Bilingual Information Retrieval: CLEF-2000 Experiments

Jacques Savoy

Institut interfacultaire d'informatique, Université de Neuchâtel, Switzerland
To appear in "ECSQARU-2001 Workshop on Management of uncertainty and imprecision in multimedia information systems"

Abstract. Given that in the Web environment, non-English documents and users tend to represent the majority, our main objective is to analyze and evaluate the retrieval effectiveness of various indexing and search strategies based on test-collections written in four different languages (English, French, German and Italian). Our second aim is to describe and evaluate various approaches that might be implemented in order to effectively access bilingual collections.

1 Introduction

As the amount of information stored on the Web increases at an amazing pace, we need an effective access to documents written in languages other than English. Moreover, for the majority of Internet users, English is not their first language. Thus, there is a real need to promote retrieval systems that can provide access to information without being hampered by language or cultural barriers.

In order to promote multilingual retrieval, the Cross-Language Evaluation Forum (CLEF) was founded and one of its goals was to elaborate various non-English test-collections, some of which will be used in this paper. This paper is organized as follows: Chap. 2 describes monolingual information retrieval systems dealing with document corpora written in English, French, Italian and German. Chapter 3 illustrates and evaluates various approaches used to resolve bilingual information retrieval problems. In this case, original queries written in English are translated into another language and the search is done on documents written in this target language.

2 Evaluation of Indexing and Searching Strategies

Most European languages (including French, Italian, German) share many of the same characteristics with the language of Shakespeare. Any adaptation of indexing or search strategies thus means the elaboration of general stopword lists and fast stemming procedures. This chapter will deal with these issues, and is organized as follows: Sect. 2.1 contains an overview of our four test-collections while Sect. 2.2 describes our general approach to build stopword lists and stemmers for use with languages other than English. Section 2.3 depicts the various vector-processing schemes used in this paper together with the Okapi probabilistic model. Finally, Sect. 2.4 evaluates these search models using four test-collections and queries written in English, French, Italian and German.

2.1 Overview of the Test-Collections

One of the main outcomes of the first CLEF workshop was the elaboration of various test-collections that would be available in four different languages (a few statistics describing these test-collections are listed in Table 1). With these corpora are included from 33 to 40 queries (see Table 2), where the requests, instead of being limited to a narrow subject range, reflect a diversity of information needs (such as "architecture in Berlin", "drug use and soccer" or "solar temple"). Given that most Web queries are relatively short, our experiments are mainly based on information contained in the title section, and our requests have an average length of 2.75 indexing terms. This limited query size also reflects the situation experienced by users whose foreign language knowledge is not enough to effectively formulate long requests. Of course in other environments, the length of queries submitted may be longer, and thus, we also studied queries based on the content of the descriptive and narrative logical sections.

The corpora used in our experiments included newspapers such as the *Los Angeles Times*, *Le Monde* (French), *La Stampa* (Italian) and *Der Spiegel* and *Frankfurter Rundschau* (German). They are similar in content and subject matter and were extracted during the same year (1994). As shown in Table 1, these corpora are of various sizes, with the English and German collections being twice the volume of the French and Italian sources. On the other hand, the mean number of distinct indexing terms per document is relatively similar across the corpora (around 180), and this number is a little bit smaller for the German corpus (145.66).

The data in Table 1 also illustrates that the mean number of relevant documents per request can vary across the corpora. The German collection resulted in the greatest number of relevant items and the greatest mean per query (22.19). The Italian corpus on the other hand contained fewer relevant papers (338) and the mean number of pertinent items per request (9.94) was also smaller.

During indexing, we retained only the following logical sections: <TITLE>, <HEADLINE>, <TEXT>, <LEAD> and <LEAD1> and we ignored other logical sections from the original documents. From topic descriptions, we automatically removed certain phrases such as "Relevant document report ", "Find documents that give ", "Trouver des documents qui parlent ", "Sono valide le discussioni e le decisioni " or "Relevante Dokumente berichten ".

