Translingual Information Retrieval: A Comparative Evaluation

Carbonell, Jaime G.; Yang, Yiming; Frederking, Robert E.; Brown, Ralf D.; Geng, Yibing; Lee, Danny

Language Technologies Institute, Carnegie Mellon University
5000 Forbes Avenue, Pittsburgh, PA 15213 USA

Abstract

Translingual information retrieval (TIR) consists of providing a query in one language and searching document collections in one or more different languages. This paper introduces new TIR methods and reports on comparative TIR experiments with these new methods and with previously reported ones in a realistic setting. Methods fall into two categories: query translation based, and statistical-IR approaches establishing translingual associations. The results show that using bilingual corpora for automated extraction of term equivalences in context outperforms other methods. Translingual versions of the Generalized Vector Space Model (GVSM) and Latent Semantic Indexing (LSI) perform relatively well, as does translingual pseudo relevance feedback (PRF). All showed relatively small performance loss between monolingual and translingual versions. Query translation based on a general machine-readable bilingual dictionary – heretofore the most popular method – did not match the performance of other, more sophisticated methods. Also, the previous very high LSI results in the literature were disconfirmed by more realistic relevance-based evaluations.

1 Introduction

Translingual information retrieval (TIR) is starting to receive considerable attention in recent years with the increased accessibility of ever-more-diverse on-line international text collections, including centrally the World Wide Web. In spite of some pioneering work [Salton, 1970; Dumais *et al.*, 1996; Davis and Dunning, 1996; Hull and Grefenstette, 1996], evaluations of different TIR techniques in realistic retrieval tasks are rare. This paper reports our evaluation results of both newly developed TIR techniques and reimplementations of previously reported techniques.

Translingual information retrieval (aka "multilingual" or "crosslingual" IR) consists of providing a query in one language and searching document collections in one or more different languages. One can envision many ways to bridge the language barrier between query and collection. In this paper, we focus on query translation and methods based on automatically establishing translingual associations between queries and documents without need of translating either.

2 MT-Based Methods for TIR

The *machine translation methods* for TIR require that either the query be translated into the target language, and the translation be used to search the target-language collection, or the collection be translated into the source language, and the original query be used to search. Let us consider the pros and cons of each approach:

- *Translation Accuracy* – Both human and machine translation [Carbonell, 1985; Nirenburg *et al.*, 1991] require context to achieve accuracy. Translating isolated words in a query is unreliable, due to unresolved lexical ambiguity. Translating documents should yield greater accuracy.
- *Retrieval Accuracy* – Since documents contain far more information than queries, random translation errors should cause less degradation for the IR task in documents than in queries. Hence for both this reason and the above, document translation is in principle preferable. In fact, preliminary findings by Dumais *et al* [Dumais *et al.*, 1996] support this line of reasoning.
- *Practicality* – Many document collections are very large. Most are searched remotely. Some are proprietary; individual documents may be read or downloaded, but the entire collection may not be copied or translated. Even if these problems were surmountable, translating the collection may require inordinately long computation and massive storage, not to mention re-indexing the translated collection.

Because of the above practicality constraint, we report only on translating the query for TIR. If the query were formulated as phrases, as a full sentence, or as a paragraph, we could apply MT systems far more reliably. However, experience shows that users typically prefer to give isolated words, or at best short phrases to an IR system.[1] The question is how to best translate a set of isolated words. Full fledged MT is not applicable. Instead, we investigated three approaches:

1. *Dictionary Term Translation* – Look up each query term in a general-purpose bilingual dictionary, and use all its possible translations. This is a form of query expansion upon translation.
2. *Corpus-based Term Translation* – Use a sentence-aligned bilingual training corpus to find the terms that co-occur in context across languages, thus creating a corpus-based term-equivalence matrix.
3. *Corpus-based Term-to-Sentence* – Use the same type of aligned bilingual training corpus to extract full sentences that in the target language co-occur with query terms in the source language. Term-to-sentence expansion may enhance recall, but at a cost in precision.

All three MT-based methods used variations of the Pangloss Example-Based Machine Translation engine (PanEBMT) [Brown, 1996], however only corpus-based term translation (called EBT below, for "Example-Based Term" translation) is further described since it produced better results.

