Evaluating a Probabilistic Model for Cross-lingual Information Retrieval

Jinxi Xu
BBN Technologies
70 Fawcett Street
Cambridge, MA 02138
jxu@bbn.com

Ralph Weischedel
BBN Technologies
70 Fawcett Street
Cambridge, MA 02138
weischedel@bbn.com

Chanh Nguyen
BBN Technologies
70 Fawcett Street
Cambridge, MA 02138
chnguyen@bbn.com

ABSTRACT

This work proposes and evaluates a probabilistic cross-lingual retrieval system. The system uses a generative model to estimate the probability that a document in one language is relevant, given a query in another language. An important component of the model is translation probabilities from terms in documents to terms in a query. Our approach is evaluated when 1) the only resource is a manually generated bilingual word list, 2) the only resource is a parallel corpus, and 3) both resources are combined in a mixture model. The combined resources produce about 90% of monolingual performance in retrieving Chinese documents. For Spanish the system achieves 85% of monolingual performance using only a “pseudo-parallel” Spanish-English corpus. Retrieval results are comparable with those of the structural query translation technique (Pirkola, 1998) when bilingual lexicons are used for query translation. When parallel texts in addition to conventional lexicons are used, it achieves better retrieval results but requires more computation than the structural query translation technique. It also produces slightly better results than using a machine translation system for CLIR, but the improvement over the MT system is not significant.

1. INTRODUCTION

The goal of cross-lingual information retrieval (CLIR) is to find documents in one language for queries in another language. We use a probabilistic cross-lingual retrieval system, whose theoretical basis is probabilistic generation of a query in one language from a document in another. Hidden Markov Models (HMMs) (Rabiner, 1989) were used to approximate the query generation process. A key component of the retrieval model is probabilistic translation from terms in a document to terms in a query. The retrieval model integrates term translation probabilities with corpus statistics of query terms and statistics of term occurrences in a document to produce a probability of relevance for the document to the query. Similar approaches have been proposed for both monolingual IR (Ponte and Croft 1998; Berger and Lafferty 1999) and for CLIR (Hiemstra and de Jong, 1999); the differences are discussed later in the paper.

The focus of this study is on empirical evaluation of the proposed system. The probabilistic approach will be compared empirically with two popular CLIR techniques, structural query translation and machine translation (MT). The major difference between our approach and structural query translation is that ours uses translation probabilities while the other treats all translations as equals. A comparison between the two approaches will show the advantages and disadvantages of using probabilistic term translation for CLIR. The major difference between the MT-based technique and our approach is that the former does not use multiple translations for a term while the latter does. A comparison between them will show the advantages and disadvantages of using multiple translations in CLIR. The basic idea of structural query translation was used by a number of studies, including (Pirkola, 1998; Ballesteros and Croft, 1998; Sperer and Oard 2000; Hull 1997). Past studies that used MT systems for CLIR include (Oard, 1998; Ballesteros and Croft, 1998).

A common problem with past research on MT-based CLIR is that a direct comparison of retrieval results with other approaches is difficult because the lexical resources inside most commercial MT systems cannot be directly accessed. To overcome the problem we will use a technique to hypothesize the term translations inside a MT system based on the text it translated. By treating the translated text as a pseudo-parallel corpus, we can automatically induce a bilingual lexicon and use it with our system for cross-lingual retrieval. That will establish a lower bound on the performance of our system if it had direct access to the linguistic knowledge in the MT system.

In the next section we describe our retrieval model, including its limitations and potential extensions. Section 3 discusses related work. Section 4 describes the lexical resources used in this work. Section 5 describes the test collections used in our experiments and how they were processed. The test collections are the TREC5 Chinese track, the TREC9 cross-lingual track and the TREC5 Spanish track (Voorhees and Harman, 1997; Voorhees and Harman, 2000). Section 6 compares CLIR performance of our system with monolingual IR performance. Section 7 and 8 compare our system with structural query translation and MT-based CLIR. The last section summarizes this work and outlines directions for future work.

