

# Improving Cross-Language Text Retrieval with Human Interactions

William C. Ogden and Mark W. Davis  
*Computing Research Lab*  
*New Mexico State University*  
*Las Cruces NM, 88003*  
*ogden@crl.nmsu.edu*

## Abstract

*Can we expect people to be able to get information from texts in languages they cannot read? In this paper we review two relevant lines of research bearing on this question and will show how our results are being used in the design of a new Web interface for cross-language text retrieval. One line of research, "Interactive IR", is concerned with the user interface issues for information retrieval systems such as how best to display the results of a text search. We review our current research, on "document thumbnail" visualizations, and discuss current Web conventions, practices and folklore. The other area of research, "Cross-Language Text Retrieval", is concerned with the design of automatic techniques, including Machine Translation, to retrieve texts in languages other than the language of the query. We review work we have done concerning query translation and multilingual text summarization. We then describe how these results are being applied and extended in the design a new demonstration interface, Keizai, an end-to-end, Web-based, cross-language text retrieval system. Beginning with an English query, the system will search Japanese and Korean Web data and display English summaries of the top ranking documents. A user should be able to accurately judge which foreign language documents are relevant to their information need and glean necessary information from the translation to schedule specific documents for human translation and subsequent analysis.*

## 1. Introduction

A Cross-Language Text Retrieval (CLTR) system is designed to retrieve text documents in a language different from the language used to specify the information needed. Several compelling research questions have emerged from our work and that of others on CLTR systems. Among these has been the basic question of whether

machine translation technology in the broadest sense is required to translate queries and documents, or whether alternative techniques can substitute for machine translation in many circumstances. The latter question becomes particularly important for languages for which there is little support from commercial machine translation vendors, especially languages that are spoken by small numbers of people. This question can be broken down further by asking whether simple bilingual dictionaries can be useful for query and document translation, and if the limitations of all dictionaries in terms of lexical coverage can be supplemented by using inductive methods to infer term equivalents over parallel texts, from the Web or other sources. We provide an overview of our work in cross-language technology and show how research has begun to answer some of these questions in recent years.

Another important aspect of our research has been focused on trying to understand how to create interfaces and systems that are useful to people. This has included iterative development of systems like our current prototype, Keizai, as well as empirical user studies evaluating document visualization interfaces for monolingual and cross-language text retrieval systems.

When conducting empirical user studies, we have found that careful consideration of how users process the task sheds enormous light on developing effective user interfaces. There are at least a couple of ways a CLTR system may be used. A bilingual user who has good reading skills in their second language may have poorer language productive skills and thus cannot express their information need in their second language as well as they can in their first language. Thus, the CLTR system allows them to find documents in their second language using their first language. A second type of user is a person who is monolingual but has an interest in finding information in documents that are written in foreign languages. They may have access to translation resources but want to limit their use to control costs. Thus, they want to be able to evaluate the relevance of a

document to their query before committing resources for a full translation.

Two elements of the CLTR system interface could benefit the two types of users identified above. The first element, the cross-language query formulation interface, should help the user construct good retrieval requests in the target foreign language. The second element, the retrieval results display, should help users judge the relevance of the retrieved documents set.

## 2. Research in Cross-language Text Retrieval

Although it is possible to translate all of the documents into the query language, for large collections the most economical approach to CLTR is to simply translate the query at retrieval time into the document languages. This presupposes that the query can be translated in a reasonably accurate fashion and that monolingual retrieval systems are available for all of the document languages.

As with Machine Translation (MT) in general, query translation in a CLTR system can be done many different ways. An advanced MT system might, for example, perform sophisticated parsing and analysis of the query, derive an inter-lingual semantic representation and generate a new query from it. At the opposite end of the spectrum, a system could use shallow translation techniques to simply substitute terms from a transfer dictionary, ignoring the ambiguities of polysemous candidates. In shallow translation approaches, the monolingual text retrieval engine operating on the translated query bears the burden of weighting the query terms by virtue of their co-occurrences, hopefully reducing the effect of poorly translated terms. Between these two extremes are a range of approaches.

To overcome the limitations of general-purpose transfer dictionaries, Salton [1] used tuned lexicons and thesauri built from controlled vocabulary to good success in specific text retrieval problems. Despite the growing availability of machine-readable dictionaries, however, preparing special-purpose lexical resources remains a daunting task.

Using massive bilingual and multi-lingual corpora as translation resources is another approach that has the potential to overcome the limitations of the shallow methods, while still requiring less resources than the tuned lexical methods or the deep semantic MT approaches. Text corpora contain examples of usage patterns in the query language that can be matched to examples in the target language if the sentences or paragraphs of the texts are aligned to one another. Although text corpora offer an intriguing possibility for CLTR query translation, the lack of domain-specific texts

or a suitably large range of texts means that general purpose query translation systems remain elusive.

