Question answering systems: the story till the Arabic linked data

Abdelghani Bouziane

EEDIS Lab,
Faculty of Sciences,
Dep. of Computer Science,
University Djillali Liabes of Sidi Bel-abbes., Algeria
E-mail: ghani.ab1@gmail.com

Djelloul Bouchiha*

Dept. Mathematics and Computer Science,
Inst. Sciences and Technologies,
Ctr Univ Naama, Algeria
E-mail: bouchiha.dj@gmail.com
*Corresponding author

Noureddine Doumi

Department of Computer Science,
University of Saida, Algeria
Email: noureddine.doumi@univ-saida.dz

Mimoun Malki

LabRI-Sidi Bel-Abbes,
Ecole Supérieure en Informatique de Sidi Bel-abbes, Algeria
Email: m.malki@esi-sba.dz

Abstract: Question answering system (QAS) is essential to satisfy the need to query information available in various formats, including structured data (ontology, databases) or unstructured data (document, web). The QAS provides a correct response to the question asked by a user in natural language. QAS uses natural language processing (NLP) techniques to interface with the system user. In this paper, we survey various QAS such as Natural Language Interfacing to DataBases (NLIDB), ontology-based question answering and question answering systems for unstructured data. We give also statistics and analysis. This can help researchers to choose an appropriate solution to their issues. In case of insufficiency, they can propose new systems for complex queries and adapt or reuse QAS techniques for specific research issues. We give also our point of view on how can QAS deal with Arabic linked data.

Keywords: question answering system; QAS; natural language processing; NLP; information retrieval; SPARQL; Semantic Web, Arabic linked data; Natural Language Interfacing to DataBases; NLIDB.

Reference to this paper should be made as follows: Bouziane, A., Bouchiha, D., Doumi, N. and Malki, M. (2017) 'Question answering systems: the story till the Arabic linked data', *Int. J. Artificial Intelligence and Soft Computing*, Vol. 6, No. 1, pp.24–42.

Biographical notes: Abdelghani Bouziane received his Engineer and Magister degree in Computer Science from the University of Mascara in 2007 and 2011. He is currently a doctoral student at the University of Sidi Bel-Abbes Algeria, and a Lecturer at the University Center of Naama. His current research interests include natural language processing, artificial intelligence, Semantic Web and Arabic linked data.

Djelloul Bouchiha received his Engineering degree in Computer Science from Sidi Bel-Abbes University, Algeria, in 2002, and MSc in Computer Science from Sidi Bel Abbes University, Algeria, in 2005, and PhD in 2011. Between 2005 and 2010, he joined the Department of Computer Science, Saida University, Algeria, as a Lecturer. He became an Assistant Professor since January 2011. Currently, he is an Associate Professor at the University Center of Naama. His research interests include Semantic Web services, web reverse-engineering, ontology engineering, knowledge management and information systems.

Noureddine Doumi is currently a Lecturer at Computer Science Department in Tahar Moulay University of Saida. He received his Magister degree in Computer Science from University of Sidi Bel-Abbes in 2005. He is a member of EEDIS lab in UDL-SBA and an active member as a developer in Unitex/GramLab project in University of Paris-Est Marne-La-Vallée. His research interest includes Arabic NLP, linguistic resources, finite-state machines and machine learning.

Mimoun Malki graduated with Engineer degree in Computer Science from National Institute of Computer Science, Algiers, in 1983. He received his MSc and PhD in Computer Science from the University of Sidi Bel-Abbes, Algeria, in 1992 and 2002, respectively. He was an Associate Professor in the Department of Computer Science at the University of Sidi Bel-Abbes from 2003 to 2010. Currently, he is a Full Professor at Djillali Liabes University of, Sidi Bel-Abbes, Algeria. He has published more than 50 papers in the fields of web technologies, ontology and reverse engineering. He is the Head of the Evolutionary Engineering and Distributed Information Systems Laboratory. Currently, he serves as an editorial board member for the *International Journal of Web Science*. His research interests include databases, information systems interoperability, ontology engineering, web-based information systems, Semantic Web services, web reengineering, enterprise mash up and cloud computing.

This paper is a revised and expanded version of a paper entitled 'Question answering systems: survey and trends' presented at Advanced Wireless, Information, and Communication Technologies (AWICT 2015), Tunisia, 5–7 October 2015.

1 Introduction

The rapid increase in massive information storage and the popularity of using the web allow researchers to store data and make them available to the public. However, tools used to explore these data, like search engines and query languages (SQL, SPARQL, NRQL), should be more intelligent to give a precise answer to the user who ignores both the structure of data and the complicated query language. These difficulties motivate the development of new adapted technology, such as question answering systems (QASs).

In fact, this kind of system allows the user to ask a question in natural language (NL) and return the right answer to his question instead of a set of documents deemed relevant, as for search engines.

However, for QASs aiming texts and web documents, the structure of the required information affects the accuracy of these systems. QAS are most effective to interact with structured knowledge bases.

Due to the importance of QAS, other surveys are available in the literature like Allam and Haggag (2012) and Kalaivani and Duraiswamy (2012). In our survey paper:

- we count and classify QASs and analyse the propositions according to different points of view
- we refresh existing surveys by adding recent works
- motivated by the development of Arabic QAS over linked data, we give a classification based, in particularly, on language and data-structure dimensions
- statistics presented through graphical histograms give clear view to researchers working in this field
- we present the current situation of the Arabic Semantic Web.

The rest of the paper is organised as follows: Section 2 describes some notions related to the discussed issue in the paper. Section 3 cites and classifies QASs. Section 4 provides statistics on the QAS. In Section 5, we present works on the Arabic Semantic Web. Finally, Section 6 concludes our work.

2 Background

2.1 What is a QAS?

A QAS allows for interaction, using NL, between a human user and a machine. QAS is able to provide answers for simple or complex questions. The importance of this research field is well justified, because finding the right answer to a question often leads to use several techniques in different research fields, such as information retrieval, knowledge bases, databases, web of document, ontology, Semantic Web, linked data and natural language processing (NLP).

2.2 Question answering research and subtasks

The arrival of the computer with a boundless capacity to store information in different ways and forms, and artificial intelligence research in the 1960/70s, have given a possibility to build a question answering machine, in simple way, store information and then retrieve it on demand. Exploring the information stored in databases necessitates the use of an interactive programming language, such as SQL, which has less expressivity and translation power. To overcome these difficulties, the NLP is the solution. The first work in NLP has focused on the development of the first automatic translator (very basic), the translation of few simple phrases form Russian to English in 1954. In 1962, the first conference on automatic translation is held at MIT. The development of the NLP research is progressing towards creating the first QAS (Tomek and Sanda, 2008).

