Available online at www.sciencedirect.com
SCIENCE DIRECT®
Information Processing and Management 41 (2005) 433–455
INFORMATION PROCESSING & MANAGEMENT
www.elsevier.com/locate/infoproman

Technical issues of cross-language information retrieval: a review

Kazuaki Kishida *

Faculty of Cultural Information Resources, Surugadai University, 698 Azu, Hanno, Saitama 357-8555, Japan

Received 10 June 2004; accepted 14 June 2004
Available online 23 August 2004

Abstract

This paper reviews state-of-the-art techniques and methods for enhancing effectiveness of cross-language information retrieval (CLIR). The following research issues are covered: (1) matching strategies and translation techniques, (2) methods for solving the problem of translation ambiguity, (3) formal models for CLIR such as application of the language model, (4) the pivot language approach, (5) methods for searching multilingual document collection, (6) techniques for combining multiple language resources, etc.

Keywords: Cross-language information retrieval; Machine translation; Word sense disambiguation; Language model

1. Introduction

Cross-language information retrieval (CLIR) is the circumstance in which a user tries to search a set of documents written in one language for a query in another language. The issues of CLIR have been discussed for several decades. As widely recognized, research efforts for developing CLIR techniques can be traced back to Gerard Salton's articles in the early 1970s (e.g., Salton, 1970).

Especially after the advent of the World Wide Web in the 1990s, CLIR has become more important, allowing users to access information resources written in a variety of languages on the Internet. Since then, the research community of IR has begun to tackle problems of CLIR extensively and intensively. The *Workshop on Cross-Linguistic Information Retrieval* held in August 1996 during the SIGIR'96 Conference is frequently cited as an epochal event for promoting research on CLIR.

* Fax: +81 426 35 9872.
E-mail address: kishida@surugadai.ac.jp

0306-4573/$ - see front matter
doi:10.1016/j.ipm.2004.06.007

Currently, CLIR issues are addressed in workshops of large-scale retrieval experiments such as TREC, CLEF and NTCIR. As described in the introductory paper to this issue, each workshop has been concerned with languages other than English as follows:

TREC: Spanish, Chinese, German, French, Italian, and Arabic.
CLEF: French, German, Italian, Swedish, Spanish, Dutch, Finnish, and Russian so far.
NTCIR: Japanese, Chinese and Korean.

Various research findings on CLIR have been reported at the meetings of TREC, CLEF and NTCIR, and many papers have been published in scientific journals and proceedings.

This article aims at reviewing techniques and methods for enhancing performance of CLIR. We already have a comprehensive review on this topic (Oard & Diekema, 1998). In addition, Peters and Sheridan (2001) cover a wide range of literature and topics on CLIR. The main purpose of this article is to examine literature subsequent to the review by Oard and Diekema and to attempt to organize research results since the mid-1990s in the CLIR field from a technical point of view. For this purpose, some works listed in Oard and Diekema (1998) will be referred to again in this article.

However, it should be noted that this review cannot be completely comprehensive because of the large number of papers on CLIR published in various research areas. The purpose here is to provide a useful map of technical issues of CLIR, rather than extensively enumerating research papers on CLIR. This paper is mainly concerned with "document retrieval," or "text retrieval" issues. For example, CLIR for multimedia data is outside our scope.

The rest of the paper is organized as follows. First, in Section 2, we discuss techniques to match query terms with document representations in the CLIR. More specifically, various methods of translation are described. Section 3 is dedicated to explaining some techniques for solving the problem of term ambiguity, which may occur in the process of translation. Some formal models for CLIR are introduced in Section 4. In particular, we describe the application of the language model (LM), which enables us to combine the retrieval model and the translation model. In Section 5, other important CLIR research topics are discussed: the pivot language approach, search of multilingual document collections, combination of language resources, issues on processing of individual language, user interface for interactive CLIR and evaluation of CLIR. Finally, Section 6 briefly discusses the future direction of CLIR research.

2. Matching strategies and translation

2.1. Matching strategies

2.1.1. Types of matching strategies

The most basic approach to CLIR is to automatically translate the query into an equivalent in the language of the target documents. The translation makes it possible to execute matching operations between the query and each document, and subsequently, compute document scores according to a standard retrieval model such as the vector space or probabilistic model.

However, this is only the starting point. Oard and Diekema (1998) have identified four types of strategies for matching a query with a set of documents in the context of CLIR (Oard & Diekema, 1998, pp. 230–232):

- No translation
  (1) Cognate matching

- Translation
  (2) Query translation
  (3) Document translation
  (4) Interlingual techniques

2.1.2. Cognate matching

In the case of the most naïve *cognate matching*, untranslatable terms such as proper nouns or technical terminology are left unchanged through the stage of translation. The unchanged term can be expected to match successfully with a corresponding term in another language if the two languages have a close linguistic relationship.

Some researchers have tried to add devices for matching cognates more effectively. For example, Davis (1997) introduced fuzzy matching based on edit distance between Spanish cognates and English words. Interestingly, for CLIR from English to French, Buckley, Walz, Mitra, and Cardie (1998) pointed out that "English query words are treated as potentially misspelled French words," and attempted to treat English words as variations of French words according to lexicographical rules. Similarly, Hiemstra and Kraaij (1999) applied fuzzy matching to every query term—not just untranslatable terms.

An alternative approach to cognate matching may be to decompose words in both the query and document into *n-grams* (more specifically, character-based overlapping *n*-grams), and to perform matching operations between the two sets of *n*-grams. Hedlund, Keskustalo, Pirkola, Airio, and Järvelin (2002) employed the *n*-gram technique for dealing with untranslatable query words, and McNamee and Mayfield (2002b) similarly applied *n*-grams to cognate matching.

When two languages are very different, e.g., English and Japanese, the techniques of edit distance and *n*-grams may not work well. However, in such cases, we can utilize phonetic transliteration from English words for cognate matching. Gey (2001) stated that "...we can often find that many words, particularly in technology areas, have been borrowed phonetically from English and are pronounced similarly, yet with phonetic customization in the borrowing language." By exploring a method for measuring similarity between transliteration and its original word, we may make cognate matching feasible. For example, Knight and Graehl (1998) studied automatic recognition of Japanese transliteration. The problem of Arabic transliteration was tackled by Stalls and Knight (1998). In general, the problem of matching between a source word and its transliteration has been frequently discussed for the cases of English–Japanese, English–Chinese, English–Korean, and so on (see Fujii & Ishikawa, 2001).

2.1.3. Query translation

Query translation is the most widely used matching strategy for CLIR due to its tractability. That is, the retrieval system does not have to change its inverted files of index terms in any way against queries in any language if a translation module enabling it to deal with the language of the query is incorporated. Furthermore, it is less computationally costly to process the translation of a query than that of a large set of documents (although it should be noted that, if we focus on only real-time online settings, query translation may take more time because the query must always be translated after it is entered by a user).

However, as many researchers have pointed out, it is relatively difficult to resolve term ambiguity arising from the process of translation because "queries are often short and short queries provide little context for disambiguation" (Oard & Diekema, 1998, p. 231). Term disambiguation will be discussed later.

2.1.4. Document translation

Document translation has opposite advantages and disadvantages from query translation. In CLIR experiments, this approach is not usually utilized, and query translation is dominant. However, some researchers have used it to translate large sets of documents (e.g., Braschler & Schäuble, 2001; Franz, Scott

McCarley, & Todd Ward, 2000; Oard & Hackett, 1998) since more varied context within each document is available for translation, which can improve translation quality.

Oard and Hackett (1998) reported that automatic machine translation of a set of documents using a commercial MT system outperforms query translation in an experiment of CLIR from German to English. However, the process incurred significant computational time and resources. Braschler and Schäuble (2001) similarly translated a document collection by MT software. On the other hand, Franz et al. (2000) proposed a "Fast Document Translation" algorithm for dealing with a large set of documents within a reasonable amount of processing time. The algorithm is based on a statistical approach to machine translation developed by the IBM group (Brown, Della Pietra, Della Pietra, & Mercer, 1993).

2.1.5. Interlingual techniques

Finally, in *interlingual techniques*, an intermediate space of subject representation into which both the query and the documents are converted is used to compare them. Oard and Diekema (1998) categorized latent semantic indexing (LSI) and controlled-vocabulary techniques based on multilingual thesauri as interlingual techniques. In an early work, Landauer and Littman (1990) used the LSI method to create a multidimensional indexing space for a parallel corpus of English and French documents.

Suppose that a parallel corpus includes N documents and each document has a pair of equivalent texts in two languages. We denote an $M_1 \times N$ term-document frequency matrix in one language by $\mathbf{X}_1$, where M_1 is the distinct number of terms in the language, and similarly assume that $\mathbf{X}_2$ is an $M_2 \times N$ term-document frequency matrix for another language. Each element of $\mathbf{X}_1$ and $\mathbf{X}_2$ is a frequency (or normalized frequency) of a term within the text of a document. Then we constitute an $M \times N$ matrix such that

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix},$$

where $M = M_1 + M_2$. If a singular value decomposition (SVD) such that $\mathbf{X} = \mathbf{U}\Lambda\mathbf{V}^{\mathrm{T}}$, can be computed, it allows us to calculate each document score for a query in both of the two languages as a value of inner product $\langle \mathbf{U}^{\mathrm{T}}\mathbf{d}, \mathbf{U}^{\mathrm{T}}\mathbf{q} \rangle$, where $\mathbf{d}$ is a document vector and $\mathbf{q}$ is a query vector (Rehder, Littman, Dumais, & Landauer, 1998).

A similar approach was also employed by Berry and Young (1995), Dumais, Landauer, and Littman (1996), Littman, Dumais, and Landauer (1998), Mori, Kokubu, and Tanaka (2001), etc. It should be noted that, in order to apply LSI to CLIR, a parallel (or comparable) corpus is needed to construct a language-independent indexing space.