For example, one method of using bilingual corpora, automatically constructs a multilingual semantic space using Lexical Semantic Indexing (LSI). This approach does very well within domains similar to the domain of the bilingual corpora used to train the system but do considerable worse when searching outside of these domains [2].

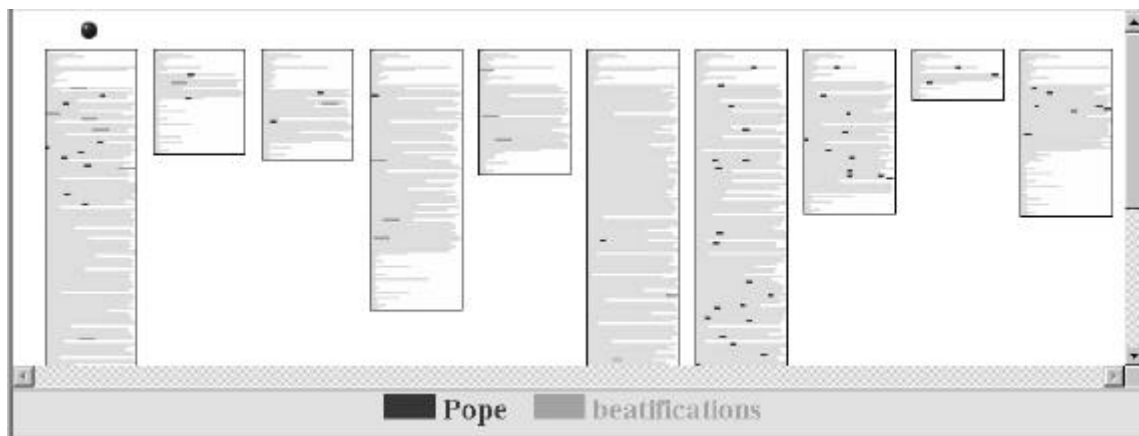
Several other researches have explored methods to improve the CLTR performance when using general bilingual lexicons with automatic disambiguation techniques. For example, Ballesteros and Croft [3] improved CLTR performance using query expansion through automatic relevance feedback. Their best performance, however, was still 32% less than a monolingual baseline.

In our own work, [4][5][6] we have used combinations of shallow dictionary-based and corpus-based method. We use the corpora to derive query translations directly by electing terminology from the target language portion of the corpus. We have also used bilingual corpora to try to optimize translated queries by eliminating ambiguous terms incrementally from the target query according to the similarity of the query and target query's retrieval characteristics. In the best cases, we can approach 80-90% of the performance of a monolingual baseline system.

To briefly summarize the work on CLTR to date, we feel that minimal bilingual dictionaries are useful and plentiful but the quality of these resources make a big difference and vary considerably. The problem of selecting the right translation term can be partially overcome by automatic techniques including some corpora based methods. Parallel corpora are hard to come by, however, and produce noisy results outside of the subject domain of the parallel texts. We believe that the CLTR user can play an important role in aiding the disambiguation process through an interactive user interface. We have begun to explore interactive methods for query formation that could help solve some of these problems and describe this work below.

In addition we have recently conducted a set of studies which address the second important element of the interactive CLTR interface, the retrieval results display. First, we discuss a set of experiments we have conducted using a monolingual system to address an issue with visualizing the relationship between query terms and retrieved documents.

## 3. Evaluating Document Thumbnail Visualizations



**Figure 1. The document thumbnail portion of the J24 interface**

One of the goals of studying interactive information retrieval is to evaluate the effectiveness of new user interface techniques on users' ability to find information relevant to their needs. In this section, we describe our efforts to investigate the effectiveness of an interface technique we call "Document Thumbnail Visualizations," which represents retrieved documents with small graphical images. We attempt to evaluate these effects in several experiments with real users. In the process, it became clear that modeling the user's task helps to identify the locus for any potential effects of the interface technique.

In a recent experiment, Kaugars [7] developed a document visualization system that displayed the results of a keyword text search as a set of small graphical representations (thumbnails) of the top 20 retrieved documents laid out in one window and a single document viewer in a second window. Document thumbnails presented color-coded highlighting indicating positions and identities of search terms in the document, and the single document viewer showed a multiple focus+context fisheye view [8] of the document focused on multiple regions defined by the sentences holding search terms. He compared this system to another more traditional system that represented documents as lists of short titles and a simple scrolled window document viewer. Kaugars found that people were faster and better at making relevance judgments for a fixed set of retrieved documents when using the thumbnail/fisheye system.

