

Arabic Question Answering: A Study on Challenges, Systems, and Techniques

Heba Samy
Faculty of Computers &
Information
Cairo University, Egypt

Ehab E. Hassanein
Faculty of Computers &
Information
Cairo University, Egypt

Khaled Shaalan
Faculty of Engineering & IT,
The British University in Dubai,
UAE

ABSTRACT

The enormous increase of the amount of information available on the web creates the need for systems like Question Answering to bridge the gap between general end users and the web with its different data representations. A considerable portion of the available data on the web is written in Arabic for and by Arabic users. This paper provides a review of the Arabic Question Answering Systems building processes and the challenges met by the researchers in this topic due to the Arabic language special characteristics. A general architecture is represented for the Question Answering task on both structured and unstructured data. Then, an overview of the work done in Arabic Question Answering Systems is presented. Finally, a number of tools and linguistic resources are recommended for researchers to develop Arabic question answering systems.

General Terms

Question Answering, Question Answering Systems, Natural Language Processing, Information Retrieval

Keywords

QA, QAS, Ontology, linked data

1. INTRODUCTION

In real time, the Web is the main source of information for all, general users to experts, students to researchers. Due to the enormous amount of information found on the web, there is a persistent need to make it easier and more natural for a user to ask a question and get specific answer. This is on the contrary to just using search engines, indicating a set of keywords and get back a set of relevant documents, to search manually for the related answer. This process is time consuming, not preferred by a user as sometimes the retrieved documents may not contain the relevant answer. This motivates the need for building Question Answering Systems, as a more natural user-web communicating means [1][2]. Many corporations and workshops applied Question Answering, such as CLEF,

MUC, NTCIR, TREC, among others [3].

Question Answering (QA) is defined, in the field of Natural Language Processing (NLP) and Information Retrieval, as the ability of a computer to answer a natural language question entered by a human. In most cases, Question Answering Systems (QAS) contain mainly three modules that may contain sub modules. These main modules are; Question Processing, Document Processing and Answer Processing [4]. Fig. 1 from [5] depicts the general architecture for a Question Answering System.

The Question Processing module is the one that is responsible for the question analysis to generate syntactic and, in some advanced cases, semantic representation of the natural language question, so that the system can understand what the user is asking about and how to return back a precise answer to his/her question. So, the input to this module is an unstructured text question and the output is structured information that are used by the following other components. [2]

The Question Processing module includes sub modules, such as the question classification sub-module that that is connected with the expected answer-type generation sub-module, keyword extraction, and query expansion sub-modules [6].

In Question classification, questions are assigned classes such as Factoid, definition, list and Boolean. Question classification taxonomies may be flat or hierarchical. In [7] a flat taxonomy is introduced in which seventeen classes are used, while [8] presented a hierarchical taxonomy with a nine main classes with a number of subclasses in each [9]. Two approaches were used for the question classification: machine learning based and rule based classification. Passage or document retrieval or processing module, retrieve, analyze and rank documents then passages, that are expected to contain the required answer. Answer extraction module use retrieved relevant documents or passages from the document processing module, extracts and ranks candidate answers [9]

