Towards Logical Inference for Arabic Question-Answering

Wided Bakari,[1] Patrice Bellot,[2] Omar Trigui,[1] Mahmoud Neji[1]

[1] Faculty of Economics and Management, 3018, Sfax Tunisia
[2] Aix-Marseille University, F-13397, Marseille Cedex 20

{wided.bakkari, omar.trigui, Mahmoud.Neji}@fsegs.rnu.tn
patrice.bellot@gmail.com

Abstract. This article constitutes an opening to think of the modeling and the analysis of Arabic texts within a question-answering system. It is a question of exceeding the traditional investigations focused on morpho-syntactic approaches. We present a new approach that analyzes a text, transforms it to logical predicates and extracts the accurate answer. In addition, we represent different levels of information within a text and choose an answer among several proposed. To do so, we transform the question and the text into logical forms. Then, recognize all entailments between them. So, the results of this recognizing are a set of text sentences that can implicate the user's question. Now, our work is concentrated on an implementation step to develop a question-answering system in Arabic using the techniques of textual entailment recognition. Text features extraction (keywords, named entities, relationships that link them) is actually considered the first step in our text modeling process. The second one is the use of textual entailment techniques that relies on inference and logic representation to extract the candidate answer. The last step is the extraction and selection of this answer.

Keywords: NLP, Arabic language, question-answering, recognition text entailment and logic forms.

1 Introduction

Today, we usually weigh multiple questions and need accurate information to answer this question. With the large volume of information found on the internet and the increasing of user's demands, retrieving accurate information is considered a tedious task although the researches in this area do not cease to increase. So, the question-answer process has become one of the most important researches to deal with trustworthy information. It is considered as the heart of the concerns in information retrieval (Sitbon et al., 2006). Generally, the common architecture of a question-answering system (Q-AS[1]) developed is practically a process of three phases (see Figure 1) (Athenikos and Han, 2010), including (question process, document process and answer process).

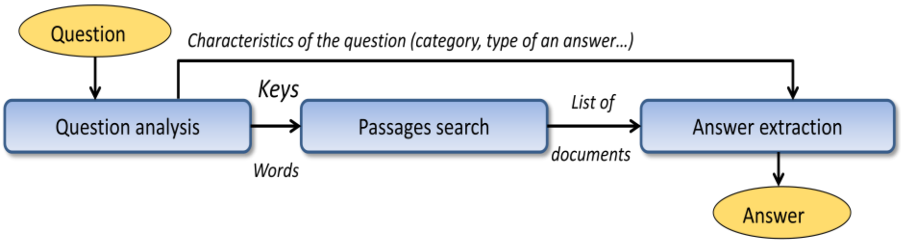


Fig. 1. Architecture of a Q-AS

The potential of question-answering (Wren, 2011) is implemented with the latest success of IBM Watson2 on Jeopardy3. Indeed, Watson analyzes the questions to get what is asked. In less than three seconds, it launched 200 million pages of natural language that contain his memory to find the correct answer and provides evidence for the accuracy of the answer. Apart from the television games, the question-answering technology provided such performance medicine, meteorology, travel, etc.

This article is structured on four sections: Section 2 introduces past works in question-answering systems. Section 3 details our proposed approach for the Arabic and its different steps. Finally, section 4 presents conclusions and potential future directions.

2 Related Works in Question-Answering Systems

Question-answering is a multidisciplinary field where a question-answering system often incorporates techniques and eventual resources, including Information Retrieval IR, Natural Language Processing (NLP), Information Extraction (IE), Machine Learning (ML), etc (Lee et al., 2005). The researches in this area were beginning since 1960. There are many approaches and methodologies that are proposed to many World languages (Bulgarian, Dutch, English, Finnish, French, German, Indonesian, Italian, Japanese, Portuguese, Arabic and Spanish). The question-answering was often occured in conferences and workshops, such as, TREC, CLEF, NTCIR, MUC, etc.

2.1 Question-Answering in English Language

In Natural Language Processing and Information Retrieval, English is the most studied in terms of resources, corpus and systems (Ligozat, 2013). The first question-answering systems focus on English questions came out in 2000. These systems deal with a deep analysis of questions on a restricted or open domain.