At least two important aspects of the Kaugars interface could have led to its superior performance, either the document-thumbnails view or the fish-eye document view (or both). The thumbnail view of the documents allows users to quickly scan the returned document set for instances of query terms and query term collocations and their distribution in and between documents. It gives the user in a single glance

information about which documents contain which query terms and which query terms are missing or how often they appear in the documents. Because the thumbnails retain the familiar shape and format of a document, the user can easily see how the query terms are distributed in the actual document. They could help users to locate information within a document and could help them answer the question "why was this document retrieved?"

Alternatively, the unique document viewer could have led to the superior performance. A fish-eye focus+context view presents the area of current interest in normal scale but as distance from the interest area increases, information scale decreases. In Kaugars's document viewer, there could be multiple areas of interest each defined by the presence of a search term in a sentence. These areas would be displayed in a normal font. Other sentences and paragraphs were shown in a smaller font. This allowed the user to very easily find and read relevant passages while ignoring intervening and mostly irrelevant text. Users could define new areas of interest in the document with a mouse click thereby returning those areas to a normal sized font. Because the documents were much smaller in size than the normal full sized scrolled view, much more of the relevant text fit within a single window on the computer screen perhaps making it easier to make relevance judgements.

From this experiment by Kaugars, we could not determine the relative importance of these two interface features. The goal our initial study [9] was to investigate the role of the thumbnail-document view feature exclusive of the fish-eye document view feature. In addition, we wanted to test the potential advantage of the thumbnail view within the context of an interactive text retrieval task that engaged the user more than the relevance judgement task. Thus, we attempted to replicate Kaugars result using an interactive, WEB version of a thumbnail document set viewer and using the prescribed TREC-7 interactive track methodology

[10]. The system (named J24 after the July 24 deadline for its completion) provides a WEB interface for entering search terms and displaying results with thumbnails for 10 documents at a time. Figure 1 shows how the document thumbnails were presented. From this example, it is easy for users to eliminate documents that do not include the search term “beatification”. The J24 system was compared to a control system, ZPRISE [11] that displays lists of documents with titles only.

The details of the study are presented in [9]. The data we collected using the TREC-7 interactive methodology failed to replicate any advantage for the thumbnail displays. We discovered that it is very difficult to determine what effect document thumbnail views may have on the interactive retrieval task from the measures available in the TREC data. Summative data like average recall and precision may mask the underlying effects of our interface technique. The following steps represent a stage model of the user’s task.

- 1) Read Topic
- 2) Choose query terms
- 3) Enter terms
- 4) Scan returned document list
- 5) Choose documents to read
- 6) Read relevant passages
- 7) Decide relevance
- 8) Determine if new instance or aspect
- 9) Save document
- 10) Modify query?
- 11) Finished

Most of the user’s time and decision-making effort is spent reading the document for content. Remember that original finding by Kaugars suggested that there was a benefit for document thumbnails in the context of a relevance judgement experiment. Thus, participants in his experiment were engaged in only two of the stages enumerated above, document selection, and relevance judgement. Also remember that Kaugars had two interface techniques. One technique, document thumbnails, could affect the document selection stage and the other technique, document fisheye views, could have affected the relevance judgement stage. In the two experiments, we wanted to test the relative importance of document thumbnails in these two subtasks more directly by changing the test user’s tasks.

#### 4. Relevance Judgement Task

In this study, we created two versions of the J24 interface, one showing the document thumbnails without the document title view and one showing the title view

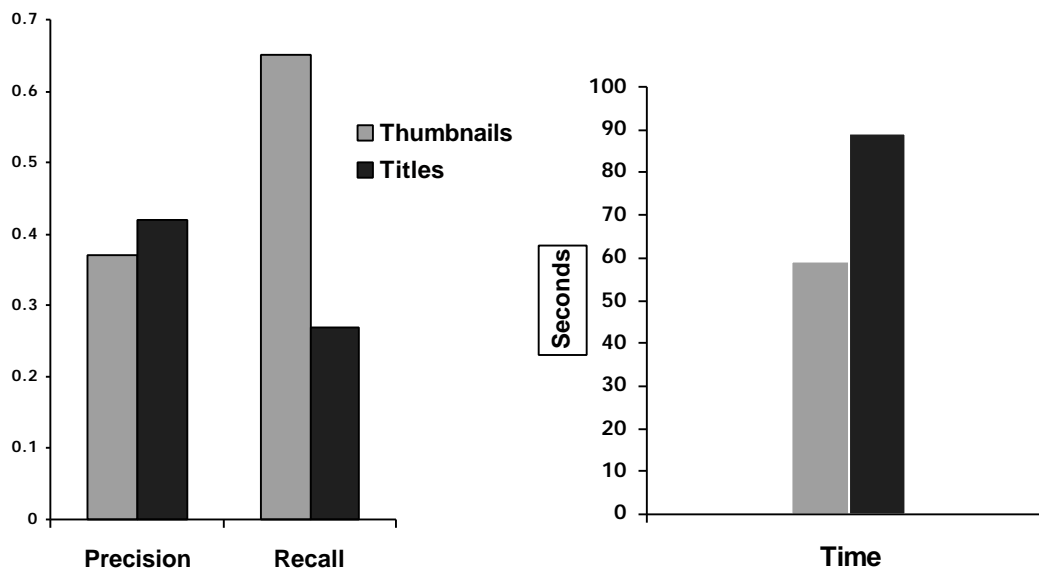
without the document thumbnails. Eight participants used each version of the interface to judge the relevance of a set of 20 documents for each of four TREC-7 topics. The participants were novices, recruited from the college student population and paid for their time. The participants using the thumbnail version given were pointers on how to use them to select documents but otherwise training was minimal in both conditions. Again, the results show that there were no significant differences in the accuracy of relevant judgements or in the time used.