Another type of interlingual approach is to use the "synsets" provided in WordNet, which is a well-known machine-readable thesaurus. For example, Diekema, Oroumchian, Sheridan, and Liddy (1999) employed the WordNet synset numbers as language-independent representations for CLIR. Since a synset number (label) representing a concept is corresponded to a set of concrete words in each of languages supported (e.g., English and French), it is possible that a query term in the source languages is linked to words in the target language via the synset number. Bian and Lin (2001) also utilized synsets of WordNet for English–Chinese IR.

2.2. Translation techniques

It is widely recognized that there are three main approaches to translation in CLIR:

- Machine translation (MT).
- Translation by bilingual machine-readable dictionary (MRD).
- Parallel or comparable corpora-based methods.

In addition some researchers have recently attempted to make use of WWW resources for obtaining translation equivalents.

2.2.1. Machine translation techniques

Intuitively, the MT system seems to be a fine tool for CLIR, and actually, if good MT software is available, the CLIR task becomes easier. However, in the case of query translation, the MT approach has not always shown better performance than that of dictionary-based translation. For example, Ballesteros and Croft (1998) reported that dictionary-based techniques outperformed a popular commercial MT system.

One of the reasons is that, as mentioned above, queries are often short and do not provide sufficient contextual information for translation. More practically, a query is often represented as only a set of terms, and it is difficult to expect MT systems to work well against such a poor representation. Also, MT systems usually try to select only one translation from the many candidates that the source words may have. Nie, Simard, Isabelle, and Durand (1999) pointed out that "by limiting the selection to only one word, the MT process prevents the IR system from expanding the original query by synonyms or related words."

2.2.2. Dictionary-based methods

Using a bilingual MRD is the general approach for CLIR when no commercial MT system with an established reputation is available. In general, most retrieval systems are still based on so-called "bag-of-words" architectures, in which both query statements and document texts are decomposed into a set of words (or phrases) through a process of indexing. Thus we can translate a query easily by replacing each query term with its translation equivalents appearing in a bilingual dictionary or a bilingual term list.

Ballesteros and Croft (1997) first pointed out problems with this method as follows:

- Specialized vocabulary not contained in the dictionary will not be translated.
- Dictionary translations are inherently ambiguous and add extraneous information.
- Failure to translate multiterm concepts such as phrases reduces effectiveness.

These defects can be considered the main reasons for the degradation of CLIR performance in comparison with that of mono-lingual retrieval, reported so far in the literature. For example, Hull and Grefenstette (1996) stated that "...we learn that translation ambiguity and missing terminology are the two primary sources of error..." Also, they reported that manual translation of multiword noun phrases contributes to improvement of retrieval performance. This suggests the importance of translation of multiterm concepts.

Many methods have been proposed for solving problems of term ambiguity and phrasal translation as discussed later. Furthermore, the limitation of the coverage of dictionaries can be alleviated to a certain degree by combining multiple resources such as other dictionaries or term lists generated from parallel (or comparable) corpora (see also Section 5.3).

2.2.3. Parallel corpora-based method

Parallel or comparable corpora are useful resources enabling us to extract beneficial information for CLIR. As mentioned already, the cross-language LSI approach makes use of this kind of corpus for constructing a multidimensional indexing space. We can also obtain translation equivalents directly from a parallel or comparable corpus. For example, in order to translate English queries into Spanish, Davis and Dunning (1995) extracted moderately frequent Spanish terms from Spanish documents aligned with English documents which had been searched using an English query (source query). A similar technique was tested by Yang, Carbonell, Brown, and Frederking (1998) in which they applied a technique of pseudo-relevance feedback to enhance the effectiveness of the search of the parallel corpus.

Sheridan and Ballerini (1996), Braschler and Schäuble (2000, 2001) and Molina-Salgado, Moulinier, Knudson, Lund, and Sekhon (2002) attempted to generate a *similarity thesaurus* from a comparable or parallel corpus based on associations between each pair of terms, which are computed from statistics on the number of documents in which both of the terms appear. The similarity thesaurus was used for obtaining translation equivalents of source query terms through a kind of query expansion. Similarly, Yang et al. (1998) empirically investigated the search performance of a method using a co-occurrence matrix of source and target terms for the translation of query terms.

Similar approaches of generating a bilingual term list from a parallel or comparable corpus have also been utilized by other researchers. For example, Chen, Gey, Kishida, Jiang, and Liang (1999) and Chen (2002) employed the logarithm of the likelihood ratio, $-2\log\lambda$ (Dunning, 1993) for measuring the association between a source term and a target term. McNamee, Mayfield, and Piatko (2001) seem to use a measure similar to mutual information (MI) for generating a bilingual term list from a parallel corpus. In Adriani (2002), weights of English–Dutch term pairs were computed based on Jaccard's coefficient formula.

On the other hand, some researchers in the CLIR field have attempted to estimate *translation probability* from a parallel corpus according to a well-known algorithm developed by a research group at IBM (Brown et al., 1993). The algorithm can automatically generate a bilingual term list with a set of probabilities that a term is translated into equivalents in another language from a set of sentence alignments included in a parallel corpus. The IBM algorithm includes five models, Model 1 through Model 5, of which Model 1 is the simplest and is often used for CLIR. The fundamental idea of Model 1 is to estimate each translation probability so that the probability represented such that

$$P(\mathbf{t}|\mathbf{s}) = \frac{\varepsilon}{(l+1)^m}\prod_{j=1}^{m}\sum_{i=0}^{l}P(t_j|s_i)$$

is maximized, where $\mathbf{t}$ is a sequence of terms $t_1,\ldots,t_m$ in the target language, $\mathbf{s}$ is a sequence of terms $s_1,\ldots,s_l$ in the source language, $P(t_j|s_i)$ is the translation probability, and ε is a parameter (Brown et al., 1993).

Franz, Scott McCarley, and Roukos (1999) employed a similar algorithm for estimating translation probabilities between a pair of terms. Nie (1999) attempted to compare performance between the probabilistic model, the dictionary-based approach and an MT system. Specifically, Nie et al. (1999) and Nie (2000) explored a method for estimating translation probabilities using parallel texts extracted automatically from the WWW. In addition, research groups employing the language model approach to CLIR (see below) have often utilized the IBM algorithm for computing translation probabilities (e.g., Xu, Weischedel, & Nguyen, 2001; Kraaij, 2002, etc.).

Some issues on the use of the IBM model for CLIR were described extensively with actual examples in Nie, Simard, and Foster (2001) and Nie and Simard (2002), e.g., translation of ambiguous words or compound terms, and so on. In particular, Nie and Simard (2002) explored a technique of "two-direction translation" for solving the problem that some common words are often included as higher-ranked translations of each query term. When these words are translated back to the source language, it is likely that they will be translated in various ways. Thus, a combination of translation in both directions could eliminate useless common words (however, the technique did not show good performance).

2.2.4. Use of WWW resources

As mentioned above, Nie et al. (1999) and Nie (2000) have employed WWW resources effectively for constructing parallel corpora. Resnik (1999) has also challenged the issue of automatic generation of a parallel corpus from the Web. The WWW can provide rich and ubiquitous machine-readable resources, from which we may be able to automatically extract information useful for CLIR. For example, Chen (2002) and Chen and Gey (2003) made use of a general search engine on the Internet and tried to find English trans-

lation equivalents of Chinese or Japanese terms (mainly proper nouns) by analyzing contexts of these terms in Chinese and Japanese Web documents returned by the engine. This method employs the fact that these documents often include the English equivalents of certain terms, e.g., names of places or people. It seems that there is room for applying a more sophisticated technique of natural language understanding in order to extract useful information on translation from texts of Web pages.

3. Term disambiguation techniques

3.1. Translation ambiguity

Disambiguation from among multiple alternative term translations, for which many researchers have attempted to develop effective techniques, is one of the most challenging issues in CLIR research. In general, word sense disambiguation (WSD) plays an important role in various applications such as machine translation, information retrieval and hypertext navigation, content and thematic analysis, grammatical analysis, speech processing, text processing and so on (Ide & Véronis, 1998). For CLIR, it is often necessary to disambiguate translations enumerated under each headword in a bilingual MRD or a term list generated from a parallel corpus.

If all translations listed in bilingual resources are straightforwardly employed as search terms, extraneous or superfluous terms irrelevant to the original query usually diminish the effectiveness of the search. Thus it is desirable that only relevant terms will be automatically or semi-automatically selected from a set of translations.

A simple method is to take only the translations corresponding to the first sense listed in the dictionary. Alternatively, we could investigate the frequency of each translation within a corpus and use only the most frequent translation. However, the application of such simple techniques may be insufficient to resolve the ambiguity of words having essentially different senses.

Several more sophisticated methods have been explored in the field of CLIR:

(1) Use of part-of-speech (POS) tags.
(2) Use of parallel corpus.
(3) Use of co-occurrence statistics in the target corpus.
(4) Use of the query expansion technique.

It should be noted that the term ''corpus-based disambiguation'' is often used in literature when collectively referring to techniques (2) through (4). The ''structured query model'' has also been investigated for use in improving search performance in cases where multiple translations are obtained from a bilingual dictionary (see Section 3.6).

3.2. Use of part-of-speech tags

The basic idea of using part-of-speech (POS) tags for translation disambiguation is to select only translations having the same POS with that of the source query term. Davis (1997) and Davis and Ogden (1997) applied a part-of-speech tagger to English queries for solving translation ambiguity in CLIR tasks from English to Spanish, i.e., only if the POS tag of a Spanish equivalent listed in an English–Spanish dictionary is coincident with that of the English query term, the translation was selected as a search term. A similar method was tested by Ballesteros and Croft (1998). This method requires that POS tagging software is available for both languages.

3.3. Parallel corpus-based disambiguation

A parallel corpus was used for determining the "best" translation or set of translations by Davis (1997, 1998), where a single translation for each source term was selected from a set of translations listed in an MRD according to the result of searching a parallel corpus. The procedure is as follows:

(1) Pick a set of translations for each term in the source (English) query from an MRD. (2) Search a part of the parallel corpus written in the target language (Spanish) for each translation respectively, and save each set of Spanish documents. (3) Search the English part of the parallel corpus for the source query. (4) Select a single translation of which set of documents is closest to the set of documents searched by the source query.

