Arabic-English Question Answering

P. Rosso[1], A. Lyhyaoui[2], J. Peñarrubia[3],
M. Montes y Gómez[4] , Y. Benajiba[2], and N. Raissouni[2]

[1] *Dpto. Sistemas Informáticos y Computación, Universidad Politécnica Valencia, Spain*
[2] *Abdelmalek Essaadi University, Ecole Nationales de Sciences Appliquées de Tanger, Morocco*
[3] *Facultad de Informática, Universidad Politécnica Valencia, Spain*
[4] *Laboratorio de Tecnologías del Lenguaje, Instituto Nacional de Astrofísica, Óptica y Electrónica, Mexico*

Emails: prosso@dsic.upv.es; abdel@tsc.uc3m.es; jlpenarr@upvnet.upv.es; mmontesg@inaoep.mx; benajibayassine@yahoo.fr; raissouni@ensat.ac.ma

ABSTRACT - ***The goal of a Question Answering (QA) system is to provide inexperienced users with a flexible access to the information allowing them for writing a query in natural language and obtaining a concise answer. QA systems are mainly suited to English as the target language. In this paper we will investigate how much the translation of the queries, from the Arabic into the English language, could reduce the accuracy of the QA task.***

1. INTRODUCTION

Nowadays, the Web has become our main information repository: nearly all kind of information (digital libraries, newspapers collections, etc.) in more than 1,500 languages is available on the Web in electronic format. These documents may satisfy almost every information need. Nevertheless, without suitable tools which could help the user, the great amount of retrieved information is nearly useless.

In *Information Retrieval (IR)* the user is interested in finding the most relevant documents which partially match a certain request (Baeza, 1999). Therefore, IR addresses the problems associated with the retrieval of documents from a collection in response to a user query and its goal is to search into a text collection (e.g. the Web) in order to return as result a subset of documents ordered by decreasing likelihood of being relevant to the given query. The most popular IR systems are the search engines for the Web (e.g. Google, Altavista and Yahoo). The aim of *Cross-Language Information Retrieval* is instead to retrieve documents written in a certain language (e.g. English) when the user query is written in another specific language (e.g. Arabic) (Y. Benajiba, 2004). In fact, if for instance the user is interested in investigating documents which are written in English, it would be nearly impossible to translate all of them into Arabic.

The goal of a *Question Answering (QA)* system is to provide inexperienced users with a flexible access to the information allowing them for writing a query in natural language and obtaining not the documents which contain the answer, but the concise answer itself (Vicedo, 2004). In recent years, the combination of the Web growth and the explosive demand for better information access has motivated the interest in Web-based QA systems. Due to the difficulty of the task, the last developments in QA (e.g. the prototype of the Piquant[1] (Practical Intelligent Question Answering Technology) IBM search engine) are mainly focused on answering factual queries (i.e., those having a simple named entity as the answer) (Del Castillo, 2004).

QA systems are often suited to English as the target language. *Cross-Language Question Answering* allows for querying the system in a language (e.g. Arabic) which is not the language of the documents (e.g English). In this paper we approach the challenging Arabic-English QA task. The main goal is to investigate how much the translation of the queries, from the Arabic into the English language, could reduce the accuracy of a QA system.

2. THE QA WEB-BASED APPROACH

The language-independent approach we used is supported by data redundancy (Brill, 2001) rather than sophisticated linguistic analyses of both questions and candidate answers. The main idea of the system we used, which is primarily based

[1] http://www.research.ibm.com/compsci/spotlight/nlp/

on (Del Castillo, 2004), is that the questions and their answers are commonly expressed using the same words, and that the probability of finding a simple (lexical) matching between them increases with the redundancy of the Web (Hovy, 2000), (Kwok, 2001), (Lin, 2002).

2.1. Query Reformulation

Given a question, the system first generates several query reformulations manipulating the order of the words of the question. The possible reformulations are illustrated for the question: *Where is the ICTIS Conference in 2005?*

- *Bag of words*: the set of words of the question different than prepositions, conjuntions and article (i.e., stopwords); e.g. *"is ICTIS Conference 2005"*
- *Verb movement*: in order to transform an interrogative sentence into a declarative one is necessary to eliminate the verb, or to move it to the final position of the sentence (a second word movement to the end was also investigated to consider the cases when an auxiliar verb exists), e.g. *"the ICTIS Conference in 2005 is"*
- *Components:* the question is divided in components (each component is an expression delimited by a preposition) and new reformulations are defined combining these components; e.g. *"is the ICTIS Conference" "in 2005" "in 2005 is the ICTIS Conference"*
- *Componenets without the first word*: in order to construct this set of reformulations we eliminate the main verb of the question, and then we apply the method of reformulations by components; e.g. *"in 2005 the ICTIS Conference" "the ICTIS Conference" "in 2005"*
- *Componenets without the first and the second words*: we suppose the presence of an auxiliar verb (not in the above example) and then we apply the method of reformulations by components.