Both document selection and relevance judgement stages were involved in this task. Users need to read the documents to make relevance judgements and this again is their primary activity. Users were not being selective when choosing documents to read, as most participants tried to read as many documents as possible in the time allowed. Thus, again, the study is not sensitive to the advantages thumbnail views may give to a user of an interactive system. In our final study, we use a task that only involves document selection.

#### 5. Document Selection Study

In this study, participants were asked to judge the relevance of a document without reading it. The J24 interface was adapted again to only show either the document thumbnails or the document titles. The document text view was not shown. Thus, in this study participants were asked to judge the relevance of documents based only on the information contained in the thumbnail views or in the title fields. Four participants each used both versions of the system to judge 20 documents for relevance to each of 11 TREC-7 topics. The topics were chosen if the top 20 documents had a heterogeneous distribution of query terms among them. In this way, the thumbnail views would be at least guaranteed to provide some information with which to distinguish documents.

In this study, there was a clear advantage for the thumbnail view condition. Average time to judge 20 documents was much less with the thumbnail view and participants judged a higher proportion of relevant documents correctly (Recall). These results can be seen in Figure 2. Participants using the thumbnails views were on average 30 seconds faster,  $t(3) = 3.82, p < .01$ , and had a .38 higher average Recall score,  $t(3) = 7.16, p < .001$ . There was no significant difference in the Precision score. The higher Recall scores were primarily due to participants judging more documents as being relevant. They judged that an average 4.62 documents were relevant when viewing the thumbnails but an average 2.07 documents when viewing titles,  $t(3) = 4.77, p < .005$ .



**Figure 2. Results of the document selection study**

The combined results of these interactive studies show that the thumbnail views clearly have better information than the titles for judging document relevance. However, when put into the context of interactive tasks that demand document reading, the document thumbnail views have very little impact on the overall performance of our test users.

Document thumbnail visualizations have proven to be a valuable interface element in the document selection task. They provide much of the same information that the TileBars interface [12] provide. Like TileBars they allow the user to make informed decisions about which documents contain relevant text and which regions of those documents to view. Both interfaces provide information about the relative length of the document, the frequency of the query terms in the document and the distribution of the query terms in the document. Unlike TileBars, however, thumbnail views present this information in an immediately recognizable and familiar format. The thumbnail image actually looks like a small document and important query word collocations are easy to find. This advantage comes at the expense of compactness however. TileBars take less screen real estate.

Finally, we can conclude that the primary effect Kaugars found in the relevance judgement task was not due to the thumbnail view but in most likelihood was due to the multiple fisheye context+focus document viewer. Query term highlighting is an important component of these types of displays and has been shown to be useful over and over again [13]. Other examples include the

SuperBook system [14], the SeeSoft visualization system [15], and the Cha-Cha interface [16].

These conclusions have several implications for cross-language text retrieval system design. Choosing documents without reading is an important activity in these systems when documents are in languages unfamiliar to the user. So, the important features of the thumbnail display (e.g. highlighting query term distribution in the documents) should be preserved in our CLTR interface. In addition, when reading is necessary, it is important to focus the user on relevant sections of the document. In the next sections, we begin to explore ways of accomplishing this.

## 6. Cross-language relevance judgement

Until now we have been discussing the improvements that can be made to an interactive retrieval system to aid the decisions a user must make about the documents returned by the system when the documents are in a language the user can read. In an important study, Resnick [17] showed how users might be able to make these types of judgements about information contained in documents in languages they cannot read. People who knew no Japanese, were asked to sort Yellow Page entries that had been translated from Japanese to English using a very simple word-for-word “gist” method. The sorting was compared with sorts done with same entries

in original English and with a random sort. The “gist” sorts were more consistent with the original English than with the random sorts indicating that people were able to glean useful information from simple translations. This result gives hope for designing CLTR systems that can be used by the monolingual person. In the following section, we evaluate a system to meet these needs.

