

REINA at the iCLEF 2004

Carlos G. Figuerola, Angel F. Zazo, José L. Alonso Berrocal, 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 interaction with the user. Users can provide aid to the system reformulating questions, adding information to these or selecting the documents with which it must operate. This year our group has explored the effects of the interaction with the user in suggesting terms to add to the question; also, operating only with passages of documents or with complete documents. The experiments carried out are described and the results are discussed

1 Introduction

Several tasks on IR require interaction with the user. In the case of the so called Question Answering, the aim is to provide concrete answers to concrete informational needs. Some approaches to this problem are based on the retrieval of text passages or fragments [7, 8], in that one assumes that is the answer. This one can be extracted by means of automatic systems or interacting with the user. If the system is not able to provide a valid answer in a totally autonomous way, a feedback with the user must be taken off, so that this one can achieve its informational need.

Of another side, the proposed task in this year in the *i-track* is cross-lingual. In this case, the interaction with the user usually deals with the translation of the questions, not with documents directly [2]. The questions are in a language, and the documents (or passages) are in (at least) another one different. This implies the necessity to translate the question to the same one language of documents. An alternative approach is to translate documents to the language of the questions [3], although usually this is more expensive.

The activity of our group has been centered, this year, in exploring the effects of two ways of interaction with the user: of a side, doing that the system it suggests to him a set of terms translated to the language of documents. Of another one, allowing the system operate not only with passages, but, on demand of the user, with complete documents. In both cases the target is to evaluate not only the number of correct answers, but also the subjective appreciation on the part of the user of this class of aids.

This work is organized as follows: first, it is provided a description of the task to make; later, the used system of retrieval is described, as well as the design of the experiment. Finally, the results are described.

2 Design of the experiment

2.1 The proposed task

The task for this year is Cross Lingual Question Answering. The initial stage is the following one: a document collection in English, and questions in Spanish. The users have as native language the Spanish and have passive knowledge of the English; this passive knowledge allows them to partially understand documents in English, as well as to include/understand some words or terms in this language.

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

Table 1: Averages of Pre-Experiments Questionnaire

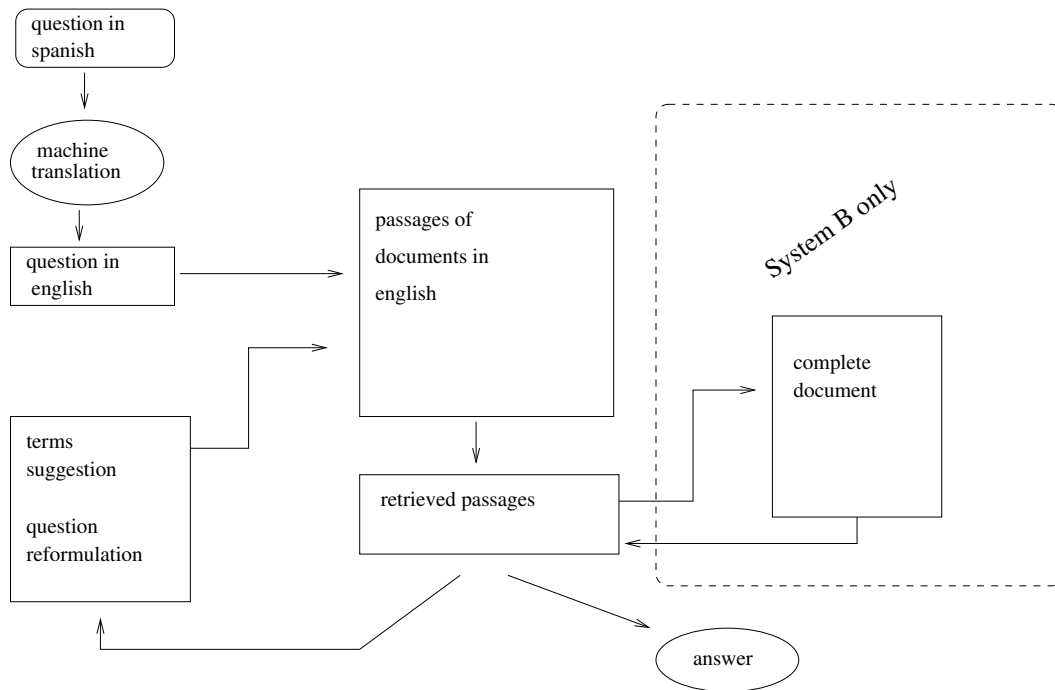


Figure 1: Layout of the Retrieval System

All the users selected for the experiment (8 altogether) were students of Information Science, accustomed to work with point-and-click interfaces, as well as to achieve searches on automated catalogue of libraries and to use World Wide Web search services. Its experience in the use of Machine Translation software, nevertheless, was enough minor (see Table 1).