This procedure is repeated for each query term so as to obtain a final set of the "best" translations. Boughanem and Nassr (2001) and Boughanem, Chrisment, and Nassr (2002) employed a parallel corpus in a similar way for disambiguating translations obtained by dictionary search.

This technique was also tested by Ballesteros and Croft (1998) with a slight modification in which at first, some top-ranked documents were determined by searching the part of the corpus written in the source language for the original query. Next, some top-ranked terms were extracted directly from the target language text of each top-ranked document based on the Rocchio method (see below). Finally, the "best" translation was selected according to the scores in the ranked term list.

3.4. Disambiguation based on co-occurrence statistics

Ballesteros and Croft (1998) pointed out that "the correct translations of query terms should co-occur in target language documents and incorrect translations should tend not to co-occur." Suppose that two terms, "Mercury" and "planet," are included in the source language (English) query. Although the term "Mercury" (or "mercury") has multiple senses, it is clear that "Mercury" is unrelated to mythology or chemistry in the context of the given query due to an existence of the other term "planet." Thus it is reasonably expected that a correct translation equivalent of "Mercury" will tend to co-occur with a correct equivalent of "planet" in the target document collection. In order to detect a set of correct translations from the target collection, Ballesteros and Croft (1998) employed a variation of *mutual information* (MI) computed from co-occurrence of a pair of two terms. According to the values for all possible pairs of two translation equivalents, we may choose a correct combination of translations. In Ballesteros and Croft (1998), this technique was actually applied to the disambiguation of phrasal translation (see below).

In order to disambiguate translation equivalents using MI or other similarity measures, we first have to compute each degree of similarity for all pairs of possible translations, and secondly must select a pair of two translations that have the maximum value. After that, the source query terms for which translations were determined are excluded, and the same procedure is repeated until translations of all source terms are selected. Similar techniques with some modifications have been described by many subsequent researchers, e.g., Lin, Lin, Bian, and Chen (1999), Gao et al. (2001a, 2001b, 2002), Maeda, Sadat, Yoshikawa, and Uemura (2000), Adriani (2001), Sadat, Maeda, Yoshikawa, and Uemura (2002), Qu, Grefenstette, and Evans (2003), Seo, Kim, Kim, Rim, and Lee (2003), etc. For example, Maeda et al. (2000) selected all combinations of translations for which the average of similarity degrees for pairs included in the combination exceeds a threshold.

However, it is computationally costly to find the "best" translations by term similarity if many terms are included in the source query because a large number of possible combinations of translation candidates must be considered. To alleviate this problem, Gao et al. (2001a, 2001b) proposed an approximate algorithm using "cohesion" of a term t with a set T of other terms such that

$$\mathrm{Cohesion}(t, T) = \max_{t' \in T} \mathrm{SIM}(t, t'),$$

where $\mathrm{SIM}(t, t')$ is a similarity measure between a pair of terms. The basic idea is to avoid processing similarity values for all combinations of translations by considering only its maximum within each set of translations for the other source terms. A similar approach with modification was also employed by Seo et al. (2003). Furthermore, Gao et al. (2002) introduced the "decaying co-occurrence model" in order to take into account the distance between a pair of terms and also tried to incorporate a syntactic relation into each pair of terms such as "(have, sub-verb, I)."

However, it is possible to unexpectedly generate false combinations by simply using co-occurrence statistics because translation equivalents of two unrelated terms in the context of a given source query may occur frequently together in the target corpus (Yamabana, Muraki, Doi, & Kamei, 1998). As a result, pairs of "incorrect" translations can be selected, and would significantly degrade retrieval performance. Yamabana et al. (1998) employed co-occurrence statistics in a set of texts written in the source language in order to reduce false combinations. First, the two most related terms in the query were determined based on co-occurrence statistics in the source language corpus, and then the "best" translations were selected from all pairs of translations of these two terms according to co-occurrence statistics in the target language corpus. It should be noted that these two corpora do not have to be parallel or comparable.

Federico and Bertoldi (2002) proposed a query translation model incorporating a hidden Markov model (HMM) for estimating joint probability of two sequences of source terms and translations. In the model, a transitive probability $P(s|s')$, where s and s' are two consecutive source terms, is estimated from co-occurrences in the target corpus. We can expect that this probability works as a device for translation disambiguation.

Finally, we should pay attention to other sources from which co-occurrence statistics can be generated. For example, Maeda et al. (2000) and Qu et al. (2003) have attempted to make use of a Web search engine in order to obtain co-occurrence frequencies. They sent a combination of possible translations to a Web portal (e.g., AltaVista), and employed the number of pages it returned as a score for selecting translations.

3.5. Query expansion for disambiguation

Pseudo relevance feedback (PRF), also known as blind feedback, is widely recognized as an effective technique for enhancing performance of information retrieval. Usually, a set of search terms in a query given by a user is expanded by adding terms automatically selected from top-ranked documents that are searched for the original query. Ballesteros and Croft (1997, 1998) have empirically shown that PRF also works effectively for CLIR tasks.

In the case of CLIR, two kinds of PRF are feasible:

- Pre-translation feedback and
- Post-translation feedback

First, documents from a corpus in the source language can be retrieved prior to translation in order to add a set of new terms to the source query (*pre-translation feedback*) if such a corpus is available. Second, after translation, standard PRF can be applied using the target document collection (*post-translation feedback*). Ballesteros and Croft (1997) suggested that pre-translation feedback may contribute to improvement of precision. This is due to the fact that the PRF is basically done using the entire query—not each source term respectively. That is, synonyms or related terms corresponding to the "correct" meaning of each source term within a context of the query are expected to be automatically added through the PRF process. Thus it is possible that "correct" translations will be obtained from an MRD or a bilingual term list in the subsequent process of translation. Recently, McNamee and Mayfield (2002a) reported that pre-translation

query expansion is very useful when lexical coverage of translation resources is poor. On the other hand, post-translation feedback can be considered a device for improving recall ratio, as shown in standard experiments of monolingual retrieval.

In CLIR, two well-known methods for weighting terms in the top-ranked documents are often utilized for selecting "good" terms, i.e., the Rocchio method and the probabilistic method. In the Rocchio method, which is based on the vector space model, an original weight of a term is modified by adding the average of weights of the term in the set of relevant documents and by subtracting the average of weights in the set of irrelevant documents. It should be noted that in PRF, some number of top-ranked documents are presupposed to be relevant.

On the other hand, based on the probabilistic model of Robertson and Sparck Jones (1976), a weight w_j of term t_j is computed such that

$$w_j = r_j \times \tau_j$$

and

$$\tau_j = \log \frac{(r_j + 0.5)(N - R - n_j + r_j + 0.5)}{(N - n_j + 0.5)(R - r_j + 0.5)},$$

where r_j is the number of relevant documents including t_j, n_j is the number of all documents including t_j, R is the number of all relevant documents, and N is the total number of documents included in the database.

Some attempts have made to improve performance of the probabilistic feedback method. For example, Sakai, Koyama, Suzuki, and Manabe (2003) used an alternative form such that

$$w_j = \sqrt{r_j} \times \tau_j$$

and also proposed other variations incorporating scores of relevant documents.

3.6. Structured query model

Another idea for dealing with multiple translations for each source query term is to consider them as a set of synonyms. From this point of view, conventional Boolean logic may be useful. Hull (1998) suggested that "Boolean disjunction (the OR operator) is a natural way to link together many translation equivalents without dramatically increasing the weight of the underlying concept." Hull (1998) actually proposed a probabilistic indexing model enabling us to rank documents based on a query in which Boolean operators are incorporated.

Meanwhile, Pirkola (1998) attempted to use the synonym operator #SYN( ) provided in the INQUERY system (Callan, Croft, & Broglio, 1995; Turtle & Croft, 1991) for each set of translation equivalents of source query terms. Use of #SYN( ) can prevent overweighting certain source query terms with multiple translations of source terms, which are often non-specific terms useless for search. Pirkola (1998) used both the #SYN operator and the #UWN operator, a kind of proximity operator for treating compound terms, and showed empirically the effectiveness of the approach. Pirkola's structured query model (*Pirkola's method*) has been subsequently used by other research groups (Darwish & Oard, 2003; Hedlund, Keskustalo, Pirkola, Sepponen, & Järvelin, 2001a; Pirkola, Hedlund, Keskustalo, & Järvelin, 2001; Sperer & Oard, 2000). Darwish and Oard (2003) developed a method for incorporating translation probability into Pirkola's method.

To be precise, the structured query model should not be classified with disambiguation techniques, but it has the same effect with that of disambiguation in that the model can enhance performance of search for a query including many ambiguous translations.

3.7. Another method for disambiguation

Boughanem et al. (2002), explored a "bi-directional translation" technique in which a form of backward translation is used for ranking translation candidates. Suppose that we need to translate English query terms into French ones. In "bi-directional translation," first a set of French equivalents for an English term is found in an English–French dictionary. Next, using a French–English dictionary, each French equivalent is reversely translated into a set of English terms. Basically, if the set includes the original source term, the French translation equivalent is chosen as a preferred translation.

3.8. Phrasal translation techniques

As Ballesteros and Croft (1997) pointed out that "...failure to translate multiterm concepts as phrases reduces effectiveness," *phrasal translation* is certainly significant for CLIR. The basic technique is to search a bilingual dictionary or a term list including phrases or compound words as headwords. Phrases or compound words can be automatically identified in the source query by matching operations against headwords of an MRD, or by using part-of-speech tags. For example, we can assume a word combination of "noun–noun" or "adjective–noun" to be a compound word. Thus the identified compound word can be replaced with the corresponding term in the target language by searching in a bilingual dictionary.

