

DEVELOPMENT AND EVALUATION OF A WEB BASED QUESTION ANSWERING SYSTEM FOR ARABIC LANGUAGE

Heba Kurdi, Sara Alkhaider, Nada Alfaiifi

Department of Computer Science, Al Imam Muhammad Ibn Saud Islamic
University
Riyadh, SA

hakurdi@imamu.edu.sa, csce.t.sara@gmail.com,
n.alfaiifi@hotmail.com

ABSTRACT

Question Answering (QA) systems are gaining great importance due to the increasing amount of web content and the high demand for digital information that regular information retrieval techniques cannot satisfy. A question answering system enables users to have a natural language dialog with the machine, which is required for virtually all emerging online service systems on the Internet. The need for such systems is higher in the context of the Arabic language. This is because of the scarcity of Arabic QA systems, which can be attributed to the great challenges they present to the research community, including the particularities of Arabic, such as short vowels, absence of capital letters, complex morphology, etc. In this paper, we report the design and implementation of an Arabic web-based question answering system, which we called "JAWEB", the Arabic word for the verb "answer". Unlike all Arabic question-answering systems, JAWEB is a web-based application, so it can be accessed at any time and from anywhere. Evaluating JAWEB showed when compared to ask.com, the well-established web-based QA system, JAWEB provided 15-20% higher recall. These promising results give clear evidence that JAWEB has great potential as a QA platform and is much needed by Arabic-speaking Internet users across the world.

KEYWORDS

Question Answering system, Natural Language Processing , Arabic language tools.

1. INTRODUCTION

Question-answering is a challenging task in general. It involves state-of-the-art techniques in various fields, such as Information Retrieval (IR), Natural Language Processing (NLP), Artificial Intelligence (AI) and software technologies. The main goal of a question-answering system, is to give a precise answer to users' queries in natural language [1] [2]. This has the great advantage of helping users to get desired answers without searching large pools of information. Unlike search engines, such as Google and Yahoo, which allow a user to retrieve webpages or documents that are partially relevant to given keyword(s), while leaving the job of excluding irrelevant hyperlinks and locating desired passages to users, QA systems provide the user with highly relevant information in answer to their questions at the passage or sentence level [3]. QA systems categorize questions into three main groups: first, fact-seeking questions, i.e. factoid questions,

asking about (who, when, where, what and how much) which relate to different types of entities: person, location, organization, time and quantity. This type of question is the most common form of questions and therefore is the focus of this work. Second is list questions (e.g. List all the countries of the European Union), asking about multiple pieces of information with a common relation between them. Third is why-type questions (e.g. Why is the sky blue?) which seeks explanations of causes for a certain phenomenon or event. The second and third categories are less common in QA systems and require more advanced NLP and AI techniques and were therefore excluded from this study.

Based on the knowledge domain, QA systems classify questions as either closed-domain questions, which ask about information related to a specific domain (for example medical or sport, etc.), or open-domain questions which ask about information in any knowledge discipline. They require wider knowledge and advanced techniques to extract the most relevant answers [4]. The latter is the focus of this paper.

A QA system consists of three main components: a question-classification component, an information-retrieval component, and an answer-extraction component. The question-classification component plays a primary role as it categorizes questions based on their keywords: who, when, where, what and how much. The information-retrieval component searches for relevant answers in information sources by looking for matching keywords. Finally, the answer-extraction component selects the most relevant answers and ranks them accordingly. The quality of a QA system heavily depends on the effectiveness of this module as it makes the decision about the final list of answers to be presented to the user [5] [6].

Although the systems have common core components, they differ vastly in the techniques utilized by each component. This can be attributed mainly to the supported language(s) by the system, since each language has special features that affect how it should be analyzed, searched and retrieved.

Arabic is among the most popular languages with nearly 300 million speakers across the globe. It is a Semitic language. These are well known for their non-concatenative morphology, in which roots consist of an isolated set of constants, rather than syllables or words. This feature among many others has rated NLP and IR as grand challenging tasks for the Arabic language, despite some attempts such as [7]-[15]. Consequently, the huge Arabic content on the Internet is still underutilized [16]. Important emerging applications, question-answering systems for online services in particular, are simply not available in Arabic, despite their abundance in many other languages, especially English and other Latin-based languages. The challenging features of the Arabic language, scarcity of Arabic QA systems and lack of web-based Arabic QA systems are the main drivers for this paper. It adds several contributions to this important field, including comprehensively surveying Arabic QA systems and a comparing between them, developing and evaluating of JAWEB, the first Arabic web-based QA system, and extending the Arabic Corpus provided in [17].

The remaining of this paper is organized as follows: In section 2, the challenges of the Arabic language are highlighted. Section 3 reviews related work in QA systems. The system architecture and functional components of JAWEB are introduced in section 4 while the implementation process is described in section 5. In section 6, the experimental results are presented and discussed. Finally, section 7 concludes the paper and indicates future work.

2. CHALLENGES OF ARABIC LANGUAGE

Arabic is highly inflectional and derivational. This results in sparseness of terms in text, which leads to inefficiency in many statistical IR and NLP techniques. The absence of diacritics in Modern Standard Arabic also adds a lot of ambiguity to Question Analysis and Answer Extraction [16].

Many aspects are involved in the slow progress of Arabic NLP and IR, compared to the accomplishments in English and other languages. These aspects, as highlighted in [18], include:

1. Arabic has a complex morphology. It is highly derivational and inflectional, which extremely complicates morphology analysis.
 - Arabic is a derivational language, to find a word in an Arabic dictionary, we start by extracting the root and then we search for the root in the dictionary.
 - Arabic is an inflectional language, which means that the construction of a word involves finding the root and adding affixes (prefix, infix and suffix) to it.
2. The correct meaning of an Arabic word can be ambiguous in the absence of diacritical marks (short vowels); so a written word may hold different meanings.
3. Arabic does not use capital letters which makes differentiating between named entities and other words difficult.
4. Numerals are written from left to right, while alphabets are written from right to left. This makes editing Arabic text difficult when both numbers and letters are presented on the same line.
5. The most used encodings for Arabic text, UTF-8 and Unicode, present many problems when processing Arabic texts.
6. Lack of Arabic corpora, lexicon and electronic dictionaries.