## 7. Interactive CLTR

In an effort to understand how CLTR systems might be improved, we have recently conducted some preliminary experiments on an interactive, cross-language text retrieval system. The system, Arctos, provides a user with a browser-based interface with which to enter English queries. After an initial query is entered, the query is translated using a simple word-for-word or phrasal translator. The user can then interactively improve the query translation using links to on-line bilingual translation resources and then submit the query for retrieval against document collections in the target language. The retrieved documents are presented using document thumbnails and query term highlighting. Further, the user can have the returned document translated from the target language to English by the Babelfish translation engine from Systran and made available by Alta Vista.

The interactive task in this preliminary study was for one user to use English TREC CLTR track topics to retrieve and judge the relevance of German documents. The user, who judged himself to have no German language knowledge, formed his own English query based on the TREC topic statement in English. He then modified and improved the German query while examining documents and using the on-line dictionary resources, submitted the modified query to the URSA1 engine for retrieval, examined the retrieved documents using the German equivalents and document thumbnail interface, submitted the documents to the Babelfish translation engine to translate to English, and judged the top 10 documents retrieved as either relevant or non-relevant. No time limit was set and the user spent approximately 8 hours working on 22 TREC topics.

The results are shown in Figure 3 and consist of a comparison of the relevance judgements made by the user in the study to the “correct” judgements provided by NIST for the TREC-6 Cross-Language evaluation track. Documents were excluded if this system retrieved them but they were not in the TREC judgments for these queries. From these numbers, we summed the counts and calculated a False Hit Ratio, that is, the ratio of the number of documents that were irrelevant but judged relevant to the number relevant and judged so. The false

		User Judgements	
		relevant	not relevant
NIST Judgements	relevant	69	1
	not relevant	11	43

**Figure 3. Number of judged documents in the Arctos study**

hit ratio for this experiment is only 15.9% We further calculated the False Drop Ratio at 2.32%, which indicates the relative number of times a relevant document was incorrectly judged irrelevant to the number of times an irrelevant document was also judged irrelevant. The combined performance figures for this system indicate a very low percentage chance of error in using this cross-language retrieval system.

What led to the relatively good performance by this user who could search for and identify relevant German documents with no German language knowledge? The user reported using a very wide range of techniques to modify queries, and reported particular difficulties with generating good phrasal equivalents in German, where compound nouns are often extremely important as query elements. The primary resource used to improve and evaluate queries was an on-line bilingual dictionary that was used to “back-translate” query terms selected by the URSA query translation process. The dictionary interface listed all matching entries with their English definitions and the user could easily pick correct query terms to expand the search. This process of user-aided query expansion has been streamlined in the Keizai prototype described below. Whereas the Arctos interface required cutting and pasting from the dictionary pages to the query entry pages, the Keizai interface integrates the two processes.

The Arctos user was able to use the document thumbnail and query term highlighting features to identify potentially relevant German language passages. To make relevance judgements he then selected these passages for translation with the Babelfish translation engine. This process has also been streamlined in the Keizai prototype with automatic summarization and translation of the retrieved documents.

## 8. The Keizai CLTR system prototype

The Keizai project is focused on providing a demonstration of an end-to-end Web-based cross-language text retrieval system. Beginning with an English query, the system will search Japanese and Korean Web data and display English summaries of the top ranking documents. User should be able to accurately judge which foreign language documents are