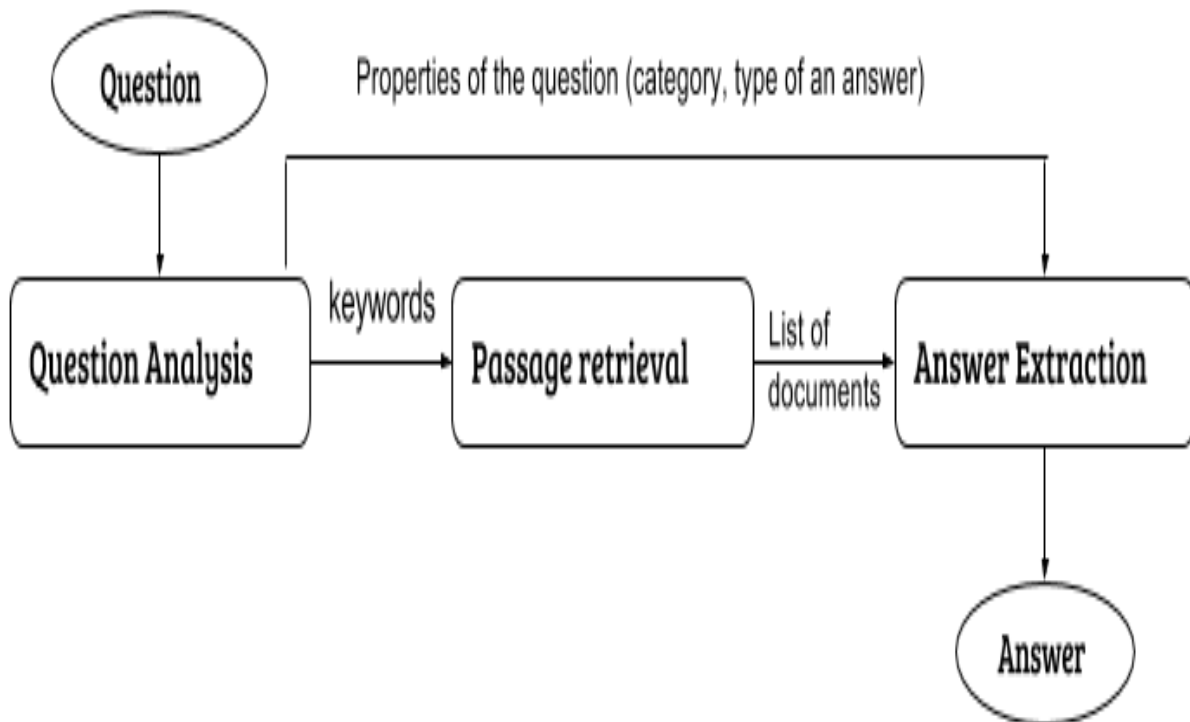


Fig 1: A QAS general architecture

The structure of the survey is as follows. Section 2 describes background information about Arabic Question Answering. Section 3 provides a brief description of the question answering system architecture on both structured and unstructured knowledge bases. Section 4 provides some characteristics of Arabic language that presents the difficulties specific to Arabic Question Answering. Section 5 provides an overview of the work done by researchers in this field. Section 6 briefly describe the language-specific tools and techniques used in Arabic QAS. Finally, the conclusion in Section 7.

2. BACKGROUND

Question Answering field overlaps with other fields such as information retrieval, information extraction and database querying that it has a different aim but may use methods by these fields to achieve its main target. Question answering applications can be classified according to different aspects. For example, the domain it covers, or the structure of its knowledge base. This section describes the ways QA overlaps with other fields and how it can be classified accordingly.

2.1 Question Answering vs. Information Retrieval

Information retrieval is the process of searching for information in documents and databases, whether these databases are relational databases or networked databases. The information that can be retrieved are documents, information within these documents, or metadata about documents [2]. While Question answering is the process of searching in the documents or databases, for a precise answer to a question asked by a human user in his natural language.

A Question Answering System may contain an information retrieval module that fetches related data to the user question keywords, from an open or closed domain of documents or databases, and the output can be used by a following answer extraction module that extracts the exact answer.

2.2 Question Answering vs. Database Querying

While both Question answering and Database Querying give users the ability to access a big volume of information and extract information relevant to their needs, each one has a special goal and architecture. Question answering systems first input is a natural language human question, but Database Querying systems first input is a database query written in a specific query language [10]. A Question Answering system can query over a database by translating the user natural language question keywords to a database query, thus a Question Answering system may include a database querying system as a subsystem.

2.3 Types of QAS

Question Answering Systems can be divided into two main groups according to the methods used in building them. The first one is based on simple natural language processing techniques with information retrieval methods, while the second one depends on natural language reasoning techniques. These two major groups of Question Answering systems can be distinguished according to different dimensions such as the domain it accesses, the techniques it uses, the type of data it analyzes [2]. Table 1 from [2] contains a good comparison between these two groups.

3. THE ANATOMY OF QAS

This section introduces the sub processes included in the question answering process, as shown in Fig 1. Then, describes both the general architecture of QA systems both on structured and unstructured data. These general architectures are used by most of the systems, although they differ in implementation methods [11].

Table 1 Characterization of QA System

Dimensions	QA system based on NLP and IR	QA systems Reasoning with NLP
Techniques	Syntax processing, Named Entity tagging and Information Retrieval	Semantic Analysis or high reasoning
Data Resources	Free text documents	Knowledge Base
Domain	Domain Independent	Domain Oriented
Responses	Extracted Snippets	Synthesized Responses
Questions Deals with	Mostly wh- type of Questions	Beyond of wh-type of questions
Evaluations	Uses existing Information Retrieval	N/A

3.1 Dimensions of QAS

To decide which components and modules to be used when designing a QAS, a number of factors should be taken into consideration, including: The QAS context, the questions as well as expected answer type, and the format of the data to be analyzed [1]. This section presents some of the important dimensions of the question answering task.

3.1.1 Question and Answer Type

Most Arabic Question Answering Systems are built for factoid and definition questions. This can be truer when talking about QAS based on linked data. This is because it is the available datasets which contain these kind of information [11]. Question in most cases is classified according to the expected answer type. Table 2 represents a question depicted from [11], with answer taxonomy from [12] and the Arabic question class word added to it.

