

EXETER AT CLEF 2002: Experiments with Machine Translation for Monolingual and Bilingual Retrieval

Adenike M. Lam-Adesina, Gareth J. F. Jones
Department of Computer Science
University of Exeter EX4 4QF
United Kingdom
{A.M.Lam-Adesina, G.J.F.Jones}@ex.ac.uk

Abstract

This year, the University of Exeter participated in both the CLEF 2002 monolingual and bilingual task for two languages: Italian and Spanish. We submitted 4 ranked results each for both Italian and Spanish Monolingual tasks and 5 each for the bilingual tasks. We report experimental results from our investigations of merging topic translations from two machine translation (MT) systems and recent experimental results for query expansion and term weighting from alternative collections. Our results show that although, query expansion and term weighting from a pilot collection has been shown to be effective in improving retrieval performance in information retrieval, the performance can be affected negatively if the lexicon of the pilot and the test collection differ.

1 Introduction

The main objective of our participation in CLEF 2002 was to test the effectiveness of some of our methods developed after CLEF 2001, and also to investigate the retrieval behavior of document collection in Italian and Spanish using topic sets in several other languages. Our official submissions included Italian and Spanish monolingual tasks as well as bilingual tasks using topic sets in German, French, English, Italian and Spanish. To present a fair comparison of results across all language pairs and methods for the bilingual runs, we used the same translation resources for each pair and we also present results for each retrieval method for each pair. Both the collections and the topics were translated from the source language into English. Firstly, because we had intended to participate in the multilingual task and secondly because the retrieval system we used could not deal with accented words. We were unable to submit results for the multilingual task because of time constraints.

Our general approach was to use the collection and topic translation strategy for CLIR. The document collection and the topic statements were submitted to the selected MT system, the output was then collected and applied on the information retrieval (IR) system. For all our submissions and subsequent runs presented in this paper, we used both the Systran Version: 3.0 and the Globalink Power Translation Pro Version: 6.4 MT systems for topic translation. It should be noted that the two collections used in our experiments were translated using only Systran Version 3.0.

Pseudo-relevance feedback (PRF) has been shown to be an effective approach to improving retrieval performance in IR and also in CLIR [1][2][3]. In our experimental work in [4][5] we demonstrated the effectiveness of a new PRF method using the Okapi BM25 probabilistic model [6]. In this work we investigated the idea of selecting expansion terms for document summaries and found this method to be more reliable than query expansion from full documents. Since CLEF 2001, we have also explored data combination techniques that merge the output of the two MT systems for the topics, and use this as the initial query set. Furthermore, we have also been investigating the use of a comparable collection (pilot) for generating expansion terms and term weighting. The method is described fully below. Our experiments for CLEF 2002 explore the effectiveness of these methods with automatically translated documents and topics.

The remainder of this paper is organized as follows: Section 2 reviews the information retrieval methods used, Section 3 gives a brief description of the data processing techniques used, Section 4 describes the different methods of PRF, Section 5 gives the experimental results and section 6 concludes the paper.

2 Retrieval Approach

The experiments were carried out using the City University research distribution version of the Okapi system. All stopwords were removed from the documents and search queries. All remaining terms were then suffix stripped using Porter stemming [7] and then indexed using a small set of synonyms.

Documents terms are weighted using the Okapi BM25 formula [6] reproduced below.

$$cw(i, j) = \frac{cfw(i) \times tf(i, j) \times (K1 + 1)}{K1 \times ((1 - b) + (b \times ndl(j))) + tf(ij)}$$

where $cw(i, j)$ = the weight of term i in document (j) ,
 $cfw(i)$ = the standard collection frequency weight
 $tf(i, j)$ = the document term frequency
 $ndl(j)$ = the normalized document length calculated as follows

$$ndl(j) = \frac{dl(j)}{\text{Averaged } dl \text{ for all documents}}$$

where $dl(j)$ = the length of j

$K1$ and b are empirically selected tuning constants for a particular collection. $K1$ modifies the effect of term frequency and b modifies the effect of document length. All our experiments were done with $K1$ and b set to 1.4 and 0.6. The parameters were set using the CLEF 2001 data sets.

2.1 Relevance Feedback

Relevance feedback is a method used to improve retrieval effectiveness by either improving the query terms (Query modification) or the term weights (term-reweighting). All our experiments used query expansion to modify the query to attempt to improve the quality of the initial query by adding new terms selected from a pool of potential expansion terms from the initial retrieval run.