2.2 The Retrieval System

In fact we lack a true system of Question Answering. In its place we used a conventional IR system, based on the vectorial model, with some adaptations. The general idea is:

- to translate the question by means of a machine translation program
- to carry out a conventional retrieval from the translated question
- the user reads retrieved documents and deduces the answer to the question

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

Table 2: Documents Collection and Passage Division

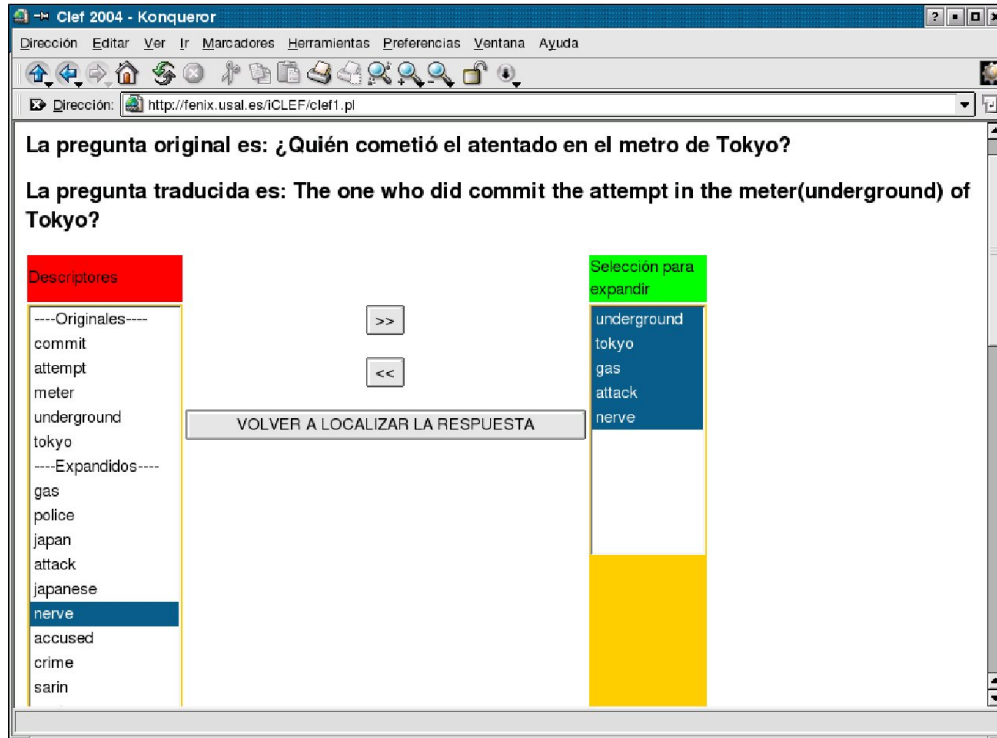


Figure 2: Screen of Suggesting Terms

The base IR system is the same one that we have used in previous participation in CLEF [11], with a classic term weight calculation based on $tf \times idf$ [5]. On this base they have been added and modified some things:

- The software of automatic translation used is Reverso [6] (<http://www.reverso.com>). The version we use its the offered in a free way to the public from the web page of a well known newspaper (<http://www.elmundo.es/traductor>); the quality of his translations is lightly superior to that of other similar programs, and has the characteristic of offering several alternative translations of the same term. No training nor specialized dictionaries are available in this version, of course.
- the basic unit of indexation is not the document, but the passages or fragments of this one. The division in passages has been made dividing the documents in windows of 100 words, including stop words. Only the field TEXT of documents has been used. The list of stop words is the standard list of SMART [4] plus the words that they appear in more of the 15 % of documents. The final average number of words by document is smaller than 100 due to final fragments of documents; it is necessary to consider that the documents of collections CLEF, being press news, they are relatively small (see Table 2).

