

# A Multiple Ontologies Based System for Answering Natural Language Questions

Anas El-Ansari, Abderrahim Beni-Hssane and Mostafa Saadi

**Abstract** Due to the massive growth of information on the Web, information retrieval systems come to play a more critical role. Most of these systems are based on content matching rather than the meaning, therefore the returned results are not always relevant to the user. To solve this problem, the next generation of information retrieval systems focus on the meaning of the user query and search data using ontologies that provide the vocabulary and structure associated with meta-data. In this work we present a Question Answering system which combines multiple knowledge bases, with a Natural Language parser to transform questions into SPARQL queries or other query language. We demonstrate the feasibility to build such a semantic QA system and the accuracy and relevance of the returned results.

**Keywords** Semantic Web · Ontology · Question answering system · Natural language processing

## 1 Introduction

The Web is a global information space. With A rapidly increasing rate of information available to users through the Web, there is a pressing need for efficient information retrieval systems such as search engines, question answering systems, etc. Nowadays,

---

A. El-Ansari (✉) · A. Beni-Hssane

LAROSERI Laboratory, Computer Science Department Sciences Faculty,  
Chouab Doukkali University, El-Jadida, Morocco  
e-mail: anas.elansari@gmail.com

A. Beni-Hssane

e-mail: abenihssane@yahoo.fr

M. Saadi

Departement Informatique & Telecoms, Ecole Nationale des Sciences  
Appliquées (ENSA), Université Hassan 1er - Settat, Khouribga, Morocco  
e-mail: saadi\_mo@yahoo.fr

© Springer International Publishing AG 2017

Á. Rocha et al. (eds.), *Europe and MENA Cooperation Advances in Information and Communication Technologies*, Advances in Intelligent Systems and Computing 520, DOI 10.1007/978-3-319-46568-5\_18

177

those systems develop very fast and successfully. However they still suffer from a lack of accuracy and the relevance of the provided results is just not up to the mark. To solve this problem, Ontologies and semantic web are becoming a pivotal methodology to represent any domain-specific conceptual knowledge in order to promote the semantic capability of an information retrieval system [1].

The Semantic Web aims to extend the current web standards and technology so that all the Web contents and information can be processed by machines. The use of ontology in the search process provides an interaction between machine and human.

The traditional search is based on term matching techniques, which helps retrieving all resources containing the user's query terms. While in a semantic search, queries that can be expressed in several ways, will be mapped on the semantic level to define topics related to the user informational need that must be retrieved from the web [1].

The question answering (QA) systems [2] aims at providing precise answers to the user's questions. For example, for a question such as (What is the capital of Morocco?), traditional term matching search systems might return a large number of web pages about Morocco and the user would have to dig into these web pages to find the answer. While an efficient QA system would directly answer the question with the name of the capital "Rabat". For that, a QA system needs an efficient natural language question processing mechanism to understand the users question and a semantic data source to get the exact answer for the question.

The QA system we propose in this paper, transforms the user's questions in natural language to a query language (SPARQL or MQL). The last is then used to interrogate different online Knowledge Bases to return an exact answer to the user.

The paper is organized as follows: Section 2 describes the NL question processing. Section 3 presents the semantic data sources we used (the knowledge bases). Section 4 explains the proposed QA system. And finally, the conclusion and directions for future work.

## 2 Related Work

Researches in the field of Question Answering has been advanced in the past couple of years [2]. With the semantic web technologies a domain-specific QA system working on a specific technical domain can make use of the specific domain-dependent ontology to recognize the true meaning included in a natural language text. So we realize that the ontology plays a pivotal role in a technical domain.

One of the typical examples of a QA system is Jeeves [3] which allows users to ask natural language questions and returns a list of matching questions to which it knows the answer. Another example [4], is a research in information processing that has focused on health care consumers. These users often have a frustrated experience while seeking online information.

Other works such as [5, 6], have proved the feasibility of implementing an ontology based question answering system. However, the degree of complicity of these systems

is considerable. And also the relevance of the answers is not optimal. GINSENG [7], a guided-input natural language search engine, and Cuebee [8] progressively guide the scientists by suggesting concepts and relationships that decompose the question into an RDF triple, which is then internally translated into a SPARQL query. This process demands more effort from the user and is a time consuming task.