The Natural Language Interfacing to DataBases (NLIDB) was the first subtask of QAS. The first QASs were developed, such as BASEBALL (Green et al., 1961), PARRY (Colby, 1971), LUNAR (Woods et al., 1972), and PROTOSYNTHEX (Simmons, 1965).

Since, several issues have been discussed:

1. NL interfaces to databases
2. open domain question answering over text
   - document-based question answering
   - question answering on the web
3. semantic ontology-based question answering.

3 Classifying QAS

Vanessa et al. (2011) give four criteria to classify a QAS, according to interlinked dimensions (cf. Figure 1).

1. the input type (question type): facts, dialogs, etc.
2. the sources from which can derive the answers: structured vs. unstructured data
3. the scope: domain specific vs. domain independent.

How it copes with the traditional intrinsic problems that the search environment imposes in any non-trivial search system.

3.1 Works on NLIDB

User, who does not know the databases structure and does not understand complicated interactive programming languages, can use NLIDB for asking questions and getting answers from these databases.

The research in NLIDB leads to deal with two sub-components (cf. Figure 2):

1. linguistic component
2. database component.

Figure 1 Interlinked dimensions of QASs

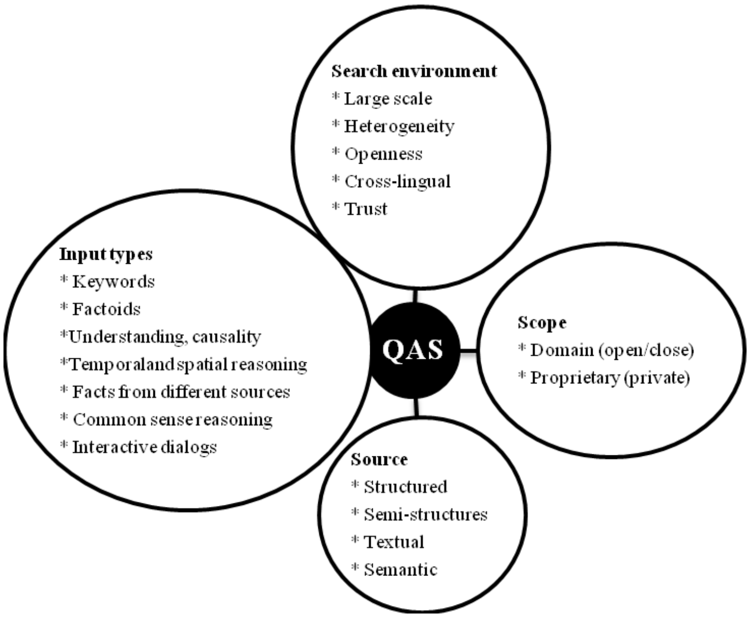
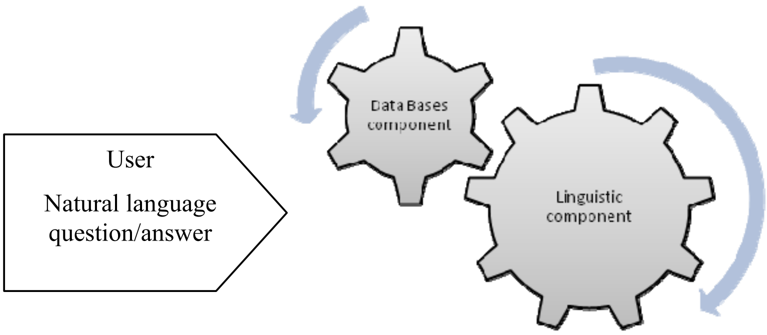


Figure 2 NLIDB components (see online version for colours)



In what follows, we list the most known NLIDB research works sorted by their appearance date.

- *LUNAR* was introduced in 1971. It is a QA Specific domain system which answers questions about samples of rocks brought back from the moon. To accomplish its function, this QAS uses two databases for literature references and chemical analysis (Woods et al., 1972).
- *RENDEZVOUS* is a NL interface where the user query is given through dialogs and special terms to clarify the input of the system (Codd, 1974).
- *PHILIQA* allows semantic comprehension of the question with three layers 'English Formal Language', 'World Model Language' and 'DataBase Language'. It is also known as a Philips QAS (Scha, 1977).

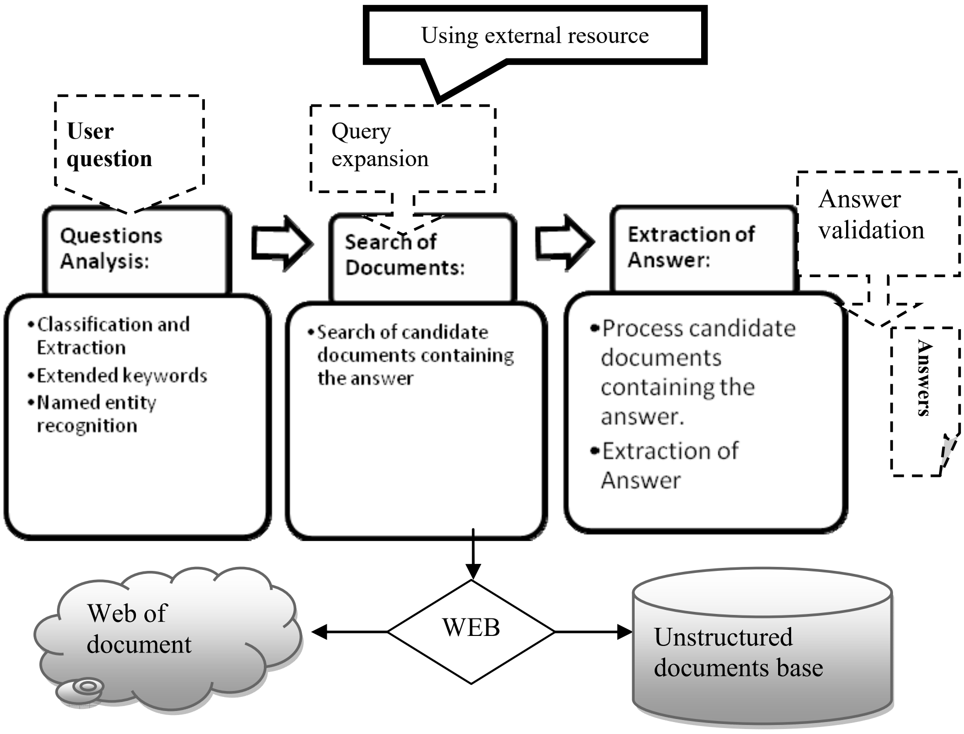
- *CHAT-80* is one of the most famous NLIBD systems in the eighties. CHAT-80 uses a logical concept and interpretation developed with the logical programming language PROLOG. The NL query is transformed into prolog logical expression and into a logical query with the Logical Query Language (LQL) to explore the database of CHAT-80 (Warren and Pereira, 1982).
- *LADDER* uses a distributed database. It is a NL interface in which the user query is parsed using a semantic grammar (Hendrix et al., 1978).
- *JANUS* is an interface with multiple target sources (databases, expert system, graphics devices). The heterogeneities of the system sources are hidden. The user does not need to know the real structure of the target source (Resnik, 1989).
- *MASQUE/SQL* is a portable NL interface, semi-configured for databases queried with SQL language (Androutsopoulos et al., 1993)
- *NALIX* is a NL interface for XML Databases. The main idea of this system is using Meaningful Query Focus (MQF) to find the relationship between the keyword and the XML element, and it is not necessary to map the databases element (Yunyao et al., 2006).
- *PRECISE* is for relational databases. It uses SQL query language. It classifies the user question and translates it to unique semantic form, formulated as SQL query. PRECISE uses a formal notion of semantically tractable questions (Popescu et al., 2003).
- *Natural Language Web Interface for Database (NLWIDB)* system aims to facilitate communicating with the computer in a natural way over the web using:
  1. normalised MySQL database
  2. NLP techniques to translate NL query into SQL query (Rukshan et al., 2013).

