

Survey on Challenges of Question Answering in the Semantic Web

Konrad Höffner ^a, Sebastian Walter ^b, Edgard Marx ^a, Ricardo Usbeck ^a, Jens Lehmann ^a, Axel-Cyrille Ngonga Ngomo ^a

^a *Leipzig University, Institute of Computer Science, AKSW Group*

Augustusplatz 10, D-04109 Leipzig, Germany

E-mail: {hoeffner,marx,lehmann,ngonga,usbeck}@informatik.uni-leipzig.de

^b *CITEC, Bielefeld University*

Inspiration 1, D - 33615 Bielefeld, Germany

E-mail: swalter@techfak.uni-bielefeld.de

Abstract. Semantic Question Answering (SQA) removes two major access requirements to the Semantic Web: the mastery of a formal query language like SPARQL and knowledge of a specific vocabulary. Because of the complexity of natural language, SQA presents difficult challenges and many research opportunities. Instead of a shared effort, however, many essential components are redeveloped, which is an inefficient use of researcher's time and resources. This survey analyzes 62 different SQA systems, which are systematically and manually selected using predefined inclusion and exclusion criteria, leading to 72 selected publications out of 1960 candidates. We identify common challenges, structure solutions, and provide recommendations for future systems. This work is based on publications from the end of 2010 to July 2015 and is also compared to older but similar surveys.

Keywords: Question Answering, Semantic Web, Survey

1. Introduction

Semantic Question Answering (SQA) is defined by users (1) asking questions in natural language (NL) (2) using their own terminology to which they (3) receive a concise answer generated by querying a RDF knowledge base.¹ Users are thus freed from two major access requirements to the Semantic Web: (1) the mastery of a formal query language like SPARQL and (2) knowledge about the specific vocabularies of the knowledge base they want to query. Since natural language is complex and ambiguous, reliable SQA systems require many different steps. While for some of them, like part-of-speech tagging and parsing, mature high-precision solutions exist, most of the others still present difficult challenges. While the massive research effort has led to major advances, as shown by the yearly Question Answering over Linked Data (QALD) evaluation cam-

paign, it suffers from several problems: Instead of a shared effort, many essential components are redeveloped. While shared practices emerge over time, they are not systematically collected. Furthermore, most systems focus on a specific aspect while the others are quickly implemented, which leads to low benchmark scores and thus undervalues the contribution. This survey aims to alleviate these problems by systematically collecting and structuring methods of dealing with common challenges faced by these approaches. Our contributions are threefold: First, we complement existing work with 72 publications about 62 systems developed from 2010 to 2015. Second, we identify challenges faced by those approaches and collect solutions for them from the 72 publications. Finally, we draw conclusions and make recommendations on how to develop future SQA systems. The structure of the paper is as follows: Section 2 states the methodology used to find and filter surveyed publications. Section 3 compares this work to older, similar surveys as well as eval-

¹Definition based on Hirschman and Gaizauskas [73].

uation campaigns and work outside the SQA field. Section 4 introduces the surveyed systems. Section 5 identifies challenges faced by SQA approaches and presents approaches that tackle them. Section 6 summarizes the efforts made to face challenges to SQA and their implication for further development in this area.

2. Methodology

This survey follows a strict discovery methodology: Objective inclusion and exclusion criteria are used to find and restrict publications on SQA.

Inclusion Criteria Candidate articles for inclusion in the survey need to be part of relevant conference proceedings or searchable via Google Scholar (see Table 1). The included papers from the publication search engine Google Scholar are the first 300 results in the chosen timespan (see exclusion criteria) that contain “question answering’ AND (‘Semantic Web’ OR ‘data web’)” in the article including title, abstract and text body. Conference candidates are all publications in our examined time frame in the proceedings of the major Semantic Web Conferences ISWC, ESWC, WWW, NLDB, and the proceedings which contain the annual QALD challenge participants.

Exclusion Criteria Works published before November 2010² or after July 2015 are excluded, as well as those that are not related to SQA, determined in a manual inspection in the following manner: First, proceeding tracks are excluded that clearly do not contain SQA related publications. Next, publications both from proceedings and from Google Scholar are excluded based on their title and finally on their content.

Notable exclusions We exclude the following approaches since they do not fit our definition of SQA (see Section 1): *Swoogle* [52] is independent on any specific knowledge base but instead builds its own index and knowledge base using RDF documents found by multiple web crawlers. Discovered ontologies are ranked based on their usage intensity and RDF documents are ranked using authority scoring. Swoogle can only find single terms and cannot answer natural language queries and is thus not a SQA system. *WolframAlpha* is a natural language interface based on the computational platform *Mathematica* [143] and aggregates a large number of structured sources and a al-

gorithms. However, it does not support Semantic Web knowledge bases and the source code and the algorithm is are not published. Thus, we cannot identify whether it corresponds to our definition of a SQA system.

Result The inspection of the titles of the Google Scholar results by two authors of this survey led to 153 publications, 39 of which remained after inspecting the full text (see Table 1). The selected proceedings contain 1660 publications, which were narrowed down to 980 by excluding tracks that have no relation to SQA. Based on their titles, 62 of them were selected and inspected, resulting in 33 publications that were categorized and listed in this survey. Table 1 shows the number of publications in each step for each source. In total, 1960 candidates were found using the inclusion criteria in Google Scholar and conference proceedings and then reduced using track names (conference proceedings only, 1280 remaining), then titles (214) and finally the full text, resulting in 72 publications describing 62 distinct SQA systems.

3. Related Work

This section gives an overview of recent QA and SQA surveys and differences to this work, as well as QA and SQA evaluation campaigns, which quantitatively compare systems.

3.1. Other Surveys

QA Surveys Cimiano and Minock [33] present a data-driven problem analysis of QA on the Geobase dataset. The authors identify eleven challenges that QA has to solve and which inspired the problem categories of this survey: question types, language “light”³, lexical ambiguities, syntactic ambiguities, scope ambiguities, spatial prepositions, adjective modifiers and superlatives, aggregation, comparison and negation operators, non-compositionality, and out of scope⁴. In contrast to our work, they identify challenges by manually inspecting user provided questions instead of existing systems. Mishra and Jain [99] propose eight classification criteria, such as application domain, types of questions and type of data. For each criterion, the different classifications are given along with their advantages, disadvantages and exemplary systems.

³semantically weak constructions

⁴cannot be answered as the information required is not contained in the knowledge base

²The time before is already covered in Cimiano and Minock [33].

Table 1

Sources of publication candidates along with the number of publications in total, after excluding based on conference tracks (I), based on the title (II), and finally based on the full text (selected). Works that are found both in a conference's proceedings and in Google Scholar are only counted once, as selected for that conference. The QALD 2 proceedings are included in ILD 2012, QALD 3 [25] and QALD 4 [137] in the CLEF 2013 and 2014 working notes.

Venue	All	I	II	Selected
Google Scholar Top 300	300	300	153	39
ISWC 2010 [110]	70	70	1	1
ISWC 2011 [8]	68	68	4	3
ISWC 2012 [36]	66	66	4	2
ISWC 2013 [5]	72	72	4	0
ISWC 2014 [96]	31	4	2	0
WWW 2011 [78]	81	9	0	0
WWW 2012 [79]	108	6	2	1
WWW 2013 [80]	137	137	2	1
WWW 2014 [81]	84	33	3	0
WWW 2015 [82]	131	131	1	1
ESWC 2011 [7]	67	58	3	0
ESWC 2012 [126]	53	43	0	0
ESWC 2013 [34]	42	34	0	0
ESWC 2014 [112]	51	31	2	1
ESWC 2015 [60]	42	42	1	1
NLDB 2011 [101]	21	21	2	2
NLDB 2012 [22]	36	36	0	0
NLDB 2013 [138]	36	36	1	1
NLDB 2014 [93]	39	30	1	2
NLDB 2015 [17]	45	10	2	1
QALD 1 [113]	3	3	3	2
ILD 2012 [136]	9	9	9	3
CLEF 2013 [53]	208	7	6	5
CLEF 2014 [28]	160	24	8	6
$\Sigma(\text{conference})$	1660	980	61	33
$\Sigma(\text{all})$	1960	1280	214	72

SQA Surveys For each participant, problems and their solution strategies are given: Athenikos and Han [9] give an overview of domain specific QA systems for biomedicine. After summarising the current state of the art by September 2009 for biomedical QA systems, the authors describe different approaches from the point of view of medical and biological QA. The authors of this survey only describe approaches, but do not identify the differences between those two main categories. In contrast to our survey, the authors hereby do not sort the presented approaches through problems, but more broader terms such as "Non-semantic knowledge base medical QA systems and approaches" or "Inference-

Table 2

Other surveys by year of publication. Surveyed years are given except when a dataset is theoretically analyzed. Approaches addressing specific types of data are also indicated.

QA Survey	Year	Coverage	Data
Cimiano and Minock [33]	2010	—	geobase
Mishra and Jain [99]	2015	2000–2014	general
SQA Survey	Year	Coverage	Data
Athenikos and Han [9]	2010	2000–2009	biomedical
Lopez et al. [91]	2010	2004–2010	general
Freitas et al. [57]	2012	2004–2011	general
Lopez et al. [92]	2013	2005–2012	general

based biological QA systems and approaches". Lopez et al. [91] presents an overview similar to Athenikos and Han [9] but with a wider scope. After defining the goals and dimensions of QA and presenting some related and historic work, the authors summarize the achievements of SQA so far and the challenges that are still open. Another related survey from 2012, Freitas et al. [57], gives a broad overview of the challenges involved in constructing effective query mechanisms for Web-scale data. The authors analyze different approaches, such as Treo [56], for five different challenges: usability, query expressivity, vocabulary-level semantic matching, entity recognition and improvement of semantic tractability. The same is done for architectural elements such as user interaction and interfaces and the impact on these challenges is reported. Lopez et al. [92] analyze the SQA systems of the participants of the QALD 1 and 2 evaluation challenge, see Section 3.2. While there is an overlap in the surveyed approaches between Lopez et al. [92] and our paper, our survey has a broader scope as it also analyzes approaches that do not take part in the QALD challenges.

In contrast to the surveys mentioned above, we do not focus on the overall performance or domain of a system, but on analyzing and categorizing methods that tackle specific problems. Additionally, we build upon the existing surveys and describe the new state of the art systems which were published after the before mentioned surveys in order to keep track of new research ideas.

3.2. Evaluation Campaigns

Contrary to QA surveys, which qualitatively compare systems, there are also evaluation campaigns,

which quantitatively compare them using benchmarks. Those campaigns show how different open-domain QA systems perform on realistic questions on real-world knowledge bases. This accelerates the evolution of QA in four different ways: Firstly, new systems don't have to include their own benchmark, shortening system development. Secondly, standardized evaluation allows for better research resource allocation as it is easier to determine, which approaches are worthwhile to develop further. Thirdly, the addition of new challenges to the questions of each new benchmark iteration motivates addressing those challenges. And finally, the competitive pressure to keep pace with the top scoring systems compels emergence and integration of shared best practises. On the other hand, evaluation campaign proceedings do not describe single components of those systems in great detail. By focussing on complete systems, research effort gets spread around multiple components, possibly duplicating existing efforts, instead of being focussed on a single one.

Question Answering on Linked Data (QALD) is the most well-known all-purpose evaluation campaign with its core task of open domain SQA on lexicographic facts of DBpedia [88]. Since its inception in 2011, the yearly benchmark has been made progressively more difficult. Additionally, the general core task has been joined by special tasks providing challenges like multilinguality, hybrid (textual and Linked Data) and its newest addition, SQA on statistical data in the form of RDF Data Cubes [75].

BioASQ [132,1,10,11] is a benchmark challenge which ran until September 2015 and consists of semantic indexing as well as an SQA part on biomedical data. In the SQA part, systems are expected to be hybrids, returning matching triples as well as text snippets but partial evaluation (text or triples only) is possible as well. The introductory task separates the process into annotation which is equivalent to named entity recognition (NER) and disambiguation (NED) as well as the answering itself. The second task combines these two steps.

TREC LiveQA, starting in 2015 [4], gives systems unanswered Yahoo Answers questions intended for other humans. As such, the campaign contains the most realistic questions with the least restrictions, in contrast to the solely factual questions of QALD, BioASQ and TREC's old QA track [40].

3.3. Frameworks

Framework may refer to architecture or to a system. System framework provides an abstraction in which an generic functionality can be selectively changed by additional third-party code. In this section we refer to system frameworks. Different from document retrieval where's there are many existing frameworks as Lucene⁵, Solr⁶ and Elastic Search⁷, there is still a lack in tools to facilitate implementation and evaluation process of SQA systems.

