Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Library & Information Science Research 27 (2005) 249–263

Library & Information Science Research

Web searching across languages: Preference and behavior of bilingual academic users in Korea

Hae-young Rieh[a,*], Soo Young Rieh[b]

[a]*Center for Liberal Arts and Instructional Development, Myongji University, San 38-2 Nam-dong, Yongin-si, Kyonggi-do 449-728, Korea*
[b]*School of Information, University of Michigan, 550 East University Avenue, Ann Arbor, MI 48109-1092, USA*

Available online 14 April 2005

Abstract

The problem of language in Web searching has been discussed primarily in the area of cross-language information retrieval (CLIR). However, much CLIR research centers on investigation of the effectiveness of automatic translation techniques. The case study reported here explored bilingual user behaviors, perceptions, and preferences with respect to the capability of the Web as a multilingual information resource. Twenty-eight bilingual academic users from Myongji University in Korea were recruited for the study. Findings show that the subjects did not use Web search engines as multilingual tools. For search queries, they selected a language that represents their information need most accurately depending on the types of information task rather than choosing their first language. Subjects expressed concerns about the accuracy of machine translation of scholarly terminologies and preferred to have user control over multilingual Web searches.

1. Problem statement

The Internet has made it possible for people to gain access to more information in various languages more quickly than ever before. They can reach a wide range of information globally. However, such searching creates difficulties in terms of the language chosen for the

* Corresponding author.
E-mail addresses: hyrieh@mju.ac.kr (H.Y. Rieh), rieh@umich.edu (S.Y. Rieh).

0740-8188/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.lisr.2005.01.006

search and the amount of information available in foreign languages. Numerous studies of Web searching behavior have inquired into who searches the Web, what kinds of search task they perform, and how they search (Hsieh-Yee, 2001). Still, a number of areas of Web searching behavior remain unexplored. One such area is the problem of multilingual searching on the Web. People who speak more than one language often face the problem of choosing the appropriate language for their search. Apparently they make a decision about language with regard to two aspects: language for (1) query terms and (2) retrieved documents. The choice of language for query and documents is not necessarily identical or consistent. Some users enter queries in their native language and prefer to retrieve documents in a foreign language, while others prefer the opposite. Or, users enter queries in one language but want to retrieve the documents in multiple languages. Their decisions may be related to such factors as individual user ability to speak a foreign language, nature of user goals and search tasks, and other situational aspects.

According to Global Reach (2004), which has tracked non-English online populations since 1995, 64.2% of the world's online populations as of March 2004 are non-English speakers. Non-English speaking online users include, among others, Chinese (14.1%), Japanese (9.6%), Spanish (9.0%), German (7.3%), and Korean (4.1%). English is still the dominant language on the Web, accounting for 68.4% of the total. Other languages available on the Web include, among others, Japanese (5.9%), German (5.8%), Chinese (3.9%), French (3.0%), and Spanish (2.4%). Given these statistics, it can be presumed that a substantial number of non-English speaking users can access Web pages written in English. However, little is known on how users make their decision about the language to use, how they utilize multilingual resources, and how they consider integrated multilingual searches in the Web.

This study explores user behaviors, perceptions, and preferences with respect to the capability of the Web as a multilingual information resource. The purpose of this study is to identify the implications for designing information retrieval systems that support user interactions with multilingual collections on the Web. Specifically, this research addresses the following questions:

- To what extent do bilingual Korean scholars conduct multilingual searches on the Web?
- How do the Korean scholars decide which search engines to use when searching for multilingual information?
- What preferences do the Korean scholars have for integrated multilingual search tools?

2. Literature review

A considerable number of recent studies have examined Web searching behavior using various methods including query analysis (e.g., Rieh & Xie, 2001; Spink & Jansen, 2004; Wang, Berry, & Yang, 2003), laboratory experiments (e.g., Hargittai, 2002; Kim, 2002; Wang, Hawk, & Tenopir, 2000), and interviews in natural settings (e.g., Choo, Detlor, & Turnbull, 2000; Rieh, 2004). Despite the worldwide nature of the Web, however, international user searching behaviors have rarely been investigated. Studies outside the

United States such as in Australia (Applebee, Clayton, & Pascoe, 1997), Israel (Lazinger, Bar-Ilan, & Peritz, 1997), and Netherlands (Voorbij, 1999) focus on Internet use and activities rather than on Web searching behavior. In reviewing multilingual access issues and problems on the Web, Large and Moukdad (2001) pointed out that non-English speakers joined the Web faster than English speakers: the former had grown from 10% in 1995 to nearly 50% in 1999. They also noted that Web pages in languages other than English were growing faster than English-language pages. Consequently, they viewed the Web as a multilingual resource searchable by a multilingual user community but lacking a common interlingua. They maintained that the language itself became a barrier to full exploitation of Web resources.