SIGIR'01, September 9-12, 2001, New Orleans, Louisiana, USA
Copyright 2001 ACM 1-58113-331-6/01/0009…$5.00.

2. RETRIVEAL MODEL

The basic function of an IR system is to rank documents against a query according to relevance. By Bayes' rule,

$$P(Doc\ is\ rel\,|Q) = \frac{P(Doc\ is\ rel)P(Q\mid Doc\ is\ rel)}{P(Q)}$$

Here *Doc* is a document and *Q* is a query. *P*(*Doc is rel*) is the prior probability of relevance for *Doc*, which we assume to be a constant.[1] *P*(*Q*) is the prior probability that *Q* is generated; since *Q* is a constant, *P*(*Q*) has no effect on document ranking. We can therefore rank documents by *P*(*Q* | *Doc is rel*), the probability that query *Q* is generated given document *Doc*.

We use Hidden Markov Models to simulate the process of query generation. (Rabiner, 1989) contains an excellent introduction to HMM theory. For convenience, we will assume that queries are in English and documents are in Chinese. We assume two states, the *General English state* and the *document state*. In the General English state, an English word for the query is generated; it may or may not describe the content of the document. In the document state, a word from the Chinese document is chosen and translated to an English word for the query. The following pseudo-code describes the query generation process.

```
Until all query words are generated
{
Toss a biased coin with probabilities $\alpha$ for
heads and $1-\alpha$ for tails. Enter the General
English state if it is heads and the
document state otherwise.

General English state: Pick an English word
from the English vocabulary according to a
probability distribution.

Document state: Pick a Chinese word from the
document according to a probability
distribution and translate it to an English
word according to another probability
distribution.
}
```

To minimize the need for training data, we estimate the parameters as follows:

1. The parameter α is a constant. We fix it at 0.3 in this study, based on prior experience.
2. In the General English (*GE*) state, we estimate the probability distribution as follows:

   $$P(e\mid GE) = freq(e,GE)/\mid GE\mid$$

   where *freq*(*e, GE*) is the frequency of English word *e* in an English corpus and |*GE*| is the size of the English corpus. Any large English corpus can be used for this purpose. In this study, we used TREC volumes 1-5 of English data.
3. In the document state (*Doc*), we estimate the probability distribution as follows:

   $$P(c\mid Doc) = freq(c,Doc)/\mid Doc\mid$$

   where *freq*(*c, Doc*) is the frequency of Chinese word *c* in *Doc* and |*Doc*| is the length of the document.
4. The probability of translation to an English word *e* given a Chinese word *c, P*(*e*|*c*), depends on *c* and *e* only. In section 4, we will discuss how to estimate the translation probabilities from parallel texts and from bilingual lexicons.

With these assumptions, it is easy to verify that:

$$P(Q|Doc) = \prod_{e\ in\ Q} \left(aP(e|GE) + (1-a) \sum_{chinese\ words\ c} P(c|Doc)\,P(e\mid c)\right)$$

This cross-lingual retrieval model is an extension of the monolingual retrieval model proposed by (Miller et al, 1999). In our discussion, we assume that the translation of a term is independent of the document and independent of the query in order to deal with data sparseness. The assumption dramatically reduces the number of parameters we need to estimate. If more data (such as a very large parallel corpus) becomes available in the future for parameter estimation, the independence assumption can be weakened to make the model more powerful. One possible technique is to employ bigram and trigram information to improve term translation.