2.2 Stopword Lists and Stemming Procedures

We defined a general stopword list made up of words determined to be of no use during retrieval, but very frequently found in document content. These stopword lists were developed for two main reasons: Firstly, we hoped that each match between a query and a document would be based on pertinent indexing

Table 1. Test-collection statistics

	English	French	Italian	German
size (in MB)	425 MB	157 MB	193 MB	383 MB
# of documents	113,005	44,013	58,051	153,694
# of distinct indexing terms / document				
mean	167.33	189.34	182.23	145.66
maximum	1,812	1,724	1,405	2,593
minimum	2	4	12	1
# of queries	33	34	34	37
# of relevant documents	579	528	338	821
mean	17.55	15.53	9.94	22.19
maximum	51 (#q:11)	62 (#q:5)	42 (#q:7)	101 (#q:5)
minimum	1 (#q:4)	1 (#q:22)	1 (#q:21)	1 (#q:6)

Table 2. Query examples

<num> C012	(both query translations failed in German)
<E-title>	Solar Temple
<D-title manually translated>	Sonnentempel
<D-title SYSTRAN>	SolarBügel
<D-title BABYLON>	solar Heiligtum
<num> C026	(both query translations failed in Italian)
<E-title>	Use of Wind Power
<I-title manually translated>	Impiego dell'energia eolica
<I-title SYSTRAN>	Uso di pontenza del vento
<I-title BABYLON>	usare di vento efficacia
<num> C037	(only SYSTRAN query failed in French)
<E-title>	Sinking of the Estonia
<F-title manually translated>	Naufrage du ferry-boat Estonia
<F-title SYSTRAN>	Descente de l'Estonie
<F-title BABYLON>	naufrage de le Estonie

terms. Non-significant words represent noise, and may actually damage retrieval performance because they do not discriminate between relevant and irrelevant documents. Secondly, by using them we could reduce the size of the inverted file, hopefully in the range of 30% to 50%.

English and French stopword lists were already available [1], [2]. For German and Italian, we established a general stopword list by following the guidelines described in [1]. Firstly, we sorted all word forms appearing in our corpora according to their frequency of occurrence and extracted the 200 most frequently occurring words. Secondly, we inspected these lists in order to remove all numbers, plus all nouns and adjectives more or less directly related to the main subjects of the underlying collections (e.g., the German word "Prozent" (ranking 69) or the Italian noun "Italia" (ranking 87)). Thirdly, we included some words containing no information, even if they did not appear in the first 200 most frequent words. For example, we added various personal or possessive pronouns (such as "meine" ("my" in German), prepositions ("nello" ("in the" in Italian)) and conjunctions ("où" ("where" in French)). The presence of homographs represents another debatable issue, and to some extent, we had to make arbitrary decisions concerning their inclusion in stopword lists. For example, the French word "son" can be translated as "sound" or "his". In our experiments we used the stoplist provided by the SMART system (571 English words), 217 French words, 431 Italian words, 294 German words (available at http://www.unine.ch/info/clef/).

After removing high frequency words, an indexing procedure tries to conflate word variants into the same stem or root using a stemming algorithm. In developing this procedure for the French, Italian and German languages, it is important to remember that these languages have more rich and complex morphologies than does the English language [3]. As a first approach, we intended to remove only inflectional suffixes so that singular and plural word forms or feminine and masculine variants conflate to the same root. Such a "quick and dirty" stemming procedure has already been developed for the French language [2]. Based on these same ideas, we implemented a stemming algorithm for the Italian and German languages (C code available at http://www.unine.ch/info/clef/). In Italian, the main inflectional rule is to modify the final character (e.g., "-o", "-a" or "-e") into another (e.g., "-i", "-e"). As a second rule, Italian morphology can also alter the final two letters (e.g., "-io" in "-o", "-co" in "-chi", "-ga" in "-ghe"). In German, a few rules can be applied to obtain the plural form of words (e.g., "Lied" into "Lieder" (song), "Mutter" into "Mütter" (mother), "Buch" into "Bücher" (book)). The suggested algorithms do not account for the morphological variations found in verbs or other derivational suffixes. In a previous study [2], we have tried various more complex stemmers for French languge without being able to improve the retrieval effectiveness over our simple suffix-stripping approach.

2.3 Indexing and Searching Strategies

In order to define a retrieval model, we will first explain how documents and queries are represented and then how these representations are compared, thus resulting in a ranked list of retrieved items. Given that French and Italian morphology is comparable to that of English, we decided to index French and Italian documents based on word stems. For the German language and its more complex compounding morphology, we decided to use a 5-gram approach [4]. This value of 5 was chosen for two reasons; it returns a better performance, and, on the other hand, it is closed to the mean word length of our German corpora (mean word length: 5.87; standard error: 3.7).

As a first approach, we adopted a binary indexing scheme within which each document or request are represented by a set of keywords without any weight. To measure the similarity between documents and requests, we compute the inner product (retrieval model denoted "document=bnn, query=bnn" ou "bnn–bnn"), see Table 3 for details.