Document retrieval frameworks usually split the retrieval process in tree steps (1) query processing, (2) retrieval and (3) ranking. The (1) query processing step consist in apply query analyzers in order to better identify documents in the data store. Thereafter, the query is used to (2) retrieve documents that match the query terms resulting from the query processing. Later, the retrieved documents are (3) ranked according to some ranking function, commonly TFIDF [127]. Developing SQA framework is a hard task because most of the systems work with a mixture of NL techniques on top of traditional IR systems. Some systems make use of the syntactic graph behind the question [135] to deduce the query intention whereas others, the knowledge graph [122]. There are systems that propose to work in either structured and unstructured data [139] or in a combination of systems [64]. Therefore, they contain very peculiar steps. Thus, a new research sub field focuses on question answering frameworks, i.e., design and development of common features for SQA systems.

openQA [94]⁸ is a modular open-source framework for implementing and instantiating SQA approaches. The framework's main work-flow consists of four stages (interpretation, retrieval, synthesis, rendering) and adjacent modules (context and service). The adjacent modules are intended to be accessed by any of the components of the main work-flow to share common features to the different modules e.g. cache. The framework proposes the *answer formulation* process in a very likely traditional document retrieval fashion where the query processing and ranking steps are replaced by the more general *Interpretation* and *Synthesis*. According to the authors, the interpretation step comprises all the pre-processing and matching techniques required to de-

⁵<https://lucene.apache.org>

⁶<https://solr.apache.org>

⁷<https://www.elastic.co>

⁸<http://openqa.aksw.org>

duce the question whereas the syntheses is the process of ranking, merging and confidence estimation required to produce the answer. The authors claims that *openQA* enables a conciliation of different architectures and methods.

4. Systems

The 72 surveyed publications describe 62 distinct systems or approaches. The implementation of a SQA system can be very complex and depending on, thus reusing, several known techniques. SQA systems are typically composed of two stages: (1) the query analyzer and (2) retrieval. The query analyzer generates or formats the query that will be used to recover the answer at the retrieval stage. There is a wide variety of techniques that can be applied at the analyzer stage, such as tokenization, disambiguation, internationalization, logical forms, semantic role labels, question reformulation, coreference resolution, relation extraction and named entity recognition amongst others. For some of those techniques, such as natural language (NL) parsing and part-of-speech (POS) tagging, mature all-purpose methods are available and commonly reused. Other techniques, such as the disambiguating between multiple possible answers candidates, are not available at hand in a domain independent fashion. Thus, high quality solutions can only be obtained by the development of new components. This section exemplifies some of the reviewed systems and their novelties to highlight current research questions, while the next section presents the contributions of all analyzed papers to specific challenges.

Hakimov et al. [65] proposes a SQA system using syntactic dependency trees of input questions. The method consists of three main steps: (1) Triple patterns are extracted using the dependency tree and POS tags of the questions. (2) Entities, properties and classes are extracted and mapped to the underlying knowledge base. Recognized entities are disambiguated using page links between all spotted named entities as well as string similarity. Properties are disambiguated by using relational linguistic patterns from PATTY [102] which

allows a more flexible mapping, e.g., “die” is mapped to DBpedia properties like `dbo:deathPlace`⁹ Finally, (3) question words are matched to the respective answer type, e.g., Who matches person, organization, company while Where matches places) and ranked. The best result is returned as answer.

PARALEX [48] only answers questions for subjects or objects of property-object or subject-property pairs, respectively. It contains phrase to concept mappings in a lexicon that is trained from a corpus of paraphrases, which is constructed from the question-answer site WikiAnswers¹⁰. If one of the paraphrases can be mapped to a query, this query is the correct answer for the paraphrases as well. By mapping phrases between those paraphrases, the linguistic patterns are extended. For example, “what is the *r* of *e*” leads to “how *r* is *e*”, so that “What is the population of New York” can be mapped to “How big is NYC”. There is a variety of other systems, such as Bordes et al. [19], that make use of paraphrase learning methods and integrate linguistic generalization with knowledge graph biases. They are however not included here as they do query RDF knowledge bases and thus do not fit the inclusion criteria.

Xser [144] is based on the observation that SQA contains two independent steps. First, Xser determines the question structure solely based on a phrase level dependency graph and second uses the target knowledge base to instantiate the generated template. For instance, moving to another domain based on a different knowledge base thus only affects parts of the approach so that the conversion effort is lessened.

QuASE [129] is a three stage open domain approach based on web search and the Freebase knowledge base¹¹. First, QuASE uses entity linking, semantic feature construction and candidate ranking on the input question. Then, it selects the documents and according sentences from a web search with a high probability to match the question and presents them as answers to the user.

DEV-NLQ [128] is based on lambda calculus and an event-based triple store¹² using only triple based retrieval operations. DEV-NLQ claims to be the only QA system able to solve chained, arbitrarily-nested, complex, prepositional phrases.

dbo	http://dbpedia.org/ontology/
dbr	http://dbpedia.org/resource/
owl	http://www.w3.org/2002/07/owl#

Table 3

URL prefixes used throughout this work.

⁹URL prefixes are defined in Table 3.

¹⁰<http://wiki.answers.com/>

¹¹<https://www.freebase.com/>

¹²<http://www.w3.org/wiki/LargeTripleStores>

CubeQA [74] is a novel approach of SQA over multi-dimensional statistical Linked Data using the RDF Data Cube Vocabulary¹³, which existing approaches cannot process. Using a corpus of questions with open domain statistical information needs, the authors analyze how those questions differ from others, which additional verbalizations are commonly used and how this influences design decisions for SQA on statistical data.

QAKiS [24,35,26] queries several multilingual versions of DBpedia at the same time by filling the produced SPARQL query with the corresponding language-dependent properties and classes. Thus, QAKiS can retrieve correct answers even in cases of missing information in the language-dependent knowledge base.

Freitas and Curry [54] evaluate a distributional-compositional semantics approach that is independent from manually created dictionaries but instead relies on co-occurring words in text corpora. The vector space over the set of terms in the corpus is used to create a distributional vector space based on the weighted term vectors for each concept. An inverted Lucene index is adapted to the chosen model.

Instead of querying a specific knowledge base, Sun et al. [129] use web search engines to extract relevant text snippets, which are then linked to Freebase, where a ranking function is applied and the highest ranked entity is returned as the answer.

HAWK [139] is the first hybrid source SQA system which processes Linked Data as well as textual information to answer one input query. HAWK uses an eight-fold pipeline comprising part-of-speech tagging, entity annotation, dependency parsing, linguistic pruning heuristics for an in-depth analysis of the natural language input, semantic annotation of properties and classes, the generation of basic triple patterns for each component of the input query as well as discarding queries containing not connected query graphs and ranking them afterwards.

SWIP (Semantic Web intercase using Pattern) [111] generates a pivot query, a hybrid structure between the natural language question and the formal SPARQL target query. Generating the pivot queries consists of three main steps: (1) Named entity identification, (2) Query focus identification and (3) subquery generation. To formalize the pivot queries, the query is mapped to linguistic patterns, which are created by hand from domain experts. If there are multiple applicable linguistic patterns for a pivot query, the user chooses between them.

¹³<http://www.w3.org/TR/vocab-data-cube/>

Hakimov et al. [66] adapt a semantic parsing algorithm to SQA which achieves a high performance but relies on large amounts of training data which is not practical when the domain is large or unspecified.

Several industry-driven SQA-related projects have emerged over the last years. For example, DeepQA of IBM Watson [64], which was able to win the Jeopardy! challenge against human experts.

YodaQA [13] is a modular open source hybrid approach built on top of the Apache UIMA framework¹⁴ that is part of the Brmsion platform and is inspired by DeepQA. YodaQA allows easy parallelization and leverage of pre-existing NLP UIMA components by representing each artifact (question, search result, passage, candidate answer) as a separate UIMA CAS. Yoda pipeline is divided in five different stages: (1) Question Analysis, (2) Answer Production, (3) Answer Analysis, (4) Answer Merging and Scoring as well as (5) Successive Refining.

Further, KAIST's Exobrain¹⁵ project aims to learn from large amounts of data while ensuring a natural interaction with end users. However, it is yet limited to Korean for the moment.

Answer Presentation Another, important part of SQA systems outside the SQA research challenges is result presentation. Verbose descriptions or plain URIs are uncomfortable for human reading. *Entity summarization* deals with different types and levels of abstractions.

Cheng et al. [31] proposes a random surfer model extended by a notion of centrality, i.e., a computation of the central elements involving similarity (or relatedness) between them as well as their informativeness. The similarity is given by a combination of the relatedness between their properties and their values.

Ngonga Ngomo et al. [106] present another approach that automatically generates natural language description of resources using their attributes. The rationale behind SPARQL2NL is to verbalize¹⁶ RDF data by applying templates together with the metadata of the schema itself (label, description, type). Entities can have multiple types as well as different levels of hierarchy which can lead to different levels of abstractions. The verbalization of the DBpedia entity `dbo:Microsoft` can vary depending on the type `dbo:Agent` rather than `dbo:Company`.

¹⁴<https://uima.apache.org/>

¹⁵<http://exobrain.kr/>

¹⁶For example, `"123"<http://dbpedia.org/datatype/squareKilometre>` can be verbalized as *123 square kilometres*.

Table 4

Different techniques for bridging the lexical gap along as well as an example of a deviation of the word “running”.

Identity	running
Similarity Measure	runnign
Stemming/Lemmatizing	run
AQE—Synonyms	sprint
Pattern libraries	<i>X</i> made a break for <i>Y</i>

5. Challenges

In this section, we address seven challenges that have to be faced by state-of-the-art SQA systems. All mentioned challenges are currently open research fields. For each challenge, we describe efforts mentioned in the 72 selected publications. Challenges that affect SQA, but that are not to be solved by SQA systems, such as speech interfaces, data quality and system interoperability, are analyzed in Shekarpour et al. [123].

5.1. Lexical Gap

In a natural language, the same meaning can be expressed in different ways. Natural language descriptions of RDF resources are provided by values of the `rdfs:label` property (*label* in the following). While synonyms for the same RDF resource can be modeled using multiple labels for that resource, knowledge bases typically don’t contain all the different terms that can refer to a certain entity. If the vocabulary used in a question is different from the one used in the labels of the knowledge base, we call this the *lexical gap*¹⁷ [66]. Because a question can usually only be answered if every referred concept is identified, bridging this gap significantly increases the proportion of questions that can be answered by a system. Table 4 shows the methods employed by the 72 selected publications for bridging the lexical gap along with examples. As an example of how the lexical gap is bridged outside of SQA, see Lee et al. [86].

String Normalization and Similarity Functions Normalizations, such as conversion to lower case or to base forms, such as “é,é,ê” to “e”, allow matching of slightly different forms (problem 3) and some simple mistakes (problem 4), such as “Deja Vu” for “déjà vu”, and are quickly implemented and executed. More elaborate normalizations use natural language programming (NLP)

techniques for stemming (both “running” and “ran” to “run”).

If normalizations are not enough, the distance—and its complementary concept, similarity—can be quantified using a *similarity function* and a threshold can be applied. Common examples of similarity functions are Jaro-Winkler, an edit-distance that measures transpositions and n-grams, which compares sets of substrings of length *n* of two strings. Also, one of the surveyed publications, Zhang et al. [150], uses the largest common substring, both between Japanese and translated English words. However, applying such similarity functions can carry harsh performance penalties. While an exact string match can be efficiently executed in a SPARQL triple pattern, similarity scores generally need to be calculated between a phrase and every entity label, which is infeasible on large knowledge bases [139]. For instance, edit distances of two characters or less can be mitigated by using the fuzzy query implementation of an Apache Lucene index¹⁸ which implements a Levenshtein Automaton [117]. Furthermore, Ngonga Ngomo [104] provides a different approach to efficiently calculating similarity scores that could be applied to QA. It uses *similarity metrics* where a triangle inequality holds that allows for a large portion of potential matches to be discarded early in the process. This solution is not as fast as using a Levenshtein Automaton but does not place such a tight limit on the edit distance.

Automatic Query Expansion Normalization and string similarity methods match different forms of the same word but not different words with similar meaning. Synonyms, like *design* and *plan*, are pairs of words that, either always or only in a specific context, have the same meaning. In hyper-hyponym-pairs, like *chemical process* and *photosynthesis*, the first word is less specific than the second one. These word pairs, taken from lexical databases such as WordNet [97], are used as additional labels in *Automatic query expansion (AQE)*. AQE is commonly used in information retrieval and traditional search engines, as summarized in Carpineto and Romano [29]. These additional surface forms allow for more matches and thus increase recall but lead to mismatches between related words and thus can decrease the precision.

In traditional document-based search engines with high recall and low precision, this trade-off is more common than in SQA. SQA is typically optimized for

¹⁷In linguistics, the term *lexical gap* has a different meaning, referring to a word that has no equivalent in another language.

¹⁸<http://lucene.apache.org>

concise answers and a high precision, since a SPARQL query with an incorrectly identified concept mostly results in a wrong set of answer resources. However, AQE can be used as a backup method in case there is no direct match. One of the surveyed publications is an experimental study [120] that evaluates the impact of AQE on SQA. It has analyzed different lexical¹⁹ and semantic²⁰ expansion features and used machine learning to optimize weightings for combinations of them. Both lexical and semantic features were shown to be beneficial on a benchmark dataset consisting only of sentences where direct matching is not sufficient.