Rieh and Rieh (2001) reported that Korean science and engineering scholars participating in their study adopted various search strategies for their research-related information seeking as well as personal pursuit. While the subjects rarely turned to search engines to look for research information, they did use search engines when searching for personal information. A major reason for accessing known Web sites such as academic research labs, professional organizations, or journal publishers directly for their research-related Web searching, rather than using search engines, was that they perceived the Web's scholarly information resources to be of little value. Rieh (2002) presented the selection criteria of search engines and the use of advanced search features in which subjects tended to select search engines based on their knowledge of search engines features or their familiarity with an engine rather than on previous experience with search results. The problem of multilingual Web searching was not included in Rieh's (2002) article.

Iivonen and White (2001) compared the choice of initial search strategies (direct address, subject directory, and search engines) of Finnish and American Web searchers. Twenty-seven graduate students from each country participated in the study, which found that Finnish searchers used search engines more and relied on directories less than the Americans. The groups were similar in their use of the direct address strategy. Iivonen and White speculated that the American searchers' familiarity with the English language led them to use subject directories more extensively, while Finnish searchers' lesser familiarity with English led them to use search engines more. They also found that the reasons for choosing a certain strategy differed in the two groups. Finnish participants focused more on question-related reasons such as judgments about the presence or absence of good search terms. American participants showed greater sensitivity to specificity, currency, and familiarity of a topic than did the Finns. The Americans paid greater attention to source-related reasons and strategy-related reasons. Although their studies (Iivonen & White, 2001; White & Iivonen, 2001) discussed the role of language familiarity in the process of Web searching, they did not specifically deal with multilingual searching behavior.

Web search in multilingual collections has long been discussed primarily within the context of cross-language information retrieval (CLIR). In recent years, information retrieval researchers have focused on CLIR, partly due to the growth of the Web. CLIR deals with retrieval situations in which users post queries in one language but expect to receive search results in another language. In general, the underlying CLIR assumption is that users start with a query in their native tongue; the software translates the query from the user's language into the language of the documents; and users obtain relevant documents in other languages. In most cases the translation is done in a word-by-word fashion, using a dictionary or a machine translation system (Lavrenko, Choquette, & Croft, 2002).

While the volume of reported research on CLIR is exploding very rapidly (Oard, 1997), most CLIR research efforts have focused on developing techniques that translate queries from a user's language into the language of the documents and then evaluating the effectiveness of the translation techniques (e.g., Maeda, Sadat, Yoshikawa, & Uemura, 2000; Pirkola, Hedlund, Keskustalo, & Järvelin, 2001). The techniques can be characterized as dictionary based (e.g., Ballesteros & Croft, 1996; Hull & Grefenstette, 1996) or corpus based (e.g., Davis, 1997; Davis & Dunning, 1996; Dumais, Letsche, Littman, & Landauer, 1997). The dictionary-based approach uses a bilingual electronic dictionary to replace source language query words with their target language equivalents. In this approach, translation poses the problem of irrelevant words generated in translation when there is lexical ambiguity in the source language or target language. Therefore, research efforts have been devoted to disambiguating translations through the use of contextual clues (Xu, Weischedel, & Nguyen, 2001), statistical co-occurrence, or triangulated translation (Ballesteros & Croft, 1998; Gollins & Sanderson, 2001).

Various CLIR models have been proposed and tested to improve translation performance. For example, Pirkola et al. (2001) proposed a structured-query model in which the structuring of queries referred to the grouping of search keys and the use of proper query operators. Lavrenko et al. (2002) proposed a CLIR relevance model that did not rely on word-for-word translation of the query. Their model attempted to construct an accurate relevance model in the target language, used that model to rank the documents in the collection, and integrated such techniques as query expansion for dealing with synonymy and translation disambiguation to handle polysemy. Bian and Chen (2000) developed a model that adopted both bilingual dictionary and monolingual corpus-based approaches to select suitably translated query terms.

In general, only a few studies have addressed user interactions with CLIR systems. Even when users were included in the research, their involvement was limited to the role of evaluators for interface design or for relevance assessment. For instance, Ogden and Davis (2000) did research primarily concerned with query translation and multilingual text summarization. They conducted user studies to evaluate the effectiveness of user interface techniques called "Document Thumbnail Visualization," a Web-based cross-language text retrieval system. Their experiments showed that the thumbnail views were effective only when the users had tasks what did not require reading documents. When users had interactive tasks, document thumbnail views had little impact on overall user performance. Karlgren and Hansen (2002) reported on an experimental study on cross-language relevance assessment with Swedish-speaking subjects fluent in English. They compared performance and relevance assessment of subjects to produce Swedish and English search results in which they used simulated domain and work–task scenarios and found that relevance assessment in a foreign language (English) was more time consuming and arduous than in a first language (Swedish). They also revealed that not only topicality but also task was related to relevance assessment.