[1] Question-Answering System
[2] Watson: http://www-03.ibm.com/innovation/us/watson/.
[3] http://fr.wikipedia.org/wiki/Jeopardy!

QALC (Question Answering program of the Language and Cognition group) (Ferret et al., 2000), (De Chalendar et al., 2002), was developed in the LIMSI laboratory in TREC 1999. This system enables to extract among several candidate sentences the 10 first that provide a suitable and accurate answer. This case is suitable if the selected sentences have different weights. Else, the accurate answer is founded in one of the last sentences as well as in the first ones. Recently, QALC presents an adaptation of an existing question answering system for a machine reading task. Grau and his colleagues proposed an approach for selecting correct answers relies on textual entailment recognition between hypotheses and texts (Grau et al., 2012).

Voorhees provided a question-answering system for factoid questions (Voorhees, 2004). It is the best system in TREC 2004 with a precision of 77% (e.g How many calories are there in a Big Mac?) and a score of 62.2% to answer the question list (e.g list the names of chewing gum).

Somewhere else, AskHERMES[4] (Ask Help clinicians to Extract and aRticulate Multimedia information for answering clinical quEstionS) (Cao et al., 2011) is an online system allows the doctors to ask complex medical questions in consultation and quickly identify the specific answers. This system analyses large volumes of documents to produce short texts as answers. The authors evaluated their system via 4654 medical questions collected in practice. AskHERMES achieved the following two scores 76.0% and 58.0% to classify general themes and extract terms.

Qakis [5] (Question Answering wiKiframework-based System) is a question-answering system in an open field. It generates SPARQL queries from questions, submits them to DBpedia[6] and compares the question to the pattern base to identify the relationship of DBpedia to identify the answer. Also, it provides an integration of multilingual chapters to a DBpedia query with natural language queries. Moreover, QAKiS allows the users to submit a query to a RDF triple store in English and get the answer in the same language. Its architecture is composed of four components (query generation, named entity recognition, pattern-matching and SPARQL package) (Cabrio et al., 2012).

Recently, Tahri and Tibermacine proposed SELNI system which is based on the DBpedia Infobox ontology (Tahri and Tibermacine, 2013). It is focused on the SVM algorithm to train the factoid questions and used SPARQL queries to interrogate the DBpedia server in order to extract the accurate answer. SELNI affected a precision of 86% based on TREC 10 test set.

Finally, the system proposed by (Bhaskar et al., 2012) participates in the main task of QA4MRE@CLEF 2012 and QA4MRE@CLEF 2013. The authors combine the question and each answer option to form the Hypothesis (H) and removed stop words from each hypothesis. This task carried out the associated document based on the TF-IDF of the matching query words along with n-gram overlap of the sentence with hypothesis H. Each sentence allows defining the Text T and the pair (T, H) allocates a ranking score based on textual entailment; each sentence is assigned an inference score with respect to each answer pattern. Based on those ranking a weighing validation is automatically assigned to each answer. The chosen answer is

[4] http://www.askhermes.org/
[5] http://qakis.org/qakis/
[6] http://dbpedia.org

the one that receives the highest score among the list of option answers. This system has a precision of 0.53 and c@1 of 0.65.

2.2 Question-Answering in Chinese and Japanese Languages

Chinese is the second most popular language in question-answering; it was performed for the first time in 2005 at NTCIR (Lee et al., 2005).

Marsha, question-answering system, relies on the same techniques used in English systems developed from TREC (Li and Croft, 2001). Marsha focused on a method based on the TREC question-answering. It has the same performance as some English question-answering systems in the TREC 8 track. Li and Croft have used 51 queries from which 51 they selected 26 queries out of 240 questions collected from Chinese students. The rest of queries are specified to reformulate a question or ask other slightly different. Marsha contains specific techniques dealing with the Chinese characteristics (word segmentation, ordinals processing). It is composed of three modules, such as, query processing, Hanquery search engine and answer extraction.

Some researchers are interested in both English and Chinese questions. They investigate an English-Chinese and Chinese-English translation using Machine Translation systems. This investigations participate in the CLQA (Cross Language Question Answering) task at NTCIR (Kwok et al., 2005). However, ASQA (Academia Sinica Question Answering) deals with Chinese to Chinese factoid question-answering (Lee et al., 2005). It is based on hybrid architecture evaluated by six types of factoid questions: personal names, location names, organization names, artifacts, times, and numbers. ASQA combines the Machine Learning and knowledge-based approaches. It achieved respectively 37.5% and 44.5% Top1 precision for correct and unsupported answers.

