Information Retrieval, 3, 273–284, 2000

Using Corpus-Based Approaches in a System for Multilingual Information Retrieval

MARTIN BRASCHLER braschler@eurospider.com
PETER SCHÄUBLE schauble@eurospider.com
Eurospider Information Technology AG, Schaffhauserstrasse 18, CH-8006 Zürich, Switzerland

Received August 30, 1999; Revised May 30, 2000; Accepted June 12, 2000

Abstract. We present a system for multilingual information retrieval that allows users to formulate queries in their preferred language and retrieve relevant information from a collection containing documents in multiple languages. The system is based on a process of document level alignments, where documents of different languages are paired according to their similarity. The resulting mapping allows us to produce a multilingual comparable corpus. Such a corpus has multiple interesting applications. It allows us to build a data structure for query translation in cross-language information retrieval (CLIR). Moreover, we also perform pseudo relevance feedback on the alignments to improve our retrieval results. And finally, multiple retrieval runs can be merged into one unified result list. The resulting system is inexpensive, adaptable to domain-specific collections and new languages and has performed very well at the TREC-7 conference CLIR system comparison.

Keywords: multilingual information retrieval, cross-language information retrieval, corpus-based approaches, document alignments

1. Introduction

We describe our efforts to develop a system for truly *multilingual information retrieval*, allowing users to formulate the query in their preferred language, in order to retrieve relevant documents in any of the languages contained in a multilingual document collection. This is an extension to the classical *cross-language information retrieval* (CLIR) problem, where the user can retrieve documents in a language different from the one used for query formulation, but only one language at a time. Since 1998, the CLIR track at the TREC text retrieval conference series also uses this generalization of CLIR (Voorhees and Harman 1999, Braschler et al. 2000). We therefore evaluated our methods at the TREC-7 conference and present an analysis of the results.

The quickly growing amount of multilingual information available to the public dictates a growing need for this general type of multilingual information retrieval. The World Wide Web is a good example of a multilingual document collection. On the web, information relevant to a user's information need is potentially available in a multitude of languages. Likewise, the need for multilingual information access is increasing in multilingual countries, organizations and enterprises.

We discuss methods in this paper that are based on data structures automatically constructed from training corpora ("*corpus-based approaches*"). At the center of discussion are so-called *document alignments*. We show that document alignments can be used for query translation, pseudo relevance feedback and merging of retrieval results.

Document alignments define a mapping between two collections, in our case in two different languages, relating the documents of one collection to their most similar counterparts in the other collection. Table 1 gives an example of such a pair of aligned documents:

Table 1. Example of an alignment pair (German and French documents, only titles are shown).

Takeshita zu Antrittsbesuch in Bonn eingetroffen (Takeshita has arrived in Bonn for first visit)
Arrivée en RFA du premier ministre japonais (Arrival of the Japanese prime minister in the FRG)

Clearly, the two documents are closely related.

We use this mapping between the document collections to form a *multilingual comparable corpus*. This is very similar to a *parallel corpus*, with the difference that a parallel corpus contains actual translations of the documents. One of the benefits of the presented method is the option to automatically build such comparable corpora, whereas parallel corpora tend to be rare and expensive to obtain.

In the field of linguistics, work on alignment has been going on for a long time. However, most of this work relies on parallel corpora and aligns at sentence or even word level and the amount of data processed is usually smaller by several magnitudes. For an example of sentence alignment, see e.g. Gale and Church (1993). There is also work on comparable corpora for lexicon extraction (see e.g. Fung and McKeown 1997). For CLIR, use of comparable corpora has also been proposed for statistical machine translation by IBM (Franz et al. 1999).

Other corpus-based approaches to CLIR have been proposed in the past few years, among them latent semantic indexing (LSI) (Landauer and Littman 1990) and the generalized vector space model (GVSM) (Carbonell et al. 1997). Apart from corpus-based approaches, further CLIR methods based on machine readable dictionaries and on machine translation systems have been developed. An overview of the state of the art in cross-language IR and predictions for future developments can be found in Klavans and Hovy (1999).