The Clarity project researchers took the user-centric design approach in developing a CLIR system (Hansen, Petrelli, Karlgren, Beaulieu, & Sanderson, 2002; Petrelli, Beaulieu, Sanderson, Demetriou, & Herring, 2004; Petrelli, Hansen, Beaulieu, & Sanderson, 2002). They argued that designers of CLIR systems should examine cross-lingual information search tasks in real environments with real users to overcome the mismatch between user goals and system mechanisms. Based on interviews and at-work observations with 10 subjects (business analysts, journalists, librarians, translators, etc.), they identified a number of user requirements for CLIR systems including the capacity to search multiple languages simultaneously, to change query languages within the same search session, to support multilingual queries, and to filter results by language, genre, date, or other features. Petrelli et al. (2004) reported that the users used the most suitable language they were familiar with for their task, which was not always their native language. It was found that English was used as a pivot in searching because of its international dominance in technical jargon. It was also found that search behaviors were dependent upon user goals and purposes for searching as well as on "language knowledge of individuals and cognitive demands of the cross-language task itself" (p. 928).

A brief summary of CLIR research to date shows that techniques of developing bilingual dictionaries and improving translation ambiguity in dictionaries have advanced in general although the quality of the techniques varies considerably. Most CLIR studies fail to address the importance of user interaction in CLIR systems. Even if users are recruited in research design, their roles tend to be those of passive evaluators rather than active information seekers. Petrelli et al. (2002, 2004) may be the first attempt in the CLIR research community to develop a process consisting of four phases, including scenarios, requirements specification, design and formative evaluation, and redesign.

3. Procedures

The study participants were 28 volunteers including faculty members, doctoral students, and a post-doctoral fellow recruited from the Science and Engineering Campus of Myongji University in Korea. Judgmental sampling (Krathwohl, 1993), which is a frequent strategy of qualitative research, was employed. According to Krathwohl (1993), the judgmental sampling researcher selects individuals "who are presumed to be typical of certain segments of the population and therefore representative of it" (p. 137). The composite of the sample represents the range of academic users by including full professors (n = 8), associate professors (n = 7), assistant professors (n = 5), post-doctoral fellow (n = 1), and doctoral students (n = 7). The sample also includes a range of disciplinary areas including physics, chemistry, mathematics, electrical engineering, electronic engineering, mechanical engineering, industrial engineering, computer science, telecommunications engineering, architecture, and clothing and textiles. Twenty-two of the subjects were male and six were female; their ages ranged from 26 to 51.

Most subjects were heavy Internet users who have used the Internet for more than 3–5 years. All subjects except one had used commercial databases for their research. However, only one subject had attended a formal training session on Web searching. Table 1 summarizes other characteristics of the subjects.

The subjects were bilingual academic users who could read, write, and speak both Korean and English. Korean scholars, especially in the field of science and engineering, are expected to publish their research in English-language journals to gain tenure and promotion. Doctoral students also have requirements to publish their research in English. These scholars in science and engineering are more likely to publish in international journals (most of these being published in English) and interact with foreign scholars and foreign information sources, while scholars in humanities and social sciences more likely focus on research problems in Korea and publish in Korean journals. Sandelin and Sarafoglou (2004) analyzed articles in *Science Citation Index*, *Social Science Citation Index*, and *Arts and Humanities Citation Index* databases between 1998 and 2000 and reported that Korean journal articles per million inhabitants in the natural sciences (in SCI) was 862 versus 20 in the social sciences (in SSCI) and 1 in the arts and humanities (in A & HCI). The rank of article numbers in the natural sciences among countries was 26, while the social sciences' rank was 28 and arts and humanities' 30.

Again, science and engineering scholars were chosen as the sample population by virtue of their being scholars in that field and thus more confident about searching for Web information in foreign languages than scholars in the social science or humanities. Previous research has revealed that members of the science faculty used the Internet more heavily than did members of the humanities or social sciences faculties (Lazinger et al., 1997). And scholars in humanities have adopted new technologies relatively slowly as compared to other fields (Wiberley, 1991; Wiberley & Jones, 1994). In addition, science and engineering scholars require more current information than do scholars in other disciplines (Ellis, Cox, & Hall, 1993).

The data were collected in natural settings by conducting semi-structured interviews and taking observations from July to August 2001. The faculty member and post-doctorate fellow interviews were conducted in the subjects' offices whereas doctoral students came to the

Table 1
Characteristics of subjects

Question	Responses	Frequency (%)
Internet experience	5–10 years	23 (82.1)
	3–5 years	5 (17.9)
	Less than 3 years	0 (0)
Internet use	Over 5 hours/day	5 (17.9)
	More than 1 hour but less than 5 hours/day	12 (42.9)
	Less than 1 hour/day	11 (39.3)
Search engine use	At least 2–3 times a day	13 (46.4)
	About once a day	6 (21.4)
	More than once a week but less than once a day	5 (17.9)
	Less than once a week	4 (14.3)

researcher's office to participate in the study. Every interview was conducted at a computer station with a high-speed Internet connection so that the subjects could readily demonstrate their search behaviors from time to time during the interview.