Binary logical restrictions are often too limiting, and the use of term frequency (denoted by tf) allows a better term distinction and increases indexing flexibility (retrieval model notation: "nnn–nnn"). In a third IR model [5], those terms that do occur very frequently in the collection are not believed to be too helpful in discriminating between relevant and non-relevant items. Thus we might count their frequency in the collection, or more precisely the inverse document frequency (denoted by *idf*), resulting in a larger weight for sparse words and a smaller weight for more frequent ones. Moreover, using a cosine normalization (retrieval model notation: "ntc–ntc"), may prove beneficial and each indexing weight may vary within the range of 0 to 1.

Considering that the occurrence of a given term in a document is a rare event, it may be a good practice to give more importance to the first occurrence of this word as compared to any successive occurrences. Therefore, the tf component may be computed as $0.5 + 0.5 \cdot [tf\ /\ max\ tf\ in\ a\ document]$. In this case, the normalization procedure is obtained by dividing tf by the maximum tf value for any term in the document (retrieval model denoted "atn"). Different weighting formulae may of course be used for documents and requests, leading to other different weighting combinations.

Finally we should consider that a term's presence in a shorter document provides stronger evidence than it does in a longer document. To account for this, we integrate document length within the weighting formula, leading to a more complex IR models, e.g., "Lnu"[6] (with the constant *pivot* fixed at 125 and *slope* at 0.1), "dtu-dtc"[7] or the Okapi probabilistic search model [8] (with $K = k_1 \cdot [(1-b) + ((b \cdot nt_i)/advl)]$ where nt_i indicates the length of document i, $advl$ is a constant fixed at 900, b at 0.75, and k_1 at 1.2). The question that then arises is: How will these retrieval models behave when used with our corpora?

Table 3. Weighting schemes

bnn	$w_{ij} = 1$	nnn	$w_{ij} = tf_{ij}$
ntc	$w_{ij} = \frac{tf_{ij} \cdot idf_j}{\sqrt{\sum_{k=1}^{t} (tf_{ik} \cdot idf_k)^2}}$	atn	$w_{ij} = idf_j \cdot \left[\frac{0.5 + 0.5 \cdot tf_{ij}}{max\ tf_{i.}}\right]$
Okapi	$w_{ij} = \frac{(k_1+1) \cdot tf_{ij}}{K + tf_{ij}}$	npn	$w_{ij} = tf_{ij} \cdot ln\left[\frac{n - df_j}{df_j}\right]$
dtu	$w_{ij} = \frac{(ln(ln(tf_{ij})+1)+1) \cdot idf_j}{(1-slope) \cdot pivot + (slope \cdot nt_i)}$		
dtc	$w_{ij} = \frac{(ln(ln(tf_{ij})+1)+1) \cdot idf_j}{\sqrt{\sum_{k=1}^{t} [(ln(ln(tf_{ik})+1)+1) \cdot idf_k]^2}}$		
Lnu	$w_{ij} = \frac{\frac{ln(tf_{ij})+1}{pivot + 1}}{(1-slope) \cdot pivot + (slope \cdot nt_i)}$		

2.4 Evaluation of Various Monolingual Corpora

As a retrieval effectiveness indicator, we adopted non-interpolated average precision as a retrieval effectiveness measure (computed on the basis of 1,000 retrieved items per request using TREC-EVAL). A decision rule is required to determine whether or not a given search strategy is better than another. The following rule of thumb could serve this purpose: a difference of at least 5% in average precision is generally considered significant.

The results in Table 4 show that the Okapi probabilistic model provides the best performance, significantly better than the vector-scheme ("Lnu–ltc"). The IR models "atn–ntc" or "dtu–dtc" perform well, yet not as well as the Okapi search approach. Finally, the traditional $tf \cdot idf$ weighting scheme ("ntc–ntc") does not exhibit very satisfactory results, and the simple term-frequency weighting scheme ("nnn–nnn") or the simple coordinate match ("bnn–bnn") results in poor retrieval performance.