For some type of questions, the answer type can be detected from the *focus* of the question, and sometimes the focus with the *topic*. For example, the two questions “what is the number of red balls in the basket?” and “what is the number of digits of 10?” the focus here for both questions is the number, while the topic is red balls at the first and digits at the second.

Table 2. Question and Answer Taxonomy

Question Class	English Question word	Arabic Question word	Expected Answer Type
Predicative questions	When	متى	Date , Time , Duration
	Where	اين	Location
	Which	اى	Person , Location , Organization , Date, Group of People
	What	ما	Person , Location , Organization , Date, Number, Reason, Explanation
	Who/Whose/Whom	من / لمن	Person , Organization, Group of People
	How Many / How Much	كم	Number, Percentage, Measure, Monetary Values
	Function words	اذكر اسم ... عرف معنى ... اذكر سبب ...	Person , Location , Organization , Date, Number, Reason, Explanation, Group of People
	List questions	Give me	اذكر اسماء ...
Yes/No Questions	Is	هل	Yes or No
Definition	Who is ... What is the...	من هو ما هو	Definition Sentence
Evaluation or comparative	What is the difference between ...	ما هو الفرق بين ...	Comparative Sentence
Association	What is the connection between ...	ما علاقة ...	Association Sentence
Explanation/Justification	Why	لماذا	Reason, Explanation
Process	How	كيف	Process

3.1.2 Data Sources

QA systems differ according to the data sources they access to retrieve an answer. Some systems use structured data, like

relational databases and linked data. Others use semi-structured data such as documents in XML. Other systems consume unstructured data such as plain text [11].

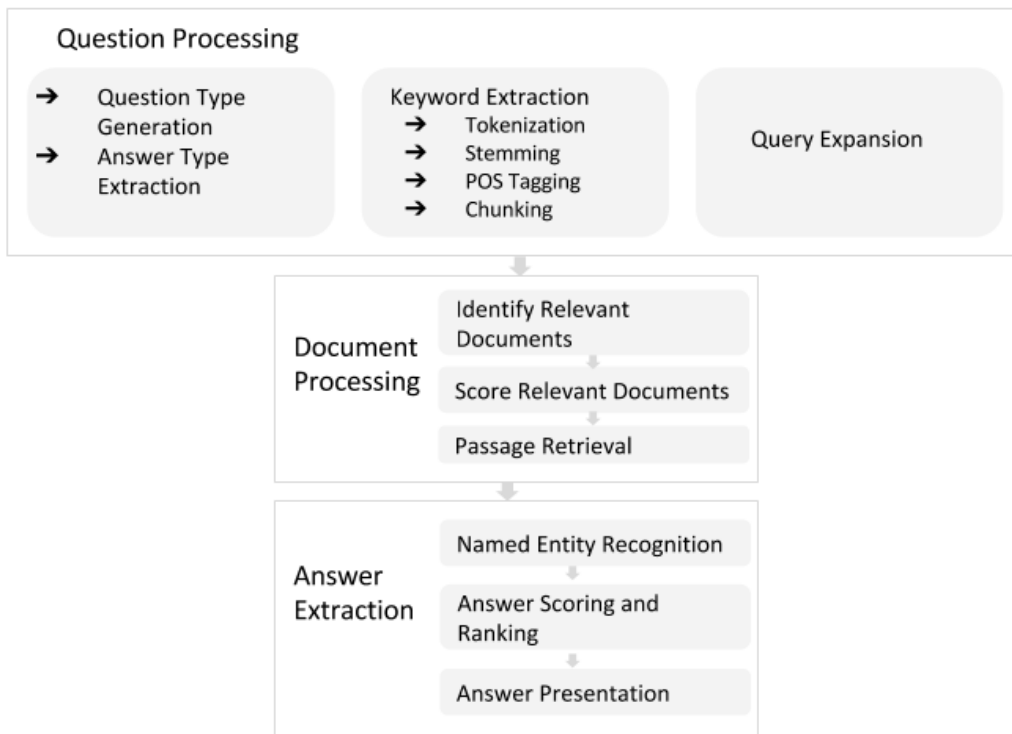


Fig 2: Phases of question answering system

3.2 Components of a QAS

3.2.1 Arabic QAS on Unstructured Data

A huge amount of Arabic unstructured or textual data is available online, such as Arabic Wikipedia. Hence there is a need for directly answering questions from these documents. [13] The framework for Arabic QAS on unstructured data is shown in Fig 2. Then a following explanation of each stage in the architecture, deduced from [12]. For unstructured data, the Arabic QA process passes through three main phases, which are: Question Processing, Document Processing, Answer Processing [4]. These three phases are optional, that is some QAS, like statistical Question Answering Systems do not include the question processing module and begin by the document retrieval phase. Others such as Definitional QAS skip the document processing phase and jump from the question processing to the answer processing phase [12].

