Different approaches to Cross Language Information Retrieval

Wessel Kraaij and Renée Pohlmann

TNO TPD

Abstract

This paper describes two experiments in the domain of Cross Language Information Retrieval. Our basic approach is to translate queries word by word using machine readable dictionaries. The first experiment compared different strategies to deal with word sense ambiguity: i) keeping all translations and integrate translation probabilities in the model, ii) a single translation is selected on the basis of the number of occurrences in the dictionary iii) word by word translation after word sense disambiguation in the source language. In a second experiment we constructed parallel corpora from web documents in order to construct bilingual dictionaries or improve translation probability estimates. We conclude that the best dictionary based CLIR approach is based on keeping all possible translations, not by simple substitution of a query term by its translations but by including reverse translation probabilities in the retrieval model.

1 Introduction

Within the framework of the TREC and recently also the CLEF information retrieval evaluation initiatives, TNO TPD has tested several approaches to cross language information retrieval (CLIR). Our basic approach is to translate queries word by word using machine readable dictionaries. The first experiment compared different strategies to deal with word sense ambiguity: i) keeping all translations and integrate translation probabilities in the model, ii) a single translation is selected on the basis of the number of occurrences in the dictionary iii) word by word translation after word sense disambiguation in the source language. In a second experiment we constructed parallel corpora from the web in order to construct bilingual dictionaries or improve translation probability estimates.

1.1 CLIR

Cross-Language Information Retrieval is receiving an increasing amount of attention in IR research. The goal of a CLIR system is to retrieve relevant documents from a multilingual document base in response to a query, irrespective of the language the documents are written in. Most CLIR systems either use query translation or document translation, cf. (Oard 1997). A third option would be to translate both queries and documents into a language independent representation (interlingua). Although this seems an attractive option, since queries and/or documents only need to be translated once and only one index needs to be maintained, in practice this last option is hardly ever used in other than very small scale, semi-automatic systems for well-defined domains, e.g. (Ruiz, Diekema and Sheridan 2000), because devising and maintaining such an interlingua for applications with very diverse documents, e.g. WWW search engines, would be infeasible. Both query translation and document translation have (dis)advantages. Theoretically, it seems that document translation would be superior to query translation. Documents provide more context for resolving ambiguities and the translation of source documents into all the languages supported by the IR system effectively reduces cross language retrieval to a monolingual task. Furthermore, document translation has the added advantage that document content is accessible to users in different languages (one of which may even be their mother tongue). Document translation, however, is inherently slower than query translation but, unlike query translation, it can be done off-line and translation speed may therefore not be crucial. Document translations need to be stored for indexing though, and storage space may be a limiting factor, especially if many languages are involved. Query translation on the other hand can be improved by consulting the user during translation, an option that is clearly not available for document translation. For realistically sized CLIR document collections like, for instance, the TREC CLIR collection which consists of 2 Gb of text, document translation is usually not considered a viable option, the majority of CLIR systems therefore apply a form of query translation, cf. (Braschler, Peters and Schäuble 2000), although two research groups have demonstrated the great potential of document translation: IBM (Franz, McCarley and Roukos 1999) with a fast statistical MT system optimised for CLIR and Eurospider (Braschler and Schäuble 2001) who translated the full CLEF collection with a commercial MT system.

1.2 CLIR evaluation conferences

Evaluation is a key activity for IR research. It gives researchers the opportunity to test new ideas on new data, while minimising the risk of tuning systems to a specific data set. The development of test corpora is a time consuming task, because human assessors are employed to set a 'gold standard'. In IR experiments, assessors decide whether retrieved documents are relevant for a certain query or not. The size of current test collections makes it impossible to do complete relevance judgements, so usually it is assumed that most relevant documents have been retrieved by a set of diverse systems. The quality of this *pool* is to a large extent dependent on the number and variety of retrieval systems that contribute to it (Hiemstra and Kraaij 1999). Since 1992 the Text REtrieval Conference (TREC) organised by NIST[1] has built a tradition of carefully controlled IR experiments. The first years were aimed at developing test procedures for two main tasks: Ad-Hoc queries and Routing queries. In later years, new tasks were introduced. The bilingual Spanish-English task at TREC-5 in 1995 can be considered the first small scale comparative CLIR evaluation experiment. In 1996, SIGIR hosted a successful CLIR workshop which stimulated groups to participate in the new CLIR task (the CLIR track in TREC terminology) at TREC-6. This first CLIR track was based on a new data set with French and German documents, originating from the Swiss News Agency SDA, a Swiss German newspaper and the AP document