Most of the studies focusing on ontology based QA systems use a domain local ontology which cannot answer a wide range of questions. In our research we made use of some global knowledge bases offering a huge amount of data to be interrogated and that are available for online access. Those data sources offers a wider range of relevant answers. We also focus on the simplicity of the system, using a simple graphical user interface assisted with autocomplete feature and an error handling component. Our system also offers the user the possibility to interrogate other knowledge base by transforming his question to a SPARQL query that he can copy to the other endpoints.

### 3 Natural Language Question Processing

The Natural Language Processing (NLP) is a research field that explores how computers can be used to understand and process natural language text or even speech to do useful things [9]. Once the user enter his question in a natural language, the system must process it and transform it to the query language. The question can be classified as follows:

- What—objects specification or an activity definition
- Who—object or person specification
- When—date
- Where—geographical location; ...

There are some frameworks that can process natural language question and transform it to a query language such as NLTK [10] and Quepy [11].

In our system we made use of the last one, Quepy, which is a python framework because it can be easily adapted to different question types and query languages. Quepy uses NLTK tagger which is a linguistic tool to analyze natural language questions. It's composed of: a tokenizer, a part-of-speech tagger and a lemmatizer.

So, once the user enters his question, for example: "Who is Bill Gates?" the framework runs NLTK tagger on the string and returns a list of quepy.tagger.Word objects. The transformation from natural language text to the SPARQL query is done by first using a special form of regular expressions:

```
person_name = Group(Plus(Pos("NNP")), "person_name")
regex = Lemma("who") + Lemma("be") + person_name + Question(Pos("."))
```

And then using a convenient way to express semantic relations:

```
person = IsPerson() + HasKeyword(person_name)
definition = DefinitionOf(person)
```

The rest is handled automatically by the framework to finally produce this SPARQL:

```
SELECT DISTINCT ?x2 WHERE {
  ?x0 rdf:type foaf:Person.
  ?x0 rdfs:label "Bill Gates"@en.
  ?x0 rdfs:comment ?x1.
}
```

The SPARQL query is then sent to the knowledge base server which will return the answer. The system offer the user the possibility to choose and search in multiple knowledge bases. The next section will describe those knowledge bases.

## 4 Semantic Data Sources

Semantic knowledge base is a machine-readable resource for the dissemination of information, generally online or with the capacity to be online. A knowledge base is not a static collection of information, but a dynamic resource that may have a learning capacity, as part of an artificial intelligence [12].

Knowledge bases are playing a major role in optimizing the intelligence of Web and search systems and in supporting information integration [12]. Today, most knowledge bases cover only specific domains, created by relatively small groups of knowledge engineers and specialists, and are very cost intensive to keep updated as domains change.

In this system we used multiple knowledge bases in order to increase the chance of getting answers to every user's question. For that, the system interrogates tree large scale publicly available knowledge bases that cover a wide range of domains.

The first one is DBPedia [13]; The English version of this knowledge base describes 4.58 million things, out of which 4.22 million are classified in a consistent ontology, including 251,000 species, 1,445,000 persons, 735,000 places, 411,000 creative works, 241,000 organizations and 6,000 diseases.

Another important knowledge base used in our system is the Freebase [14]. A large collaborative knowledge base composed mainly by its own community members. This online collection of structured data was collected from many sources, including individual contributions, wikis, etc. Freebase aimed to create a global resource that allow people and machines to access and process information more effectively [15]. We also made use of the LOD knowledge base [16]. The

used knowledge bases are connected with other data sets to cover even a wider range of domains. Other knowledge bases can be easily added to the system to cover more domains if needed.

## 5 Proposed Semantic QA System

Our Question Answering System includes three components: question processing based on the Qeepy framework, knowledge base Interrogator and answer processing.

The Question Processing component's job is to analyze and transform a natural language question to a SPARQL query. SPARQL (Protocol and RDF Query Language) [17] is an RDF query language, that is, a semantic query language for knowledge bases, able to retrieve and manipulate data stored in Resource Description Framework (RDF) format.

The knowledge base Interrogator is responsible for sending the query to the SPARQL endpoint of the knowledge base selected by the user. This model returns a set of RDF results. The Answer Processing component handles the returned results to filtrate and transform into a natural language result.

### 5.1 System Architecture

In this section, we present a high level overview of how the whole system works as pictured in Fig. 2. The aim of QA systems is to find exact and correct answers for user's questions. In addition to the graphical user interface (GUI), our QA system contain three main components:

1. Question processing
2. Knowledge base Interrogator
3. Answer processing

Using the GUI the user enters a question in natural language and hits the search button. The Question Processing component analyses the user's question to extract the keywords to be used in the query and prepares an adapted SPARQL and MQL query. The user then is asked to choose the knowledge base to interrogate. According to the user's choice, the Question Processing component will send either the SPARQL query (For DBpedia and Linked Open Data) or the MQL query (For Freebase

The knowledge base Interrogator uses the SPARQL/MQL query to interrogate the selected knowledge base via the endpoint and then returns an RDF answer.