Our query expansion method selects terms from summaries of the top 5 ranked documents. The summaries were generated using the method described in [4]. The summary generation method combines the Luhn's Keyword Cluster Method [8], Title terms frequency method [4], Location/header method [9] and the Query-bias method [10] to form an overall significance score for each sentence. For all our experiments we used the top 6 ranked sentences as the summary of each document. From this summary we collected all non-stopwords and ranked them using a slightly modified version of the Robertson selection value (rsv) [11] reproduced below. The top 20 terms was then selected in all our experiments.

$$rsv(i) = rw(i) \times r(i)$$

where $r(i)$ = number of relevant documents containing term i
 $rw(i)$ is the standard Robertson/Sparck Jones relevance weight [12] reproduced below

$$rw(i) = \log \frac{(r(i) + 0.5)(N - n(i) - R + r(i) + 0.5)}{(n(i) - r(i) + 0.5)(R - r(i) + 0.5)}$$

where $n(i)$ = the total number of documents containing term i
 $r(i)$ = the total number of relevant documents term i occurs in
 R = the total number of relevant documents for this query
 N = the total number of documents

In our modified version, although potential expansion terms are selected from the summaries of the top 5 ranked documents, they are ranked using the top 20 ranked documents from the initial run.

3 Data Processing

The two document collections used in our experiments, Italian and Spanish were translated to English using the Systran Translation Software version 3.0. This was necessitated by the inability of the retrieval system (Okapi) used for our experiments to deal with accented terms as well as languages other than English. All queries were translated from the source language into English using both the Systran Version 3.0 and Globalink Power Translation Pro version 6.4 MT software. All our experiments were done using both the title and description fields of the CLEF topics.

4 Procedures

Our submissions to CLEF 2002 investigated a number of approaches to term weighting and query expansion as described below.

4.1 Standard Method

This method is the same as that used in our CLEF 2001 official submissions [5]. Initial retrieval run using translated queries was performed. The top 5 assumed relevant documents were summarized and the pool of potential expansion terms was generated from the summaries. The top 20 terms was then added to the initial query for the feedback run. (Indicated “Test coll. terms and weight” in the results).

4.2 Pilot searching

Query expansion is aimed at improving initial search topic in order to make it a better expression of user’s information need. This is normally achieved by adding terms selected from assumed relevant document retrieved from the test collection, to the initial query. Another approach that has been shown to be effective is the selection of expansion terms from a larger collection, a subset of which would be the test collection. Based on the assumption that if additional documents from the same corpus as the test collection are available, these can be used for improved query expansion, we explore the idea of pilot searching [13]. The larger data collection is likely to enable more accurate parameter estimation and hopefully better retrieval and document ranking. The Okapi submissions for the TREC-7 [13] adhoc tasks used the TREC disks 1-5 of which the TREC-8 data is a subset, for parameter estimation and query expansion. The method was found to be very effective. Our post CLEF 2001 results for bilingual English also demonstrated the effectiveness of this approach [14]. The TREC-8 data collection consisting of more than half a million documents was used as “pilot collection” in our experiments. The CLEF 2002 English collection is a subset of the TREC-8 data collection. Two different approaches were taken to the pilot searching procedure. They are as follows

- 1 Apply the original query terms on the pilot collection using the Okapi system without feedback. Extract terms from the top R assumed relevant documents; rank the extracted terms and select the desired number of expansion terms from the top of the list. The corresponding $cfw(i)$ term weights are also stored along with the expansion terms. The expansion terms are added to the initial query terms and applied on the test collection (Test coll. weight and Pilot coll. expansion terms). This approach is shown to give an improvement for the CLEF 2001 bilingual task [1].
- 2 The second method involve using the expansion terms from the pilot collection as above but this time the $cfw(i)$ weights from the pilot collection are used instead of the term weights from the test collection (Pilot coll. terms and weight). This method gave a further improvement for the CLEF 2001 bilingual task.

4.3 Combination Methods

MT systems sometimes make translation errors due to the limitations of the dictionaries used in them. However, this problem can be tackled by combining the outputs of multiple MT systems. This idea is based on the proven notion that combination of evidence from multiple information sources is beneficial to text retrieval. Thus, in this method, for each untranslated query, two different translations of the query from the two different translators used in these experiments are merged into a single query.

Furthermore, the merged queries are then used in two different ways as follows.

- 1 The first method uses only the combined queries as the initial query (Combined MT queries)
- 2 The second (exemgcnt) uses the combined queries and upweight the weight of terms occurring in both translation by 2 (Combined MT upweighted).