3.2.1.1 Question Processing

Given a natural language question as input, the question processing module is the module that determines what the user is asking for. This is made by the following common subtasks: Question classification and answer type expectation, Keyword Extraction, and Query Expansion.

- **Question classification and Expected Answer detection:** By classifying the user question into a question type class, the expected answer type can be detected, thus provides an increase in the accuracy of the retrieved answer [2].
- **Keyword extraction:** The process of extracting the list of keywords from the question, can be done with

the help of some techniques such as named entity recognition, stop-word lists, and POS tagging and a set of heuristics such as the quoted expressions in the question, proper nouns, nouns, verbs, nominals and complex nominals (detected by named-entity recognition), and the question focus.

- **Tokenization:** breaking up the question into a set of meaningful words, so-called tokens. For example, in the question “متى ذهبت خديجة لوالدتها؟” “when did Khadija go to her mother?”, the tokens will be (متى - ل - والده - ها - ذهبت - خديجة - ل - والدته - ها) which are (when - did go - Khadija - to - mother - her)
- **POS tagging:** is the processes of assigning a part of speech (such as noun, verb, adjective, etc.) tag (word category) to each word or token in the question sentence.
- **Stemming:** is the process of reducing the inflected or derived words to their root (or primitive) form.
- **Query expansion:** Adding synonyms of keywords to the original query to improve passage retrieval performance. An example that shows the addition of synonym to a query, consider the keyword “كأس العالم” (the World Cup) which could be used in another document as the keyword “مونديال”, (borrowed from the French “Mondial”).

3.2.1.2 Document Processing

- **Identify relevant documents:** Using the set of keywords and their synonyms, the knowledge base is searched to get the documents that contain the set

of keywords or some of them.

- *Ranking relevant documents:* The set of documents received are then ranked according to their degree of relevancy to the set of keywords, to get the top N documents that can be searched more intensively to get the relevant passages in the passage retrieval phase.
- *Passage retrieval:* From the top N ranked documents, the system searches for paragraph-sized passages that are more relevant to the query terms.

3.2.1.3 Answer extraction

- *Named entity recognition (NER):* The candidate answers are looked for as named entities. NER is the process of extracting named entities and tag them with a predefined category, such as name, organization, number.[8]
- *Answer scoring and ranking:* The retrieved candidate answers are scored and ranked to get the most relevant answer.
- *Answer presentation:* For factoid questions, the answer may be presented as a named entity. While others like those for definition questions, can be presented as a sentence or even a paragraph.

3.2.2 QA general architecture on linked data

The main tasks in QAS based on structured data are Data preprocessing, Question Analysis, Data Matching, Query construction, Scoring, and Answer retrieval and assessment. The history of Question answering on structured data goes back to 1960s 1970s such as Baseball (1961) and Lunar (1977) [14]. The rest of this section presents an architecture in Fig. 3 with a deduced explanation from [11]. *Data Preprocessing:* The process of preprocessing the information

in the datasets, can be indexing, data cleaning, and feature extraction.

- *Question Analysis:* It is usually the first step in any question answering system. Where a linguistic analysis is carried up that can be syntactic or semantic analysis. It aims also for detection and extraction of question features as well. Main tools that are used for this purpose are the POS taggers, stemmers, and parsers. Also some resources are used such as WordNet and Wiktionary. These techniques all aim mainly at detecting the question type, question focus and expected answer type.
- *Data Matching:* It is the component that deals with the differences between the question vocabularies and the dataset elements. It can be simple term lookup to more complex semantic matching.
- *Query Construction:* The process of transforming the input question to a query according to the results of the question analysis and data matching processes.
- *Scoring:* A number of candidate answers are retrieved and need to be evaluated. Scoring can use different techniques such as similarity measures, frequency of data elements.
- *Answer Retrieval and assessment:* By executing the constructed query across the database, the system extracts the answer that may need checking for its type if fits the expected answer type.
- *Answer Representation:* Finally, the answer is processed to be represented in natural language format rather than in links or triples from linked data.

