2007.

[45] Saeedeh Shekarpour, Edgard Marx, Axel-Cyrille Ngonga Ngomo, and Sören Auer. Sina: Semantic interpretation of user queries for question answering on interlinked data. *Web Semantics: Science, Services and Agents on the World Wide Web*, 30:39–51, 2015.

[46] Saeedeh Shekarpour, Axel-Cyrille Ngonga Ngomo, and Sören Auer. Question Answering on interlinked data. In *WWW*, pages 1145–1156, 2013.

[47] Sameer Singh, Sebastian Riedel, Brian Martin, Jiaping Zheng, and Andrew McCallum. Joint inference of entities, relations, and coreference. In *Proceedings of the 2013 workshop on Automated knowledge base construction*, pages 1–6. ACM, 2013.

[48] Richard Socher, Danqi Chen, Christopher D Manning, and Andrew Ng. Reasoning with neural tensor networks for knowledge base completion. In *Advances in Neural Information Processing Systems*, pages 926–934, 2013.

[49] Mark Steedman and Jason Baldridge. Combinatory categorial grammar. *Non-Transformational Syntax: Formal and Explicit Models of Grammar. Wiley-Blackwell*, 2011.

[50] Ilya Sutskever, James Martens, and Geoffrey E Hinton. Generating text with recurrent neural networks. In *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, pages

1017–1024, 2011.

[51] Giovanni Tummarello, Richard Cyganiak, Michele Catasta, Szymon Danielczyk, Renaud Delbru, and Stefan Decker. Sig. ma: Live views on the web of data. *Web Semantics: Science, Services and Agents on the World Wide Web*, 8(4):355–364, 2010.

[52] Christina Unger, Lorenz Bühmann, Jens Lehmann, Axel-Cyrille Ngonga Ngomo, Daniel Gerber, and Philipp Cimiano. Template-based question answering over rdf data. *Proceedings of the 21st international conference on World Wide Web*, (639–648), 2012.

[53] Ricardo Usbeck, Axel-Cyrille Ngonga Ngomo, Lorenz Bühmann, and Christina Unger. HAWK âĂŞ Hybrid Question Answering using Linked Data. 2014.

[54] Denny Vrandečić and Markus Krötzsch. Wikidata: a free collaborative knowledgebase. *Communications of the ACM*, 57(10):78–85, 2014.

[55] Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen. Knowledge graph embedding by translating on hyperplanes. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*, pages 1112–1119. Citeseer, 2014.

[56] Zi Yang, Ying Li, James Cai, and Eric Nyberg. QUADS : Question Answering for Decision Support. *Proc. SIGIR 2014*, pages 375–384, 2014.

[57] Amrapali Zaveri, Dimitris Kontokostas, Mohamed A Sherif, Lorenz Bühmann, Mohamed Morsey, Sören Auer, and Jens Lehmann. User-driven quality evaluation of dbpedia. In *Proceedings of the 9th International Conference on Semantic Systems*, pages 97–104. ACM, 2013.

[58] Amrapali Zaveri, Anisa Rula, Andrea Maurino, Ricardo Pietrobon, Jens Lehmann, and Sören Auer. Quality assessment for linked data. *Semantic Web Journal*, preprint(preprint), 2015.

[59] Luke S Zettlemoyer and Michael Collins. Online learning of relaxed ccg grammars for parsing to logical form. In *EMNLP-CoNLL*, pages 678–687, 2007.