The Answer processing component gets the RDF answer and process it to extract the exact answer and any related information possible and transform it into a natural language answer. The answer is finally returned to the user via the GUI (Fig. 1).

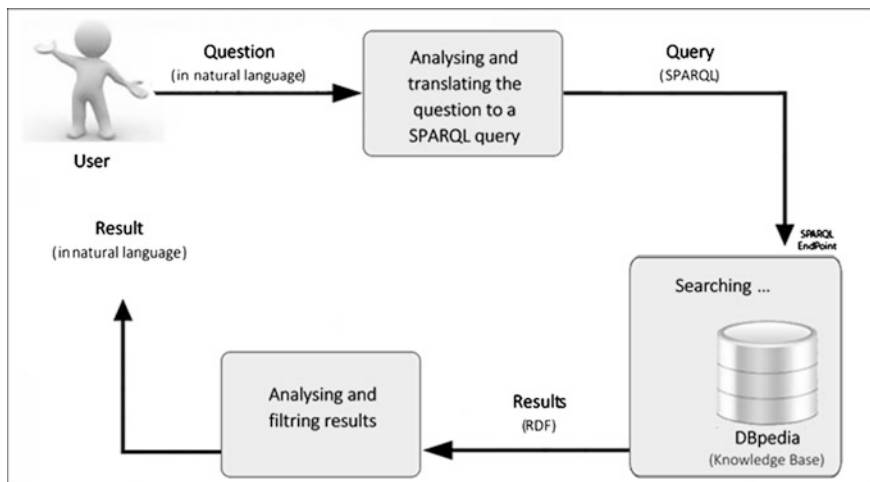


Fig. 1 Design Architecture of the QA system

## 5.2 Simulation

By accessing the interface of the application, you will find, like all search engines, a text field where you type your question (in English) and you search with the little orange button. You can also enjoy the “autocomplete” feature that helps you type the question (Fig. 2).

Once you click the search button, three other buttons appear asking you to choose which knowledge base you want to interrogate: Freebase—DBpedia—Linked Open Data (LOD) (Fig. 3).

The answer to a question is not necessarily the same in all three knowledge bases it is also possible that you find the right answer in a single database or two. This is the reason why we used the three in this system instead of single one.

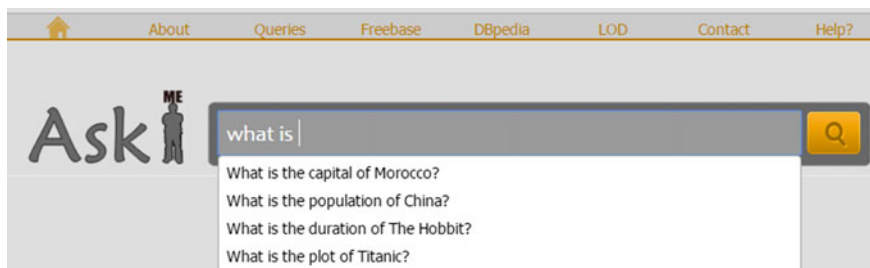
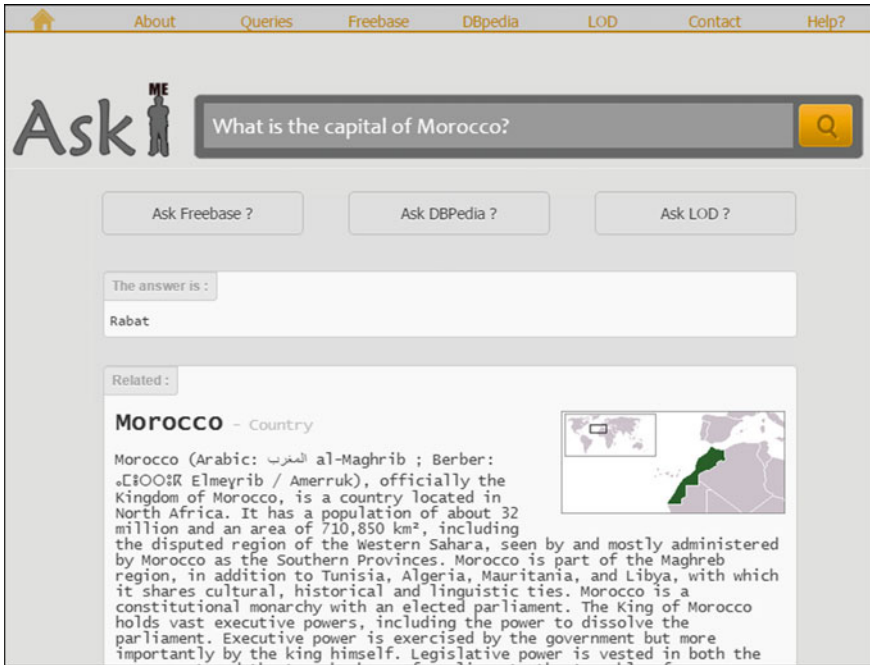