However, it is inevitable that we will be confronted with insufficient coverage of lexical resources to be used. When an untranslatable phrase is found in the source query, we are forced to execute word-by-word translation, which, incidentally, causes a term ambiguity problem. Suppose that a phrase consisting of two words appears in the source query where one has two translations and the other has three from an MRD. In this case, we have to exclude erroneous phrasal translations from the six possible combinations. This is an example of the problem of ambiguity in phrasal translation.

A simple way is to automatically adopt all combinations as multiword expressions. Alternatively, Ballesteros and Croft (1997, 1998) proposed a sophisticated disambiguation method based on co-occurrence statistics mentioned above and reported that the statistical method shows good performance. Theoretically, phrases or compounds can be categorized as compositional or non-compositional. According to Pirkola et al. (2001), "compounds whose meaning can be derived from the meanings of component words are called compositional compounds," and "a compound whose meaning cannot be deduced on the basis of its component is called a non-compositional compound." Thus, clearly the co-occurrence statistics-based method has limitations for detecting translations of non-compositional phrases. It would be promising to combine both dictionary-based and statistical methods in order to compensate for the drawbacks of each method.

4. Formal model for CLIR

4.1. Retrieval model for estimating document scores

If the query is translated into the language used in the document collection and the weight of each query term is estimated in the translation process, the following procedure is almost the same as that of usual ad hoc retrieval. Typically, document scores for ranked output are calculated using inverted files, and documents are sorted in decreasing order of scores.

Thus, in CLIR, standard retrieval models or algorithms for estimating document scores such as the vector space model (Buckley, Allan, & Salton, 1994), Okapi formula (Robertson, Walker, Jones, Hancock-Beaulieu, & Gatford, 1995), INQUERY (Turtle & Croft, 1991), logistic regression model (Cooper, Chen, & Gey, 1994), etc., have been employed. In particular, as mentioned above, the INQUERY system may hold a unique position in CLIR because the structured query approach can be implemented by using special

functions of the INQUERY system (#SYN or #UWN operators). PIRCS (Kwok, 1996), a kind of probabilistic model, has often been used specifically for Chinese retrieval tasks (Kwok, Grunfeld, Dinstl, & Chan, 2001).

Recently, some researchers have attempted to apply the so-called language model (LM) to CLIR tasks. Originally, the LM has been explored in the field of statistical natural language processing as a potentially useful tool for the data-driven approach. Since IR tasks can be essentially considered a kind of natural language processing of query statements and document texts, it may be natural to apply the LM to problems of IR.

4.2. Language model for CLIR

The *language model* is basically the probabilistic distribution of sequences of "words" (Manning & Shütze, 1999, p. 71). A way to apply the model to document ranking is to estimate "the probability of producing the query given a language model of a document," and to use the value as a document score (Hiemstra, 1998; Ponte & Croft, 1998). A simple form of the probability can be written such that

$$P(Q|d) = \prod_{t \in Q} (1 - \lambda)P(t) + \lambda P(t|d),$$

where Q is a set of query terms, d is a given document and λ is a parameter ($0 \leqslant \lambda \leqslant 1$). Intuitively, it is enough to make use of only a conditional probability $p(t|d)$ for estimating the probability of producing the query given the document. However, the probability that the term is generated in general, $p(t)$, is incorporated in order to prevent the probability of a query term not appearing in the document from becoming zero. When weights of each query term are incorporated, the above formula is slightly changed (see Miller, Leek, & Schwartz, 1999b for details).

One of the advantages of the language model approach to CLIR tasks is to enable us to put translation probability $p(t|s)$ directly into the formula where s is a source term and t is a target term. There are two ways to incorporate the translation probability:

$$P(Q|d) = \prod_{t \in Q} (1 - \lambda)P(t) + \lambda \sum_{s} P(t|s)P(s|d), \tag{I}$$

$$P(Q|d) = \prod_{t \in Q} \sum_{s} P(t|s)[(1 - \lambda)P(s) + \lambda P(s|d)]. \tag{II}$$

The formula (I) has been used in Xu et al. (2001), Fraser, Xu, and Weischedel (2003) and Franz and Scott McCarley (2003). It should be noted that formula (I) was originally derived from an application of the Hidden Markov Model (HMM) to an IR problem (Miller, Leek, & Schwartz, 1999a, 1999b). Since then, formula (II) has been employed by Hiemstra and Kraaij (1999), Kraaij, Pohlmann, and Hiemstra (2000), etc. With respect to (II), Hiemstra, Kraaij, Pohlmann, and Westerveld (2001) proposed a new relevance feedback method for CLIR using the language model. The main difference between the two formulas is that in formula (I), a corpus in the language of the query is needed for estimating probability $P(t)$.

Kraaij (2002) also examined a variation of LM estimating $P(Q, d)$, not $P(Q|d)$, for modeling an empirical fact that longer documents have a higher probability of relevance (i.e., $P(Q, d) = P(Q|d)P(d)$). Furthermore, Lavrenko, Choquette, and Croft (2002) explored another formal method for applying language modeling in which a model of the source topic in the target document is estimated based on a "relevance model," which specifies how often we expect to see any given word in the documents relevant to the query.

The translation probabilities can be estimated by the following methods:

- EM algorithm developed by IBM group (Brown et al., 1993).
- Use of information in a bilingual dictionary.

A simple method of using information in a dictionary is to count the number of translations for each source term. For example, if a source term s has n translations, $t_1, \ldots, t_n$, the translation probabilities can be assumed such that $P(t_i|s) = 1/n$ $(i = 1, \ldots, n)$ uniformly (Xu et al., 2001). On the other hand, for weighting translations, Hiemstra and Kraaij (1999) seem to use a more complicated procedure based on the number of distinct "senses" of the source term that each translation covers.

5. Other research topics in CLIR

5.1. Pivot language approach

So many languages are spoken in the world, and it is not always possible to obtain the bilingual resources we need for a particular pair of languages. A promising technique to circumvent the problem of limited availability of linguistic resources would be the *pivot language approach*, in which an intermediate language acts as a mediator between two languages for which no bilingual resource is available. Suppose that a CLIR task between Japanese and Dutch is requested by a user. In this case, machine-readable resources of Japanese–Dutch pairs may be unavailable, and it would be easier to find Japanese–English and Dutch–English resources since English is such a widely used language. Thus CLIR between Japanese and Dutch can be performed via English (as an intermediary) without direct bilingual resources of Japanese and Dutch.

The pivot language approach may also alleviate the problem of explosive combinations of languages, i.e., if we have to perform CLIR between each pair of n languages, $O(n^2)$ resources are needed. However, the pivot language approach enables us to handle the complex tasks with only $O(n)$ resources (Gey, 2001).

A basic way of using the pivot language approach would be a *transitive translation* of a query using two bilingual dictionaries (Ballesteros, 2000). In the case of search from Japanese to Dutch via English, if Japanese–English and English–Dutch dictionaries are available, CLIR can be performed by replacing Japanese query terms with the corresponding English equivalents and successively substituting the English equivalents with the Dutch equivalents. Of course, if Japanese–English and English–Dutch MT systems can be used, a similar transitive translation is also feasible.

With the case of dictionary-based transitive translation, translation ambiguity can become a more serious problem. It is possible that resulting translations become doubly ambiguous if each replacement stage yields ambiguity: (1) from the source language to the intermediate language and (2) from the intermediate language to the target language. Suppose, for example, that a Japanese source query consists of four words, and every word has four English equivalents. In addition, if every English equivalent has four Dutch equivalents, simple replacements are going to produce 64 $(=4^3)$ search terms in total from only 4 source terms, which would inevitably contain some irrelevant translations. To solve this problem, Ballesteros (2000) attempted to apply the disambiguation methods mentioned above (co-occurrence-based method, query expansion, etc.) to transitive translation and attained a substantial improvement in search performance. Gollins and Sanderson (2001) also proposed a technique called "lexical triangulation" to alleviate the translation ambiguity problem in which two pivot languages are used independently and removal of erroneous translations is attempted by taking only translations in common from two ways of transitive translation using two pivot languages.

The pivot language approach has been utilized in TREC, NTCIR, and CLEF due to unavailability of bilingual resources. For example, the following transitive combinations of languages have been explored:

- English > French > German (Franz et al., 1999)
- French > English > German, etc. (Gey, Jiang, Chen, & Larson, 1999)

- German > English > Italian (Hiemstra & Kraaij, 1999)
- Japanese > English > Chinese (Lin & Chen, 2003)
- Chinese > English > Japanese (Chen & Gey, 2003)

In particular, Franz et al. (1999) proposed some interesting techniques for searching German documents with English queries:

(1) Convolution of translation probability: Estimating translation probability from an English term e to a German term g through French terms f such that

$$P(g|e) = \sum_f P(g|f)P(f|e).$$

(2) Automatic query generation from the intermediate language corpus: Generating French queries automatically by simply merging all non-stopwords in the top-ranked French documents searched by the English–French CLIR system, and putting the French query into the French–German CLIR system.

5.2. Merging strategy for multilingual information retrieval

Suppose that we have a multilingual document collection in which two or more languages are mixed (not a parallel corpus), and a user wishes to search the collection for a query expressed in a single language. This task is more complicated than simple bilingual CLIR. In CLEF and NTCIR, multilingual CLIR has been adopted as a research task, and many research groups have worked on the issue.

Basically, there are two approaches for multilingual IR (Lin & Chen, 2003):

- *Distributed architecture* in which the document collection is separated by language, and each part is indexed and retrieved independently.
- *Centralized architecture* in which the document collection in various languages is viewed as a single document collection and is indexed in one huge index file.

