

Interactive and Bilingual Question Answering using Term Suggestion and Passage Retrieval

Carlos G. Figuerola, Angel F. Zazo, José L. Alonso Berrocal, and Emilio Rodríguez Vázquez de Aldana

University of Salamanca, REINA Research Group,
figue|afzazo|berrocal|aldana@usal.es,
<http://reina.usal.es>

Abstract. The Question Answering Task requires user interaction. Users can help the system by reformulating the questions, adding information to them or selecting the documents on which the system should work to obtain the answers. Our group has researched the effects on user interaction of suggesting terms to be added to the question, and the differences between using fragments or complete documents. This article describes the experiments we carried out and discusses the results we obtained.

1 Introduction

Several tasks in Information Retrieval require user interaction. In the case of the task known as Question Answering, the objective is to provide specific answers to specific information needs. Some approaches to this problem are based on the retrieval of passages or fragments of text [1, 2], assuming that the answer is to be found in these passages. The answer can then be extracted either through an automatic process or through user interaction. If the system is not able to provide a valid answer immediately, some kind of feedback process should be provided so that the user can further express his/her information needs.

Moreover, the task proposed in CLEF 2004 in the *i-track* was multilingual. Thus user interaction must work on a translation of the questions, and not on the target documents [3]. The questions are in one language and the documents (or passages from them) are in at least one other language. Thus, the questions must be translated to the same language as the documents. An alternative approach is to translate the documents to the language of the questions [4], but this is more expensive in terms of processing capability.

This year, the activity of our group has focused on exploring the effects of two kinds of user interaction: on the one hand, making the system suggest a set of terms translated into the language of the documents to the user, on the other hand, allowing the system to work not only on passages but, at the user's request, also on complete documents. In both cases the goal was to evaluate not only the number of correct answers, but also the subjective evaluation that the user makes of this type of help.

This article is organised as follows: first a description of the task to be performed is provided; then the system used for the exercise is described together with the experimental design; finally the results are discussed.

2 Experimental Design

2.1 Task Proposed

This year's task was Cross Language Question Answering. The initial scenario was the following: a collection of documents in English and questions in Spanish. The users' native language was Spanish and they had a passive knowledge of English, which allowed them to understand the content, albeit only partially, and to interact using some simple terms in that language.

All the users selected for the experiments (8 in all) were students of information science, accustomed to working with point-and-click interfaces, as well as to making searches in computerised library catalogues and using the search services of the WWW. Their experience with machine translation programs, however, was scarce (see Table 1).

Table 1. Pre-Experiments Questionnaire Averages

Age	24.13
Experience in using a point-and-click interface	4.38
Exp. in computerized OPACs	4.26
Experience in searching commercial systems	3.25
Searching on www search services	4.5
Using Machine Translation software	2.38
How often conduct a search	4.25
Enjoy carryng out information searches	3.75
Reading skills in document language (english)	3.13

2.2 Retrieval System

Since we did not have a true Question Answering system available, we used a conventional retrieval system based on the vectorial model, but with some adaptations. The process was the following:

- translate the question using a machine translation software
- carry out a conventional retrieval based on the translated question
- let a user read the retrieved documents and deduce the answer to the question

Our base system for retrieval was the same as the one we used in some prior editions of CLEF [5], with a classic scheme for calculating the weights of the terms based on $tf \times idf$ [6]. We made some additions and changes to this base system:

