

Towards an Automatic Text Comprehension for the Arabic Question-Answering: Semantic and Logical Representation of Texts

Wided Bakari, Patrice Bellot and Mahmoud Neji.

Abstract

Automatic text comprehension is an arduous task of automatic natural language processing. This task has been addressed in several NLP applications, namely, question-answering, information retrieval, etc. In some cases, these applications require a minimal understanding of the text. The use of this task in question-answering requires that the result must be the accurate answer to a question, instead of a number of references to documents that contain the answer. It makes it possible to construct a coherent representation of the content of a text to answer a question in natural language. In this paper, we propose to construct a semantic and logical representation of Arabic texts (the question and the passages of texts). Therefore, we propose a new approach for the Arabic question-answering. This approach is based on the automatic understanding of Arabic texts (question or passages of texts) to transform them into semantic and logical representations. The experiments results of our approach achieved an accuracy of 74%.

1 Introduction

With the growth of electronic media, including the Web, which has become the main source of information for everyone, the amount of information has grown exponentially. In the existence of this large-scale digitization, finding high-precision information is a challenge. Thus, the question-answering research is developing every day; Research in the field of Arabic question-answering began in the 1990s. To our knowledge, there are many systems that have been developed for many languages of the world (i.e. English, French, Chinese, Japanese, etc.). However, there are few systems that have been developed for the Arabic language. These systems provide answers in the form of short passages,

extracts of document collections. Therefore, the performance of these systems is limited by the difficulty of the language processing and the considerable lack of effective NLP tools that take into consideration the Arabic language [Alagha & Abu-Taha, 2015], [Al-Khalifa & Al-Wabil, 2007]. Arabic is an inflected and derivational Semitic language where from one root, several forms could be derived. This language poses challenges for the tasks of the NLP, as opposed to other languages, because of its morphological richness, relatively free words order and diglossic nature (where standard and dialects mix in most genres data) [Almarwani and Diab, 2017]. Moreover, with the absence of the capitalization in the Arabic language, we cannot distinguish the verbs from the nouns. The absence of voyellation in the Arabic texts is also a great source of ambiguity for morphological, syntactic and semantic analysis. Thus, the Arabic language still lacks large-scale computing resources that are very useful for English as the WordNet [Miller, 1995], or a resource, such as, VerbOcean [Chklovski and Pantel, 2004].

In this paper, we focus on a few dimensions whose development seems more particularly to the automatic understanding of Arabic texts. Specifically, we introduce a real analysis of text passages which may contain an answer to a question. Indeed, the automatic understanding of a text is one of the basic components of such a question-answering system to choose and obtain a precise answer. Consequently, the approach presented in this paper makes it possible to construct a semantic and logical representation of texts via the conceptual graphs. As a result, text comprehension can be defined as a cognitive process of understanding concepts from a text and relations between them.

The rest of this paper is organized as follows. In Section 2, we describe the automatic text comprehension in order to transform the texts

(question, text passages) into semantic and logical representations. Some related works presenting the most common of Arabic question-answering systems in Arabic language are presented in Section 3. Our proposed approach for the Arabic question-answering system is presented in section 4. The experimental results are presented and discussed in Section 5. Finally, we offer some concluding remarks in Section 6.

2 Automatic text comprehension

Automatic text comprehension has emerged in the 60s and 70s since the beginning of the automatic language processing. Indeed, to understand a text is quite different than to perform a simple reading. It is the task of achieving an in-depth understanding of one or a few texts. In fact, the task is focused on reading simple documents where the right answers require some inference and a review of the basic knowledge previously acquired [Banerjee et al., 2013]. It is, in fact, a way in which human comprehension of text has been measured to answer questions about the text [Vanderwende, 2007].