3.2 Works on question answering over documents

In information retrieval (IR) and NLP, question qnswering (QA) is the task of automatically providing an answer for a question asked by a human in NL. QA as a task can be divided into three main distinct subtasks (Vanessa et al., 2011), which are question analysis, document retrieval and answer extraction (cf. Figure 3). Most QASs follow these three subtasks. However, they may differ in how they implement every subtask. Further modules can be added in the QA pipeline subtasks, for example query expansion and answer validation.

The NL processing techniques are used for interfacing the QAS at the side of user who asks many kinds of questions. In particular, factoid questions are those asked mainly about named entity (NE), using for example the words: when, where, how much/many, who and what, which ask respectively about date/time, place, person, and organisation. The Second type is the questions that ask about the definition of term or concept. Questions that use the words ‘why’ or ‘how’ are another type that is hard to answer, and there are very little if any attempts done to answer this type of questions.

Figure 3 Architecture of QAS dedicated to the web of documents or text



Most current work on QA has been rekindled largely by the TREC Text Retrieval Conference (sponsored by the American National Institute, NIST, and the Defense Advanced Research Projects Agency, DARPA) and by the cross-lingual QA track at CLEF.

As pointed out by Hirschman and Gaizauskas (2001), QAS for text involve basically two steps: In the first step, the entity to find by the question is semantically defined. In the second, the answer entity is enriched by supplement constraints. The below list of research works, gives a concise overview on the most popular QAS for text in the literature:

- *LASSO* allows four steps:
  1. introducing the question
  2. expecting the answer
  3. identifying the question focus
  4. giving the relevant key-word present in the question and not in the answer (Moldovan et al., 1999).
- In *FALCON* a recognised NE is used for mapping the semantic categories of the answers. After this step, the category of the question is identified and it is mapped into answers taxonomy (Harabagiu et al., 2000).

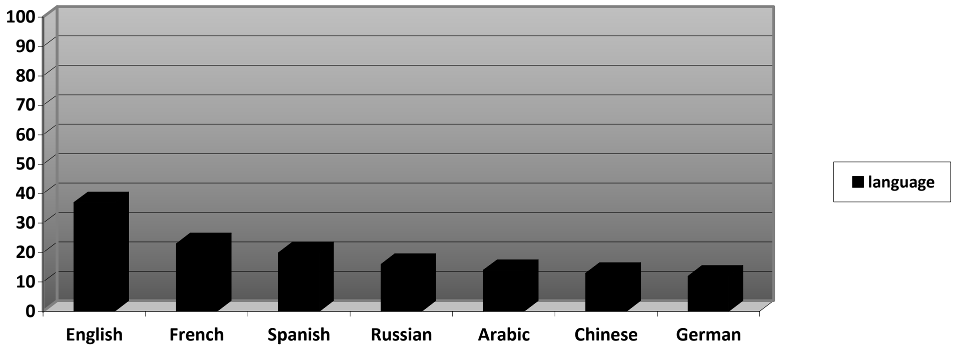
- In *DIMAP* the document is parsed and converted into triples (semantic relation triples). These triples are stored in a structured form, creating a triples database in order to be used to answer the question. The semantic relational triples are extracted using semantic techniques (Litkowski, 2001).
- *Power Answer* (Moldovan et al., 2004) developed at Language Computer Corporation (LCC) looks for answers in large collection of texts by combining syntactic, semantic, lexical and word knowledge information sources. This system consists of three main parts: question processing, document retrieval and answer extraction.
- *PALANTIR* (Harabagiu et al., 2005) QAS was designed with two primary goals: to be a test platform for QA, and to be a dialog-friendly system. PLANTIR QAS uses several extraction techniques applied for:
  1. detection of collocations
  2. recognition of the expected answer type
  3. indexing the document collection based on a very large set of NE classes
  4. answer extraction, based on a set of filters and an extended set of features
  5. answer ranking, based on several strategies.
- *StoQA* uses the NE recognition NLP technique, stop-word lists, and parts-of-speech taggers to extract phrase from the user question. This phrase will be used like input in the search engine working on the documents corpus to find the related document containing the exact answer (Stoyanchev et al., 2008).
- *Mulder* is a QAS for factual question. The user query is extended to multiple queries sent to Google search engine. A linguistic processing using WordNet is done to classify the query, and then a formulation module converts the query to a set of keywords (Kwok et al., 2001).
- *QALC* provides answers to English factoid questions based on syntactic and semantic analysis, using a consistence NLP. The QALC system uses seven modules (NL question analysis, terms extraction, search engine, automatic indexing, NE recognition, question sentences pairing) (Ferret et al., 1999).
- *QRISTAL* uses massively NLP techniques. It is a multi-language system based on the NLP techniques: syntactical technical parsing, semantics disambiguation, conceptual and thematic analysis and named entities recognition (Laurent et al., 2007).
- *WebQA* uses the template-mapping technique to define the question type, and the clustering technique to extract multiple answer blocks (Parthasarathy and Chen, 2007).
- *Ask.com* looks for the user's question in its database and returns a list of questions that it knows how to answer (Ask, 2015).
- Youzheng et al. use question-type-specific method (*QTSM*) that extracts answers from social Q&A pairs. The input question type is compared to social Q&A pairs which have the same type of input question (Youzheng et al., 2015).

3.3 QAS for Arabic language

Arabic language is the official or co-official language of 26 countries in Middle East and North Africa and the most spoken language in the Semitic language group. Spoken by more than 422 million, Arabic is one of the most common languages in the world, and is the religious language of all Muslims of various ethnicities around the world.