[1] `http://trec.nist.gov`

set from TIPSTER. The topic set consisted of 24 queries, which were available in 5 languages. Groups were allowed to do any combination of topic and document language except EN-EN. The evaluation was quite successful, because a lot of new groups participated. The organisation of the track proved to be more difficult than monolingual evaluations. Firstly, the topic development had to be synchronised over several languages, secondly, relevance judgements were spread over different languages and carried out at different institutes, because NIST lacked enough native speakers of German and French. In terms of cross group comparability, the CLIR task structure had some problems. Because the availability of corpora in 3 languages and topics in 5 languages, groups from different nationalities generally chose to work in their own languages. Apart from lack of comparability, this also had an adverse effect on the reliability of the evaluation, because the number of runs per document language pool was quite low. But TREC-6 proved to be the starting point of a new stream of IR research for non English languages, also drawing attention from statistical MT researchers. The organisation decided to have a more controlled evaluation at TREC-7. The TREC-7 task showed three major changes:

1. The extension with Italian as a new document language. The Italian document collection also originated from the Swiss news agency SDA.

2. Instead of a free choice of tasks, groups were stimulated to do a multilingual run, i.e. retrieving relevant documents in multiple languages based on a query in a single language.

3. The start of the "GIRT" subtask, which focused on CLIR in a domain specific document collection. GIRT is a document collection consisting of documents from the social sciences, which are indexed by a domain specific multilingual thesaurus.

A similar set-up was maintained at the CLIR task of TREC-8. The new set-up was successful, although there were still some problems. First of all, only a few groups were able to do the multilingual task, because it required a lot of resources. Comparability of the runs improved considerably, but it is still a question whether one can really compare the performance of an English query on the document collection in 4 languages with the performance of a German query on the same collection. This was caused by the fact that the English document collection was much larger than the other subcollections, and yielded most of the relevant documents. There were also problems with quality control of the topics in the different languages, because sometimes translations were done by non-native speakers, or some query translations were not done from the source language. But, the quality of the evaluation matured steadily every year.

In 2000, the organisation of the CLIR evaluation moved to Europe, in order to acquire independent European funding and to attract more European participants. The new name of the evaluation is "Cross Language Evaluation Forum (CLEF)". Not surprisingly, CLEF focuses on European languages. The organisation stimulated participation of new groups by including bilingual and monolingual tasks for languages other than English. The number of participants has grown indeed while improving the quality of the evaluation: CLEF had more topics and larger pools for the relevance assessments. CLEF 2001 seems to continue the growth curve with 30 registered participants. Apart from CLEF, several other Cross Language Evaluation forums exist: NTCIR which focuses on Asian languages, Chinese–English is also the focus of cross language tasks at TREC and TDT sponsored by the American TIDES program, and Amaryllis, a French CLIR research program. Links to these activities can be found on the CLEF webpage: `http://www4.eurospider.ch/CLEF/resources.html`

In this paper we will present results from experiments run in the context of the CLIR track at TREC-8 and at CLEF 2000. Table 1 gives an overview of the two document collections.

