Towards an Arabic Question Answering system on the web

Dr. Mohammed M. A. Ibrahim Sakre
The Shorouk Academy
Higher Institute for Computers and Information Systems
m_sakre2001@yahoo.com

Abstract

The World Wide Web today is so huge that it has become more and more complicated to find answers to questions using regular search engines. Current search engines can return ranked lists of documents, but they do not bring direct answers to the user. The goal of Open Domain Question Answering (QA) systems is to take a natural language question, understand the meaning of the question, and present a short answer as a response based on a stored data. This paper presents an Arabic QA system model that is experimented on the World Wide Web. Techniques from Information Retrieval and Natural Language Processing are used in this research. The proposed model is tested and shows a success compared to other Latin systems.

Keywords: Artificial Intelligence, Arabic Natural Language Processing, Information Retrieval, Information Extraction , Question-Answering , semantic tagging.

1- Introduction and Motivation:

During the recent decades, the Web has become the main source of information, as nearly all kinds of data (digital libraries, newspapers collections, etc...) are stored in an electronic format. The data available is likely to satisfy most requests, nevertheless without the appropriate search facilities, the great amount of retrieved information is practically useless. Tools such as Search Engines (SEs) and Information Retrieval (IR) systems have been developed and are being essential to help users in their searching processes. In fact, these used mechanisms for instance, search engines such as Google [21], Yahoo [22] or MSN [23] allow a user only to retrieve the relevant documents which (partially) match a given query [1], [2]. Indeed, the use of SE presents a constraint for users as they have to manually filter a long set of returned documents. There is an urgent need for tools that would reduce the amount of text one might have to read in order to obtain the desired information. This paper aims at doing exactly that for a special (and popular) class of information seeking behavior: QUESTION ANSWERING. People have questions and they need answers, not documents. Automatic question answering will definitely be a significant advance in the state-of-art information

Research in the field of Q/A has known significant progress for languages such as English, Spanish, French or Italian [3]. In the context of the Arabic language there are few attempts for building Q/A systems. Unlike the Latin languages the Arabic Q/A task presents additional challenges to researchers in this field [6] [9]. This is mainly due to the particularities of the Arabic language (short vowels, absence of capital letters, complex morphology, etc.). However, the Q/A process modules are generally the same but the details differ with different languages. As figure (1) explains, there are three main modules [11],[24]:

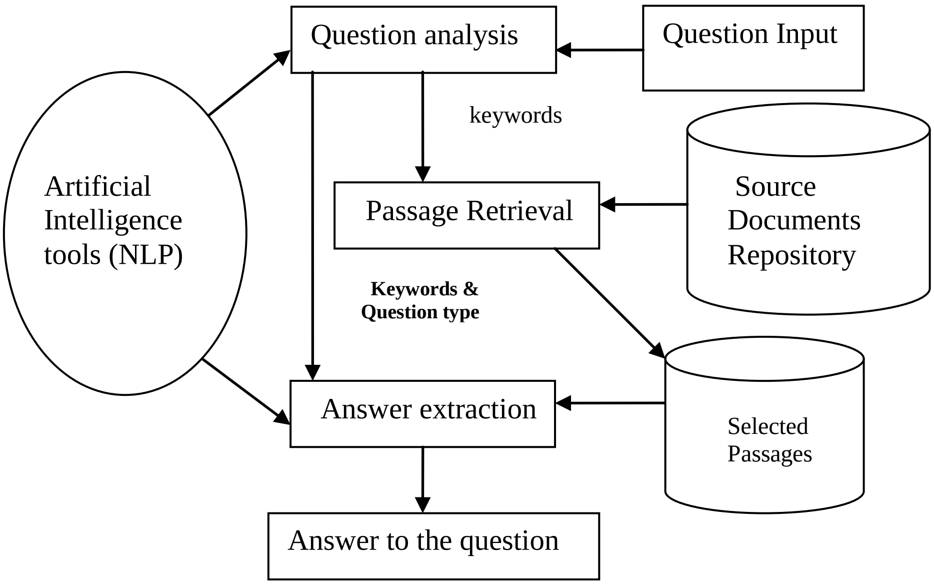


Figure (1) Generic diagram of question answering systems.

I- **The Question analysis module**: This module is responsible for tokenizing the input question; the question's keywords are extracted and the question/answer type is recognized. Natural Language Processing (NLP) tools, such as morphological, syntactical, and semantical analyzers, of the question language, may be used.

II- Passage Retrieval (PR) module: The keywords and the results of the analysis module, are used to retrieve the related documents from the set of Source Documents (Repository), then the related passages (paragraphs) in these documents are selected.

III- Answer extraction module: The results of the analysis module, which are the keywords and the question/answer types, are used to retrieve the question answer(s) from the set of related passages (paragraphs). NLP tools, such as morphological, syntactical and semantical analyzers may be used to find and form the answers.