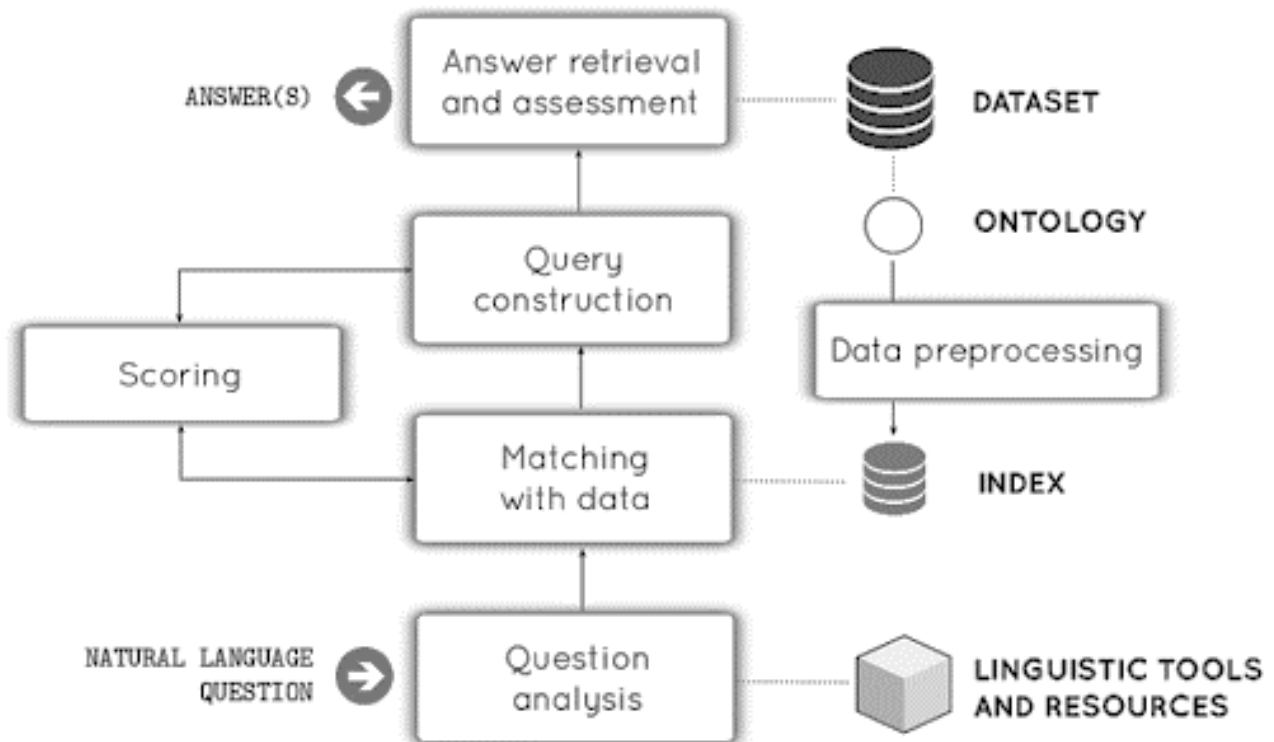


Fig 3: A general Architecture of QAS on Linked Data

4. CHALLENGES OF ARABIC QAS

Arabic is a Semitic language [9]. Its population is over 422 million people in the world. It is the first language of the Arab League and an official language in the United Nations [15]. It is considered as the third international important language that comes after English and French [9]. Arabic is a ritual language of 1.6 billion Muslims in the world [16]. It is the language which Holy Quran and Sunnah was written by [9].

Arabic has a very rich syntax with some particular features that are hard for a computer to be able to process them [17]. This richness resulted in a number of challenges that needed to be treated by researchers in a different way [4]. This section addresses a number of these challenges.

4.1 No Capital Letters

A main process in factoid question answering, is Named Entity Recognition, where named entities like proper names or names of locations and persons, is detected using a named entity recognizer [8]. This process is much easier in Latin languages whose named entities begin by capital letters, than in Arabic that has no capital letters and so cannot use capitalization as an orthographic feature to differentiate between Arabic proper nouns than other word forms such as adjectives and common nouns. For example, the Arabic proper name محسن (Mohsen) can be used as a person name, a noun (improver), or adjective (improving) [12] [18].

4.2 Optional short vowels

Short vowels in written Arabic are represented by diacritic signs. These diacritics has an impact on the same orthographic form of the Arabic word phonetic representation and its meaning. The Arabic modern version can be written without diacritics, and it can be understood by Arabs but creates diverse types of orthographic ambiguity. For example, the sentence “من أجمل البنات” can be diacriticed in two ways, the first is a question “who is the prettiest girl?” and the second as a part of a sentence that can be the answer to the previous question “..... is among the most pretty girls.” [12]

4.3 Free word order

Arabic has a sort of free word order for a sentence with the same meaning. For instance, the sentence “Ahmed wrote an article” can be written in three different word orders: 1) “احمد اكتب مقالا” Subject-Verb-Object, 2) “كتب احمد مقالا” Verb-Subject-Object. This lead to more implications that need a deeper level of question analysis.