Various NLP applications need to understand the meaning of a text (although they often determine it superficially). Understanding is typing, recognizing, and relating the relevant elements to a task. In addition, these applications involve semantic processing of the natural language. In some cases, these applications require a minimal understanding of the text. The semantic annotation highlights in the same text relevant information elements (keywords, proper names, etc.). Information extraction makes it possible to extract structured information to fill a database (i.e. concerning company buybacks, networks of gene interactions, etc.). The automatic summary of texts aims at providing a target text summarizing the main information contained in a source text (s). The target is usually a simple selection of relevant sentences from the source text (s). In question-answering (QA), it is actually building a coherent representation of the content of a text to answer a question in natural language.

The texts build the mass of information the most present on the web (sound and images are more recent). As a result, a text is generally composed of sentences of equivalent size before their extraction. The sentence is a sequence of words and

punctuations that are combined to add semantics to the text [Patil et al., 2016]. In this context, researchers, such as, [Gomez-Adorno et al., 2013] define the automatic understanding through the ability to understand and read the main ideas written implicitly in a given text. In our work, the process of understanding is done in two ways; one way to represent the text in conceptual graphs and a second way to transform these graphs into logical forms of first order.

2.1 A semantic representation for the automatic text comprehension

The comprehension of texts is intended to give an account of the meaning of the text. Indeed, understanding a text supposes that we are able to form a cohesive and unified representation of the information emitted by the text. We propose to provide a semantic representation of texts if we want to determine the necessary entailments to answer a given question. To a certain extent, the task is to manage the linguistic variation in order to provide a representation of the text that makes it possible to manipulate it. In addition, the main challenge is how to convert the information found in the text into the language with which a machine can think and make decisions, and correctly answer the questions of the users. To do this, we suggest using conceptual graph formalism, which allows a richer representation of textual content. This formalism represents the sentence of the text and of the question with a structure formed by vertices and edges [Sowa, 1984]. It provides a higher level of text comprehension by capturing the semantics in the text.

The conceptual graph is a connected, bipartite and finite graph [Sowa, 1984]. It is used as an intermediate language to interpret the object-oriented formalism and the natural language. The representation of texts semantically via the conceptual graphs takes into account one of the current techniques that facilitate the process of handling different NLP applications, such as, information retrieval, question-answering, machine translation, automatic summary, reasoning, etc. The graph is constituted of a set of nodes connected by links; the nodes of the graph are either concepts (denoted by rectangles) or conceptual relations (denoted by ovals). Conceptual relations nodes indicate a relationship involving one or more concepts (they consist of a

type of relations). Concepts consist of a type of concept and a referent (instantiation of the type of concept). In natural language processing, the conceptual modeling is a way of modeling the semantics. The semantics of texts turns into semantics of conceptual models at a higher level of abstraction, in terms of concepts and relations. An example of a conceptual graph relating to the sentence "توجد بحيرة قارون في محافظة الفيوم" ("Qaroun Lake is located in Fayoum Governorate") is shown in Figure 1.

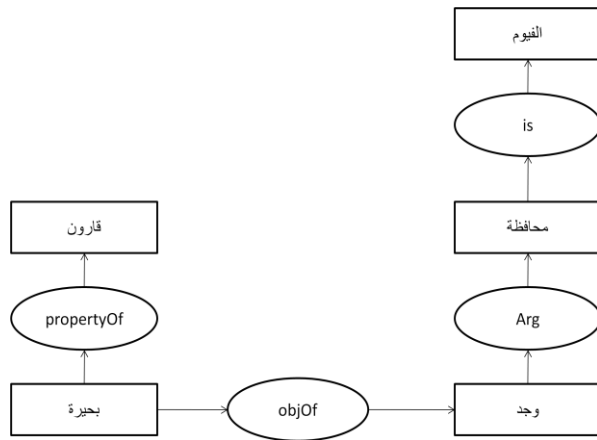


Figure 1: Example of a conceptual graph for "توجد بحيرة قارون في محافظة الفيوم"

2.2 A logical interpretation for the automatic text comprehension

The logic is the most important and difficult level of natural language processing. Indeed, one of the great challenges in understanding the text, such as, constructing a coherent overall representation of the text, is that much of the information needed in this representation is implicit. Thus, in order to fill the gaps and achieve a global representation, the language processing systems require a great deal of knowledge of the world, and the creation of these knowledge resources remains a fundamental challenge [Clark et al., 2008]. Otherwise, being able to represent a problem is interesting only if we can use this representation to perform intelligent tasks, such as, the logical reasoning. Thus, another important role of the logic is to develop a mechanism that allows a machine to perform reasoning.