2.1 PanEBMT Translations

In general, EBMT systems [Brown, 1996; Nagao, 1984] use a large corpus of example pairs of previously translated sentences, in order to find close matches and translations of words and phrases in context. The PanEBMT parallel corpus was derived primarily from the Spanish and English portions of the UN Multilingual Corpus [Graff and Finch, 1994], with an admixture of texts from the Pan-American Health Organization and ARPA MT evaluations. The total corpus contains some 685,000 sentence pairs after removing duplicated Spanish sentences. PanEBMT translates by finding the set of matches to a new text string (word, phrase or sentence) in the indexed bilingual corpus. Then the translations corresponding to these matches are combined into candidate translations of the new text. Because queries contain more isolated terms than phrases or sentences, our query-translation experiment is unable to exploit the power of EBMT.

2.2 Corpus-based Term Translation

In order to create domain-specific or corpus-specific bilingual dictionaries automatically, we start from a large sentence-aligned bilingual corpus and generate a large term co-occurrence table combined with a thresholding scheme[Brown, 1997]. The result was used as the dictionary for corpus-based (example-based) term substitution MT (EBT).

Co-occurrence dictionary generation is performed in two phases: First the co-occurrence matrix (indexed by source-language words on one axis and target-language words on the other) is generated. Each cell in the matrix represents the number of times the source-language word occurred in the same sentence pair as the target-language word. Given this matrix, we compute the conditional probability that if the term occurs in one language its counterpart (i.e. its candidate translation) also occurs in the other language. If this probability is above a threshold, then the term translation is added into the dictionary. Should a term in one language co-occur with several terms in the other language with sufficient frequency to pass the conditional probability threshold, *all* are stored as candidate translations.

3 IR-based Methods for TIR

We extended three monolingual retrieval methods to translingual retrieval: pseudo-relevance feedback (PRF)[Buckley *et al.*, 1995], the general vector space model (GVSM)[Wong *et al.*, 1985], and the latent semantic indexing (LSI) approach[Deerwester *et al.*, 1990]. In each case, a translingual semantic correspondence between queries and documents is established based on a document-aligned bilingual training corpus, without requiring sentence-level alignment, bilingual dictionaries or machine translation.

3.1 Pseudo-Relevance Feedback

Relevance feedback (RF) is an approach to query expansion in monolingual text retrieval[Salton, 1989]. It requires a user to judge interactively which retrieved documents are relevant, and uses the relevance judgments to expand the original query for additional search. By "pseudo-relevance feedback" (PRF) we mean using the top-ranking documents obtained in an initial retrieval without human judgements, assuming that a significant fraction of top-ranked documents will be relevant. Both RF and PRF are query expansion techniques similar to case-based IR [Rissland and Daniels, 1995], and both typically improve performance in monolingual retrieval compared to not using them. The adaptation of PRF to translingual retrieval is relatively simple if a bilingual corpus is available. That is, we find the top-ranking documents for a query in the source language, substitute the corresponding documents in the target language, and use these documents to form the corresponding query in the target language. Figure 1 illustrates the data flow for translingual RF and PRF.

All the IR techniques discussed here, including the PRF approach are variants of the vector space model (VSM)[Salton, 1989] where both queries and documents are represented using vectors of term weights. To allow clear theoretical comparison of these IR-based methods, let us define the notation for VSM (including PRF):

$$\vec{q} = (q_1, q_2, \ldots, q_m)^t$$
$$\vec{d} = (d_1, d_2, \ldots, d_m)^t$$
$$sim(\vec{q}, \vec{d}) = \cos(\vec{q}, \vec{d}) = \frac{\sum_{i=1}^m q_i d_i}{\sqrt{\sum_{i=1}^m q_i^2}\sqrt{\sum_{i=1}^m d_i^2}}$$

[1] LYCOS reports that their typical user queries for general web search are only one to three words long, although they are occasionally reformulated into longer queries.