Apart from Chinese and English, in NTCIR-6, Mitamura and his associates participated in four CLQA subtasks (J-J, E-J, C-C, and E-C) and introduced the Japanese language (Mitamura et al., 2007). They developped their JAVELIN system where the answers of English questions are extracted from Chinese and Japanese documents. JAVELIN III was an extension of a their previous system JAVELIN II that was initially designed to monolingual English question answering (Nyberg et al., 2005). JAVELIN III is specified by a modular, extensible and a language-independent architecture. The best run in answer precision obtained 13% for E-J and 19% for E-C.

Recently, NTCIR-9 introduced for the first time the new task RITE@NTCIR (Recognizing Inference in TExt). RITE is a generic task that handles a major understanding of the text in various research areas, such as, Information Retrieval, Question-Answering (Harabagiu and Andrew, 2006), Text Summarization, Opinion Analysis, etc. It is proposed to ensure advancements in textual inference research (Shima et al., 2011). According to these authors, RITE4QA subtask is inspired by a series of Answer Validation tasks at CLEF (Peñas et al., 2007).

2.3 Question-Answering in Arabic Language

The first Arabic question-answering system, called AQAS (Mohammed et al., 1993), has emerged in the 90s. AQAS introduces an approach based on a Human-Computer Interface. It is based on querying databases to extract the answer; transforms the question into a query to retrieve the answer and seeks answers from structured data bases that focused on a knowledge model. However, QARAB (Hammo et al., 2004) seeks answers from unstructured documents extracted from the newspaper Al-RAYA.

The majority of Arabic studies in question-answering focus on NLP techniques to extract the accurate answer, they deal with morpho-syntactic approaches. Few of these studies provide logic and inference based-approaches. In addition, ArabiQA (Benajiba et al., 2007) uses named entities techniques; QASAL (Brini et al., 2009) uses NooJ platform to identify answers from an education book. AQUASYS (Bekhti et al., 2011) uses the NLP techniques to analyze question and extract answer from an Arabic corpus and JAWEB (Kurdi et al., 2014) was constructed on the basis of AQUASYS by providing a user interface as an extension. JAWEB was exposed a formal model for a lightweight semantic-based open domain yes/no Arabic question-answering system, it is based on a paragraph retrieval (with variable length). The study of (N Bdour and Gharaibeh, 2013) proposed a constrained semantic representation using an explicit unification framework based on semantic similarities and query expansion (synonyms and antonyms). The system of Trigui and his colleges (Trigui et al., 2010), called DefArabiQA, is based on linguistic patterns, it is considered the first system provides answers to definition questions. It is based on a set of lexical patterns and uses heuristic rules to filter the candidate definitions.

Recently, Abouenour and his associates (Abouenour et al., 2012) proposed IDRAAQ system, it is based on Arabic WordNet and semantic expansion. Without using the database collections of CLEF, IDRAAQ achieved a precision 0.31 and a c@1 0, 21. ALQASIM proposed by (Ezzeldin et al., 2013) is based on selection and validation of the answer, it answers multiple choice questions. This system analysed reading test comprehension instead of questions. It achieved a performance of 0.31 precision and 0.36 c@1 without using any database collection tests.

Table 1. Comparison of arabic question-answering systems

Criteria / System	Data source	Answer	Performance	Question analysis / Text analysis	Question type	Features	Technique and tools used
AQAS (mohamed et al.,1993)	Structured data base	A sentence	Not Mentioned	No / No	several forms (declarative statements)	Stop words Removal, tokenization	Not Mentionned
QARAB (Hammo et al., 2004)	Unstructured data base	A sentence	Precision : 97.3% Recall 97.3 %	Yes / No	Question started by : من،متى، أين، كم	Type and category of expected answer.	Information retrieval IR Natural language processing NLP