	TREC-8		CLEF2000	
Nr. topics	28		40	
doc language	source	total docs	source	total docs
English	AP	242,866	LA Times	110,250
German	SDA	185,099	Frankfurter Rundschau, Der Spiegel	153,694
French	SDA	141,637	Le Monde	44,013
Italian	SDA	62,359	La Stampa	58,051

Table 1: Description of test collections

1.3 TNO engine & Retrieval Model

IR research at TNO started with the development of the Twenty-One retrieval system, a cross language retrieval system initially developed for dissemination and retrieval of documents in the field of sustainable development (Agenda 21). The development of the Twenty-One system was started in the context of an EU project in the Telematics Application Programme. Besides TNO TPD, project partners included the Universities of Twente and Tübingen, DFKI, Xerox, Getronics and several environmental organisations. Both document translation and query translation approaches to CLIR were explored in the development of the Twenty-One retrieval system. The first prototype was largely based on document translation. Using existing Machine Translation resources (Logos), source documents were translated and stored in the database. This early prototype of the Twenty-One system was not tested as such in the TREC CLIR evaluation task. Instead, all experiments were carried out with an information retrieval system based on a simple unigram language model (Hiemstra and Kraaij 1999). The basic idea is that documents can be represented by simple statistical language models. Now, if a query is more probable given a language model based on document d_1, than given a language model based on document d_2, then we hypothesise that document d_1 is more relevant to the query than document d_2. Thus the probability of generating a certain query given a document-based language model can serve as a score to rank documents with respect to relevance.

$$(1)\quad P(T_1, T_2, \cdots, T_n|D_k)P(D_k) \;=\; P(D_k)\prod_{i=1}^{n}(1-\lambda_i)P(T_i) + \lambda_i P(T_i|D_k)$$

Formula 1 shows the basic idea of this approach to information retrieval, where the document-based language model is interpolated with a background language model to compensate for sparseness. In the formula, each query term is modeled by a random variable T_i ($1 \leq i \leq n$, where n is the query length), whose sample space is the set $\{t^{(0)}, t^{(1)}, \cdots, t^{(m)}\}$ of all terms in the collection. The probability measure $P(T_i)$ defines the probability of drawing a term at random from the collection, $P(T_i|D_k)$ defines the probability of drawing a term at random from the document; and λ_i defines the importance of each query term. For our experiments we worked with a simplified model where we used the same constant λ_i for each query term. The optimal λ (0.15) was found by tuning on several test collections. The a-priori probability of relevance $P(D_k)$ is usually taken to be a linear function of the document length, modelling the empirical fact that longer documents have a higher probability of relevance.

The retrieval model has been extended for the CLIR task, by integrating a statistical translation step into the model (Hiemstra 2001). The CLIR extension is presented in the following formula:

$$(2)\quad \begin{aligned} &P(D_k, S_1, S_2, \cdots, S_n) \;= \\ &\quad P(D_k)\prod_{i=1}^{n}\sum_{j=1}^{m} P(S_i|T_i{=}t^{(j)})((1{-}\lambda_i)P(T_i{=}t^{(j)}) + \lambda_i P(T_i{=}t^{(j)}|D_k)) \end{aligned}$$

Here S_i refers to terms in the source (query) language and T_i refers to terms in the target (document) language, $P(S_i|T_i = t^{(j)})$ represents the probability of translating a term from the target language $t^{(j)}$ to a source language term S_i.[2]

An informal paraphrase of the extension is: the relevance of a document in a target language with respect to a query in a different source language can be modelled by the probability that the document generates the query. We know that several words T_j in the target language can be translated into the query term S_i, we also assume for the moment that we know their respective translation probabilities. The calculation of the probability involves an extra step: the probability of generating a certain query term is the sum of the probabilities that a document in the target language generates a word which in turn is translated to the query term. These probabilities are a product of the probability $P(T_j)$ as in Formula 1 with the translation probability $P(S_i|T_j)$. We refer to (Kraaij, Pohlmann and Hiemstra

[2] Note that the notions of source and target language are a bit confusing here, because the CLIR retrieval model contains a translation component, which translates *target* language terms to *source* language terms.

2000) and (Hiemstra 2001) for a technical description of the model. Section 2.1.1 explains how these translation probabilities are estimated. The retrieval model is implemented in the TNO retrieval engine, allowing for a fast and efficient retrieval procedure.

