



A Survey of Users and Their Use of Information in Different Languages

Technical Report
LIS 566 Digital Information Retrieval

Erin Larucci

elarucci@gmail.com



Executive Summary

To: Dr. Jianqiang Wang
From: Erin Larucci
Subject: A Survey of Users and Their Use of Information in Different Languages
Date: December 10, 2012
CC: LIS 566 Digital Information Retrieval

Attached please find citations and abstracts for documents relating to users and how they use information in foreign languages. I used a wide variety of databases and search engines to find relevant documents. These are detailed in an appendix along with search terms and queries used (and their effectiveness) in retrieving documents. The queries in Revision 1 and Revision 3 were the most successful in identifying relevant documents.

There are some articles that are “possibly relevant”; these are indicated with a blue font. Hyperlinks have been included where applicable for your convenience. The section that contains “previously reported documents” is a copy of the Revision 1 query results I sent to you on October 27; I included them here so you have all the documents in one place. (All documents retrieved from the preliminary search have been omitted as they were not relevant.)

Throughout the course of searching, I found the following search terms (and their variants) to be the most helpful in locating relevant documents:

User study

Multilingual information access

I was unfortunately unable to find a good quantity of documents in news or multimedia (there is one full-text news item included). The Revision 2 queries focused on searching for news and multimedia; they were not successful. The majority of documents are therefore scholarly articles. I do not think my search terms were conducive to finding older articles (due to advances in technology, the terminology might not be relevant to older documents); most of the documents retrieved date within the past fifteen years.

Please let me know if there is any way I can further assist you with this project. I would be more than willing to continue working on it if it would be helpful to you.

User Selection

For this project, I wanted to do research in the area of library and information science. When students were notified that some professors from the Graduate School of Education at University at Buffalo would be interested in working with students for the term project, I indicated my interest. As a result, I was paired with Dr. Jianqiang Wang, an Associate Professor and Interim Chair in the Department of Library and Information Studies.

User Needs

Dr. Wang required information relating to his research topic and requested that I find articles, news, web resources, multimedia, conference proceedings, etc. relating to users and their use of information in foreign languages during the past 40 years (though it was noted he had not done much research in the past two years making it important that I find recent documents). The retrieved documents were to be in English, but could originate from any country. Documents that only discussed the technological side of cross-language information retrieval were not to be considered relevant. He preferred the results to be cited and put into a Microsoft Word document to be delivered to him at the end of the term.

- needs all documents relating to topic (high recall)
- documents should cover previous 40 years; past two years very important
- material requested: scholarly articles, news, web resources, multimedia, conference proceedings
- technology-only is not relevant
- “foreign language” is relative (English is a foreign language in France, Germany, Japan, etc.)
- only articles in English are to be provided
- results to be cited (used APA)

User Model

Dr. Wang has extensive experience in the area of research I was asked to investigate. As a result, the request was direct and did not vary over the course of the term. Being an expert in the field, he was helpful in directing the course of the queries by evaluating two sets of results I sent to him - a preliminary search and a revised search. Over the course of the project, I was able to put together a user profile:

- expert in topic, able to quickly evaluate results for relevance
- indicated he will tolerate some irrelevant documents due to need for high recall
- as an Associate Professor and Interim Chair of Department, might not have a lot of free time to spend on task
- able to suggest useful databases, resources
- results to be used for current and future research

Search Strategies

The search strategy I used evolved considerably over the course of the project. Four sets of searches were performed with the first two receiving evaluation.

Preliminary Search

After an email exchange where some clarification was provided, I offered to perform a preliminary search using the terms in the original search task and terms I thought would be relevant. I did this for two reasons: it would help me to further clarify the search task and the evaluation of the results would help to direct future queries. In order to begin searching, I constructed an outline of concepts and their terms.

Concept Outline 1

CONCEPTS	TERMS
Information Retrieval	Information retrieval
Foreign Language	Cross-language Cross-lingual Foreign language Multilingual
Technology	Technique Technology System

For the preliminary search, I chose to use Dialog ClassicWeb, Google, and Google Scholar. I chose these sources because Google and Google Scholar would provide a general feel for the topic and Dialog would allow me to find very specific items (and hopefully provide me with additional terms and/or information).

Dialog

Query 1: **Databases: Library & Information Services**
(information? AND use?) AND (foreign(W)language OR cross-language OR cross(w)language OR cross-lingual OR cross(w)lingual OR multilingual)

Query 2: **Databases: sf allhuman, allnews, allpaper, allscience, allsoc, allsochu**
(information(w)retriev? and use?) and (foreign(w)language or cross-language or cross(w)language or cross-lingual or cross(w)lingual or multilingual) and (system? or technolog? or technique?)