5 Experimental Results

In this section we report the results of our investigation for all methods described above. Baseline results without feedback and results after the application of the different methods of feedback are presented for Italian and Spanish Monolingual and Bilingual tasks. All our official submissions are indicated by a *. In all cases the results use the Title and Description fields of the search topics and we present the average precision (Avep), the % change in average precision relative to the baseline for the MT system used (% chg) and the total number of relevant documents retrieved (R-ret). For feedback runs all initial query terms are upweighted by multiplying the original term weights by 3.5.

5.1 Italian Monolingual runs

Run-id	Systran MT			Globalink MT		
	Avep	% chg	R-ret	Avep	% chg	R-ret
<i>Baseline no feedback</i>	388	-	966	349	-	853
<i>Test coll. weight, pilot coll. expansion terms</i>	*414	6.70%	993	378	8.31%	910
<i>Test coll. term and weight</i>	<u>*453</u>	16.75%	1004	<u>394</u>	12.89%	905
<i>Pilot coll.term and weight</i>	376	-3.09%	910	375	7.45%	880

Table 1: Retrieval results for topic translation for Systran and Globalink MT, showing results before and after application of various method of feedback.

<i>Run-id</i>	<i>Avep</i>	<i>% chg systran</i>	<i>% chg Globalink</i>
<i>Combined MT queries</i>	*421	8.51%	20.63%
<i>Combined MT queries upweighted</i>	*411	5.93%	17.77%

Table 2: Retrieval results for combined queries from both translators with summary-based expansion term selection and cfw(i) weight from test collection.

Table 1 above shows the results for Italian Monolingual task before and after various method of feedback are applied. Application of the different feedback methods resulted in improvement in precision except for method using the pilot collection for term selection and weighting for Systran translated topics. Investigation of this result shows that the lexicon of the translated query using Systran is usually different from that of the TREC-8 data used as the pilot collection. Also there were some query terms, which were not translated and were returned as the original Italian word. This resulted in a large number of mismatches between the query and the documents in the pilot collection. The Globalink translator fared better in this aspect because the untranslated query was closer after translation to the original English query and hence resulted in better selection and parameter estimation from the pilot collection. The feedback result was however affected by the differences in the lexicon used in the expanded query and the target collection.

Systran translated topics gave better performance compared to Globalink topics, this is perhaps attributable to the fact that the document collection was translated using Systran.

The best result (underlined) for both translations is achieved using the test collection generated expansion terms and their corresponding weights.

Table 2 shows the effect of the application of merged queries from the two MT systems, on the retrieval system. Although the method resulted in about 8% improvement relative to the baseline result for Systran and about 21% relative to the baseline for Globalink, the result is still lower than that achieve by the best method. The result however suggests that merged queries may be useful in achieving better retrieval performance for poor MT systems.

5.2 Italian Bilingual runs

Run-id	Systran MT			Globalink MT		
	Avep	% chg	R-ret	Avep	% chg	R-ret
<i>Baseline no feedback</i>						
<i>German</i>	293	-	830	305	-	777
<i>Spanish</i>	319	-	856	337	-	874
<i>French</i>	324	-	917	310	-	908
<i>English</i>	330	-	898			

Table 3: Results showing baseline results for Italian bilingual tasks for both MT systems.

Run-id	Systran MT			Globalink MT		
<i>Test coll. weight, pilot expansion terms)</i>	Avep	% chg	R-ret	Avep	% chg	R-ret
<i>German</i>	347	18.43%	914	364	19.34%	867
<i>Spanish</i>	362	13.47%	904	379	12.46%	936
<i>French</i>	373	15.12%	965	359	15.81%	968
<i>English</i>	407	23.33%	976			

Table 4: Retrieval results for topic translation using both MT systems with summary-based expansion term selection from pilot collection and cfw(i) from test collection.

Run-id	Avep	% chg Systran	% chg Globalink
<i>Combined MT queries for</i>			
<i>German</i>	377	28.67%	23.60%
<i>Spanish</i>	*373	16.93%	10.68%
<i>French</i>	*348	7.41%	12.26%

Table 5: Retrieval results for merged queries from the two MT systems with summary-based expansion term selection and cfw(i) from test collection.

Run-id	Avep	% chg Systran	% chg Globalink
<i>Combined MT queries upweighted</i>			
<i>German</i>	*368	25.60%	20.66%
<i>Spanish</i>	365	14.42%	8.31%
<i>French</i>	377	16.36%	21.61%

Table 6: Retrieval results for merged queries from the two MT systems with summary-based expansion term selection and cfw(i) from test collection. Terms occurring in both translations are upweighted by 2.