We mean a semantic representation including a logical processing of texts in natural language. Our goal is twofold, from a mathematical point of

view; it is to better understand the structure of the first-order logical formulas, especially, if they are used as semantic representations of texts in natural language. From a linguistic point of view, it is a question of combining two modes of representation used in the semantic and the logic of natural language. We make the logical representation into a form of predicate-arguments that allows transforming the semantic representations in conceptual graphs (of the question and the text passage) into logical forms. The implementation of logical forms is based on information provided by the conceptual graphs. Obviously, we have proposed an algorithm that provides the transformation process and creates predicates and assigns arguments to them. The logical forms generated from the conceptual graphs are taken into consideration during the textual entailment determination step that relies on the semantic and the logic representation to extract the desired answer.

In particular, [Sowa, 1984] defines an operator Φ (Phi) that transforms every element of a conceptual graph model into an element of first-order logic. This interpretation gives a particular semantic to the generic marker. The transformation of a conceptual graph into a first order logic formula is as follows. At each node of the graph corresponds a quantized variable to which is applied a unary predicate having for name the word associated with the node. The arcs introduce a binary predicate with the name of the arc tag and for arguments the variables associated with its source and destination nodes. The final formula is the conjunction of the predicates thus obtained. For the graph of figure 1, we obtain the following logical representation:

$$\exists X \exists Y \exists Z \exists W \exists T : \text{بحيرة}(X) \wedge \text{وجد}(Y) \wedge \text{objOf}(X,Y) \wedge \text{قارون}(Z) \wedge \text{propertyOf}(X,Z) \wedge \text{محافظة}(W) \wedge \text{Arg}(Y,W) \wedge \text{الفيوم}(T) \wedge \text{is}(W,T).$$

3 Related works

In this section, we include a review about the Arabic question-answering systems. The research in area has been catalyzed by the Text Retrieval Conference (TREC) series since 1999. Question-answering systems are complex systems that, given a question asked in natural language, can find an answer to this question, in a corpus or in the Web,

and justify it by quoting their source(s). The first works in this research field go back to the 1960s for the English language and 1990s for the Arabic language. Research on Arabic Natural Language Processing (NLP) is facing a lot of problems due to language complexity, lack of machine readable resources and lack of interest among Arabic researchers [Al-Shawakfa, 2016].

Since then, in Arabic language, there are some attempts of question-answering systems that have started to appear since 1990s by introducing various approaches that can be used to extract an answer to a question. These systems have obtained acceptable results for finding the precise answer. Despite these attempts, systems in Arabic are still few and immature because of the unique aspects of this language. In the remainder of this section, we have presented most of these systems that are performed in this language since their birth with AQAS from [Mohamed et al., 1993] to Lemaza [Azmi and Alshenaifi, 2017] which is recently presented. Figure 2 shows the review of the most common question-answering systems in Arabic language.

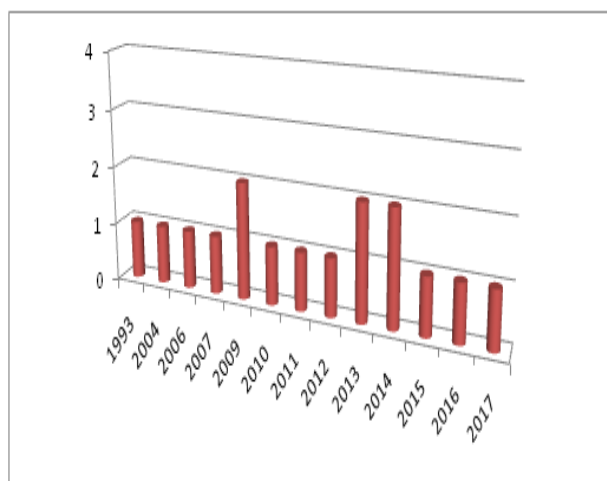


Figure 2: Question-answering systems in Arabic

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; it transforms the question into a query to retrieve the answer and seeks answers from structured data bases that focused on a knowledge model.