Some of the above reformulations may not be syntactically correct and, therefore, not so likely to be found on the Web. On the contrary, the right ones will have a higher redundancy on the Web.

2.2. Snippets Recollection and Answer Extraction

After the query reformulation, the QA system sends each reformulation to a search engine (e.g. Google), and collects the returned snippets (document summaries) which were retrieved from the Web. This is an example of a snippet retrieved with the reformulation "*the ICTIS Conference in 2005*":

MyBusinessCommunities
... 11th international **Conference** on Concurrent Enterprising - ICE **2005** (103 **... ICTIS**'**2005** (131 visitors) Tetuan - Morocco, **2005**-06-03 till **2005**-06-06 **...**
www.prolearn-online.com/events.php?sort1=1&sort2=1&offset=60&newlanguage=1 - 24k - Cached - Similar pages

The right answer (*Tetuan – Morocco*) is in the retrieved snippet and it is only a matter to extract it.

To extract the most frequent n-grams (sequences of n words) from the snippets (each n-gram is defined as a possible answer to the given question), we used a statistical criterium which ranks them by decreasing likelihood of being the correct answer. The method which is used for the n-gram extraction and ranking is based on regular expressions A compensation factor is applied in order to avoid favoring short n-grams with respect to larges ones. The method extracts the twenty most frequent unigrams which satisfy a given typografic criteria (i.e., words starting with an uppercase letter, numbers and names of months), determines all the n-grams (from bigrams to pentagrams, built from the set frequent unigrams), ranks the n-grams based on their compensated relative frequency, and finally selects the top five ngrams (candidates as possible answes).

The compensated relative frequency of a **n**-**g**ram **g**(**n**) **=** ($w_1 \ldots w_n$) is computed as follows (Del Castilo, 2004):

$$P_{g(n)} = \sum_{i=1}^{n} \sum_{j=1}^{n-i} \frac{f_{j(i)}}{\sum_{\forall x \in G_i} f_{x(i)}} \qquad (1)$$

where G_i is the set of n-grams of size i, $|G_i|$ indicates the cardinality of this set, $j(i)$ is an –gram j of size i contained in $g(n)$, and $f_{j(i)}$ is the frequency of occurrence of this n-gram.

3. PRELIMINARY EXPERIMENTS

3.1. The CLEF-2003 Query Corpus

Some preliminary experiments were carried out using the queries corpus of the CLEF[2]-2003 competition. The Cross-Language Evaluation Forum (CLEF) is a European consortium that

[2]www.clef-campaign.org

organises an international competition regarding of IR and QA systems, operating on European languages in both monolingual and cross-language contexts. In our case, we used the questions in English and compared the answers with those obtained after the translation process into English from an Arabic corpus which was manually created by a linguist. For the Arabic-English translation process, an automatic machine translator was used.

3.2. The Query Translation Process

One of the nowadays challenge is writing a question in a language (e.g. Arabic) and query a collection of documents which are written in another language (e.g. English). In fact, it would be technically impossible to translate all the target documents into the query's source language.

The main aim of our preliminary experiments was to investigate how much the translation of the queries, from the Arabic into the English language, could reduce the accuracy of the QA task. For the translation of the questions the *TARJIM*[3] Arabic-English machine translation system was used.

3.3. Experimental Results

For each question we generated the five different kinds of query reformulations, and for each reformulation we collected, if posible, 50 snippets. Table 1 shows the precision (i.e., the proportion of the questions which were correctly answered) of the preliminary experiments we carried out comparing the performance of the QA system when the original English questions and those obtained after the Arabic-English translation were used. The best results were generally obtained with the "verb movement" reformulation. Nevertheless, a more detailed analysis of the results showed us that there are cases (e.g. questions like "What is the capital of…" or "In what year…") in which other reformulations (e.g. the components ones) allow to obtain the right answer.

Table 1. Precision of correct answers (over 450).