Run-id	Systran MT			Globalink MT		
<i>Test coll. terms and weight</i>	Avep	% chg	<i>R-ret</i>	Avep	% chg	<i>R-ret</i>
<i>German</i>	341	16.38%	869	377	28.67%	825
<i>Spanish</i>	363	13.79%	906	371	16.30%	922
<i>French</i>	375	15.74%	947	358	11.11%	955
<i>English*</i>	374	13.33%	938			

Table 7: Retrieval results for topic translation for both MT systems with summary-based expansion term selection and cfw(i) from test collection .

Run-id	Systran MT			Globalink MT		
<i>Pilot coll. terms and weight</i>	Avep	% chg	R-ret	Avep	% chg	R-ret
<i>German</i>	335	14.33%	872	325	6.56%	851
<i>Spanish</i>	333	4.39%	830	360	6.82%	917
<i>French</i>	365	12.65%	929	352	13.55%	948
<i>English*</i>	399	20.91%	968			

Table 8: Retrieval results for topic translation for both MT systems with summary-based expansion term selection and cfw(i) from pilot collection

Table 3-8 shows the retrieval behavior for the language pairs before and after various methods of feedback for Italian bilingual runs. Result using pilot collection for query expansion and term weighting again gives the worst result for all pairs overall. This is almost certainly due to the differences in the lexicon used in the pilot and the test collection, which reduces the query-document matching and subsequently result in reduction in retrieval effectiveness. The result for the combined queries further strengthens the idea that combining the output from two MT systems might help in reducing the effect of poor MT systems on retrieval. The best result overall is given by the Italian-English run using test collection weight and pilot collection expansion terms. Globalink translated topics gave better performance overall compared to the results using Systran translated topics although French topics using Systran performed better than that using Globalink.

5.3 Spanish Monolingual runs

Run-id	Systran MT			Globalink MT		
	Avep	% chg	R-ret	Avep	% chg	R-ret
<i>Baseline no feedback</i>	442	-	2413	419	-	2235
<i>Test coll. weight, pilot coll. expansion terms</i>	*473	7.01%	2538	466	11.22%	2412
<i>Test coll. terms and weight</i>	*475	7.46%	2517	445	6.21%	2266
<i>Pilot coll. terms and weight</i>	420	-4.98%	2249	431	2.86%	2295

Table 9: Retrieval results for topic translation for Spanish bilingual runs using Systran and Globalink MT, showing results before and after application of various method of feedback.

Run-id	Avep	%chg Systran	%chg Globalink
Combined MT queries	*470	6.33%	2479
Combined MT queries upweighted	*468	5.88%	2460

Table 10: Retrieval results for combined queries from both translators using summary-based expansion term selection and cfw(i) weight from test collection

Table 9 and 10 monolingual Spanish results shows the same trend as the Italian monolingual results in Table1-2 above. Results again show that using pilot collection to estimate term weight and query expansion for Systran topics is not very effective, it resulted in about 5% reduction in average precision compared to the baseline. The combined query method (Table 10) is also shown to reduce the negative effect of translation output on retrieval.

5.4 Spanish Bilingual runs

Run-id	Systran MT			Globalink MT		
	Avep	% chg	R-ret	Avep	% chg	R-ret
<i>Baseline no feedback</i>						
<i>German</i>	298	-	1853	340	-	1868
<i>Italy</i>	331	-	1926	339	-	1847
<i>French</i>	357	-	2041	377	-	2144
<i>English</i>	371	-	2149			

Table 11: Spanish Bilingual baseline retrieval results for topic translation from both MT systems.

Run-id	Systran MT			Globalink MT		
	Avep	% chg	R-ret	Avep	% chg	R-ret
<i>Test coll. Weight and pilot coll. Expansion terms</i>						
<i>German</i>	329	10.40%	2180	379	11.47%	2156
<i>Italian</i>	369	11.48%	2110	399	17.70%	2108
<i>French</i>	390	9.24%	2237	411	9.01%	2345
<i>English</i>	426	14.82%	2345			

Table 12 above: Retrieval results for topic translation using both MT systems with summary-based expansion term selection from pilot collection and cfw(i) from test collection

Run-id	Avep	% chg Systran	% chg Globalink
<i>Combined MT queries</i>			
<i>German</i>	359	20.47%	5.59%
<i>Italian</i>	354	6.95%	4.42%
<i>French</i>	419	17.36%	11.14%

Table 13: Retrieval results for topic translation using merged queries from both MT systems with summary-based expansion term selection from test collection and cfw(i) from test collection.

