ISSN (e): 2250 – 3005 || Volume, 06 || Issue, 12|| December – 2016 ||
International Journal of Computational Engineering Research (IJCER)

Answer Extraction for how and why Questions in Question Answering Systems

Waheeb Ahmed[1], Dr.BabuAnto P[2]
[1]Research Scholar, Department of Information Technology, Kannur University, Kerala, India,
[2]Associate Professor, Department of Information Technology, Kannur University, Kerala, India

ABSTRACT
With the increasing amount of Arabic text on the web and in the information repositories and the demand of users to have specific answers to their questions, the need for Question Answering (QA) Systems became a necessity. Our Question Answering System answers two types of Questions: How and Why Questions. The system takes a question given in natural language expressed in the Arabic language and attempts to produce concise answers. The system's main source of knowledge is a collection of Arabic text documents extracted from the Arabic Wikipedia. The reasons behind developing this system is due to the absence of Arabic Questions Answering Systems(QASs) which deals with How and Why questions and this is because of the complexity of extracting the answers that satisfy this type of questions. Information Retrieval (IR) module is used to retrieve the target document from the corpus. The IR is coupled with Natural Language (NLP) Tools to process the given question and to extract the answer. The major goal of the proposed system is to extract the passage which is likely to contain the answer based on the semantic similarity between question keywords and the sentences of the passage. We used Precision, Recall and F1 Measure to calculate the accuracy of the system.

Keywords:*Answer Extraction, Artificial Intelligence, Information Retrieval, Information Extraction, Natural Language Computing,Question Answering System, Question Analysis.*

I. INTRODUCTION

Question Answering is a popular application of Natural language processing. It is concerned with building systems that accepts questions given in natural language by humans and tries to produce the required answer. This field is emerged due to the high demand for systems that accept a question from user in natural language rather than a set of keywords and consequently supply a concise answer. Traditional search engines like Google and Yahoo usually return a list of links [1]. However, they do not give specific answers to users. It is the task of the user to look for the answer in these links by browsing them and searching for it and this may consume a considerable amount of time. Recently, both of the information growth and the high demand for an efficient access to information has increased the motivation of research in QASs[2].

1.1 Categories of Questions

The research in QA deals with a variety of questions including:

- Factual: Questions that ask for factual information [who, what, where, when].This type of questions require a short answer in the form of a single word or phrase. e.g. "Who invented the Piano?"(*من اخترع البيانو؟*)
- Definition: Questions that looks for definition of a term. e.g."What is Geoinformatics?"(*ما هي نظم المعلومات الجغرافية؟*)
- Listing: Questions that requirelists of facts or entities. e.g. "List the action movies of 2016?"(*اذكر أفلام الأكشن لعام 2016؟*)
- Causal questions[why,how]: Questions that seek for explanations about an entity.e.g. "How can we measure the speed of light?"(*كيف نقيس سرعة لضوء؟*)
- Yes/No questions: Questions that require a yes/no answer. e.g. "Does the water have color?"(*هل للماء لون؟*)

QASs are classified into two domains depending on the source of information from which the QA returns the answer: open domain and closed domain. Open domain QASs return the answer from the web and they are not restricted to a specific field of knowledge. In contrary, closed domain QASs retrieves the answer from a database or knowledge base which is limited to a specific field or area like Medicine, Biology, Weather forecasting etc. Many QAs has been developed for answering factoid questions like who, what, where and

when. However, questions like how and why that need descriptive answers need complex processing. Answering How and Why questions is considered hard since these questions may need long answers.

1.2 Arabic Language Challenges

There are several challenges posed by the Arabic language which makes Arabic language processing a hard task[3][4]:

- Morphological complexity
- Lack of basic NLP tools for processing the language like (morphological analyzers, information extraction tools) and lack of other linguistic resources like specialized dictionaries,corpora,lexicon etc.
- Highly inflectional and highly derivational. This means the same context may appear in several forms, which impose the need for a huge corpus in order to get a representative frequency of all the forms in which a context might appear or to make a solution to minimize the number of these forms into a smaller one.
- The direction of writing is from Right-To-Left and a group of its letters change their forms according to their position/appearance in the word.

Ambiguity where the same word has different meanings.Lack of capitalization that makes it difficult to extract named entities.The above challenges slowed down the development of Arabic QASs especially for questions which requires explanations as answers like How and Why questions.