According to the world's 10 most influential languages by Weber (1999), Arabic is one of the world's 10 most influential languages. Precisely, the Arabic language is the fifth most influential language in the world as shown in Figure 4.

Figure 4 Most influential languages in the world (see online version for colours)



Arabic is the fourth language used in the web with 168, 1 million users (Internet World Stats, 2016). Thus, it is necessary to develop tools for helping the users to exploit the content of the web in this language.

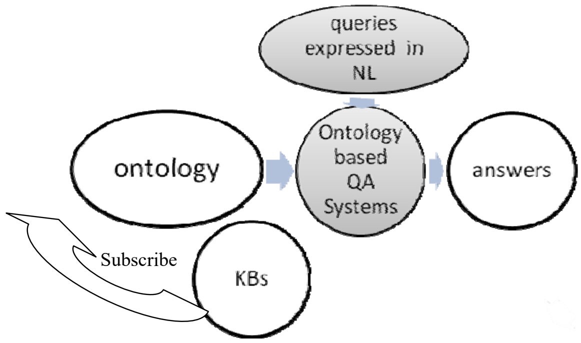
Despite the wide spread of Arabic language, the situation is less bright for Arabic QASs. Although, research in this field has started in Alshalabi (2005) and Black et al. (2006), it is slow progressing and has limited results. Generally, tools and resources are lacking in Arabic. This has reflected negatively on the quality and the number of Arabic QASs. Next are given works on Arabic QAS:

- *AQSA* extracts answers from structured data. It is the first system for the Arabic language. Knowledge from the radiation domain is presented using the frames technique. There is no published evaluation about AQSA (Mohammed et al., 1993).
- *QARAB* is an unconnected (non-web-based) QAS for only factoid question. Any other type of question is supported. It uses IR and NLP techniques to extract answers from a collection of Arabic newspaper texts (Hammo et al., 2002).
- *QASAL* is not web-based (stand-alone) Arabic QAS. It uses Nooj platform as linguistic development environment. The system is only used for factoid questions (Brini et al., 2009).
- *ArQA* is a stand-alone QAS that provides answers to only factoid questions, combining the techniques of IR, NLP and tools for the validation of resulted answers (Badawy et al., 2011).
- *Aquasys* is a stand-alone Arabic QAS, which uses NLP techniques for answering the factoid questions (Bekhti et al., 2011).

3.4 Works on semantic ontology-based QASs

Ontology-based semantic QASs take queries expressed in NL and a given ontology as input, and return answers drawn from one or more KBs. Therefore, they do not require the user to learn the vocabulary or the structure of the ontology (cf. Figure 5).

Figure 5 Ontology-based QASs input/output (see online version for colours)



The user asks question using a NL. The process begins by linguistic analysis (dependency graphs using a syntactical parser with a step of named entities recognition NER). The next step is to classify the question respecting to the defined categories of questions. The SPARQL query is generated according to the two steps (linguistic analysis and question classification), and an external ontology resource is used for matching items generated in the process. Finally, when the SPARQL query is generated, we move to interrogate the linked data and generate the exact answer (cf. Figure 6).

Figure 6 Global architecture of QAS over linked data (see online version for colours)

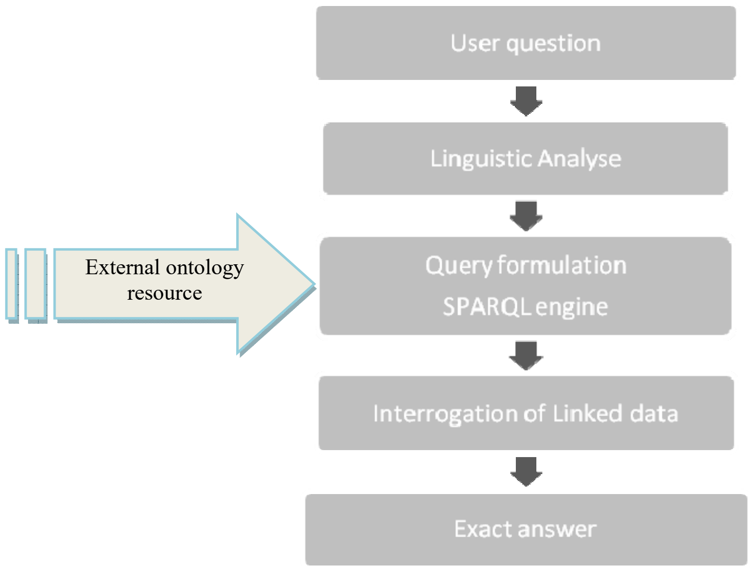


Table 1 summarises features and techniques of works on ontology-based question answering.

Table 1 Works on semantic ontology-based QASs

Work	*Features and techniques*
AquaLog (Lopez et al., 2007)	Allows the user to choose an ontology, and then ask NL queries with respect to the universe of discourse covered by the ontology.
PowerAqua (Lopez et al., 2012)	QAS focusing on querying multiple Semantic Web resources
QACID (Fernandez et al., 2009)	Relies on an ontology, a collection of user queries, and an entailment engine that associates new queries to a cluster of existing queries.
ORAKEL (Cimiano et al., 2007)	translates factual wh-queries into F-logic or SPARQL, and evaluates them with respect to a given KB
E-librarian (Linckels and Meinel, 2005)	Understands the sense of the user query to retrieve multimedia resources from a KB.
GINSENG (Bernstein et al., 2006)	Controls user's input via a fixed vocabulary and predefined sentence structures through menu-based options.
PANTO (Wang et al., 2007)	Portable NLI that takes a NL question as input and executes a corresponding SPARQL query on a given ontology model.
QuestIO (Tablan et al., 2008)	NL queries are translated into formal queries but the system is reliant on the use of gazetteers initialised for the domain ontology.
FREyA (Damljanovic et al., 2010)	Providing improvements with respect to a deeper understanding of a question's semantic meaning.
QAKIS (Cabrio et al., 2012)	NL technique for matching fragments and textual patterns auto collected from Wikipedia.
SPARQL2NL (Axel-Cyrille et al., 2013)	In the side of converting a SPARQL query into NL.
SWIP (Pradel et al., 2014)	The processing of the NL query is based on the use of the pivot query: from the NL user query into a pivot query, and the formalisation of this pivot query.
Pythia (Unger and Cimiano, 2011)	Using ontology in the process of interpretation of the user query.
Yahya et al. (2013)	Extends the user query in variants relaxed query using mapping linguistic structures to ontology-compliant semantic structures.
SQUALL (Ferré, 2014)	Using a controlled NL for translation to SPARQL query.
TBSL (Unger et al., 2012)	The user question is transformed to a template query (a mirror template).
LODQA (Kim and Cohen, 2013)	From the NL query, it generates the SPARQL query using the template model.
RTV (Giannone et al., 2013)	Integrates a lexical semantic modelling and statistical inference with the using of hidden Markov models (HMM) to match ontology triples with the input user query.
DeepQA IBM Watson's system (Kalyanpur et al., 2012)	Using unstructured and structured data (RDF format) to extract and score evidence.