3. RELATED WORK

Our retrieval model is similar to a number of existing ones. One such model was proposed in (Hiemstra and de Jong, 1999). A significant difference is that our model makes use of corpus statistics of the query language (English) while Hiemstra's does not. Roughly speaking, corpus statistics of a term can indicate the importance of a term in a query. In general, frequent terms are less useful than rare terms. This fact has been exploited by the traditional TF.IDF model as inverse document frequency (IDF). Instead of using the corpus statistics of the English terms (query terms), Hiemstra's model uses the corpus statistics of the Chinese terms (terms in documents). This is an attempt to model the importance of an English term based on the corpus statistics of its Chinese translations. This is a reasonable approximation if we do not have sufficient English text at our disposal. But given the vast amount of available textual data nowadays, we think a direct estimation procedure is more reliable because it avoids the noise introduced by translation.

Our model is an alternative to the structural query translation technique proposed in (Pirkola, 1998), whose basic idea can be traced to an earlier study in (Hull, 1997). It has been used in a number of studies, including (Sperer and Oard, 2000; Ballesteros and Croft, 1998; Kwok, 2000). This technique treats translations of a query term as synonyms of the term: occurrences of the Chinese translations of an English term in the Chinese documents are treated as instances of the English term. The technique is typically applied with a TF.IDF retrieval model. This technique treats all translations as equals while our model does not.

(Berger and Lafferty, 1999) views query generation as a translation process. So far, the model has only been used for

[1] Previous studies show that all documents are not equal. Longer documents in the TREC corpora, for example, are more likely to be relevant than short ones (Singhal, 1996). We ignore this issue because it is not a concern in this study.

monolingual retrieval, but potentially it can be applied to CLIR as well.

Studies that used MT systems for CLIR include (Ballesteros and Croft 1998; Oard 1998). As discussed earlier, direct comparisons with other techniques have been a problem because lexicons in most MT systems are inaccessible. (McCarley, 1999) studied both query and document translations and concluded the combination of the two translations can improve retrieval performance. (Levow and Oard, 1999) studied the impact of lexicon coverage on CLIR performance.

4. LEXICAL SOURCES

Two manual lexicons and one parallel corpus were used for English and Chinese CLIR experiments:

1. The LDC lexicon. It contains 86,000 English entries, 137,000 Chinese entries and 240,000 translation pairs. It is available from the Linguistic Data Consortium (LDC).
2. The CETA lexicon. It contains 35,000 English entries, 202,000 Chinese entries and 517,000 translation pairs. It can be obtained through the MRM Corporation, Kensingston, MD.
3. HKNews (Hong Kong SAR News) corpus. This parallel corpus consists of 18,000 pairs of documents in English and Chinese, with about 6 million English words. An algorithm developed in-house was used to align the corpus, resulting in 230,000 pairs of sentences. The corpus is available from LDC.

We use two techniques to estimate translation probabilities. For the manual bilingual lexicons, we assume uniform translation probabilities. That is, if a Chinese word *c* has *n* translations e_1 to e_n, we assume $P(e_i|c) = 1/n$.

For a parallel corpus, we use Brown et al's statistical machine translation models (Brown et al, 1993) to automatically induce a probabilistic bilingual lexicon. We used the WEAVER system developed by John Lafferty for this purpose (Lafferty, 1999). The WEAVER system implemented three of the five models proposed by Brown et al. Model 1 was used in this work for its efficiency. In order to keep the size of the induced lexicon manageable, a threshold (0.01) was used to discard low probability translations.

In order to increase lexicon coverage and to produce more robust probability estimates, different lexicons (including manual and induced) were combined to produce a single lexicon. Translation probabilities from different sources were linearly combined with equal weights:

$$P(e \mid c) = (P_{ldc}(e \mid c) + P_{ceta}(e \mid c) + P_{hknews}(e \mid c))/3$$

An exception is that if *c* does not occur in a source, the weight for that source will be equally distributed to the remaining sources. This ensures that the sum of the translation probabilities given a Chinese term is equal to 1. We should note that the weights given to the lexical sources could be adjusted to optimize retrieval performance. We will not explore this issue because it is not the focus of this work.