	Time per Question
System A	146.6
System B	115.8
Total	198.9

Table 3: Averaged Time per Question in seconds (excluding 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 4: Terms Added per Question (excluding question during more than 5 minutes)

- the capacity to suggest terms to the user has been added to the system, so that this one can add these terms to the question. The suggested terms are in English (the language of documents); the idea is that these terms could improve the automatic translation of the questions; one hopes that, since the users they have a passive knowledge of the language of documents, are in conditions to include/understand part of those terms. The terms suggested for each question are obtained by means of techniques of expansion of terms. Our group of investigation has a great experience in expansion of queries applied to classic tasks of IR [11, 9, 10]. For that reason, we have used the expansion technique that offers the best results: the use of thesaurus of local association. The relations of co-occurrences of terms in the first retrieved documents are used to construct the thesaurus. Using this technique of expansion the best terms related to all the terms of the original question are obtained. For each question the best 30 related terms are obtained, and they are showed to the user so that it can, if it wishes it, incorporate them to the original question. The mechanism of suggestion of terms are frequently used in interactive experiments [1].
- The access and interaction with the system are carried out through a Web server and several forms. The most important characteristics are the possibility of making several iterations (to reformulate the translated question, to examine recovered passages, etc) and the possibility to obtain and to read the complete document from the retrieved passages.

2.3 The experiment

All the users were undergone a process of previous training. For the experiment itself, the retrieval system was prepared in two different ways (**system A** and **system B**); the second way allows to accede to complete documents, while first no.

All the users were asked to make the 16 questions: half with system A and other half with system B. Nevertheless, the order of the questions and systems was organized so that half of the users began by the system A and the other half by system B. Each user worked with a sequence of different questions.

All the operations of the users were logged by the system. The sessions or questions of more than 5 minutes of duration were considered like nonvalid (19.5 %) The duration of each session seems to be in relation to the number of iterations (reformulations of questions), and this one with the one of terms added to the original translated question. In all these cases, the System B (that allows to visualize the complete documents) required less resources.

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

Table 5: Reformulation or Iterations por Question (excluding questions during more than 5 minutes)

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

Table 6: Terms added per Question

3 Results

The evaluation of *i-track* comes in two flavours: strict and lenient. Lenient is more favorable, but also more realistic; for example, the answer to question 12 can take diverse forms, all of them correct ones: UDF, Union for French Democracy, UDF center-right party, etc. Since the users were not instructed on a concrete way of express the answer, it seems that the lenient results are more convenient.

In any case, a superiority of **System B** is appraised clearly; that is to say, the possibility of acceding to the complete document, starting off of the retrieved passages, it produces more successes (an improvement of the 28.75 %), which was awaited, although perhaps not of a so important way. But there are more interesting data: in that concerns about the possibility of adding suggested terms, we see that, in general, the users have used little this possibility; and when they have done it, they have added few terms. Of the 128 questions (16×8 users) in 91 of them terms were not added, and in 35 they were added only 2 terms. Of the 80 correct questions (with both systems), only in 9 were added terms; in the failed questions, nevertheless, more than half it had added terms.

At first sight, this seems to indicate that suggesting terms is no useful to obtain answers. Nevertheless, of the 48 failures, 28 are NIL or timeout. Of these 28, in 17, 2 terms were added; and in 2 of them up to 6 terms. What it seems to indicate that the users do not appreciate the usefulness of suggesting terms, and they only use this possibility when they have difficulties in find out the answers. Iterations are related with adding terms, as the only sense of reformulating a question is to change terms.

Results, nevertheless, must be viewed in the lighth of the nature of documents and the questions themshelves. Documents, as press news, are in general short and mono-topic; fragmentation in passages can result of minor interest. Question are short, too; in addition, they have proper nouns or other terms without or with obvious translation. In fact, a manual examination of the automatic translations of the questions showed their correctness. So, adding or removing terms is of minor interest.