Run-id	Avep	% chg Systran	% chg Globalink
<i>Combined MT queries upweighted</i>			
<i>German</i>	*354	18.79%	4.12%
<i>Italian</i>	*379	14.50%	11.80%
<i>French</i>	*414	15.97%	9.81%

Table 14: Retrieval results for topic translation using merged queries from both MT systems with summary-based expansion term selection from test collection and cfw(i) from test collection. Terms that occur in both translations are upweighted by 2.

Run-id	Systran MT			Globalink MT		
<i>Test coll. terms and weight</i>	<i>Avep</i>	<i>% chg</i>	<i>R-ret</i>	<i>Avep</i>	<i>% chg</i>	<i>R-ret</i>
<i>German</i>	318	6.71%	1901	387	29.87%	1975
<i>Italian</i>	359	8.46%	2025	369	16.81%	2040
<i>French</i>	382	7.00%	2069	414	11.48%	2326
<i>English*</i>	412	11.05%	2289			

Table 15: Retrieval results for topic translation using both MT systems with summary-based expansion term selection from test collection and cfw(i) from test collection.

Run-id	Systran MT			Globalink MT		
<i>Pilot coll. terms and weight</i>	<i>Avep</i>	<i>% chg</i>	<i>R-ret</i>	<i>Avep</i>	<i>% chg</i>	<i>R-ret</i>
<i>German</i>	334	12.08%	2152	350	2.94%	2124
<i>Italian</i>	342	3.55%	1914	370	9.14%	2075
<i>French</i>	373	4.48%	2268	396	5.04%	2316
<i>English*</i>	2372	13.20%	2372			

Table 16: Retrieval results for topic translation using both MT systems with summary-based expansion term selection and cfw(i) from pilot collection.

Results for Spanish bilingual runs (Table 11-16) show that Globalink translated topics gave better performance overall compared to results using Babelfish translated topics. The Spanish to English pair using summary-based expansion term selection from pilot collection and cfw(i) from test collection gave the best results overall. Combining the queries from the two MT systems (Table 13 and 14) is again shown to be effective in reducing degradation in performance brought about by poor MT output.

5.5 Further Runs

Query expansion and term weighting from a pilot collection has been shown to be very effective in information retrieval. The results above however suggests otherwise. This is probably due to the differences in the language of the pilot and the test collection. To test this theory, we did some further runs; the pilot and the test collection were merged to form a single collection. This merged collection was then used as the pilot collection, i.e. for query expansion and term weighting. The expanded query and the corresponding weight is then applied on the test collection. The tables below show the effect of this method on retrieval. In all cases we present the result for both Systran and Globalink translated topics.

Run-id	Systran MT			Globalink MT		
<i>Merged coll. terms and weight</i>	<i>Avep</i>	<i>% chg</i>	<i>R-ret</i>	<i>Avep</i>	<i>% chg</i>	<i>R-ret</i>
<i>Monolingual Italian</i>	442	13.92%	980	384	10.02%	872
<i>Monolingual Spanish</i>	490	10.86%	2513	478	14.08%	2478

Table 17 Retrieval results for Monolingual Italian and Spanish using summary-based term selection and cfw(i) from merged collection.

Run-id	Systran MT			Globalink MT		
<i>Merged coll. weight and expansion terms</i>	<i>Avep</i>	<i>% chg</i>	<i>R-ret</i>	<i>Avep</i>	<i>% chg</i>	<i>R-ret</i>
<i>German</i>	346	18.09%	906	372	21.97%	874
<i>Spanish</i>	371	16.30%	904	378	12.17%	922
<i>French</i>	371	14.51%	952	366	18.06%	970
<i>English</i>	406	23.03%	978			

Table 18 Retrieval results for Bilingual Italian using summary-based term selection and cfw(i) from merged collection.

Run-id	Systran MT			Globalink MT		
	Avep	% chg	R-ret	Avep	% chg	R-ret
<i>Merged coll. terms and weight</i>						
<i>German</i>	358	20.13%	2151	392	15.29%	2083
<i>Italian</i>	345	4.23%	2097	359	5.90%	2070
<i>French</i>	383	7.28%	2186	407	7.96%	2288
<i>English</i>	443	19.41%	2384			

Table 19 Retrieval results for Bilingual Spanish using summary based term selection and cfw(i) from merged collection.