Query 3: **Databases: sf allsochu, allscience**
information(w)retriev? and (usage OR use?) and (foreign(w)language OR cross-
language OR cross(w)language OR cross-lingual OR cross(W)lingual OR
multilingual) and (system? OR technolog? OR technique?)

Google

Query 1: (“information retrieval” AND (use OR usage OR user)) AND (“foreign language”
OR foreign OR multilingual OR cross-lingual OR “cross language” OR “cross
lingual”) AND (system OR technology OR technological OR technique)

Google Scholar

Query 1: (“information retrieval” AND (use OR usage OR user)) AND (“foreign language”
OR foreign OR multilingual OR cross-lingual OR “cross language” OR “cross
lingual”) AND (system OR technology OR technological OR technique)

Query 2: ((“information retrieval” OR information OR “use of information”) AND (use OR
usage OR user)) AND (“foreign language” OR foreign OR multilingual OR cross-
lingual OR “cross language” OR “cross lingual”) AND (system OR technology
OR technological OR technique)

The queries did not provide relevant results for the following reasons: 1) Most of the articles retrieved had already been retrieved by the user, and 2) the articles focused on technology without discussing users. It became clear to me that I needed to revise not only the queries but the concepts as well.

Revision 1

I requested feedback on the search terms I used in the queries and the user indicated he would review them. I did not receive a response, so I spent some time reviewing the tags from the retrieved articles to see if any of the keywords or index terms could be used to find different results. I also researched technical terms in information science (mainly through Google) relating to the search task. After this, I revised the concept outline and created new queries.

Concept Outline 2

CONCEPTS	TERMS
Information Retrieval	Information retrieval
Foreign Language	Cross-language Cross-lingual Foreign language MLIA Multilingual Multilingual information access
Users	Usage Use User studies

For the revised search, I chose to use Google, University at Buffalo Libraries, Yahoo!, Infomine, YouTube, Web of Science, and the New York Times (through Google). Infomine and Web of Science are databases that were introduced to students through exercises. The federated search at UB Libraries searches all the databases available (but didn't work as well as I thought it would). My aim was to get a wide variety of sources to maximize my chances of retrieval. Overall, this was a successful approach.

Google

Query 1: "multilingual information access"

Query 2: ("multilingual information access" OR "MLIA") AND ("user studies" OR "user study")

UB Libraries

Query 1: "user behavior and evaluation of multilingual information access in digital libraries" [Note: I was trying to find one specific article to find more information, but I got some other results with this]

Yahoo

Query 1: ("user study" or "user studies") and "information retrieval" and (language or lingual or multilingual)

Infomine

Query 1: |user stud*| and |information retrieval| and (language or lingu*) [no results]

Query 2: |user stud*| and MLIA [no results]

Query 3: |user stud*| and |multilingual information retrieval| [no results]

Query 4: |user stud*| and |information retrieval| [no results]

YouTube

- Query 1: ("user studies" or "user study") and MLIA
Query 2: ("user studies" or "user study") and (multilingual or language) and "information retrieval"
Query 3: MLIA user studies [no relevant results]

Web of Science

- Query 1: "user stud*" and "information retrieval" and multilingual [no relevant results]
Query 2: "user stud*" and "information access" and multilingual [no results]
Query 3: "user stud*" and "information access" and language [1 result, possibly relevant]

NY Times (searched through Google)

- Query 1: MLIA site:www.nytimes.com [no results]
Query 2: multilingual information retrieval site:www.nytimes.com [no relevant results]
Query 3: multilingual information access site:www.nytimes.com
Query 4: digital multilingual information access site:www.nytimes.com

The user indicated the new results were much better and all the retrieved articles were relevant as well as a web resource; two more results (a video and news article) were not relevant. I provided the citations and abstracts for the ten relevant items. I discovered that "user studies" (and variations of that term) was key to finding relevant documents. "Multilingual information access" (MLIA) was also helpful. I employed these terms throughout the rest of my searching.

Revision 2

The majority of documents I retrieved were scholarly articles, which was within the scope of the assignment, but the user also wanted multimedia and news. I had a great deal of difficulty in finding relevant items (the ones found were deemed not relevant by the user.) After more reviewing and suggestions, I added more search terms to the outline.