II. RELATED WORK

AQAS is knowledge-based system which returns answers from structured data but not from plain text (unstructured text). AQAS tries to answer simple factoid questions like Who, What, Where and When[5];Besides that no results for their system are reported. QARAB is a closed domain simple factoid question answering that answers questions like Who, Whom, When, What, Where but it does not address How and Why questions and the corpus consists of documents which are extracted from a newspaper called the Al-Raya published in Qatar[6].QASAL is a QA system for Arabic language for answering factoid questions. It is built on the NooJ platform[7], and no experimental results or performance has been published for this system [8].Bdour and Gharaibeh developed a system for Yes/No questions only [9].Our proposed work concentrates onprocessing and answering causal questions [How(كيف), Why(لماذا)] for Arabic language.

III. METHODOLOGY

We used natural language tools for processing the question and IR module using the term frequency-inverse document frequency(tf-idf) weighing for retrieving the relevant documents from the corpus. Our corpus consists of 500 documents extracted from the Arabic Wikipedia. The question set consists of 80 questions which is divided into two sets: one set consist of 40 How questions and the other set consists of 40 Why questions. The user will supply a question in Natural Language to the QA system. The QAS will process the question and deliver the answer. The following steps are performed to analyze the given question and retrieve the candidate answer:

1. Question Analysis.
2. Question Expansion.
3. Document Retrieval.
4. Answer Extraction.

3.1 Question Analysis

The question analysis phase consists of three steps:

1. Question classification.
2. Tokenization
3. Identification of Question Focus.

Question Classification: Question Classification seeks identifying what the question is looking for. If a question starts with Why(لماذا), then the question is classified as REASON. That is, the question is looking for reason. For example, (لماذا تبدوا السماء زرقاء أثناء النهار؟) "Why does the sky look blue during day?"

The question is classified as REASON. If the question starts with How(كيف), it is classified as MANNER. That is, the question is seeking an answer of type MANNER. The main purpose of classifying the question is that this information(Question Class either MANNER or REASON) will be sent to the Answer Extraction(AE) module to extract the proper answer from the retrieved document.

Tokenization: The question is tokenized into individual tokens and these tokens are stored in a list. Stop-words are removed. Stop-words are words that appears very frequently and have less important meaning like prepositions and conjunctions(in, from, to, about, on , and, or)(من، الى ، عن ، على ، و ، أو).These words are removed from the question. After that, a chunker is used to get the named entities and noun phrases. For

example: **"**Why did the Egyptian scientist "Ahmed Zewail" become famous?(" *لماذا أصبح العالم المصري أحمد زويل مشهورا؟***").** We have developed a simple rule-based the named entities based on the output of Stanford Part-Of-Speech (POS) Tagger for Arabic language. The chunker will extract "Ahmed Zewail"(*أحمد زويل***)** as a named entity.The list of keywords after tokenization and chunking ["Ahmed Zewail", "Egyptian", "scientist", "become", "famous"]. That is, ["*أحمد زويل*"**,**"*المصري*"**,**"*العالم*"**, "***أصبح***", "***مشهورا***"**].**

Identification of Question Focus: Question focus is a word or a phrase extracted from the question that helps in identifying the type of the expected answer. The question class along with the question focus will benefit the AE module in ranking the candidate answers. For example, the question (*لماذا منح نجيب محفوظ جائزة نوبل في الأدب 1988***)"**Why was Naguib Mahfouz awarded the Noble Prize in Literature 1988?". The focus of this question is looking for something related to "Naguib Mahfouz". The focus here is the Noun Phrase(NP) "the Noble Prize in Literature"(*جائزة نوبل في الأدب***)** and this is done using the chunker. The answer type in figure-1 is the defined by the combination of the question classification and the question focus.