[Hammo et al., 2004] presented QARAB (Question answering for Arabic), which is an Arabic QA system that uses both natural language processing and information retrieval techniques. Its data was extracted from the Alraya newspaper published in Qatar. It retrieves only short passages that contain an answer, not necessarily the exact answer.

In 2006, [Rosso et al., 2006] introduced an ongoing implementation of a Factoid Arabic QA system that was used for the purposes of QA tracks in the Cross Language Evaluation Forum (CLEF) and Text Retrieval Conference (TREC) competitions. In their paper, the authors have only introduced partially implemented modules of the system in which the Named Entity Recognition (NER) module as well as a Java Information Retrieval System (JIRS) module were embedded.

This system was completed and introduced later on by [Benajiba et al., 2007], in which the effect of correctly identifying a Named Entity (NE) to produce correct answers is emphasized. This system is called ArabiQA (Arabic question answering system), is an Arabic question answering system. This system is based on a generic architecture made up of three modules: a passage retrieval system (JIRS), an Arabic Named Entities Recognition system (NER) and an Answer Extraction (AE) module. This system is making a precision near 83.3%.

QASAL [Brini et al., 2009] is a prototype to build an Arabic factual Question Answering system using Nooj platform to identify answers from a set of education books. The Experimentations have been conducted and showed that for a test data of 50 questions the system reached 67.65% as precision, 91% as recall and 72.85% as F-measure.

[Kanaan et al., 2009] propose a question answering system that provides short answers to questions expressed in the Arabic language. The system utilizes techniques from Information Retrieval and Natural Language Processing to process a collection of Arabic text documents as its primary source of knowledge. The authors used an existing tagger to identify proper names and other crucial lexical items and build lexical entries.

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.

AQUASYS (Arabic Question-Answering System) [Bekhti et al., 2011] is a system focused on named entities and designed to answer factoid questions related that can be of any type: person, location, organization, time, quantity, etc. This system addresses the questions that start with interrogative pronouns (من who, ما what, أين where, متى when, كم كم العددية how many, كم الكمية how much).

IDRAAQ system [Abouenour et al., 2012] is implemented through a three-level approach to enhance the Passage Retrieval (PR) stage. It includes the Recognizing Textual Entailment (RTE) and Answer Validation tasks. Without using the database collections of CLEF and relying on the density of the distance N-gram model, semantic expansion and Arabic WordNet, IDRAAQ achieved a precision of 0, 13 and c@1 equals 0, 21. The system of [Bdour and Gharaibeh, 2013] is an open domain yes/no Arabic question answering system based on paragraph retrieval; it aims to retrieve paragraphs (with variable length) that contain answers to the question.

ALQASIM proposed by [Ezzeldin et al., 2013] is based on selection and validation of the answer, it answers multiple choice questions. This system analyzed 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.

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).

Al-Bayan [Abdelnasser et al., 2014], an Arabic question-answering system specialized for the Holy Quran. It accepts an Arabic question about the Quran, retrieves the most relevant Quran verses, then it extracts the passage that contains the answer from the Quran and its interpretation books (Tafseer).

An Entailment-based Why Arabic Question Answering (EWAQ) system was introduced by [Al-Khawaldeh, 2015]. According to the author, EWAQ enhanced the accuracy of “Why” questions by improving the re-ranking of passages that is

relevant and retrieved by many search engines as possible answers. She claimed that the accuracy of her system has improved over that of search engines; Google, Yahoo and Ask.com.