For English and Spanish CLIR, we used a lexicon induced from a translated corpus by a MT system (SYSTRAN). We will discuss that in detail in section 8. Table 1 summarizes the statistics about the lexical sources.

Table 1: Statistics about lexical sources. HKNews is a statistically derived lexicon. The combined lexicon is a combination of LDC, CETA and HKNews. English words are stemmed.

Lexical Source	English Terms	Chinese Terms	Translation Pairs
LDC	86,000	137,000	240,000
CETA	35,000	202,000	517,000
HKNews	21,000	75,000	860,000
Combined	104,997	305,103	1,490,000

5. TEST COLLECTIONS

Three test corpora were used in our experiments: TREC5 Chinese track (TREC5C), TREC9 cross-lingual track (TREC9X) and TREC5 Spanish track (TREC5S). TREC5C and TREC9X consist of Chinese documents with queries in English and Chinese. Having two versions of the same queries allows both monolingual and cross-lingual experiments. TREC5S consists of Spanish documents with queries in English and Spanish. English stemming used the Porter stemmer (Porter, 1980) and Spanish stemming used the stemmer by (Xu and Croft, 1998). All three fields (title, description and narrative) of the TREC topics were used in query formulation. Table 2 shows statistics about the test corpora.

For Chinese text segmentation, we used a simple dictionary-based algorithm. A list of valid Chinese words was obtained by combining the Chinese entries in the LDC and CETA lexicons. To segment Chinese text, the algorithm examines every substring of 2 or more characters and treats it as a word if it appears in the Chinese word list. In addition, a single Chinese character is also treated as a word if it is not part of any of the words recognized in the first step. The goal of the algorithm is to optimize cross-lingual performance, since it allows as many matches between English terms and Chinese terms as possible. For monolingual retrieval in Chinese, however, it has been shown that the best search strategy is to use a combination of bigrams and unigrams of Chinese characters (Kwok, 1997). That strategy was used in our monolingual experiments in order to produce the strongest monolingual baseline.

Table 2: Statistics about test collections. TREC5C=TREC5 Chinese track. TREC5S=TREC5 Spanish track. TREC9X=TREC9 Cross-lingual track

Corpus	TREC5C	TREC5S	TREC9X
Query language	English	English	English
Document language	Chinese	Spanish	Chinese
Query count	28	25	25
Document count	164,789	172,952	127,938
Query length	35	22	21

Throughout this paper, we will use the TREC average non-interpolated precision to measure retrieval performance (Voorhees, 1997).

6. CHINESE RETRIEVAL RESULTS

Table 3 shows the retrieval results of our CLIR system on TREC5C and TREC9X. Our monolingual results were obtained using Miller et al's HMM monolingual retrieval system (Miller et al, 1999). The monolingual results form a strong baseline; they are better than the best official monolingual results in the TREC5 and TREC9 proceedings (Voorhees and Harman, 1997, 2000). Given the strong baseline, the cross-lingual results using the combined lexicon are very impressive because they are around 90% of monolingual results (87% on TREC5C and 92% on TREC9X).

Table 3: Retrieval results on TREC5C and TREC9X.

Corpora	TREC5C	TREC9X
Monolingual	0.3910	0.3362
LDC	0.2886	0.1725
CETA	0.3067	0.2126
HKNews	0.2530	0.2418
Combined	0.3391	0.3100

Retrieval results using individual lexicons are significantly worse than those using the combination of the three lexical resources, confirming findings by other researchers that lexicon coverage is critical for CLIR performance (Levow and Oard, 1999). The results show that dialect similarity can also affect retrieval performance. Both the TREC9X corpus and the HKNews parallel corpus are in Cantonese (a Chinese dialect). Therefore, HKNews is more effective on TREC9X than LDC and CETA, which have a strong bias toward Mandarin (standard Chinese). On the other hand, since TREC5C is a Mandarin corpus, LDC and CETA are better than HKNews on TREC5C.