ArabiQA (Benajiba et al.,2007)	Corpus	A sentence	Not Mentionned	Yes / No	Factual question	Keyword and named entities recognized.	Not Mentionned
QASAL (Brini et al.,2009)	Book of education	A sentence	Not Mentionned	Yes / No	Factual question	Type of expected answer, focus, keywords.	NOOJ platform
DefArabiQA (Trigui et al.,2010)	Web	A sentence	MRR: 0.81	Yes / No	Definition question	Question topic, type of expected answer.	Not Mentionned
AquASys (Bekhti & Al-Harbi, 2011)	Corpus	Short answer	Precision: 66.25% Recall : 97.5% F1-Score: 87.89%	Yes / No	Factual question	Keywords ; type of expected answer	Not Mentionned
IDRAAQ (Abouenour et al.,2012)	Not Mentionned	Not Mentionned	Not Mentionned	Yes / Yes	QMC	Keywords ; type of expected answer	Not Mentionned
ALQASIM (Ezzeldin et al., 2013)	Not Mentionned	A sentence	Precision: 0.31 C@1: 0.36	Yes / Yes	QMC	Not Mentionned	Not Mentionned
System of (NBdour and Gharaibeh, 2013)	Not Mentionned	A paragraph	Not Mentionned	Yes / No	Yes/No question	Stop words	IR, NLP, artificiel intelligence (AI),
JAWEB (Kurdi et al., 2014)	Corpus	A sentence	Recall: 15-20%	Yes / No	Factual question	Answer type ; Keywords	NLP

In Table 1, we present a summary of Arabic studies; the columns represent our evaluation criteria and the rows are the works studied. We conclude that the most of these studies ensure analysis of the question but do not manipulate the information associated in the text.

3 Approach

We propose to deal with the automatic text comprehension. The previous studies are based on a superficial analysis of the texts. The originality of our approach lies in the in-depth analysis of Arabic texts for the generation of textual inferences. Our method consists of five main steps, namely, analysis of the text and the question; transformation of the Arabic statements (text + question) into logical forms; the recognition of implications generated between these forms and the selection of desired answer. In the following subsections we present each one of these tasks in detail.

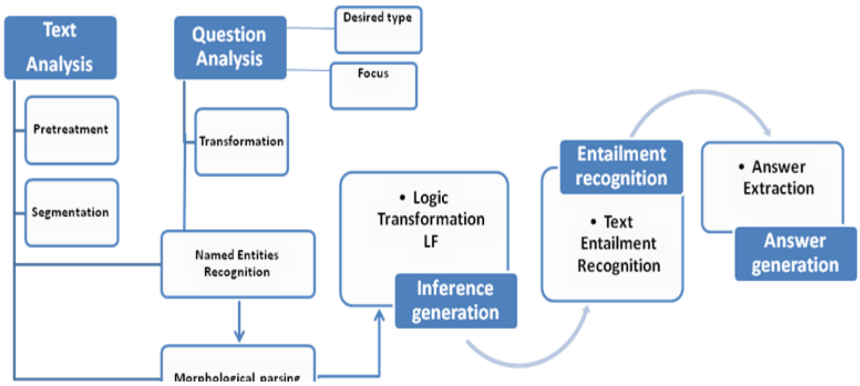


Fig. 2. Proposed approach

3.1 Preprocessing and Analysis

Our approach begins by an analysis step, it accepts as input an Arabic text in html format and generates an annotated and analyzed text. First, a preprocessing step is performed to clean the html text retrieved from the website "Euronews" to produce an Arabic text in txt format. Then, a segmentation step determines the division of text into tokens (sentences, words). A text analysis without segmentation lead to unreliable results (Ghassan, 2001). Afterward, the recognition of named entities is performed by ArNER (Zribi et al., 2010) to determine the set of named entities. This task is so harder in Arabic than other languages due to the lack of capital letters (Shaalan, 2014). Finally, we use Al-khalil parser to carry out the morphological analysis in order to identify the grammatical category of text words. Segmentation and Morphological analysis play a very important role in most applications of Natural Language Processing (e.g. information extraction, automatic summarization, etc.). The results of these two tools are presented successively in Figures 3 and 4.