2 CLIR Experiments

Within the framework of the TREC and recently also the CLEF information retrieval evaluation initiatives, TNO TPD has tested several approaches to cross language information retrieval. Our main approach to CLIR for TREC and CLEF has been query translation. We experimented with two basic variants:

- Dictionary-based query translation using the VLIS lexical database developed by Van Dale Lexicography
- Corpus-based translation using parallel corpora

We will describe our experiments with these query translation techniques in the next sections.

2.1 Dictionary-based query translation

Our dictionary-based query translation strategies are based on the Van Dale VLIS database. The VLIS database is a relational database which contains the lexical material that is used for publishing several bilingual translation dictionaries, i.e. Dutch $\rightarrow$ German, French, English, Spanish. The database contains 270k simple and composite lemmas for Dutch corresponding to about 513k concepts. These concepts, Lexical Entities (LEs) in Van Dale terminology, are linked by several typed semantical relations, e.g. hyperonymy, synonymy, antonymy, effectively forming a concept hierarchy. All concepts have corresponding translations in French, Spanish, German and English. In Table 2 below, some statistics for the VLIS database are given.

language	simple lemmas	composite lemmas	total
English	260k	40k	300k
German	224k	24k	248k
French	241k	23k	264k
Spanish	139k	28k	167k

Table 2: number of translation relations in the VLIS database

Before translation, queries are pre-processed in a series of steps:

1. Tokenizing: The query string is separated into individual words and punctuation characters.

2. Part of speech tagging: Words are annotated with their part of speech. We use the the Xelda toolkit developed by Xerox Research Centre in Grenoble for tagging and lemmatisation.

3. Lemmatisation: Inflected word forms are lemmatised (replaced with their base form).

4. Stopword removal: So-called stopwords, i.e. frequent non-content bearing words like articles, auxilliaries etc, are removed.

The remaining query terms are subsequently translated into the different source languages. We used three different strategies for selecting translations from the VLIS database: all translations, the "most probable"[3] translation without using context information and translation after disambiguation in the source language. These three strategies will be discussed in the next sections.

2.1.1 All translations

For almost every lemma the VLIS lexical database lists a number of senses, each again possibly with several translations. In one experiment we decided to use all possible translations to search for relevant documents as this might at least lead to higher recall. We used disjunction to combine all possible translations of each query term, whereas conjunction was used to link the translations in a way that reflects the original query. For example:

bosbranden Sydney → (forest OR wood) AND fire AND Sydney

These "Boolean" queries are generated automatically in the translation process, no hand-coding of operators is required. We do not actually use the Boolean operators "OR" and "AND" but they are implicitly encoded in the structure of the translated query. We developed an algorithm that inputs queries in conjunctive normal form and assigns a probablility of relevance to documents given these queries. The algorithm takes into account the relative probabilities of translations. These probabilities are estimated in the following way. Some lemmas have identical translations for different senses. The Dutch lemma *bank*, for example, translates to *bank* in English in five different senses: "institution", "building", "sand bank", "hard layer of earth" and "dark cloud formation". Other translations include *bench*, *couch*, *pew*, etc. Since our retrieval model is based on the probability that a document (in the target language) generates a query (in the source language), cf. Section 1.3 above, translation probabilities are computed in the following way. First, we select all lemmas in the target language that translate to the query term in the source language. We subsequently translate the target language lemmas to the source language and count the number of times that the target lemma translates to the literal query term, e.g.

[3]No real translation probabilities are used but an approximation strategy.

query: bank (Dutch)	
bank (English) →	bank (2x), oever, reserve, rij etc.
pew (English) →	(kerk)bank, stoel
couch (English) →	bank, sponde, (hazen)leger, etc.

In the example above, the probability that *bank* (E) translates to *bank* (NL) is twice as high as the probability that *bank* (E) translates to *oever* (NL). Furthermore, some combinations of translations of query terms are more likely to occur together in documents than others. Documents containing such combinations of query terms will be ranked higher than others by the retrieval model. In this way the document collection itself is used for implicit disambiguation of possible translations (Hull 1997).