2- Related work for Arabic QA system:

In comparison to other languages, there are few NLP tools and resources (e.g. lexicon, Arabic morphological analysis/generation tool, an Arabic lexicon supported with semantic features, Arabic synonym dictionary, etc…) which are available for the Arabic language. This is especially true for the QA task. In the international Text Retrieval Conference (TREC4) and Cross Language Evaluation Forum (CLEF5) competitions, there is not a QA task at the moment which includes the use of the Arabic language, and so far, only two Arabic Cross-Language IR (CLIR) tasks were organized in TREC 2001 and 2002 competitions (but neither proposed a QA task in Arabic). Therefore, QA systems are often developed for English as the target language because English is the language of the majority of documents on the Web. In [12] preliminary Cross-Language Question Answering (CLQA) experiments were carried out in order to allow for querying a system in Arabic and translating each question into English. The main aim was to investigate the effect of using a translator (from Arabic into English) in a QA system [2].

The most well-known Arabic Q/A systems are:

• QARAB [4] is a system that takes natural language questions expressed in Arabic and attempts to provide short answers. The system's primary source of knowledge is a collection of Arabic newspaper texts extracted from Al-Raya1, a newspaper published in Qatar. QARAB uses shallow language understanding to process questions and it does not attempt to understand the content of the question at a deep, semantic level.

• AQAS [5] is knowledge-based and, therefore, extracts answers only from structured data and not from raw text (non structured text written in natural language).

• ArabiQA [6] is an Arabic Q/A prototype based on the Java Information Retrieval System (JIRS) [7], Passage Retrieval (PR) system, and a Named Entities Recognition (NER) module. It embeds an Answer Extraction module dedicated especially to factoid questions. In order to implement this module authors developed an Arabic NER system [8] and a set of patterns manually built for each type of question.

• QASAL [9] is a recent attempt for building an Arabic Q/A which process factoid questions (e.g. questions that have NE answers). Experiments have been conducted and showed that for a test data of 50 questions the system reached 67.65% precision, 91% recall and 72.85% F-measure.

AQAS and QARAB offered the Arabic Natural Language Processing (NLP) research community the first prototypes of Arabic Q/A systems. However, AQAS processes only structured data whereas QARAB provides passages instead of precise answers. ArabiQA and QASAL target only factoid questions. The former integrates an NER system that has been evaluated and tested, the latter has also been tested; however, both have used a relatively low number of questions in their tests. The use of all these systems in an open domain such as the Web has not been tested.

Tritus is a Q/A system that automatically learns to transform natural language questions into queries containing terms and phrases expected to appear in documents containing answers to the questions [10]. A prototype search engine, Tritus, that applies the method to web search engines is presented. The researchers claimed that a blind evaluation on a set of real queries from a web search engine log shows that the method significantly outperforms the underlying web search engines as well as a commercial search engine specialized in question answering [10].

Answering multiple-choice questions, where a set of possible answers is provided together with the question, constitutes a simplified but nevertheless challenging area in question answering research**[13].**

3- The proposed system for Arabic Question Answering on the Web:

The proposed system model accepts user question, analyzes it, search the web for the related pages and extracts the proper answers from these pages. The question types managed in the proposed system model are the questions which start with any of the following question words:

(من ، كم ، أين ، متى ، ماذا ، ما هي ، ما هو، لماذا ، هل)

(who, how many, where, When, what, why)

The proposed system model consists of three main phases, as shown in figure (2), which are:

The Question analysis phase:

In this phase the keywords (mainly the non-stop words) are extracted and the question type is determined. This phase contains a special module which is the morphological analysis and generation module. This module is coded in Prolog in another research [14], [15] and despite it does not cover all the aspects of the Arabic language it is used in our system. Each keyword is analyzed and its Arabic root (stem) is extracted. Using the root, a set of the derivatives " مشتقات " of that root are generated. A semantic filter is used to select the semantically related words to the original keyword. These words are added to the list of keywords. The set of key words are passed to the next phase which is the Web search phase. Another set of words can be added to the list by searching a synonyms dictionary of the original set as in figure (3).

The Web search phase:

In this phase the resultant set of keywords are sent to Google search engine to get a set of related pages. The number of the keywords in the searched set should be not more than 32 words according to Google restrictions. So, the set of words basically contains the original words and the generated or synonym words. The output of this phase is a set of related pages.

The Information and Answer extraction phase:

In this phase the resultant set of related pages and the set of key words are passed to the information extraction module which extracts all the related text paragraphs and pass this set with the related URLs to the answer extraction module. The answer extraction module receives the question type and the set of key words from the question analysis phase as well. Using the received data, the answer extraction module finds the answer for the question.