In distributed architectures, a standard bilingual search is repeatedly performed for each separate language sub-collection respectively, and several ranked document lists are generated by each run. Then the problem becomes how to merge the results of each run into a single ranked list so that all relevant documents in any language are successfully ranked. Essentially, the *merging strategy* is a general research issue of IR when searching distributed resources (i.e., distributed IR), in which it is inevitably necessary to merge ranked lists obtained from each resource. In CLIR, the following merging strategies have been investigated:

- *Raw score*: straightforwardly using document scores estimated in each run.
- *Round robin*: interleaving each document list in a round robin fashion by assuming that distribution of relevant documents is identical among the lists.
- *Normalized score*: normalizing document scores by each run in order to remove effects of collection-dependent statistics on estimation of the scores.
- *Rank-based score*: mathematically converting ranks in each run into scores by assuming a relationship between the rank and probability of relevance.
- *Modified score*: modifying raw scores in each run so as to reduce effects of collection-size dependency, translation ambiguity, etc.

If the retrieval model employed for each run can estimate relevance probability of each document correctly, it would be reasonable to re-rank all documents together according to values of the probability (i.e., raw scores). For example, Chen and Gey (2003) simply merged the results from Chinese, Japanese and English collections according to values of probability of relevance estimated by the logistic regression model.

However, in most cases, it would be difficult to consider each document score to be a pure probability of relevance even if a probabilistic retrieval model was actually used. In this case, if we can assume that relevant documents are distributed in the same way in every separate language sub-collection, a simple strategy is round robin-based merging, in which only the rank of each document is taken into account. Otherwise, an alternative method is to use normalized document scores such that

$$v = (v - v_{\min})/(v_{\max} - v_{\min}),$$

where v is a raw score, and $v_{\min}$ and $v_{\max}$ are the minimum and maximum in each run respectively (Powell, French, Callan, Connell, & Viles, 2000). Savoy (2002) has empirically compared search performance among the four strategies of round robin, raw score, normalized score and the CORI approach (see Callan et al., 1995 for details) using the CLEF test collection and reported that normalized score is dominant among them. Similarly, Moulinier and Molina-Salgado (2002) tried to conduct comparisons among round robin, raw score, CORI, normalized score and collection-weighted normalized score (a variation of normalized score), and reported that collection-weighted normalized score showed higher mean average precision.

Other techniques for estimating optimal scores for merging ranked lists have been proposed. Franz et al. (2000) empirically found a linear relationship between log of rank and precision at the rank and used scores that are converted according to the relationship for merging results from each run. Similarly, the strategy of rank-based scoring was investigated in Kraaij et al. (2000). Hiemstra et al. (2001) also examined the effectiveness of modifying raw scores so as to remove effects of collection-size dependency in the process of estimating raw scores. Meanwhile, Lin and Chen (2003) proposed a method of modifying raw scores based on the degree of ambiguity when each source query was translated, according to an assumption that a good translation may give much more relevant documents. Savoy (2003a) tested a logistic regression formula for predicting a relevance probability of a document given a rank and a score of the document.

On the other hand, for the centralized architecture, the set of multilingual documents is not divided into sub-collections for each language. In order to search such a heterogeneous collection, we need either

(1) to translate the source query into all languages included in the document collection and to merge all translations into a single query, or
(2) to translate the documents into a single language used in the query.

Gey et al. (1999), Chen (2002) and Nie and Jin (2003) employed the first method for searching the CLEF test collection. With this method, it may be necessary to adjust idf factors because documents in a language having fewer documents may take advantage of weighting by document frequency (Lin & Chen, 2003).

5.3. Combination of some language resources

Needless to say, quality and coverage of language resources for translation significantly affect search performance of CLIR. Specifically, in the case of searches between two unrelated languages in which cognate matching has almost no effect, e.g., Japanese and German, poor lexical coverage of the bilingual dictionary or term list to be used could be a fatal factor leading to low performance because there is almost no means to deal with untranslatable terms. Actually, McNamee and Mayfield (2002a) experimentally confirmed a conjecture that retrieval performance drops with decreased lexical coverage when using the CLEF test collection.

A promising approach to the problem of poor lexical coverage is to merge results of translations from multiple language resources. Xu et al. (2001) and Darwish and Oard (2003) combined three distinct types of translation resources assuming equal weights of each resource, i.e., final translation probability of each pair of terms was calculated as an average of three values of probability obtained from each resource.

Jones and Lam-Adesina (2002) also explored techniques of data fusion and query combination (Belkin, Kantor, Fox, & Shaw, 1995) for putting together different translated outputs from several MT systems. In the case of data fusion, two document scores computed from outputs by two MT systems were simply summed for each document. In the case of query combination, before the estimation of document scores, a single query was formed by taking the unique terms from two outputs. The authors discuss the mathematical characteristics of the two techniques.

5.4. Language processing issues

5.4.1. Text processing for languages other than English

Traditionally, research articles on IR in international journals have largely concentrated on searching English documents for English queries. It seems that research findings of IR techniques for other languages have been mainly published in the countries in which the language is spoken as the mother tongue. However, in order to attack issues of CLIR tasks, IR researchers need to share knowledge on text processing of various languages other than English. Thus papers on this topic have been published in international conferences or journals, and many language resources (e.g., MRDs, stopword lists, stemmers, morphological analyzers, etc.) for various languages have become available on the Internet.

A typical procedure for processing the text of queries and documents in each language is as follows.

(1) Tokenization of text into a set of terms
(2) Assignment of part-of-speech tag
(3) Stopword removal
(4) Lemmatization
(5) Stemming
(6) Noun phrase extraction

5.4.2. Tokenization

Complexity of text processing depends on which language is targeted. For example, as well-known in the case of the Chinese and Japanese languages, there is no explicit boundary between words in each sentence. Thus tokenization of text in such languages is more complicated than that of European languages.

However, it might be necessary to solve similar problems in some European languages. For example, decomposition of compound words is significant for mono- or cross-language retrieval of German or Swedish (Hedlund, Pirkola, & Järvelin, 2001b). Specifically, it seems that techniques for decomposing German compound words have been explored by many researchers.

5.4.3. Stopword list

There are two methods for creating a stopword list in a new language: statistical and linguistic. The statistical approach, taken by Buckley et al. (1998), simply does a frequency count for all words in some document collection in that language and chooses the top n words by frequency to be the stop list. The linguistic approach, assuming a bilingual dictionary is available, would be to translate an English stopword list into the target language and use that list in processing the new language.

A general guideline for creating a stopword list was given by Fox (1990), and English and French stopword lists are available in Fox (1990) and Savoy (1999). Savoy (2003b) also described a procedure for creating German and Italian stopword lists in line with the guidelines by Fox (1990).

5.4.4. Stemming

Porter's algorithm (Porter, 1980) is widely used to stem English words in IR. Although effectiveness of normalization by stemming for English monolingual IR has not yet been shown explicitly (Frakes, 1992), in the case of morphologically rich or lexically complex languages other than English, it seems that the use of stemmers brings about greater improvement of retrieval performance (Savoy, 2003b; Sheridan & Ballerini, 1996). For CLIR task in which matching operations between terms of different languages are needed at various stages of processing, the development of effective stemmers is certainly important to enhance search performance. We can obtain a number of rule-based stemmers for European languages from Porter's SNOWBALL project (http://snowball.tartarus.org/), in which Porter (2001) provides an excellent description of the components which are essential in the creation of a good rule-based stemmer.

Oard, Levow, and Cabezas (2001) employed a four-stage "backoff translation" for locating a term within a translation lexicon, in which four matching operations are performed: (1) matching of the surface form of a term to surface forms of headwords in the lexicon, (2) matching of the stem of a term to surface forms of headwords, (3) matching of the surface form of a term to stems of headwords, (4) matching of the stem of a term to stems of headwords.

In addition, Oard et al. (2001) have proposed a "statistical stemming" approach in order to automatically extract information on suffixes from a text collection. This approach can be taken as a special case of unsupervised acquisition of morphology in the field of computational linguistics (Oard et al., 2001).

5.5. User interfaces for interactive CLIR

Although most of the research literature on CLIR implicitly treats the search as a task to be performed automatically by a machine, a practical approach for providing better search results would be to develop systems in which humans and machines interact (Oard, 2001, p. 58). In order to accomplish better *interactive CLIR*, a well-designed user interface would play an important role. An early interactive CLIR system for English–Spanish, QUILT, accommodates some functions and GUI for supporting query translation (Davis & Ogden, 1997):

- Options for displaying the query translation terms from a bilingual lexicon.
- Pop-up windows that show the retrieved Spanish document with translated Spanish query terms highlighted.
- A Pop-up window that shows the variant translations of each English term.

QUILT also seems to have a feature that displays English gloss translations of Spanish documents retrieved by the system. A more recent version of QUILT was described by Davis and Ogden (2000). Some other systems have been developed such as FromTo-CLIR (Kim et al., 1999), MULINEX (Capstick et al., 2000) and so on. Peters and Sheridan (2001) also listed several working CLIR systems.

In CLEF-2001, a challenging track for exploring interactive applications of CLIR, iCLEF, was included. The track was concerned with current technology for supporting interactive relevance assessment, and several research groups participated (Oard & Gonzalo, 2002). The interactive CLLR track has been continued in following CLEF campaigns.

5.6. Evaluation of CLIR

In order to develop better techniques or methods for automatic or interactive CLIR, a continuing sequence of experimental evaluations is indispensable. From this viewpoint, we must admire the efforts of TREC, CLEF and NTCIR and their huge contributions toward significantly promoting and enhancing CLIR research. Descriptions of the systems and findings of these retrieval experiments can be found in the working notes and proceedings of these activities.

Many techniques described in this article were proposed and experimentally tested in the campaigns organized by these three initiatives. Very useful research findings on the performance of the CLIR techniques have been cumulated through the evaluation process. Unfortunately, the provision of details of performance levels is outside the scope of this review.

Methodology used for evaluation is also an important topic for CLIR research. Standard Cranfield-type methods have been basically used to assess CLIR experiments in TREC, CLEF and NTCIR. However, it should be noted that CLIR experiments have a unique characteristic, that is, the performance of search runs executed using queries in the language of document collections can be employed as a baseline for the evaluation. For example, results of English to Japanese CLIR runs can be evaluated by comparing them with those of Japanese monolingual runs if corresponding Japanese queries are correctly prepared by human translators. We can usually assume that the monolingual runs give us an upper limit of performance. The overviews of the TREC, CLEF and NTCIR, and the introductory paper of this special issue provide us with more useful information on evaluation methodologies for CLIR tasks.