The remainder of the paper is structured as follows: first, we will outline how to produce document alignments (Section 2), We then show how our system uses the alignments to build a so-called similarity thesaurus for query translation (Section 3), to merge retrieval results (Section 4), and for doing pseudo-relevance feedback during cross-language retrieval (Section 5.1). The system has been very competitive in the CLIR track of the TREC-7 conference, and we analyze the results in (Section 5.2). We finish with an outlook on future work in (Section 6).

2. Document alignment

2.1. Using indicators for alignment

This paper deals with alignments on the document level, i.e. document similarity is calculated based on entire documents, as opposed to alignment on a finer level such as paragraphs, sentences or words.

Alignments are produced by using so-called "*indicators*" to find similarities between pairs of documents from the collections involved. Indications of such a similarity include shared proper nouns and numbers, classifiers, similar dates and terms that can be mapped using a dictionary.

The basic concept underlying the alignment process is to use texts from the first collection as queries and run them against the documents from the other collection, thus retrieving their most similar counterparts. These pairs of similar documents form a mapping between the two collections.

2.2. Producing the alignments

We produced alignments for most collections used in the TREC-7 cross-language IR track, in particular the AP collection (from the Associated Press, English documents from the years 1988–1990) and the three SDA collections (from the Schweizerische Depeschenagentur SDA, in German, French and Italian. German and French texts are from the years 1988–1990, while for Italian, only November 1989 to end of 1990 is available). All these collections contain texts from news wires. We did not produce alignments for the NZZ German collection (different type of data—this collection contains newspaper articles by the Neue Zürcher Zeitung NZZ. Also, the texts are dated from the year 1994, and are therefore from a different time period). We were however able to obtain extra data from the SDA news wire that is not contained in the TREC collections. This way we added an additional roughly 7 years of data for all three languages covered by SDA.

Note that the SDA collections in the different languages do not contain translations. They were produced by different editorial staff in the same organization. However, the various SDA collections contain a fairly similar choice of topics. They are also roughly classified with a small set of manually-assigned, language-independent descriptors (some 300+ different descriptors). The SDA texts are therefore very well suited for alignment.

The texts of the AP collection address a substantially different set of topics, and do not contain descriptors. This makes it much harder to align them to the SDA texts.

We aligned the following pairs of collections:

- SDA German with AP English, using a simplistic bilingual wordlist gathered from various free Internet sources, coupled with frequency-based term elimination.
- SDA German with SDA French, using proper nouns and numbers, combined with descriptors
- SDA German with SDA Italian, using proper nouns and numbers, combined with descriptors

When we align a pair of collections, the documents of one collection are assigned to take the role of queries, which are then run against the documents of the other collection. The "query" documents are "transferred" into the target language of the other collection by using the indicators, as listed above for the individual pairs of collections. This means that for the SDA German-AP English alignment, we used simple word-by-word dictionary lookup. The resulting "translated" document is probably illegible to a human reader, but can be used

to retrieve similar items in the target collection. For SDA German-SDA French and SDA German-SDA Italian, we opted for an even more simplistic "translation"—consisting of a list of proper nouns and numbers contained in the original document, plus the classifiers that the document was assigned by SDA.

The effort invested in this "translation" step therefore can be adjusted depending on how similar the collections are. In particular, we believe that provided the collections are similar enough, good alignment is obtainable without using slow and costly methods like machine translation. The only manually built resource we used for our alignment experiments was a free, simplistic German-English wordlist. This not only makes the approach cost-efficient, it also makes it applicable to language pairs where it is difficult to obtain high quality linguistic resources.

For running the queries against the other collection, we introduced refinements such as thresholding (including query length normalization) and using sliding date windows. Since not every document has a good counterpart in the other collection, a thresholding mechanism is needed to decide whether to form a pair. The date windows are introduced because news agencies tend to publish stories about the same events on or near the same date. Therefore, a small date distance is an indication of a good pair. By only considering documents within a narrow date window, we can also speed up the alignment process. For an in-depth discussion of these specific effects that goes beyond the scope of this paper, including techniques for visualization, see Braschler and Schäuble (1998).