Table 17-19 shows the results for using the merged collection (pilot and test) for query expansion and term weighting. Monolingual Results for Italian and Spanish shows that merging the two collections result in better estimation of term weight, which results in improved retrieval. This method resulted in about 14% improvement in retrieval compared to the baseline results and about 18% improvements over using the pilot collection for expansion and weighting for the Monolingual runs.

For the bilingual runs, an improvement of about 23% compared to the baseline runs for the Italian Bilingual run and about 20% for the Spanish bilingual runs

6 Conclusions and Further Work

In this paper we have presented our results for the CLEF 2002 monolingual and bilingual Italian and Spanish retrieval tasks. The results suggest that good retrieval results can be achieved by merging the output of two commercially available MT systems. It also shows that all language pairs behave very differently to different feedback method, this requires further investigation to determine the causes of such behavior and how they can be tackled. The combined query method is very effective in smoothing out the negative effects of bad translations in most cases. Using pilot collection to estimate term weight and for query expansion although shown to be very effective in [13], the results shown here suggests that when there is a difference in the language of the pilot and the test collection the method might not be as effective. We show that further improvements can be achieved by merging the two collections to form a pilot collection.

Further investigation is needed to determine the reason for the slightly poor performance of the Systran translated queries compared to the Globalink translated queries. We also noticed that some terms were left untranslated by the MT systems, this is more predominant in the Systran translations, and might have been reason for the lower performance achieved using the topics translated using Systran MT compared to the performance for Globalink topics in the bilingual results.

References

- [1] G.J.F. Jones, T. Sakai, N. H. Collier, A. Kumano and K. Sumita. A Comparison of Query Translation Methods for English-Japanese Cross-Language Information Retrieval. In Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pages 269-270, San Francisco, 1999. ACM.
- [2] L. Ballesteros and W. B. Croft. Phrasal Translation and Query Expansion Techniques for Cross-Language Information Retrieval. In Proceedings of the 20th Annual International ACM SIGIR conference on Research and Development in Information Retrieval, pages 84-91, Philadelphia, 1997. ACM.
- [3] G.Salton and C. Buckley. Improving Retrieval performance by Relevance Feedback. Journal of the American Society for Information Science, pages 288-297, 1990.
- [4] A.M. Lam-Adesina and G.J.F. Jones. Applying Summarization Techniques for Term Selection in Relevance Feedback. In Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pages 1-9, New Orleans, 2001. ACM.
- [5] G.J.F. Jones and A.M. Lam-Adesina. Exeter at CLEF 2001: Experiments with Machine Translation for Bilingual Retrieval. In Proceedings of the CLEF 2001: Workshop on Cross-Language Information Retrieval and Evaluation, pages 59-77, Darmstadt, Germany, 2002.

- [6] S.E. Robertson and S. Walker. Some simple effective approximations to the 2-Poisson model for probabilistic weighted retrieval. In Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pages 232-241, Dublin, 1994. A.C.M.
- [7] M.F. Porter. An algorithm for suffix stripping. *Program*, 14:10-137, 1980.
- [8] H.P. Luhn. The Automatic Creation of Literature Abstracts. *IBM Journal of Research and Development*, 2(2):159-165, 1958.
- [9] H.P. Edmundson. New Methods in Automatic Abstracting. *Journal of the ACM*, 16(2):264-285, 1969
- [10] A. Tombros and M. Sanderson. The Advantages of Query-Biased Summaries in Information Retrieval. In proceedings of the 21st Annual International ACM SIGIR Conference Research and Development in Information Retrieval, pages 2-10, Melbourne, 1998. ACM.
- [11] S.E. Robertson. On term selection for query expansion. *Journal of Documentation*, 46:359-364, 1990.
- [12] S.E. Robertson and K. Sparck Jones. Relevance weighting of search terms. *Journal of the American Society for Information Science*, 27(3):129-146, 1976.
- [13] S.E. Robertson, S. Walker, and M. M. Beaulieu. Okapi at TREC-7: automatic ad hoc, filtering, VLS and interactive track. In E. Voorhees and D.K. Harman, editors, Overview of the Seventh Text REtrieval Conference (TREC-7), pages 253-264. NIST, 1999.
- [14] G.J.F. Jones and A.M. Lam-Adesina. Combination Methods for improving the Reliability of Machine Translation Based Cross-Language Information Retrieval. Proceedings of the 13th Irish Conference on Artificial Intelligence and Cognitive Science, September 2002. To appear.