Table 4. Average precision of various indexing and searching strategies based on monolingual requests and documents

	average precision (% change)			
title only model	English 33 queries 579 relevant	French 34 queries 528 relevant	Italian 34 queries 338 relevant	German 37 queries 821 relevant
Okapi–npn	37.26	41.62	33.98	31.64
Lnu–ltc	32.69 (-12.3%)	36.59 (-12.1%)	32.47 (-4.4%)	27.66 (-12.6%)
atn–ntc	31.40 (-15.7%)	39.04 (-6.2%)	28.96 (-14.8%)	31.30 (-1.1%)
dtu–dtc	31.96 (-14.2%)	37.89 (-9.0%)	31.04 (-8.7%)	28.23 (-10.8%)
ntc–ntc	18.11 (-51.4%)	25.02 (-39.9%)	20.35 (-40.1%)	23.42 (-26.0%)
bnn–bnn	12.54 (-66.3%)	22.85 (-45.1%)	19.63 (-42.2%)	23.44 (-25.9%)
nnn–nnn	9.69 (-74%)	14.56 (-65.0%)	15.15 (-55.4%)	9.78 (-69.1%)

For longer requests however these findings may be altered. To analyze this proposition, Table 5 demonstrates the impact of query length on search performance improvement, listing three different query formulations: (1) title section only, (2) both the title and descriptive sections or (3) all three sections (title, descriptive and narrative). Table 5 shows that retrieval effectiveness is enhanced when topics include more search terms, leading to significant enhancement, when comparing retrieval schemes and using queries based on the title section with those built using the title, descriptive and narrative sections. This finding does not hold however when there is a simple coordinate match ("bnn–bnn") or a simple term-frequency weighting scheme ("nnn–nnn"), thus tending to demonstrate that search keywords extracted from the descriptive or narrative sections are less likely to discriminate. Similar conclusions can be drawn with more statistically-based evidence when considering Italian, German and English collections.

Table 5. Average precision of various monolingual search models using different query formulations (French corpus, 34 queries)

	average precision (% change)		
query model \ mean terms per request	title only 2.88 terms	title-desc 6.95 terms	title-desc-narr 16.9 terms
Okapi–npn	41.62	46.29 (+11.2%)	46.73 (+12.3%)
Lnu–ltc	36.59	40.11 (+9.6%)	42.17 (+15.2%)
atn–ntc	39.04	41.68 (+6.8%)	43.98 (+12.7%)
dtu–dtc	37.89	40.12 (+5.9%)	43.93 (+15.9%)
ntc–ntc	25.02	27.40 (+9.5%)	29.65 (+18.5%)
bnn–bnn	22.85	16.55 (-27.6%)	13.76 (-39.8%)
nnn–nnn	14.56	14.09 (-3.2%)	13.69 (-6.0%)

3 Bilingual Information Retrieval

In Chap. 2, we described monolingual information retrieval systems within which documents and queries were written in the same language. In the current chapter, we will take a look at bilingual search models that, based on queries written in English, can retrieve pertinent information from document collections written in French, Italian or German [9]. To cross this language barrier, we have based our approach on free and readily available translation resources that can automatically provide translations of queries submitted in the desired target language. The first section of this chapter describes some related works while Sect. 3.2 presents our combined strategy and compares the retrieval effectiveness of our approach to other solutions proposed.

3.1 Related Work

In an early work, Salton [10] showed that by using a thesaurus manually constructed with care, cross-language retrieval could be as effective as that of monolingual retrieval. In this case, each thesaurus class groups related words in both German and English, thus serving as an interlingual link between English terms and their related German translations. Derived from this work, we might consider manually indexing case law using a manually built controlled thesaurus, a solution used by the Swiss Supreme Court retrieval system.

Sheridan & Ballerini [11] suggested generating co-occurrence information from comparable corpora, but unaligned, in order to find statistically related terms in the target language for a better translation quality. The average precision obtained by this similarity thesaurus is still considerably below that of single-language retrieval. Moreover, comparable corpora are not readily available.

Hull and Grefenstette [12] proposed an approach that resulted in a drop of 50% in average precision when translating words based on bilingual dictionaries, compared to monolingual performance. These authors found that the correct identification and translation of multi-word expressions can make the biggest difference in average performance compared to the problem of resolving translation ambiguity or when faced with missing terminology (e.g., a given word does not appear in the bilingual dictionary).

Ballesteros & Croft [13] suggested using a bilingual dictionary to perform word-by-word translation and then adding terms to the query through pre-translation and post-translation query modification, using blind relevance feedback.

Braschler & Schäuble [14] suggested using available machine translation software to automatically translated queries, documents or both. In this study, the document translation-based approach performed better than the query translation retrieval scheme. Moreover, the combination of similarity thesaurus, document and query translation-based resulted in the best performance. However when searching on the Web, query translation alone seems to be more realistic.