The flow of our QA system is shown in the following figure:

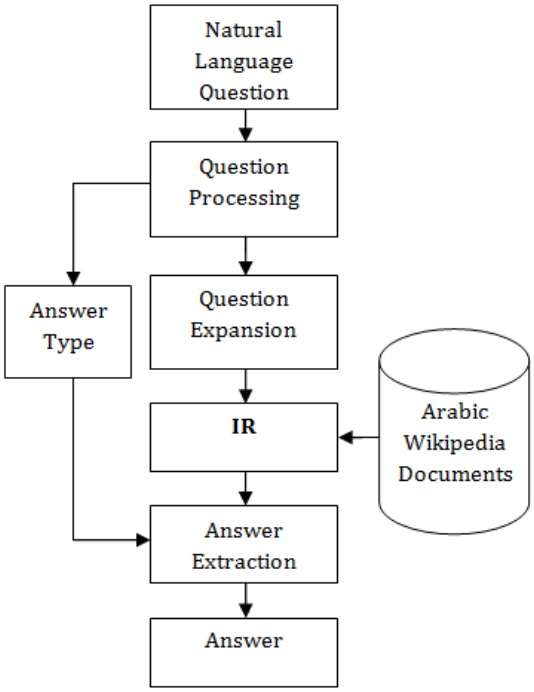

Figure1.QA Architecture

3.2 Question Expansion

In question expansion alternative synonyms for some keywords in the question(verbs and adjectives) are used. We used Arabic WordNet(AWN)[10] (available as open source software) to extract the synonyms for the verbs and adjectives in the question. The reason for question expansion is that the same verb/adjective in the question may not be available in the answer. So, we have to expand the question by adding synonyms for some words in the question. These synonyms are fed into the list of question terms that will be sent to the IR module and this will increase the chance of getting the answer. For example, (*لماذا تغني الطيور؟*) "Why do birds sing?" The synonyms for (*تُغني*sing) include (*تُغرد, تُبلبل*) are added to the question keywords list.

3.3 Documents Retrieval

We used Vector Space Model for developing our IR module for retrieving the relevant documents from ArabicWikipedia corpus. Vector Space Model is an algebraic model that represents query strings and text documents as vectors [11]. After getting the available named entities and the noun phrases and other keywords extracted from the question, these extracted keywords are received by the IR module which search for them in the index to retrieve the relevant document which contains all or most of the question keywords.

3.4 Answer Extraction

Our proposed method for extracting the answer from the top ranked document retrieved by the IR module is implemented in the following procedures:

1. If the question class is REASON. The keywords [(because, due to , reason)*لأنه,بسبب,لهذا,لذلك*] are added to the list of question keywords. If the question class is MANNER, the keywords [(by, using) *عن,بواسطة,باستخدام طريق*] are added to the list of question keywords.
2. The top ranked document which is retrieved by the IR module is divided into passages at the discourse level.
3. Passage which contains the question focus is given weight=1 and passages that do not contain the question focus is given weight=0.
4. Cosine similarity between the question and every sentence in the passage is calculated using the following formula:

$$S(q,s) = \frac{A}{\sqrt{B}\sqrt{C}}$$

A=Sum($q_i s_i$), B=Sum(q_i^2) , C=Sum(s_i^2)

Where,

q_i is representing the tf-idf of the term i in the question.

s_i is the tf-idf of the term i in the sentence.

5. Total similarity between the question and every sentence S in the passage p is calculated by $S(p)=S_1+S_2+\ldots+S_n+$weight
6. S(p) is calculated using the equation in step 4 for all passages.
7. The passage with the highest S(p) score is extracted as answer and presented to the user.

IV. RESULTS AND PERFORMANCE EVALUATION

There are many evaluation metrics that are used for evaluating question QA systems. The following metrics are used inText Retrieval Conference(TREC-8) project: Precision, Recall and F-measure. Where,

$$\text{Precision}=\frac{\text{Number of correctly answered questions}}{\text{Total number of system responses}}$$

$$\text{Recall}=\frac{\text{Number of correct answers}}{\text{Total number of questions to be answered}}.$$

F measure is the combination of the precision and recall with equal weight given to both of them:

$$\text{F1 measure}=\frac{2*(\text{Precision}*\text{recall})}{(\text{Precision}+\text{recall})}$$ [12].

The above measures are the common measures used for evaluating any QA system including TREC project series and many other question answering systems on different languages in the literature.