6. Concluding remarks: future directions for research

Through a review of the literature, this paper has described research issues on CLIR and discussed various techniques that can be adopted. As mentioned in the introduction, this review does not cover all research works. A number of papers or articles on CLIR not referred to in this review have been published in various research fields and communities. We can cite Oard and Dorr (1996), Oard and Diekema (1998), Peters and Sheridan (2001), and Fujii and Ishikawa (2001) as sources for identifying additional papers.

The last issue we should discuss is the future direction of CLIR research. What is the goal of CLIR? What should the next steps be to achieve the goal? In the workshop on CLIR held at SIGIR 2002, the organizers presented three challenges (Gey, Kando, & Peters, 2002):

1. Where to get resources for resource-poor languages?
2. Why do we not have a sizeable Web corpus in multiple languages?
3. Why aren't search engines using our research?

As a possible answer for the third question, they stated that "if users are presented with a ranked list of documents that they cannot read, what is the utility?" This is a crucial point for considering the future direction of CLIR research. That is, we may need to make a plan after having a clear grasp of information needs of users on CLIR and explicitly delineating realistic utility when applications of CLIR are employed by the actual users.

Meanwhile, various interesting areas for CLIR research seem to remain, e.g., CLIR for multimedia data, cross-language question answering, cross-language filtering, cross-language topic detection and tracking, cross-language summarization, cross-language document clustering, and so on. This review article cannot cover all the state-of-the-art research in these areas where substantive research has already been performed. The CLIR researchers may have to carefully select future directions from many possibilities in order to enable the actual users to effectively and efficiently satisfy their information needs.

Acknowledgement

The author would like to thank the referees and the editors for useful suggestions.

References

Adriani, M. (2001). Ambiguity problem in multilingual information retrieval. In C. Peters (Ed.), *Cross-language information retrieval evaluation*. *LNCS* (2069, pp. 156–165). Berlin: Springer-Verlag.

Adriani, M. (2002). English–Dutch CLIR using query translation techniques. In C. Peters et al. (Eds.), *Evaluation of cross-language information retrieval systems*. *LNCS* (2406, pp. 219–225). Berlin: Springer.

Ballesteros, L. A. (2000). Cross-language retrieval via transitive translation. In W. B. Croft et al. (Eds.), *Advances in information retrieval: recent research from the center for intelligent information retrieval* (pp. 203–234). Boston, MA: Kluwer.

Ballesteros, L., & Croft, W. B. (1997). Phrasal translation and query expansion techniques for cross-language information retrieval. In *Proceedings of the 20th ACM SIGIR conference on research and development in information retrieval* (pp. 84–91).

Ballesteros, L., & Croft, W. B. (1998). Resolving ambiguity for cross-language retrieval. In *Proceedings of the 21st ACM SIGIR conference on research and development in information retrieval* (pp. 64–71).

Belkin, N. J., Kantor, P., Fox, E. A., & Shaw, J. A. (1995). Combining the evidence of multiple query representations for information retrieval. *Information Processing & Management, 31*, 431–448.

Berry, M. W., & Young, P. G. (1995). Using latent semantic indexing for multilanguage information retrieval. *Computers and the Humanities, 29*(6), 413–429.

Bian, G. W., & Lin, C. C. (2001). Trans-EZ at NTCIR-2: synset co-occurrence method for English–Chinese cross-lingual information retrieval. In *Proceedings of the second NTCIR workshop on research in Chinese & Japanese text retrieval and text summarization*. Tokyo: National Institute of Informatics. Available: http://research.nii.ac.jp/ntcir/workshop/.

Boughanem, M., & Nassr, N. (2001). Mercure at CLEF-1. In C. Peters (Ed.), *Cross-language information retrieval evaluation*. *LNCS* (2069, pp. 202–209). Berlin: Springer-Verlag.

Boughanem, M., Chrisment, C., & Nassr, N. (2002). Investigation on disambiguation in CLIR: aligned corpus and bi-directional translation-based strategies. In C. Peters, et al. (Eds.), *Evaluation of cross-language information retrieval systems*. *LNCS* (2406, pp. 158–168). Berlin: Springer-Verlag.

Braschler, M., & Schäuble, P. (2000). Using corpus-based approaches in a system for multilingual information retrieval. *Information Retrieval, 3*, 273–284.

Braschler, M., & Schäuble, P. (2001). Experiments with the Eurospider retrieval system for CLEF 2000. In C. Peters (Ed.), *Cross-language information retrieval evaluation*. *LNCS* (2069, pp. 140–148). Berlin: Springer-Verlag.

Brown, P. F., Della Pietra, S. A., Della Pietra, V. J., & Mercer, R. L. (1993). The mathematics of statistical machine translation: parameter estimation. *Computational Linguistics, 19*(2), 263–311.

Buckley, C., Allan, J., & Salton, G. (1994). Automatic routing and ad-hoc retrieval using SMART: TREC2. In *Proceedings of TREC-2*. Gaithersburg, MD: National Institute of Standards and Technology. Available: http://trec.nist.gov/pubs/.

Buckley, C., Walz, J., Mitra, M., & Cardie, C. (1998). Using clustering and superconcepts within SMART: TREC-6. In *Proceedings of TREC-6*. Gaithersburg, MD: National Institute of Standards and Technology. Available: http://trec.nist.gov/pubs/.

Callan, J. P., Croft, W. B., & Broglio, J. (1995). TREC and TIPSTER experiments with INQUERY. *Information Processing & Management, 31*(3), 327–343.

Capstick, J., Diagne, A. K., Erbach, G., Uszkoreit, H., Leisenberg, A., & Leisenberg, M. (2000). A system for supporting cross-lingual information retrieval. *Information Processing & Management, 36*, 275–289.

Chen, A. (2002). Multilingual information retrieval using English and Chinese queries. In C. Peters et al. (Eds.), *Evaluation of cross-language information retrieval systems*. *LNCS* (2406, pp. 44–58). Berlin: Springer-Verlag.

Chen, A., & Gey, F. C. (2003). Experiments on cross-language and patent retrieval at NTCIR-3 workshop. In *Proceedings of the third NTCIR workshop on research in information retrieval, automatic text summarization and question answering*. Tokyo: National Institute of Informatics. Available: http://research.nii.ac.jp/ntcir/workshop/.

Chen, A., Gey, F. C., Kishida, K., Jiang, H., & Liang, Q. (1999). Comparing multiple methods for Japanese and Japanese–English text retrieval. In *Proceedings of the first NTCIR workshop on research in Japanese text retrieval and term recognition*. Tokyo: National Institute of Informatics. Available: http://research.nii.ac.jp/ntcir/workshop/.

Cooper, W. S., Chen, A., & Gey, F. C. (1994). Full text retrieval based on probabilistic equations with coefficients fitted by logistic regression. In *Proceedings of TREC-2*. Gaithersburg, MD: National Institute of Standards and Technology. Available: http://trec.nist.gov/pubs/.

Darwish, K., & Oard, D. W. (2003). CLIR experiments at Maryland for TREC-2002: evidence combination for Arabic–English retrieval. In *Proceedings of TREC-2002*. MD: National Institute of Standards and Technology. Available: http://trec.nist.gov/pubs/.

Davis, M. (1997). New experiments in cross-language text retrieval at NMSU's Computing Research Lab. In *Proceedings of TREC-5*. Gaithersburg, MD: National Institute of Standards and Technology. Available: http://trec.nist.gov/pubs/.

Davis, M. W. (1998). On the effective use of large parallel corpora in cross-language text retrieval. In G. Grefenstette (Ed.), *Cross-language information retrieval* (pp. 12–22). Boston, MA: Kluwer.

Davis, M., & Dunning, T. (1995). A TREC evaluation of query translation methods for multi-lingual text retrieval. In *Proceedings of TREC-4*. Gaithersburg, MD: National Institute of Standards and Technology. Available: http://trec.nist.gov/pubs/.

Davis, M. W., & Ogden, W. C. (1997). QUILT: implementing a large-scale cross-language text retrieval system. In *Proceedings of the 20th ACM SIGIR conference on research and development in information retrieval* (pp. 92–98).

Davis, M. W., & Ogden, W. (2000). Towards universal text retrieval: Tipster text retrieval research at New Mexico State University. *Information Retrieval, 3*, 339–356.

Diekema, A., Oroumchian, F., Sheridan, P., & Liddy, E. D. (1999). TREC-7 evaluation of conceptual interlingua document retrieval (CINDOR) in English and French. In *Proceedings of TREC-7*. Gaithersburg, MD: National Institute of Standards and Technology. Available: http://trec.nist.gov/pubs/.

Dumais, S. T., Landauer, T. K., & Littman, M. L. (1996). Automatic cross-linguistic information retrieval using latent semantic indexing. In *Proceedings of the 19th ACM SIGIR conference on research and development in information retrieval* (pp. 16–23).

Dunning, T. (1993). Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics, 19*(1), 61–74.

Federico, M., & Bertoldi, N. (2002). Statistical cross-language information retrieval using *n*-best query translations. In *Proceedings of the 25th ACM SIGIR conference on research and development in information retrieval* (pp. 167–174).

Fox, C. (1990). A stop list for general text. *ACM-SIGIR Forum, 24*, 19–35.

Frakes, W. B. (1992). Stemming algorithms. In W. B. Frakes & R. Baeza-Yates (Eds.), *Information retrieval: data structures & algorithms* (pp. 131–160). Englewood Cliffs, NJ: PTR–Prentice-Hall.

Franz, M., & Scott McCarley, J. (2003). Arabic information retrieval at IBM. In *Proceedings of TREC-2002*. Gaithersburg, MD: National Institute of Standards and Technology. Available: http://trec.nist.gov/pubs/.