3.2 Bilingual Information Retrieval

In our bilingual experiments, we were faced with the following situation. The original queries were written in English and we did not have any parallel or aligned corpora to derive statistically or semantically related words in the target language. In order to develop a fully automatically approach, we chose to translate the requests using the SYSTRAN system (http://www.systran.com) and to translate query terms word-by-word using the BABYLON bilingual dictionary (http://www.babylon.com). In this case, the bilingual dictionary may suggest not only one, but several candidates for each word revealing the underlying ambiguity of a given term.

Of course, various errors can result from translating automatically a query formulation due to too small coverage by the bilingual dictionary, the underlying ambiguity of a term, the correct identification of multi-word concepts and their appropriate translation, and the translation of proper names. Examples of unsuccessful query translations are given in Table 2.

As reported by previous works, our experiments shown in Table 6 indicate that the average precision produced by using machine-translation systems or a bilingual dictionaries is clearly below that achieved by a monolingual run (label "monolin."). Using two translated words instead of one decreases the retrieval effectiveness to a greater extent (-47.27% in mean for the French collection). When we observed Italian of German bilingual retrieval performances (see Tables 7 and 8), we could draw similar conclusions. However, the retrieval effectiveness obtained with the SYSTRAN system approach seems better than the performance achieved by BABYLON bilingual dictionary.

In order to improve search performance, we tried combining the machine translation approach results with those produced by a bilingual dictionary. In order to verify this hypothesis, we added the first translated word produced by a bilingual dictionary look-up to each translated query generated by the SYSTRAN system. The average precision obtained by this combined strategy is shown in Table 6 under the label "combined". As depicted in this table, this approach results in better performance than the other bilingual search strategies. However, retrieval effectiveness was still below the average precision obtained by manually translating the requests into French. For the Italian and German corpora, we might draw identical conclusions.

Table 6. Precision of translating strategies (French corpus, title only, 34 queries)

	average precision (% change)				
model	monolin.	SYSTRAN	BABYLON 1	BABYLON 2	combined
Okapi–npn	41.62	29.64 (-28.8%)	29.45 (-29.4%)	21.37 (-48.7%)	33.14 (-20.4%)
Lnu–ltc	36.59	25.64 (-29.9%)	24.44 (-33.2%)	20.40 (-44.2%)	28.34 (-22.6%)
atn–ntc	39.04	26.37 (-32.5%)	28.58 (-26.8%)	20.15 (-48.4%)	30.87 (-20.9%)
dtu–dtc	37.89	26.66 (-29.6%)	28.99 (-23.5%)	21.13 (-44.2%)	30.01 (-20.8%)
ntc–ntc	25.02	14.69 (-41.3%)	14.93 (-40.3%)	13.08 (-47.7%)	17.14 (-31.5%)
bnn–bnn	22.85	14.59 (-36.1%)	12.10 (-47.1%)	8.49 (-62.8%)	16.69 (-30.0%)
nnn–nnn	14.56	8.60 (-40.9%)	10.14 (-30.4%)	9.49 (-34.8%)	10.88 (-25.3%)
mean difference		-34.17%	-32.92%	-47.27%	-24.05%

Table 7. Precision of translating strategies (Italian corpus, title only, 34 queries)

	average precision (% change)				
model	monolin.	SYSTRAN	BABYLON 1	BABYLON 2	combined
Okapi–npn	33.98	20.79 (-38.8%)	19.93 (-41.3%)	17.47 (-48.6%)	25.78 (-24.1%)
Lnu–ltc	32.47	19.70 (-39.3%)	18.96 (-41.6%)	18.16 (-44.1%)	24.62 (-24.2%)
atn–ntc	28.96	16.77 (-42.1%)	15.94 (-45.0%)	11.80 (-59.3%)	21.48 (-25.8%)
dtu–dtc	31.04	20.08 (-35.3%)	18.92 (-39.0%)	15.09 (-51.4%)	22.04 (-29.0%)
ntc–ntc	20.35	14.02 (-31.1%)	13.16 (-35.3%)	13.14 (-35.4%)	15.58 (-23.4%)
bnn–bnn	19.63	11.31 (-42.4%)	8.03 (-59.1%)	7.22 (-63.2%)	9.88 (-49.7%)
nnn–nnn	15.15	11.72 (-22.6%)	12.50 (-17.5%)	11.77 (-22.3%)	12.45 (-17.8%)
mean difference		-35.95%	-39.84%	-46.32%	-27.72%