Table 1 Works on semantic ontology-based QASs (continued)

Work	*Features and techniques*
Xser (Kun et al., 2014)	The system operates in two steps: 1 using a semantic parser for the linguistic analysis in order to detect the predicate argument structures 2 the query is instantiated with respect to the structure of knowledge base.
Ganswer (Lei et al., 2014)	The question answering process is graph driven, and consists of two steps: 1 a dependency parsing of the question results a semantic structure of the question 2 the resulting graph is matched with RDF subgraphs triples, a disambiguation is necessary for matching the subgraphs.
CASIA (Shizhu et al., 2014)	A Markov logic networks algorithm is used for learning a joint model, detecting phrases and mapping semantic items. For these phrases, the semantic items are grouped into a graph.
Intui3 (Corina, 2014)	NLP techniques are used: the question is syntactically analysed, chunked, and the named entities are identified. Then each chunk receives one or more interpretation depending on its type and on additional semantic and syntactic information available for that chunk. Using a combination of rules that are attached for each type of interpretation chunk, the question interpretation is mapped to a corresponding SPARQL query.
ISOFT (Seonyeong et al., 2014)	Transforming NL questions into SPARQL queries using a template-based approach. A linguistic analysis of the input question, query templates and slots are determined, searching for appropriate concepts in the knowledge base, based on string similarity and explicit semantic analysis.
Metafrastes (Embregts et al., 2013)	The system retrieves information from Semantic Web knowledge bases, using NL engine, to translate the NL query to SPARQL query.
AR2SPARQL (AlAgha and Abu-Taha, 2015)	Translates the user question expressed in Arabic towards a SPARQL request. This system uses an intermediate representation in the form of RDF triple.

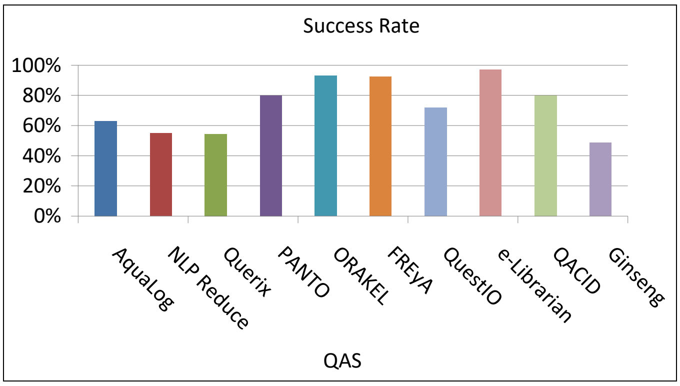
4 QAS performance

In this section we present statistics about two types of QASs: qntology-based QAS and text-based QAS.

4.1 Ontology-based QASs

To show the performance of the ontology-based QASs, we looked at the evaluation results carried out in the literature, notably those summarised in the survey paper (Vanessa et al., 2011). Then we establish the histogram of Figure 7.

Figure 7 Performance results of the ontology-based QASs (see online version for colours)



Performance of the ontology-based QAS is represented by the success rate (correct answers to questions) in the graph above.

We found that the success rate of these QASs varies between 49% and 89%. These results depend on two criteria:

1 the algorithms and methods of NLP

2 the specified domain to be questioned.

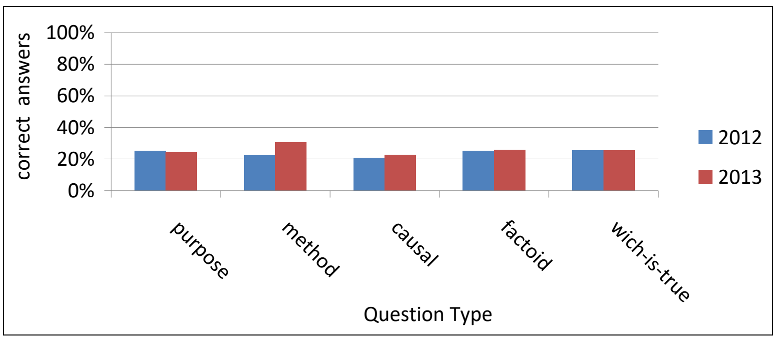
4.2 *Text-based QASs*

To evaluate the text-based QASs, we looked at the results given in the Question Answering for Machine Reading (QA4MRE), the main task at the 2013 Cross Language Evaluation Forum (Sutcliffe et al., 2013).

QA4MRE reads single documents and identifies correct and NoA answers to a set of questions, over the two years 2012 and 2013. NoA means that the system decided not to answer the question.

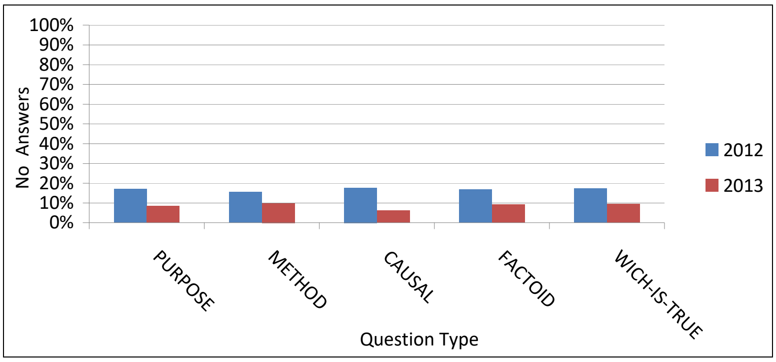
Figure 8 shows the percentage of correct answers for different question types (purpose, method, causal, factoid and which-is-true) in the 2012/2013 versions of the QA4MRE challenge.

Figure 8 Percentage of correct answers according to different question types shown over the years 2012 and 2013 (see online version for colours)



Using the same types of questions, Figure 9 shows the percentage for no answered questions in 2012–2013. A low number of no answered question means that the system is more reliable.

Figure 9 Percentage of NOA answers according to different question types shown over the years 2012 and 2013 (see online version for colours)



We can say that QAS reliability increases in direct proportion to percentage of correct answers, and is in inverse proportion to percentage of NoA answers. These results depend on three criteria: questions type, friability of searching candidate document algorithm and the extraction of the correct answers module.

However, the fact that a QAS does not return an answer does not necessarily mean that it is not able to find an answer as sometimes the answer does not exist in the used corpus. It very much depends on the setup of the experiment.