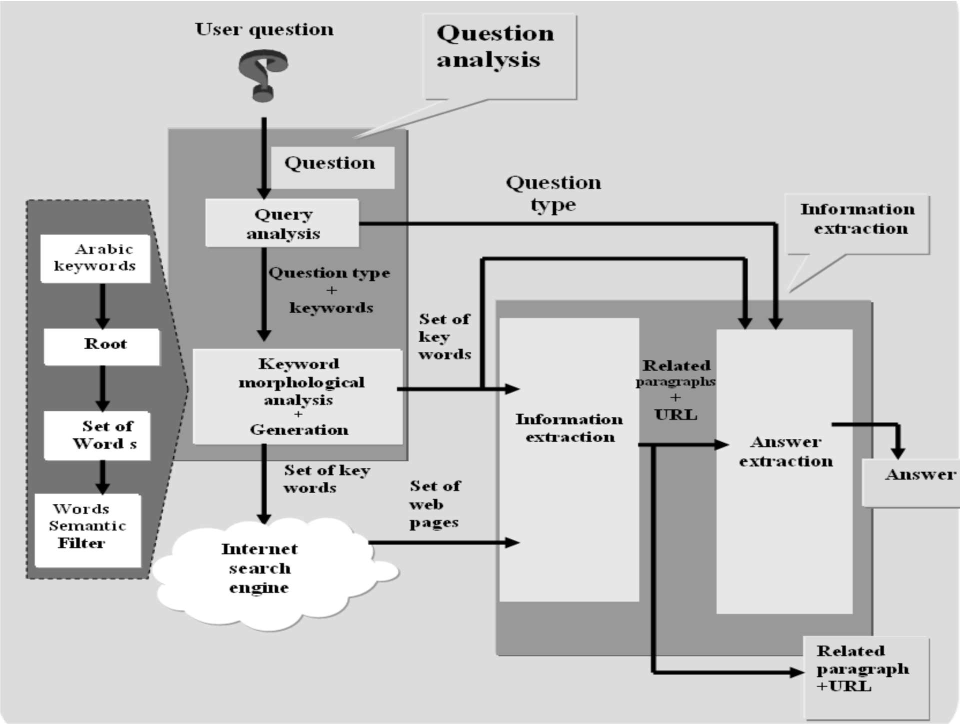


Figure (2) The main phases of the proposed system model

The answer extraction module uses a statistical approach with four built in tables to find the answers for the question types when, where, who and how many. The four tables are:

1) The time and date table which contains data about dates and time.
2) The locations table which contains data about names of locations and places all over the world.
3) The names table which contains data about people proper names.
4) The numbers table which contain data about numbers.

Using a suitable arrangement of the question’s keywords, in addition to the detected data in the text paragraphs, the answer is formulated and output to the user.

Figure (3) shows the possible execution paths of the proposed system model and Algorithm (1) clarify the execution steps in more details.

Figure (4-a) and Figure (4-b) show a sample of the user interface with the result answer of the proposed system model.