4 Conclusion

Convinced that isolated retrieval effectiveness evaluations are not very useful, we have carried out various mono- and bilingual experiments based on various search strategies showing that:

1. French, Italian of German collections can be accessed with the same retrieval models developed for the English corpora (see Table 4);
2. the best retrieval model for the English collection is also the best for the three other languages (see Table 4);

Table 8. Precision of translating strategies (German corpus, title only, 37 queries)

	average precision (% change)				
model	monolin.	SYSTRAN	BABYLON 1	BABYLON 2	combined
Okapi–npn	31.64	22.59 (-28.6%)	17.39 (-45.1%)	18.61 (-41.2%)	25.43 (-19.6%)
Lnu–ltc	27.66	18.74 (-32.2%)	15.01 (-45.7%)	16.30 (-41.1%)	21.79 (-21.2%)
atn–ntc	31.30	21.78 (-30.4%)	14.42 (-53.9%)	14.63 (-53.3%)	25.25 (-19.3%)
dtu–dtc	28.23	18.12 (-35.8%)	13.28 (-53.0%)	14.15 (-49.9%)	22.70 (-19.6%)
ntc–ntc	23.42	15.63 (-33.3%)	12.01 (-48.7%)	11.91 (-49.1%)	16.05 (-31.5%)
bnn–bnn	23.44	12.75 (-45.6%)	10.05 (-57.1%)	7.30 (-68.9%)	14.17 (-39.5%)
nnn–nnn	9.78	6.82 (-30.3%)	7.63 (-22.0%)	4.65 (-52.5%)	6.07 (-37.9%)
mean difference		-33.74%	-46.50%	-50.83%	-26.96%

3. using more search terms may improve retrieval effectiveness (see Table 5);
4. bilingual retrieval based on the query translation approach using only one source of evidence (machine translation or bilingual dictionary) is not really effective (see Tables 6 to 8);
5. combining a bilingual dictionary and a machine translation approach may significantly enhance retrieval effectiveness (see Tables 6 to 8).

Of course, these findings need still be confirmed using other languages or other test-collections and we should analyze other sources of evidence when automatically translating requests.

Acknowledgments. This research was supported by the SNF (grant 21-58 813.99).

References

1. Fox C.: A stop list for general text. ACM-SIGIR Forum **24** (1999) 19–35
2. Savoy, J.: A stemming procedure and stopword list for general French corpora. JASIS **50** (1999) 944–952
3. Sproat, R.: Morphology and computation. The MIT Press Cambridge (1988)
4. McNamee, P., Mayfield, J.: A language-independent approach to European text retrieval. Proceedings CLEF-2000 (2000) www.iei.pi.cnr.it/DELOS/CLEF/apl.doc
5. Salton, G.: Automatic text processing: The transformation, analysis, and retrieval of information by computer. Addison-Wesley, Reading (1989)
6. Buckley, C., Singhal, A., Mitra, M., Salton, G.: New retrieval approaches using SMART. Proceedings TREC'4 (1996) 25–48
7. Singhal, A., Choi, J., Hindle, D., Lewis, D. D., Pereira, F.: AT&T at TREC-7. Proceedings TREC-7 (1999) 239–251
8. Robertson, S. E., Walker, S., Beaulieu, M.: Experimentation as a way of life: Okapi at TREC. Information Processing & Management **36** (2000) 95–108
9. Grefenstette, G. (ed.): Cross-language information retrieval. Kluwer (1998)
10. Salton, G.: Automatic processing of foreign language documents. In: Salton, G. (ed.): The SMART retrieval system, experiments in automatic document processing. Prentice-Hall, Englewood Cliffs (1971) 206–219

11. Sheridan, P., Ballerini, J. P.: Experiments in multilingual information retrieval using SPIDER system. Procedings SIGIR'96 (1996) 58–65
12. Hull, D., Grefenstette, G.: Querying across languages. Proceedings SIGIR'96 (1996) 49–57
13. Ballesteros, L., Croft, B. W.: Statistical methods for cross-language information retrieval. In: Grefensette, G. (ed.): Cross-language information retrieval. Kluwer, Boston (1998), 23–40
14. Braschler, M. Schäuble, P.: Experiments with the Eurospider retrieval system for CLEF 2000. Proceedings CLEF-2000 (2000) www.iei.pi.cnr.it/DELOS/CLEF/braschler.pdf