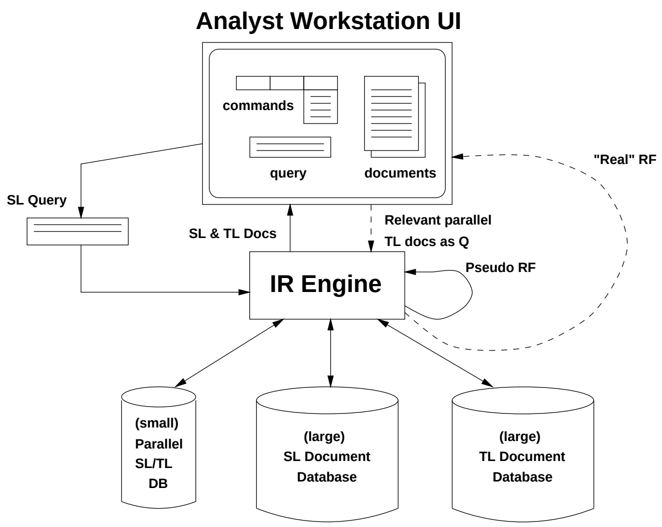

Figure 1: Data flow for translingual relevance feedback and pseudo-relevance feedback

where $\vec{q}$ is the query, $\vec{d}$ is a document in a corpus, m is the number of unique terms (words or phrases) in the corpus, and q_i and d_i are the term weights in the query and the document, respectively.

3.2 Generalized Vector Space Model

A criticism of conventional VSM is that it uses terms as an orthogonal basis of the vector space, but terms are often not semantically independent. Wong *et al* proposed an alternative, namely the "generalized vector space model" (GVSM) [Wong *et al.*, 1985], also referred to as "the dual space" [Sheridan and Ballerini, 1996]. The idea is to use documents as the basis for representing terms instead of using terms.

Consider a term-document matrix, $A_{m\times n}$, as a training corpus where m is the size of the vocabulary, and n is the number of unique documents in this corpus. One can view this matrix as a way to represent documents (the columns) using terms, and to represent terms (the rows) using documents. The former view corresponds to the conventional vector space model, and the latter view corresponds to the dual space. Each row vector of A reflects the term usage in the corpus, i.e., the *pattern* of this term distributed over documents.

Matrix A can be used for query transformation by computing $\vec{q'} = A^t\vec{q}$ where $\vec{q}$ is the original query vector whose dimensions are unique terms, and $\vec{q'}$ is the transformed vector whose dimensions are unique documents. The transformation is equivalent to weighting the *distribution pattern* of each term using its weight in the original query, and summing up the weighted patterns to obtain a new representation of the query. Similar to query transformation, a document can also be transformed into a vector in the dual space by computing $\vec{d'} = A^t\vec{d}$ where $\vec{d}$ is the document vector in the conventional VSM. The retrieval criterion in GVSM for monolingual retrieval is defined to be:

$$sim(\vec{q},\vec{d}) = \cos(A^t\vec{q}, A^t\vec{d}).$$

Here we propose a novel extension of the monolingual GVSM for translingual retrieval. Assuming a bilingual corpus for training, we form two matrices, A and B, where A is a term-document matrix for the training documents in the source language (also the language of the queries), B is a term-document matrix for the training documents in the target language, and the corresponding columns of A and B are the matching pairs of documents in the bilingual corpus. We use A for query transformation and B target-language document transformation. The retrieval criterion is defined to be:

$$sim(\vec{q},\vec{d}) = \cos(A^t\vec{q}, B^t\vec{d})$$

Since matrix A and B share the same dual space, the transformations $A^t\vec{q}$ and $B^t\vec{d}$ give the query and the document a common basis (presenting distribution patterns of terms over documents) on which they can be compared. This is how the translingual correspondence is established.

The computation in GVSM consists of the transformation ($A^t\vec{q}$ and $B^t\vec{d}$) and the cosine computation. The time complexity of the first part is similar to the computation in VSM. It is proportional to the number of non-zero elements in a query or document vector, $O(kn)$ where k is the average number of unique terms per query or document, and n is the number of document pairs in the bilingual training corpus. The time complexity in the second part, is $O(n)$ per document, or $O(nl)$ for a test corpus of l documents. It is possible to significantly reduce this complexity in large problems by aggressively removing non-influential elements from the transformed document vectors [Yang, 1995].