Each interview was structured around the following topic areas:

- information-seeking habits on the Web for professional activities and tasks;
- Web searches for research tasks and personal pursuits (scholars were asked to demonstrate their typical behaviors);
- favorite search engines and reasons for such preferences;
- Web searches when both Korean and English documents are needed; and
- perceptions of and preferences for multilingual Web searching.

Questions about the language choice were deliberately avoided so that they could show their natural search behaviors without paying specific attention to the language used. The one-hour interviews were audiotaped for subsequent transcription and analysis; the interviewer also took detailed notes during the course of the interviews.

All of the interviews were transcribed from the audio-tape recordings, and the interview transcripts were collated along with notes taken during the interviews. The transcripts were then organized around various themes: kinds of search engines used, nationality of the search engines, reasons for choosing particular search engines, search terms entered, language of search queries, language of the Web sites, opinions on multilingual searches, use patterns of foreign documents for research and other purposes, reasons for using foreign documents, and so on. All the interviews were conducted in Korean and were transcribed in Korean as well. The responses of subjects quoted below are translated into English by the first author and then verified by the second author for accuracy.

4. Results

4.1. Multilingual searches on the Web

More than three-quarters of the subjects (*n* = 22; 78.6%) relied on foreign documents in their multilingual searches on the Web to a much greater extent than on Korean documents in their research. Some (*n* = 6; 21.4%) even alluded to the fact that all the documents they read for their research were foreign in origin. The term foreign documents most often referred to documents written in English. Doctoral students were more likely to read Korean than foreign documents.

Subjects turned to foreign documents due to the amount of information, ease of access, and quality of information. Six subjects (21.4%) specifically mentioned that the volume of Korean documents was smaller than that of foreign documents. Some subjects mentioned that the Internet has tremendously changed the ways in which they look for information during their research process, pointing out advantages including obtaining in-progress papers (Faculty 16) and saving time (Doctoral Student 05).

The subjects perceived that the overall value of information in foreign documents was as follows:

- better (Faculty 07);
- higher quality (Faculty 05, Faculty 10);
- more credible (Faculty 02);
- more professional (Doctoral Student 01, Doctoral Student 05);
- more current (Doctoral Student 01, Doctoral Student 03, Faculty 17); and
- more advanced (Faculty 06, Doctoral Student 04, Faculty 09, Faculty 20).

While the subjects concentrated on looking for foreign documents for their research, they mentioned that they did not need to "search hard" for Korean documents. Unlike with foreign documents, these scholars tended to obtain the documents in Korean by subscribing to journals, by asking their colleagues, or by directly contacting the researchers in the fields.

The following comments illustrate subjects' perceptions of foreign documents:

> There is more information available in foreign documents than in Korean documents. Korean documents tend to be less professional. In addition, foreign Web sites provide numerous links which lead me to another set of useful information while Korean sites do not provide enough links (Doctoral Student 05).
>
> I am searching for only foreign language documents and rarely search for Korean documents because there are relatively small numbers of people who are doing the research in my area, so I know who is doing what in Korea (Faculty 18).
>
> When I need new ideas, I don't search in Korean documents in general. That's because I can find more references about novel research from English documents. There is less information, and, accordingly, fewer novel ideas in Korean documents (Faculty 08).

While subjects showed strong preferences for foreign documents for their research, they did not show similar patterns for other kinds of tasks. By request, each subject during the interview searched the Web for two types of information (research and personal topics) in three search engines: one chosen by the subject and the other two given by the interviewer including Mochanni (http://www.mochani.com) and KINDS (http://www.kinds.or.kr) sites. The interviewer provided no instructions about the language to be used for subjects' searches. Of the 25 subjects who demonstrated the search process for their research project, 23 entered their search queries in English and only two in Korean. The subjects who entered English terms chose English as their query language because they wanted to retrieve the documents in English; in contrast, when they conducted searches for personal topics, all of them entered search queries in Korean and wanted to retrieve documents in Korean only.

4.2. Selection of search engines for multilingual searches

The second research question examined whether the language feature offered in search engines would make any difference in scholars' choice of search engines. This question also attempted to identify other criteria that influence user choice of search engines.

Two kinds of search engines are available in Korea: those engines developed in Korea (e.g., Naver, Empas, Simmani) and those developed in other countries, mostly in the United States, in Korean language versions (e.g., Yahoo! Korea, Altavista Korea, and Google Korea). The major difference between these two kinds of search engines is that most foreign-brand search engines have multilingual search capabilities, while Korean-brand engines offer only monolingual search in Korean. Both Altavista and Google, for example, have search technology to support searching in Korean and provide a search feature enabling users to choose their search language(s). In the case of Google (http://www.google.co.kr), users can choose either "All Languages" or "Korean Language." Altavista (http://kr.altavista.com) users also can choose "All Languages" or "Korean" for their search results if they wish to search only for pages written in these language(s). Yahoo! Korea (http://kr.yahoo.com), on the other hand, does not offer a choice of languages, and users can retrieve the documents in Korean regardless of the language they employed in their query.