2.1.2 Most probable translation

In our ”most probable” translation strategy we select a single translation for each query term based on the number of occurrences of translations in the dictionary. When a lemma has several identical translations for different senses, e.g. in the “bank” example in Section 2.1.1 above, this ”most probable” translation is selected. If no translation occurs more than once, the first translation is chosen by default. The implicit assumption in this strategy is that the number of occurrences of a translation in the dictionary may serve as a rough estimate of an actual translation probability. Ideally, these probabilities should of course be estimated from actual corpus data.

2.1.3 Word sense disambiguation

We also experimented with a rather crude word sense disambiguation technique. In this approach, dictionary-based word senses are disambiguated in the source language using corpus information. First, the original query is used for retrieval on a monolingual corpus in the source language. All unique non-stopwords in the top N documents produced by this run are saved. We experimented with different values of N for this initial monolingual retrieval run, 20 turned out to be the best choice. Subsequently, all query terms are looked up in the VLIS database. The semantic relations defined in VLIS are used to look up synonyms, hyponyms and hyperonyms of each different sense of a query term. In this way we gather a group of words associated with each particular sense of a query term. These groups are expanded further using words from example sentences which illustrate the use of a particular word sense, which are also provided in the VLIS database. See Table 3 for examples of these word sense groups.

The groups of words associated with each possible sense of a query term are subsequently compared with the words from the monolingual retrieval run and ”evidence” for each sense is computed based on the overlap between the two sets of words. The sense for which the most evidence is found is selected. If no evidence is found at all or all senses score equally, the first sense is selected by default.

LEs	word sense groups
bank_1	concern business enterprise deposit mortgage loan
bank_2	rise elevation mound sandbank shoal aground stuck
pipe_1	duct funnel nozzle tube supply drain eustachian
pipe_2	tobacco peace clay water hookah opium

Table 3: example word sense groups

Query translation is now fairly straightforward. The translations for the selected word senses are looked up in the VLIS database, if more than one translation is given for a particular sense they are all included in a structured query (c.f. section 2.1.1 above).

2.1.4 Results

strategy	avp E-E	avp E-F	avp E-G	avp E-I	average	merged
alltrans	0.313	0.367	0.251	0.312	0.308	0.279
mprobtrans	0.313	0.332	0.205	0.312	0.288	0.252
disamtrans	0.313	0.310	0.181	0.312	0.276	0.241
monoling	0.313	0.551	0.410	0.362	0.409	0.323

Table 4: Results of the cross-lingual runs

In Table 4 the results of our submission for the TREC-8 CLIR track are presented. We chose to submit CLIR runs with English as the source language. English queries were run on the four target language document collections: English, German, French and Italian. Note that the E-E runs are monolingual runs without any form of translation. Because Italian is not included in the VLIS database, we did not use any of the translation strategies described above for the E-I runs, we used the Systran MT system instead. For comparison, we also include the results for the monolingual counterparts of all the cross-language runs. They provide an indication of the upper bound in performance that can be reached using our retrieval model. We present two different scores for the final results of the CLIR runs: average and merged. Average is simply the average score of the individual runs for the different target languages. The CLIR task, however, requires that the result list of a CLIR run consists of the top 100 documents in the four target languages *ordered by relevance*, *irrespective of language*. The result lists for the four different target languages therefore need to be merged in some way in order to obtain the final result list. A whole range of possible merging strategies have been proposed so far and research is still very much going on in this area. For TREC-8 we used a merging strategy based on document rank. We will not go into the details here but refer to (Kraaij et al. 2000).

If we look at the results for the different translation strategies we can conclude that the strategy where all translations are kept performs best for both French and German (the results for English and Italian are not relevant for this comparison). The second best strategy is the "most probable" strategy and the disambiguation strategy performs worst. We tentatively conclude that it is best not to select a particular translation unless one is very sure it is the correct one. Apparently, the CLIR retrieval process is not damaged nearly as much by adding extra incorrect translations as it is by leaving out correct ones. We were somewhat surprised by the results for disambiguation compared to the most probable translation strategy. It seems counterintuitive that simply picking the most probable translation, irrespective of context, should outperform the context-sensitive disambiguation strategy. More experimentation and error analysis are needed to explain this result.