3.3 Latent Semantic Indexing

Latent Semantic Indexing[Deerwester *et al.*, 1990] (LSI) is a one-step extension of GVSM. The claim is that neither terms nor documents are the optimal choice for the orthogonal basis of a semantic space, and that a reduced vector space consisting of the most meaningful linear combinations of documents would be a better representative basis for the content of documents.

In monolingual retrieval, LSI uses the term-document matrix (A) for training, the same as in GVSM. It computes the orthogonal dimensions ("the latent semantic structures") in matrix A, and selects the principal dimensions as the new basis for a reduce vector space. The monolingual LSI retrieval criterion is defined to be:

$$A = U\Sigma V^t$$
$$L = U\Sigma^{-1}$$
$$sim(\vec{q},\vec{d}) = \cos(L^t\vec{q}, L^t\vec{d})$$

where matrices U and V contain a set of p orthogonal singular vectors each (one for the representation of terms, and another for the representation of documents). Matrix Σ is p-diagonal, containing the singular values indicating the importance of the corresponding singular vectors in matrices U and V. Matrix L contains the weighted singular vectors in U, where the weights are the corresponding singular values in Σ. This matrix can be viewed as a reduced version of matrix A. That is, both A and L use their row vectors to represent terms, but the term vectors in L are much shorter than the term vectors in A. The dimensions in L are linear combinations of documents, while the dimensions in A are individual documents.

The translingual LSI model [Dumais *et al.*, 1996] is similar to the model for monolingual LSI, except that a bilingual document corpus is needed for training instead of a monolingual corpus. Let $\vec{q}$ be a query in the source language, $\vec{d}$ be a document in the target language, and $\begin{bmatrix} A \\ B \end{bmatrix}$ be the matrix of bilingual document pairs where A and B are the same as defined in GVSM. Then the translingual LSI retrieval criterion is defined to be:

$$\begin{bmatrix} A \\ B \end{bmatrix} = U_2 \Sigma_2 V_2^t$$
$$L_2 = U_2 \Sigma_2^{-1}$$
$$sim(\vec{q}, \vec{d}) = \cos(L_2^t \vec{q}, L_2^t \vec{d})$$

where U_2, V_2, Σ_2 and L_2 are the matrices computed using the singular value decomposition of the bilingual input matrix .

LSI has a quadratic time complexity of $O(n'p)$ where $n' = \max\{m, n\}$ is the larger number between the size (m) of the joint vocabulary of both languages and the number (n) of document pairs in the bilingual training corpus; p is the number of orthogonal dimensions (singular vectors) computed in the singular value decomposition. Thus the scalability of this method to a large corpus would be much more limited than the VSM or GVSM approach if a large number of singular vectors is necessary for good retrieval performance.

3.4 The Scientific Challenge

The similarities and differences between the three models mentioned above can be seen in their retrieval criteria:

$$\begin{aligned} VSM: &\quad sim(\vec{q}, \vec{d}) = \cos(\vec{q}, \vec{d}) \\ GVSM: &\quad sim(\vec{q}, \vec{d}) = \cos(A^t \vec{q}, B^t \vec{d}) \\ LSI: &\quad sim(\vec{q}, \vec{d}) = \cos(L_2^t \vec{q}, L_2^t \vec{d}) \end{aligned}$$

The fundamental difference, in theory, is the choice of the basis for the similarity comparison between queries and documents. VSM assumes semantic independence of terms in its basis. GVSM uses documents instead, assuming documents are semantically independent. LSI computes the orthogonal dimensions in a training corpus, and chooses the principal dimensions as the basis of a reduced vector space. GVSM and LSI are close variants in the sense that both exploit the dual space. The only difference is whether to use the original dimensions (document vectors) or the reduced dimensions (the orthogonal singular vectors) as the basis for the vector space. Which model best represents the semantic space of documents and queries is a scientifically challenging question.

Given the methods, empirical validation is important. For monolingual retrieval, performance improvement of GVSM over VSM was observed on small collections [Wong *et al.*, 1985]; improvement of LSI over VSM was observed sometimes but not always [Deerwester *et al.*, 1990]. Until our work reported below, there has not been a comparison between GVSM and LSI, in either monolingual or translingual retrieval.