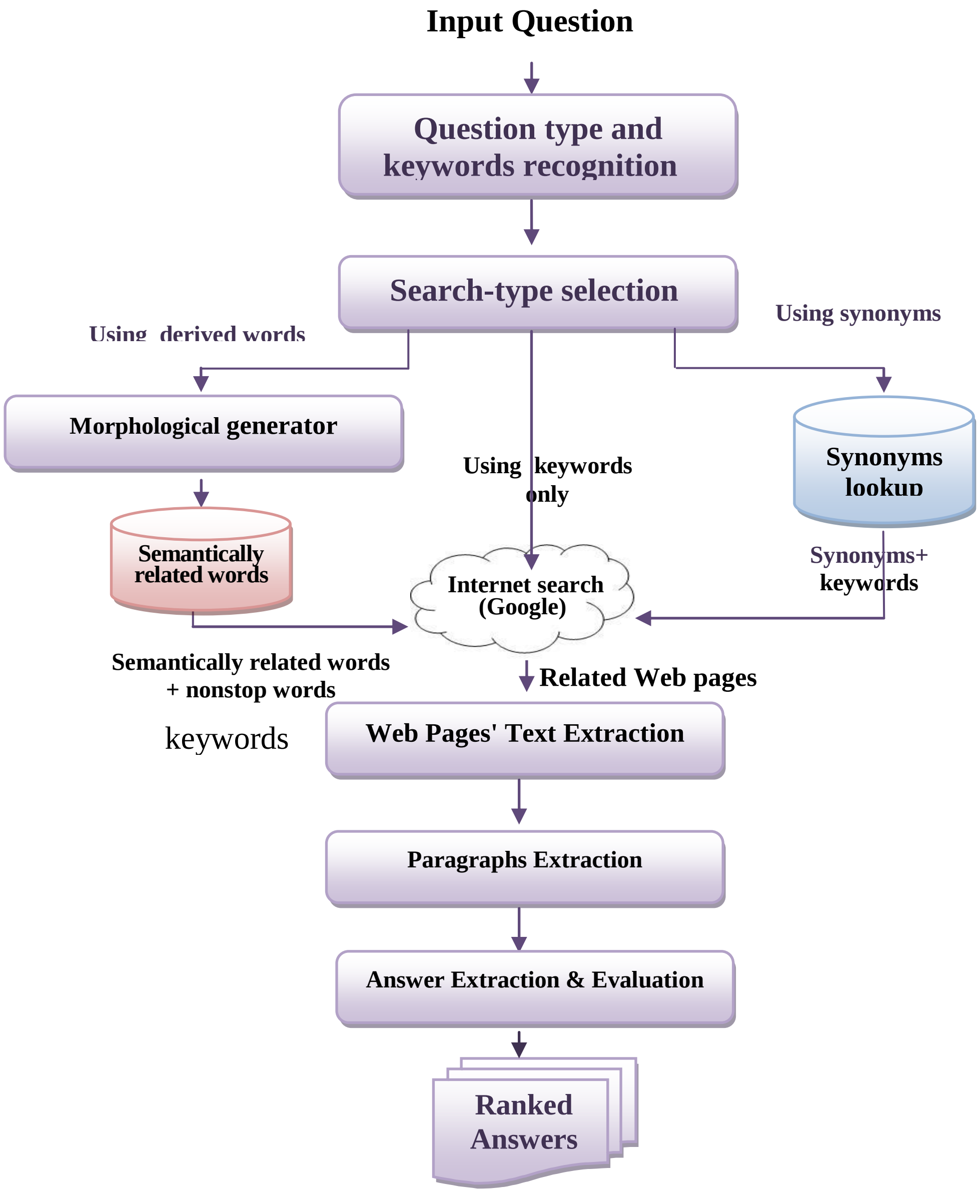


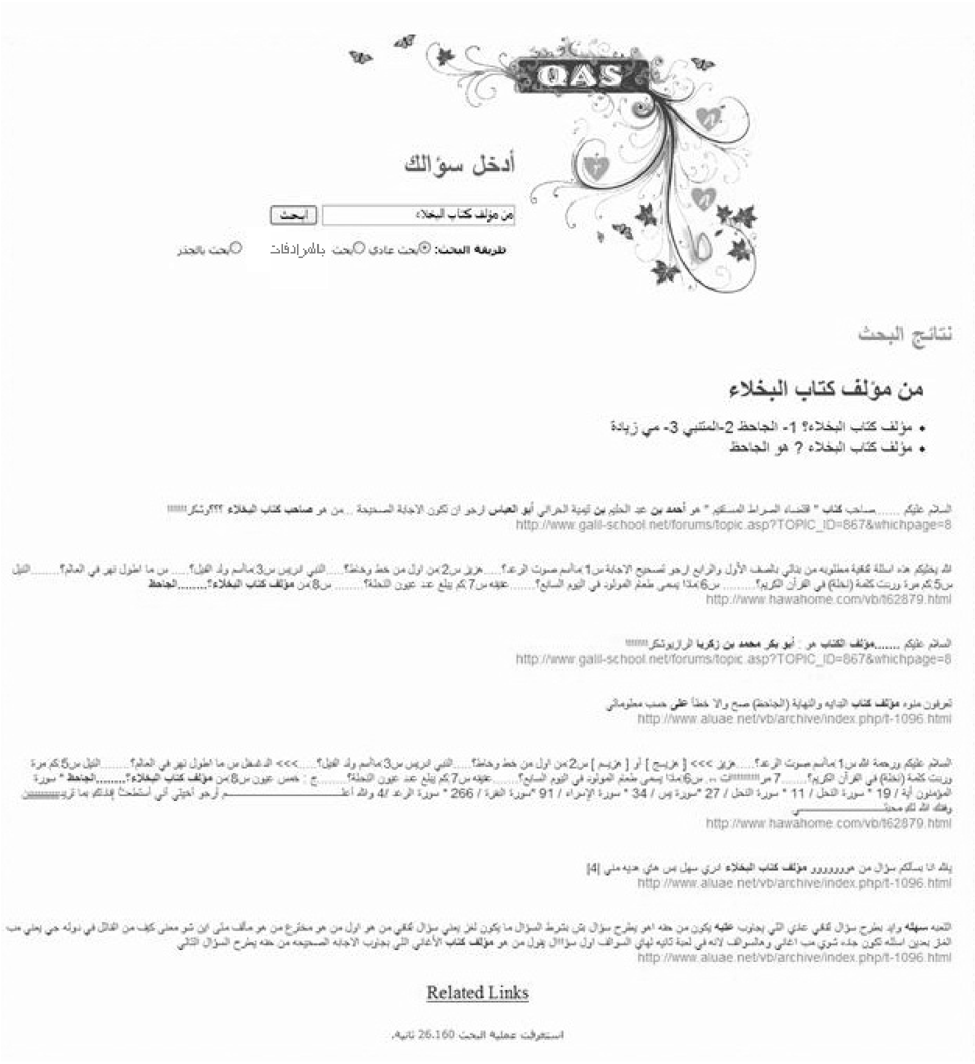
Figure (3) The execution chart of the proposed system model

Algorithm (1) :

```
1- Receive the input question, identify the question type and extract the
   possible keywords (non-stop words).
2- Receive from the user input the recommended search type (TYPE).
3- Case ( TYPE):
       Use keywords only (Non-stop words) only: go to step (4)
       Use keywords & synonyms: for each keyword find its synonyms
                                and add them to the set of the
                                current keywords.
                                Go to step (4)
       Use keywords & derived words:
           • For each keyword do :
               o Get all the possible derived words using an Arabic
                 Morphological Generator.
               o Select the semantically related words to the
                 current keyword.
               o Add the new words to the set of keywords.

           •   Go to step (4)

4- Search for the WebPages related to the set of keywords using Google
   search engine.
5- For each webpage, extract the text paragraphs that contain any of the
   keywords, and keep the page URL with each paragraph.
6- Give evaluation scores to paragraphs according to the number of different
   keywords existing in each one.
7- Sort paragraphs in a descending order.
8- Extract the most possible answer sentences from the sorted paragraphs
   using the DB (Database of places, proper nouns, time …), keywords, and
   the question type ( e.g. ask about place, time or person names…).
9- Output the ranked sentence, paragraphs and URLs.
```



أدخل سؤالك

من مؤلف كتاب البخلاء

طريقة البحث: ● بحث عادي ○ بحث بالمرادفات ○ بحث بالجذر

نتائج البحث

من مؤلف كتاب البخلاء

- مؤلف كتاب البخلاء؟ 1- الجاحظ 2-المتنبي 3- مي زيادة
- مؤلف كتاب البخلاء ? هو الجاحظ