Franz, M., Scott McCarley, J., & Roukos, S. (1999). Ad hoc and multilingual information retrieval at IBM. In *Proceedings of TREC-7*. Gaithersburg, MD: National Institute of Standards and Technology. Available: http://trec.nist.gov/pubs/.

Franz, M., Scott McCarley, J., & Todd Ward, R. (2000). Ad hoc, cross-language and spoken document information retrieval at IBM. In *Proceedings of TREC-8*. Gaithersburg, MD: National Institute of Standards and Technology. Available: http://trec.nist.gov/pubs/.

Fraser, A., Xu, J., & Weischedel, R. (2003). TREC 2002 cross-lingual retrieval at BBN. In *Proceedings of TREC-2002*. Gaithersburg, MD: National Institute of Standards and Technology. Available: http://trec.nist.gov/pubs/.

Fujii, A., & Ishikawa, T. (2001). Japanese/English cross-language information retrieval: exploration of query translation and transliteration. *Computers and the Humanities, 35*, 389–420.

Gao, J., Nie, J. Y., Zhang, J., Xun, E., Su,Y., Zhou, M., & Huang, C. (2001a). TREC-9 CLIR experiments at MSRCN. In *Proceedings of TREC-9*. Gaithersburg, MD: National Institute of Standards and Technology. Available: http://trec.nist.gov/pubs/.

Gao, J., Nie, J. Y., Xun, E. X., Zhang, J., Zhou, M., & Huang, C. (2001b). Improving query translation for cross-language information retrieval using statistical models. In *Proceedings of 24th ACM SIGIR conference on research and development in information retrieval* (pp. 96–104).

Gao, J., Nie, J. Y., He, H., Chen, W., & Zhou, M. (2002). Resolving query translation ambiguity using a decaying co-occurrence model and syntactic dependence relations. In *Proceedings of 25th ACM SIGIR conference on research and development in information retrieval* (pp. 183–190).

Gey, F. C. (2001). Research to improve cross-language retrieval: position paper for CLEF. In C. Peters (Ed.), *Cross-language information retrieval evaluation. LNCS* (2069, pp. 83–88). Berlin: Springer-Verlag.

Gey, F. C., Jiang, H., Chen, A., & Larson, R. R. (1999). Manual queries and machine translation in cross-language retrieval and interactive retrieval with Cheshire II at TREC-7. In *Proceedings of TREC-7*. Gaithersburg, MD: National Institute of Standards and Technology. Available: http://trec.nist.gov/pubs/.

Gey, F., Kando, N., & Peters, C. (2002). Cross-language information retrieval: a research roadmap: In *Summary of a workshop at SIGIR-2002*. Available: http://ucdata.berkeley.edu/sigir-2002/.

Gollins, T., & Sanderson, M. (2001). Improving cross language information retrieval with triangulated translation. In *Proceedings of the 24th ACM SIGIR conference on research and development in information retrieval* (pp. 90–95).

Hedlund, T., Keskustalo, H., Pirkola, A., Sepponen, M., & Järvelin, K. (2001a). Bilingual tests with Swedish, Finish, and German queries: dealing with morphology, compound words, and query structure. In C. Peters (Ed.), *Cross-language information retrieval evaluation. LNCS* (2069, pp. 210–223). Berlin: Springer-Verlag.

Hedlund, T., Pirkola, A., & Järvelin, K. (2001b). Aspects of Swedish morphology and semantics from the perspective of mono- and cross-language information retrieval. *Information & Processing Management, 37*, 147–161.

Hedlund, T., Keskustalo, H., Pirkola, A., Airio, E., & Järvelin, K. (2002). Utaclir@CLEF2001: effects of compound splitting and *n*-gram techniques. In C. Peters, et al. (Eds.), *Evaluation of cross-language information retrieval systems. LNCS* (2406, pp. 118–136). Berlin: Springer-Verlag.

Hiemstra, D. (1998). A linguistically motivated probabilistic model of information retrieval. In C. Nikolaou & C. Stephanidis (Eds.), *Research and advanced technology for digital libraries*. *LNCS* (1513, pp. 569–584). Berlin: Springer-Verlag.

Hiemstra, D., & Kraaij, W. (1999). Twenty-one at TREC-7: ad-hoc and cross-language track. In *Proceedings of TREC-7*. Gaithersburg, MD: National Institute of Standards and Technology. Available: http://trec.nist.gov/pubs/.

Hiemstra, D., Kraaij, W., Pohlmann, R., & Westerveld, T. (2001). Translation resources, merging strategies, and relevance feedback for cross-language information retrieval. In C. Peters (Ed.), *Cross-language information retrieval evaluation*. *LNCS* (2069, pp. 102–115). Berlin: Springer-Verlag.

Hull, D. (1998). A weighted Boolean model for cross-language text retrieval. In G. Grefenstette (Ed.), *Cross-language information retrieval* (pp. 119–136). Boston, MA: Kluwer.

Hull, D. A., & Grefenstette, G. (1996). Querying across languages: a dictionary-based approach to multilingual information retrieval. In *Proceedings of the 19th ACM SIGIR conference on research and development in information retrieval* (pp. 49–57).

Ide, N., & Véronis, J. (1998). Introduction to the special issue on word sense disambiguation: the state of the art. *Computational Linguistics, 24*(1), 1–40.

Jones, G. J. F., & Lam-Adesina, A. M. (2002). Experiments with machine translation for bilingual retrieval. In C. Peters, et al. (Eds.), *Evaluation of cross-language information retrieval systems*. *LNCS* (2406, pp. 59–77). Berlin: Springer-Verlag.

Kim, T., Sim, C. M., Yuh, S., Jung, H., Kim, Y. K., Choi, S. K., Park, D. I., & Choi, K. S. (1999). FromTo-CLIR: web-based natural language interface for cross-language information retrieval. *Information Processing & Management, 35*, 559–586.

Knight, K., & Graehl, J. (1998). Machine transliteration. *Computational Linguistics, 24*(4), 599–612.

Kraaij, W., Pohlmann, R., & Hiemstra, D. (2000). Twenty-one at TREC-8: using language technology for information retrieval. In *Proceedings of TREC-8*. Gaithersburg, MD: National Institute of Standards and Technology. Available: http://trec.nist.gov/pubs/.

Kraaij, W. (2002). TNO at CLEF-2001: comparing translation resources. In C. Peters et al. (Eds.), *Evaluation of cross-language information retrieval systems*. *LNCS* (2406, pp. 78–93). Berlin: Springer-Verlag.

Kwok, K. L. (1996). A network approach to probabilistic information retrieval. *ACM Transaction on Information Systems, 12*, 325–353.

Kwok, K. L., Grunfeld, N., Dinstl, N., & Chan, M. (2001). TREC-9 cross language, web and question–answering track experiments using PIRCS. In *Proceedings of TREC-9*. Gaithersburg, MD: National Institute of Standards and Technology. Available: http://trec.nist.gov/pubs/.

Landauer, T., & Littman, M. L. (1990). A statistical method for language-independent representation of the topical content of text segments. In *Proceedings of the 6th annual conference of the UW centre for the new oxford English dictionary and text research*. Available: http://www.cs.duke.edu/~mlittman/docs/x-lang.ps.

Lavrenko, V., Choquette, M., & Croft, W. B. (2002). Cross-lingual relevance models. In *Proceedings of the 25th ACM SIGIR conference on research and development in information retrieval* (pp. 175–182).

Lin, C. J., Lin, W. C., Bian, G. W., & Chen, H. H. (1999). Description of the NTU Japanese–English cross-lingual information retrieval system used for NTCIR workshop. In *Proceedings of the first NTCIR workshop on research in Japanese text retrieval and term recognition*. Tokyo: National Institute of Informatics. Available: http://research.nii.ac.jp/ntcir/workshop/.

Lin, W. C., & Chen, H. H. (2003). Description of NTU approach to NTCIR3 multilingual information retrieval. In *Proceedings of the third NTCIR workshop on research in information retrieval, automatic text summarization and question answering*. Tokyo: National Institute of Informatics. Available: http://research.nii.ac.jp/ntcir/workshop/.

Littman, M. L., Dumais, S. T., & Landauer, T. K. (1998). Automatic cross-language information retrieval using latent semantic indexing. In G. Grefenstette (Ed.), *Cross-language information retrieval* (pp. 51–62). Boston, MA: Kluwer.

Maeda, A., Sadat, F., Yoshikawa, M., & Uemura, S. (2000). Query term disambiguation for Web cross-language information retrieval using a search engine. In *Proceedings of the 5th international workshop information retrieval with Asian languages* (pp. 25–32).

Manning, C. D., & Shütze, H. (1999). *Foundations of statistical natural language processing*. Cambridge, MA: MIT Press.

McNamee, P., & Mayfield, J. (2002a). Comparing cross-language query expansion techniques by degrading translation resources. In *Proceedings of the 25th ACM SIGIR conference on research and development in information retrieval* (pp. 159–166).

McNamee, P., & Mayfield, J. (2002b). JHU/APL experiments at CLEF: translation resources and score normalization. In C. Peters, et al. (Eds.), *Evaluation of cross-language information retrieval systems*. *LNCS* (2406, pp. 193–208). Berlin: Springer-Verlag.

McNamee, P., Mayfield, J., & Piatko, C. (2001). A language-independent approach to European text retrieval. In C. Peters (Ed.), *Cross-language information retrieval evaluation*. *LNCS* (2069, pp. 129–139). Berlin: Springer-Verlag.

Miller, D. R. H., Leek, T., & Schwartz, R. M. (1999a). BBN at TREC7: using hidden Markov models for information retrieval. In *Proceedings of TREC-7*. Gaithersburg, MD: National Institute of Standards and Technology. Available: http://trec.nist.gov/pubs/.

Miller, D. R. H., Leek, T., & Schwartz, R. M. (1999b). A hidden Markov model information retrieval system. In *Proceedings of the 22nd ACM SIGIR conference on research and development in information retrieval* (pp. 214–221).