The results reveal that the language feature offered in foreign-brand search engines failed to influence the subjects' choice of search engines. Rather, the subjects used multilingual search engines as a monolingual search tool, completely ignoring the capability of multilingual searching. Most subjects identified two favorite search engines, separating "foreign search engines" from "Korean search engines." That is, when they searched for foreign language documents, they went to the U.S. version of search engines (e.g., http://www.google.com; http://www.lycos.com; http://www.yahoo.com) rather than to the Korean version of foreign search engines (e.g., http://www.google.co.kr; http://www.lycos.co.kr; http://kr.yahoo.com) although most of these Korean versions of foreign search engines (except Yahoo! Korea) offered multilingual search tools. When they intended to search for Korean documents, they chose Korean search engines. Rather than the language-selection feature, the subjects identified three other major criteria for selecting a particular search engine: familiarity (*n* = 10; 35.8%), satisfaction with search results (*n* = 7; 25%), and functionality (*n* = 3; 10.7%). Others included the use of a search engine for the subject's personal e-mail account and word of mouth. Three subjects said they did not choose the engine to search; rather, they used the default engine given by a Web browser or an initial homepage set up by other members of the family.

Table 2 presents the criteria and reasons for choosing search engines. No subject mentioned the feature of language selection as a criterion, as the following comments illustrate:

> It is just my habit to use it. It is the most popular one, and it is the very first search engine that came out (Faculty 01).
> With this one, I can search for more diverse things than with any other search engines available... In other words, diversity is more important than the search results themselves. When I enter [major league baseball player] 'Chanho Park,' I can get images, audio, and other multimedia results, as well as relevant news articles (Doctoral Student 02).
> I use this one often... I began to use it when I heard that it used natural language retrieval, and I like it a lot. I started to use it before it became so popular. It is better than other engines in terms of features. It even indicates redundant items among the results (Faculty 08).

Some subjects reported that they rarely used search engines, so they were not able to specify reasons for selection. One subject (Faculty 13) said that when he wanted to find personal information, such as for travel, he tended to go to newspaper sites. Another

Table 2
Major criteria for choosing search engines

Criteria	Reason	Subject number
Familiarity (*n* = 10; 35.8%)	The first engine used	Faculty 01
		Doctoral student 01
	Most familiar	Faculty 06
	Most popular and famous	Faculty 05
		Faculty 19
	Once started using it, continue to use	Doctoral student 03
		Faculty 20
	Use it for a long time	Faculty 07
	First engine developed	Faculty 11
		Faculty 18
Satisfaction with search results (*n* = 7; 25%)	Better results	Post-doc fellow 01
		Doctoral student 07
	More information	Faculty 03
	More recent information	Doctoral student 06
	Results are fine	Faculty 09
		Faculty 17
	Shows results from foreign Web sites as well	Doctoral student 04
Functionality (*n* = 3; 10.7%)	Availability of image search	Doctoral student 02
	Capability of natural language retrieval	Faculty 08
	Capability of entering queries as sentences	Faculty 12
Others (*n* = 7; 25%)	Word of mouth	Faculty 16
		Doctoral student 05
	Has personal e-mail account on search engine site	Faculty 14
		Faculty 10
	Default engine given by Web browser or initial homepage set up by other (not personal choice)	Faculty 02
		Faculty 15
		Faculty 18

(Faculty 04) explained that he had never become used to conducting searches on search engines and used professional organizations' Web sites most often. When subjects were asked the problems associated with using search engines, they sometimes described frustrating experiences, making numerous comments as to how search engines could be improved. Most comments were centered around the difficulty of coming up with proper keywords. Some subjects mentioned that they wanted a more structured cataloging system on the Web (Faculty 14) or a term suggestion feature (Faculty 18) because it was difficult to find information only by entering keywords. However, no subject cited problems with language selection or difficulties with searching for foreign documents.

4.3. User preferences for integrated multilingual searches

With respect to the third research question, user needs and requirements for multilingual search tools were investigated. Here subjects were asked whether they would like to retrieve integrated results of foreign and Korean documents when they entered their search query in either language. The responses were split between preferring an integrated multilingual search feature (*n* = 14; 50%) and not preferring such a feature (*n* = 11; 39.3%). For subjects who preferred integrated multilingual search results, retrieving "more results" (*n* = 4; 14.3%) and getting the "results easily" (*n* = 10; 35.7%) were the primary reasons for their preference. Subjects who did not want the multilingual search feature most often mentioned "too much information" (*n* = 5; 17.9%) and "time consuming" (*n* = 3; 10.7%). In addition, at least eight subjects (28.6%) explicitly expressed doubts about the quality of "machine translation" techniques, especially for scholarly terminologies. Given these doubts, they preferred to enter search queries for their research in English, which prevails in their academic domains. Even for the language of retrieved documents, the subjects preferred foreign documents (mostly English) to Korean ones. Thus, there were low expectations for multilingual searches as far as the research information was concerned. The following quote shows the example of the subjects' concerns with the translation.