3. RELATED WORK

Question-Answering systems present a good solution for textual information retrieval and knowledge sharing and discovery. This is why a large number of QA systems have been developed especially in English, which might be related to the language features and the maturity of research in the countries speaking it. The situation is less bright for the Arabic language, despite of its wide spread. Although research in the field of Arabic QA systems has already started [7]-[15], it is slow progressing and has limited results. Computerized tools and resources in general are lacking in Arabic [16], which has reflected negatively on the number of Arabic QA systems. The list includes:

- AQSA (1993) [19] seems to be the first QA system in the language. It is a knowledge-based QA system that extracts answers from structured data only. It uses the frames technique to present the knowledge from the radiation domain. However, no published evaluation is available for the system.
- QARAB (2004) [20] is a stand-alone (non-web-based) QA systems that uses IR and NLP techniques to extract answers from a collection of Arabic newspaper texts. It provides

answers to factoid questions but does not support other types of questions, such as how or why.

- ArabiQA (2007) [21] is a stand-alone Arabic QA system that deals with factoid questions, using Named Entity Recognition techniques and Java Information Retrieval System (JIRS) for Arabic text. It is designed specifically for Arabic factoid questions. However, system implementation has not been completed yet.
- QASAL (2009) [22] is a stand-alone Arabic QA system for factoid questions which uses the NooJ platform [24] as a linguistic development environment. The system takes advantage of some linguistic techniques from IR and NLP to process Arabic text documents containing answers to factoid questions. Published work based on this systems does not include experimental results or performance metrics. The overall functionality of the system is limited to the amount of available tools developed for the Arabic language by the NooJ Arabic modules.
- ArQA (2011) [18] is a stand-alone QA system that provides answers to factoid questions expressed in Arabic. This system has a pipeline architecture with four modules: question processing, passage retrieval, answer extraction and answer validation modules. Each module is the result of combining several IR and NLP techniques and tools to improve validity of returned answers.
- AQuASys (2011) [17]: a stand-alone Arabic QA system for factoid questions. It extensively utilizes NLP techniques to analyses questions and retrieves answers from an Arabic corpus that has been developed by the authors. Retrieved answers are scored and presented based on their relevance.

The survey above indicates clearly the lack of web-based QA systems, therefore JAWEB has been developed as a web-based application to bridge this gap providing ubiquitous accessed at any time and from anywhere.

4. JAWEB ARCHITECTURE

JAWEB is an Arabic web-based QA system that focuses on factoid questions. It accepts questions related to any named entity, including person, location, organization, time etc. After analyzing the question, important information is extracted to retrieve the most relevant answers from an Arabic corpus. The system is composed of four components: user interface, question analyzer, passage retrieval and answer extractor. The user interface runs at the client side while all other modules run at the server side. As illustrated in Figure 1, each component is structured of several modules with a distinct task carried by each.

User Interface

The user interface is implemented as a simple webpage with an input form that accepts Arabic factoid questions expressed in natural language. The question handler module validates entered questions and passes them to the server side. Later on, when the answers are retrieved from the corpus, the answer-viewer module formats and presents them in descending order of relevance; the desired answer is displayed at the top, followed by a list of candidate answers.