Fig. 2 User interface of the QA system (1) while typing the question



**Fig. 3** User interface of the QA system (2) after hitting the search button and choosing the knowledge base to interrogate

Sometimes the search engine cannot transform a question in natural language to a SPARQL or MQL query. For one reason or another, either because the question contains errors, a problem of capitalized names, or it was not well formulated.

### 5.3 Evaluation

After the implementation phase, we conducted some initial experiments on two versions of the system with the help of ten volunteers. The first version was based on a single ontology (DBpedia), while the second was based on the tree knowledge bases combined (DBpedia, Freebase and LOD). To evaluate the precision and relevance of the returned results we use the precision and recall method [1]. For that, we asked ten subjects ( $S_i$ ) to use the system and ask different questions. Once the experience was done we saved the users feedback in the following table (Table 1).

Table 1 Experimentation results

Subject	User's feedback (1st version)				User's feedback (2nd version)				Precision	Recall			
	Total # of returned	Total # of no-answers	Total # of relevant returned answers	Total # of relevant given by the user	Total # of returned answers	Total # of no-answers	Total # of relevant returned answers	Total # of relevant given by the user					
S1	12	0	10	12	0.83333	0.83333	0.83333	16	0	15	16	0.93750	0.93750
S2	10	1	10	11	1.00000	0.90909	0.90909	12	0	12	12	1.00000	1.00000
S3	5	2	4	7	0.80000	0.57143	0.57143	8	1	8	9	1.00000	0.88889
S4	13	0	11	13	0.84615	0.84615	0.84615	12	1	12	13	1.00000	0.92308
S5	8	1	8	9	1.00000	0.88889	0.88889	11	0	10	10	0.90909	1.00000
S6	12	0	12	13	1.00000	0.92308	0.92308	14	0	12	13	0.85714	0.92308
S7	11	0	9	11	0.81818	0.81818	0.81818	15	0	15	15	1.00000	1.00000
S8	6	0	5	6	0.83333	0.83333	0.83333	10	0	9	10	0.90000	0.90000
S9	15	2	14	17	0.93333	0.82353	0.82353	18	1	18	19	1.00000	0.94737
S10	15	1	13	16	0.86667	0.81250	0.81250	18	0	17	18	0.94444	0.94444