4.4 Derivational nature

Arabic is a highly derivational language that its lexicon is built from 10,000 roots, each root is made of 3 to 4 letters. And to each root affixes (prefix, infix, or suffix) can be added to form lemmas. Affixes are added to the root according to 120 patterns. So Lemma = Root + Pattern [4]. For example, the two lemmas “كاتب” (writer) and “اكتب” (you write) which takes the pattern “فاعل” for the former and “أفعل” for the latter. Those lemmas are derived from the word “كتب” (wrote), a three-letter root which has a pattern “فعل”, by adding the infix “ا” to the former and the prefix “ا” for the latter. Adding to that, its agglutinative nature, that made it highly inflectional [12]. So, a word = root + affixes (prefix, infix, or suffix) [4]. All of that, causes a huge lexical variation that is very systematic but complicated morphology. This leads to data sparseness that needs morphological preprocessing or massive training corpora [4].

4.5 Lack of linguistic resources

There is a need for corpora with different types of questions and tagged documents, to train and test question answering systems. These corpora should be admitted by the research community. A lot of this sort of needed linguistic resources are made for English, but for Arabic, it still in its beginning stage, where the available resources are few and has a limited capability. Thus researchers always make their own corpora that need to be annotated and verified manually [12].

5. OVERVIEW OF STATE-OF-THE-ART SYSTEMS

A lot of work done within the English and other Latin based question answering research field. In fact, it starts from 90's. But for Arabic, it is still in its beginning era due to Arabic language specific challenges. Most of the work done in the Arabic Question Answering field, experienced over fitting and unrealistic evaluation. This is because of the lack of Arabic language resources, corpora and test-beds [4].

In 1993, Mohammed et al. [19] developed AQAS (Arabic Question Answering System). AQAS is a Question Answering System, which is based on a structured knowledge base. It was designed to search for answers found in structured data only, but failed to extract an answer from unstructured data. No evaluation reported for this system.

In 2002, QARAB was developed that used unstructured Arabic documents of Al-Raya newspaper in Qatar [4]. QARAB was the first Arabic QAS to use Natural Language Processing techniques for extracting answers. QARAB considered the question as a bag of words, without any attempt to process the question semantically [12]. It used a morphological analyzer called AraMorph, to tokenize and normalize the question and POS tagging to extract the question keywords that are used to retrieve the answer [20]. An Information Retrieval system was used to retrieve the top 10 most relevant passages from the knowledge base. The system reported 97.3 % Precision and Recall. It was evaluated by four native Arabic speakers on a set of 113 questions [12] [19].

Then, in 2007, Benajiba introduced the ArabiQA system that extracted Arabic passages using Arabic-JIRS [21]. The Arabic Named Entity Recognition ANERsys was used to extract named entities that represent the answers from the relevant retrieved passages. The system was evaluated across a test set of 200 questions from CLEF 2006 and 11638 documents from the Arabic Wikipedia. The authors reported a Precision of 83.3%. The system answered factoid questions only. It has no deep question processing that it did not get the predicted answer type nor use a query expansion to improve the Recall of the system [12].

In 2009, QASAL was an Arabic QAS for factoid and definition questions. It applied the linguistic environment Nooj to extract the question focus by identifying the question pattern, and extracted the answer by using lexical patterns. Authors reported a precision of 93% and recall of 100% over a set of 43 definition questions across the web [12].

In 2010, Trigui et al. developed DefArabicQA. DefArabicQA is an Arabic Question Answering System that targets definitional questions. The candidate definitions are identified according to some lexical patterns that was designed manually. Then they used some heuristic rules to categorize these candidate definitions. After categorizing the candidate definitions, they used a statistical approach to rank them. The

system evaluated over 2000 snippets from Google, Arabic Wikipedia, and 50 definition questions. The authors reported MRR of 0.81 [12].

In 2011, Bekhti and Al-Harbi developed AQuASys. AQuASys is a factoid based Arabic QA system. It used extensive NLP techniques in the question processing phase [19]. The system reported a Recall of 97% and Precision of 66.25%. It was evaluated across a set of 80 questions and documents from ANERgazet and ANERCorp. [12]

In 2014, Abdelnasser et al. developed Al-Bayan. Al-Bayan is An Ontology based Arabic Question Answering systems for the Holy Quran. Questions asked are about Quranic verses and the answers are retrieved associated with its interpretation from Ibn Kathir corpus. The system used the Quran ontology presented by Atwell et al. [20]. The system achieved an accuracy of 5% using the top three results [19].