Pattern libraries RDF individuals can be matched from a phrase to a resource with high accuracy using similarity functions and normalization alone. Properties however require further treatment, as (1) they determine the subject and object, which can be in different positions²¹ and (2) a single property can be expressed in many different ways, both as a noun and as a verb phrase which may not even be a continuous substring²² of the question. Because of the complex and varying structure of those linguistic patterns and the required reasoning and knowledge²³, libraries to overcome these issues have been developed.

PATTY [102] detects entities in sentences of a corpus and determines the shortest path between the entities. The path is then expanded with occurring modifiers and stored as a pattern. Thus, PATTY is able to build up a pattern library on any knowledge base with an accompanying corpus.

BOA [62] generates linguistic patterns using a corpus and a knowledge base. For each property in the knowledge base, sentences from a corpus are chosen containing examples of subjects and objects for this particular property. BOA assumes that each resource pair that is connected in a sentence exemplifies another label for this relation and thus generates a pattern from each occurrence of that word pair in the corpus.

PARALEX [48] contains phrase to concept mappings in a lexicon that is trained from a corpus of paraphrases from the QA site WikiAnswers. The advantage is that no manual templates have to be created as they are automatically learned from the paraphrases.

Entailment A corpus of already answered questions or linguistic question patterns can be used to infer the answer for new questions. A phrase *A* is said to *entail* a phrase *B*, if *B* follows from *A*. Thus, entailment is directional: Synonyms entail each other, whereas hyper- and hyponyms entail in one direction only: “birds fly” entails “sparrows fly”, but not the other way around. Ou and Zhu [107] generate possible questions for an ontology in advance and identify the most similar match to a user question based on a syntactic and semantic similarity score. The syntactic score is the cosine-similarity of the questions using bag-of-words. The semantic score also includes hypernyms, hyponyms and denormalizations based on WordNet [97]. While the preprocessing is algorithmically simple compared to the complex pipeline of NLP tools, the number of possible questions is expected to grow superlinearly with the size of the ontology so the approach is more suited to specific domain ontologies. Furthermore, the range of possible questions is quite limited which the authors aim to partially alleviate in future work by combining multiple *basic questions* into a *complex question*.

Document Retrieval Models for RDF resources Blanco et al. [18] adapt entity ranking models from traditional document retrieval algorithms to RDF data. The authors apply BM25 as well as tf-idf ranking function to an index structure with different text fields constructed from the title, object URIs, property values and RDF inlinks. The proposed adaptation is shown to be both time efficient and qualitatively superior to other state-of-the-art methods in ranking RDF resources.

Composite Approaches Elaborate approaches on bridging the lexical gap can have a high impact on the overall runtime performance of an SQA system. This can be partially mitigated by composing methods and executing each following step only if the one before did not return the expected results.

BELA [141] implements four layers. First, the question is mapped directly to the concept of the ontology using the index lookup. Second, the question is mapped based on Levenshtein distance to the ontology, if the Levenshtein distance of a word from the question and a property from an ontology exceed a certain threshold. Third, WordNet is used to find synonyms for a given word. Finally, BELA uses explicit semantic analysis (ESA) Gabilovich and Markovitch [59]. The evaluation is carried out on the QALD 2 [136] test dataset and shows that the more simple steps, like index lookup and Levenshtein distance, had the most positive influence

¹⁹lexical features include synonyms, hyper and hyponyms

²⁰semantic features making use of RDF graphs and the RDFS vocabulary, such as equivalent, sub- and superclasses

²¹E.g., “X wrote Y” and “Y is written by X”

²²E.g., “X wrote Y together with Z” for “X is a coauthor of Y”.

²³E.g., “if X writes a book, X is called the author of it.”

on answering questions so that many questions can be answered with simple mechanisms.

Park et al. [109] answer natural language questions via regular expressions and keyword queries with a Lucene-based index. Furthermore, the approach uses DBpedia [90] as well as their own triple extraction method on the English Wikipedia.

5.2. Ambiguity

Ambiguity is the phenomenon of the same phrase having different meanings; this can be structural and syntactic (like “flying planes”) or lexical and semantic (like “bank”). We distinguish between homonymy, where the same string accidentally refers to different concepts (as in money bank vs. river bank) and polysemy, where the same string refers to different but related concepts (as in bank as a company vs. bank as a building). We distinguish between synonymy and taxonomic relations such as metonymy and hypernymy. In contrast to the lexical gap, which impedes the recall of a SQA system, ambiguity negatively effects its precision. Ambiguity is the flipside of the lexical gap.

This problem is aggravated by the very methods used for overcoming the lexical gap. The more loose the matching criteria become (increase in recall), the more candidates are found which are generally less likely to be correct than closer ones. *Disambiguation* is the process of selecting one of multiple candidate concepts for an ambiguous phrase. We differentiate between two types of disambiguation based on the source and type of information used to solve this mapping:

Corpus-based methods are traditionally used and rely on counts, often used as probabilities, from unstructured text corpora. Such statistical approaches [125] are based on the *distributional hypothesis*, which states that “difference of meaning correlates with difference of [contextual] distribution” [69]. The *context* of a phrase is identified here as its central characteristic [98]. Common context features used are word co-occurrences, such as left or right neighbours, but also synonyms, hyponyms, POS-tags and the parse tree structure. More elaborate approaches also take advantage of the context outside of the question, such as past queries of the user [124].

In SQA, *Resource-based methods* exploit the fact that the candidate concepts are RDF resources. Resources are compared using different scoring schemes based of their properties and the connections between them. The assumption is that high score between all the resources chosen in the mapping implies a higher prob-

ability of those resources being related, and that this implies a higher probability of those resource being correctly chosen. RVT [63] uses *Hidden Markov Models (HMM)* to select the proper ontological triples according to the graph nature of DBpedia. CASIA [71] employs *Markov Logic Networks (MLN)*: First-order logic statements are assigned a numerical penalty, which is used to define hard constraints, like “each phrase can map to only one resource”, alongside soft constraints, like “the larger the semantic similarity is between two resources, the higher the chance is that they are connected by a relation in the question”. *Underspecification* [133] discards certain combinations of possible meanings before the time consuming querying step, by combining restrictions for each meaning. talks about their Each term is mapped to a *Dependency-based Underspecified Discourse REpresentation Structure (DUDE [32])*, which captures its possible meanings along with their class restrictions. Treo [56,55] performs entity recognition and disambiguation using Wikipedia-based semantic relatedness and spreading activation. Semantic relatedness calculates similarity values between pairs of RDF resources. Determining semantic relatedness between entity candidates associated to words in a sentence allows to find the most probable entity by maximizing the total relatedness. EasyESA [30] is based on distributional semantic models which allow to represent an entity by a vector of target words and thus compresses its representation. The distributional semantic models allow to bridge the lexical gap and resolve ambiguity by avoiding the explicit structures of RDF-based entity descriptions for entity linking and relatedness. gAnswer [77] tackles ambiguity with *RDF fragments*, i.e., star-like RDF subgraphs. The number of connections between the fragments of the resource candidates is then used to score and select them. Wikimantic [20] can be used to disambiguate short questions or even sentences. It uses Wikipedia article interlinks for a generative model, where the probability of an article to generate a term is set to the terms relative occurrence in the article. Disambiguation is then an optimization problem to locally maximize each article’s (and thus DBpedia resource’s) term probability along with a global ranking method. Shekarpour et al. [118,121] disambiguate resource candidates using segments consisting of one or more words from a keyword query. The aim is to maximize the high textual similarity of keywords to resources along with relatedness between the resources (classes, properties and entities). The problem is cast as a Hidden Markov Model (HMM) with the states representing the set of candi-

date resources extended by OWL reasoning. The transition probabilities are based on the shortest path between the resources. The Viterbi algorithm generates an optimal path though the HMM that is used for disambiguation. DEANNA [145,146] manages phrase detection, entity recognition and entity disambiguation by formulating the SQA task as an integer linear programming (ILP) problem. It employs *semantic coherence* which measures co-occurrence of resources in the same context. DEANNA constructs a disambiguation graph which encodes the selection of candidates for resources and properties. The chosen objective function maximizes the combined similarity while constraints guarantee that the selections are valid. The resulting problem is NP-hard but it is efficiently solvable in approximations by existing ILP solvers. The follow-up approach [147] uses DBpedia and Yago with a mapping of input queries to semantic relations based on text search. At QALD 2, it outperformed almost every other system on factoid questions and every other system on list questions. However, the approach requires detailed textual descriptions of entities and only creates basic graph pattern queries. LOD-Query [119] is a keyword-based SQA system that tackles both ambiguity and the lexical gap by selecting candidate concepts based on a combination of a string similarity score and the connectivity degree. The string similarity is the normalized edit distance between a labels and a keyword. The connectivity degree of a concept is approximated by the occurrence of that concept in all the triples of the knowledge base. *Pomelo* [67] answers biomedical questions on the combination of Drugbank, Diseases and Sider using owl:sameAs links between them. Properties are disambiguated using predefined rewriting rules which are categorized by context. Rani et al. [115] use fuzzy logic co-clustering algorithms to retrieve documents based on their ontology similarity. Possible senses for a word are assigned a probability depending on the context. Zhang et al. [150] translates RDF resources to the English DBpedia. It uses feedback learning in the disambiguation step to refine the resource mapping

Instead of trying to resolve ambiguity automatically, some approaches let the user clarify the exact intent, either in all cases or only for ambiguous phrases: SQUALL [50,51] defines controlled, English-based, vocabulary that is enhanced with knowledge from a given triple store. While this ideally results in a high performance, it moves the problem of the lexical gap and disambiguation fully to the user. As such, it covers a middle ground between SPARQL and full-fledged SQA with the author's intent that learning the grammatical

structure of this proposed language is easier for a non-expert than to learn SPARQL. A cooperative approach that places less of a burden on the user is proposed in [95], which transforms the question into a discourse representation structure and starts a dialogue with the user for all occurring ambiguities. CrowdQ [43] is a SQA system that decomposes complex queries into simple parts (keyword queries) and uses crowdsourcing for disambiguation. It avoids excessive usage of crowd resources by creating general templates as an intermediate step. FREyA (Feedback, Refinement and Extended Vocabulary Aggregation) [37] represents phrases as potential ontology concepts which are identified by heuristics on the syntactic parse tree. Ontology concepts are identified by matching their labels with phrases from the question without regarding its structure. A consolidation algorithm then matches both potential and ontology concepts. In case of ambiguities, feedback from the user is asked. Disambiguation candidates are created using string similarity in combination with WordNet synonym detection. The system learns from the user selections, thereby improving the precision over time. TBSL [135] uses both an domain independent and a domain dependent lexicon so that it performs well on specific topic but is still adaptable to a different domain. It uses AutoSPARQL [87] to refine the learned SPARQL using the QTL algorithm for supervised machine learning. The user marks certain answers as correct or incorrect and triggers a refinement. This is repeated until the user is satisfied with the result. An extension of TBSL is DEQA [89] which combines Web extraction with OXPath [58], interlinking with LIMES [105] and SQA with TBSL. It can thus answer complex questions about objects which are only available as HTML. Another extension of TBSL is ISOFT [108], which uses explicit semantic analysis to help bridging the lexical gap. NL-Graphs [47] combines SQA with an interactive visualization of the graph of triple patterns in the query which is close to the SPARQL query structure yet still intuitive to the user. Users that find errors in the query structure can either reformulate the query or modify the query graph. KOIOS [16] answers queries on natural environment indicators and allows the user to refine the answer to a keyword query by faceted search. Instead of relying on a given ontology, a schema index is generated from the triples and then connected with the keywords of the query. Ambiguity is resolved by user feedback on the top ranked results.

A different way to restrict the set of answer candidates and thus handle ambiguity is to determine the expected answer type of a factual question. The stan-

standard approach to determine this type is to identify the focus of the question and to map this type to an ontology class. In the example “Which books are written by Dan Brown?”, the focus is “books” which is mapped to `dbo:Book`. There is however a long tail of rare answer types that are not as easily alignable to an ontology, which, for instance, Watson [64] tackles using the TyCor [85] framework for type coercion. Instead of the standard approach, candidates are first generated using multiple interpretations and then selected based on a combination of scores. Besides trying to align the answer type directly, it is *coerced* into other types by calculating the probability of an entity of class A to also be in class B. DBpedia, Wikipedia and WordNet are used to determine link anchors, list memberships, synonyms, hyper- and hyponyms. The follow-up, Welty et al. [142] compare two different approaches for answer typing. Type-and-Generate (TaG) approaches restrict candidate answers to the expected answer types using predictive annotation, which requires manual analysis of a domain. TyCor on the other hand employs multiple strategies using generate-and-type (GaT), i.e., it generates all answers regardless of answer type and tries to coerce them into the expected answer type. Experimental results hint that GaT outperforms TaG when accuracy is higher than 50%. The significantly higher performance of TyCor when using GaT is explained by its robustness to incorrect candidates while there is no recovery from excluded answers from TaG.