> I would prefer to search in Korean and in foreign languages separately. That is because I am not sure whether the translation will be done accurately. In fact, scholarly terminology has specific meaning, which sometimes could be different from a dictionary-based definition for the word. Therefore, I would be suspicious as to how well the search results match with what I look for (Faculty 04).

On the other hand, subjects who responded that they preferred integrated multilingual searches expressed the desire to have control over language selection, not only the language for the results but also for a display format in which they could decide whether they wanted to see multiple language documents on the same page or on separate pages as seen in the following comments:

> I think that searching in multiple languages could be good because I can compare the results. Of course, it is always a problem as to how many valuable documents will be retrieved. I would have trouble if too many documents are retrieved. It would be very difficult for me if I have to organize the results. It would be good if the results were well organized (Faculty 06).
>
> I think that a multiple language search should show the results separately for each language. If the results are all together, it would be problematic. It should allow users to choose display options for search results. Probably the biggest problem would be how to express English words in Korean characters. As there is no standard for this, it would pose difficult problems (Faculty 18).

5. Discussion and Conclusion

The premise of this study was that bilingual people living in multilingual environments would take advantage of various information resources written in multiple languages on the Web. The results reveal that the Korean bilingual academic users who participated in this study used both Korean and English information resources available on the Web; however, in general they did not use Web search engines as multilingual tools. No subject utilized the feature of language selection offered by most foreign-brand search engines (except Yahoo! Korea). Rather, the language was a determining factor when the subjects made their initial selection of search engine: most subjects had two distinct search engines as their favorites—one for Korean documents and another for English. Furthermore, those Korean scholars who were capable of speaking, reading, and writing multiple languages did not conduct multilingual searches across all kinds of information task. Although they used the Web to find a variety of information spanning work-related and personal information problems, they were interested in foreign documents for very limited search topics, mostly related to their research projects. For everything else including teaching, hobbies, sports, and news, they rarely looked for information in foreign languages. While the subjects used foreign search engines for their research interests, they used Korean search engines for their personal information needs.

This study demonstrates the value of studying multilingual search behavior on the Web in natural settings by identifying user needs and preferences for integrated multilingual search systems. An assumption upon which most CLIR research is based is that bilingual users who have reading skills in a second language are still likely to submit their search queries in their first language as they cannot express their information need well enough in their second language (Ogden & Davis, 2000). The results of this study indicate that user choice of language is dependent upon types of search task, rather than familiarity with the language. When searching for information for their research, most of the subjects entered queries in English (second language), rather than in Korean (first language). This finding is consistent with that of Petrelli et al. (2004), who reported that users chose the most appropriate language for their task, one that was not necessarily their native language. The reasons for choice of language appear to stem from two causes. One is that the subjects believed that English queries would enable them to specify scholarly terms accurately; the other is that they intended to obtain English documents because they believed that documents published in English were more current, better, novel, and more credible than those published in Korean.

One of the major concerns expressed with respect to multilingual search systems was the accuracy of machine translations of scholarly terminologies. Some subjects pointed out that the dictionary-based definitions of certain words sometimes mean something totally different from their scholarly usage. This implies that it is important for CLIR systems to gain the trust of users with respect to automatic translation. To gain such trust, interaction with the functionality of language seems to be key. The subjects expressed their desire to engage in the search process actively and to be able to choose their preferences depending on their search task and familiarity with the language in the domain. Most of the subjects in the Clarity project (Petrelli et al., 2004) were not interested in seeing how the system was translating the query; concerns were focused on search outcomes. However, their subjects were more likely to be motivated to reformulate their query when asked to see the query translation process.

When subjects were asked about desirable features of integrated multilingual searches, they preferred to have user control over multilingual parts of their searches both when entering search queries and when obtaining search results. Many subjects emphasized that they wanted to exercise control in the presentation of search results by switching between a separated results list and an integrated list, since they considered efficiency to be an important success factor in Web searching. Many users did not like the idea of integrating the results simply because it would present too many results and it would be time-consuming for them to review. Oard (2003) pointed out that the difference between monolingual and cross-language retrieval is that it is easy to see how a ranked list can be used in a monolingual process, but an equally good list may be of no use at all to a cross-language user. The findings of this study indicate that a simple ranked list in current Web search engines may not be a useful format for multilingual Web searches.

Future studies in CLIR should focus on the search behavior of users in the entire process of system development from the identification of user requirements to the evaluation of the system. No matter how advanced new CLIR techniques may be, users will eventually make a decision to use such functionalities. Thus, it is important to enhance the understanding of multilingual Web searching from the user perspective. In the future, it would be worthwhile to conduct research in controlled experimental settings for the purpose of observing how the selection of language is related to the effectiveness of search results as well as user perception of search success. The subjects in this study were bilingual Web users who were sufficiently competent to enter search terms and to understand Web content written in English. Future studies can examine user preferences and search behaviors that depend on foreign language ability by studying Web users who have a low proficiency in foreign language skills but who have an interest in finding documents written in foreign languages.