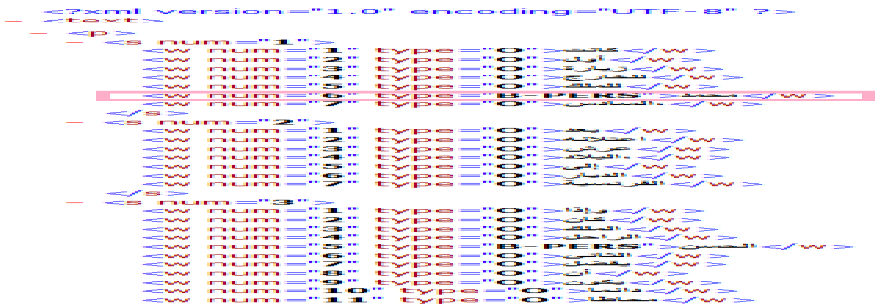


Fig. 3. Annotated text with ArNER

Fig. 4. : Morphological parser with Al-Khalil

The second step of our approach consists in analyzing the user's question. This is a preliminary step of the answer research process. Generally, the features extracted from this step facilitate the answer extraction. In Arabic, the majority of studies focuses on the extraction of keywords and named entity recognition. In our case, we focus on the reformulation of the question into a declarative form. This must be used shortly for generating logical forms.

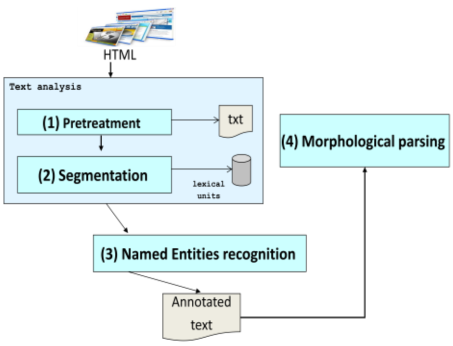


Fig. 5. Steps of text analysis

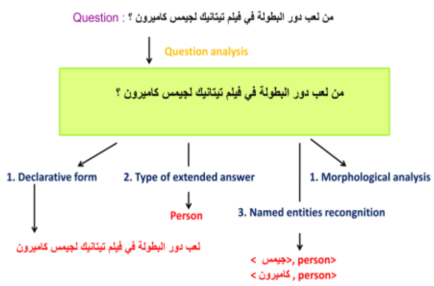


Fig. 6. Steps of question analysis

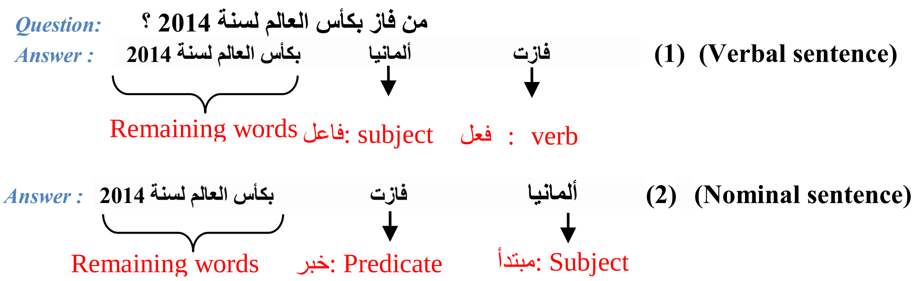
3.2 Transformation of the Question and the Text into Logic Forms

In this section, we look at the transformation of Arabic statements (question and text) into logic forms which affect the logical meaning of an Arabic sentence. Logic form is an intermediary step between the syntactic parse and the deep semantic form (Moldovan and Rus, 2001). Semantics is the most difficult level at which the language makes a contact with the real world. It is the most important part of Natural Language Processing. Practically all researches on the question answering in Arabic focus on morpho-syntactic approaches. Although, several approaches had been proposed for semantics, very few of them applied to Arabic. Logic-based approaches were applied to English and some other languages but none of them has been applied to Arabic yet.

Our approach deals with verbal and nominal simple sentences using AL-Khalil morphological parser to determine the grammatical category words in the Arabic statements. In the transformation of the Arabic statements into logical forms, we create our logic representations from a text and question data which are divided into two classes:

- A verbal sentence begins with a verb, and it has at least a verb (فعل) and a subject (فاعل). The subject can be indicated by the conjugation of the verb, and not written separately.
- A nominal sentence begins with a noun, it have two parts: a subject (مبتدأ) and a predicate (خبر). In this case, only nominal elements are used as constituents. The subject of the nominal sentence is a noun or a pronoun, while the predicate can be a noun, adjective, preposition and noun, or verb.