The mapping given by the alignment pairs is *not* bijective. Several documents in one collection may be aligned to the same text in the other collection and not all documents are members of a pair. This is a consequence of different amounts of coverage of the same events by the two collections. If alignments in the opposite direction are needed, the roles of the collections must be swapped, and the pairs recomputed.

Aligning TREC-sized collections takes a few days to a week on reasonably sized Sun desktop workstations. The alignment process is also parallelizable, with different computers working on different parts of the collections. And, most importantly, since date windows are used, old alignments can be kept without recalculation when new documents are added to the collections. Aligning a new day's worth of news stories takes only a few minutes this way.

2.3. Evaluation of alignments

To assess the quality of alignments, we use a process similar to relevance assessments for result lists in information retrieval. A *sample* of the pairs is judged by a human assessor. We are using five levels of similarity, which are assigned to the document pairs. This is due to the fact that we found it hard to make binary decisions like in relevance judgments for retrieval runs: whereas there is a short, concise query for such assessments, in evaluating alignments, two whole documents have to be compared for similarity. They can be similar as a whole, only in parts, or not similar at all. We used the following five levels of similarity: "same story", "related story", "shared aspect", "common terminology" and "unrelated" (Braschler and Schäuble 1998).

We now discuss the results of trying this strategy on a 1% sample of the alignments between English AP and German SDA. This is the most interesting combination, because the alignment quality is likely to be questionable, since these are the most difficult collections

to align that we used for our experiments. We found that 58.8 ± 3.3% (95% confidence interval) of the pairs are made of documents that share one or more common topics, and 75.1 ± 2.9% of the pairs are made of documents that share at least common terminology. For the applications in this paper, we decided that this is unsatisfactory for building a similarity thesaurus for query translation. However, the collections are still well enough aligned to use them for pseudo relevance feedback and merging. The SDA collection pairs are a lot more similar, so they were used for all three application areas.

3. Similarity thesaurus

We can use the comparable corpus produced through the alignment process to calculate a so-called *similarity thesaurus*. Such a similarity thesaurus is an information structure representing term similarities which reflect domain knowledge of the collection over which the thesaurus is constructed. To construct the thesaurus, the roles of document and terms are exchanged, with the documents serving as indexing features and the terms as retrievable items (Schäuble and Knaus 1992, Schäuble 1997). Therefore, a similarity thesaurus represents a dual space to the document space and various retrieval methods developed for the document space can be easily transferred to this dual space. Monolingual similarity thesauri can be used for query expansion (Qiu 1995), and their multilingual counterparts have applications in query translation for CLIR (Sheridan and Ballerini 1996).

In our implementation of the similarity thesaurus, we use the dual space version of the classical "*tf***idf*" weighting method. The term frequencies *tf* are then replaced with feature frequencies *ff*, and the inverse document frequencies *idf* turn into inverse item frequencies *iif*. The resulting "*ff***iif*" weighting scheme is used to calculate the term—term similarities $sim(\varphi_i, \varphi_j)$ as follows:

$$sim'(\varphi_i, \varphi_j) := \sum_{d_j \in \varphi_h \cap \varphi_i} a_{j,h} \times a_{j,i} \tag{1}$$

where

$$a_{j,i} := ff(d_j, \varphi_i) \times iif(d_j) \tag{2}$$

and φ_i = feature i, d_j = document j.

The final similarity values $sim(\varphi_i, \varphi_j)$ are obtained after cosine length normalization.

We calculated similarity thesauri over the comparable corpora we constructed for SDA German-SDA French and for SDA German-SDA Italian. We did not build a thesaurus for SDA German-AP English, because the quality of alignments was not satisfactory (see Section 2.3). Moreover, the fact that we had the wordlist that we used for alignment available reduced the immediate need for a thesaurus for this language combination.

4. Merging the results of mono- and cross-language searches

In order to allow multilingual information retrieval, with the user querying a collection containing documents from many languages, our system merges the results from individual

bilingual cross-language runs. We use the "Lnu.ltn" (Singhal et al. 1996) weighting scheme to calculate the scores of individual documents (the "retrieval status values" $RSV_i(q, d_j)$, giving the probability of relevance of document d_j with regard to query q). In this case, and also for almost all popular alternative weighting schemes, merging is a non-trivial problem, since the scores of the bilingual runs i are on different scales. Additionally, the numbers of relevant documents in the different languages are unknown. It is possible that the same number of relevant documents exist in each language; but is also possible that all relevant documents are in a single language.