السلام عليكم ......صاحب **كتاب** " اقتضاء الصراط المستقيم " هو **أحمد بن** عبد الحليم **بن** تيمية الحراني **أبو العباس** ارجو ان تكون الاجابة الصحيحة ...من هو **صاحب كتاب البخلاء** ؟؟؟وشكرا!!!!!
http://www.galil-school.net/forums/topic.asp?TOPIC_ID=867&whichpage=8

الله يخليكم هذه اسئلة ثقافية مطلوبه من بناتي بالصف الأول والرابع ارجو تصحيح الاجابة س1 ما اسم صوت الرعد؟....هزيز س2 من اول من خط وخاط؟....النبي ادريس س3 ما اسم ولد الفيل؟.... س ما اطول نهر في العالم؟.......النيل س5 كم مرة وردت كلمة (نحلة) في القرآن الكريم؟........ س6 ماذا يسمى طعم المولود في اليوم السابع؟......عقيقه س7 كم يبلغ عدد عيون النحلة؟....... س8 من **مؤلف كتاب البخلاء**؟.......**الجاحظ**
http://www.hawahome.com/vb/t62879.html

السلام عليكم ......**مؤلف الكتاب** هو : **أبو بكر محمد بن زكريا** الرازيوشكرا!!!!!
http://www.galil-school.net/forums/topic.asp?TOPIC_ID=867&whichpage=8

تعرفون منوه **مؤلف كتاب** البدايه والنهاية (الجاحظ) صح والا خطأ **على** حسب معلوماتي
http://www.aluae.net/vb/archive/index.php/t-1096.html

السلام عليكم ورحمة الله س1 ما اسم صوت الرعد؟....هزيز >>> [هزيم] أو [هزيم] س2 من اول من خط وخاط؟....النبي ادريس س3 ما اسم ولد الفيل؟....>>> الدغفل س ما اطول نهر في العالم؟.......النيل س5 كم مرة وردت كلمة (نحلة) في القرآن الكريم؟.......7 مرااااااات .. س6 ماذا يسمى طعم المولود في اليوم السابع؟......عقيقه س7 كم يبلغ عدد عيون النحلة؟.......ج : خمس عيون س8 من **مؤلف كتاب البخلاء**؟.......**الجاحظ** * سورة المؤمنون آية / 19 * سورة النحل / 11 * سورة النحل / 27 *سورة يس / 34 * سورة الإسراء / 91 *سورة البقرة / 266 * سورة الرعد /4 والله أعلـــــــــــــــــــم أرجو أختي اني استطعت إفادتكم بما تريديييييين وفقك الله لكم محبتـــــــــــــي
http://www.hawahome.com/vb/t62879.html

يلا انا بسالكم سؤال من هوووووو **مؤلف كتاب البخلاء** اتري سهل بس هاي هديه مني [4]
http://www.aluae.net/vb/archive/index.php/t-1096.html

اللعبه **سهله** واي يطرح سؤال ثقافي علاي اللي يجاوب **عليه** يكون من حقه انو يطرح سؤال بس بشرط السؤال ما يكون لغز يعني سؤال ثقافي من هو اول من هو مخترع من هو مالف ملي اين شو معني كيف من القائل في دوله جي يعني مب الغاز بدين اسئله تكون جلد شوي مب اغاني ومشوالف لانه في لعبة ثانيه لهاي السوالف اول سؤال يقول من هو **مؤلف كتاب** الأغاني اللي بجاوب الاجابه الصحيحه من حقه يطرح السؤال التالي
http://www.aluae.net/vb/archive/index.php/t-1096.html

Related Links

استغرقت عملية البحث 26.160 ثانية.

Figure (4-a) A sample of the user interface with the resulting answer of the proposed system.

4- Challenges in Arabic computational linguistics:

Arabic is a major and a highly inflected language, and thus requires good stemming for effective text mining. Yet no standard approach to stemming has emerged [4],[6].

Despite In this research, the need for the following Arabic linguistics tools appeared:

a- A complete Arabic morphological analysis/generation tool, to extract keywords roots and generate derivatives.

b- An Arabic lexicon supported with semantic features, to support in selecting the proper words generated as derivatives. The used semantic filter depends on a partial lexicon.

c- A complete Arabic synonym dictionary. The used synonym dictionary is a limited one.

There are many research efforts which cover the Arabic morphology [14]. Regarding the second requirement, up to the author knowledge there is no complete Arabic lexicon tagged with semantic features which is available in the market till the time of publishing this research.

5- Results and evaluations:

Information retrieval systems are usually compared on the basis of the "quality" of the retrieved document sets. This "quality" is traditionally quantified using two metrics, Recall (R) and Precision (P) [16]. Recall and Precision can be defined as:

$$R= r / K \quad …….(1) \qquad P = r/ N \quad ……..(2)$$

Where : r = Number relevant facts retrieved.

N = Total number of facts retrieved.

K = Total number of facts that in the answer key.

To measure recall over a collection, we need to mark every document in the collection as either relevant or non-relevant for each evaluation question. This, of course, is a daunting task for any large document collection, and is essentially impossible for the web, which contains billions of documents. Researchers have addressed this problem by developing standard document collections with queries and associated relevance judgments, and by limiting the domain of documents that are judged [17].