For a logical representation, we refer to a semantic representation of statements in natural language, especially the logical structure. A predicate expression is a graph of predicate-argument relationship, we work with the following example:



Arabic is different from English and other languages in the words order, the verb and noun criteria and the sentences type. In our case, we deal with verbal and nominal sentences. In verbal sentence, we handle the transitive and intransitive verb in order to assume the logical representations. In nominal sentence, we care the common and proper nouns. The identification of such predicate depends on its context. The predictive structure is a graph of predicate-argument relations.

Table 2. Mapping the words of a sentence to logical forms

Sentence words	***Rules***	***Logic predicates***
Example 1:	**فازت ألمانيا بكأس العالم لسنة 2014**	**(case of a transitive verb)**
Proper noun German almania ألمانيا	***logical constant*** Germany Almania ألمانيا	Germany Almania ألمانيا
Transitive verb win faza فاز	***Two predicates*** (λY) (λX) win (X,Y) (λY) (λX) faza (X,Y) (λY) (λX) فاز (X,Y)	Y^ X^ win (X,Y) Y^ X^ faza (X,Y) Y ^X^ فاز (X,Y)
Example 2:	**نام الطفل**	**(case of an intransitive verb)**
Common noun student tilmidh تلميذ	***One predicate*** (λX) student (X) (λX) tilmidh (X) (λX) تلميذ (X)	X^ student (X) X^ tilmidh (X) X^تلميذ (X)
Intransitive verb sleep naama نام	***One predicate*** (λX) sleep (X) (λX) naama (X) (λX) نام (X)	X^ sleep (X) X^ naama (X) X^ نام (X)

3.3 Implementation of the proposed approach

We have implemented our approach with the Java language, in NetBeans. We use ArNER to determine the named entities and Al-Khalil parser to identify the grammatical category of words. The results of these two steps are grouped into XML files.

In our work, the question analysis determines the expected answer type and reformulates the question in the affirmative form in order to generate hypotheses to be passed to the logical transformation module.

Question Analysis module is shown in Fig. 7.

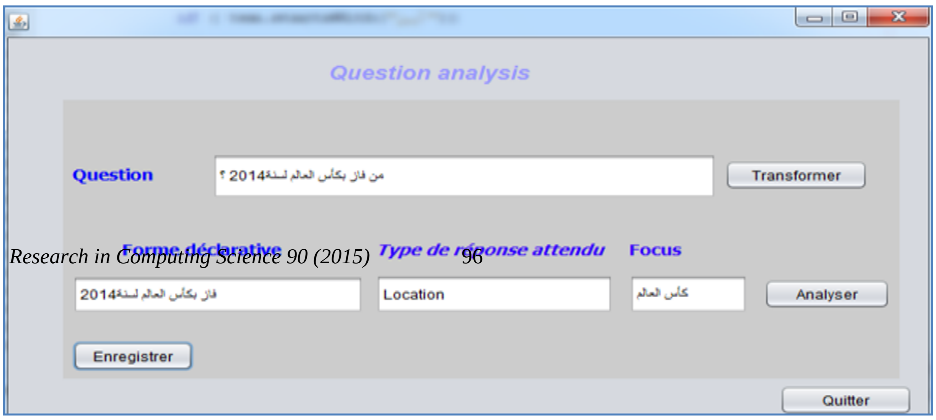


Fig. 7. Question analysis

4 Conclusion

Logic is a level between the syntactic and the deep analysis, it is the most important and difficult level in Natural Language Processing. Logic-based approaches are a rich research topic even though there is still room for improvement. This task was applied for many other languages (English, French...) but not yet to Arabic. This is due to the lack of necessary tools of Arabic and the specificities of this language. Our approach is a representative solution for Arabic.

In this paper, we proposed an original method intended for the Arabic within the framework of the automatic Arab statements comprehension to generate logical predicates. It is based on logic formulas. This method involves four steps: the textual statements analysis, the transformation of the question and text into logic forms, the textual entailments recognition and the desired answer retrieval.

For future work, we will extend this work to cover others cases in Arabic. We plan to study the phenomenon of the textual entailment recognition from logical forms generated. Then, we address these implications in the process of the desired answer selection. We investigate all cases studied logic formulas to generate the logical representations of textual statements. Also, we plan to implement this proposed model for the generation of the correct answer.