5.3. Multilingualism

Knowledge on the Web is expressed in various languages. While RDF resources can be described in multiple languages at once using language tags, there is not a single language that is always used in Web documents. Partially because users want to use their native language in search queries. A more flexible approach is to have SQA systems that can handle multiple input languages, which may even differ from the language used to encode the knowledge. Deines and Krechel [41] use *GermaNet* [68] which is integrated into the multilingual knowledge base *EuroWordNet* [140] together with *lemon-LexInfo* [23], to answer German questions. Agarwal et al. [3] only need to successfully translate part of the query after which the recognition of the other entities is aided using semantic similarity and relatedness measures between resources connected to the initial ones in the knowledge base. *QAKiS (Question Answering wiKiframework-based system)* [35] automatically extends existing mappings between different lan-

guage versions of Wikipedia, which is carried over to DBpedia.

5.4. Complex Queries

Simple questions can most often be answered by translating into a set of simple triple pattern. Problems arise when several facts have to be found out, connected and then combined respectively the resulting query has to obey certain restrictions or modalities like a result order, aggregated or filtered results.

YAGO-QA [2] allows nested queries when the sub-query has already been answered, for example “Who is the governor of the state of New York?” after “What is the state of New York?” YAGO-QA extracts facts from Wikipedia (categories and infoboxes), WordNet and GeoNames. It contains different surface forms such as abbreviations and paraphrases for named entities.

PYTHIA [134] is an ontology-based SQA system with an automatically build ontology-specific lexicon. Due to the linguistic representation, the system is able to answer natural language question with linguistically more complex queries, involving quantifiers, numerals, comparisons and superlatives, negations and so on.

IBM’s Watson System [64] handles complex questions by first determining the focus element, which represents the searched entity. The information about the focus element is used to predict the lexical answer type and thus restrict the range of possible answers. This approach allows for indirect questions and multiple sentences.

Shekarpour et al. [118,121], as mentioned in Section 5.2, propose a model that use a combination of knowledge base concepts with a HMM model to handle complex queries.

Intui2 [44] is an SQA system based on DBpedia based on *synfragments* which map to a subtree of the syntactic parse tree. Semantically, a synfragment is a minimal span of text that can be interpreted as a RDF triple or complex RDF query. Synfragments interoperate with their parent synfragment by combining all combinations of child synfragments, ordered by syntactic and semantic characteristics. The authors assume that an interpretation of a question in any RDF query language can be obtained by the recursively interpretation of its synfragments. Intui3 [45] replaces self-made components with robust libraries such as the neural networks-based NLP toolkit SENNA and the DBpedia Lookup service. It drops the parser determined interpretation combination method of its predecessor that

suffered from bad sentence parses and instead uses a fixed order right-to-left combination.

GETARUNS [42] first creates a logical form out of a query which consists of a focus, a predicate and arguments. The focus element identifies the expected answer type. For example, the focus of “Who is the major of New York?” is “person”, the predicate “be” and the arguments “major of New York”. If no focus element is detected, a yes/no question is assumed. In the second step, the logical form is converted to a SPARQL query by mapping elements to resources via label matching. The resulting triple patterns are then split up again as properties are referenced by unions over both possible directions, as in $(\{?x ?p ?o\} \cup \{?o ?p ?x\})$ because the direction is not known beforehand. Additionally, there are filters to handle additional restrictions which cannot be expressed in a SPARQL query, such as “Who has been the 5th president of the USA”.

5.5. Distributed Knowledge

If concept information—which is referred to in a query—is represented by distributed RDF resources, information needed for answering it may be missing if only a single one or not all of the knowledge bases are found. In single datasets with a single source, such as DBpedia, however, most of the concepts have at most one corresponding resource. In case of combined datasets, this problem can be dealt with by creating *sameAs*, *equivalentClass* or *equivalentProperty* links, respectively. However, interlinking while answering a semantic query is a separate research area and thus not covered here.

Some questions are only answerable with multiple knowledge bases and we assume already created links for the sake of this survey. The ALOQUS [84] system tackles this problem by using the PROTON [38] upper level ontology first to phrase the queries. The ontology is then aligned to those of other knowledge bases using the BLOOMS [83] system. Complex queries are decomposed into separately handled subqueries after coreferences²⁴ are extracted and substituted. Finally, these alignments are used to execute the query on the target systems. In order to improve the speed and quality of the results, the alignments are filtered using a threshold on the confidence measure.

²⁴Such as “List the Semantic Web *people* and *their* affiliation.”, where the coreferent *their* refers to the entity *people*.

Herzig et al. [72] search for entities and consolidate results from multiple knowledge bases. Similarity metrics are used both to determine and rank results candidates of each datasource and to identify matches between entities from different datasources.

5.6. Procedural, Temporal and Spatial Questions

Procedural Questions Factual, list and yes-no questions are easiest to answer as they conform directly to SPARQL queries using SELECT and ASK. Others, such as why (causal) or how (procedural) questions require more additional processing. Procedural QA can currently not be solved by SQA, since, to the best of our knowledge, there are no existing knowledge bases that contain procedural knowledge. While it is not an SQA system, we describe the document-retrieval based KOMODO [27] to motivate further research in this area. Instead of an answer sentence, KOMODO returns a Web page with step-by-step instructions on how to reach the goal specified by the user. This reduces the problem difficulty as it is much easier to find a Web page which contains instructions on how to, for example, assemble a “Ikea Billy bookcase” than it would be to extract, parse and present the required steps to the user. Additionally, there are arguments explaining reasons for taking a step and warnings against deviation. Instead of extracting the sense of the question using an RDF knowledge base, KOMODO submits the question to a traditional search engine. The highest ranked returned pages are then cleaned and procedural text is identified using statistical distributions of certain POS tags.

In basic RDF, each fact, which is expressed by a triple, is assumed to be true, regardless of circumstances. In the real world and in natural language however, the truth value of many statements is not a constant but a function of either or both the location or time.

Temporal Questions Tao et al. [131] answer temporal question on clinical narratives. They introduce the Clinical Narrative Temporal Relation Ontology (CNTRO), which is based on Allen’s Interval Based Temporal Logic [6] but allows usage of time instants as well as intervals. This allows inferring the temporal relation of events from those of others, for example by using the transitivity of *before* and *after*. In CNTRO, measurement, results or actions done on patients are modeled as events whose time is either absolutely specified in date and optionally time of day or alternatively in relations to other events and times. The framework also

includes an SWRL [76] based reasoner that can deduce additional time information. This allows the detection of possible causalities, such as between a therapy for a disease and its cure in a patient.

Melo et al. [95] propose to include the implicit temporal and spatial context of the user in a dialog in order to resolve ambiguities. It also includes spatial, temporal and other implicit information.

QALL-ME [49] is a multilingual framework based on description logics and uses the spatial and temporal context of the question. If this context is not explicitly given, the location and time are of the user posing the question are added to the query. This context is also used to determine the language used for the answer, which can differ from the language of the question.

Spatial Questions In RDF, a location can be expressed as 2-dimensional geocoordinates with latitude and longitude, while three-dimensional representations (e.g. with additional height) are not supported by the most often used schema²⁵. Alternatively, spatial relationships can be modeled which are easier to answer as users typically ask for relationships and not exact geocoordinates.

Younis et al. [149] employ an inverted index for named entity recognition that enriches semantic data with spatial relationships such as crossing, inclusion and nearness. This information is then made available for SPARQL queries.

5.7. Templates

For complex questions, where the resulting SPARQL query contains more than one basic graph pattern, sophisticated approaches are required to capture the structure of the underlying query. Current research follows two paths, namely (1) template based approaches, which map input questions to either manually or automatically created SPARQL query templates or (2) template-free approaches that try to build SPARQL queries based on the given syntactic structure of the input question.

For the first solution, many (1) template-driven approaches have been proposed like TBSL [135] or SINA [118,121]. Furthermore, Casia [70] generates the graph pattern templates by using the question type, named entities and POS tags techniques. The generated graph patterns are then mapped to resources using

WordNet, PATTY and similarity measures. Finally, the possible graph pattern combinations are used to build SPARQL queries. The system focuses in the generation of SPARQL queries that do not need filter conditions, aggregations and superlatives.

Ben Abacha and Zweigenbaum [14] focus on a narrow medical patients-treatment domain and use manually created templates alongside machine learning.

Damova et al. [39] return well formulated natural language sentences that are created using a template with optional parameters for the domain of paintings. Between the input query and the SPARQL query, the system places the intermediate step of a multilingual description using the Grammatical Framework [116], which enables the system to support 15 languages.

Rahoman and Ichise [114] propose a template based approach using keywords as input. Templates are automatically constructed from the knowledge base.

However, (2) template-free approaches require additional effort of making sure to cover every possible basic graph pattern [139]. Thus, only a few SQA systems tackle this approach so far.

Xser [144] first assigns semantic labels, i.e., variables, entities, relations and categories, to phrases by casting them to a sequence labelling pattern recognition problem which is then solved by a structured perceptron. The perceptron is trained using features including n-grams of POS tags, NER tags and words. Thus, Xser is capable of covering any complex basic graph pattern.

Going beyond SPARQL queries is TPSM, the open domain Three-Phases Semantic Mapping [61] framework. It maps natural language questions to OWL queries using Fuzzy Constraint Satisfaction Problems. Constraints include surface text matching, preference of POS tags and the similarity degree of surface forms. The set of correct mapping elements acquired using the FCSP-SM algorithm is combined into a model using predefined templates.

An extension of gAnswer [151] (see Section 5.2) is based on question understanding and query evaluation. First, their approach uses a relation mining algorithm to find triple patterns in queries as well as relation extraction, POS-tagging and dependency parsing. Second, the approach tries to find a matching subgraph for the extracted triples and scores them based on a confidence score. Finally, the top-k subgraph matches are returned. Their evaluation on QALD 3 shows that mapping NL questions to graph pattern is not as powerful as generating SPARQL (template) queries with respect to aggregation and filter functions needed to answer several benchmark input questions.

²⁵see http://www.w3.org/2003/01/geo/wgs84_pos at <http://lodstats.aksw.org>

6. Conclusion

In this survey, we analyzed 62 systems and their contributions to seven challenges for SQA systems. Semantic question answering is an active and upcoming research field with many existing and diverse approaches covering a multitude of research challenges, domains and knowledge bases.

We only cover QA on the Semantic Web, that is, approaches that retrieve resources as Linked Data from RDF knowledge bases. As similar challenges are faced by QA unrelated to the Semantic Web, we refer to Section 3. We choose to not go into detail for approaches that do not retrieve resources from RDF knowledge bases. Moreover, our consensus can be found in Table 6 for best practices. The upcoming HOBBIT²⁶

²⁶<http://project-hobbit.eu/>

Table 5

Number of publications per year per addressed challenge. Percentages are given for the fully covered years 2011–2014 separately and for the whole covered timespan, with 1 decimal place. For a full list, see Table 7.

Year	Total	Lexical Gap	Ambiguity	Multilingualism	Complex Operators	Distributed Knowledge	Procedural, Temporal or Spatial	Templates
absolute								
2010	1	0	0	0	0	0	1	0
2011	16	11	12	1	3	1	2	2
2012	14	6	7	1	2	1	1	4
2013	20	18	12	2	5	1	1	5
2014	13	7	8	1	2	0	1	0
2015	6	5	3	1	0	1	0	0
all	70	46	42	6	12	4	6	11
percentage								
2011		68.8	75.0	6.3	18.8	6.3	12.5	12.5
2012		42.9	50.0	7.1	14.3	7.1	7.1	28.6
2013		85.0	60.0	10.0	25.0	5.0	5.0	25.0
2014		53.8	61.5	7.7	15.4	7.7	7.7	0.0
all		65.7	60.0	8.6	17.1	5.7	8.6	15.7

project will clarify, which modules can be aligned with state-of-the-art performance and will quantify the impact of those modules. To cover the field of SQA in depth, we excluded works solely about similarity [12] or paraphrases [15]. The existence of common SQA challenges implies that a unifying architecture can improve the precision as well as increase the number of answered questions [94]. Research into such architectures, includes openQA [94], OAQA [148], QALLME [49] and QANUS [103] (see Section 3.3). Our goal, however, is not to quantify submodule performance or interplay. That will be the task of upcoming projects of large consortiums. A new community²⁷ is forming in that field and did not find a satisfying solution yet.²⁸ In this section, we discuss each of the seven research challenges and give a short overview of already established as well as future research directions per challenge, see Table 6.

Overall, the authors of this survey cannot observe a research drift to any of the challenges. The number of publications in a certain research challenge does not decrease significantly, which can be seen as an indicator that none of the challenges is solved yet – see Table 5. Naturally, since only a small number of publications addressed each challenge in a given year, one cannot draw statistically valid conclusions. The challenges proposed by Cimiano and Minock [33] and reduced within this survey appear to be still valid.