Molina-Salgado, H., Moulinier, I., Knudson, M., Lund, E., & Sekhon, K. (2002). Thomson legal and regulatory at CLEF 2001: monolingual and bilingual experiments. In C. Peters, et al. (Eds.), *Evaluation of cross-language information retrieval systems*. *LNCS* (2406, pp. 226–234). Berlin: Springer-Verlag.

Mori, T., Kokubu, T., & Tanaka, T. (2001). Cross-lingual information retrieval based on LSI with multiple word spaces. In *Proceedings of 2nd NTCIR workshop meeting on evaluation of Chinese & Japanese text retrieval and text summarization*. Tokyo: National Institute of Informatics. Available: http://research.nii.ac.jp/ntcir/workshop/.

Moulinier, I., & Molina-Salgado, H. (2002). Thomson legal and regulatory experiments for CLEF2002. In C. Peters (Ed.), *Working notes for the CLEF-2002 Workshop* (pp. 91–96).

Nie, J. Y. (1999). TREC-7 CLIR using a probabilistic translation model. In *Proceedings of TREC-7*. Gaithersburg, MD: National Institute of Standards and Technology. Available: http://trec.nist.gov/pubs/.

Nie, J. Y. (2000). CLIR using a probabilistic translation model based on Web documents. In *Proceedings of TREC-8*. Gaithersburg, MD: National Institute of Standards and Technology. Available: http://trec.nist.gov/pubs/.

Nie, J. Y., & Jin, F. (2003). A multilingual approach to multilingual retrieval. In C. Peters, et al. (Eds.), *Advances in cross-language information retrieval. LNCS* (2785, pp. 101–110). Berlin: Springer-Verlag.

Nie, J. Y., & Simard, M. (2002). Using statistical translation model for bilingual IR. In C. Peters, et al. (Eds.), *Evaluation of cross-language information retrieval systems. LNCS* (2406, pp. 137–150). Berlin: Springer-Verlag.

Nie, J. Y., Simard, M., Isabelle, P., & Durand, R. (1999). Cross-language information retrieval based on parallel texts and automatic mining of parallel texts from the Web. In *Proceedings of the 22nd ACM SIGIR conference on research and development in information retrieval* (pp. 74–81).

Nie, J. Y., Simard, M., & Foster, G. (2001). Multilingual information retrieval based on parallel texts from the Web. In C. Peters (Ed.), *Cross-language information retrieval evaluation. LNCS* (2069, pp. 188–201). Berlin: Springer-Verlag.

Oard, D. W. (2001). Evaluating interactive cross-language information retrieval: document selection. In C. Peters (Ed.), *Cross-language information retrieval evaluation. LNCS* (2069, pp. 57–71). Berlin: Springer-Verlag.

Oard, D. W., & Diekema, A. R. (1998). Cross-language information retrieval. *Annual Review of Information Science and Technology, 33*, 223–256.

Oard, D. W., & Dorr, B. J. (1996). A survey of multilingual text retrieval. College Park, MD: University of Maryland, Institute for Advanced Computer Studies. Available: http://www.cs.umd.edu/TRs/authors/Douglas_W_Oard.html.

Oard, D. W., & Gonzalo, J. (2002). The CLEF 2001 interactive track. In C. Peters, et al. (Eds.), *Evaluation of cross-language information retrieval systems. LNCS* (2406, pp. 308–319). Berlin: Springer-Verlag.

Oard, D. W., & Hackett, P. (1998). Document translation for cross-language text retrieval at the University of Maryland. In *Proceedings of TREC-6*. Gaithersburg, MD: National Institute of Standards and Technology. Available: http://trec.nist.gov/pubs/.

Oard, D. W., Levow, G. A., & Cabezas, C. I. (2001). CLEF experiments at Maryland: statistical stemming and backoff translation. In C. Peters (Ed.), *Cross-language information retrieval evaluation. LNCS* (2069, pp. 176–187). Berlin: Springer-Verlag.

Peters, C., & Sheridan, P. (2001). Multilingual information access. In M. Agosti, F. Crestani, & G. Pasi (Eds.), *Lectures on information retrieval. LCSN* (1980, pp. 51–80). Berlin: Springer-Verlag.

Pirkola, A. (1998). The effects of query structure and dictionary setups in dictionary-based cross-language information retrieval. In *Proceedings of the 21st ACM SIGIR conference on research and development in information retrieval* (pp. 55–63).

Pirkola, A., Hedlund, T., Keskustalo, H., & Järvelin, K. (2001). Dictionary-based cross-language information retrieval: problems, methods, and research findings. *Information Retrieval, 4*, 209–230.

Ponte, J. M., & Croft, W. B. (1998). A language modeling approach to information retrieval. In *Proceedings of the 21st ACM SIGIR conference on research and development in information retrieval* (pp. 275–281).

Powell, A. L., French, J. C., Callan, J., Connell, M., & Viles, C. L. (2000). The impact of database selection on distributed searching. In *Proceeding of the 23rd ACM SIGIR conference on research and development in information retrieval* (pp. 232–239).

Porter, M. F. (1980). An algorithm for suffix stripping. *Program, 14*(3), 130–137.

Porter, M. F. (2001). Snowball: a language for stemming algorithms. Available: http://snowball.tartarus.org/texts/introduction.html. 2004/Mar./10.

Qu, Y., Grefenstette, G., & Evans, D. A. (2003). Resolving translation ambiguity using monolingual corpora. In C. Peters, et al. (Eds.), *Advances in cross-language information retrieval. LNCS* (2785, pp. 223–241). Berlin: Springer-Verlag.

Rehder, B., Littman, M. L., Dumais, S., & Landauer, T. K. (1998). Automatic 3-language cross-language information retrieval with latent semantic indexing. In *Proceedings of the TREC-6*. Gaithersburg, MD: National Institute of Standards and Technology. Available: http://trec.nist.gov/pubs/.

Resnik, P. (1999). Mining the web for bilingual text. In *37th annual meeting of the association for computational linguistics (ACL'99)*, College Park, MD.

Robertson, S. E., & Sparck Jones, K. (1976). On relevance probabilistic indexing and information retrieval. *Journal of the American Society for Information Science, 27*(3), 129–146.

Robertson, S. E., Walker, S., Jones, S., Hancock-Beaulieu, M. M., & Gatford, M. (1995). Okapi at TREC-3. In *Overview of TREC-3*. Gaithersburg, MD: National Institute of Standards and Technology. Available: http://trec.nist.gov/pubs/.

Sadat, F., Maeda, A., Yoshikawa, M., & Uemura, S. (2002). Query expansion techniques for the CLEF Bilingual track. In C. Peters, et al. (Eds.), *Evaluation of cross-language information retrieval systems. LNCS* (2406, pp. 177–184). Berlin: Springer-Verlag.

Sakai, T., Koyama, M., Suzuki, M., & Manabe, T. (2003). Toshiba KIDS at NTCIR-3: Japanese and English–Japanese IR. In *Proceedings of the third NTCIR workshop on research in information retrieval, automatic text summarization and question answering*. Tokyo: National Institute of Informatics. Available: http://research.nii.ac.jp/ntcir/workshop/.

Salton, G. (1970). Automatic processing of foreign language documents. *Journal of the American Society for Information Science, 21*(3), 187–194.

Savoy, J. (1999). A stemming procedure and stopword list for general French corpora. *Journal of the American Society for Information Science, 5*(10), 944–952.

Savoy, J. (2002). Report on CLEF-2001 experiments: effective combined query-translation approach. In C. Peters et al. (Eds.), *Evaluation of cross-language information retrieval systems. LNCS* (2406, pp. 27–43). Berlin: Springer-Verlag.

Savoy, J. (2003a). Report on CLEF 2002 experiments: combining multiple sources of evidence. In C. Peters et al. (Eds.), *Advances in cross-language information retrieval. LNCS* (2785, pp. 66–90). Berlin: Springer-Verlag.

Savoy, J. (2003b). Cross-language information retrieval: experiments based on CLEF 2000 corpora. *Information Processing & Management, 39*, 75–115.

Seo, H. C., Kim, S. B., Kim, B. I., Rim, H. C., & Lee, S. Z. (2003). KUNLP system for NTCIR-3 English–Korean cross-language information retrieval. In *Proceedings of the third NTCIR workshop on research in information retrieval, automatic text summarization and question answering*. Tokyo: National Institute of Informatics. Available: http://research.nii.ac.jp/ntcir/workshop/.

Sheridan, P., & Ballerini, J. P. (1996). Experiments in multilingual information retrieval using the SPIDER system. In *Proceedings of the 19th ACM SIGIR conference on research and development in information retrieval* (pp. 58–64).

Sperer, R., & Oard, D. W. (2000). Structured translation for cross-language information retrieval. In *Proceedings of the 23rd ACM SIGIR conference on research and development in information retrieval* (pp. 120–127).

Stalls, B., & Knight, K. (1998). Translating names and technical terms in Arabic text. In *Proceedings of the COLING/ACL workshop on computational approaches to semantic languages*. Available: http://www.isi.edu/natural-language/people/knight.html.

Turtle, H. R., & Croft, W. B. (1991). Evaluation of an inference network-based retrieval model. *ACM Transactions on Information Systems, 9*(3), 187–222.

Xu, J., Weischedel, R., & Nguyen, C. (2001). Evaluating a probabilistic model for cross-lingual information retrieval. In *Proceedings of the 24th ACM SIGIR conference on research and development in information retrieval* (pp. 105–110).

Yamabana, K., Muraki, K., Doi, S., & Kamei, S. (1998). A language conversion front-end for cross-language information retrieval. In G. Grefenstette (Ed.), *Cross-language information retrieval* (pp. 93–104). Boston, MA: Kluwer.

Yang, Y., Carbonell, J., Brown, R. D., & Frederking, R. E. (1998). Translingual information retrieval: learning from bilingual corpora. *Artificial Intelligence, 103*, 323–345.