4 Empirical Evaluation

We carried out a comparative evaluation of the six translingual IR methods described above (the three term-based MT methods, PRF, GVSM, and LSI) on a realistic retrieval task. The large UN Multilingual Corpus[Graff and Finch, 1994] from the Linguistic Data Consortium was available to us, but, among other problems, there were no queries or human relevance judgements available for training and evaluation. We conducted our experiments on a subset of this corpus, consisting of 2255 document pairs pertaining to UNICEF reports and deliberations. Each document pair consists of an English document and its corresponding Spanish translation. 1134 document pairs were randomly selected and used for translingual training. The remaining 1121 pairs were set aside for testing. The average (monolingual) document is 9 paragraphs long. Altogether, the training and test sets in both languages consist of almost 2 million words of text.

We conducted our experiments as follows: First, we created 30 queries in English, germane to the UNICEF subcollection. The average query length was 11 words. Second, we obtained human relevance judgements on the cross product of the 30 queries and 1121 test documents (33,630 samples in all), and used these as our gold standard for testing[2]. Third, we trained each method to find translingual equivalences using paired documents, without queries; hence no relevance judgements were required for training. Fourth, we tested each method *monolingually* on the test set to obtain ranked lists of retrieved documents. Fifth, we applied the translingual version of our methods. Finally, we evaluated the results by comparing the retrieval degradation when moving from monolingual to translingual IR for all the methods.

We optimized each method for monolingual retrieval, with respect to its performance on 11-point average precision using the full human relevance judgements on the 30 queries for the test corpus. Optimizations include the settings on TF and IDF weights for cosine-similarity scoring, setting thresholds on pseudo-relevance feedback, setting cutoff levels for number of non-influential elements for GVSM (200 was optimal), and determining the optimal number of singular vectors in LSI (also 200, coincidentally). When each retrieval method was performing at optimum on monolingual retrieval, we tested

[2]An initial set of experiments was conducted before these judgements were available, as described below.

that method with exactly the same parameter settings on translingual retrieval.

We carried out two sets of experiments. The first comparison focuses on overlap between monolingually-retrieved documents and translingual retrievals, based on the parallel corpus to establish correspondences. The central tenet of the first evaluation was that perfect translingual retrieval would retrieve exactly the corresponding set of documents as monolingual retrieval, but in the target language. Hence, monolingual versus translingual comparison was our primary effectiveness measure. The second comparison is a direct evaluation of monolingual and of translingual retrieval using human relevance judgements and computing precision and recall scores in the traditional IR manner [Salton and McGill, 1983].

4.1 Initial Evaluation

Before we had relevance judgements available, we carried out an initial set of experiments, using two novel evaluation methods. Each of the retrieval methods was trained on the same bilingual training data. We then ran each of the translingual methods using English queries to retrieve Spanish documents, and compared the Spanish documents retrieved translingually to the English documents retrieved monolingually by each method.

The results were compared using both an 11-point average precision and an overlap measure. "Precision" in this case was measured by comparing the top $L2$ of the translingually-retrieved ranked list against the top $L1$ of the corresponding monolingually-retrieved ranked list, and taking the monolingually-retrieved documents as relevant to the query. The overlap measure was simply the percentage of identical documents present in the top $L1$ ranked documents retrieved monolingually and the top $L2$ translingually, for a given $L1$ and $L2$. We calculated both measures in order to minimize any artifact of our evaluation method.

The translingual results are presented in Figures 2 and 3. Figure 2 presents the 11-point average precision for each method plotted against the number of reference documents used ($L1$). $L2$ is four times $L1$ in each case, in an attempt to make the second ranked list sufficiently long. Figure 3 presents the overlap measure for each method, again versus $L1$. Here we use $L2 = L1$, as per our definition of overlap. As mentioned above, we only report the results of the corpus-based EBT method in these graphs, for brevity, given that it performed best among the MT methods.

From these figures, we see that GVSM outperformed the other methods in terms of these document overlap measures. All the methods have a large overlap between their monolingual and translingual retrievals, which is a good sign. Moreover, the graphs of the two evaluation methods are remarkably consistent.