Table 1. Experiment results for our QAS

QUESTION TYPE	PRECISION(%)	RECALL(%)	F-MEASURE(%)
How(كيف)	61	52	56
Why(لماذا)	67	62	64

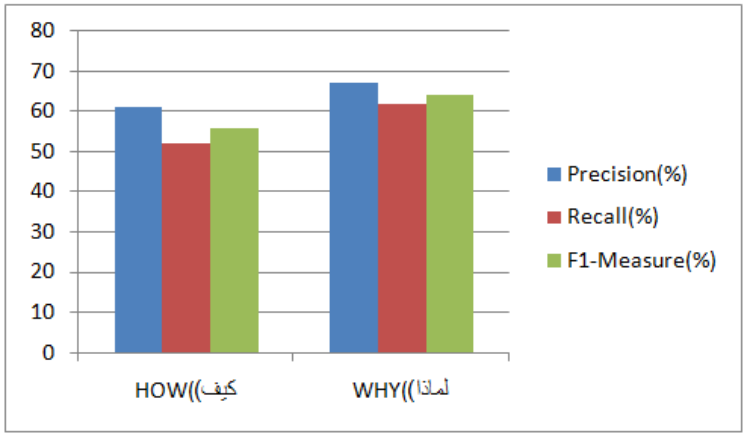

Figure 2. Distribution of accuracy of the QAS for HOW & WHY Questions

The obtained Precision of the system for total 40 How questions is 61% and the Recall is 52%. The F1 measure is 56%.For the total 40 Why questions the obtained precision is 67% and the Recall is 62%. The F1 measure is 64%. The performance of the QAS for answering the Why questions was 64% which is higher than the result got for the How questions by 8%. The result is promising and it is the first system that deals with Arabic How & Why questions comparing to the literature on Arabic QASs[5][6][8][9].

V. CONCLUSION

Our QAS attempts to answer Arabic Why and How) questions. The proposed system uses NLP tools for question analysis and IR for document retrieval. The process of retrieving the candidate passage which is likely to contain the answer is done by computing the similarity between the How/Why question and the sentences in all the passages in the retrieved document. Passage with the highest score is extracted and presented to the user. This system is the first attempt to answer complex how & why questions. As a future work more features will be used to increase the system accuracy.

REFERENCES

[1] P. Rosso, Y. Benajiba and A. Lyhyaoui , "Towards an Arabic Question Answering system," In Proc. of the 4th Conference on Scientific Research Outlook & Technology Development in the Arab world, pages. 11-14, Dec. 2006.
[2] J. Burger et. al,"Issues, Tasks, and Program Structures to roadmap research in question & answering," In Document Understanding Conferences Roadmapping Documents, pages. 1-35, Jan. 2001.
[3] W. Brini , M. Ellouze , S. Mesfar , and L. Belguith,"An Arabic Question Answering System for Factoid Questions," In Proc. of IEEE International Conference on Natural Language Processing and Knowledge Engineering, pages.1-7, Sept. 2009.
[4] A. Bodor, A. Mohammed and M. Sherif, "Arabic Text Question Answering from an Answer Retrieval Point of View: a survey," International Journal of Advanced Computer Science and Applications, Vol. 7, No. 7, pages. 478-484, Jan. 2016.
[5] F. Mohammed, K. Nasser and H. Harb, "A Knowledge-Based Arabic Question Answering System (AQAS)," ACM SIGART Bulletin, pages. 21-33, Oct. 1993.
[6] B. Hammou, H. Abu-salem and S. Lytinen, "QARAB:QuestionAnswering System to support the Arabic Language," In Proceedings of the ACL-02 workshop on Computational approaches to semitic languages, ACL, pages. 1-11, Jan. 2002.
[7] Nooj web site: http://www.nooj4nlp.net-Last visited-September, 2016.
[8] K. Al-Daimi and M. Abdel-Amir,"The Syntactic Analysis of Arabic by Machine," Computers and Humanities, Springer,Vol. 28, No. 1, pages. 29-37, Jan. 1994.
[9] W. Bdour and N. Gharaibeh, "Development of Yes/No Arabic QA System," International Journal of Artificial Intelligence & Applications, Vol. 4, No. 1, pages. 51-63, Jan. 2013.
[10] Global WordNetwebsite:http://globalwordnet.org/ Arabic Wordnet-Last visited-September, 2016.
[11] G. Salton, A. Wong and C. Yang, "A vector space model for automatic indexing," In the Communications of the ACM, Vol. 18, No. 11, pages. 613-620, Nov. 1975.
[12] E. Voorhees, "The TREC-8 QA Track Report," In Proc. of the 8th Text Retrieval Conference (TREC-8) , Nov. 2000.