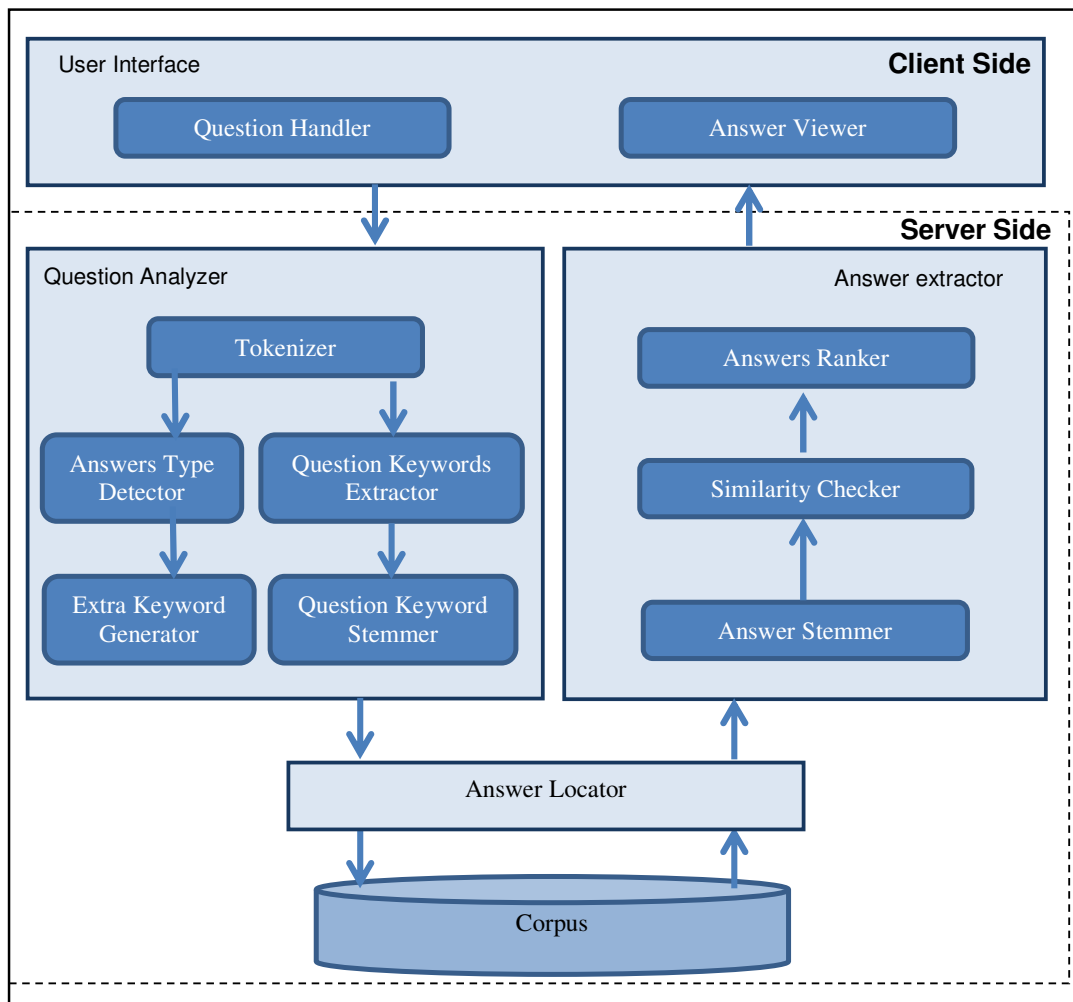


Figure 1: JAWEB system detailed architecture

Question-Analyzer

This component aims at identifying the type of the question. It has five modules: tokenizer, answer-type detector, question keyword extractor, extra keywords generator and question words stemmer. The user enters a question in a web-based form at the client side. The question is sent to the server side, where the tokenizer breaks the question into separate words. Afterwards, the answer-type detector identifies the category of the answer, which can be time, organization, person, location etc., based on the interrogative nouns (من who, ما what, أين where and متى when). The question keywords are the rest of the words in the question, after excluding stop-words and interrogative nouns; they are identified by the question keywords extractor module. The extra keywords generator produces synonyms of the keywords to expand the range of related answers. For example, for a question like: متى استقلت تونس (When has Tunisia become independent?) the extra keywords are (عام, year, سنة, month, etc.). Extracted keywords are stemmed, if necessary, so affixes are removed by the question keywords stemmer module to ease computing similarity with the keywords in the extracted answers later.

Answers Retriever

This component searches for potential answers and retrieves them from the corpus. It locates all sentences that contains a pattern that matches any word from the list of question keywords and extra keywords.

Answers Extractor

This component is responsible for selecting the most relevant answers from the potential answers list. It is composed of three modules. The first module is the answer keywords stemmer which gets roots of the keywords in the retrieved answer to ease checking similarity between a user question and potential answers, a task carried out by the second module, which is the answer-similarity checker. Similarity is measured by counting the number of matching keywords between the question and each retrieved answer. Based on this, the third module, answers ranker, sorts answers in descending order of their relevance. The most relevant answer is considered as the desired answer while the rest of answers are candidate answers.

5. SYSTEM IMPLEMENTATION

JAWEB has been built based on AQuASys [17] and as a natural extension to it. It provides a web interface to the system, an additional support for Arabic language presentation in web browsers, an extended corpus as well as an extensive evaluation and testing framework. The user interface was developed as a JSP webpage that accepts Arabic factoid questions. It provides a user-friendly interface where a user can type his/her questions. The question should begin with an interrogative noun: (متى, أين, من, ما, كم الكمية, كم العددية)

Dreamweaver, an Adobe proprietary web development application, was used to design the interface. The servlet and java classes on the server side were implemented using NetBeans IDE. The server was a GlassFish Server 3.1.2, which is an open-source application server project started by Sun Microsystems for the Java EE platform and now sponsored by Oracle Corporation. This server ran on 2.50 GHz CPU and an internal memory of 4 GB.

An extended version of the Arabic corpus developed by [17], containing 39,660 words with size of 457 KB, was used as the information pool to retrieve answers. The Arabic Khoja's stemmer [23] was adopted for keywords stemming. The system was tested thoroughly to insure correctness of obtained results and then an extensive evaluation have been carried out to compare JAWEB performance with a well established web-based QA system that supports Arabic Language, Ask.com [10], as described in the following section.

6. PERFORMANCE EVALUATION

The end objective of JAWEB is to develop a web-based QA system that answers Arabic factoid questions with high accuracy and speed. To evaluate this system, three performance measures were considered:

- *RECALL* [32]: the ratio of the number of relevant sentences retrieved to the total number of relevant sentences in the corpus. It is usually expressed as a percentage and calculated as:

$$RECALL = A / (A + B) * 100\% \quad (1)$$

Where *A*: number of retrieved relevant answers, *B*: number of relevant answer not retrieved

- *PRECISION* [32]: the ratio of the number of relevant sentences retrieved to the total number of sentences retrieved. It is usually expressed as a percentage and calculated as:

$$PRECISION = A/A+C * 100\% \quad (2)$$

Where A: number of retrieved relevant answers, C: number of retrieved irrelevant answers

- *RESPONSE TIME*: the time from when a user hits the search button until a response is displayed on the browser.

To have clear insight into JAWEB performance, we used a well-established QA system that supports the Arabic Language, Ask.com [10], as a benchmark. Ask.com is a famous web-based question answering system with multilingual support. It accepts colloquially-expressed questions and retrieves hyperlinks to webpages that contain similar keywords to those in the questions.

We ran several experiments with each of the supported type of factoid questions to get answers from Ask.com and JAWEB. Snapshots of answers to some example questions, listed in Table 1, are presented in Figure 2 – Figure 7. All figures illustrate that JAWEB has consistently given the right and most accurate answer as the first responses which can be attributed to the use of the stemmer (in both question and answer modules), the extra keywords generator and the similarity checker which has resulted in accurate ranking of retrieved answers. In contrast, Ask.com provided the correct answer in the second or third hyperlinks.

Comparisons between Ask.com and JAWEB in terms of *RECALL*, *PRECISION* and *RESPONSE TIME* are depicted in the graphs in Figure 8, Figure 9 and Figure 10, respectively. It is important to note that having no access to the knowledge base of Ask .com and in order to be able to calculate the *RECALL and PRECISION*, we have only considered the first five retrieved pages, which may slightly affect the accuracy of the experiments.

Figure 8 shows that *RECALL* of JAWEB was always higher than that of Ask.com and maintained at 100%, as the former has successfully retrieved all relevant answers from the corpus. The JAWEB *PRECISION* was admittedly less than ask.com, as shown in Figure 9, but it should be noted that it performed as well as the famous website, scoring over 90% in precision, despite the large difference in corpus sizes. In addition, even at its worst, for a question type that ask.com had the lowest precision too, JAWEB was not less precise than 70%. Regardless, the precision can easily be improved as the project launches, when feedback from users is attained and the corpus is grown in size and capacity. We believe that this is clear evidence that JAWEB has great potential as a QA platform and is much needed by Arabic-speaking Internet users across the world.