JAWEB is a web based Arabic Question Answering system. It used NLP methods in question processing and named entity recognition. The authors evaluated the system with a Recall of 100% and a precision of 80%. It reported a 15-20% higher Recall when compared with ask.com [19].

6. DO IT YOURSELF

This section contains some of the recommended tools for building an Arabic QAS. These tools and resources are well-known in the field and are tried to get good results.

Tool	Link
Text Processing Tools	
<i>Text Segmentation Tools:</i> are tools that identify the boundaries between linguistic meaningful units of text such as words or phrases.	
Stanford Word Segmenter	http://nlp.stanford.edu/software/segmenter.shtml
	http://sergeytihon.github.io/Stanford.NLP.NET/StanfordWordSegmenter.html
<i>Word tagging and text annotation tools:</i> are tools that annotates words in a piece of text with parts of speech, such as verb, noun, adjective... etc.	
Stanford Log-linear POS Tagger	http://nlp.stanford.edu/software/tagger.shtml#Download
<i>Word analysis tools:</i> are tools that process words to their primary linguistic forms to produce linguistic features.[12]	
Khoja Stemmer	http://zeus.cs.pacificu.edu/shereen/research.htm
ISRI	https://nltk.googlecode.com/svn/trunk/doc/a

Stemmer	pi/nltk.stem.isri.ISRIStemmer-class.html
MADAMI RA	http://innovation.columbia.edu/technologies/cu14012_arabic-language-disambiguation-for-natural-language-processing-applications
Farasa	http://qatsdemo.cloudapp.net/farasa/
AlKhalil Analyzer	http://oujda-nlp-team.net/?p=1299&lang=en
AraMorph	http://www.nongnu.org/aramorph/english/principles.html
Stanford CoreNLP	http://nlp.stanford.edu/software/corenlp.shtml#Demo
AraNLP	https://sites.google.com/site/mahajalthobaiti/resources
<i>Syntax analysis tools:</i> are tools that analyze the text sentences and detect its syntactic structure and validate it.[12]	
Stanford Parser	http://nlp.stanford.edu/software/lexparser.shtml
Information retrieval tool	
Lucene	http://lucene.apache.org/core/
JIRS	http://sourceforge.net/projects/jirs/
Linguistic development environment tools	
NooJ	http://www.nooj4nlp.net/pages/download.html
GATE	http://gate.ac.uk/
SAFAR	http://arabic.emi.ac.ma/safar/
FassiehR	N/A
Linguistic Resources	
Knowledge-based resources	

Arabic Wikipedia	http://ar.wikipedia.org/wiki
Arabic WordNet	http://globalwordnet.org/arabic-wordnet/awnbrowser/#BrowserDownload .
DBpedia	http://dbpedia.org/About
Linguistic corpora: for training and testing questions	
Arabic Stopwords	http://www.ranks.nl/stopwords/arabic
ANERcorp	http://www1.ccls.columbia.edu/~ybenajiba/downloads.html
Penn Arabic Treebank (PATB)	https://catalog.ldc.upenn.edu/LDC2010T08
Extensible Frameworks	
QANUS	http://www.qanus.com/

7. EVALUATION

The process of conducting this survey involved at first searching for proper papers to cite. For this goal, there was a need to define the search keywords precisely. Those keywords were Question Answering Systems, Question Answering Systems survey, Question Answering Systems review, types of Question Answering Systems, Arabic Question Answering, Arabic Question Answering Systems, Challenges of Arabic Language on QAS, Arabic language tools and techniques used in QAS, Arabic QAS survey, Arabic QAS review, simple QAS, QAS on linked data, ontology based QAS, ontology based Arabic QAS, and Arabic QAS on linked data.

According to these keyword search hundreds of papers were retrieved. But because this survey aims at helping new Arabic researchers in the field to get the big picture about what the question answering is and how to build it in Arabic, there was a need to include only the English and Arabic papers that explained the question answering processes as well as sub processes, and its architecture in details. Also the papers that talk about different types of QASs in English and how to apply it in other Semitic languages. Surveys about the QASs in both English and Arabic were included as well. We excluded papers that were deeply specialized like papers that describe the building process of a stemmer or a parser. A lot of old papers and surveys we also excluded so that the survey would be updated as possible.