[Hamdelsayed and Atwell, 2016] proposed a rule-based system for the Holy Quran that retrieves the right verse(s) from the Holy Quran instead of generating NL answers. This system is applied for only two chapters: Al-Bagrah and Al-Fatihah (البقرة والفاحة). The authors use a java program to extract the answer from a MS-Access database which contains our knowledge base for our Quran question-answering system.

Lemaza [Azmi and Alshenaifi, 2017] is a question-answering system that exclusively handles why questions expressed in Arabic language. Most of the existing Arabic question-answering systems do not handle why-questions, as these are more difficult to handle. The authors tested our system using 110 why-question and answer pairs using a corpus of 700 documents extracted from Open Source Arabic Corpora.

4 Proposed Approach

We propose a new approach for the Arabic question-answering. This approach is based on the automatic understanding of Arabic texts (question or passages of texts) to transform them into semantic and logical representations. More precisely, we manage to introduce a real analysis of the text that can contain an answer to a question. Isolated from the domain studied in question-answering, the problem of understanding a text and ensuring its in-depth analysis, have not had sufficient priority except in the last five years in comparison with that given to the analysis of the question. We suggest that automatic understanding of a text is one of the basic components of such a question-answering system for choosing and obtaining a precise answer. Therefore, our approach allows us to construct a semantic and logical representation of texts via the conceptual graphs. As a result, text comprehension can be defined as a cognitive process of understanding concepts from a text and relationships between them.

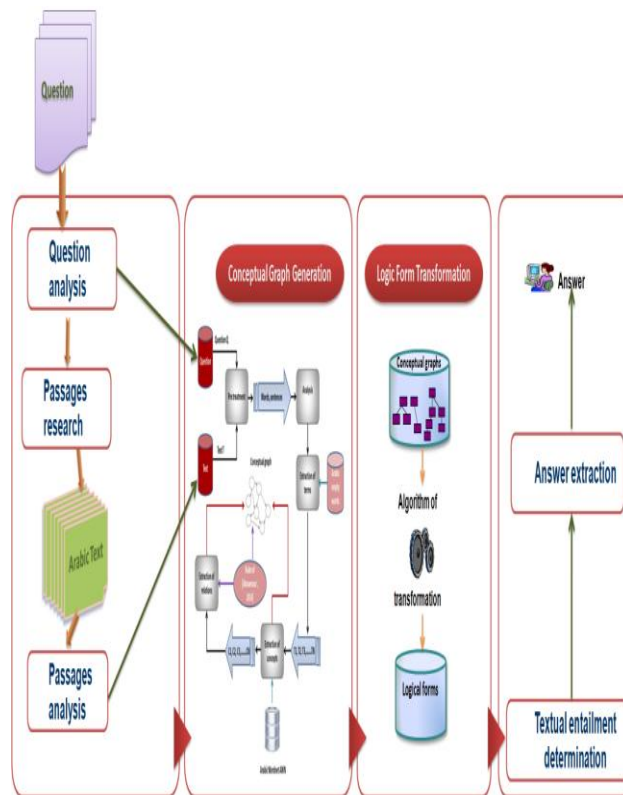


Figure 3: Proposed approach

The proposed approach involves seven main phases (Figure 2), namely : (1) question analysis, (2) passages research, (3) passages analysis, (4) conceptual graph generation, (5) logic form transformation, (6) textual entailment determination, finally (7) answer extraction. In the following subsections, we will discuss the different phases of our proposed approach for the Arabic question-answering system.

4.1 Question analysis

This step consists of presenting the various characteristics that can help us to find the passages of texts and to select the precise answer. Although that the techniques differ from one system to another, the majority of these systems are based on a step of analysis of the question. This is a preliminary step in the process of finding specific answers to questions in natural language. In the literature, there are several types of questions that are often related to a particular type of answer as well as different approaches to answering to it. In our work, we focus on the type of factual questions that primarily require a named entity's answer. This

type of questions begins with When, Where, How much, Who and What, etc., and asks for a date / time, place, person and organization, etc.