Bridging the (1) lexical gap has to be tackled by every SQA system in order to retrieve results with a high recall. For named entities, this is commonly achieved using a combination of the reliable and mature natural language processing algorithms for string similarity and either stemming or lemmatization, see Table 6. Automatic Query Expansion (AQE), for example with WordNet synonyms, is prevalent in information retrieval but only rarely used in SQA. Despite its potential negative effects on precision²⁹, we consider it a net benefit to SQA systems. Current SQA systems duplicate already existing efforts or fail to decide on the right technique. Thus, reusable libraries to lower the entrance effort to SQA systems are needed. Mapping to RDF properties from verb phrases is much harder, as they show more variation and often occur at multiple places of a question. Pattern libraries, such as

²⁷<https://www.w3.org/community/nli/>

²⁸<http://eis.iai.uni-bonn.de/blog/2015/11/the-2nd-annual-meetup-of-question-answering-community/>

²⁹Synonyms and other related words almost never have exactly the same meaning.

Table 6
Established and actively researched as well as envisioned techniques
for solving each challenge.

Challenge	Established	Future
Lexical Gap	stemming, lemmatization, string similarity, synonyms, vector space model, indexing, pattern libraries, explicit semantic analysis	combined efforts, reuse of libraries
Ambiguity	user information (history, time, location), underspecification, machine learning, spreading activation, semantic similarity, crowdsourcing, Markov Logic Network	holistic, knowledge-base aware systems
Multilingualism	translation to core language, language-dependent grammar	usage of multilingual knowledge bases
Complex Operators	reuse of former answers, syntactic tree-based formulation, answer type orientation, HMM, logic	non-factual questions, domain-independence
Distributed Knowledge and Procedural, Temporal, Spatial	temporal logic	domain specific adaptors, procedural SQA
Templates	fixed SPARQL templates, template generation, syntactic tree based generation	complex questions

BOA [62], can improve property identification, however they are still an active research topic and are specific to a knowledge base.

The next challenge, (2) ambiguity is addressed by the majority of the publications but the percentage does not increase over time, presumably because of use cases with small knowledge bases, where its impact is minuscule. For systems intended for longtime usage by the same persons, we regard as promising integration of previous questions, time and location, as is already common in web of document search engines. There is a variety of established disambiguation methods, which use the context of a phrase to determine the most likely RDF resource, some of which are based on unstructured text collections and others on RDF resources. As we could make out no clear winner, we recommend system developers to make their decisions based on the resources (such as query logs, ontologies, thesauri) available to them. Many approaches reinvent disambiguation efforts and thus—like for the lexical gap—holistic, knowledge-base aware, reusable systems are needed to facilitate faster research.

Despite its inclusion since QALD 3 and following, publications dealing with (3) multilingualism remain a small minority. Automatic translation of parts of or the whole query requires the least development effort, but suffers from imperfect translations. A higher quality can be achieved by using components, such as parsers and synonym libraries, for multiple languages. A possible future research direction is to make use of various language versions at once to use the power of a unified graph [35]. For instance, DBpedia [90] provides

a knowledge base in more than 100 languages which could form the base of a next multilingual SQA system.

Complex operators (4) seem to be used only in specific tasks or factual questions. Most systems either use the syntactic structure of the question or some form of knowledge-base aware logic. Future research will be directed towards domain-independence as well as non-factual queries.

Approaches using (5) distributed knowledge as well as those incorporating (6) procedural, temporal and spatial data remain niches. Procedural SQA does not exist yet as present approaches return unstructured text in the form of already written step-by-step instructions. While we consider future development of procedural SQA as feasible with the existing techniques, as far as we know there is no RDF vocabulary for and knowledge base with procedural knowledge yet.

The (7) templates challenge which subsumes the question of mapping a question to a query structure is still unsolved. Although the development of template based approaches seems to have decreased in 2014, presumably because of their low flexibility on open domain tasks, this still presents the fastest way to develop a novel SQA system but the limitation to simple query structures has yet to be overcome.

Future research should be directed at more modularization, automatic reuse, self-wiring and encapsulated modules with their own benchmarks and evaluations. Thus, novel research field can be tackled by reusing already existing parts and focusing on the research core problem itself. A step in this direction is QANARY [21], which describes how to modularize QA systems by providing a core QA vocabulary

against which existing vocabularies are bound. Another research direction are SQA systems as aggregators or framework for other systems or algorithms to benefit of the set of existing approaches. Furthermore, benchmarking will move to single algorithmic modules instead of benchmarking a system as a whole. The target of local optimization is benchmarking a process at the individual steps, but global benchmarking is still needed to measure the impact of error propagation across the chain. A Turing test-like spirit would suggest that the latter is more important, as the local measure are never fully representative. Additionally, we foresee the move from factual benchmarks over common sense knowledge to more domain specific questions without purely factual answers. Thus, there is a movement towards multilingual, multi-knowledge-source SQA systems that are capable of understanding noisy, human natural language input.

Acknowledgements This work has been supported by the FP7 project GeoKnow (GA No. 318159), the BMWI Project SAKE (Project No. 01MD15006E) and by the Eurostars projects DIESEL (E!9367) and QAMEL (E!9725) as well as the European Union's H2020 research and innovation action HOBBIT (GA 688227).

Table 7: Surveyed publications from November 2010 to July of 2015, inclusive, along with the challenges they explicitly address and the approach or system they belong to. Additionally annotated is the use light expressions as well as the use of intermediate templates. In case the system or approach is not named in the publication, a name is generated using the last name of the first author and the year of the first included publication.

publication	system or approach	Year	Lexical Gap	Ambiguity	Multilingualism	Complex Operators	Distributed Knowledge	Procedural, Temporal or Spatial Templates	QALD 1	QALD 2	QALD 3	QALD 4
Tao et al. [131]	Tao10	2010						✓				
Adolphs et al. [2]	YAGO-QA	2011	✓	✓								
Blanco et al. [18]	Blanco11	2011	✓									
Canitrot et al. [27]	KOMODO	2011	✓	✓								
Damljanovic et al. [37]	FREyA	2011	✓	✓						✓		
Ferrandez et al. [49]	QALL-ME	2011			✓			✓				
Freitas et al. [56]	Treo	2011	✓	✓		✓	✓					
Gao et al. [61]	TPSM	2011	✓	✓								
Kalyanpur et al. [85]	Watson	2011		✓								
Melo et al. [95]	Melo11	2011	✓	✓								
Moussa and Abdel-Kader [100]	QASYO	2011	✓					✓				
Ou and Zhu [107]	Ou11	2011	✓	✓		✓		✓				
Shen et al. [124]	Shen11	2011		✓								
Unger and Cimiano [133]	Pythia	2011		✓								
Unger and Cimiano [134]	Pythia	2011	✓	✓		✓		✓				
Bicer et al. [16]	KOIOS	2011	✓	✓								
Freitas et al. [55]	Treo	2011								✓		
Ben Abacha and Zweigenbaum [14]	MM+/BIO-CRF-H	2012						✓				
Boston et al. [20]	Wikimantic	2012		✓								
Gliozzo and Kalyanpur [64]	Watson	2012				✓						
Joshi et al. [84]	ALOQUS	2012				✓	✓					
Lehmann et al. [89]	DEQA	2012	✓	✓				✓				
Yahya et al. [145]	DEANNA	2012										
Yahya et al. [146]	DEANNA	2012	✓	✓								
Shekarpour et al. [118]	SINA	2012	✓	✓								
Unger et al. [135]	TBSL	2012	✓	✓				✓				
Walter et al. [141]	BELA	2012	✓	✓	✓			✓				
Younis et al. [149]	Younis12	2012	✓					✓				
Welty et al. [142]	Watson	2012		✓								
Elbedweihy et al. [46]	Elbedweihy12	2012									✓	
Cabrio et al. [24]	QAKiS	2012									✓	
Demartini et al. [43]	CrowdQ	2013	✓			✓						

Aggarwal et al. [3]	Aggarwal12	2013	✓	✓	✓			
Deines and Krechel [41]	GermanNLI	2013	✓					
Dima [44]	Intui2	2013	✓			✓		
Fader et al. [48]	PARALEX	2013	✓	✓			✓	✓
Ferre [50]	SQUALL2SPARQL	2013	✓			✓		✓
Giannone et al. [63]	RTV	2013	✓					✓
Hakimov et al. [65]	Hakimov13	2013	✓	✓				✓
He et al. [70]	CASIA	2013	✓	✓		✓		✓
Herzig et al. [72]	CRM	2013	✓	✓		✓		
Huang and Zou [77]	gAnswer	2013	✓	✓				
Pradel et al. [111]	SWIP	2013	✓					✓
Rahoman and Ichise [114]	Rahoman13	2013					✓	
Shekarpour et al. [121]	SINA	2013	✓	✓			✓	
Shekarpour et al. [120]	SINA	2013	✓					
Shekarpour et al. [119]	SINA	2013	✓	✓			✓	
Delmonte [42]	GETARUNS	2013	✓	✓		✓		
Cojan et al. [35]	QAKiS	2013				✓		
Yahya et al. [147]	SPOX	2013	✓	✓				
Zhang et al. [150]	Kawamura13	2013	✓	✓				
Carvalho et al. [30]	EasyESA	2014	✓	✓				
Rani et al. [115]	Rani14	2014				✓		
Zou et al. [151]	Zhou14	2014				✓		
Stewart [128]	DEV-NLQ	2014				✓		
Höffner and Lehmann [75]	CubeQA	2014	✓	✓			✓	
Cabrio et al. [26]	QAKiS	2014				✓		
Freitas and Curry [54]	Freitas14	2014	✓	✓				
Dima [45]	Intui3	2014	✓			✓		✓
Hamon et al. [67]	POMELO	2014		✓				✓
Park et al. [108]	ISOFT	2014	✓					✓
He et al. [71]	CASIA	2014	✓	✓				✓
Xu et al. [144]	Xser	2014	✓					✓
Elbedweihy et al. [47]	NL-Graphs	2014		✓				
Sun et al. [130]	QuASE	2015	✓	✓				
Park et al. [109]	Park15	2015	✓	✓				
Damova et al. [39]	MOLTO	2015				✓		
Sun et al. [129]	QuASE	2015	✓	✓				
Usbeck et al. [139]	HAWK	2015	✓			✓		✓
Hakimov et al. [66]	Hakimov15	2015	✓					