8. CONCLUSION

A question answering system is the system that processes a human natural language question and retrieves the related answer from an open or closed domain. This survey attempts to serve as a simple as well as comprehensive tutorial of building an Arabic Question Answering System on structured

and unstructured data. It began by defining the different types of Question Answering Systems, presenting the general architecture for QAS on structured as well as unstructured data. Then an illustration of some of the Arabic language challenges to be taken into consideration when building question answering systems, the impacts they have upon the Question Answering building process. At the end it collected and classified the Arabic language tools and techniques needed for each stage of building the QAS on both structured and unstructured data. Arabic question answering that are based on unstructured data (i.e. plain text) are more studied rather than those based on linked data. This was the motivation point to propose a new architecture for Arabic Question Answering System that is ontology based.

9. REFERENCES

- [1] Unger, Christina, et al. "An Introduction to Question Answering over Linked Data." Reasoning Web (2014).
- [2] Al Chalabi, Hani. "Question Processing for Arabic Question Answering System" <http://bsp.space.buid.ac.ae/bitstream/1234/760/1/120149.pdf>
- [3] Pundge, Ajitkumar et al. "Question Answering System, Approaches and Techniques: A Review" International Journal of Computer Applications 141 (2016): 0975-8887
- [4] Ezzeldina, Amed Magdy and Shaheen, Mohamed. "Survey of Arabic Question Answering: Challenges, Tasks, Approaches, Tools, And Future Trends" ACIT (2012): 1812-0857
- [5] Dahab, Mohamed Y. et al. "A Comparative Study on Arabic Stemmers" International Journal of Computer Applications 125 (2015): 0975 – 8887
- [6] K. Ray, Santosh and Shaalan, Khaled. "A Review and Future Perspectives of Arabic Question Answering Systems" IEEE Transactions on Knowledge and Data Engineering 28 (2016): 3169 - 3190
- [7] Brini, Wissal et al. "Factoid and definitional Arabic Question Answering system" NooJ 10 (2010)
- [8] Shaalan, Khaled. "A Survey of Arabic Named Entity Recognition and Classification" Computational Linguistics. 40 (2014): 469-51
- [9] Gupta, Poonam and Gupta, Vishal. "A Survey of Text Question Answering Techniques" International Journal of Computer Applications. 53 (2012): 0975 – 8887
- [10] Refaee, Eshrag and Rieser, Verena. "Subjectivity and Sentiment Analysis of Arabic Twitter Feeds with Limited Resources" Workshop on Free/Open-Source Arabic Corpora and Corpora Processing Tools Workshop Programme. (2014)
- [11] Hamdelsayed, MA and Atwell, ES "Islamic Applications of Automatic Question-Answering" Journal of Engineering and Computer Science, 17 (2016).
- [12] Al-Shargi , Faisal and Rambow, Owen. "DIWAN: A Dialectal Word Annotation Tool for Arabic" (2015).
- [13] N. Bdour, Wafa and Gharaibeh, Natheer K. "Development of Yes/No Arabic Question Answering System" IJAIA 4 (2013).
- [14] Badia, Antonio. "Question answering and database

- querying: Bridging the gap with generalized quantification” *Journal of Applied Logic*. 5 (2007).
- [15] W. AbuTaha, Alaa. “An Ontology-Based Arabic Question Answering System” <http://library.iugaza.edu.ps/thesis/115181.pdf>
- [16] Sheker, Mustefa. “Domain-Specific Ontology-Based Approach For Arabic Question Answering” *JATIT* 83 (2015): 1992-8645 .
- [17] Al-Chalabi, Hani et al. “Semantic based Query Expansion for Arabic Question Answering Systems” *First International Conference on Arabic Computational Linguistics*. (2015).
- [18] Bakari, Wided et al. “Researches and Reviews in Arabic Question Answering: principal approaches and systems with classification” *ACIT*. (2016).
- [19] Al-Chalabi, Hani et al. “Question Classification for Arabic Question Answering Systems” *ICTRC*. (2015).
- [20] Radev, Dragomir et al. “Probabilistic Question Answering on the Web,” *J. Am. Soc. Inf. Sci. Technol.* 56 (2005): 571–583.
- [21] Diekema, Anne et al. “Question Answering: CNLP at the TREC-9 Question Answering Track” *TREC-9*. (2000).
- [22] Yao, Xuchen and Durme, Benjamin. “Information Extraction over Structured Data: Question Answering with Freebase” *ACL*. (2014).
- [23] Watanabe, Yusuke et al. “Question Answering from Unstructured Text by Retrieval and Comprehension” *cs.CL*. (2015).
- [24] Bakari, Wided et al. “Towards Logical Inference for Arabic Question-Answering” *International Conference on Intelligent Text Processing and Computational Linguistics*. (2015).
- [25] Haoda Huang, Benyu Zhang “Text Segmentation” *Encyclopedia of Database Systems* (2009).