4.2 Passages research

Text passages are an important intermediary between complete documents and exact answers [Brini et al., 2009]. The keywords of the question are provided to the Google search engine to get the text passages. Moreover, the existence of these words in the passages can show the presence of the answer to this question. The method for finding text passages for each question is described in [Bakari et al., 2016]. First, we look for URLs that match those words. Then, for each address, we propose to recover the Web page that suits him. Finally, we transform each selected web page into a txt format (".html" to ".txt"). We obtain for each question a text comprising several passages of texts.

4.3 Passages analysis

Text generated from the Web contains information that does not make sense (i.e. transliterations, etc.). First, we propose to clean the texts, to transform them into an appropriate form and to prepare them for future treatments. Then, we transform these texts into a standard format that can be easily manipulated. Thus, we propose to add the UTF-8 in-coding, to filter the non-Arabic letters, etc. Then, the texts extracted from the Web are often long; their treatment consists of segmenting them into sentences. Finally, a list of linguistic analysis, including, the named entity recognition, the syntactic and the morphological analysis is implemented. This analysis increases the chance of finding the precise answer to a question in natural language.

4.4 Conceptual graph generation

In this step, we propose to transform the questions and the text passages into conceptual graphs. Indeed, a conceptual modeling in NLP is a way of modeling the semantics. The text semantic is transformed into semantics of conceptual models at a high level of abstraction, in terms of concepts and relations. Moreover, a concept represents an object of interest also called object of knowledge. Conceptual relations connect the concepts. A conceptual graph is a connected set of concepts

and conceptual relations. To transform a text (passage, question) into a conceptual graph, we first introduce the list of terms. Then, we extract the list of concepts associated to these terms. Then, we identify the list of the relations between these concepts. Finally, we construct the graph corresponding to these concepts and relations.

4.5 Logic form transformation

Representing the question and the text passages with a conceptual graph can facilitate the generation of logical representations. In our work, we propose a conceptual graph transformation algorithm based on the Φ operator principle proposed in [Sowa, 1984] in order to obtain a logical representation. This algorithm transforms each element in a conceptual graph into an element of logic. It associates with any type of concept a predicate (i.e. Person (X)), and with any type of relation, a predicate of the same arity as the type (i.e. Agentof (X, Y)). It also associates with individual markers constants (i.e. ALI) and generic markers variables (i.e. *) that can be considered as arguments (i.e. (Person: ALI), (Person: *) \rightarrow (Person (ALI), Person (X)). Finally, the result is a logical formula obtained by conjunction of predicates associated with the concepts and the relations (i.e. Person (X) \wedge Agentof (X, Y), ..., etc.).

4.6 Textual entailment determination

In our work, we represent the questions (hypotheses H) and the passages of texts (text T) in logical forms to determine the relation of textual entailment between them. To do this, we propose a semantic method that is based on the extraction and the combination of features to determine the relation of entailment between pairs of logical forms. We discover three features (i.e. (i) overlap of predicate-arguments, (ii) matching between named entities, and (iii) semantic similarity). Since the text T and the hypothesis H are respectively represented by logical representations FOLT and FOLH. For each feature, we assign a score and we take as input the logical representations FOLT and FOLH. For the determination of textual entailment, we use different measures (i.e. the measure overlap, a correspondence score, the measure Wu-Palmer, ..., etc.). The problem of entailment is considered a problem of classification with two

classes, a class "YES" and a class "NO." We use the values of features like data of learning to form characteristic vectors. These vectors are transmitted to a WEKA J48 decision tree classifier [Witten and Frank, 1999] which classifies them as "yes" or "no" entailments.

4.7 Answer extraction

After having obtained the results of the textual entailment, we extract only the passages having a relation of entailment with the question (yes) and we discard all the passages which have no relation of entailment with the question. The passages retained with entailment are considered as the candidate answers. Subsequently, a confidence score for each entailment is calculated by determining the weighted average. The passages are ordered in descending order. Then, the passage of text with the highest confidence score is the best candidate passage that contains the answer to the question. In our case, the answer is either a passage that contains an answer or a named entity.