Our system uses linear transformations of the retrieval status values to cope with these problems. We decided for a linear approach since little is known about the distribution of RSVs, which is also highly dependent on the particular weighting scheme that is employed. Therefore, if a document d_j in language i has been retrieved by run i, its retrieval status value $RSV_i(q, d_j)$ is mapped to a common scale $RSV(q, d_j)$ in the following way:

$$RSV(q, d_j) := \alpha_i + \beta_i * RSV_i(q, d_j). \tag{3}$$

The parameters α_i and β_i are determined by means of aligned documents and a least square fit which minimizes the sum of the squares of the error of aligned pairs. For instance, assume that d_j and d_k were aligned because d_j covers a story in language h and d_k covers the same or a similar story in language i. These two documents obtained the scores $RSV_h(q, d_j)$ and $RSV_i(q, d_k)$, respectively. Because they were aligned, they should be mapped to similar scores,

$$\alpha_h + \beta_h * RSV_h(q, d_j) \approx \alpha_i + \beta_i * RSV_i(q, d_k), \tag{4}$$

or in other words: the square of the difference

$$\Delta_{jk}^2 := (\alpha_h + \beta_h * RSV_h(q, d_j) - \alpha_i - \beta_i * RSV_i(q, d_h))^2 \tag{5}$$

should be minimized, which is achieved by a least square fit. The advantage of this approach is that not only relevant but also irrelevant pairs of aligned documents are used for merging. Of course, non-aligned documents can also be mapped to the common scale using the mappings that were determined by means of the aligned pairs.

We tried this strategy to produce our merged TREC-7 runs. In doing so, we identified the following problems:

- Since only a certain percentage of the documents that were retrieved are aligned, some of the queries have very few data points to calculate the regression. The fewer the data points, the less reliable the regression parameters obviously are.
- In rare cases, the regression slope turned out to be negative. This is possible if there are extremely few data points, the count of relevant documents is very unbalanced with respect to the collections, or if there are very many relevant documents. We solved the problem by introducing an extra data point (0, 0) for every query, which ensured the slope to be positive at all times. A more careful analysis of this problem is desirable.

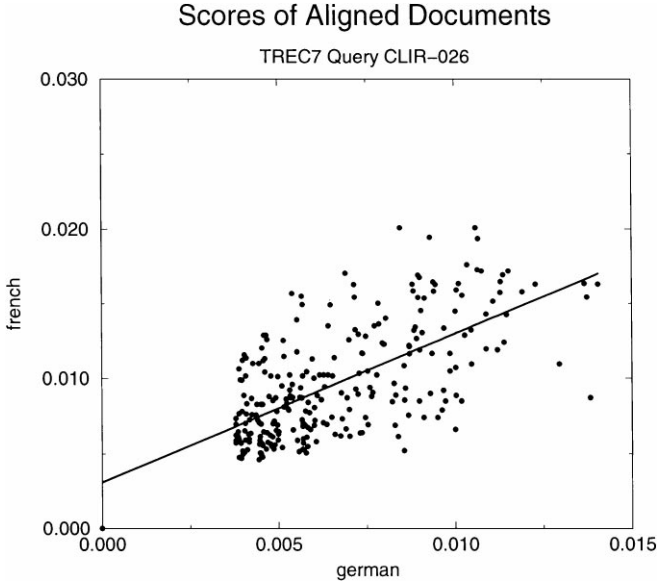

Figure 1. Plot of the scores of aligned documents. For all documents that were retrieved from the German and French collections, the score is plotted if the documents were paired during the alignment process. Also shown is the result from linear regression, which is used for rescaling of the scores during merging.

- By rescaling using the regression parameters, one collection is preferred over the other since there are data points above the regression line. This is a problem if the variation is high: one of the collections dominates the top ranks in this case.

Figure 1 shows an example for linear regression on the scores of aligned documents.