Acknowledgements

I give my sincere thanks for my collaborators Professor Patrice Bellot (University of Aix Marseille, France) and Mr Mahmoud NEJI (University of Sfax-Tunisia) and Mr Omar Trigui (University of Sfax-Tunisia) that I have benefited greatly by working with them.

References

1. Abouenour L., Bouzoubaa K. and Rosso P. (2012). "IDRAAQ: New Arabic Question Answering System Basedon Query Expansion and Passage Retrieval". *CLEF (Online Working Notes/Labs/Workshop).*
2. Athenikos SJ., Han H. (2010)."Biomedical question answering: A survey", Computer Methods and Programs in Biomedecine 99 (1):24, *PMID*, 19913938.
3. Bdour W. N and Gharaibeh N. K. (2013) "Development of Yes/No Arabic Question Answering System", *In International Journal of Artificial Intelligence & Applications (IJAIA),*. Vol.4. No.1, January 2013. DOI: 10.5121/ijaia.2013.4105 51.
4. 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", In *International Journal of Academic Research*; Jul2011, Vol. 3 Issue 4, p45.
5. Benajiba Y., Rosso P. and Lyhyaoui A. (2007). "Implementation of the ArabiQA Question Answering System's components", In: Proc. *Workshop on Arabic Natural Language Processing, 2nd Information Communication Technologies Int. Symposium. ICTIS*-2007, Fez, Morroco, April 3-5.

6. Bhaskar P., Pakray P., Banerjee S., Bandyopadhyay S. and Gelbukh A. (2012). "Question Answering System for QA4MRE@CLEF, *Workshop on Question Answering For Machine Reading Evaluation, QA4MRE*, September.
7. Brini W., Ellouze M., Mesfar S. and Belguith L. H. (2009)"An Arabic Question-Answering system for factoid questions", In *IEEE International Conference on Natural Language Procesing and Knowledge Engineering (IEEE NLP-KE'09),* Dalian, China.
8. 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", *11th International Semantic Web Conference (ISWC 2012),* Demo paper. Boston, USA.
9. Cao Y., Liu F., Simpson P., Antieau L., Bennett A., Cimino JJ., Ely J. and Yu H. (2011) "AskHERMES: an online question answering system for complex clinical questions", *J Biomed Inform*, 44(2):277–288.
10. De Chalendar G., Dalmas T., Elkateb-Gara F., Ferret O., Grau B., Hurault-Plantet M., Illouz G., Monceaux L., Robba I., Vilnat A. (2002). "The Question Answering System QALC at LIMSI, Experiments in Using Web and WordNet", *TREC11*, NIST special publication SP.
11. Ezzeldin A. M., Kholief M. H. and El-Sonbaty Y. (2013). "ALQASIM: Arabic Language Question Answer Selection in Machines", *CLEF 2013*, 2013. P 100-103.
12. Ferret O., Grau B., Hurault-Plantet M., Illouz G., Jacquemin C., Masson N. and Lecuyer P. (2000) "QALC-the question answering system of LIMSI-CNRS", Technical report: LIMSI-CNRS, *TREC 9*, evaluation.
13. Ghassan M. (2001) Analyse informatique des signes typographiques pour la segmentation de textes et l'extraction automatique des citations. Réalisation des Applications informatiques : SegATex et CitaRE, PhD thesis, Paris-Sorbonne.
14. Grau B., Pho V-M., Ligozat A-L., Ben Abacha A., Zweigenbaum P. and Chowdhury Md.F.M. (2012). "Adaptation of LIMSI's QALC for QA4MRE", *CLEF (Online Working Notes/Labs/Workshop).*
15. Hammo B., Ableil S., Lytinen S. and Evens M. (2004). "Experimenting with a Question Answering system for the Arabic language", In *Computers and the Humanities*. Vol. 38, N°4. Pages 397—415.
16. Harabagiu S. and Andrew H. (2006). "Methods for using textual entailment in open-domain question answering", In Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the ACL.
17. Kurdi H., Alkhaider S. and Alfaif N. (2014). "Development and evaluation of a web based question answering system for arabic language", In *Computer Science & Information Technology (CS & IT)*, 04(02), 187 – 202.
18. Kwok K-L., Choi S., Dinstl N. and Deng P. (2005). "NTCIR-5 Chinese, English, Korean Cross Language Retrieval Experiments using PIRCS", *In: Proc. of the Fifth NTCIR Workshop Meeting. NII*, Tokyo, pp.88-95.
19. Lee C-W., Shih C-W., Day M-Y., Tsai T-H., Jiang T-J., Wu C-W., Sung C-L., Chen Y-R., Wu S-H., and Hsu W-L. (2005). "Perspectives on Chinese Question Answering Systems", *in Proceedings of the Workshop on the Sciences of the Artificial (WSA 2005)*, Hualien, Taiwan.
20. Lee C-W., Shih C-W., Day M-Y., Tsai T-H., Jiang T-J., Wu C-W., Sung C-L., Chen Y-R., Wu S-H. and Hsu W-L. (2005) "ASQA: Academia Sinica Question Answering System for NTCIR-5 CLQA", *in Proceedings of NTCIR-5 Workshop*, Tokyo, Japan, 202-208.
21. Li X. and Croft W.B. (2001). "Evaluating Question Answering Techniques in Chinese" HLT *Notebook Proceedings*, pp. 201-206.