Naturally, ask.com was faster than our system, as illustrated in Figure 10, thanks to the use of high-power server CPUs for ask.com, as opposed to JAWEB 2.50 GHz. Again, if the project launches, the response time is expected to improve dramatically, when servers that are more powerful are provided.

Table 1: Experimental questions

No.	Type	Arabic	English
Q1	Who	من هو محمد طنجة؟	Who is Muhammad Tangier?
Q2	When	متى توحدت المملكة العربية السعودية؟	When was Kingdom of Saudi Arabia united?
Q3	What	ماهي الأهرامات المصرية؟	What are Egyptian pyramids?
Q4	Where	أين تقع المملكة العربية السعودية؟	Where is the Kingdom of Saudi Arabia located?
Q5	How much	كم تبلغ درجة حرارة القشرة الأرضية؟	How much is the temperature of the Earth's crust?
Q6	How many	كم عدد سكان الرياض؟	How many residents are there in Riyadh?



Figure 2.a: JAWEB answers for a who question

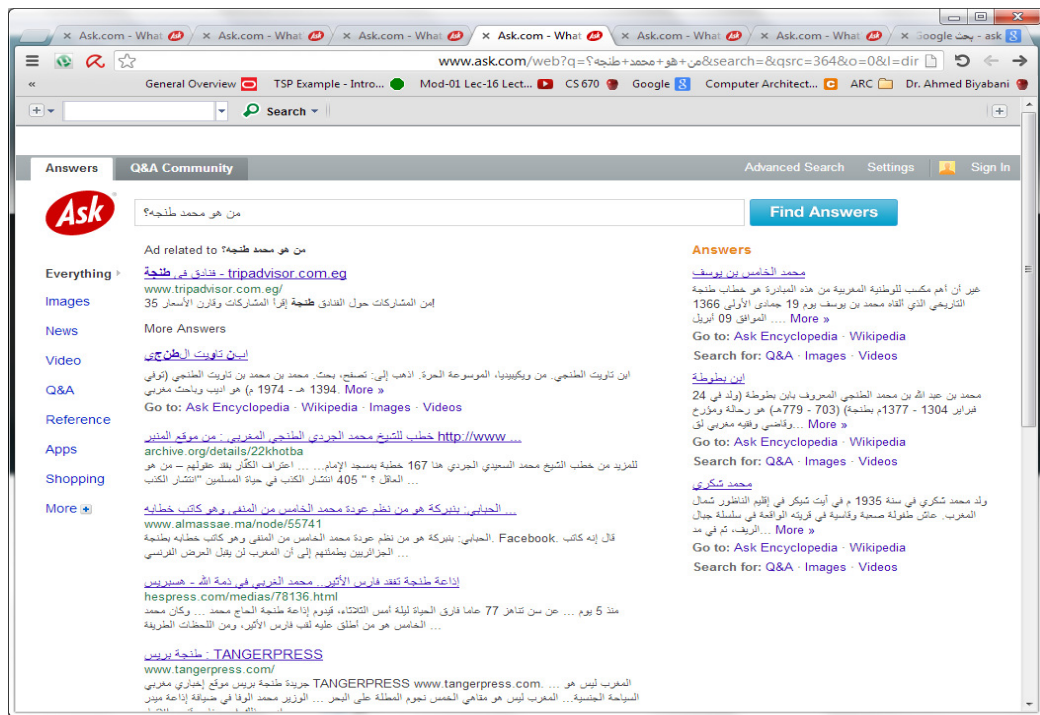


Figure 2.b: Ask answers for a who question

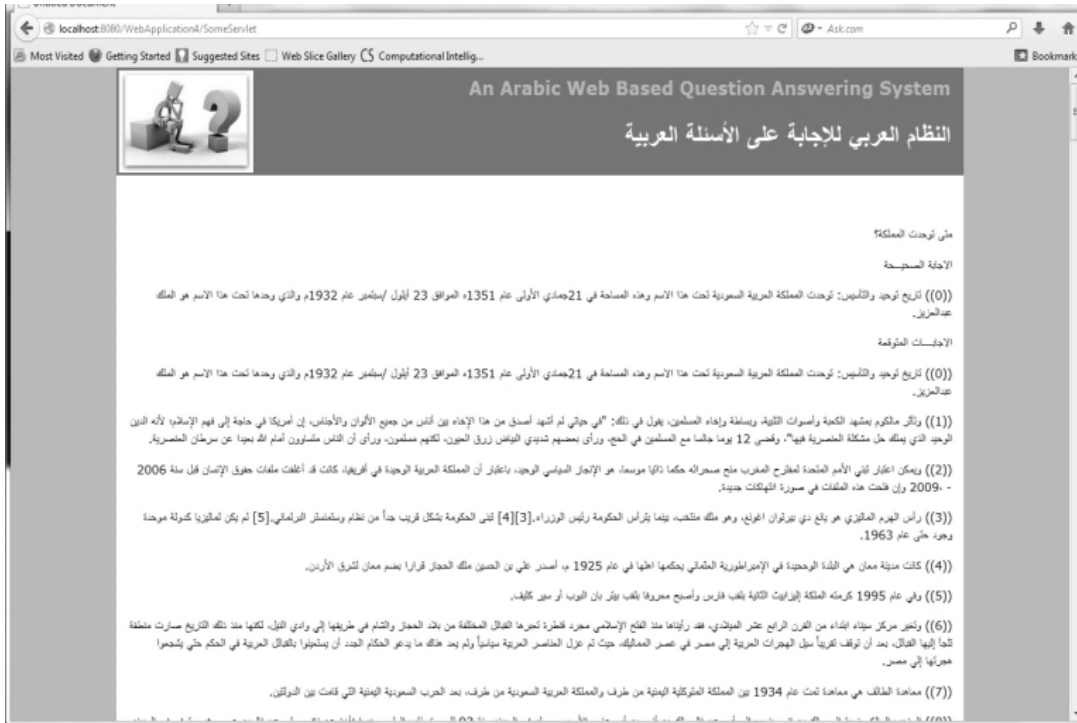


Figure 3.a: JAWEB answers for a when question

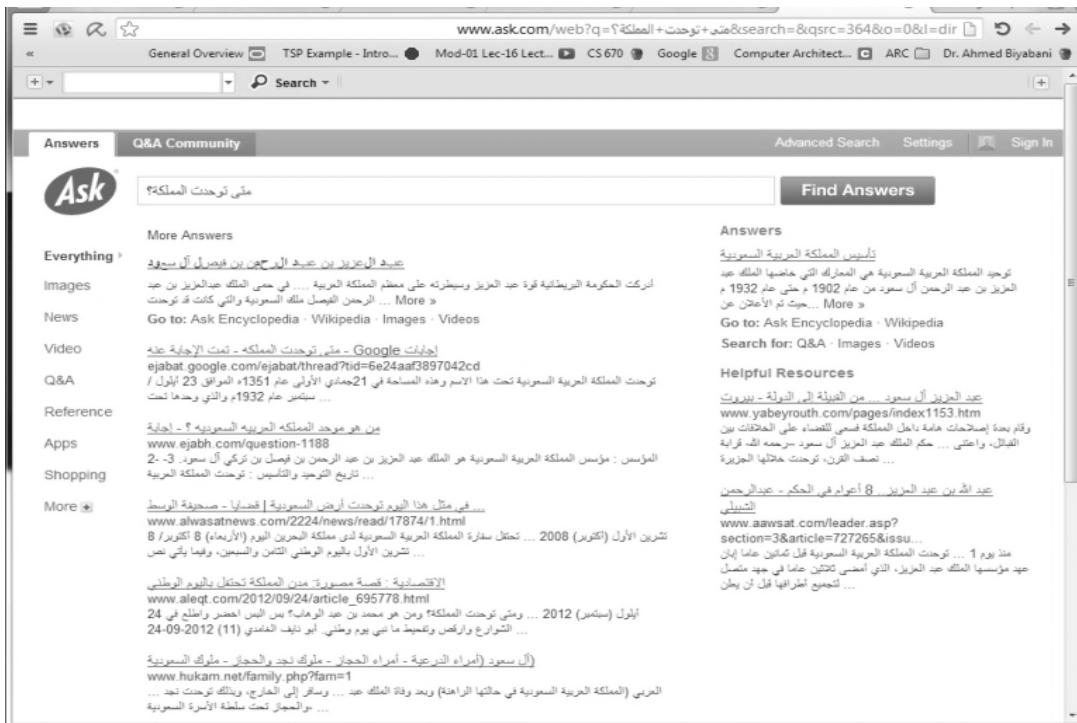


Figure 3.b: Ask.com Answers for when question



Figure 4.a: JAWEB answers for a what question

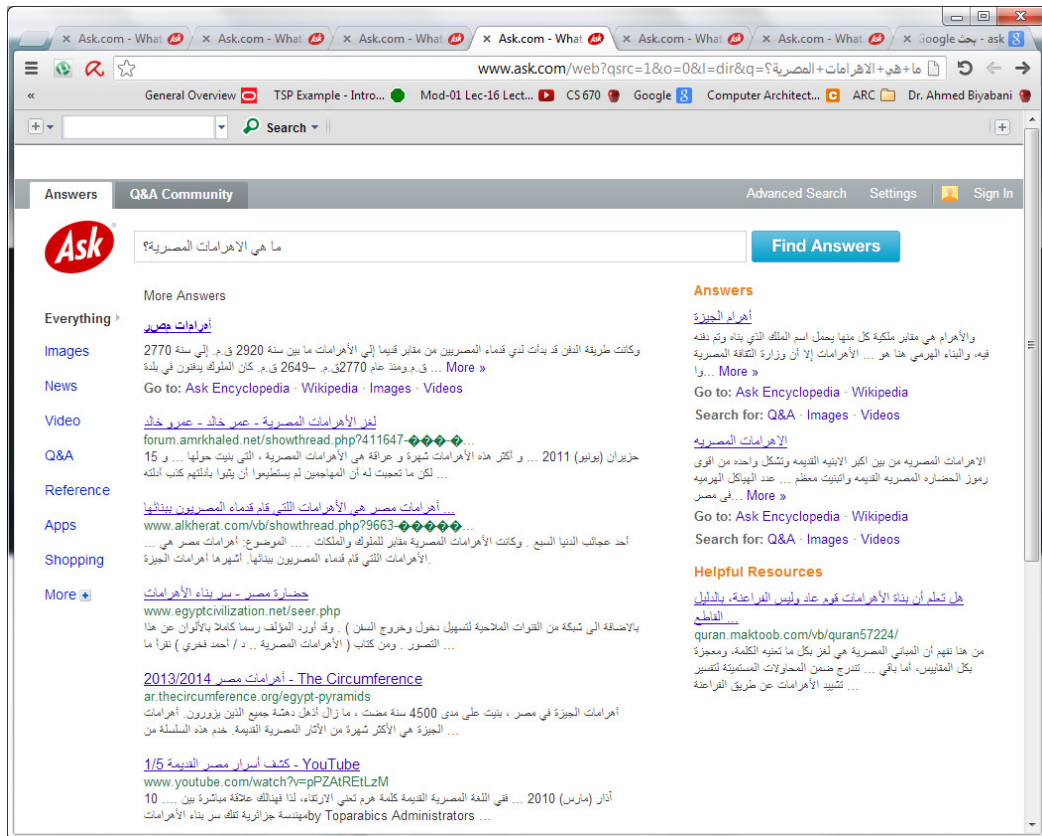


Figure 4.b: Ask answers for a what question



Figure 5.a: JAWEB answers for where question

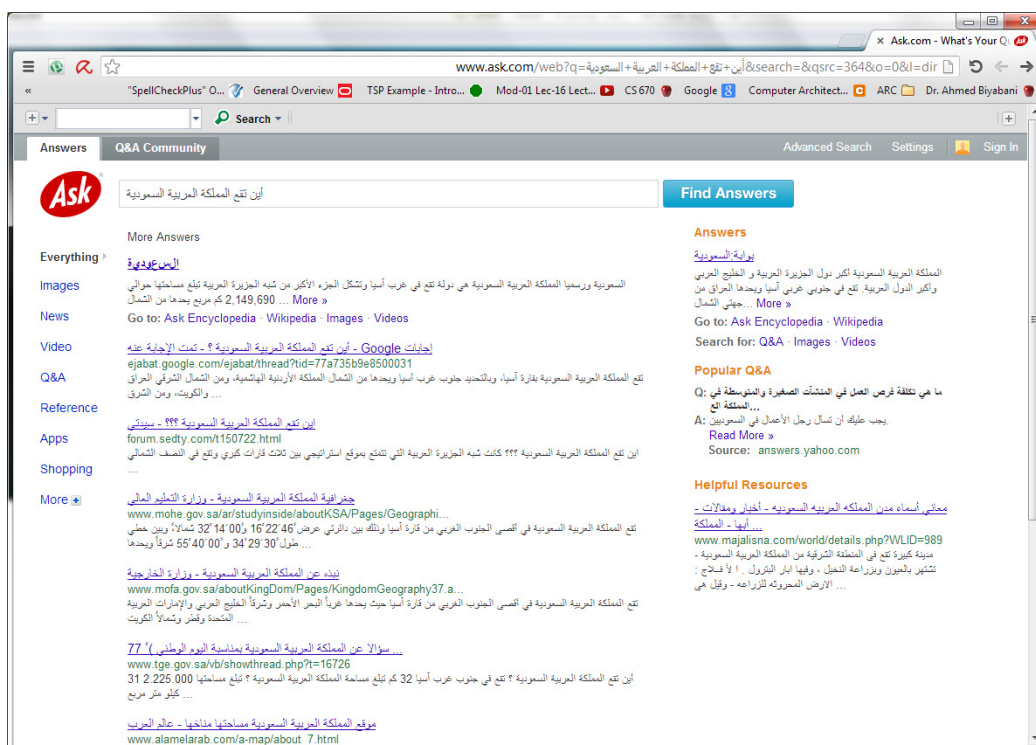


Figure 5.b: Ask answers for a where question



Figure 6.a: JAWEB answers for a how much question

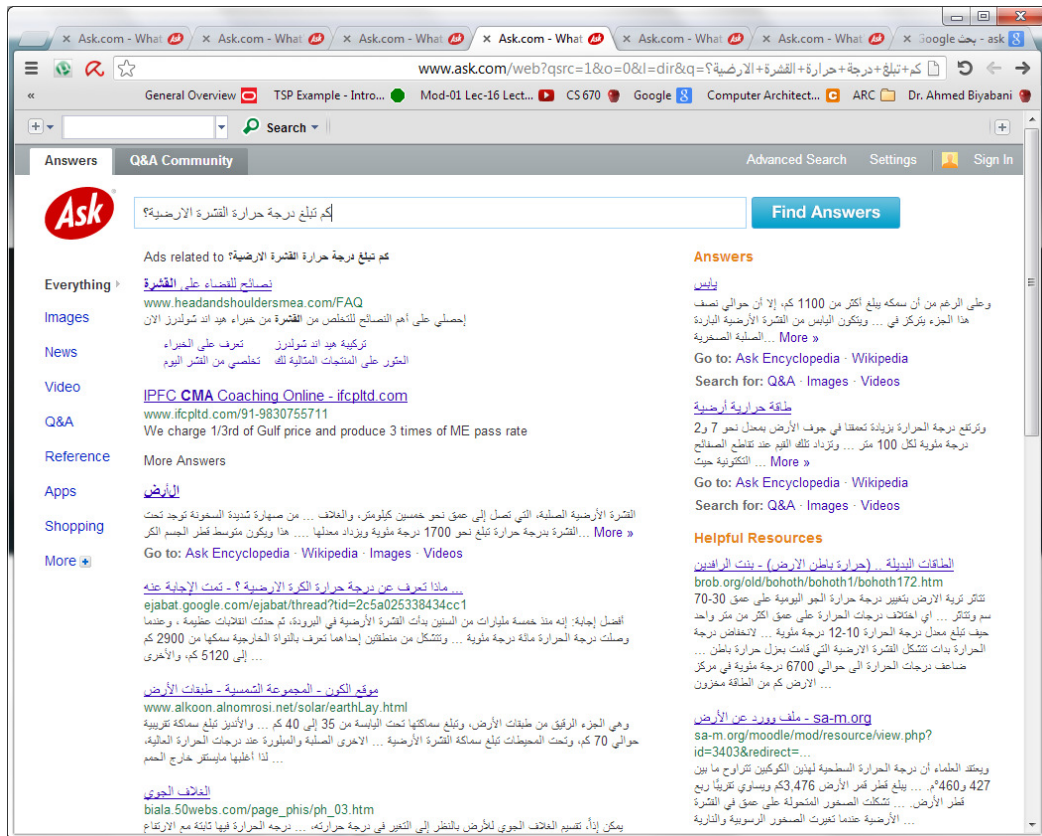


Figure 6.b: Ask answers for a how much question



Figure 7.a: JAWEB answers for a how many question

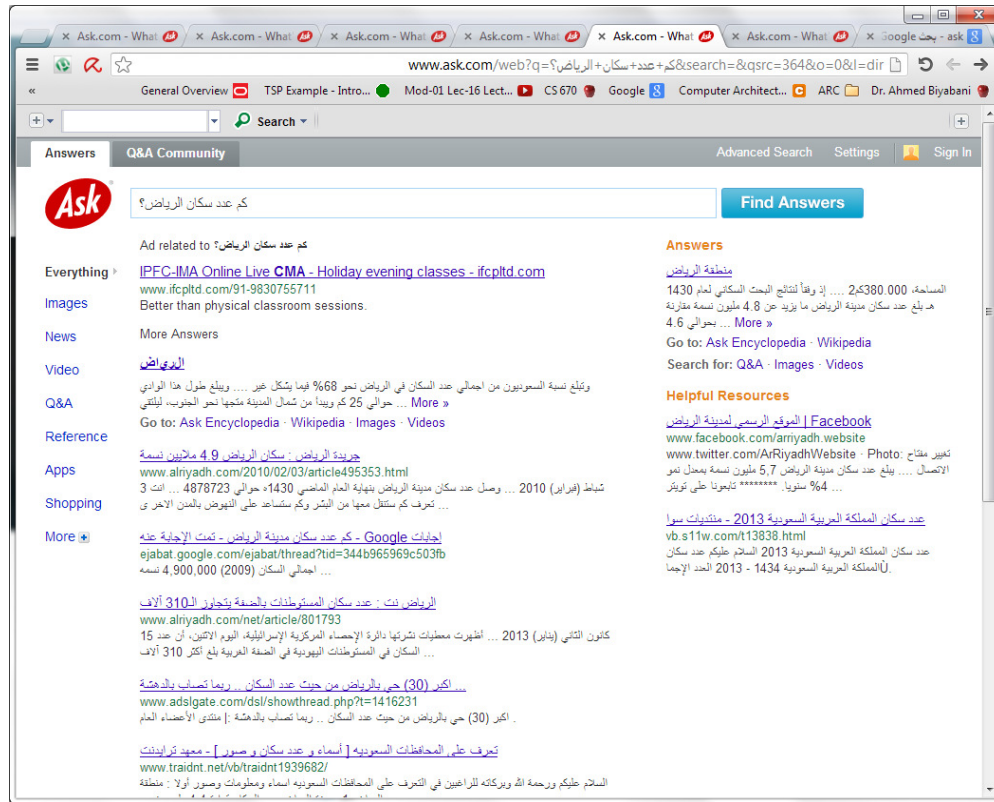