We compared our alignment merging strategy to two simple merging strategies. The two alternative runs where produced as follows: we computed a run constructed by taking one document in turn from each collection (document 1 from collection 1, document 1 from collection 2, document 2 from collection l, document 2 from collection 2, . . . , i.e. merging based on ranks). Finally, we also computed a run by merging everything *without* rescaling.

Merging without rescaling is expected to work very poorly, since the score ranges of the retrieval method we employed (Lnu.ltn) are not comparable across different queries and collections. And indeed, average precision is a full 44% lower than with the alignment merging method. Consequently, we did not investigate this strategy further. The difference between the run with merging using alignments compared to the run using rank-based merging is much smaller: the two runs are pretty much on par, with the alignment merging run only having a slight 5% edge. However, when doing comparison on a query-by-query basis, twice as many queries perform substantially better in terms of average precision for the alignment method than is the case for the rank-based merge. The benefit is most pronounced in the important high-precision range.

5. A multilingual information retrieval system using alignments

5.1. Cross-language information retrieval

We now describe the system for multilingual information retrieval that we used for our TREC-7 cross-language experiments. Since the system relies mainly on document alignments, building on the techniques we introduced in Sections 3 and 4, as opposed to e.g. machine translation, it is comparatively simple and very inexpensive. It is also fairly language independent, which is a big plus for languages with no sophisticated linguistic resources widely available.

The system uses a combination of two strategies. First, the query is translated, either using a similarity thesaurus if we had one available (German-French, German-Italian) or a wordlist (German-English). In both cases, very crude "*pseudo-translations*" are produced. The translation process consists of little more than word lookups in the thesauri or wordlist.

Since it has been shown repeatedly how poorly simple dictionary lookup performs for the problem of cross-language retrieval, our system combines the pseudo-translation with a strategy of *relevance feedback* on aligned documents. For more information on relevance feedback in general, see e.g. Harman (1992). For our application, consider the case that the two collections to be searched were parallel, i.e., real translations of each other. In such a setting, it would be possible to search the collection that corresponds to the language of the query, and return a result list produced by replacing every document found by its counterpart in the other collection. While this is not a very interesting case, since the collection to be searched is seldom available in translated form, we can replace the requirement for a parallel corpus with one for a comparable corpus. This requirement can be met in a lot of interesting real-world applications. One such example are the TREC cross-language collections that we aligned.

First, the user's query is run against the source collection, thus obtaining a ranked list. Instead of replacing the documents by their translations, the document-level mapping produced by the alignment process is used to replace them with their most similar counterparts from the other collection, if available. This produces a new result list containing documents in the target language. Because a lot of documents are not part of an alignment pair, however, they would never be retrieved using this strategy.

We addressed this problem by using a *pseudo relevance feedback* (also known as *local feedback*) (Xu and Croft 1996) loop *after* the replacement step (but *before* a search in the target language takes place). A certain number of the highest ranked documents are *assumed* relevant and the terms that best statistically discriminate these documents from the rest of the collection are extracted (we use the standard Rocchio method). These terms form a query used for a new search. Because the documents are already in the target language, so is the query produced. This strategy bears only *superficial* similarity to approaches such as proposed in Ballesteros and Croft (1997), since in our case, the feedback is used to effectively produce the translation *itself*, whereas in these other approaches, feedback is used *before* or *after* the translation. Unlike usual applications of relevance feedback, the new terms cannot be combined with the original query because their languages do not match. Only terms coming from the feedback process form the new query.

Relevance feedback alone works surprisingly well for certain queries. It fails however if the initial query does not retrieve any relevant documents. These poorly performing queries benefit enormously from a combination of the pseudo relevance feedback method with the query pseudo-translation obtained as outlined above. As was shown through query-by-query analysis of earlier experiments for TREC-6 CLIR topics (Braschler and Schäuble 1998), the queries that perform well using relevance feedback alone are little affected either positively or negatively by adding in query pseudo-translation. Consequently, the two methods mix well.

The actual combination of the terms from the two methods is facilitated by the fact that the relevance feedback mechanism reweights the terms, as is standard practice for Rocchio-based feedback methods.