References

Applebee, A., Clayton, P., & Pascoe, C. (1997). Internet and academic work in Australian universities: A quantitative study. *Internet Research: Electronic Networking Applications and Policy*, *7*, 67–74.

Ballesteros, L., & Croft, W. B. (1996). Dictionary-based methods for cross-lingual information retrieval. In R. Wagner, & H. Thoma (Eds.), *Proceedings of the 7th International DEXA Conference on Database and Expert Systems Applications* (pp. 791–801). New York: Springer-Verlag.

Ballesteros, L., & Croft, W. B. (1998). Resolving ambiguity for cross-language retrieval. In W. B. Croft, A. Moffat, C. J. van Rijsbergen, R. Wilkinson, & J. Zobel (Eds.), *Proceedings of the 21st annual International ACM S1G1R Conference on Research and Development in Information Retrieval* (pp. 64–71). New York: ACM Press.

Bian, G. W., & Chen, H. H. (2000). Cross-language information access to multilingual collections on the Internet. *Journal of the American Society for Information Science*, *51*, 281–296.

Choo, C. W., Detlor, B., & Turnbull, D. (2000). *Web work: Information seeking and knowledge work on the World Wide Web*. Boston: Kluwer Academic Publishers.

Davis, M. (1997). New experiments in cross-language text retrieval at NMSU's computing research lab. In D. K. Harman, & E. M. Voorhees (Eds.), *Proceedings of the fifth Text Retrieval Conference (TREC-5)* (pp. 447–454). Gaithersburg, MD: Department of Commerce, National Institute of Standards and Technology.

Davis, M., & Dunning, T. (1996). A TREC evaluation of query translation methods for multi-lingual text retrieval. In D. K. Harman (Ed.), *Proceedings of the fifth Text Retrieval Conference (TREC-4)* (pp. 483–498). Gaithersburg, MD: Department of Commerce, National Institute of Standards and Technology.

Dumais, S. T., Letsche, T. A., Littman, M. L., & Landauer, T. K. (1997). Automatic cross-language retrieval using latent semantic indexing. In D. Hull, & D. Oard (Eds.), *Working notes of the AAAI-97 spring symposium technical report on cross-language text and speech retrieval* (pp. 18–24). Menlo Park, CA: AAAI.

Ellis, D., Cox, D., & Hall, K. (1993). A comparison of the information seeking patterns of researchers in the physical and social sciences. *Journal of Documentation*, *49*, 356–369.

Global Reach. (2004). *Global internet statistics by language*. San Francisco, CA: Global Reach. Retrieved June 1, 2004 from http://www.global-reach.biz/globstats

Gollins, T., & Sanderson, M. (2001). Improving cross language information retrieval with triangulated translation. In W. B. Croft, D. J. Harper, D. H. Kraft, & J. Zobel (Eds.), *Proceedings of the 24th annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 90–95). New York: ACM Press.

Hansen, P., Petrelli, D., Karlgren, J., Beaulieu, M., & Sanderson, M. (2002). User-centered interface design for cross-language information retrieval. *Proceedings of the twenty-fifth annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 383–384). New York: ACM Press.

Hargittai, E. (2002). Beyond logs and surveys: In-depth measures of people's Web use skills. *Journal of the American Society for Information Science and Technology*, *53*, 1239–1244.

Hsieh-Yee, I. (2001). Research on Web search behavior. *Library and Information Science Research*, *23*, 167–185.

Hull, D. A., & Grefenstette, G. (1996). Querying across languages: A dictionary-based approach to multilingual information retrieval. *Proceedings of the 19th annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 49–57). New York: ACM Press.

Iivonen, M., & White, M. D. (2001). The choice of initial Web search strategies: A comparison between Finnish and American searchers. *Journal of Documentation*, *57*, 465–491.

Karlgren, J., & Hansen, P. (2002). Cross-language relevance assessment and task context. *Proceedings of the CLEF workshop.* Retrieved November 15, 2002 from http://www.dcs.shef.ac.uk/research/groups/nlp/clarity/papers/iclef02-sics.pdf

Kim, K.-S. (2002). Information seeking on the Web: Effects of user and task variables. *Library and Information Science Research*, *23*, 233–255.

Krathwohl, D. R. (1993). *Methods of educational and social science research: An integrated approach*. White Plains, NY: Longman.

Large, A., & Moukdad, H. (2001). Multilingual access to Web resources: An overview. *Program*, *34*, 43–58.

Lavrenko, V., Choquette, M., & Croft, W. B. (2002). Cross-lingual relevance models. In M. Beaulieu, R. Baeza-Yates, S. H. Myaeng, & K. Järvelin (Eds.), *Proceedings of the 25th annual International ACM-SIGIR Conference on Research and Development in Information Retrieval* (pp. 175–182). New York: ACM Press.