The system is going to be evaluated based on precision measure only (P) and observe enhances on the Recall measure (R). "K" in equation (1) can be considered to have a constant value for each question at specific time. So ‘r’ can be considered as a measure for enhancement in Recall.

Table (1) shows a question sample of the input to the system with the precision evaluation. For a sample of 78 Arabic questions, according to SPSS 13.0 statistical package, Mode is equal to 100 which means that most of the answers are correct as in table(2).

Statistics

QAS

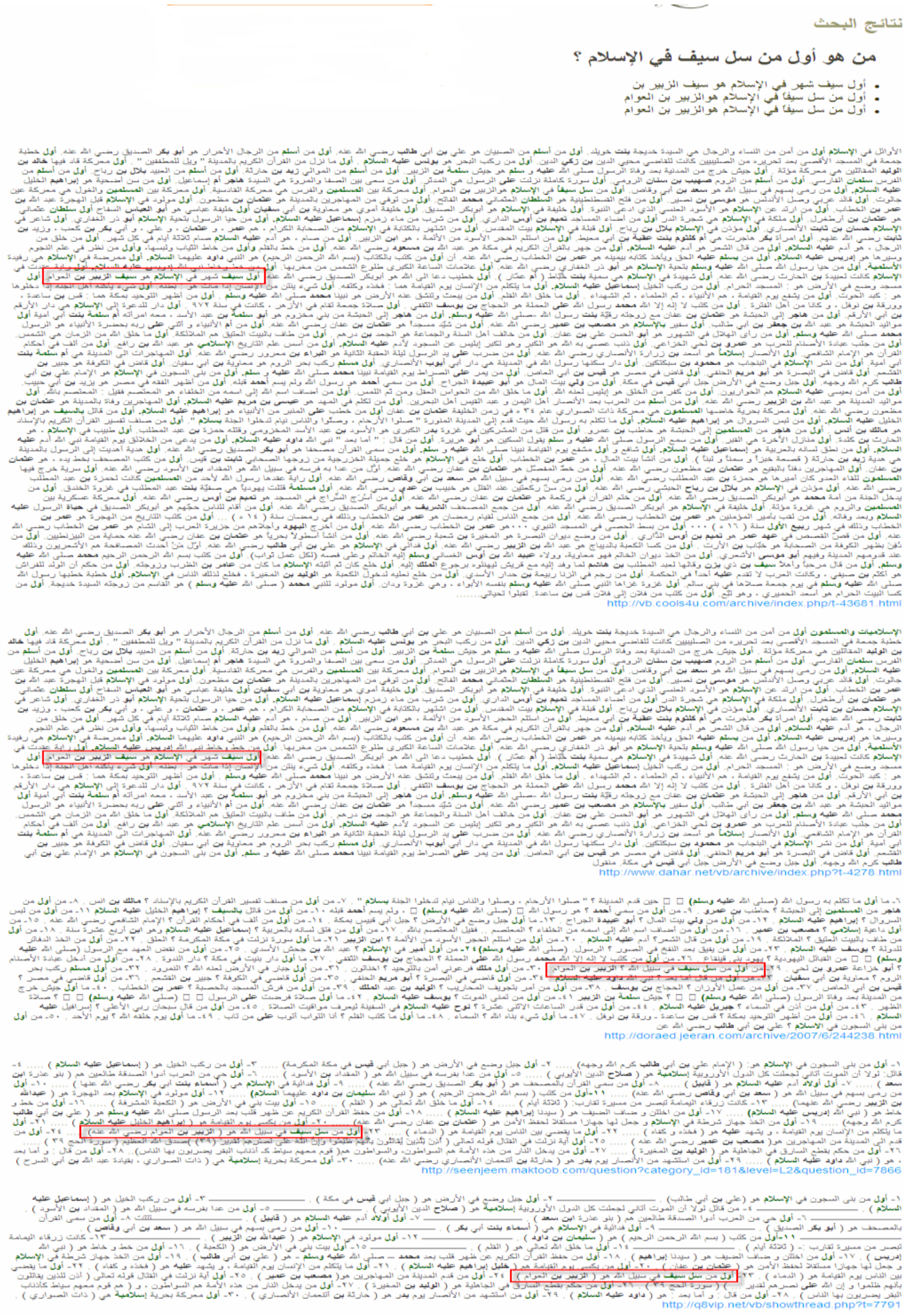
N	Valid	78
	Missing	0
	Mean	63.54
	Mode	100
	Std. Deviation	34.740
	Variance	1206.901

Table (2): result of SPSS 13.0

Since there are no other Arabic QA systems developed and working on the web yet, two Latin question answering systems are used in this evaluation. These two systems are: Ask [19] and Answers [20]. A sample of 100 questions is used as an input for the three systems and their precision result values are compared. It is to be noted that input questions to these two systems are in English and they search for the answer in an English sources while our system searches in Arabic sources. Table (3) shows the Precision measure of the proposed system compared to that of (Ask and Answers) systems for a sample of questions. Figure (5) shows a graphical representation of the proposed system with the Ask and Answers systems.

