University of Ottawa's participation in the CL-SR task at CLEF 2006

Muath Alzghool and Diana Inkpen

School of Information Technology and Engineering
University of Ottawa
{alzghool,diana}@site.uottawa.ca

Abstract This paper presents the second participation of the University of Ottawa group in CLEF, the Cross-Language Spoken Retrieval (CL-SR) task. We present the results of the submitted runs for the English collection and very briefly for the Czech collection, followed by many additional experiments. We have used two Information Retrieval systems in our experiments: SMART and Terrier were tested with many different weighting schemes for indexing the documents and the queries and with several query expansion techniques (including a new method based on log-likelihood scores for collocations). Our experiments showed that query expansion methods do not help much for this collection. We tested whether the new Automatic Speech Recognition transcripts improve the retrieval results; we also tested combinations of different automatic transcripts (with different estimated word error rates). The retrieval results did not improve, probably because the speech recognition errors happened for the words that are important in retrieval, even in the newer ASR2006 transcripts. By using different system settings, we improved on our submitted result for the required run (English queries, title and description) on automatic transcripts plus automatic keywords. We present cross-language experiments, where the queries are automatically translated by combining the results of several online machine translation tools. Our experiments showed that high quality automatic translations (for French) led to results comparable with monolingual English, while the performance decreased for the other languages. Experiments on indexing the manual summaries and keywords gave the best retrieval results.

Categories and Subject Descriptors

H.3 [Information Storage and Retrieval]: H.3.1 Content Analysis and Indexing; H.3.3 Information Search and Retrieval; H.3.4 Systems and Software

General Terms

Measurement, Performance, Experimentation

Keywords

Information retrieval, speech recognition transcripts, indexing schemes, automatic translation

1 Introduction

This paper presents the second participation of the University of Ottawa group in CLEF, the Cross-Language Spoken Retrieval (CL-SR) track. We briefly describe the task [10]. Then, we present our systems, followed by results for the submitted runs for the English collection and very briefly for the Czech collection. We present results for many additional runs for the English collection. We experiment with many possible weighting schemes for indexing the documents and the queries, and with several query expansion techniques. We test with different speech recognition transcripts to see if the word error rate has an impact on the retrieval performance. We describe cross-language experiments, where the queries are automatically translated from French, Spanish, German and Czech into English, by combining the results of several online machine translation (MT) tools. At the end we present the best results when summaries and manual keywords were indexed.

The CLEF-2006 CL-SR collection includes 8104 English segments, and 105 topics (queries). Relevance judgments were provided for 63 training topics, and later for 33 test topics. In each document (segment), there are six fields that can be used for the official runs: ASRTEXT2003A, ASRTEXT2004A, ASRTEXT2006A, ASRTEXT2006B, AUTOKEYWORD2004A1, and AUTOKEYWORD2004A2. The first four fields are transcripts produced using Automatic Speech Recognition (ASR) systems developed by the IBM T. J. Watson Research Center in three successive years 2003, 2004, and 2006, with different estimated mean word error rates of 44%, 38%, and 25% respectively.

Among the 8104 segments covered by the test collection, only 7377 segments have the ASRTEXT2006A field. The ASRTEXT2006B field content is identical to the ASRTEXT2006A field if there is ASR output produced by the 2006 system for the segment, or identical to the ASRTEXT2004A if not. Moreover just 7034 segments have ASRTEXT2003A field. The AUTOKEYWORD2004A1 and AUTOKEYWORD2004A2 field contain a set of thesaurus keywords that were assigned automatically using two different k-Nearest Neighbor (kNN) classifiers using only words from the ASRTEXT2004A field of the segment. Among the 8104 segments covered by the test collection, 8071 and 8090 segments have AUTOKEYWORD2004A1 and AUTOKEYWORD2004A2, respectively

There is also a Czech collection for this year's CL-SR track; the document collection consists of ASR transcripts for 354 interviews in Czech, together with some manually assigned metadata and some automatically generated metadata, and 115 search topics in two languages (Czech and English). The task for this collection is to return a ranked list of time stamps marking the beginning of sections that are relevant to a topic.

2 System Overview

The University of Ottawa Cross-Language Information Retrieval (IR) systems were built with off-the-shelf components. For translating the queries from French, Spanish, German, and Czech into English, several free online machine translation tools were used. Their output was merged in order to allow for variety in lexical choices. All the translations of a title made the title of the translated query; the same was done for the description and narrative fields. For the retrieval part, the SMART [2,9] IR system and the Terrier [1,6] IR system were tested with many different weighting schemes for indexing the collection and the queries.