References

- [1] *BioASQ 2013—Biomedical Semantic Indexing and Question Answering*, volume 1094 of *CEUR Workshop Proceedings*, 2013.
- [2] P. Adolphs, M. Theobald, U. Schäfer, H. Uszkoreit, and G. Weikum. YAGO-QA: Answering questions by structured knowledge queries. In *ICSC 2011 IEEE Fifth International Conference on Semantic Computing*, pages 158–161, Los Alamitos, CA, USA, 2011. IEEE Computer Society.
- [3] N. Aggarwal, T. Polajnar, and P. Buitelaar. Cross-lingual natural language querying over the Web of Data. In E. Métais, F. Meziane, M. S. V. Sugumaran, and S. Vadera, editors, *18th International Conference on Applications of Natural Language to Information Systems, NLDB 2013*, pages 152–163. Springer-Verlag, Berlin Heidelberg, Germany, 2013.
- [4] E. Agichtein, D. Carmel, D. Pelleg, Y. Pinter, and D. Harman. Overview of the TREC 2015 LiveQA track. In E. M. Voorhees and A. Ellis, editors, *Proceedings of The Twenty-Fourth Text REtrieval Conference, TREC 2015, Gaithersburg, Maryland, USA, November 17-20, 2015*, volume Special Publication 500-319. National Institute of Standards and Technology (NIST), 2015. URL <http://trec.nist.gov/pubs/trec24/papers/Overview-QA.pdf>.
- [5] H. Alani, L. Kagal, A. Fokoue, P. Groth, C. Biemann, J. X. Parreira, L. Aroyo, N. Noy, C. Welty, and K. Janowicz, editors. *The Semantic Web—ISWC 2013*, Berlin Heidelberg, Germany, 2013. Springer-Verlag.
- [6] J. F. Allen. Maintaining knowledge about temporal intervals. *Communications of the ACM*, 26(11):832–843, 1983.
- [7] G. Antoniou, M. Grobelnik, E. Simperl, B. Parsia, D. Plexousakis, P. de Leenheer, and J. Z. Pan, editors. *The Semantic Web: Research and applications*, volume 6643 of *Lecture Notes in Computer Science*, Berlin Heidelberg, Germany, 2011. Semantic Technology Institute International (STI2), Springer-Verlag.
- [8] L. Aroyo, H. Welty, Chris Alani, J. Taylor, A. Bernstein, L. Kagal, N. Noy, and E. Blomqvist, editors. *The Semantic Web—ISWC 2011*, Berlin Heidelberg, Germany, 2011. Springer-Verlag.
- [9] S. Athenikos and H. Han. Biomedical Question Answering: A survey. *Computer methods and programs in biomedicine*, 99(1):1–24, 2010.
- [10] G. Balikas, A. Kosmopoulos, A. Krithara, G. Paliouras, and I. Kakadiaris. Results of the BioASQ track of the question answering lab at clef 2014. In *CLEF 2014 Working Notes*, volume 1180 of *CEUR Workshop Proceedings*, 2014.
- [11] G. Balikas, A. Kosmopoulos, A. Krithara, G. Paliouras, and I. Kakadiaris. Results of the BioASQ tasks of the question answering lab at clef 2015. In *CEUR Workshop Proceedings*, volume 1391 of *CEUR Workshop Proceedings*, 2015.
- [12] D. Bär, C. Biemann, I. Gurevych, and T. Zesch. Ukp: Computing semantic textual similarity by combining multiple content similarity measures. In *Proceedings of the First Joint Conference on Lexical and Computational Semantics—Volume 1: Proceedings of the Main Conference and the Shared Task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation, SemEval ’12*, pages 435–440, Stroudsburg, PA, USA, 2012. Association for Computational Linguistics. URL <http://dl.acm.org/citation.cfm?id=2387636.2387707>.
- [13] P. Baudiš and J. Šedivý. Modeling of the question answering task in the yodaqa system. In *Experimental IR Meets Multilinguality, Multimodality, and Interaction*, pages 222–228. Springer, 2015.
- [14] A. Ben Abacha and P. Zweigenbaum. Medical Question Answering: Translating medical questions into SPARQL queries. In *IHI ’12: Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium*, pages 41–50, New York, NY, USA, 2012. Association for Computational Linguistics.
- [15] J. Berant and P. Liang. Semantic parsing via paraphrasing. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1415–1425, Baltimore, Maryland, June 2014. Association for Computational Linguistics. URL <http://www.aclweb.org/anthology/P/P14/P14-1133>.
- [16] V. Bicer, T. Tran, A. Abecker, and R. Nedkov. Koios: Utilizing semantic search for easy-access and visualization of structured environmental data. In *The Semantic Web—ISWC 2011*, pages 1–16. Springer, 2011.
- [17] C. Biemann, S. Handschuh, A. Freitas, F. Meziane, and E. Métais. *Natural Language Processing and Information Systems: 20th International Conference on Applications of Natural Language to Information Systems, NLDB 2015, Passau, Germany, June 17-19, 2015, Proceedings*, volume 9103 of *Lecture Notes in Computer Science*. Springer, Berlin Heidelberg, Germany, 2015.
- [18] R. Blanco, P. Mika, and S. Vigna. Effective and efficient entity search in RDF data. In L. Aroyo, C. Welty, H. Alani, J. Taylor, and A. Bernstein, editors, *Proceedings of the 10th International Conference on The Semantic Web—Volume Part I*, volume 7031 of *Lecture Notes in Computer Science*, pages 83–97, Berlin Heidelberg, Germany, 2011. Springer-Verlag. ISBN 978-3-642-25072-9. URL <http://dl.acm.org/citation.cfm?id=2063016.2063023>.
- [19] A. Bordes, J. Weston, and N. Usunier. Open question answering with weakly supervised embedding models. In *Machine Learning and Knowledge Discovery in Databases*, pages 165–180. Springer, 2014.
- [20] C. Boston, S. Carberry, and H. Fang. Wikimantic: Disambiguation for short queries. In G. Bouma, A. Ittoo, E. Métais, and H. Wortmann, editors, *17th International Conference on Applications of Natural Language to Information Systems, NLDB 2012*, pages 140–151, Berlin Heidelberg, Germany, 2012. Springer-Verlag.
- [21] A. Both, D. Diefenbach, K. Singh, S. Shekarpour, D. Cherix, and C. Lange34. Qanary—An extensible vocabulary for open Question Answering systems. ESWC, 2016.
- [22] G. Bouma, A. Ittoo, E. Métais, and H. Wortmann, editors. *Natural Language Processing and Information Systems: 17th International Conference on Applications of Natural Language to Information Systems, NLDB 2012, Groningen, The Netherlands, June 26-28, 2012, Proceedings*, volume 7337 of *Lecture Notes in Computer Science*, Berlin Heidelberg, Germany, 2012. Springer.
- [23] P. Buitelaar, P. Cimiano, P. Haase, and M. Sintek. Towards linguistically grounded ontologies. In L. Aroyo, P. Traverso, F. Ciravegna, P. Cimiano, T. Heath, E. Hyvönen, R. Mizoguchi, E. Oren, M. Sabou, and E. Simperl, editors, *The Semantic Web: Research and Applications, 6th European Semantic Web Conference, ESWC 2009*, volume 6643 of *Lecture Notes in Computer Science*, pages 111–125, Berlin Heidelberg, Germany,

2009. Springer-Verlag.
- [24] E. Cabrio, A. P. Aprosio, J. Cojan, B. Magnini, F. Gandon, and A. Lavelli. QAKiS @ QALD-2. In *Interacting with Linked Data (ILD 2012)*, page 61, 2012.
- [25] E. Cabrio, P. Cimiano, V. Lopez, A. N. Ngomo, C. Unger, and S. Walter. QALD-3: Multilingual Question Answering over Linked Data. In *Working Notes for CLEF 2013 Conference*, 2013.
- [26] E. Cabrio, J. Cojan, F. Gandon, and A. Hallili. Querying multilingual DBpedia with QAKiS. In *The Semantic Web: ESWC 2013 Satellite Events*, pages 194–198. Springer-Verlag, Berlin Heidelberg, Germany, 2013.
- [27] M. Canitrot, T. de Filippo, P.-Y. Roger, and P. Saint-Dizier. The KOMODO system: Getting recommendations on how to realize an action via Question-Answering. In *IJCNLP 2011, Proceedings of the KRAQ11 Workshop: Knowledge and Reasoning for Answering Questions*, pages 1–9, 2011.
- [28] L. Cappellato, N. Ferro, M. Halvey, and W. Kraaij, editors. *Working Notes for CLEF 2014 Conference*, volume 1180 of *CEUR Workshop Proceedings*, 2014. URL <http://ceur-ws.org/Vol-1180>.
- [29] C. Carpineto and G. Romano. A survey of Automatic Query Expansion in Information Retrieval. *ACM Computing Surveys*, 44(1):1:1–1:50, Jan. 2012. ISSN 0360-0300. .
- [30] D. Carvalho, C. Calli, A. Freitas, and E. Curry. Easyesa: A low-effort infrastructure for explicit semantic analysis. In *Proc. of 13th International Semantic Web Conf*, pages 177–180, 2014.
- [31] G. Cheng, T. Tran, and Y. Qu. RELIN: Relatedness and informativeness-based centrality for entity summarization. In L. Aroyo, C. Welty, H. Alani, J. Taylor, A. Bernstein, L. Kagal, N. F. Noy, and E. Blomqvist, editors, *The Semantic Web – ISWC 2011*, volume 7031 of *Lecture Notes in Computer Science*, pages 114–129, Berlin Heidelberg, Germany, 2011. Springer.
- [32] P. Cimiano. Flexible semantic composition with dudes. In *Proceedings of the Eighth International Conference on Computational Semantics*, pages 272–276. Association for Computational Linguistics, 2009.
- [33] P. Cimiano and M. Minock. Natural language interfaces: What is the problem?—A data-driven quantitative analysis. In *Natural Language Processing and Information Systems, 15th International Conference on Applications of Natural Language to Information Systems, NLDB 2010*, volume 6177 of *Lecture Notes in Computer Science*, pages 192–206, Berlin Heidelberg, Germany, 2010. Springer-Verlag.
- [34] P. Cimiano, O. Corcho, V. Presutti, L. Hollink, and S. Rudolph, editors. *The Semantic Web: Semantics and Big Data*, volume 7882 of *Lecture Notes in Computer Science*, Berlin Heidelberg, Germany, 2013. Semantic Technology Institute International (STI2), Springer-Verlag.
- [35] J. Cojan, E. Cabrio, and F. Gandon. Filling the gaps among dbpedia multilingual chapters for question answering. In *Proceedings of the 5th Annual ACM Web Science Conference*, pages 33–42. ACM, 2013.
- [36] P. Cudré-Mauroux, J. Hefflin, E. Sirin, T. Tudorache, J. Euzenat, M. Hauswirth, J. X. Parreira, J. Hendler, G. Schreiber, A. Bernstein, and E. Blomqvist, editors. *The Semantic Web—ISWC 2012*, Berlin Heidelberg, Germany, 2012. Springer-Verlag.
- [37] D. Damljanovic, M. Agatonovic, and H. Cunningham. FREYA: An interactive way of querying Linked Data using natural language. In *Proceedings of the 1st Workshop on Question Answering Over Linked Data (QALD-1)*, Co-located with the 8th Extended Semantic Web Conference, pages 10–23, 2011.
- [38] M. Damova, A. Kiryakov, K. Simov, and S. Petrov. Mapping the central LOD ontologies to PROTON upper-level ontology. In P. Shvaiko, J. Euzenat, F. Giunchiglia, H. Stuckenschmidt, M. Mao, and I. Cruz, editors, *Proceedings of the 5th International Workshop on Ontology Matching (OM-2010) collocated with the 9th International Semantic Web Conference (ISWC-2010)*, 2010.
- [39] M. Damova, D. Dannells, R. Enache, M. Mateva, and A. Ranta. Natural language interaction with semantic web knowledge bases and LOD. *Towards the Multilingual Semantic Web*, 2013.
- [40] H. T. Dang, D. Kelly, and J. J. Lin. Overview of the TREC 2007 question answering track. In E. M. Voorhees and L. P. Buckland, editors, *Proceedings of The Sixteenth Text REtrieval Conference, TREC 2007, Gaithersburg, Maryland, USA, November 5-9, 2007*, volume Special Publication 500-274. National Institute of Standards and Technology (NIST), 2007. URL <http://trec.nist.gov/pubs/trec16/papers/QA.OVERVIEW16.pdf>.
- [41] I. Deines and D. Krechel. A German natural language interface for Semantic Search. In H. Takeda, Y. Qu, R. Mizoguchi, and Y. Kitamura, editors, *Semantic Technology*, volume 7774 of *Lecture Notes in Computer Science*, pages 278–289. Springer-Verlag, Berlin Heidelberg, Germany, 2013. .
- [42] R. Delmonte. Computational linguistic text processing—Lexicon, grammar, parsing and anaphora resolution. 2008.
- [43] G. Demartini, B. Trushkowsky, T. Kraska, and M. J. Franklin. Crowdsql: Crowdsourced query understanding. In *CIDR 2013, Sixth Biennial Conference on Innovative Data Systems Research, Asilomar, CA, USA, January 6-9, 2013, Online Proceedings*. www.cidrdb.org, 2013. URL http://www.cidrdb.org/cidr2013/Papers/CIDR13_Paper137.pdf.
- [44] C. Dima. Intui2: A prototype system for Question Answering over Linked Data. In *Working Notes for CLEF 2013 Conference*, 2013.
- [45] C. Dima. Answering natural language questions with Intui3. In *Working Notes for CLEF 2014 Conference*, 2014.
- [46] K. Elbedweihy, S. N. Wrigley, and F. Ciravegna. Improving semantic search using query log analysis. In *Interacting with Linked Data (ILD 2012)*, page 61, 2012.
- [47] K. Elbedweihy, S. Mazumdar, S. N. Wrigley, and F. Ciravegna. NL-Graphs: A hybrid approach toward interactively querying semantic data. In *The Semantic Web: Trends and Challenges*, pages 565–579. Springer, 2014.
- [48] A. Fader, L. Zettlemoyer, and O. Etzioni. Paraphrase-Driven Learning for Open Question Answering. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1608–1618, 2013.
- [49] O. Ferrandez, C. Spurk, M. Kouylekov, I. Dornescu, S. Ferrandez, M. Negri, R. Izquierdo, D. Tomas, C. Orasan, G. Neumann, et al. The QALL-ME framework: A specifiable-domain multilingual Question Answering architecture. *Journal of Web Semantics*, 9(2):137–145, 2011.
- [50] S. Ferre. SQUALL2SPARQL: A translator from controlled english to full SPARQL 1.1. In *Working Notes for CLEF 2013 Conference*, 2013.