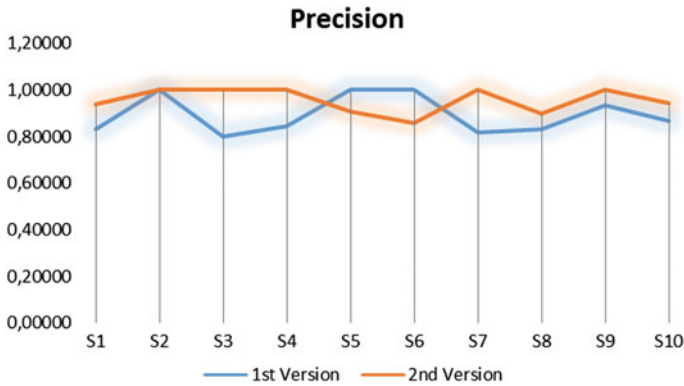


Fig. 4 Precision—calculated for both versions based on each user’s feedback

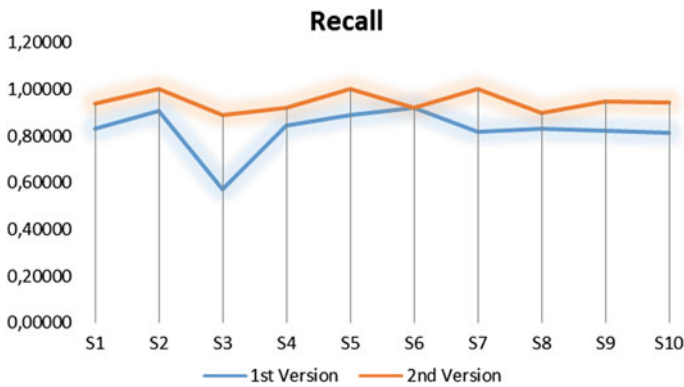


Fig. 5 Recall—calculated for both versions based on each user’s feedback

The goal of our ontology based semantic search system is to maximize precision and accuracy of the results by combining different knowledge bases. And we can see from the experimentation results in the following figures that the precision and recall values increases in the second version where we combined different knowledge bases (Figs. 4 and 5).

The precision graph shows how useful the search results are, while the recall graph shows how complete the results are. The experimentation results shows that when combining multiple ontologies, the system relevance rate increases, however an important number of questions didn’t have answers in the used knowledge bases so there is still room for improvements.

## 6 Conclusion and Future Work

The initial evaluation result shows the feasibility and benefits of building a semantic QA system based on Ontology. And other experiments in related works do prove that it is feasible to use the Ontology-based method to develop Question Answering Systems. Comparing our system with other Question Answering Systems in the Related Work section. Our system offers an easy to use user interface and offers the user to enter questions in natural language. Also combines multiple ontologies to increase the relevance and range of answers so our system can answer a wider range of user's question in multiple domains. Another advantage of this system is that it returns exact answers to the user's questions.

We have implemented a natural language question answering system based on multiple ontologies however there are still many features to implement, on the realized system, such as error handling, or a result relevance rating component.

## References

1. Sudha Ramkumar, A., Poorna, B.: Ontology based semantic search: an introduction and a survey of current approaches. In: ICICA, 2014, 2014 International Conference on Intelligent Computing Applications (ICICA), 2014 International Conference on Intelligent Computing Applications (ICICA), pp. 372–376 (2014). doi:[10.1109/ICICA.2014.82](https://doi.org/10.1109/ICICA.2014.82)
2. Voorhees, E.: The TREC question answering track. *Nat. Lang. Eng.* **7**(4), 361–378 (2006)
3. Askjeeves. <http://askjeeves.com/> (2000)
4. Laura, A., Dagobert, S., Thomas, C.: Semantic representation of consumer questions and physician answers. *Int. J. Med. Inf.* **22**(10), 513–529 (2006)
5. Liu, L., Qi, Q., Li, F.: Ontology-based interactive question and answering system. In: 2010 International Conference on Internet Technology and Applications, Wuhan, pp. 1–4 (2010). doi:[10.1109/ITAPP.2010.5566132](https://doi.org/10.1109/ITAPP.2010.5566132)
6. Guo, Q., Zhang, M.: Question answering based on pervasive agent ontology and Semantic Web. *J. Knowl.-Based Syst.* **22**(6), 443 (2009). doi:[10.1016/j.knosys.2009.06.003](https://doi.org/10.1016/j.knosys.2009.06.003)
7. Bernstein, A., Kaufmann, E., Kaiser, C.: Querying the semantic web with Ginseng: a guided input natural language search engine. In: 15th Workshop on Information Technologies and Systems. Las Vegas, NV: SSRN (2005)
8. Mendes, P.N., McKnight, B., Sheth, A.P., Kissinger, J.C.: TcruziKB: Enabling complex queries for genomic data exploration. In: 2008 IEEE International Conference on Semantic Computing, pp. 432–439. IEEE (2008)
9. Chowdhury, G.G.: Natural language processing. *Ann. Rev. Inf. Sci. Technol.* **37**, 51–89 (2003). doi:[10.1002/aris.1440370103](https://doi.org/10.1002/aris.1440370103)
10. NLTK Project. <http://www.nltk.org/> (2015)
11. Quepy Framework. <https://quepy.readthedocs.org/en/latest/> (2012)
12. Swartout, B., et al.: Toward distributed use of large-scale ontologies. In: Proceedings of the Tenth Workshop on Knowledge Acquisition for Knowledge-Based Systems (1996)
13. DBpedia Project. <http://dbpedia.org/about>
14. MetaWeb, FreeBase. <https://www.freebase.com/>
15. WikiPedia: Freebase. <https://en.wikipedia.org/wiki/Freebase>
16. LOD diagram. <http://richard.cyganiak.de/2007/10/lof>
17. WikiPedia: SPARQL. <https://en.wikipedia.org/wiki/SPARQL>