For translating the topics into English we used several online MT tools. The idea behind using multiple translations is that they might provide more variety of words and phrases, therefore improving the retrieval performance. The seven online MT systems that we used for translating from Spanish, French, and German were:

1. http://www.google.com/language_tools?hl=en
2. http://www.babelfish.altavista.com
3. http://freetranslation.com
4. http://www.wordlingo.com/en/products_services/wordlingo_translator.html
5. http://www.systranet.com/systran/net
6. http://www.online-translator.com/srvurl.asp?lang=en
7. http://www.freetranslation.paralink.com

For translation the Czech language topics into English we were able to find only one online MT system: http://intertran.tranexp.com/Translate/result.shtml.

We combined the outputs of the MT systems by simply concatenating all the translations. All seven translations of a title made the title of the translated query; the same was done for the description and narrative fields. We used the combined topics for all the cross-language experiments reported in this paper.

3 Retrieval

We used two systems in our participation: SMART and Terrier. SMART was originally developed at Cornell University in the 1960s. SMART is based on the vector space model of information retrieval [2]. It generates weighted term vectors for the document collection. SMART preprocesses the documents by tokenizing the text into words, removing common words that appear on its stop-list, and performing stemming on the remaining words to derive a set of terms. When the IR server executes a user query, the query terms are also converted into weighted term vectors. Vector inner-product similarity computation is then used to rank documents in decreasing order of their similarity to the user query. The newest version of SMART (version 11) offers many state-of-the-art options for weighting the terms in the vectors. Each term-weighting scheme is described as a combination of term frequency, collection frequency, and length normalization components [8].

In this paper we employ the notation used in SMART to describe the combined schemes: xxx.xxx. The first three characters refer to the weighting scheme used to index the document collection and the last three characters refer to the weighting scheme used to index the query fields. In SMART, we used mainly the lnn.ntn weighting scheme which performs very well in CLEF-CLSR 2005 [4]; lnn.ntn means that lnn was used for documents and ntn for queries according to the following formulas:

$$weight_{\ln n} = \ln(tf) + 1.0$$

$$weight_{\text{ntn}} = tf \times \log \frac{N}{n_t}$$

Where *tf* denote the term frequency of a term *t* in the document or query, N denotes the number of documents in the collection, and n_t denotes the number of documents in which the term t occurs.

We have also used a query expansion mechanism with SMART, which follows the idea of extracting related words for each word in the topics using the Ngram Statistics Package (NSP) [7]. We extracted the top 6412 pairs of related words based on log likelihood ratios (high collocation scores in the corpus of ASR transcripts), using a window size of 10 words. We chose log-likelihood scores because they are known to work well even when the text corpus is small. For each word in the topics, we added the related words according to this list. We call this approach to relevance feedback SMARTnsp.

Terrier was originally developed at University of Glasgow. It is based on Divergence from Randomness models (DFR) where IR is seen as a probabilistic process [1, 6]. We experimented with the In(exp)C2 weighting model, one of Terrier's DFR-based document weighting models. Using the In(exp)C2 model, the relevance score of a document *d* for a query *q* is given by the formula:

$$sim(d,q) = \sum_{t \in q} qtf \,.\, w(t,d)$$

where
- *qtf* is the frequency of term *t* in the query *q*,
- *w(t,d)* is the relevance score of a document *d* for the query term *t*, given by:

$$w(t,d) = \left(\frac{F+1}{n_t \times (tfn_e + 1)}\right) \times \left(tfn_e \times \log_2 \frac{N+1}{n_e + 0.5}\right)$$

where

-*F* is the term frequency of t in the whole collection.

-*N* is the number of document in the whole collection.

-n_t is the document frequency of t.

-n_e is given by $n_e = N \times (1 - (\frac{1-n_t}{N})^F)$

- ***tfn**$_e$* is the normalized within-document frequency of the term *t* in the document *d.* It is given by the normalization 2 [1, 3]:

$$tfn_e = tf \times \log_e \left(1 + c \times \frac{avg\_l}{l}\right)$$

where
- *c* is a parameter, for the submitted run, we fix this parameter to 1.
- *tf* is the within-document frequency of the term *t* in the document *d*.
- *l* is the document length and *avg_l* is the average document length in the whole collection.