Figure 7.b: Ask answers for a how many question

Computer Science & Information Technology (CS & IT)

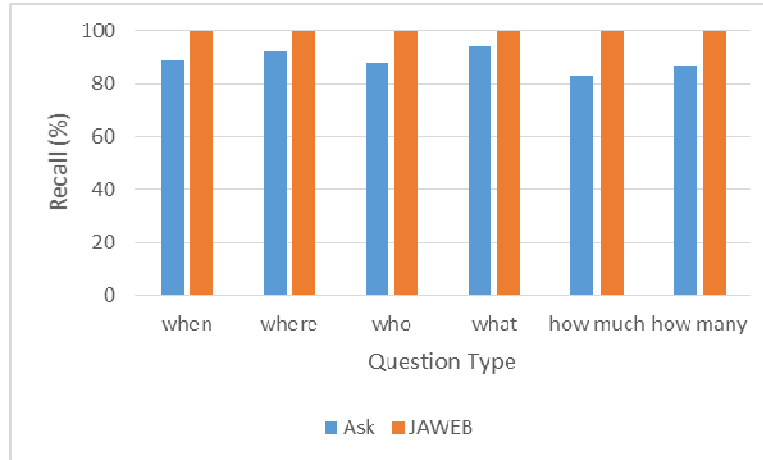


Figure 8: Comparison of the Recall

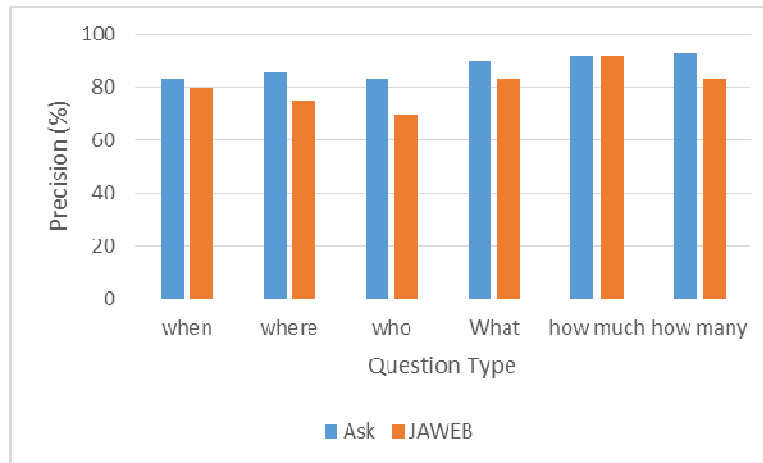


Figure 9: Comparison of the Precision

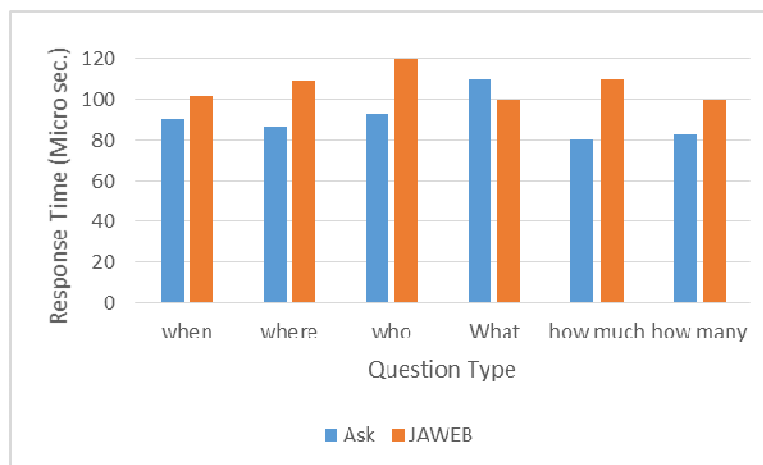


Figure 10: Comparison of the performance-Time

7. CONCLUSION & FUTURE WORK

JAWEB is a web-based Arabic question answering application system. It takes native Arabic questions from the end user and processes them in the server. The processing consists of three modules: question analysis, passage retrieval, and answer extraction. JAWEB finds and extracts accurate answers for the user. It also retrieves the most relevant potential answers and ranks them on the web page.

The evaluation experiment shows the high recall performance of the system, which is attributable to stemming. In the future, we expect to improve user interaction to incorporate user feedback for more precise results. We also plan on using the Arabic Named Entity Recognition to provide more accurate answers. In addition, we will consider the "why" and "how" questions, which require deep linguistic and semantic analysis.

ACKNOWLEDGEMENTS

This work was partially funded by the Long-Term Comprehensive National Plan for Science, Technology and Innovation of the Kingdom of Saudi Arabia, grant number 11- INF1895-08.

We would like to thank Dr. Sman Bekhti and his colleagues for their valuable support and comments based on their Arabic QA system AQUASYS. We would also like to thank Dr. Shereen Khoja for providing us with the Arabic stemmer package that she has developed.