In case the collection to be searched is not aligned, two independent aligned collections can be used for the relevance feedback step. The aligned collections are then only used for the transfer of the query into the target language, whereas the search takes place on the third collection.

5.2. Results on the TREC-7 CLIR collection

We now give the results we obtained for our system on the TREC-7 CLIR data collection. We submitted three official runs, all using the German topics (in TREC terminology, a topic denotes the "raw" form of a query) to search the whole TREC-7 CLIR multilingual document collection (German, French, Italian and English). All TREC topics contain several fields that allow to simulate queries of different lengths. Our runs differ only in the topic fields used: title only, title and description, or all fields (i.e. the full topics). The runs were produced by using our cross-language IR strategy just described, and then merging the individual cross-language runs to form a single, multilingual result list. Technical details about the setup of the experiments can be found in Braschler et al. (1999).

Our results compare very favorably to other submissions in TREC-7, especially considering that no costly linguistic resources were used to produce the runs. The average precision of our run using the full topics was roughly 12% less than the best run submitted. As CLIR was still a fairly new task at TREC-7, the number of participants was limited. This means some caution is appropriate when interpreting the results, because the evaluation methodology employed at TREC relies on having a sufficient number of dissimilar runs from multiple sources (Voorhees 1998). While an analysis of the evaluation process used in the TREC CLIR track would go beyond the scope this paper, a recent discussion gives strong indication that the results are reliable at least for the official runs submitted for the TREC conference (Braschler et al. 2000). We therefore deduce that our method performs well and is competitive.

Clearly, the run using the full topics is performing the best. We believe that long queries are beneficial to corpus-based techniques like the similarity thesaurus that have a broad vocabulary coverage, but also contain bad entries ("artifacts" from the training data) that may have an adverse effect if little input is available for the translation process.

The fact that the full queries perform so much better than the other two query sets also shows that our corpus-based techniques suffer less from a word ambiguity problem than

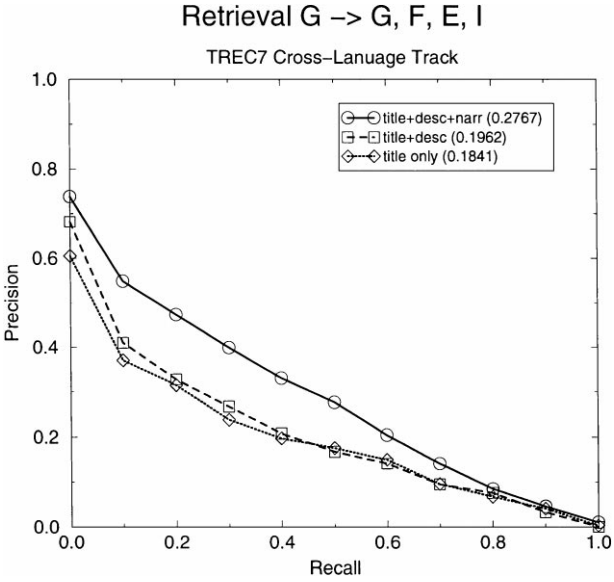

Figure 2. Results from our TREC-7 experiments. Shown are precision/recall curves for multilingual retrieval using German topics to match documents in German, French, English and Italian.

simple dictionary-based approaches. The latter approaches will produce very long output if the query length increases; this due to every word potentially having more than one translation. Such long queries are not likely to perform well, so this problem has to be addressed by adding additional components for word sense disambiguation or reweighting the translation alternatives. The similarity thesaurus in contrast allows to translate queries by using terms similar to the overall query concept instead of individual words, thus even allowing query reduction.

Work is clearly needed for the case of shorter queries, which is the usual case in some popular applications, such as web searching. The future direction of work in this area will likely be the incorporation of more linguistic resources.

6. Summary and outlook

We have presented a system for multilingual information retrieval, where users can formulate queries in their preferred language and retrieve documents from a multilingual collection containing many languages. The need for this kind of search systems is growing rapidly, as more and more multilingual information becomes available.

Our system is based on a process of so-called document alignment, where pairs of documents from different languages are formed according to their similarity. This alignment process is very modest in terms of the resources used; and most important, it does not need costly high-quality linguistic resources.