Questions	Bag words	Comp.	Comp no 1st word	Comp no 1st and 2nd words	Verb mov.
English (original)	9.1% (41)	17.1% (77)	14.9% (67)	10.4% (47)	**24% (108)**
English (from Arabic)	3.8% (17)	1.6% (7)	4.9% (21)	4.9% (21)	**7.2% (31)**

In the further experiments, we took into consideration the top five better answers for each question. In order to fully evaluate the performance of the QA system, the precision measure was used together with the *Mean Reciprocal Rank (MRR)* of the first correct answer (see Table 2). The MRR is computed as follow:

$$MRR = \frac{1}{n}\sum_{i=1}^{n} r_i \qquad (2)$$

where n is the total number of test questions and r_i is the reciprocal of the rank (position in the answer list) of the first correct answer. For instance, if the correct answer is in the second position, r_i = 0.5, whereas if it is in the third then r_i = 0.33. In the case the correct answer does not occur in the list of the top five n-grams, then r_i =0.

Table 2. Precision and MRR measures.

Questions	Bag words	Comp.	Comp no 1st word	Comp no 1st and 2nd words	Verb mov.
English (original)	17.1% 0.12	24.4% 0.19	26.7% 0.20	22.0% 0.16	**39.5% 0.31**
English (from Arabic)	6.0% 0.04	2.4% 0.02	7.4% 0.06	8.4% 0.06	**10.7% 0.08**

In every query reformulation, the translation process caused a decreasing of even more than 30% in the performance. Tables 3 and 4 show a couple of bad translations (in the first one the proper name "Nirvana" was also wrongly translated).

Table 3. Example in which also a proper name was badly translated.

original	What was the name of the singer and head of Nirvana?
Arabic	ما اسم المغني و رئيس نرفانا ؟
translation	What is the name of the main singer of Nirfana?

Table 4. Example of bad translation.

original	How many European countries form part of the G7?
Arabic	كم عدد الدول الأوربية المكونة لمجموعة السبع؟
translation	Quantity of an European country belongs to the group of seven?

On the other hand, in quite unusual cases (see Tables 5 and 6) with the translated question we obtained a right answer whereas, we did not obtain any (first case) or we obtained a wrong one (second case) with the original one.

Table 5. Example of wrong translation and right answer (California).

original	Which American state has the strictest environmental laws?
Arabic	ما هي الولاية الأمريكية ذات القانون البيئي الأكثر صرامة؟
translation	What she is the American state for which the environmentallaws with more stricness?

[3] http://tarjim.ajeeb.com/ajeeb/default.asp

Table 6. Example of wrong translatrion and right answer (February)

original	During what month do almond trees blossom?
Arabic	متى تزهر أشجار اللوز؟
translation	During any month the almonds trees bloom ?

4. CONCLUSIONS AND FURTHER WORK

The performance of a cross-language Arabic-English QA system is very much affected by the translation process. In the experiments we carried out the QA performance decreased of more than 30%. More machine translators should be used at the same time in order not to rely just on one translation and to choose the best one on a statistical basis (Larosa, 2005). It should be also interesting to use the query reformulation technique directly to the Arabic language without the necessity to go through the Arabic-English translation process for each query.

Acknowledgments

This work was made possible "sucran" (thanks) to the R2D2 (CICYTTIC2003-07158-C04-03), ICT EU-India (ALA/95/23/2003/077-054) research projects and Conacyt (J43990-Y).

References

Baeza, R., and Ribeiro, B., 1999, Modern Information Retrieval. ACM Press, New York, Addison-Wesley.

Brill, E., Lin, J., Banko, M., and Dumais, S., 2001, Data-intensive question answering. In TREC-10 2001.

Del Castillo, M., Montes y Gómez, M., and Villaseñor, L. 2004, QA on the web: A preliminary study for Spanish languege. Proceedings of the 5th Mexican International Conference on Computer Science (ENC04), Colima, Mexico.

Hovy, E., Gerber, L., Hermajakob, U., Junk, M., and Lin, C., 2000, Question answering in Webclopedia. In TREC-9 2000.

Kwok, C., Etzioni, O., and Weld, D. 2001, Scaling question answering to the Web. Proceedings of the WWW Conference, 2001.

Larosa, D., et. al., 2005, Best Translation for an Italian-Spanish Question Answering System. Proceedings of Information Communication Technologies International Symposium - ICTIS'05, Tetuan, Morocco, 3-6 June 2005.

Lin, J., 2002, The Web as a resource for question answering: perspectives and challenges". Proceedings of the 3rd International Conference on Language Resources and Evaluation (LREC-2002).

Y. Benajiba, M. S. Kasttet, and A. Lyhyaoui, 2004, Bilingual Information Retrieval. Internal Report. Abdelmalek Essaadi University, Ecole nationals des sciences appliqués de Tanger, Morocco.

Vicedo, J., 2004, Los Sistemas de Búsqueda de Respuestas desde una Perspectiva Actual. In Revista Iberoamericana de Inteligencia Artificial. (in Spanish).