We estimated the parameter *c* of the normalization 2 formula by running some experiments on the training data, to get the best values for c depending on the topic fields used. We obtained the following values: c=0.75 for queries using the Title only, c=1 for queries using the Title and Description fields, and c=1 for queries using the Title, Description, and Narrative fields. We select the c value that has a best MAP score according to the training data.

We have also used a query expansion mechanism in Terrier, which follows the idea of measuring divergence from randomness. In our experiments, we applied the Kullback-Leibler (KL) model for query expansion [4, 10]. It is one of the Terrier DFR-based term weighing models. Using the KL model, the weight of a term *t* in the *top*-ranked documents is given by:

$$w(t) = P_x \times \log_2 \frac{P_x}{P_c}$$

where

$$P_x = \frac{tfx}{lx} \text{ and } P_c = \frac{F}{token_c}$$

-*tfx* is the frequency of the query term in the top-ranked documents.
-*lx* is the sum of the length of the top-ranked documents,
-F is the term frequency of the query term in the whole collection.
- $token_c$ is the total number of tokens in the whole collection.

4 Experimental Results

4.1 Submitted runs

Table 1 shows the results of the submitted results on the test data (33 queries). The evaluation measure we report is the standard measure computed with the trec_eval script: MAP (Mean Average Precision). The information about what fields of the topic were indexed is given in the column named Fields: T for title only, TD for title + description, TDN for title + description + narrative. For each run we include an additional description of the experimental settings and which document fields were indexed. For the uoEnTDt04A06A and uoEnTDNtMan runs we used the indexing scheme In(exp)C2 from Terrier; and for uoEnTDNsQEx04, uoFrTDNs, and uoSpTDNs we used the indexing scheme lnn.ntn from SMART. We used SMARTnsp query expansion for the uoEnTDNsQEx04 run, KL query expansion for uoEnTDNtMan and uoEnTDt04A06A, and we didn't use any query expansion techniques for uoFrTDNs and uoSpTDNs.

Table 1. Results of the five submitted runs, for topics in English, French, and Spanish. The required run (English, title + description) is in bold.

Language	Run	MAP	Fields	Description
English	uoEnTDNtMan	0.2902	TDN	Terrier: MANUALKEYWORD + SUMMARY
English	uoEnTDNsQEx04	0.0768	TDN	SMART: NSP query expansion ASRTEXT2004A + AUTOKEYWORD2004A1, A2
French	uoFrTDNs	0.0637	TDN	SMART: ASRTEXT2004A + AUTOKEYWORD2004A1, A2
Spanish	uoSpTDNs	0.0619	TDN	SMART: ASRTEXT2004A + AUTOKEYWORD2004A1, A2
English	uoEnTDt04A06A	**0.0565**	TD	Terrier: ASRTEXT2004A + ASRTEXT2006A + AUTOKEYWORD2004A1, A2

We also participated in the task for Czech language. We indexed the Czech topics and ASR transcripts. Table 2 shows the results of the submitted runs on the test data (29 topics) for the Czech collection. The evaluation measure we report is the mean General Average Precision (GAP), which rewards retrieval of the right timestamps in the collection. MAP scores could not be used because the speech transcripts were not segmented. A default segmentation was provided: one document was produced for every minute of the interview.

From our results, that used the default segmentation, we note:

- The mean GAP is very low for all submitted runs (for all teams).
- There is a big improvement when we indexed the field ENGLISHMANUKEYWORD relative to the case when we indexed CZECHMANUKEYWORD; this means we loose a lot due to the translation tool used to translate the ENGLISHMANUKEYWORD field into Czech.
- No improvements if CZECHMANUKEYWORD in added to the ASR field.
- Terrier's results are slightly better than SMART's for the required run.

In the rest of the paper we focus only on the Eglish CL-SR collection.

Table 2. Results of the five submitted runs for Czech collection. The required run (English, title + description) is in bold.

Language	Run	GAP	Fields	Description
Czech	uoCzEnTDNsMan	0.0039	TDN	SMART: ASRTEXT, CZECHAUTOKEYWORD, CZECHMANUKEYWORD, ENGLISH MANUKEYWORD, ENGLISHAUTOKEYWORD
Czech	uoCzTDNsMan	0.0005	TDN	SMART: ASRTEXT, ZECHAUTOKEYWORD, CZECHMANUKEYWORD
Czech	uoCzTDNs	0.0004	TDN	SMART: ASRTEXT, CZECHAUTOKEYWORD
Czech	uoCzTDs	**0.0004**	TD	SMART: ASRTEXT, CZECHAUTOKEYWORD
Czech	uoCzEnTDt	**0.0005**	TD	Terrier: ASRTEXT, CZECHAUTOKEYWORD

4.2 Comparison of systems and query expansion methods

Table 3 presents results for the best weighting schemes: for SMART we chose lnn.ntn and for Terrier we chose the In(exp)C2 weighting model, because they achieved the best results on the training data. We present results with and without relevance feedback.