Concept Outline 3

CONCEPTS	TERMS
Information Retrieval	Information retrieval Search behavior
Foreign Language	Cross-language Cross-lingual Foreign language MLIA Multilingual Multilingual information access
Users	Usage Use User behavior User experience User studies User study

Following the suggestion to use non-technical terms to search for multimedia, I made a concerted effort to locate news and multimedia. For this task, I chose LexisNexis, NY Times (through Google as I could not access it through LexisNexis; NY Times was suggested by the user), Google Videos, YouTube, Google News, and Science Magazine (suggested by the user).

LexisNexis

Query 1: (“information retrieval!” OR “information access”) AND (multilingual OR MLIA)
AND (“user stud!” OR “user behavior” OR “search behavior”)

[attempted to access LexisNexis on four occasions - news search would not work properly]

NY Times (through Google)

Query 1: (“information retrieval” OR “information access”) AND (multilingual OR MLIA)
AND (“user study” OR “user studies” OR “user behavior” OR “search behavior”)
site:nytimes.com [no results]

Query 2: multilingual “information seeking” site:nytimes.com

Query 3: multilingual searching digital site:nytimes.com [no relevant results]

Query 4: multilingual online searching site:nytimes.com [1 relevant result]

Google Videos

Query 1 (“information retrieval” OR “information access”) AND (multilingual OR MLIA OR “multilingual information access”) AND (“user studies” OR “user study” OR “user behavior” OR “search behavior”) [no relevant results]

Query 2: (“information retrieval” OR “information access”) AND (multilingual OR MLIA OR “multilingual information access”) AND (“user studies” OR “user study” OR “user behavior” OR “search behavior”) -site:.biz [no relevant results]

Query 3: (“information retrieval” OR “information access”) AND (multilingual OR MLIA OR “multilingual information access”) AND (“user studies” OR “user study” OR “user behavior” OR “search behavior”) -site:.biz -site:7strategy.com [no relevant results]

YouTube

Query 1: (“information retrieval” OR “information access”) AND (multilingual OR MLIA OR “multilingual information access”) AND (“user studies” OR “user study” OR “user behavior” OR “search behavior”) [no relevant results]

Query 2: (multilingual OR MLIA OR “multilingual information access”) AND (“user studies” OR “user study” OR “user behavior” OR “search behavior” OR usage) [no relevant results]

Query 3: “digital information retrieval” and user and multilingual [no relevant results]

Google News

Query 1: (“information retrieval” OR “information access”) AND (multilingual OR MLIA OR “multilingual information access”) AND (“user studies” OR “user study” OR “user behavior” OR “search behavior”) [no results]

Query 2: (“information retrieval” OR “information access”) AND (“foreign language” OR foreign OR multilingual OR cross-lingual OR “cross language” OR “cross lingual”) AND (“user studies” OR “user study” OR “user behavior” OR “search behavior”) [no results]

Query 3: Information multilingual user study [no relevant results; new term: “user experience”]

Query 4: Multilanguage information retrieval “user experience” [“Multilanguage” not recognized by Google - search engine substituted “multilingual” instead; no results]

Science Magazine

Query 1: “information retriev*” and (“user stud*” or “user behavior” or “user experience” or “search behavior”) and (multilingual or MLIA or “multilingual information access”) [no results]

Query 2: “information retriev*” and user* and (multilingual or MLIA) [no results]

Query 3: “information retriev*” and multilingual [no relevant results]

Query 4: Multilingual and use* [no relevant results]

Query 5: use* and (multilingual or MLIA) and information [no relevant results]

[Realized that, even though the “help” section declared Boolean searching was possible, many irrelevant search results came back due to “or” and “and” being tagged]

Query 6: use* multilingual MLIA information [no results]

Query 7: multilingual information use* [no relevant results]

I became very frustrated with my lack of progress in the area of news and multimedia. Documents that *were* retrieved were wildly irrelevant. In a few cases, even though I constructed queries with proper truncation symbols and Boolean operators, the search engines wouldn't recognize that (see Science Magazine note above).

Revision 3

Seeing as how I had already spent most of my time on the search task attempting to retrieve news and multimedia, I decided to move on to scholarly articles in order to provide valuable information to my user. This proved to be far more productive. The concept outline was expanded a few more times through subsequent queries. The final outline:

Concept Outline 4

CONCEPTS	TERMS
Information Retrieval	Information retrieval Information-seeking Information-seeking behavior Search behavior Searching
Foreign Language	CLIR Cross-language Cross-lingual Foreign language Languages Linguistics MLIA Multilanguage Multilingual Multilingual information access Multilingualism
Users	Usage Use User behavior User experience User studies User study

In order to find relevant scholarly articles, I chose the following databases: Library, Information Science, and Technology Abstracts (LISTA), Association for Computing Machinery (ACM) Digital Library (suggested by the user), and Dialog ClassicWeb. For the search in Dialog, I used the Library & Information Services group of databases that consists of ERIC, Inspec, National Technical Information Service (NTIS), Social SciSearch, Dissertation Abstracts Online, Gale Group Magazine Database, British Education Index, Gale Group Trade & Industry Database, and Library Literature and Information Science.