5 Arabic QAS over linked data

Most of web QAS are dealing with documents. The structure of the required information on the web of documents affects the accuracy of these systems. These systems need to interact with structured and valid knowledge bases.

Currently, mutation of these systems to the web of data seems necessary to find the correct and accurate answers to questions. New query answering systems have to deal with linked data instead of linked documents. Linked data has been discussed under the Semantic Web technology (Berners-Lee, 2006). Linked data provides a publishing paradigm in which not only documents, but also data, can be a first class citizen of the web (Heath and Bizer, 2011).

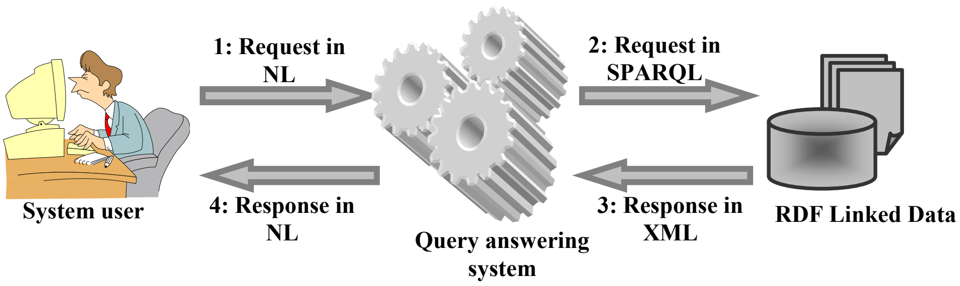
Several works on Arabic Semantic Web model were discussed in Isbaitan and Al-Wahidi (2011) and Khalid (2013). Also, Beseiso et al. (2011) proposed an Arabic language framework for Semantic Web by adding a new layer to web applications in order to get important links between web pages.

The other popular real-world example of a Semantic Web is the DBpedia project. DBpedia project aims to represent Wikipedia content in a structured from, such as RDF triples. DBpedia is considered the main web hub that links different datasets to each other. The availability of DBpedia in local languages allows to develop semantic applications that discover new knowledge from different web resources, and facilitates the making of sophisticated queries. The Arabic Chapter of DBpedia is a new one (Al-deel, 2015). It was effectively published in 11–1–2016 at the Arabian Semantic Web Research Group (ASWRG, 2016).

The challenge now is to implement QAS to explore linked data. The system user can formulate his request with Arabic NL. The system converts then the request into SPARQL request to interrogate Arabic RDF linked data and finally returns the results to the user.

Figure 10 represents the scenario of interacting with a QAS dedicated to Arabic linked data.

Figure 10 Scenario: user, query answering system and linked data (see online version for colours)



Many problems have to be solved. NLP techniques can be used to convert the user request from NL into SPARQL. Then, other APIs can be used to return the results to the user.

6 Conclusions

A QAS aims at giving precise answers to users' questions introduced in NL. The purpose of this paper is to cite and classify many QAS. This can clear the way for researchers in this domain. They can choose the appropriate system to their problem. They can also see the shortcomings and correct them, or propose new QASs.

It is important to note that one of the most important features of QASs is their ability to provide exact answers from different sources. Then, the user asks a question using a NL without knowing the structure of the sources to be queried.

Some languages are better served than others, due to the maturity of research in the countries speaking these languages. So the research in NLP is primordial for developing QASs for unstructured and structured data.

The Arabic Semantic Web resources are relatively missing. This makes hard the development of Arabic QAS. Fortunately, in the last few years, some researches started to develop Arabic RDF linked data in the context of Semantic Web. This will motivate development of QAS to interrogate Arabic RDF linked data.

References

AlAgha, I. and Abu-Taha, A. (2015) 'AR2SPARQL: an Arabic natural language interface for the Semantic Web', *International Journal of Computer Applications* (0975-8887), September, Vol. 125, No. 6, pp.19–27.

Al-feel, H. (2015) 'The roadmap for the Arabic chapter of DBpedia', Mathematical and Computational Methods in Electrical Engineering, *Proceedings of the 14th International Conference on Telecommunications and Informatics*, Sliema, Malta, 17–19 August, pp.115–125.

Allam, A. and Haggag, M. (2012) 'The question answering systems: a survey', *International Journal of Research and Reviews in Information Sciences (IJRRIS)*, September, Vol. 2, No. 3, pp.211–221.

Alshalabi, R. (2005) 'Pattern-based stemmer for finding Arabic roots information', *Technology Journal*, Vol. 4, No. 1, pp.38–43.

Androutsopoulos, I., Ritchie, G. and P. Thanisch. (1993) 'An efficient and portable natural language query interface for relational databases', *Proceedings of the 6th International Conference on Industrial & Engineering Applications of Artificial Intelligence and Expert Systems*, Gordon and Breach Publishers Inc., Langhorne, PA, USA, pp.327–330.

Arabian Semantic Web Research Group (ASWRG) (2016) [online] http://ar.dbpedia.org/index_en.html (accessed 3 May 2016).

Ask (2015) [online] http://www.Ask.com (accessed 29 November 2015).

Axel-Cyrille, N., Bühmann, L. and Unger, C. (2013) 'Sorry, i don't speak SPARQL – translating SPARQL queries into natural language', *International World Wide Web Conference Committee (IW3C2)* WWW 2013, 13–17 May 2013, Rio de Janeiro, Brazil, ACM 978-1-4503-2035-1/13/05.

Badawy, O., Shaheen, M. and Hamadene, A. (2011) 'ARQA high-performance Arabic question answering system', *Proceedings of Arabic Language Technology International Conference*, pp.129–136.

Bekhti, S., Rehman, A., Al-Harbi, M. and Saba, T. (2011) 'AQuASys: an Arabic question answering system based on extensive question analysis and answer relevance scoring', *International Journal of Academic Research*, July, Vol. 3, No. 4, pp.45–54.

Berners-Lee, T. (2006) *Linked Data*, 27 July, W3C Design Issues [online] http://www.w3.org/DesignIssues/LinkedData.html (accessed 3 May 2016).

Bernstein, A., Kauffmann, E., Kaiser, C. and Kiefer, C. (2006) 'Ginseng: a guided input natural language search engine', *Proceeding of the 15th Workshop on Information Technologies and Systems (WITS 2005)*, MV-Wissenschaft, Münster, pp.45–50.

Beseiso, M., Abdul Rahim, A. and Roslan, I. (2011) 'An Arabic language framework for Semantic Web', *2011 International Conference on Semantic Technology and Information Retrieval (STAIR)*, IEEE.