5 Experimental results

In this section, we describe our performed experiments to evaluate our approach. For the evaluation, we used a set of 250 factual questions with 2500 passages of texts that are extracted from the web. To analyze the impact of the proposed approach, we used a standard set of 250 Arabic factual questions as dataset. The experimental questions come from four sources, namely, discussion forums, commonly asked questions (FAQ), some questions of the two evaluation campaigns TREC and CLEF. These English questions are translated into Arabic for experiments. The set of 250 factual questions were fed into Google to retrieve the relevant text passages that can contain the relevant answers. These passages were also analyzed in order to extract the correct answers. We have obtained an accuracy of 74%, which is very encouraging compared to the size of the tag set used till now. In fact, good accuracy performance is mainly due to both automatic understanding of Arabic texts for question-answering and recognition of textual entailment to the question-answering in Arabic.

6 Conclusion

In this paper, we have presented the theoretical foundation for a new logical approach to the question-answering in Arabic, namely, the automatic text comprehension. We have described a new approach for the Arabic question-answering. This approach has brought together several techniques, such as, NLP techniques, logical reasoning techniques, RTE techniques, etc. It is based essentially on the automatic understanding of Arabic texts (question or passages of texts) in order to transform them into semantic and logical representations. In our work, we propose not only to obtain exact and precise answers to questions in Arabic, but also to carry out semantic and logical analyzes for the question and for text passages. For a logical representation we referred to a semantic representation. The idea was to transform these texts into conceptual graphs, a powerful logic-based formalism that is used to model textual information through concepts and relations. From this formalism, we proposed a conversion algorithm to model and determine a logical representation for each graph.