22. Ligozat A-L. (2013). "Classification de questions par traduction", *COnference en Recherche d'Informations et Applications, CORIA*.
23. Mitamura T., Lin F., Shima H., Wang M., Ko J., Betteridge J., Bilotti M., Schlaikjer A. and Nyberg E. (2007). "JAVELIN III: Cross-Lingual Question Answering from Japanese and Chinese Documents", *In Proceedings of NTCIR-6 Workshop*.
24. Mohammed F., Nasser K. and Harb H. (1993). "A knowledge based Arabic Question Answering system (AQAS)", *ACM SIGART Bulletin*, pp. 21-33.
25. Moldovan D.I. and Rus V. (2001). "Logic Form Transformation of WordNet and its Applicability to Question Answering", *ACL* . 394-401.
26. Nyberg E., Frederking R., Mitamura T., Bilotti M., Hannan K., Hiyakumoto L., Ko J., Lin F., Lita L., Pedro V. and Schlaikjer A. (2005). "JAVELIN I and II systems at TREC 2005", *In Proc. of TREC'05*.
27. Peñas A., Rodrigo Á., and Verdejo F. (2008). "Overview of the Answer Validation Exercise 2007", *CLEF 2007*, Lecture Notes in Computer Science LNCS 5152. Springer, Berlin.
28. Shaalan K. (2014), "A Survey of Arabic Named Entity Recognition and Classification", Computational Linguistics, 40 (2): 469-510, MIT Press, USA, 2014.
29. Shima H., Kanayama H., Lee Ch-W., Lin Ch-J., Mitamura T., Miyao Y., Shi S., Takeda K. (2011) "Overview of NTCIR-9 RITE: Recognizing Inference in TExt", *NTCIR-9*, 12/2011.
30. Sitbon L., Gillard L., Grivolla J., Bellot P., Blache P. (2006). "Vers une prédiction automatique de la difficulté d'une question en langue naturelle", *13ième conférence Traitement Automatique des Langues Naturelles (TALN)*, Louvain, Belgique, p. 337-346, 10-13 Avril.
31. Tahri A. and Tibermacine O. (2013). "DBPEDIA based factoid question answering system", *International Journal of Web & Semantic Technology (IJWesT),* Vol.4, No.3, July 2013.
32. Trigui O., Belguith L.H and Rosso P. (2010), "DefArabicQA: Arabic Definition Question Answering System", *In Workshop on Language Resources and Human Language Technologies for Semitic Languages, 7th LREC*. Valletta, Malta. 2010.
33. Voorhees E.M. (2004). "Overview of the TREC 2004 Question Answering Track", *TREC 2004*.
34. Wren JD. (2011). "Question answering systems in biology and medicine – the time is now", *Bioinformatics*, 27(14):2025–2026.
35. Zribi I., Hammami S. M. and Belguith L. H. (2010) "L'apport d'une approche hybride pour la reconnaissance des entités nommées en langue arabe", *In TALN'2010*, Montréal, 19-23 juillet 2010 (pp. 19–23).