Precision	Questions in Arabic
100%	من هو أول من سل سيف في الإسلام ؟
80%	أين تقع مكة المكرمة ؟
100%	متى وقعت معركة أجنادين ؟
60%	كم يبلغ قطر الشمس؟
60%	من هو مخترع الانترنت ؟
80%	أين تقع جمهورية جزر فيجي؟
80%	متى زار ابن بطوطة مقديشو ؟
60%	كم يبلغ عدد سكان محافظة الأنبار؟
80%	من هي أم مريم ابنه عمران؟
100%	ن يصب نهر الغانج؟
100%	متى هاجمت المراكب الطوربيدية المراكب الأمريكية في خليج تونكين؟
50%	من هو مخترع البنج الطبي؟
100%	من هو مخترع القلب الاصطناعي ؟
33%	متى اخترع المصباح الكهربائي؟
25%	أين تقع جاكرتا ؟
100%	أين تقع بحيرة زرزر ؟
100%	من أول من كسا الكعبة المشرفة؟
50%	من هي أول شهيدة بالإسلام ؟
100%	كم يبلغ عدد سكان كالاباكا ؟
60%	من هو مخترع الهاتف؟
60%	كم تبلغ مساحة جمهورية مصر العربية؟
100%	من هو ملك الأردن؟

Table (1) A question sample of the input to the system with the precision evaluation



نتائج البحث

من هو أول من سل سيف في الإسلام ؟

- أول سيف شهر في الإسلام هو سيف الزبير بن
- أول من سل سيفاً في الإسلام هوالزبير بن العوام
- أول من سل سيفاً في الإسلام هوالزبير بن العوام

http://vb.cools4u.com/archive/index.php/t-43681.html

http://www.dahar.net/vb/archive/index.php?t-4278.html

http://doraed.jeeran.com/archive/2007/6/244238.html

http://seenjeem.maktoob.com/question?category_id=181&level=L2&question_id=7866

http://q8vip.net/vb/showthread.php?t=7791

Figure (4-b) A sample of a result answer of the proposed system.

Question	Arabic QA	Ask.com	Answers.com
من هو مخترع الهاتف؟ Who is the telephone inventor?	60%	80%	60%
أين تقع الإمارات العربية المتحدة؟ Where is the United Arab Emirates?	40%	20%	0%
كم تبلغ مساحة جمهورية مصر العربية؟ What is the surface area of Egypt?	60%	0%	100%
من هو ملك الأردن؟ Who is the king of Jordan?	100%	60%	20%
متى توفيت الأميرة ديانا؟ When did princess Diana die?	20%	20%	40%
أين توجد الغدة الدرقية؟ Where is the thyroid gland ?	100%	60%	0%
من هي ملكة الأردن؟ Who is the queen of Jordan?	100%	80%	20%
كم عدد جزر أندونيسيا؟ How many islands in Indonesia?	20%	40%	40%
أين يزرع الشاي؟ Where is tea planted in?	40%	20%	20%

Table (3): Precision measure of the proposed system and the Ask and Answers systems.

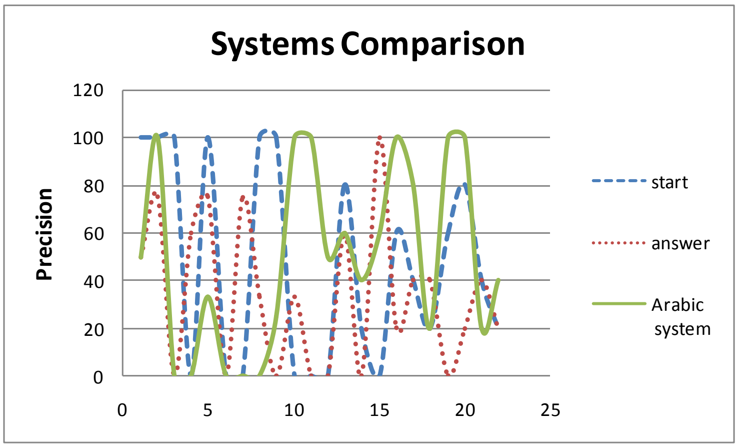


Figure (5): Evaluation statistics of the proposed system and the Ask and Answers systems.

Considering Start system as a control (base) for comparison between the three systems, according to ANOVA test, there are significant differences between the three systems as shown in table (4).

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
answer	Between Groups	11951.976	5	2390.395	2.889	.018
	Within Groups	79419.867	96	827.290		
	Total	91371.843	101			
Qas	Between Groups	52436.406	5	10487.281	10.819	.000
	Within Groups	93058.182	96	969.356		
	Total	145494.588	101			

Table (4): result of SPSS 13.0 (ANOVA test)

Using multiple comparison test (POST HOC), it is found that Answers system is better than Start except for answers with scores 60%. This is according to Benferroni and Dunnett T3. The same test shows that the proposed (QAS) is better than Start except for the answers with 60% system. In general, it is found that the proposed system acts better than the other two systems as shown in table (5).

Dependent Variable			(J) start (I) start	Mean Difference (I-J)	Std. Error	Sig.
answer	Bonferroni	0	20	19.733	9.096	.488
			40	2.400	10.138	1.000
			60	32.400(*)	9.824	.021
			80	6.400	10.503	1.000
			100	3.567	7.877	1.000
Qas	Bonferroni	0	20	13.333	9.846	1.000
			40	-7.394	10.974	1.000
			60	-54.667(*)	10.634	.000
			80	-38.667(*)	11.369	.015
			100	10.667	8.527	1.000

Table (5): result of SPSS 13.0 (Multiple Comparisons)

The author believes that this comparison is not fair for the three models, since different test beds and queries with different languages are used while they should be similar but the results demonstrates the capabilities of the proposed system.