7. COMPARISON WITH STRUCTURAL QUERY TRANSLATION FOR CHINESE

In this section we compare the retrieval results of our system with those of the structural query translation technique. Our experiments followed the query translation procedure described in (Pirkola, 1998). A term in a Chinese document is treated as an instance of an English term if it is a translation of the English term according to a bilingual lexicon. Given a Chinese corpus, the term frequency and the document frequency of an English term are computed as:

$$tf(e, Doc) = \sum tf(c_i, Doc)$$

$$df(e) = | \bigcup doc\_set(c_i) |$$

where c_i's are Chinese translations of e and $doc\_set(c_i)$ is the set of Chinese documents containing c_i. The tf and df values of English terms were used with the INQUERY $tf.idf$ function (Allan et al, 2000) to compute the retrieval score of a Chinese document for an English query.

Table 4 shows that our system and structural query translation achieved similar retrieval results when LDC and CETA were used. The exception is that on TREC9X using CETA our system is significantly better (0.2126 vs. 0.1750). When HKNews and the combined lexicon were used, our system is significantly better.

Table 4: Retrieval results of structural query translation.

Corpora	Structural Model on TREC5C	HMM on TREC5C	Structural Model on TREC9X	HMM on TREC9X
LDC	0.3009	0.2886	0.1696	0.1725
CETA	0.2924	0.3067	0.1750	0.2126
HKNews	0.1886	0.2530	0.2022	0.2418
Combined	0.2764	0.3391	0.2285	0.3100

Since the procedure we used to obtain translation pairs from parallel texts is statistically based, it is error prone for infrequent terms. Most of the incorrect translations have a small probability estimate. These bad translations are automatically discounted by our system because they have small probabilities. However, since the structural query translation technique treats all translations equally, the bad translations become a serious problem. Experiments show that removing the low probability translations significantly improves the performance of structural query translation. Figure 1 shows the performance curves when we vary the probability cut off values on TREC9. The results confirm that noisy translations from the parallel corpus are a serious problem for structural query translation. However, these noisy translations are useful information to our system; removing them hurts retrieval performance of our system. The advantage of our system seems to be its capability of utilizing noisy translations to improve retrieval performance.

The disadvantage of our system is that it is less efficient than structural query translation due to the extra computation incurred by the using of translation probabilities in our model. The efficiency issue can be addressed by pre-computing $P(e|Doc)$ of the retrieval function. Such optimization techniques have been used in previous work (Hiemstra and de Jong, 1999). They were not used in this work because they would prevent us from experimenting with different bilingual lexicons without re-indexing.

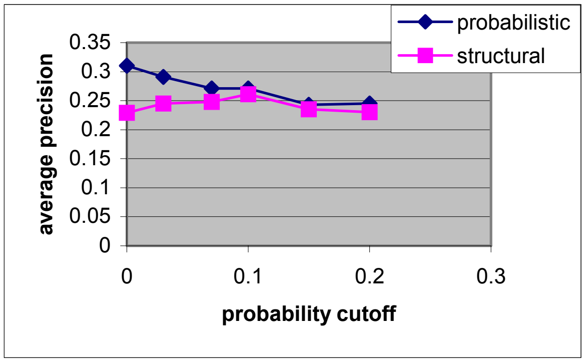

Figure 1: TREC9X, performance of the probabilistic term translation model and structural translation approach with varying thresholds on including low probability translations.

8. COMPARISON WITH MT-BASED APPROACHES FOR SPANISH

The major difference between MT-based CLIR and our approach is that the former uses one translation per term and the latter uses multiple translations. It has been suggested that CLIR can potentially utilize the multiple useful translations in a bilingual lexicon to improve retrieval performance (Klavans and Hovy, 1999). In our experiments, we used SYSTRAN version 3.0 (http://www.systransoft.com) for query and document translation. SYSTRAN is generally accepted as one of the best commercial MT systems for English-Spanish translation.