If we compare the cross-language runs with their monolingual counterparts on a per-query basis, there are a number of queries with very poor results for all three translation strategies. We have identified some factors which contributed to this effect.

- Phrases. The failure to recognise and translate phrases as a unit was especially detrimental for the English to German runs where English phrases have to be translated to German single word compounds, e.g. *World War* → *Weltkrieg*, *armed forces* → *Bundeswehr*.
- Tagging errors, e.g. *arms* (weapons) was tagged as the plural of *arm* (body part) by the Xerox tagger.
- Capitalisation. Since most words in query titles[4] were capitalised, we decided to convert titles to lower case to prevent the tagger from tagging all title words as proper nouns. This had the effect that those title words that were actually proper nouns were not tagged correctly, e.g. the proper name *Turkey* was translated as *Truthuhn* and *dindon* (bird) in German and French respectively.

2.2 Parallel corpora

We developed three parallel corpora based on web pages in close cooperation with RALI, Université de Montréal. RALI already had developed an English-French parallel corpus of web pages, so it seemed interesting to investigate the feasibility of a full multilingual system based on web derived lexical resources only. We used the PTMiner tool (Nie, Simard, Isabelle and Durand 1999) to find web pages which have a high probability to be translations of each other. The mining process consists of the following steps:

1. Query a web search engine for web pages with a hyperlink anchor text "English version" and respective variants.

[4]TREC queries, or "topics" as they are called, are fairly extensive representations of an "information need". They consist of a title, description and narrative.

2. (For each web site) Query a web search engine for all web pages on a particular site.

3. (For each web site) Try to find pairs of path names that match certain patterns, e.g.: `/department/tt/english/home.html` and `/department/tt/italian.html`.

4. (For each pair) download web pages, perform a language check using a probabilistic language classifier, remove pages which are not positively identified as being written in a particular language.

The mining process was run for three language pairs and resulted in three modestly sized parallel corpora. Table 5 lists sizes of the corpora during intermediate steps. Due to the dynamic nature of the web, a lot of pages that have been indexed, do not exist anymore. Sometimes a site is down for maintenance. Finally a lot of pages are simply place holders for images and are discarded by the language identification step.

language	# web sites	# candidate pages	# candidate pairs	# cleaned pairs
EN-IT	3651	1053649	23447	4768
EN-DE	3817	1828906	33577	5743
EN-NL	3004	1170082	24738	2907

Table 5: Intermediate sizes during corpus construction

These parallel corpora have been used in different ways: i) to refine the estimates of translation probabilities of a dictionary based translation system (corpus based probability estimation) ii) to construct simple statistical translation models (IBM model 1) (Nie et al. 1999).

2.2.1 Results

run name	avp	description
ne1	0.307	standard NL→EN
ne2	0.276	corpus frequencies NL→EN
ei1	0.320	Systran MT EN→IT
ei2	0.275	corpus translations EN→IT

Table 6: Results of the bilingual runs

Table 6 lists the results of the bilingual experiments. The base run for Dutch to English scored an average precision of 0.307. The experiment with corpus based frequencies yielded disappointing results. We first generated topic translations in a standard fashion based on VLIS. Subsequently we replaced the translation probabilities $P(w_{NL}|w_{EN})$ by rough corpus based estimates. We simply looked up all English sentences which contained the translation and determined the proportion of the corresponding (aligned) Dutch sentences that contained the original Dutch query word. If the pair was not found, the original probability was left unchanged. Unfortunately a lot of the query terms and translations were not found in the aligned corpus, because they were lemmatised whereas the corpus was not lemmatised. This mismatch probably hurt the estimates. The procedure resulted in high translation probabilities for words that did not occur in the corpus and low probabilities for words that did occur. Other bilingual experiments for Dutch to English are reported in (Hiemstra, Kraaij, Pohlmann and Westerveld 2001)

For English to Italian we compared a Systran MT run with a statistical MT run based on the small parallel web corpus. We were quite surprised by the performance of the statistical MT run, which was not much below the performance of the Systran run. Key conclusion from this run is that usable translation dictionaries can be built from parallel web corpora.