6- Conclusions and future work:

The results of the proposed model show success in using Arabic Question answering system on the Web compared to other Latin systems. However, more efforts are still required to include all the Arabic question types and to enhance the answer extraction module of the system by including more natural language processing techniques to understand the selected paragraphs and form the answer.

The need for the Arabic linguistics tools such as "Arabic morphological analysis/generation tool, an Arabic lexicon supported with semantic features, and Arabic synonym dictionary" appeared to be available for scientific research.

7- References:

[1] Chowdhury, G. “ Introduction to modern information retrieval”, second edition, Facet publishing, 2004.

[2] Rosso P., Benajiba Y., Lyhyaoui A., “Towards an Arabic Question Answering system”. In: Proc. 4th Conf. on Scientific Research Outlook & Technology Development in the Arab world, SROIV, Damascus, Syria, 11-14 December, 2006.

[3] Buscaldi D., Rosso P., Gómez J.M., Sanchis E. ”Answering Questions with an n-gram based Passage Retrieval Engine”, Journal of Intelligent Information Systems , Volume 34, Number 2, 113-134, 2009.

[4] Hammou B., Abu-salem H., Lytinen S., Evens M., 2002. “QARAB: A Question answering system to support the ARABic language”. In: Proc. of the workshop on Computational approaches to Semitic languages, ACL, pages 55-65, Philadelphia.

[5] Mohammed F.A., Nasser K., Harb H.M. (1993), “A knowledge-based Arabic Question Answering System (AQAS)”. In: ACM SIGART Bulletin, pp. 21-33.

[6] Benajiba Y., Rosso P., Lyhyaoui A., 2007. “Implementation of the ArabiQA Question Answering System's components”. In: Proc. Workshop on Arabic Natural Language Processing, 2nd Information Communication Technologies Int. Symposium, ICTIS-2007, Fez, Morroco, April 3-5.

[7] Benajiba Y., Rosso P., Gómez J.M. “Adapting JIRS Passage Retrieval System to the Arabic”. In: Proc. 8th Int. Conf. on Comput. Linguistics and Intelligent Text Processing, CICLing-2007, Springer-Verlag, LNCS(4394), pp. 530-541.

[8] Benajiba Y., Mona D., Rosso P. Using Language Independent and Language Specific Features to Enhance Arabic Named Entity Recognition. In: IEEE Transactions on Audio, Speech and Language Processing. Special Issue on Processing Morphologically Rich Languages, Vol. 17, No. 5, July 2009.

[9] Brini W., Ellouze M., Hadrich Belguith L. (2009). “QASAL : Un système de question-réponse dédié pour les questions factuelles en langue Arabe”. In: 9ème Journées Scientifiques des Jeunes Chercheurs en Génie Electrique et Informatique, Tunisia. (in French).

[10] Eugene Agichtein , Steve Lawrence , Luis Gravano, Learning search engine specific query transformations for question answering, Proceedings of the 10th international conference on World Wide Web, p.169-178, May, 2001, Hong Kong.

[11] Lahsen Abouenour, Karim Bouzoubaa and Paolo Rosso, " Three-level approach for Passage Retrieval in Arabic Question/Answering Systems ", 3rd International Conference on Arabic Language Processing (CITALA'09), May 4-5, 2009, Rabat, Morocco.

[12] Rosso, P., Lyhyaoui, A., Penarrubia, J., Montes y Gomez, M., Benajiba, Y., Raissouni, N., Arabic-English Question Answering. In: Proc. Symposium on Information Communication Technologies Int., Tetuan, Morocco, 2005.

[13] Rawia Awadallah and Andreas Rauber, "Web-Based Multiple Choice Question Answering for English and Arabic Questions", Advances in Information Retrieval Lecture Notes in Computer Science, 2006, Volume 3936/2006, 515-518.

[14] Ibrahim M. M., " Information Retrieval of Arabic text using A.I. Techniques", M.Sc. thesis, Military Technical College, Cairo, Egypt, 1987.

[15] Gerald Gazder and Chris Mellish, " Natural language processing in prolog ", Addison – Wesley Publisher Company, 1989.

[16] Chowdhury, G. " Introduction to modern information retrieval", second edition, Facet publishing, 2004.

[17] E. Voorhees, "Overview of the English Text Retrieval Conference (TERC-8)", Proceedings of TERC-8, 1999.

[19] http://www.ask.com , 1-11-2010

[20] http://www.answers.com , 1-11-2010

[21] http://www. google.com , 11-1-2010

[22] http://www.yahoo.com , 1-10-2010

[23] http://www.msn.com , 1-11-2010

[24] Richard Cooper, Richard J Cooper And Stefan M Ruger. " A Simple Question Answering System", The Ninth Text REtrieval Conference (TREC 2000), Gaithersburg, Maryland, November 13-16, 2000. National Institute of Standards and Technology (NIST), online publication:
http://trec.nist.gov/pubs/trec9/t9_proceedings.html