References

- Banerjee, S., Bhaskar, P., Pakray, P., Bandyopadhyay, S., & Gelbukh, A. F. (2013, September). Multiple Choice Question (MCQ) Answering System for Entrance Examination. In CLEF (Working Notes).
- Vanderwende, L. (2007, March). Answering and Questioning for Machine Reading. In AAI Spring Symposium: Machine Reading (p. 91).
- Patil, N. V., Patil, A. S., & Pawar, B. V. Issues and Challenges in Marathi Named Entity Recognition. *International Journal on Natural Language Computing (IJNLC)* Vol, 5, 15-30. 2016.
- Gómez-Adorno, H., Pinto, D., & Ayala, D. V. (2013). Semantic Answer Validation in Question Answering Systems for Reading Comprehension Tests. In AMW.
- Sowa John F. (1984). *Conceptual Structures: Information Processing in Mind and Machine*. Addison-Wesley Company.
- Clark, P., Harrison, P. Recognizing Textual Entailment with Logical Inference. In *Proceedings of 2008 Text Analysis Conference (TAC'08)*, Gaithersburg, Maryland, 2008.
- Mohammed, F. A., Nasser, Khaled, et Harb, H. M. A knowledge based Arabic question answering system (AQAS). *ACM SIGART Bulletin*, 1993, vol. 4, no 4, p. 21-30.
- Hammo, B., Abuleil, S., Lytinen, S., & Evens, M. (2004). Experimenting with a question answering system for the Arabic language. *Computers and the Humanities*, 38(4), 397-415.
- Benajiba, Y., Rosso, P., & Lyhyaoui, A. (2007, April). 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, Morocco, April (pp. 3-5)*.
- Brini, W., Ellouze, M., & Hadrich Belguith, L. (2009). QASAL: Un système de question-réponse dédié pour les questions factuelles en langue Arabe. 9ème Journées Scientifiques des Jeunes Chercheurs en Génie Electrique et Informatique, Tunisia.
- Trigui, O., Belguith, L. H., & 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 (pp. 40-45)*.
- Kanaan, G., Hammouri, A., Al-Shalabi, R. and Swalha, M. "A New Question Answering System for the Arabic Language", *American Journal of Applied Sciences* 6 (4). 2009. 797-805.
- Abouenour, L., Bouzoubaa, K., & Rosso, P. (2012). IDRAAQ: New Arabic Question Answering System Based on Query Expansion and Passage Retrieval. In CLEF (Online Working Notes/Labs/Workshop).
- Abouenour, L. (2014). Three-levels Approach for Arabic Question Answering Systems. Diss. Ecole Mohammadia d'Ingénieurs.
- Bdour W. N and Gharaibeh N. K. (2013) —Development of Yes/No Arabic Question Answering Systeml, In *International Journal of Artificial Intelligence & Applications (IJAIA)*,. Vol.4. No.1, January 2013. DOI: 10.5121/ijaia.2013.4105 51.
- Ezzeldin, A. M., Kholief, M. H., & El-Sonbaty, Y. (2013, September). ALQASIM: Arabic language question answer selection in machines. In *International Conference of the Cross-Language Evaluation Forum for European Languages (pp. 100-103)*. Springer, Berlin, Heidelberg.
- Bekhti, S., Rehman, A., AL-Harbi, M. and Saba, T. "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.
- Kurdi, H., Alkhaider, S., & Alfaif, N. (2014). Development and evaluation of a web based question answering system for arabic language. *Computer Science & Information Technology (CS & IT)*, 4(02), 187-202.
- Abdelnasser, H., Ragab, M., Mohamed, R., Mohamed, A., Farouk, B., El-Makky, N., & Torki, M. (2014). Al-Bayan: an arabic question answering system for the holy quran. In *Proceedings of the EMNLP 2014 Workshop on Arabic Natural Language Processing (ANLP)*(pp. 57-64).
- Bakari, W., Bellot, P., & Neji, M. (2016). AQA-WebCorp: Web-based factual questions for Arabic. *Procedia Computer Science*, 96, 275-284. ISO 690
- Ian H. Witten and Eibe Frank. 1999. *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*. Morgan Kaufmann.
- Al-Khawaldeh, F. "Answer Extraction for Why Arabic Question Answering Systems: EWAQ," *Proc. World of Computer Science and Information Technology Journal (WCSIT)*, vol. 5, no. 5, pp. 82-86, 2015.
- Al-Shawakfa, E. (2016). A Rule-Based Approach to Understand Questions in Arabic Question Answering. *Jordanian J. Comput. Inf. Technol*, 2(3), 210-231.
- Azmi, A. M., & Alshenaifi, N. A. (2017). Lemaza: An Arabic why-question answering system. *Natural Language Engineering*, 23(6), 877-903.
- Rosso, P., Benajiba, Y. and Lyhyaoui, A. “Towards an Arabic Question Answering system”, In *Proceedings of 4th Conference on Scientific Research Outlook & Technology Development in the Arab world. SROIV. Damascus, Syria. 11-14 December. 2006.*
- Hamdelsayed, M., & Atwell, E. S. (2016). Islamic Applications of Automatic Question-Answering. *Journal of Engineering and Computer Science*, 17(2), 51-57.
- Miller, George A. *WordNet: a lexical database for English*. *Communications of the ACM*, 1995, vol. 38, no 11, p. 39-41.
- AlAgha I, Abu-Taha A (2015) AR2SPARQL: An Arabic Natural Language Interface for the Semantic Web, *International Journal of Computer Applications*, 125(6), pp 19-27.
- Al-Khalifa H, Al-Wabil A (2007) The Arabic language and the semantic web: Challenges And opportunities, In: *The first international symposium on computers and the Arabic language*, November 2007, Riyadh, Saudi Arabia.
- Almarwani, Nada and Diab, Mona. Arabic Textual Entailment with Word Embeddings. In : *Proceedings of the Third Arabic Natural Language Processing Workshop*. 2017. p. 185-190.
- Chklovski, Timothy and Pantel, Patrick. Verbocean: Mining the web for fine-grained semantic verb relations. In : *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*. 2004.