- [51] S. Ferré. Squall: a controlled natural language as expressive as sparql 1.1. In *Natural Language Processing and Information Systems*, pages 114–125. Springer-Verlag, Berlin Heidelberg, Germany, 2013.
- [52] T. Finin, L. Ding, R. Pan, A. Joshi, P. Kolari, A. Java, and Y. Peng. Swoogle: Searching for knowledge on the semantic web. In *Proceedings of the National Conference on Artificial Intelligence*, volume 20, page 1682. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2005.
- [53] P. Forner, R. Navigli, D. Tufis, and N. Ferro, editors. *Working Notes for CLEF 2013 Conference*, volume 1179 of *CEUR Workshop Proceedings*, 2013. URL <http://ceur-ws.org/Vol-1179>.
- [54] A. Freitas and E. Curry. Natural language queries over heterogeneous linked data graphs: A distributional-compositional semantics approach. In *Proceedings of the 19th international conference on Intelligent User Interfaces*, pages 279–288. ACM, 2014.
- [55] A. Freitas, J. G. Oliveira, E. Curry, S. O’Riain, and J. C. P. da Silva. Treo: combining entity-search, spreading activation and semantic relatedness for querying linked data. In *Proceedings of the 1st Workshop on Question Answering Over Linked Data (QALD-1), Co-located with the 8th Extended Semantic Web Conference*, pages 24–37, 2011.
- [56] A. Freitas, J. G. Oliveira, S. O’Riain, E. Curry, and J. C. P. Da Silva. Querying Linked Data using semantic relatedness: A vocabulary independent approach. In R. Munoz, A. Montoyo, and E. Metais, editors, *Natural Language Processing and Information Systems, 16th International Conference on Applications of Natural Language to Information Systems, NLDB 2011*, volume 6716 of *Lecture Notes in Computer Science*, pages 40–51, Berlin Heidelberg, Germany, 2011. Springer-Verlag.
- [57] A. Freitas, E. Curry, J. Oliveira, and S. O’Riain. Querying heterogeneous datasets on the Linked Data Web: Challenges, approaches, and trends. *IEEE Internet Computing*, 16(1):24–33, Jan 2012. .
- [58] T. Furche, G. Gottlob, G. Grasso, C. Schallhart, and A. Sellers. OXPath: A language for scalable data extraction, automation, and crawling on the Deep Web. *The VLDB Journal*, 22(1): 47–72, 2013. .
- [59] E. Gabrilovich and S. Markovitch. Computing semantic relatedness using Wikipedia-based explicit semantic analysis. In M. M. Veloso, editor, *IJCAI-07, Proceedings of the 20th International Joint Conference on Artificial Intelligence*, volume 7, pages 1606–1611, Palo Alto, California, USA, 2007. AAAI Press.
- [60] F. Gandon, M. Sabou, H. Sack, C. d’Amato, P. Cudré-Mauroux, and A. Zimmermann, editors. *The Semantic Web. Latest Advances and New Domains*, volume 9088 of *Lecture Notes in Computer Science*, Cham, Switzerland, 2015. Semantic Technology Institute International (STI2), Springer International Publishing.
- [61] M. Gao, J. Liu, N. Zhong, F. Chen, and C. Liu. Semantic mapping from natural language questions to OWL queries. *Computational Intelligence*, 27(2):280–314, 2011.
- [62] D. Gerber and A.-C. N. Ngomo. Extracting multilingual natural-language patterns for RDF predicates. In *Knowledge Engineering and Knowledge Management*, pages 87–96. Springer-Verlag, Berlin Heidelberg, Germany, 2012.
- [63] C. Giannone, V. Bellomaria, and R. Basili. A HMM-based approach to Question Answering against Linked Data. In *Working Notes for CLEF 2013 Conference*, 2013.
- [64] A. M. Gliozzo and A. Kalyanpur. Predicting lexical answer types in open domain QA. *International Journal On Semantic Web and Information Systems*, 8(3):74–88, July 2012. .
- [65] S. Hakimov, H. Tunc, M. Akimaliev, and E. Dogdu. Semantic Question Answering system over Linked Data using relational patterns. In *Proceedings of the Joint EDBT/ICDT 2013 Workshops*, pages 83–88, New York, NY, USA, 2013. Association for Computational Linguistics.
- [66] S. Hakimov, C. Unger, S. Walter, and P. Cimiano. *Natural Language Processing and Information Systems: 20th International Conference on Applications of Natural Language to Information Systems, NLDB 2015, Passau, Germany, June 17-19, 2015, Proceedings*, chapter Applying Semantic Parsing to Question Answering Over Linked Data: Addressing the Lexical Gap, pages 103–109. Springer International Publishing, Cham, Switzerland, 2015.
- [67] T. Hamon, N. Grabar, F. Mouglin, and F. Thiessard. Description of the POMELO system for the task 2 of QALD-4. In *Working Notes for CLEF 2014 Conference*, 2014.
- [68] B. Hamp and H. Feldweg. Germanet—a lexical-semantic net for German. In I. Mani and M. Maybury, editors, *Proceedings of the ACL/EACL ’97 Workshop on Intelligent Scalable Text Summarization*, pages 9–15. Citeseer, 1997.
- [69] Z. S. Harris. *Papers in structural and transformational linguistics*. Springer, 2013.
- [70] S. He, S. Liu, Y. Chen, G. Zhou, K. Liu, and J. Zhao. CASIA@QALD-3: A Question Answering system over Linked Data. In *Working Notes for CLEF 2013 Conference*, 2013.
- [71] S. He, Y. Zhang, K. Liu, and J. Zhao. CASIA@V2: A MLN-based Question Answering system over Linked Data. In *Working Notes for CLEF 2014 Conference*, 2014.
- [72] D. M. Herzig, P. Mika, R. Blanco, and T. Tran. Federated entity search using on-the-fly consolidation. In *The Semantic Web-ISWC 2013*, pages 167–183. Springer-Verlag, Berlin Heidelberg, Germany, 2013.
- [73] L. Hirschman and R. Gaizauskas. Natural language Question Answering: The view from here. *Natural Language Engineering*, 7(4):275–300, 2001.
- [74] K. Höffner and J. Lehmann. Towards question answering on statistical linked data. In *Proceedings of the 10th International Conference on Semantic Systems*, pages 61–64. ACM, 2014.
- [75] K. Höffner and J. Lehmann. Towards Question Answering on statistical Linked Data. In *Proceedings of the 10th International Conference on Semantic Systems, SEM ’14*, pages 61–64, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-2927-9. . URL <http://svn.aksw.org/papers/2014/cubeqa/short/public.pdf>.
- [76] I. Horrocks, P. F. Patel-Schneider, H. Boley, S. Tabet, B. Groszof, and M. Dean. SWRI: A Semantic Web rule language combining OWL and RuleML. W3c member submission, World Wide Web Consortium, 2004. URL <http://www.w3.org/Submission/SWRL>.
- [77] R. Huang and L. Zou. Natural language Question Answering over RDF data. In *Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data*, pages 1289–1290, New York, NY, USA, 2013. Association for Computational Linguistics.

- [78] WWW '11: *Proceedings of the 20th International Conference on World Wide Web*, New York, NY, USA, 2011. International World-Wide Web Conference Committee, ACM. ISBN 978-1-4503-0632-4.
- [79] WWW '12: *Proceedings of the 21st International Conference on World Wide Web*, New York, NY, USA, 2012. International World-Wide Web Conference Committee, ACM. ISBN 978-1-4503-1229-5.
- [80] WWW '13 Companion: *Proceedings of the 22nd International Conference on World Wide Web Companion*, Republic and Canton of Geneva, Switzerland, 2013. International World-Wide Web Conference Committee, International World Wide Web Conferences Steering Committee. ISBN 978-1-4503-2038-2.
- [81] WWW '14: *Proceedings of the 23rd International Conference on World Wide Web*, New York, NY, USA, 2014. International World-Wide Web Conference Committee, ACM. ISBN 978-1-4503-2744-2.
- [82] WWW '15 Companion: *Proceedings of the 24th International Conference on World Wide Web Companion*, Republic and Canton of Geneva, Switzerland, 2015. International World-Wide Web Conference Committee, International World Wide Web Conferences Steering Committee. ISBN 978-1-4503-3473-0.
- [83] P. Jain, P. Hitzler, A. P. Sheth, K. Verma, and P. Z. Yeh. Ontology alignment for Linked Open Data. In *Proceedings of ISWC*, pages 402–417, Berlin, Heidelberg, 2010. Springer-Verlag.
- [84] A. K. Joshi, P. Jain, P. Hitzler, P. Z. Yeh, K. Verma, A. P. Sheth, and M. Damova. Alignment-based querying of Linked Open Data. In *On the Move to Meaningful Internet Systems: OTM 2012*, pages 807–824. Springer-Verlag, Berlin Heidelberg, Germany, 2012.
- [85] A. Kalyanpur, J. W. Murdock, J. Fan, and C. Welty. Leveraging community-built knowledge for type coercion in Question Answering. In *The Semantic Web—ISWC 2011*, pages 144–156. Springer-Verlag, Berlin Heidelberg, Germany, 2011.
- [86] J.-T. Lee, S.-B. Kim, Y.-I. Song, and H.-C. Rim. Bridging lexical gaps between queries and questions on large online q&a collections with compact translation models. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing, EMNLP '08*, pages 410–418, Stroudsburg, PA, USA, 2008. Association for Computational Linguistics. URL <http://dl.acm.org/citation.cfm?id=1613715.1613768>.
- [87] J. Lehmann and L. Bühmann. AutoSPARQL: Let users query your knowledge base. In *The Semantic Web: Research and Applications, 8th European Semantic Web Conference, ESWC 2011*, volume 6643 of *Lecture Notes in Computer Science*, 2011. URL http://jens-lehmann.org/files/2011/autosparql_eswc.pdf.
- [88] J. Lehmann, C. Bizer, G. Kobilarov, S. Auer, C. Becker, R. Cyganiak, and S. Hellmann. DBpedia - a crystallization point for the web of data. *Journal of Web Semantics*, 7(3):154–165, 2009. .
- [89] J. Lehmann, T. Furche, G. Grasso, A.-C. Ngonga Ngomo, C. Schallhart, A. Sellers, C. Unger, L. Bühmann, D. Gerber, K. Höffner, D. Liu, and S. Auer. DEQA: Deep web extraction for Question Answering. In *The Semantic Web—ISWC 2012*, Berlin Heidelberg, Germany, 2012. Springer.
- [90] J. Lehmann, R. Isele, M. Jakob, A. Jentzsch, D. Kontokostas, P. N. Mendes, S. Hellmann, M. Morsey, P. van Kleef, S. Auer, et al. DBpedia—A large-scale, multilingual knowledge base extracted from Wikipedia. *Semantic Web Journal*, 5:1–29, 2014.
- [91] V. Lopez, V. Uren, M. Sabou, and E. Motta. Is Question Answering fit for the Semantic Web?: A survey. *Semantic Web Journal*, 2(2):125–155, April 2011. .
- [92] V. Lopez, C. Unger, P. Cimiano, and E. Motta. Evaluating Question Answering over Linked Data. *Journal of Web Semantics*, 21:3–13, 2013.
- [93] *Natural Language Processing and Information Systems: 19th International Conference on Applications of Natural Language to Information Systems, NLDB 2014, Montpellier, France, June 18–20, 2014. Proceedings*, volume 8455 of *Lecture Notes in Computer Science*, Berlin Heidelberg, Germany, 2014. L'Unité Mixte de Recherche Territoires, Environnement, Télédétection et Information Spatiale (TETIS), Springer.
- [94] E. Marx, R. Usbeck, A.-C. Ngomo Ngonga, K. Höffner, J. Lehmann, and S. Auer. Towards an open Question Answering architecture. In *SEMANTICS 2014*, 2014. URL http://svn.aksw.org/papers/2014/Semantics_openQA/public.pdf.
- [95] D. Melo, I. P. Rodrigues, and V. B. Nogueira. Cooperative Question Answering for the Semantic Web. In L. Rato and T. G. Calves, editors, *Actas das Jornadas de Informática da Universidade de Évora 2011*, pages 1–6. Escola de Ciências e Tecnologia, Universidade de Évora, November, 16 2011.
- [96] P. Mika, T. Tudorache, A. Bernstein, C. Welty, C. Knoblock, D. Vrandečić, P. Groth, N. Noy, K. Janowicz, and C. Goble, editors. *The Semantic Web—ISWC 2014*, Cham, Switzerland, 2014. Springer International Publishing.
- [97] G. A. Miller. WordNet: A lexical database for english. *Communications of the ACM*, 38(11):39–41, 1995.
- [98] G. A. Miller and W. G. Charles. Contextual correlates of semantic similarity. *Language and cognitive processes*, 6(1): 1–28, 1991.
- [99] A. Mishra and S. K. Jain. A survey on Question Answering systems with classification. *Journal of King Saud University-Computer and Information Sciences*, 2015.
- [100] A. M. Moussa and R. F. Abdel-Kader. QASYO: A Question Answering system for YAGO ontology. *International Journal of Database Theory and Application*, 4(2), 2011.
- [101] R. Muñoz, A. Montoyo, and E. Metais, editors. *Natural Language Processing and Information Systems: 16th International Conference on Applications of Natural Language to Information Systems, NLDB 2011, Alicante, Spain, June 28-30, 2011, Proceedings*, volume 6716 of *Lecture Notes in Computer Science*, Berlin Heidelberg, Germany, 2011. Springer.
- [102] N. Nakashole, G. Weikum, and F. Suchanek. PATTY: A taxonomy of relational patterns with semantic types. In *EMNLP-CoNLL 2012, 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, Proceedings of the Conference*, pages 1135–1145, Stroudsburg, PA, USA, 2012. Association for Computational Linguistics.
- [103] J.-P. Ng and M.-Y. Kan. Qanus: An open-source question-answering platform. *arXiv preprint arXiv:1501.00311*, 2015.
- [104] A.-C. Ngonga Ngomo. Link discovery with guaranteed reduction ratio in affine spaces with minkowski measures. In *Proceedings of ISWC*, 2012.
- [105] A.-C. Ngonga Ngomo and S. Auer. LIMES—A time-efficient approach for large-scale link discovery on the web of data. In