3 Conclusions

Our initial conclusions from these experiments are that, so far, our best dictionary-based CLIR approach is keeping all possible translations. Our approach is not based on simple substitution of a query term by its translations but on including (reverse) translation probabilities in the retrieval model. Other researchers have published good results with similar strategies (Pirkola 1998),(Sperer and Oard 2000). Another common ingredient with these approaches is that our CLIR queries are *structured* queries, unlike standard - bag of word - expanded queries, which seem to work well for monolingual retrieval tasks but do not yield similar results in a CLIR setting (Hiemstra 2001). The results of our experiments also seem to indicate that the effectiveness of the CLIR process is not reduced nearly as much by including incorrect translations of query terms as it is by excluding correct ones. Our system could probably be improved by a model for phrase translations, which are especially important for translations from English to e.g. German (compounds). Finally, our pilot experiment seems to indicate that parallel web corpora can be used to produce reasonable translation resources.

References

Braschler, M. and Schäuble, P.(2001), Experiments with the eurospider retrieval system for CLEF 2000, *in* C. Peters (ed.), *Proceedings of CLEF 2000*, Springer. (to be published).

Braschler, M., Peters, C. and Schäuble, P.(2000), Cross-language information retrieval (CLIR) track overview, *in* E. Voorhees and D. Harman (eds), *The Eighth Text REtrieval Conference (TREC-8)*, National Institute for Standards and Technology. Special Publication 500-246.

Franz, M., McCarley, J. and Roukos, S.(1999), Ad hoc and multilingual information retrieval at IBM, *in* E. Voorhees and D. Harman (eds), *The Seventh Text REtrieval Conference (TREC-7)*, National Institute for Standards and Technology. Special Publication 500-242.

Hiemstra, D.(2001), *Using Language Models for Information Retrieval*, PhD thesis, University of Twente.

Hiemstra, D. and Kraaij, W.(1999), Twenty-one at TREC-7: Ad hoc and cross language track, *in* E. Voorhees and D. Harman (eds), *The Seventh Text REtrieval Conference (TREC-7)*, National Institute for Standards and Technology. Special Publication 500-242.

Hiemstra, D., Kraaij, W., Pohlmann, R. and Westerveld, T.(2001), Translation resources, merging strategies and relevance feedback, *in* C. Peters (ed.), *Proceedings of CLEF 2000*, Springer. (to be published).

Hull, D.(1997), Using structured queries for disambiguation in cross-language information retrieval, *in* D. Hull and D. Oard (eds), *AAAI Symposium on Cross-Language Text and Speech Retrieval*, American Association for Artificial Intelligence. http://www.clis.umd.edu/dlrg/filter/sss/papers/.

Kraaij, W., Pohlmann, R. and Hiemstra, D.(2000), Twenty-one at TREC-8: using language technology for information retrieval, *in* E. Voorhees and D. Harman (eds), *The Eighth Text Retrieval Conference (TREC-8)*, National Institute for Standards and Technology. Special Publication 500-246.

Nie, J., Simard, M., Isabelle, P. and Durand, R.(1999), Cross-language information retrieval based on parallel texts and automatic mining of parallel texts on the web, *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 74–81.

Oard, D. W.(1997), Alternative approaches for cross-language text retrieval, *in* D. Hull and D. Oard (eds), *AAAI Symposium on Cross-Language Text and Speech Retrieval*, American Association for Artificial Intelligence. http://www.clis.umd.edu/dlrg/filter/sss/papers/.

Pirkola, A.(1998), The effects of query structure and dictionary setups in dictionary-based cross-language information retrieval, *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 55–63.

Ruiz, M., Diekema, A. and Sheridan, P.(2000), CINDOR conceptual interlingua document retrieval, *in* E. Voorhees and D. Harman (eds), *The Eighth Text Retrieval Conference (TREC-8)*, National Institute for Standards and Technology. NIST Special Publication 500-246.

Sperer, R. and Oard, D. W.(2000), Structured translation for cross-language information retrieval, *Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 120–127.