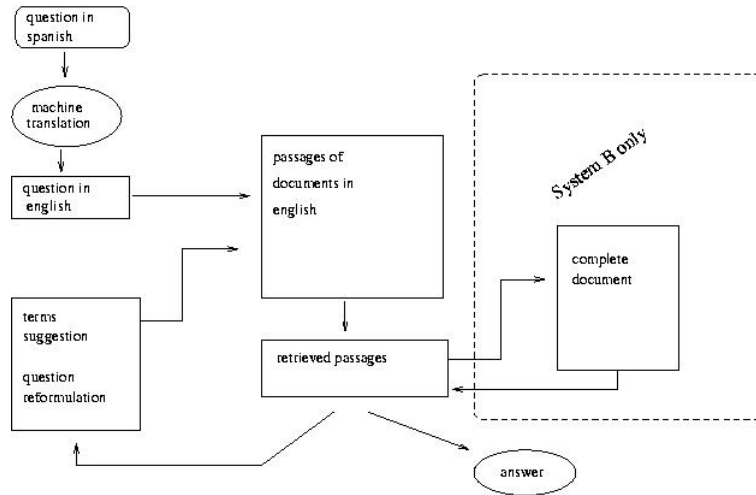


Fig. 1. Layout of the Retrieval System

- The machine translation programme used was Reverso [7]. The version we used is freely available to the public through the web page of a well-known newspaper (<http://www.elmundo.es/traductor>); The quality of its translations is slightly superior to that offered by other similar software, but one of its features is that it offers several alternative translations of the same term. In this version there are no specialised dictionaries nor, obviously, any type of training.
- the basic indexing unit is not the document but rather passages or fragments of it. The documents were divided into windows of 100 words each (including stop words). Only the TEXT field of the documents was used. The list of stop words was the standard list of SMART [8], to which the words appearing in more than 15 % of the documents were added. The final average of words per passage was less than 100, owing to the end fragments of the documents; note that the documents of the CLEF collections used are relatively short since they are news items (see Table 2).

Table 2. Documents Collection and Passage Division

Collection	LA94 & GH95	Passage Division
Documents	169,477	915,283
Total index terms	302,241	302,241
Averaged doc length (words)	229.94	42.71
Averaged doc length (unique index terms)	163.77	38.01

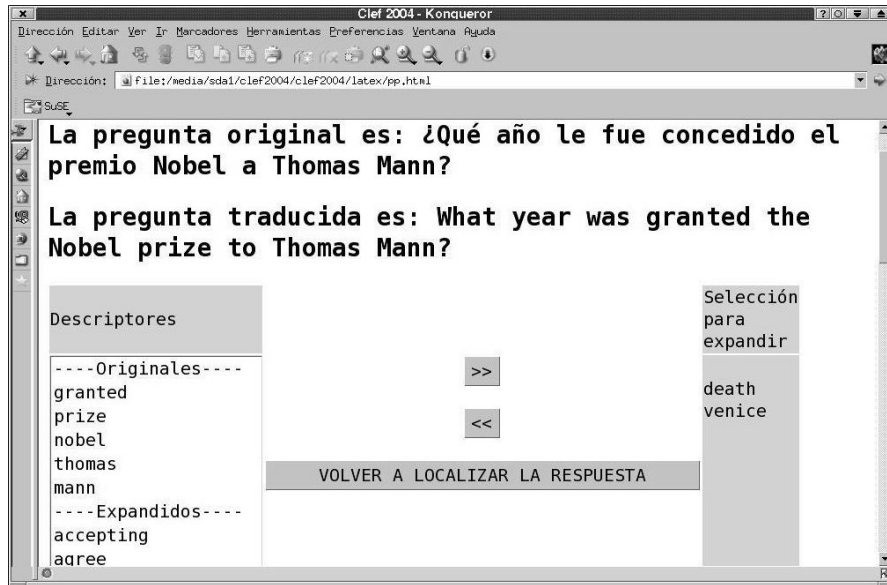


Fig. 2. Screen of Suggesting Terms

- the capability of suggesting terms to the user was added to the system so that the user could add these terms to the question translated. These suggested terms were in English (the language of the documents); the idea was that these terms could improve the translation of the questions obtained automatically. It was expected that, since the users had a passive knowledge of the language of the documents, they would be capable of understanding and including some of the terms suggested. The terms suggested for each question were obtained by means of query expansion techniques. Our group has a lot of experience in query expansions applied to classic tasks of information retrieval [5, 9, 10]. Thus, we chose the expansion technique offering the best results, i.e. the use of local association thesauri. The cooccurrence relations of the terms in the first documents retrieved were used to construct the thesaurus. Hence, the terms best related to all the terms of the original question are selected. For each question the 30 best related terms are obtained, and are shown to the user so that they can incorporate them in the original question if they wish. Term suggestion mechanisms are used frequently in interactive experiments[11].
- Both the access to and the interaction with the system were carried out through a web server and several forms. The most important features of the system are the possibility to make several iterations (reformulating the translated question, examining the passages retrieved etc.) and the possibility to obtain and read the complete document for the passage retrieved.