We performed four retrieval runs on the TREC5S corpus:

1. Query translation. English queries are translated to Spanish via SYSTRAN. Retrieval was performed using the translated queries on the Spanish corpus.
2. Document translation. The Spanish corpus is translated to English via SYSTRAN. Retrieval was performed using English queries on the translated corpus.
3. Combined run. The two retrieval scores for each document obtained in 1 and 2 were multiplied to produce a combined score for that document. Documents were then ranked based on the combined scores. Previous studies (McCarley, 1999) suggested that such a combination can improve CLIR performance.
4. Probabilistic CLIR. We induced a bilingual lexicon from the translated corpus by treating the translated corpus as a pseudo-parallel corpus. WEAVER was used to induce a bilingual lexicon for our approach to CLIR.

Table 5 shows that probabilistic CLIR using our system outperforms the three runs using SYSTRAN, but the improvement over the combined MT run is very small. Its performance is around 85% of monolingual retrieval. Please note that the induced lexicon is probably a trimmed version of the true lexicon in SYSTRAN. Had we had direct access to the relevant linguistic knowledge (including lexicon and disambiguation knowledge) in the MT system, we could probably make a better probabilistic bilingual lexicon than the one induced from a pseudo-parallel corpus. As a result, we could produce better retrieval performance. On the other hand, the test set has only 25 queries and the difference between our system and the combined MT run is very small. Therefore, we cannot draw a firm conclusion about the retrieval advantage of probabilistic CLIR without further study.

Nonetheless, the results suggest that a simple dictionary-based approach can be as effective as a sophisticated MT system for CLIR. This is particularly important for languages where MT may not be available, but where bilingual word lists may have been compiled.

Table 5: Comparing our CLIR system and MT-based CLIR.

Monolingual	0.4275
Query translation	0.2943
Doc translation	0.3197
Doc and query translation	0.3466
Probabilistic CLIR	0.3615

The goal of our experiments is not to dismiss the MT-based approach; it is viable for at least two reasons. First, it is much faster than our CLIR system. It is about 10 times as fast as our CLIR system in the above experiments. Even though pre-computation can improve the efficiency of our system (as we discussed earlier), we expect MT-based CLIR would still be faster due to a sparser term-document matrix. Second, the retrieved documents are readable by end users. These properties make it the ideal search strategy in an interactive CLIR environment. The advantage of the dictionary-based approach is also twofold. It is relatively inexpensive to build and it can potentially produce better retrieval results by using more than one translation per term.

9. CONCLUSIONS

We proposed and evaluated a probabilistic CLIR retrieval system. The system achieved roughly 90% of monolingual performance in retrieving Chinese documents and 85% in retrieving Spanish documents. We have shown how a simple mixture model combining bilingual word lists and parallel corpora can outperform either alone. It also appears that, with this approach, additional bilingual lexicons and parallel text improve performance substantially in spite of the increased ambiguity.

Experiments show that while our system is more effective than the structural query translation technique when parallel texts are available for term translation, the latter is more efficient. Our system is also slightly more effective than the combined technique of query and document translation using a commercial MT system, but the difference in retrieval performance is small.

One area for future work is to improve our retrieval model by incorporating contextual information for better term translation. Term disambiguation has been a subject of intensive study in CLIR (Ballesteros, 1998). Applying the research results in that area will be helpful. A second area is to make better use of the translation models in WEAVER. Some of the translation models allow a word to be translated to several words (e.g. a phrase) in the other language. We believe if properly used, this feature can improve retrieval performance because it more accurately accounts for the query generation process than our current retrieval model.

10. REFERENCES

[1] Allan, J., Callan, J., Feng, F-F, and Malin, D. 2000. "INQUERY at TREC8." In *TREC8 Proceedings*, Special publication by NIST, 2000.