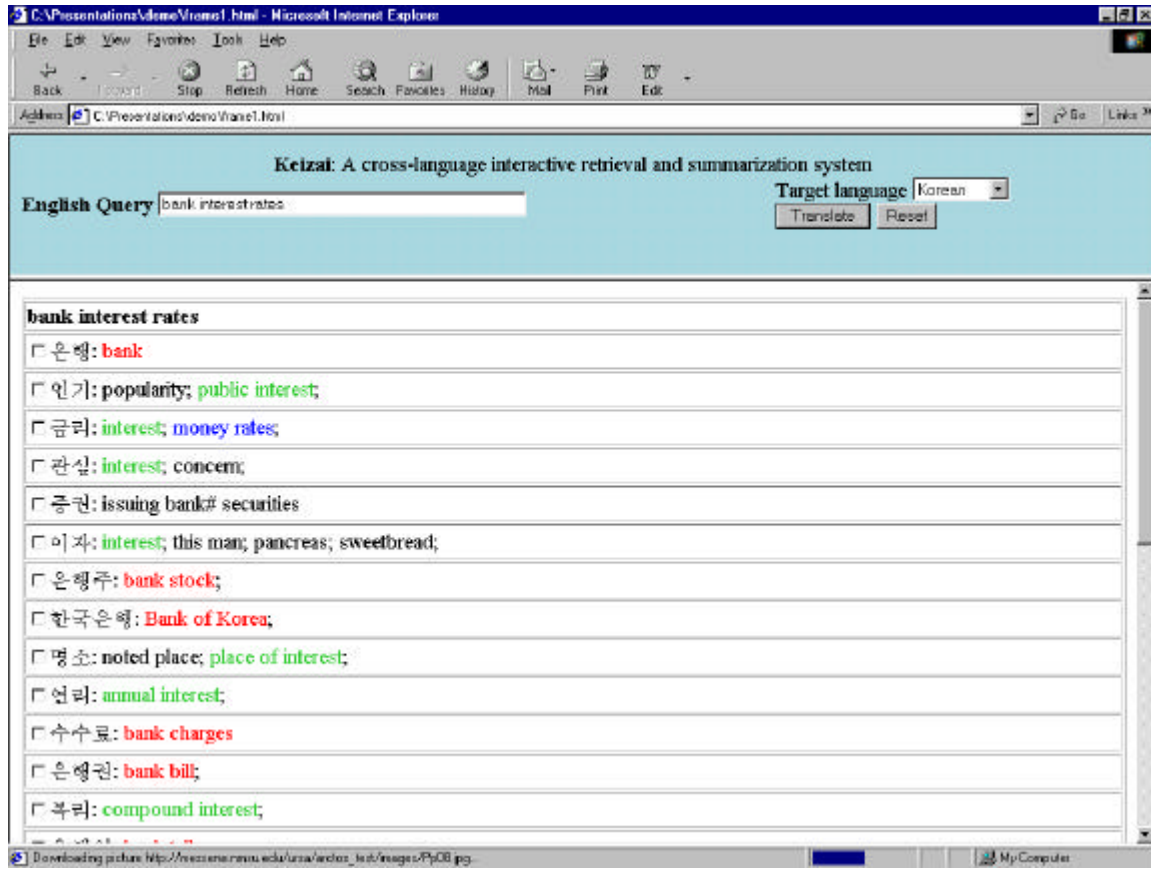


Figure 4. Keizai query term selection.

relevant to their query. The system currently searches archived Web news in Japanese from the Asahi Shimbun news service and in Korean from the Dong-A Ilbo.

The prototype system has undergone extensive revision and iterative design. The design of the query interface and query translation module has been improved through several design reviews and usability walkthroughs. As a result of these and our experience with Arctos and other CLTR research, the test system now implements a new approach to query translation that presents extended English definitions of query terms and phrases alongside their Japanese or Korean translations. The user can select the English definition that most accurately reflects their intention in the original query. Query term disambiguation is handled by user intervention without the need for morphological analysis and segmentation. A user who knows no Korean or Japanese selects the correct surface forms of the Korean or Japanese query terms. At the same time, the query translation module has been improved to only select terms that actually occur in the target data.

Figure 4 shows the current design. In this example, the user has entered an English query, "bank interest rates". The system has selected Korean query terms from a Korean-English lexicon that could translate to one or

more of the query terms. A pre-search eliminates terms that will not return any documents, and then the terms are sorted by frequency. The user can simply select those terms whose definitions are consistent with their information need.

Similar progress has been made in the WWW-based presentation interface (see Figure 5). Dynamic HTML capability is being used to quickly and compactly display document summaries and the distribution of query terms in documents are displayed in both the original and English forms. While document thumbnails are not being used, query term occurrences for each document are being represented. Documents are being summarized in the original language, Japanese and Korean, by an internally developed summarization engine. Summarization is based on a variety of statistical and symbolic techniques and is parameterized to produce a variety of summary types for different goals. In this case, query specific summaries are produced which focus on passages containing query terms. The original summaries are then translated to English using machine translation facilities for Japanese and Korean.

Documents are sorted by a relevance measure and are represented by the first line of the summary for each retrieved document which is displayed in the top portion