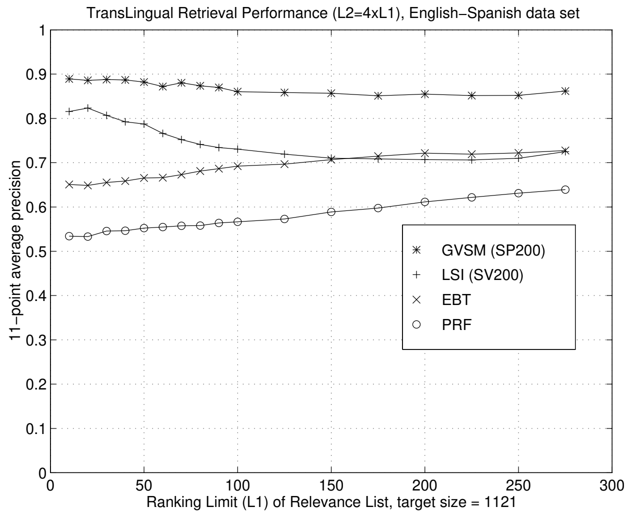

Figure 2: 11-pt. ave. precision vs. rank limit threshold

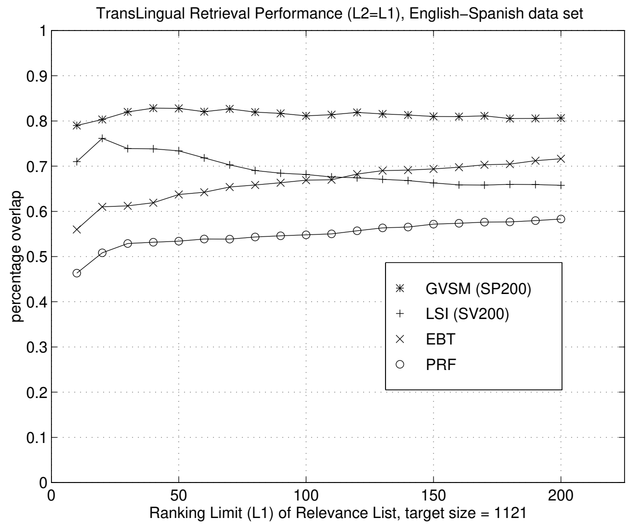

Figure 3: Overlap % vs. rank limit threshold

4.2 Evaluation with Human Judgements

For our second comparison, we evaluated each method, monolingually and translingually, using human relevance judgements. The corresponding 11-point average precision values in the table in figure 4 below. For comparison, we also include corresponding translingual results reported by other researchers. Because the methods have been run on different corpora with different queries, direct comparisons on absolute 11-point-precision recall figures are not meaningful. However, the ratio of translingual IR (TIR) over monolingual IR results may be more indicative of the relative power of the TIR methods. We encourage direct comparisons on the same corpus. We also present our results in the standard recall-precision graphs for monolingual and translingual IR in figures 5 and 6, respectively.

Site	Method	(Monolingual) IR	TIR	TIR/MIR
CMU	Dict Q-tran	.4721 (SMART.basic)	.2898	61%
CMU	EBT Q-tran	.4721 (SMART.basic)	.4318	91%
CMU	PRF Q-exp	.4487	.4043	90%
CMU	GVSM QD-tran (SP=200)	.4008	.3804	95%
CMU	LSI QD-tran (200 SVs)	.3689	.3093	84%
UMASS	CorpusPhrase Q-exp	.20	.1358	68%
ETH	LSI thesaurus Q-exp	.527	.212-.278	40-53%
	(LSI thesaurus + RF)	?	?	(68%)
XEROX	Dict Q-tran	.393	.235	60%
NMSU	Dict Q-tran	?	?	40-50%?

Figure 4: Results of TIR methods (with ntc.ntc weighting)

As the table in figure 4 shows, example-based term (EBT) translation (a.k.a. corpus-based term translations), never before tried for TIR, exhibits top absolute performance, whereas general-purpose machine-readable dictionary (MRD) query-translation exhibits the worst performance. In spite of the similarity between these methods (both translate the query), the former is trained to the corpus and exploits context, and is therefore much superior. This result indicates that the most popular TIR method reported in the literature (MRD-based query translation) may be the simplest, but its performance leaves much to be desired.