[2] Ballesteros, L., and Croft, W.B. 1998. "Resolving ambiguity for cross-language retrieval." In *Proceedings of SIGIR Conference*, pages 64-71, 1998.

[3] Berger, A. and Lafferty, J. 1999. "Information retrieval as statistical translation." In *Proceedings of SIGIR Conference*, 1999.

[4] Brown, P., Della Pietra, S., Della Pietra, V., and Mercer, R. 1993. "The mathematics of statistical machine translation: parameter estimation." *Computational Linguistics*, 19(2):263-311, 1993.

[5] Hiemstra, D. and de Jong, F. 1999. "Disambiguation strategies for cross-language information retrieval." In *Proceedings of the third European Conference on Research and Advanced Technology for Digital Libraries*, pages 274-293, 1999.

[6] Hull, D. 1993. "Using statistical testing in evaluation of retrieval experiments." In *Proceedings of SIGIR Conference*, 1993.

[7] Hull, D. 1997. “Using structured queries for disambiguation in cross-language information retrieval.” In *AAAI Symposium on Cross-Language Text and Speech Retrieval*, 1997.

[8] Klavans, J. and Hovy, E. 1999. "Multilingual (or Cross-lingual) Information Retrieval". Chapter 2, *Multilingual Information Management, current levels and future abilities*. Editors, E. Hovy, N. Ide, R. Frederking, J. Mariani and A. Zampolli, Arpil, 1999.

[9] Kwok, K. L. 1997. “Comparing representations in Chinese information retrieval.” *Proceedings of SIGIR Conference*, 1997.

[10] Kwok, K.L. 2000. “TREC9 Cross-language, question-answering track experiments using PIRCS.” *TREC9 Proceedings* published by NIST, 2000.

[11] Lafferty, J. 1999. Personal communications.

[12] Levow, G.A. and Oard, D. 1999. “Evaluating lexical coverage for cross-language information retrieval.” In *Workshop on Multilingual Information Processing and Asian Language Processing*, Beijing, 1999.

[13] McCarley, J. S. 1999. “Should we translate the documents or the queries in cross-language information retrieval.” In *Proceedings of ACL 99*, pages 208-214, June 1999.

[14] Miller, D., Leek, T., and Schwartz, R. 1999. “A hidden markov model information retrieval system.” In *Proceedings of SIGIR Conference*, 1999.

[15] Oard, D. 1998. “A comparative study of query and document translation for cross-language information retrieval.” *Third Conference of the Association for Machine Translation in the Americas (AMTA*), 1998.

[16] Pirkola, A. 1998. “The effects of query structure and dictionary setups in dictionary-based cross-language information retrieval.” In *Proceedings of SIGIR Conference*, pages 55-63, 1998.

[17] Ponte, J. and Croft, W.B. 1998. “A language modeling approach to information retrieval.” In *Proceedings of SIGIR Conference*, pages 275-281, 1998.

[18] Porter, M. 1980. “An algorithm for suffix stripping.” *Program* 14, 3(1980), pages 130-137.

[19] Rabiner, L. 1989. “A tutorial on Hidden Markov models and selected applications in speech recognition”, In *Proceedings of IEEE 77,* pages 257-286, 1989.

[20] Singhal, A. and Buckley, C. and Mitra, M. “Pivoted Document Length Normalization.” *In Proceedings of SIGIR Conference*, 1996.

[21] Sperer, R. and Oard, D. 2000. “Structured query translation for cross-language information retrieval.” In *Proceedings of SIGIR Conference, 2000*.

[22] Voorhees, E. and Harman, D. 1997. *TREC-5 Proceedings*. NIST special publication, 1997.

[23] Voorhees, E. and Harman, D. 2000. *TREC-9 Proceedings*. To be published by NIST.

[24] Xu, J. and Croft, W. B. 1998. “Corpus-based stemming using co-occurrence of word variants.” *ACM TOIS*, 18(1):79-112, January 1998.