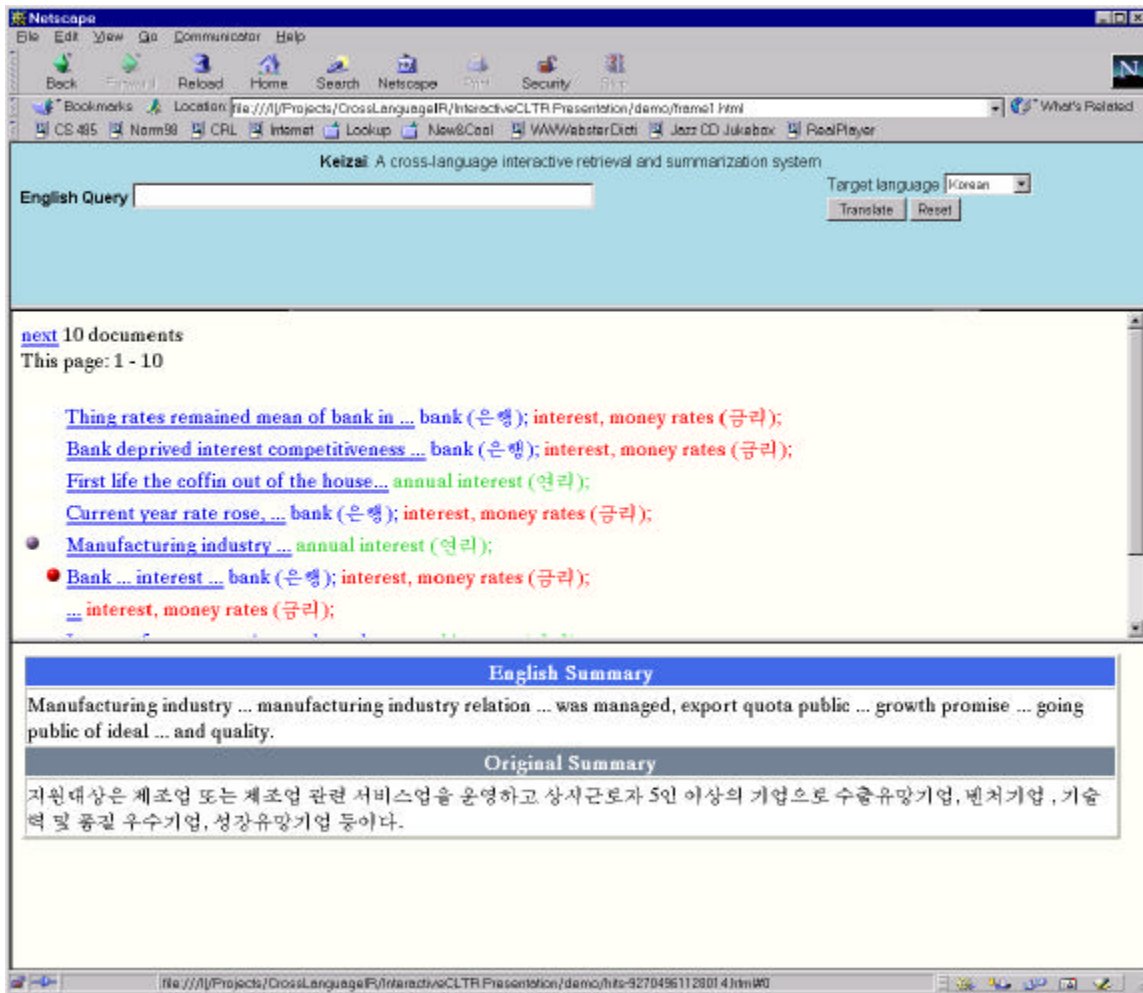


Figure 5. Keizai results display

of the results display along with the color coded query terms occurring in the document. Both the original and the English summaries are dynamically displayed in the bottom portion of the display for each document when the mouse pointer is positioned over the document line giving users a very quick way of evaluating document relevance.

The Keizai demonstration interface brings together many of the findings of our work on CLTR and Interactive IR. However, we have yet to test a number of research issues with the system. Of primary importance is the quality of the translated summaries. We need to do more work here and we plan to evaluate the effectiveness of the Keizai with test users in the near future.

A number of related projects use some of the same techniques found in Keizai. For example, the MULINEX system [18] provides query translation into German, French and English and has a "Query Assistant" which is very much like the query term selection interface in Keizai. There are plans for adding translated summaries as well.

## 9. Conclusion

The goal of our text retrieval research is to get users the texts they want in as useful a form as possible, in as efficient a manner as possible, and without regard for language. The systems we have produced---especially *Arctos*, *J24* and *Keizai* ---bring five years of research together that unite multi-lingual, cross-language and interactive technologies in an effort to provide the beginnings of a comprehensive text retrieval system. Ongoing work is aimed at improving our understanding of the effectiveness of various metaphors for visualization of retrieval results. While the work on thumbnail representations demonstrates that relatively simple mechanisms for visualization can provide advantages to an end user over standard WWW-based retrieval presentation methods, there is clearly an enormous range of work that can still be done. First, the interactive cross-language retrieval evaluations of *Arctos* demonstrates