Pseudo-relevance feedback also performed well in absolute terms, indicating that if the user were willing to provide true relevance judgements, full relevance feedback could become the top-performing method for TIR.

GVSM, never before tried for TIR, performed relatively well and showed the least degradation from MIR to TIR. LSI did not perform according to expectations from the literature. In earlier less realistic experiments (with queries formulated directly from documents), LSI had performed better [Dumais *et al.*, 1996].

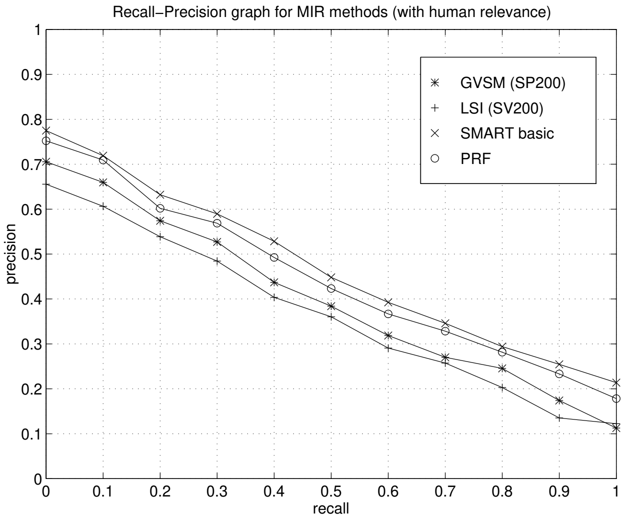

Figure 5: Recall-Precision performance of MIR methods

5 Conclusions

This paper reports a thorough evaluation of multiple methods for translingual retrieval in a query-based retrieval task. We believe that the evaluation methodology used here may be generally useful when costly human relevance judgements are unavailable. Our experimental results indicate that:

- Translingual retrieval is viable by a number of different techniques, ranging from term-based query translation and pseudo-relevance feedback to generalized vector spaces and latent semantic indexing.
- In the translingual retrieval test, example-based MT establishing corpus-based term equivalences performed best, followed by PRF, GVSM, LSI and MRD-based query translation. However, in terms of performance relative to monolingual retrieval, GVSM performed best.
- MRD-based query translation, though popular in the literature, should be re-examined as the TIR method of choice given the results in this paper.
- It appears that Translingual LSI is not as good in a realistic setting with actual queries and 11-point-average precision evaluations as in the preliminary Dumais *et al* study, although LSI does perform better that simple MRD-based query translation.

It is worth noting that GVSM is simple to compute and easy to scale up, somewhat better than LSI, and its performance is not crucially dependent on the exact value of a tuned parameter (such as the number of singular vectors of LSI). More work is clearly called for in further evaluating the GVSM method and corpus-based term-translation in other realistic contexts, and investigating whether other forms of tunable MT-based translingual IR could be made to perform reasonably well, especially in situations where translating the collection does not pose serious problems.

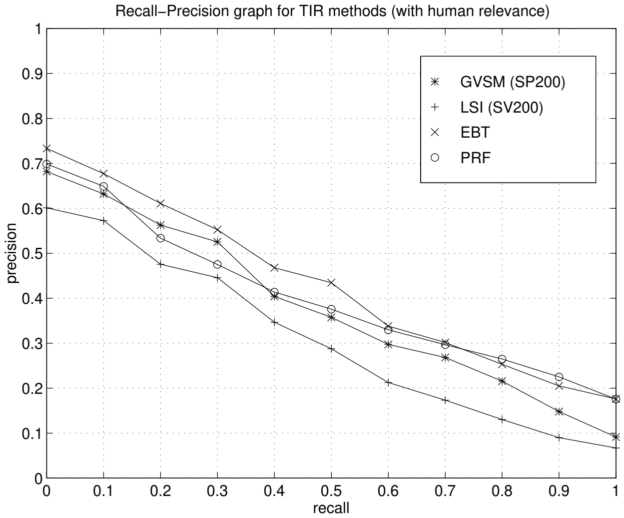

Figure 6: Recall-Precision performance of TIR methods

References

[Brown, 1996] R.D. Brown. Example-Based Machine Translation in the Pangloss System. In *Proceedings of the Sixteenth International Conference on Computation Linguistics*, pages 169–174, 1996.