Lazinger, S. S., Bar-Ilan, J., & Peritz, B. C. (1997). Internet use by faculty members in various disciplines: A comparative case study. *Journal of the American Society for Information Science*, *48*, 508–518.

Maeda, A., Sadat, F., Yoshikawa, M., & Uemura, S. (2000). Query term disambiguation for Web cross-language information retrieval using a search engine. In K. F. Wong, D. L. Lee, & J. H. Lee (Eds.), *Proceedings of the 5th International Workshop on Information Retrieval with Asian Languages (IRAL '00)* (pp. 25–32). New York: ACM Press.

Oard, D. W. (1997, December). Serving users in many languages: Cross-language information retrieval for digital libraries. *D-Lib Magazine*. Retrieved November 15, 2002 from http://www.dlib.org/dlib/december97/oard/12oard.html

Oard, D. W. (2003). When you come to a fork in the road, take it: Multiple futures for CLIR research. *Cross language information retrieval: A research roadmap. Workshop at the 22nd International Conference on Research and Development in Information Retrieval.* Retrieved April 1, 2004 from http://ucdata.berkeley.edu/sigir-2002/sigir2002CLIR-03-oard.pdf

Ogden, W. C., & Davis, M. W. (2000). Improving cross-language text retrieval with human interactions. *Proceedings of the 33rd Hawaii International Conferences on System Sciences.* V. 3 Maui, Hawaii. Retrieved November 15, 2002 from http://crl.nmsu.edu/Research/Projects/tipster/ursa/Papers/Hawaii.pdf

Petrelli, D., Beaulieu, M., Sanderson, M., Demetriou, G., Herring, P., & Hansen, P. (2004). Observing users, designing clarity: A case study on the user-centered design of a cross-language information retrieval system. *Journal of the American Society for Information Science and Technology*, *55*, 923–934.

Petrelli, D., Hansen, P., Beaulieu, M., & Sanderson, M. (2002). User requirements elicitation for cross-language information retrieval. In T. D. Wilson, & M. J. Barrulas (Eds.), *Proceedings of the International Conference on Information Needs, Seeking and Use in Different Contexts (ISIC2002).* Retrieved 15, 2002 from http://www.sics.se/~preben/papers/Hansen-Petrelli-ISIC-2002.pdf

Pirkola, A., Hedlund, T., Keskustalo, H., & Järvelin, K. (2001). Methods and problems in dictionary-based cross-language retrieval: Literature review and research at the University of Tampere. *Information Retrieval*, *4*, 209–230.

Rieh, H. Y. (2002). Analysis of search engine use, search behaviors and aptitude by Web users. *Journal of the Korean Library and Information Science Society*, *36*(3), 69–91 (in Korean).

Rieh, S. Y. (2004). On the Web at home: Information seeking and Web searching in the home environment. *Journal of the American Society for Information Science and Technology*, *55*, 743–753.

Rieh, H. Y., & Rieh, S. Y. (2001). Information seeking, evaluation, and use on the Internet: A case study of science and engineering scholars. *Journal of the Korean Society for Information Management*, *18*(4), 163–181 (in Korean).

Rieh, S. Y., & Xie, H. (2001). Patterns and sequences of multiple query reformulations in Web searching: A preliminary study. In E. Aversa, & C. Manley (Eds.), *Proceedings of the 64th Annual Meeting of the American Society for Information Science and Technology*, *vol. 38* (pp. 246–255). Medford, NJ: Information Today.

Sandelin, B., & Sarafoglou, N. (2004). Language and scientific publication statistics: A note. *Language Problems and Language Planning*, *28*(1), 1–10.

Spink, A., & Jansen, B. J. (2004). Web search: Public searching of the Web. Boston: Kluwer Academic Publishers.

Voorbij, H. J. (1999). Searching scientific information on the Internet: A Dutch academic user survey. *Journal of the American Society for Information Science*, *50*, 598–615.

Wang, P., Berry, M., & Yang, Y. (2003). Mining longitudinal Web queries: Trends and patterns. *Journal of the American Society for Information Science and Technology*, *54*, 743–758.

Wang, P., Hawk, W. B., & Tenopir, C. (2000). Users' interaction with World Wide Web resources: An exploratory study using a holistic approach. *Information Processing and Management*, *36*, 229–251.

White, M. D., & Iivonen, M. (2001). Questions as a factor in Web search strategy. *Information Processing and Management*, *37*, 721–740.

Wiberley, S. E. Jr. (1991). Habits of humanists: Scholarly behavior and new information technologies. *Library Hi-Tech*, *9*, 17–21.

Wiberley, S. E., Jr., & Jones, W. G. (1994). Humanists revisited: A longitudinal look at the adoption of information technology. *College & Research Libraries*, *55*, 499–509.

Xu, J., Weischedel, R. M., & Nguyen, C. (2001). Evaluating a probabilistic model for cross-lingual information retrieval. In W. B. Croft, D. J. Harper, D. H. Kraft, & J. Zobel (Eds.), *Proceedings of the 24th annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 105–110). New York: ACM Press.