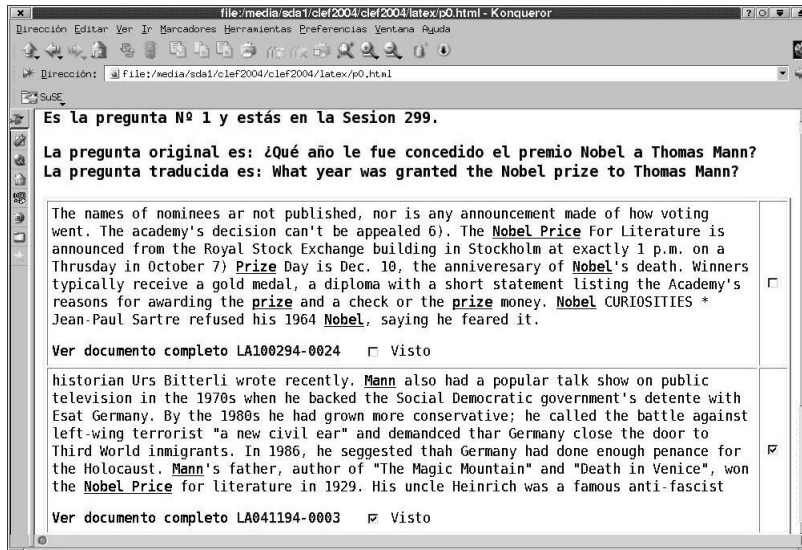


Fig. 3. Retrieved Passages from Question 1

2.3 The Experiment

All the users were given prior training. For the experiment the retrieval system was prepared in two different ways (**system A** and **system B**): System B permitted access to complete documents, while while system A did not.

All the users were asked to submit 16 questions, half using system A and the other half using system B. However, the questions and systems were ordered in such a way that half the users began with system A and the other half with system B. Each user worked with a different sequence of questions.

All the user operations were logged by the system. Each question was allowed a maximum of 5 minutes to be processed; those that took longer, were considered invalid (19.5 %). The duration of each session seemed to be related to the number of iterations (question reformulation) and each iteration, in turn, to the number of terms added to the original question. In all cases system B (which allows the user to see the complete documents) required fewer resources.

Table 3. Averaged Time per Question in seconds (excluding more than 5 minutes)

	Time per Question
System A	146.6
System B	115.8
Total	198.9

Table 4. Terms Added per Question (excluding question during more than 5 minutes)

System A
0 terms added 33
2 terms added 11
System B
0 terms added 51
2 terms added 8

Table 5. Reformulation or Iterations per Question (excluding questions during more than 5 minutes)

System A
0 iterations 33
2 iterations 11
System B
0 iterations 51
2 iterations 8

Table 6. Terms added per Question

Fails
0 terms added 20
2 terms added 26
6 terms added 2
Hits
0 terms added 71
2 terms added 9

3 Results

The evaluation of the *i-track* can be carried out in two ways: strict and lenient. The latter is more favourable, and also more realistic. For example, the answer to question 12 can take diverse forms, all of them correct: UDF, French Democratic Union, UDF party, etc. Since the users were not instructed as to a specific way of expressing the answer, it seemed more appropriate to take into consideration the lenient evaluation.

In any case, the superiority of system B is quite clear; this means that the possibility of accessing the complete document based on the passages retrieved is more successful (an improvement of 28.75 %), which was expected, but not to that degree. But there are more interesting data: we can see that the possibility of adding the terms suggested by the system to the question, was not greatly employed by the users. When they did do so, few terms were added. 91 of the 128 questions (16×8 users) had no term added to them, and only 2 terms were added in 35. Of the 80 questions answered correctly (with both systems), terms were only added in 9. However, keywords were added to more than half of the incorrect questions.