According to Table 3, we note that:

- Relevance feedback helps to improve the retrieval results in Terrier for TDN, TD, and T for the training data; the improvement was high for TD and T, but not for TDN. For the test data there is a small improvement.
- NSP relevance feedback with SMART does not help to improve the retrieval for the training data (except for TDN), but it helps for the test data (small improvement).
- SMART results are better than Terrier results for the test data, but not for the training data.

Table 3.Results (MAP scores) for Terrier and SMART, with or without relevance feedback, for English topics. In bold are the best scores for TDN, TD, and T.

	System	**Training**			**Test**		
		TDN	**TD**	**T**	**TDN**	**TD**	**T**
1	SMART	**0.0954**	0.0906	0.0873	0.0766	0.0725	0.0759
	SMARTnsp	0.0923	0.0901	0.0870	**0.0768**	**0.0754**	**0.0769**
2	Terrier	0.0913	0.0834	0.0760	0.0651	0.0560	0.0656
	TerrierKL	0.0915	**0.0952**	**0.0906**	0.0654	0.0565	0.0685

4.3 Comparison of retrieval using various ASR transcripts

In order to find the best ASR transcripts to use for indexing the segments, we compared the retrieval results when using the ASR transcripts from the years 2003, 2004, and 2006 or combinations. We also wanted to find out if adding the automatic keywords helps to improve the retrieval results. The results of the experiments using Terrier and SMART are shown in Table 4 and Table 5, respectively.

We note from the experimental results that:

- Using Terrier, the best field is ASRTEXT2006B which contains 7377 transcripts produced by the ASR system on 2006 and 727 transcripts produced by the ASR system in 2004, this improvement over using only the ASRTEXT2004A field is very. On the other hand, the best ASR field using SMART is ASRTEXT2004A.
- Any combination between two ASRTEXT fields does not help to improve the retrieval.
- Using Terrier and adding the automatic keywords to ASRTEXT2004A improved the retrieval for the training data but not for the test data. For SMART it helps for both the training and the test data.
- In general, adding the automatic keywords helps. Adding them to ASRTEXT2003A or ASRTEXT2006B improved the retrieval results for the training and test data.
- For the required submission run English TD, the maximum MAP score was obtained by the combination of ASRTEXT 2004A and 2006A plus autokeywords using Terrier (**0.0952)** or SMART (**0.0932**) on the

training data; on the test data the combination of ASRTEXT 2004A and autokeywords using SMART obtained the highest value, **0.0725**, higher than the value we report in Table 1 for the submitted run.

Table 4. Results (MAP scores) for Terrier, with various ASR transcript combinations. In bold are the best scores for TDN, TD, and T.

	Terrier					
Segment fields	**Training**			**Test**		
	TDN	TD	T	TDN	TD	T
ASRTEXT 2003A	0.0733	0.0658	0.0684	0.0560	0.0473	0.0526
ASRTEXT 2004A	0.0794	0.0742	0.0722	**0.0670**	0.0569	0.0604
ASRTEXT 2006A	0.0799	0.0731	0.0741	0.0656	0.0575	0.0576
ASRTEXT 2006B	0.0840	0.0770	0.0776	0.0665	**0.0576**	0.0591
ASRTEXT 2003A+2004A	0.0759	0.0722	0.0705	0.0596	0.0472	0.0542
ASRTEXT 2004A+2006A	0.0811	0.0743	0.0730	0.0638	0.0492	0.0559
ASRTEXT 2004A+2006B	0.0804	0.0735	0.0732	0.0628	0.0494	0.0558
ASRTEXT 2003A+ AUTOKEYWORD2004A1,A2	0.0873	0.0859	0.0789	0.0657	0.0570	0.0671
ASRTEXT 2004A+ AUTOKEYWORD2004A1, A2	0.0915	**0.0952**	0.0906	0.0654	0.0565	0.0685
ASRTEXT 2006B+ AUTOKEYWORD2004A1,A2	**0.0926**	0.0932	0.0909	0.0717	0.0608	0.0661
ASRTEXT 2004A+2006A+ AUTOKEYWORD2004A1, A2	0.0915	**0.0952**	**0.0925**	0.0654	0.0565	**0.0715**
ASRTEXT 2004A+2006B+ AUTOKEYWORD2004A1,A2	0.0899	0.0909	0.0890	0.0640	0.0556	0.0692

Table 5. Results (MAP scores) for Terrier, with various ASR transcript combinations. In bold are the best scores for TDN, TD, and T.