Black, W., Elkateb, S., Rodriguez, H., Alkhalifa, M., Vossen, P., Pease, A. and Fellbaum, C. (2006) 'Introducing the Arabic wordnet project', *Proceedings of 3rd International WordNet Conference (GWC-06).*

Brini, W., Ellouze, M., Mesfar, S. and Belguith, L. (2009) 'An Arabic question-answering system for factoid questions', *Proceedings of IEEE International Conference on Natural Language Processing and Knowledge Engineering.*

Cabrio, E., Cojan, J., Aprosio, A.P., Magnini, B., Lavelli, A. and Gandon, F. (2012) 'QAKiS: an open domain QA system based on relational patterns', *Proceedings of the ISWC 2012 Posters and Demonstrations Track*, CEUR Workshop Proceedings, Vol. 914.

Cimiano, P., Haase, P. and Heizmann, J. (2007) 'Porting natural language interfaces between domains an experimental user study with the ORAKEL system', in Chin, D.N., Zhou, M.X., Lau, T.S. and Puerta, A.R. (Eds.): *Proceeding of the International Conference on Intelligent User Interfaces*, ACM, Gran Canaria, Spain, pp.180–189.

Codd, E.F. (1974) 'Seven steps to rendezvous with the casual user', in Kimbie, J. and Koeman, K. (Eds.): *Data Base Management*, North-Holland Publishers, Amsterdam.

Colby, K.M. (1971) 'Artificial paranoia', *Artificial Intelligence*, Vol. 2, pp.1–25, North-Holland Publishing Company, Amsterdam.

Corina, D. (2014) *Answering Natural Language Questions with Intui3*, CLEF 2014 Working Notes Papers.

Damljanovic, D., Agatonovic, M. and Cunningham, H. (2010) 'Natural language interface to ontologies: combining syntactic analysis and ontology-based lookup through the user interaction', in Aroyo, L., Antoniou, G., Hyvönen, E., ten Teije, A., Stuckenschmidt, H., Cabral, L. and Tudorache, T. (Eds.): *Proceeding of the European Semantic Web Conference*, Heraklion, Greece, Springer Verlag.

Embregts, H., Milea, V. and Frasincar, F. (2013) 'Metafrastes: a news ontology-based information querying using natural language processing', *KMO 2013*, pp.313–324.

Fernandez, O., Izquierdo, R., Ferrandez, S. and Vicedo, J.L. (2009) 'Addressing ontology-based question answering with collections of user queries', *Information Processing and Management*, Vol. 45, No. 2, pp.175–188, Elsevier.

Ferré, S. (2014) 'SQUALL: a controlled natural language as expressive as SPARQL 1.1', *NLDB 2013*, LNCS 7934, pp.114–125, Springer-Verlag, Berlin, Heidelberg.

Ferret, O., Grau, B., Illouz, G., Jacquemin, C. and Masson, N. (1999) 'QUALK the question answering program of the language an cognition group', *LIMSI-CNRS in TREC-8*, Columbia NIST Special Publication, Gaithersburg, Maryland.

Giannone, C., Bellomaria, V. and Basili, R. (2013) 'A HMM-based approach to question answering against linked data', *Proceedings of the Question Answering over Linked Data lab (QALD-3) at CLEF 2013*, LNCS, Springer.

Green, B.F., Wolf, A.K., Chomsky, C. and Laughery, K. (1961) 'BASEBALL: an automatic question answering', *Proceedings Western Joint Computer Conference*, ACM New York, NY, USA, 9–11 May 1961, pp.219–224.

Hammo, B., Abu-Salem, H. and Lytinen, S. (2002) 'QARAB: a question answering system to support the Arabic language', *Proceedings of the Workshop on Computational Approaches to Semitic Languages*, pp.55–65, Philadelphia.

Harabagiu, S., Moldovan, D., Clark, C., Bowden, M., Hickl, A. and Wang, P. (2005) 'Employing two question answering systems in TREC 2005', *Proceedings of the Fourteenth Text Retrieval Conference*.

Harabagiu, S., Moldovan, D., Pasca, M., Mihalcea, R., Surdeanu, M., Bunescu, R., Girju, R., Rus, V. and Morarescu, P. (2000) 'Falcon – boosting knowledge for answer engines', *Proceeding of the 9th Text Retrieval Conference (Trec-9)*, pp.479–488.

Heath, T. and Bizer, C. (2011) 'Linked data: evolving the web into a global data space', *Synthesis Lectures on the Semantic Web: Theory and Technology*, Vol. 1, No. 1, pp.1–136, Morgan and Claypool Publishers.

Hendrix, G., Sacrdoti, E., Sagalowicz, D. and Slocum, J. (1978) 'Developing a natural language interface to complex data', *ACM Transactions on database systems*, Vol. 3, No. 2, pp.105–147.

Hirschman, L. and Gaizauskas, R. (2001) 'Natural language question answering: the view from here', *Natural Language Engineering*, Special Issue on Question Answering, Vol. 7, No. 4, pp.275–300, Cambridge University Press.

Internet World Stats (2016) [online] http://www.internetworldstats.com/stats7.htm (accessed 2 May 2016).

Isbaitan, O. and Al-Wahidi, H. (2011) 'Arabic model for Semantic Web 3.0', *Proceedings of the 2011 International Conference on Intelligent Semantic Web-Services and Applications*, ACM.

Kalaivani, S. and Duraiswamy, K. (2012) 'Comparison of question answering systems based on ontology and Semantic Web in different environment (2012)', *Journal of Computer Science*, Vol. 8, No. 9, pp.1407–1413.

Kalyanpur, A. et al. (2012) 'Structured data and inference in DeepQA', *IBM Journal of Research and Development*, Vol. 56, No. 3, pp.351–364.

Khalid, A. (2013) *An Arabic Semantic Web Model*, thesis Master of Science, M.S., Kent State University.

Kim, J.D. and Cohen, K.B. (2013) 'Natural language query processing for SPARQL generation: a prototype system for SNOMEDCT', *Proceedings of BioLINK SIG.*

Kun, X., Yansong, F. and Dongyan, Z. (2014) *Answering Natural Language Questions via Phrasal Semantic Parsing*, CLEF 2014 Working Notes Papers.

Kwok, C., Etzioni, O. and Weld, D. (2001) 'Scaling question answering to the web', *Proceeding of the 10th International Conference on World Wide Web*, pp.150–161, ACM, Hong Kong, China.

Laurent, D., Séguéla, P. and Nègre, S. (2007) 'Cross lingual question answering using QRISTAL for CLEF 2006', in Clough, P. et al. (Eds.): *Evaluation of Multilingual and Multi-modal Information Retrieval*, pp.339–350, Springer, Berlin, Heidelberg.

Lei, Z., Ruizhe, H., Haixun, W., Jerey, X.Y., Wenqiang, H. and Dongyan, Z. (2014) 'Natural language question answering over RDF – a graph data driven approach', *Proceedings of SIGMOD.*

Linckels, S. and Meinel, C. (2005) 'A simple solution for an intelligent librarian system', *Proceeding of the IADIS International Conference of Applied Computing*, pp.495–503.