- IJCAI-11, Proceedings of the 24th International Joint Conference on Artificial Intelligence*, Palo Alto, California, USA, 2011. AAAI Press.
- [106] A.-C. Ngonga Ngomo, L. Bühmann, C. Unger, J. Lehmann, and D. Gerber. SPARQL2NL—Verbalizing SPARQL queries. In D. Schwabe, V. Almeida, H. Glaser, R. Baeza-Yates, and S. Moon, editors, *Proceedings of the IW3C2 WWW 2013 Conference*, pages 329–332, 2013.
- [107] S. Ou and Z. Zhu. An entailment-based Question Answering system over Semantic Web data. In *Digital Libraries: For Cultural Heritage, Knowledge Dissemination, and Future Creation*, pages 311–320. Springer-Verlag, Berlin Heidelberg, Germany, 2011.
- [108] S. Park, H. Shim, and G. G. Lee. ISOFT at QALD-4: Semantic similarity-based question answering system over Linked Data. In *Working Notes for CLEF 2014 Conference*, 2014.
- [109] S. Park, S. Kwon, B. Kim, S. Han, H. Shim, and G. G. Lee. Question answering system using multiple information source and open type answer merge. In *Proceedings of NAACL-HLT*, pages 111–115, 2015.
- [110] P. F. Patel-Schneider, Y. Pan, P. Hitzler, P. Mika, L. Zhang, J. Z. Pan, I. Horrocks, and B. Glimm, editors. *The Semantic Web—ISWC 2010*, Berlin Heidelberg, Germany, 2010. Springer-Verlag.
- [111] C. Pradel, G. Peyet, O. Haemmerle, and N. Hernandez. SWIP at QALD-3: Results, criticisms and lesson learned. In *Working Notes for CLEF 2013 Conference*, 2013.
- [112] V. Presutti, C. d’Amato, F. Gandon, M. d’Aquin, S. Staab, and A. Tordai, editors. *The Semantic Web: Trends and Challenges*, volume 8465 of *Lecture Notes in Computer Science*, Cham, Switzerland, 2014. Semantic Technology Institute International (STI2), Springer International Publishing.
- [113] QALD 1. *Proceedings of the 1st Workshop on Question Answering Over Linked Data (QALD-1)*, Co-located with the 8th Extended Semantic Web Conference, 2011.
- [114] M.-M. Rahoman and R. Ichise. An automated template selection framework for keyword query over Linked Data. In *Semantic Technology*, pages 175–190. Springer-Verlag, Berlin Heidelberg, Germany, 2013.
- [115] M. Rani, M. K. Mueyba, and O. Vyas. A hybrid approach using ontology similarity and fuzzy logic for semantic question answering. In *Advanced Computing, Networking and Informatics—Volume 1*, pages 601–609. Springer-Verlag, Berlin Heidelberg, Germany, 2014.
- [116] A. Ranta. Grammatical framework. *Journal of Functional Programming*, 14(02):145–189, 2004.
- [117] K. U. Schulz and S. Mihov. Fast string correction with levenshtein automata. *International Journal on Document Analysis and Recognition*, 5(1):67–85, 2002.
- [118] S. Shekarpour, A.-C. Ngonga Ngomo, and S. Auer. Query segmentation and resource disambiguation leveraging background knowledge. In G. Rizzo, P. N. Mendes, E. Charton, S. Hellmann, and A. Kalyanpur, editors, *Proceedings of the Web of Linked Entities Workshop in conjunction with the 11th International Semantic Web Conference (ISWC 2012)*, 2012.
- [119] S. Shekarpour, S. Auer, A.-C. Ngonga Ngomo, D. Gerber, S. Hellmann, and C. Stadler. Generating SPARQL queries using templates. *Web Intelligence and Agent Systems*, 11(3), 2013.
- [120] S. Shekarpour, K. Höffner, J. Lehmann, and S. Auer. Keyword query expansion on Linked Data using linguistic and semantic features. In *ICSC 2013 IEEE Seventh International Conference on Semantic Computing*, Los Alamitos, CA, USA, 2013. IEEE Computer Society.
- [121] S. Shekarpour, A.-C. N. Ngomo, and S. Auer. Question Answering on interlinked data. In D. Schwabe, V. A. F. Almeida, H. Glaser, R. A. Baeza-Yates, and S. B. Moon, editors, *Proceedings of the 22st international conference on World Wide Web*, pages 1145–1156. International World Wide Web Conferences Steering Committee / ACM, 2013. ISBN 978-1-4503-2035-1.
- [122] S. Shekarpour, E. Marx, A.-C. N. Ngomo, and S. Auer. SINA: Semantic interpretation of user queries for Question Answering on interlinked data. *Journal of Web Semantics*, 30:39–51, 2015. URL http://svn.aksw.org/papers/2013/WebSemantic_SINA/public.pdf.
- [123] S. Shekarpour, K. M. Endris, A. Jaya Kumar, D. Lukovnikov, K. Singh, H. Thakkar, and C. Lange. Question answering on linked data: Challenges and future directions. In *Proceedings of the 25th International Conference Companion on World Wide Web*, WWW ’16 Companion, pages 693–698, Republic and Canton of Geneva, Switzerland, 2016. International World Wide Web Conferences Steering Committee. ISBN 978-1-4503-4144-8. . URL <http://dx.doi.org/10.1145/2872518.2890571>.
- [124] Y. Shen, J. Yan, S. Yan, L. Ji, N. Liu, and Z. Chen. Sparse hidden-dynamics conditional random fields for user intent understanding. In *Proceedings of the 20st international conference on World Wide Web*, pages 7–16, New York, NY, USA, 2011. Association for Computational Linguistics. .
- [125] K. Shirai, K. Inui, H. Tanaka, and T. Tokunaga. An empirical study on statistical disambiguation of japanese dependency structures using a lexically sensitive language model. In *Proceedings of Natural Language Pacific-Rim Symposium*, pages 215–220, 1997.
- [126] E. Simperl, P. Cimiano, A. Polleres, O. Corcho, and V. Presutti, editors. *The Semantic Web: Research and applications*, volume 7295 of *Lecture Notes in Computer Science*, Berlin Heidelberg, Germany, 2012. Semantic Technology Institute International (STI2), Springer-Verlag.
- [127] K. Sparck Jones. A statistical interpretation of term specificity and its application in retrieval. *Journal of documentation*, 28(1):11–21, 1972.
- [128] R. Stewart. A demonstration of a natural language query interface to an event-based semantic web triplestore. *The Semantic Web: ESWC 2014 Satellite Events: ESWC 2014 Satellite Events, Anissaras, Crete, Greece, May 25-29, 2014, Revised Selected Papers*, 8798:343, 2014.
- [129] H. Sun, H. Ma, W.-t. Yih, C.-T. Tsai, J. Liu, and M.-W. Chang. Open domain Question Answering via semantic enrichment. In *Proceedings of the 24th International Conference on World Wide Web*, pages 1045–1055. International World Wide Web Conferences Steering Committee, 2015.
- [130] H. Sun, H. Ma, W.-t. Yih, C.-T. Tsai, J. Liu, and M.-W. Chang. Open domain question answering via semantic enrichment. In *Proceedings of the 24th International Conference on World Wide Web*, pages 1045–1055. International World Wide Web Conferences Steering Committee, 2015.
- [131] C. Tao, H. R. Solbrig, D. K. Sharma, W.-Q. Wei, G. K. Savova, and C. G. Chute. Time-oriented Question Answering from clinical narratives using semantic-web techniques. In *The Semantic Web—ISWC 2010*, pages 241–256. Springer-Verlag,

- Berlin Heidelberg, Germany, 2010.
- [132] G. Tsatsaronis, M. Schroeder, G. Paliouras, Y. Almirantis, I. Androustopoulos, E. Gaussier, P. Gallinari, T. Artieres, M. R. Alvers, M. Zschunke, et al. BioASQ: A challenge on large-scale biomedical semantic indexing and Question Answering. In *2012 AAAI Fall Symposium Series*, 2012.
- [133] C. Unger and P. Cimiano. Representing and resolving ambiguities in ontology-based Question Answering. In S. Padó and S. Thater, editors, *Proceedings of the TextInfer 2011 Workshop on Textual Entailment*, pages 40–49, Stroudsburg, PA, USA, 2011. Association for Computational Linguistics.
- [134] C. Unger and P. Cimiano. Pythia: Compositional meaning construction for ontology-based Question Answering on the Semantic Web. In R. Munoz, A. Montoyo, and E. Metais, editors, *16th International Conference on Applications of Natural Language to Information Systems, NLDB 2011*, pages 153–160. Springer-Verlag, Berlin Heidelberg, Germany, 2011.
- [135] C. Unger, L. Bühmann, J. Lehmann, A.-C. Ngonga Ngomo, D. Gerber, and P. Cimiano. Template-based Question Answering over RDF data. In *Proceedings of the 21st international conference on World Wide Web*, pages 639–648, 2012.
- [136] C. Unger, P. Cimiano, V. Lopez, E. Motta, P. Buitelaar, and R. Cyganiak, editors. *Interacting with Linked Data (ILD 2012)*, volume 913 of *CEUR Workshop Proceedings*, 2012. URL <http://ceur-ws.org/Vol-913>.
- [137] C. Unger, C. Forascu, V. Lopez, A.-C. N. Ngomo, E. Cabrio, P. Cimiano, and S. Walter. Question Answering over Linked Data (QALD-4). In *Working Notes for CLEF 2014 Conference*, 2014.
- [138] *Natural Language Processing and Information Systems: 18th International Conference on Applications of Natural Language to Information Systems, NLDB 2013, Salford, UK. Proceedings*, volume 7934 of *Lecture Notes in Computer Science*, Berlin Heidelberg, Germany, 2013. University of Salford, Springer.
- [139] R. Usbeck, A.-C. Ngonga Ngomo, L. Bühmann, and C. Unger. HAWK - hybrid Question Answering over linked data. In *12th Extended Semantic Web Conference, Portoroz, Slovenia, 31st May - 4th June 2015*, 2015. URL http://svn.aksw.org/papers/2015/ESWC_HAWK/public.pdf.
- [140] P. Vossen. *A multilingual database with lexical semantic networks*. Springer-Verlag, Berlin Heidelberg, Germany, 1998.
- [141] S. Walter, C. Unger, P. Cimiano, and D. Bär. Evaluation of a layered approach to Question Answering over Linked Data. In *The Semantic Web—ISWC 2012*, pages 362–374. Springer-Verlag, Berlin Heidelberg, Germany, 2012.
- [142] C. Welty, J. W. Murdock, A. Kalyanpur, and J. Fan. A comparison of hard filters and soft evidence for answer typing in Watson. In *The Semantic Web—ISWC 2012*, pages 243–256. Springer, 2012.
- [143] S. Wolfram. *The Mathematica book*. Cambridge University Press and Wolfram Research, Inc., New York, NY, USA and, 100:61820–7237, 2000.
- [144] K. Xu, Y. Feng, and D. Zhao. Xser@ QALD-4: Answering natural language questions via Phrasal Semantic Parsing. In *Working Notes for CLEF 2014 Conference*, 2014.
- [145] M. Yahya, K. Berberich, S. Elbassuoni, M. Ramanath, V. Tresp, and G. Weikum. Deep answers for naturally asked questions on the Web of Data. In *Proceedings of the 21st international conference on World Wide Web*, pages 445–449, New York, NY, USA, 2012. Association for Computational Linguistics.
- [146] M. Yahya, K. Berberich, S. Elbassuoni, M. Ramanath, V. Tresp, and G. Weikum. Natural language questions for the Web of Data. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, EMNLP-CoNLL '12*, pages 379–390, Stroudsburg, PA, USA, July 2012. Association for Computational Linguistics. URL <http://dl.acm.org/citation.cfm?id=2390948.2390995>.
- [147] M. Yahya, K. Berberich, S. Elbassuoni, and G. Weikum. Robust question answering over the web of Linked Data. In *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management*, pages 1107–1116. ACM, 2013.
- [148] Z. Yang, E. Garduno, Y. Fang, A. Maiberg, C. McCormack, and E. Nyberg. Building optimal information systems automatically: Configuration space exploration for biomedical information systems. In *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management*, pages 1421–1430. ACM, 2013.
- [149] E. M. G. Younis, C. B. Jones, V. Tanasescu, and A. I. Abdelmoty. Hybrid geo-spatial query methods on the Semantic Web with a spatially-enhanced index of DBpedia. In *GIScience*, pages 340–353, 2012.
- [150] N. Zhang, J.-C. Creput, W. Hongjian, C. Meurie, and Y. Ruichek. Query Answering using user feedback and context gathering for Web of Data. In *3rd International Conference on Advanced Communications and Computation, INFOCOMP*, pages p33–38, 2013.
- [151] L. Zou, R. Huang, H. Wang, J. X. Yu, W. He, and D. Zhao. Natural language question answering over RDF: A graph data driven approach. In *Proceedings of the 2014 ACM SIGMOD international conference on Management of data*, pages 313–324. ACM, 2014.