The document alignments enable us to tackle a range of problems, notably query translation through similarity thesauri, cross-language searches with the help of pseudo-relevance

feedback and merging of retrieval runs. The resulting system proved competitive when we participated in the CLIR track at the TREC-7 conference.

Future work will likely concentrate on improving the alignment process and refining the merging strategy. With the alignment, an interesting issue will be how to integrate more or better linguistic resources, without losing the flexibility to use the method in cases where such resources are unavailable. As for the merging, our goal is making the merging strategy more stable especially in cases where only little alignment information is usable.

Acknowledgments

Work leading to some aspects of the experiments presented began during one of the authors' time at the National Institute of Standards and Technology (NIST) and the Swiss Federal Institute of Technology (ETH). Thanks go to Donna Harman, Paul Over and Paraic Sheridan. The paper benefited a lot from helpful comments by two referees. Their feedback was greatly appreciated.

References

Ballesteros L and Croft BW (1997) 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.

Braschler M and Schäuble P (1998) Multilingual information retrieval based on document alignment techniques. In: Proceedings of the 2nd European Conference on Research and Advanced Technology for Digital Libraries, pp. 183–197.

Braschler M, Mateev B, Mittendorf E, Schäuble P and Wechsler M (1999) SPIDER retrieval system at TREC7. In: Proceedings of the Seventh Text Retrieval Conference (TREC-7), pp. 509–517.

Braschler M, Harman D, Hess M, Kluck M, Peters C and Schäuble P (2000) The Evaluation of systems for cross-language information retrieval. In: Proceedings of the Second International Conference on Language Resources and Evaluation (LREC-2000).

Carbonell JG, Yang Y, Frederking RE, Brown R, Geng Y and Lee D (1997) Translingual information retrieval: A comparative evaluation. In: Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI '97).

Franz M, McCarley JS and Koukos S (1999) Ad hoc and multilingual information retrieval at IBM. In: Proceedings of the Seventh Text Retrieval Conference (TREC-7), pp. 157–168.

Fung P and McKeown K (1997) Finding terminology translations from non-parallel corpora. In: The 5th Annual Workshop on Very Large Corpora.

Gale, WA and Church, KW (1993) A program for aligning sentences in bilingual corpora. Computational Linguistics, 19(1). Special Issue on Using Large Corpora I, pp. 75–102.

Harman DK (1992) Relevance feedback and other query modification techniques. In: Frakes, WB and Baeza-Yates R, Eds., Information Retrieval, Data Structures & Algorithms. Prentice-Hall, Englewood Cliffs.

Klavans J and Hovy E (1999) Multilingual (or Cross-lingual) information retrieval. In: Multilingual Information Management: Current Levels and Future Abilities, Ch. 2. (http://www.cs.cmu.edu/∼ref/mlim/chapter2.html).

Landauer TK and Littman ML (1990) Fully automatic cross-language document retrieval using latent semantic indexing. In: Proceedings of the Sixth Annual Conference of the UW Centre for the New Oxford English Dictionary and Text Research.

Qiu Y (1995) Automatic query expansion based on a similarity thesaurus. PhD Thesis, Swiss Federal Institute of Technology (ETH), Zurich, Switzerland.

Schäuble P (1997) Multimedia Information Retrieval. Kluwer Academic Publishers, Dordrecht.

Schäuble P and Knaus D (1992) The various roles of information structures. In: Information and Classification. Springer Verlag, Berlin, pp. 282–290.

Sheridan P and Ballerini JP (1996) Experiments in multilingual information retrieval using the SPIDER system. In: Proceedings of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 58–65.

Singhal A, Buckley C and Mitra M (1996) Pivoted document length normalization. In: Proceedings of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 21–29.

Voorhees EM (1998) Variations in relevance judgments and the measurement of retrieval effectiveness. In: Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 315–323.

Voorhees EM and Harman DK (1999) Overview of the seventh text retrieval conference (TREC-7). In: Proceedings of the Seventh Text Retrieval Conference (TREC-7), pp. 1–23.

Xu J and Croft BW (1996) Query expansion using local and global document analysis. In: Proceedings of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 4–11.