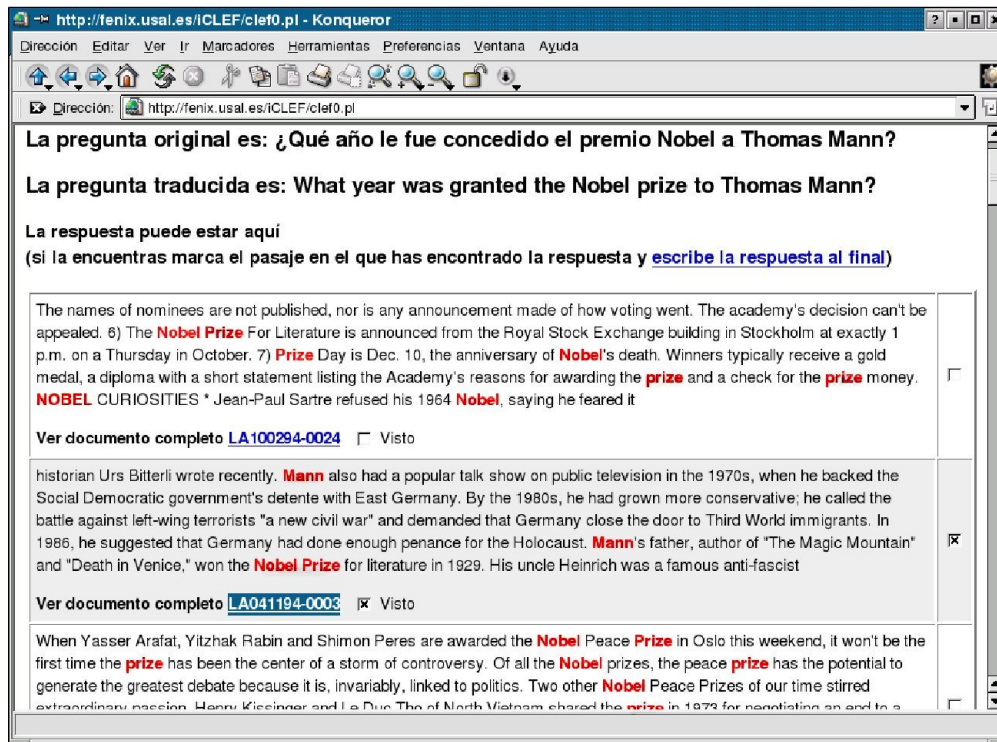


Figure 3: Retrieved Passages from Question 1

4 Conclusions

We have explored the interaction with the user towards finding related terms with the questions and suggesting them to the user; and working with passages but allowing to the user to get the complete documents. While term suggestion shows a low appreciation by the user, getting complete documents rather only passages improves not only the hits, but the time required to achieve a correct answer. Nevertheless, we must to take in account the small size of the documents used in the experiments, as well as the terseness in the questions.

References

- [1] N. J. Belkin, C. Cool, D. Kelly, S.-J. Lin, S. Y. Park, J. Pérez-Carballo, and C. Sikora. Iterative exploration, design and evaluation of support for query reformulations in interactive information retrieval. *Information Processing & Management*, 37(3):403–434, 2001.
- [2] B. J. Dorr, D. He, J. Luo, D. W. Oard, R. Schwartz, J. Wang, and D. Zajic. iCLEF 2003 at Maryland: Translation selection and document selection. In C. Peters, editor, *Results of the CLEF 2003 Cross-Language System Evaluation Campaign. Working Notes for the CLEF 2003 Workshop*, Trondheim, Norway, 2003.
- [3] F. López-Ostenero, J. Gonzalo, and F. Verdejo. UNED at CLEF 2003: Searching cross-language summaries. In C. Peters, editor, *Results of the CLEF 2003 Cross-Language System Evaluation Campaign. Working Notes for the CLEF 2003 Workshop*, Trondheim, Norway, 2003.
- [4] G. Salton. *The SMART Retrieval System. Experiments in Automatic Document Processing*. Prentice Hall, Englewoods Cliffs, N. J., 1971.

- [5] G. Salton and C. S. Yang. On the specification of term values in automatic indexing. *Journal of Documentation*, 29(4):351–372, 1973.
- [6] Softissimo. in <http://www.softissimo.com>.
- [7] J. L. Vicedo. *Recuperación de información de alta precisión: los sistemas de búsqueda de respuestas*. Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN), Alicante, 2003.
- [8] W. A. Woods, S. Green, P. Martin, and A. Houston. Halfway to question answering. In *The Ninth Text REtrieval Conference (TREC 9)*, pages 489–500, Gaithersburg, Maryland (USA), 2000. NIST Special Publication 500-249.
- [9] Á. F. Zazo. *Técnicas de expansión en los sistemas de recuperación de información*. PhD thesis, Departamento de Informática y Automática. Universidad de Salamanca, Sept. 2003.
- [10] Á. F. Zazo, C. G. Figuerola, J. L. A. Berrocal, and E. Rodríguez. Reformulation of queries using similarity thesauri. *Information Proccession & Management*, 2004. Article in press.
- [11] Á. F. Zazo, C. G. Figuerola, J. L. A. Berrocal, E. Rodríguez, and R. Gómez. Experiments in term expansion using thesauri in Spanish. In *Advances in Cross-Language Information Retrieval. Third Workshop of the Cross-Languge Evaluation Forum, CLEF 2002, Rome, Italy. September, 2002 Revised Papers*, pages 301–310. Springer-Verlag (LNCS num. 2785), 2003.