	SMART					
Segment fields	**Training**			**Test**		
	TDN	TD	T	TDN	TD	T
ASRTEXT 2003A	0.0625	0.0586	0.0585	0.0508	0.0418	0.0457
ASRTEXT 2004A	0.0701	0.0657	0.0637	0.0614	0.0546	0.0540
ASRTEXT 2006A	0.0537	0.0594	0.0608	0.0455	0.0434	0.0491
ASRTEXT 2006B	0.0582	0.0635	0.0642	0.0484	0.0459	0.0505
ASRTEXT 2003A+2004A	0.0685	0.0646	0.0636	0.0533	0.0442	0.0503
ASRTEXT 2004A+2006A	0.0686	0.0699	0.0696	0.0543	0.0490	0.0555
ASRTEXT 2004A+2006B	0.0686	0.0713	0.0702	0.0542	0.0494	0.0553
ASRTEXT 2003A + AUTOKEYWORD2004A1,A2	0.0923	0.0847	0.0839	0.0674	0.0616	0.0690
ASRTEXT 2004A+ AUTOKEYWORD2004A1,A2	**0.0954**	0.0906	0.0873	**0.0766**	**0.0725**	**0.0759**
ASRTEXT 2006B+ AUTOKEYWORD2004A1,A2	0.0869	0.0892	0.0895	0.0650	0.0659	0.0734
ASRTEXT 2004A+ 2006A + AUTOKEYWORD2004A1,A2	0.0903	**0.0932**	0.0915	0.0654	0.0654	0.0777
ASRTEXT 2004A +2006B + AUTOKEYWORD2004A1,A2	0.0895	0.0931	**0.0919**	0.0652	0.0655	0.0742

4.4 Cross-language experiments

Table 6 presents results for the combined translation produced by the seven online MT tools, from French, Spanish, and German into English, for comparison with monolingual English experiments (the first line in the table). All the results in the table are from SMART using the lnn.ntn weighting scheme.

Since the result of combined translation for each language was better than when using individual translations from each MT tool on the CLEF 2005 CL-SR data [4], we used combined translations in our experiments.

The retrieval results for French translations were very close to the monolingual English results, especially on the training data. On the test data, the results were much worse when using only the titles of the topics, probably because the translations of the short titles were less precise. For translations from the other languages, the retrieval results deteriorate rapidly in comparison with the monolingual results. We believe that the quality of the French-English translations produced by online MT tools was very good, while the quality was lower for Spanish, German and Czech, successively.

Table 6. Results of the cross-language experiments, where the indexed fields are ASRTEXT2004A, and AUTOKEYWORD2004A1, A2 using SMART with the weighting scheme lnn.ntn.

	Language	**Training**			**Test**		
		TDN	**TD**	**T**	**TDN**	**TD**	**T**
1	English	0.0954	0.0906	0.0873	0.0766	0.0725	0.0759
2	French	0.0950	0.0904	0.0814	0.0637	0.0566	0.0483
3	Spanish	0.0773	0.0702	0.0656	0.0619	0.0589	0.0488
4	German	0.0653	0.0622	0.0611	0.0674	0.0605	0.0618
5	Czech	0.0585	0.0506	0.0421	0.0400	0.0309	0.0385

4.5 Manual summaries and keywords

Table 7 presents the results when only the manual keywords and the manual summaries were used. The retrieval performance improved a lot, for topics in all the languages. The MAP score jumped from 0.0654 to 0.2902 for English test data, TDN, with the In(exp)C2 weighting model in Terrier. The results of cross-language experiments on the manual data show that the retrieval results for combined translation for French and Spanish language were very close to the monolingual English results on training data and test data. For all the experiments on manual summaries and keywords, Terrier's results are better than SMART’s.

Table 7.Results of indexing the manual keywords and summaries, using SMART with weighting scheme lnn.ntn, and Terrier with (In(exp)C2).