Litkowski, K.C. (2001) 'Syntactic clues and lexical resources in question-answering, information technology', *The Ninth Text Retrieval Conference (TREC-9),* NIST Special Publication, pp.500–249.

Lopez, V., Fernández, M., Motta, E. and Stieler, N. (2012) 'Power aqua: supporting users in querying and exploring the Semantic Web', *Semantic Web*, Vol. 3, No. 3, pp.249–265.

Lopez, V., Uren, V., Motta, E. and Pasin, M. (2007) 'Aqualog: an ontology-driven question answering system for organizational semantic intranets', *Journal of Web Semantics: Science Service and Agents on the World Wide Web*, Vol. 5, No. 2, pp.72–105.

Mohammed, F., Nasser, K. and Harb, H. (1993) 'A knowledge-based Arabic question answering system (AQAS)', *Proceedings of ACM SIGART Bulletin*, pp.21–33.

Moldovan, D., Harabagiu, S., Pasca, M., Mihalcea, R., Goodrum, R., Girju, R. and Rus, V. (1999) 'LASSO: a tool for surfing the answer net', *Proceeding of the Text Retrieval Conference (TREC-8)*.

Moldovan, D., Sanda, H., Christine, C. and Mitchell, B. (2004) *Experiments and Analysis of LCC's Two QA Systems Over TREC 2004*, Language Computer Corporation Richardson, Texas, 75080.

Parthasarathy, S. and Chen, J. (2007) 'A web-based question answering system for effective e-learning', *Proceedings of IEEE International Conference on Advanced Learning Technologies,* pp.142–146.

Popescu, A.M., Etzioni, O. and Kautz, H.A. (2003) 'Towards a theory of natural language interfaces to databases', *Proceeding of the International Conference on Intelligent User Interfaces*, ACM Press, pp.149–157.

Pradel, C., Haemmerlé, O. and Hernandez, N. (2014) 'Swip: a natural language to SPARQL interface implemented with SPARQL', *ICCS 2014, LNAI 8577*, pp.260–274, Springer International Publishing, Switzerland.

Resnik, P. (1989) *Access to Multiple Underlying Systems in JANUS*, BBN Report 7142, September, Bolt Beranek and Newman Inc., Cambridge, Massachusetts.

Rukshan, A., Prashanthi, R. and Sinnathamby, M. (2013) 'Natural language web interface for database (NLWIDB)', *Proceedings of the Third International Symposium, SEUSL*, 6–7 July 2013, Oluvil, Sri Lanka.

Scha, R.J.H. (1977) 'Philips question answering system PHILIQA1', *SIGART Newsletter*, February, No. 61, ACM, New York.

Seonyeong, P., Hyosup, S. and Gary Geunbae, L. (2014) *ISOFT at QALD-4 Semantic Similarity-based Question Answering System Over Linked Data*, CLEF 2014 Working Notes Papers.

Shizhu, H., Yuanzhe, Z., Liu, K. and Zhao, J. (2014) *CASIA@V2: A MLN-based Question Answering System Over Linked Data*, CLEF 2014 Working Notes Papers.

Simmons, R.F. (1965) 'Answering English questions by computer: a survey', *Communications of the ACM*, Vol. 8, No. 1, pp.53–70.

Stoyanchev, S., Song, Y. and Lahti, W. (2008) 'Exact phrases in information retrieval for question answering', *Proceedings of the 2nd Workshop on Information Retrieval for Question Answering*, pp.9–16.

Sutcliffe, R., Peñas, A., Hovy, E., Forner, P., Rodrigo, Á., Forascu, C., Benajiba, Y. and Osenova, P. (2013) *Overview of QA4MRE Main Task at CLEF 2013*, Working Notes.

Tablan, V., Damljanovic, D. and Bontcheva, K. (2008) 'A natural language query interface to structured information', in Bechhofer, S., Hauswirth, M., Hoffmann, J. and Koubarakis, M. (Eds.): *Proceeding of the 5th European Semantic Web Conference*, Springer Verlag, Tenerife, Spain, pp.1–15.

Tomek, S. and Sanda, H. (2008) *Advances in Open Domain Question Answering*, Springer, The Netherlands.

Unger, C. and Cimiano, P. (2011) 'Pythia: compositional meaning construction for ontology based question answering on the Semantic Web', in Muñoz, R., Montoyo, A. and Métais, E. (Eds.): *NLDB 2011*, LNCS, Vol. 6716, pp.153–160, Springer, Heidelberg.

Unger, C., Bühmann, L., Lehmann, J., Ngomo, A.C.N., Gerber, D. and Cimiano, P. (2012) 'Template-based question answering over RDF data', *Proceedings of the 21st International Conference on World Wide Web*, ACM, pp.639–648.

Vanessa, L., Victoria, U., Marta, S. and Enrico, M. (2011) *Is Question Answering Fit for the Semantic Web?: A Survey*, Universität, Bielefeld, Germany.

Wang, C., Xiong, M., Zhou, Q. and Yu, Y. (2007) 'PANTO: a portable natural language interface to ontologies', in Franconi, E., Kifer, M. and May, W. (Eds.): *Proceeding of the 4th European Semantic Web Conference*, Springer Verlag, Innsbruck, Austria, pp.473–487.

Warren, D. and Pereira, F. (1982) 'An efficient and easily adaptable system for interpreting natural language queries', *Computational Linguistics*, Vol. 8, Nos. 3–4, pp.110–122.

Weber, G. (1999) 'Top languages: the world's 10 most influential languages', *Language Today*, January, Vol. 24, No. 3, AATF National Bulletin, Cincinnati, Ohio, USA.

Woods, W., Kaplan, R. and Webber, B. (1972) *The Lunar Sciences Natural Language Information System*, Bolt Beranek and Newman Inc., Cambridge, Massachusetts Final Report, B.B.N. Report No 2378.

Yahya, M., Berberich, K., Elbassuoni, S. and Weikum, G. (2013) 'Robust question answering over the web of linked data', *Proceedings of the 22nd ACM International Conference on Information and Knowledge Management*, pp.1107–1116, ACM.

Youzheng, W., Chiori, H., Hideki, K. and Hisashi, K. (2015) 'Leveraging social Q&A collections for improving complex question answering', *Computer Speech and Language*, Vol. 29, No. 1, pp.1–19.

Yunyao, L., Huahai, Y. and Jagadish, H.V. (2006) 'Constructing a generic natural language interface for an XML database', in Loannidis, Y. et al. (Eds.): *EDBT 2006*, LNCS 3896, pp.737–754, Springer-Verlag, Berlin, Heidelberg.