REFERENCES

- [1] R. Alshalabi "Experimenting with a Question Answering System for the Arabic Language", *Computers & the Humanities*, Vol. 38, pp. 397-415, Nov 2004.
- [2] K. Arai and A. Handayani "Question Answering System for an Effective Collaborative Learning", *International Journal of Advanced Computer Science and Applications*, Vol. 3, pp.60-64,
- [3] A. Allam and M. Haggag, "The Question Answering Systems: A Survey," *International Journal of Research and Reviews in Information Sciences (IJRRIS)*, Vol. 2, No. 3, September 2012.
- [4] D. Hai, L. Kosseim, "The Problem of Precision in Restricted-Domain Question Answering. Some Proposed Methods of Improvement", In *Proceedings of the ACL 2004 Workshop on Question Answering in Restricted Domains*, Barcelona, Spain, Publisher of Association for Computational Linguistics, July 2004, PP.8-15.
- [5] A. Lamjiri, "A Syntactic Candidate Ranking Method for Answering Non-copulative Questions," Ph.D. thesis, Concordia University, 2007.
- [6] P. Walke, S. Karale, "Implementation approaches for various categories of question answering system," In *Proceedings of IEEE Conference on Information & Communication Technologies (ICT)*, 2013, pp. 402,407, 11-12 April 2013.
- [7] R. Alshalabi, "Pattern-based stemmer for finding Arabic roots," *Information Technology Journal*, Vol. 4, pp. 38-43, 2005.
- [8] R. Grishman, B. Sundheim, "Message Understanding Conference-6: A Brief History," In *Proceedings of International Conference on Computational Linguistics*, 1996.
- [9] A. Borthwick, "A Maximum Entropy Approach to Named Entity Recognition". PhD thesis, New York University, 1999.
- [10] Y. Benajiba, P. Rosso , J. Ruiz, "ANERsys: An Arabic Named Entity Recognition system based on Maximum Entropy", In *Proceedings of CICLing*, 2007
- [11] M. Maamouri, A. Bies, T. Buckwalter, "The Penn Arabic Treebank: Building a Large-scale Annotated Arabic Corpus," In *Proceedings of Arabic Language Resources and Tools*, 2004.
- [12] S. Mesfar, "Standard Arabic formalization and linguistic platform for its analysis," In *Proceedings of The Challenge of Arabic for NLP/MT Conference*, London, England, pp. 84-95, 2006.
- [13] A. Kabbaj, K. Bouzoubaa, "Amine Platform Page on SourceForge", <http://amine-platform.sourceforge.net/>, checked Aug. 2nd, 2013

- [14] JM Gómez, D. Buscaldi, P. Rosso, E. Sanchis, “JIRS Language-Independent Passage Retrieval System A Comparative Study,” In Proceedings of 5th International Conference on Natural Language Processing, ICON-2007, Hyderabad, India, January 4-6, 2007
- [15] W. Black, S. Elkateb, H. Rodriguez, M. Alkhalifa, P. Vossen, A. Pease, C. Fellbaum “Introducing the Arabic WordNet project,” In Proceedings of 3rd International WordNet Conference (GWC-06), 2006.
- [16] A. Ezzeldin, M. Shaheen “A Survey of Arabic question answering: challenges, tasks, approaches, tools, and future trends”, In Proceedings of The International Arab Conference on Information Technology, 2012, pp. 280-287.
- [17] S. Bekhti, A. Rehman, M. Al-Harbi, T. Saba “AQuASys: an Arabic Question Answering System Based on Extensive Question Analysis and Answer Relevance Scoring.” International Journal of Academic Research, Vol. 3, pp.45-54, July 2011.
- [18] O. Badawy, M. Shaheen and A. Hamadene “ARQA High-Performance Arabic Question Answering System”, In Proceedings of Arabic Language Technology International Conference, 2011, pp. 129-136.
- [19] F. Mohammed, K. Nasser, H. Harb “A knowledge-based Arabic Question Answering System (AQAS),” In Proceedings of ACM SIGART Bulletin, 1993, pp. 21-33.
- [20] B. Hammo, H. Abu-Salem, S. Lytinen. “QARAB: A Question Answering System to Support the Arabic Language”. In Proceedings of the workshop on computational approaches to Semitic languages, 2002, pp. 55-65, Philadelphia.
- [21] Y. Benajiba, P. Rosso, A. Lyhyaoui “Implementation of the ArabiQA Question Answering System's components.”, In Proceedings of Workshop on Arabic Natural Language Processing, 2nd Information Communication Technologies Int. Symposium, ICTIS, 2007.
- [22] W. Brini, M. Ellouze, S. Mesfar, L. Belguith “An Arabic Question-Answering system for factoid questionns,” In Proceedings of IEEE International Conference on Natural Language Processing and Knowledge Engineering, 2009.
- [23] S. Khoja, “Stemming Arabic Text,” <http://zeus.cs.pacificu.edu/shereen/research.htm>, checked may 6th, 2013.
- [24] <http://www.nooj4nlp.net/pages/introduction.html>, checked may 6th, 2013.

AUTHORS

Dr. Heba Kurdi received her B.Sc. and MSc. degrees in computer science from King Saud University, Riyadh, Saudi Arabia, in 1999, and 2003 respectively. In 2010, she got her PhD in wireless networks and communications from the School of Engineering and design, Brunel University, UK, where she also contributed as a visiting researcher in Summer 2011. She is currently an assistant professor at Al Imam Muhammad Ibn Saud Islamic University, Riyadh, SA. She has many publications in international conferences and journals and she is also a frequent reviewer for several computer science journals. Her research interest includes grid and cloud computing, and Bio-inspired Engineering.



Sara Alkhaider received her B.Sc. in Computer Science from Princess Noura University, Riyadh, Saudi Arabia, in 2009. She is currently an MSc student of Computer Science at Al Imam Muhammad Ibn Saud Islamic University, Riyadh, Saudi Arabia,. Her research interest includes web technologies, information retrieval .

Nada Al Faifi is an MSc student at the Computer Science Department in Al Imam Muhammad Ibn Saud Islamic University, Riyadh, Saudi Arabia. She has received her BSc. Degree in Computer sciences from the same university. Her research interest includes semantic web and artificial intelligence.