that keeping the human “in the loop” during query formulation has significant value in the CLTR scenario. There may be additional avenues and evaluation methodologies that reveal ways query improvement can be done interactively for the monolingual case as well. Among these are detecting terms with that are polysemous and soliciting greater user input about the desired meaning of the term. This same problem may also be handled by modeling the system user. If a user is known to be interested in financial information, then a query can be expanded by “investment bank” when they enter a query containing “bank” and not by “Mississippi” or “levee”. CLTR research may be approaching a standstill in that machine translation technology exists for many core languages---and is reasonably priced and available. Our changed focus on developing simplified tools and approaches to CLTR has been driven by the recognition that the less frequently used languages will remain neglected by MT vendors for purely economic reasons. Research here is focusing on developing shallow tools for leveraging small bilingual lexicons that are freely available and acquiring parallel and non-parallel corpora from the huge body of documents on the Web. Making maximum use of these tools also involves active research on interactive methods. The *Keizai* system shows that the shortcomings of noisy and small bilingual resources may actually be offset by maximizing the understanding of even a monolingual system user.

## 10. References

- [1] Salton, G. Automatic Processing of Foreign Language Documents, Prentice-Hall, Englewood Cliffs, NJ. 1971.
- [2] Susan T. Dumais, Todd A. Letsche, Michael L. Littman and Thomas K. Landauer Automatic Cross-Language Retrieval Using Latent Semantic Indexing, , in Cross-Language Text and Speech Retrieval, AAAI Technical Report SS-97-05. Available at <http://www.clis.umd.edu/dlrg/filter/sss/papers/>.
- [3] Lisa Ballesteros and W. Bruce Croft, Phrasal Translation and Query Expansion Techniques for Cross-Language Information Retrieval, in Proceedings of the 20th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Philadelphia, July 1997, pp. 84-91. Earlier work by the same authors is available at <http://ciir.cs.umass.edu/info/psfiles/irpubs/ir.html>
- [4] Davis, M., and Dunning, T., A TREC evaluation of query translation methods for multi-lingual text retrieval", In Proceedings of the 4th Text Retrieval Conference (TREC-4). 1996. Available at <http://trec.nist.gov/pubs.html>.
- [5] Davis, M. 1996. New experiments in cross-language text retrieval at NMSU's Computing Research Lab. In Harman, D. K., ed., The Fifth Text REtrieval Conference (TREC-5). NIST. 1996. Available at <http://trec.nist.gov/pubs.html>.
- [6] Mark W. Davis and William C. Ogden, QUILT: Implementing a large-scale Cross-Language Text Retrieval System. In Proceedings of the 20th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Philadelphia, July 1997, pp. 92-98.
- [7] Karlis, J. Kaugars, A Hierarchical, Approach to Detail + Context Views. Unpublished Doctoral Dissertation. New Mexico State University. 1998.
- [8] Furnas, George W., Generalized fisheye views. In Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems, ACM (April 1986), 16-23.
- [9] William C. Ogden, Mark W. Davis, and Sean Rice. Documents thumbnail visualization for rapid relevance judgments: When do they pay off? In Harman, D. K., and E. M. Voorhees ed., The Seventh Text REtrieval Conference (TREC-7). NIST. 1998. Available at <http://trec.nist.gov/pubs.html>.
- [10], TREC-7 Interactive Track Home Page [URL] <http://www-nlpir.nist.gov/projects/t7i/t7i.html>
- [11] Z39.59/PRISE [URL] <http://www-nlpir.nist.gov/works/papers/zp2/zp2.html>
- [12] Hearst, Marti A, TileBars: Visualization of Term Distribution Information in Full Text Infor-mation Access, CHI '95 Proceedings, 1995.
- [13] Marti A. Hearst. User Interfaces and Visualization. In Modern Information Retrieval, edited by Ricardo Baeza-Yates and Berthier Ribeiro-Neto, Addison Wesley/ACM Press, New York, 1999.
- [14] Landauer, Thomas K., Dennis E. Egan, Joel R. Remde, Michael Lesk, Carol C. Lochbaum, & Daniel Ketchum. Enhancing the usability of text through computer delivery and formative evaluation: the superbook project. In Hypertext: A psychological perspective, ed. by C. McKnight, A. Dillon, & J. Richardson, 71-136. Ellis Horwood. 1993.
- [15] Eick, Stephen G. Graphically displaying text. Journal of Computational and Graphical Statistics 3, 127-142, 1994.
- [16] Chen, Michael, & Marti A. Hearst. Presenting web site search results in context: A demonstration. In Proceedings of the 20th Annual International ACM/SIGIR Conference. Page 381, Melbourne, Australia, 1998. On the web at: <http://chacha.berkeley.edu>.
- [17] Philip Resnik, Evaluating Multilingual Gisting of Web Pages, in Cross-Language Text and Speech Retrieval, AAAI Technical Report SS-97-05. Available at <http://www.clis.umd.edu/dlrg/filter/sss/papers/> .
- [18] <http://mulinux.dfki.de/demo.html>