[Brown, 1997] R.D. Brown. Automated Dictionary Extraction for “Knowledge-Free” Example-Based Translation. In *Proceedings of the Seventh International Conference on Theoretical and Methodological Issues in Machine Translation*, 1997.

[Buckley *et al.*, 1995] C. Buckley, G. Salton, J. Allan, and A. Singhal. Automatic Query Expansion Using SAMRT: TREC 3. In *Overview of the Third Text REtrieval Conference (TREC-3)*, pages 69–80, 1995.

[Carbonell, 1985] J. G. Carbonell. New Approaches to Machine Translation. In *Proceedings of the conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*, Hamilton, NY, 1985.

[Davis and Dunning, 1996] M. Davis and T. Dunning. A TREC evaluation of query translation methods for multi-lingual text retrieval. In *The 4th Text Retrieval Conference (TREC-4)*, 1996.

[Deerwester *et al.*, 1990] S. Deerwester, S.T. Dumais, G.W. Furnas, T.K. Landauer, and R. Harshman. Indexing by Latent Semantic Analysis. In *J Amer Soc Inf Sci 1, 6*, pages 391–407, 1990.

[Dumais *et al.*, 1996] S. Dumais, T. Landauer, and M. Littman. Automatic Cross-Linguistic Information Retrieval using Latent Semantic Indexing. In *Proceedings of SIGIR-96*, Zurich, August 1996.

[Graff and Finch, 1994] David Graff and Rebecca Finch. Multilingual Text Resources at the Linguistic Data Consortium. In *Proceedings of the 1994 ARPA Human Language Technology Workshop*. Morgan Kaufmann, 1994.

[Hull and Grefenstette, 1996] D.A. Hull and G. Grefenstette. Querying Across Languages: a Dictionary-based Approach to Multilingual Information Retrieval. In *19th Ann Int ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'96)*, pages 49–57, 1996.

[Hutchins, 1986] W. J. Hutchins. *Machine Translation: Past, Present, Future.* Ellis Horwood, Chichdster, UK, 1986.

[Nagao, 1984] M. Nagao. A Framework of a Mechanical Translation between Japanese and English by Analogy Principle. In A. Elithorn and R. Banerji (eds), editors, *Artificial and Human Intelligence*. NATO Publications, 1984.

[Nirenburg *et al.*, 1991] S. Nirenburg, J. G. Carbonell, M. Tomita, and K. Goodman. *Knowledge-Based Machine Translation.* Morgan Kaufmann Inc, San Mateo, CA, 1991.

[Rissland and Daniels, 1995] Edwina L. Rissland and Jody J. Daniels. Using CBR to Drive IR. In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence (IJCAI-95)*, pages 400–407, 1995.

[Salton and McGill, 1983] G. Salton and M. J. McGill. *Introduction to Modern Information Retrieval.* McGraw-Hill Computer Science Series. McGraw-Hill, New York, 1983.

[Salton, 1970] G. Salton. Automatic Processing of Foreign Language Documents. *Journal of American Society for Information Sciences*, 21:187–194, 1970.

[Salton, 1989] G. Salton. *Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer.* Addison-Wesley, Reading, Pennsylvania, 1989.

[Sheridan and Ballerini, 1996] P. Sheridan and J.P. Ballerini. Experiments in Multilingual Information Retrieval using the SPIDER System. In *19th Ann Int ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'96)*, pages 58–65, 1996.

[Wong *et al.*, 1985] S.K.M. Wong, W. Ziarko, and P.C.N. Wong. Generalized Vector Space Model In Information Retrieval. In *ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'85).*, pages 18–25, 1985.

[Yang, 1995] Y. Yang. Noise Reduction in a Statistical Approach to Text Categorization. In *Proceedings of the 18th Ann Int ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'95)*, pages 256–263, 1995.