	Language and System	**Training**			**Test**		
		TDN	**TD**	**T**	**TDN**	**TD**	**T**
1	English SMART	0.3097	0.2829	0.2564	0.2654	0.2344	0.2258
2	English Terrier	**0.3242**	**0.3227**	**0.2944**	**0.2902**	**0.2710**	**0.2489**
3	French SMART	0.2920	0.2731	0.2465	0.1861	0.1582	0.1495
4	French Terrier	0.3043	0.3066	0.2896	0.1977	0.1909	0.1651
5	Spanish SMART	0.2502	0.2324	0.2108	0.2204	0.1779	0.1513
6	Spanish Terrier	0.2899	0.2711	0.2834	0.2444	0.2165	0.1740
7	German SMART	0.2232	0.2182	0.1831	0.2059	0.1811	0.1868
8	German Terrier	0.2356	0.2317	0.2055	0.2294	0.2116	0.2179
9	Czech SMART	0.1766	0.1687	0.1416	0.1275	0.1014	0.1177
10	Czech Terrier	0.1822	0.1765	0.1480	0.1411	0.1092	0.1201

5 Conclusion

We experimented with two different systems: Terrier and SMART, with various weighting scheme for indexing the document and query terms. We proposed a new approach for query expansion that uses collocations with high log-likelihood ratio. Used with SMART, the method obtained a small improvement on test data (probably not significant). The KL relevance feedback method produced only small improvements with Terrier on test data. So, query expansion methods do not seem to help for this collection.

The improvements of mean word error rates in the ASR transcripts (of ASRTEXT2006A relative to ASRTEXT2004A) did not improve the retrieval results. Also, combining different ASR transcripts (with different error rates) did not seem to help.

For some experiments, Terrier was better than SMART, for other it was not; therefore we cannot clearly choose one or another IR system for this collection.

The idea of using multiple translations proved to be good. More variety in the translations would be beneficial. The online MT systems that we used are rule-based systems. Adding translations by statistical MT tools might help, since they could produce radically different translations.

On the manual data, the best MAP score we obtained is around 29%, for the English test topics. On automatically-transcribed data the best result is around 7.6% MAP score. Since the improvement in the ASR word error rate does not improve the retrieval results, as shown from the experiments in section 4.3, we think that the justification for the difference to the manual summaries is due to the fact that summaries contain different words to represent the content of the segments. In future work we plan to investigate methods of removing or correcting some of the speech recognition errors in the ASR contents and to use speech lattices for indexing.

References

1. G. Amati and C. J. van Rijsbergen : Probabilistic models of information retrieval based on measuring the divergence from randomness. ACM Transactions on Information Systems (TOIS), 20(4):357-389, October 2002.
2. C. Buckley, G. Salton, and J. Allan : Automatic retrieval with locality information using SMART. In Proceedings of the First Text REtrieval Conference (TREC-1), pages 59–72. NIST Special Publication 500-207, March 1993.
3. C. Carpineto, R. de Mori, G. Romano, and B. Bigi. An information-theoretic approach to automatic query expansion. ACM Transactions on Information Systems (TOIS), 19(1): 1-27, January 2001.
4. D. Inkpen, M. Alzghool, and A. Islam : Using various indexing schemes and multiple translations in the CL-SR task at CLEF 2005. In Proceedings of CLEF 2005, Lecture Notes in Computer Science 4022, Springer-Verlag, 2006.
5. D. W. Oard, D. Soergel, D. Doermann, X. Huang, G. C. Murray, J. Wang, B. Ramabhadran, M. Franz and S. Gustman : Building an Information Retrieval Test Collection for Spontaneous Conversational Speech, in Proceedings of SIGIR, 2004.
6. I. Ounis, G. Amati, V. Plachouras, B. He, C. Macdonald and D. Johnson : Terrier Information Retrieval Platform. In Proceedings of the 27th European Conference on Information Retrieval (ECIR 05), 2005. http://ir.dcs.gla.ac.uk/wiki/Terrier
7. Pedersen. and S. Banerjee : The design, implementation and use of the ngram statistics package., Proceedings of the Fourth International Conference on Intelligent Text Processing and Computational Linguistics, Mexico City, Mexico, 2003.
8. G. Salton : Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer. Addison-Wesley Publishing Company, 1989.
9. G. Salton and C. Buckley : Term-weighting approaches in automatic retrieval. Information Processing and Management 24(5): 513-523, 1988.
10. R. W. White, D. W. Oard, G. J. F. Jones, D. Soergel, X. Huang : Overview of the CLEF-2005 Cross-Language Speech Retrieval Track. In Proceedings of CLEF 2005, Lecture Notes in Computer Science 4022, Springer-Verlag, 2006.