At first glance this seems to indicate that the suggestion of terms is not effective to obtaining correct answers. However, of the 48 failed questions, 28 were not answered at all or took up all 5 minutes of the time. Of these 28, 17 had 2 terms added, and in 2 of them up to 6 terms each were added. This leads us to think that the users did not value the utility of adding suggested terms and that they only applied this possibility when they had difficulty finding the answers. The iterations are related to the addition of terms, since the only way to reformulate a question is to change terms.

The results, however, should be interpreted taking into account the nature of the documents and also that of the questions. The documents, being news items from the press, tended to be short and have a single theme; their fragmentation into passages, therefore, was probably of little interest. The questions were also short; furthermore, they contained proper names or other terms that either did not need translation or the translation was obvious, even for a machine translation system. In fact, a manual revision of the machine translation showed that in general they were fairly correct. Thus, adding or removing terms was not of interest since they started from correct translations.

4 Conclusions

We have explored user interaction by finding terms related to those of the questions and suggesting them to the user. We also worked with passages but allowed the user to obtain complete documents as well. The possibility of adding related terms was not highly rated by the users, who considered it of little use. However, the possibility to obtain and see the complete documents, based on the passages retrieved, improved the number of correct answers, as well as the time required to obtain a correct answer. Nevertheless, the results should be viewed taking into account the small number of documents as well as the conciseness of the questions.

References

1. Vicedo, J.L.: Recuperación de información de alta precisión: los sistemas de búsqueda de respuestas. Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN), Alicante (2003)

2. Woods, W.A., Green, S., Martin, P., Houston, A.: Halfway to question answering. In: The Ninth Text REtrieval Conference (TREC 9), Gaithersburg, Maryland (USA), NIST Special Publication 500-249 (2000) 489–500
3. Dorr, B.J., He, D., Luo, J., Oard, D.W., Schwartz, R., Wang, J., Zajic, D.: iCLEF 2003 at Maryland: Translation selection and document selection. In Peters, C., ed.: Results of the CLEF 2003 Cross-Language System Evaluation Campaign. Working Notes for the CLEF 2003 Workshop, Trondheim, Norway (2003)
4. López-Ostenero, F., Gonzalo, J., Verdejo, F.: UNED at CLEF 2003: Searching cross-language summaries. In Peters, C., ed.: Results of the CLEF 2003 Cross-Language System Evaluation Campaign. Working Notes for the CLEF 2003 Workshop, Trondheim, Norway (2003)
5. Zazo, Á.F., Figuerola, C.G., Berrocal, J.L.A., Rodríguez, E., Gómez, R.: Experiments in term expansion using thesauri in Spanish. In: Advances in Cross-Language Information Retrieval. Third Workshop of the Cross-Language Evaluation Forum, CLEF 2002, Rome, Italy. September, 2002 Revised Papers, Springer-Verlag (LNCS num. 2785) (2003) 301–310
6. Salton, G., Yang, C.S.: On the specification of term values in automatic indexing. *Journal of Documentation* **29** (1973) 351–372
7. Softissimo. (in <http://www.softissimo.com>)
8. Salton, G.: The SMART Retrieval System. Experiments in Automatic Document Processing. Prentice Hall, Englewoods Cliffs, N. J. (1971)
9. Zazo, Á.F.: Técnicas de expansión en los sistemas de recuperación de información. PhD thesis, Departamento de Informática y Automática. Universidad de Salamanca (2003)
10. Zazo, Á.F., Figuerola, C.G., Berrocal, J.L.A., Rodríguez, E.: Reformulation of queries using similarity thesauri. *Information Processing & Management* (2004) In Press
11. Belkin, N.J., Cool, C., Kelly, D., Lin, S.J., Park, S.Y., Pérez-Carballo, J., Sikora, C.: Iterative exploration, design and evaluation of support for query reformulations in interactive information retrieval. *Information Processing & Management* **37** (2001) 403–434