LISTA

Query 1: (“information retriev?” OR “information access”) AND (multilingual OR MLIA) AND (“user stud?” OR “user behavior” OR “search behavior” OR “user experience”) [no results]

- Query 2: “cross-language information retrieval” and (“user stud?” OR “user behavior” OR “search behavior” OR “user experience”) [1 relevant result - previously retrieved in Search Revision 1] [Got term “information-seeking behavior”]
- Query 3: (“cross-language information retrieval” or multilingualism) AND “information-seeking behavior” [1 relevant result]
- Query 4: “information-seeking behavior” and (MLIA or “multilingual information access”) [no results]
- Query 5: (“user stud?” OR “user behavior” OR “search behavior” OR “user experience” OR “information seeking”) AND multilingual [2 relevant results; 1 possibly relevant result]
- Query 6: (“use? stud?” OR “use? behavior” OR “search behavior” OR searching) AND (language? OR linguistics) [1 relevant result]
- Query 7: (“use? stud?” OR “use? behavior” OR “search behavior” OR searching) AND (multilanguage? OR multiling?) [no results]
- Query 8: (“use? stud?” OR “use? behavior” OR “search behavior” OR searching) AND “cross-language?” [no results]
- Query 9: cross-language search? [4 relevant results, 1 possibly relevant result; got term: CLIR]
- Query 10: CLIR AND use? [retrieved same documents as Q9 with 1 additional possibly relevant result]
- [Publication from 1976 to 2009, but nothing older than 1999 was retrieved.]

ACM Digital Library

- Query 1: (CLIR OR “cross-language” OR MLIA OR multilingual OR multilanguage) AND (use* OR behavior OR search) [no results]
- Query 2: CLIR AND use* [retrieved many of the same documents as Q9/Q10 in LISTA]
- Query 3: (multilingual OR multilanguage) AND use* [no relevant results]
- Query 4: (CLIR OR MLIA) AND (“user stud*” OR “user behavior” OR “search behavior”) [1 relevant result]
- Query 5: “information-seeking” AND CLIR AND use* [2 relevant results]
- Query 6: CLIR and “user study” [1 relevant result; 1 possibly relevant result]

DIALOG

Query 1: (information-seeking OR search(w)behavior OR searching) AND (user(w)stud?
OR user(w)behavior OR user(w)experience OR use?) AND (CLIR OR MLIA OR
cross-language OR multilingual OR multilanguage OR
multilingual(w)information(w)access) [no relevant results]

Search 1: Library & Information Services databases
Expand “MLIA”
E3 = MLIA [43 results]
s e3 and use? [29 results]
rd [28 results]
[1 relevant result]

Search 2: Library & Information Services databases
Expand “CLIR”
E3 = CLIR [653 results]
s e3 and use? [441 results]
rd [376 results]

[Discovered that CLIR = Council on Library and Information Resources]

s s4 not CS:council on library and information resources [8 results]
[1 relevant result]

Search 3: Library & Information Services databases
Expand “multilingualism”
E3 = multilingualism [4312 results]
E8 = multilinguality [138 results]
s e3 and e8 and information(w)retriev? and use? [1 result]
[no relevant results]

Search 4: Library & Information Services databases
Expand “multilanguage”
E3 = multilanguage [1264 results]
s e3 and information(w)retriev? and use? [38 results]
rd [36 results]
[1 relevant result; 2 possibly relevant results]

Search 5: Library & Information Services databases
Expand “search(w)behavior” [no results]

Search 6: Library & Information Services databases
Expand “information-seeking”
E3 = information-seeking [557 results]

E10 = information-seeking behavior [295 results]
s (e3 or e10) and multilingual? and information(w)retriev? [1 result]
[no relevant results]

Search 7: E17 = information-seeking behaviour [70 results]
E36 = information-seeking habits [6 results]
E49 = information-seeking patterns [10 results]
s (e3 or e10 or e17 or e36 or e49) and (multilanguage OR multilingual? OR
MLIA OR CLIR OR cross-language OR cross(w)language) [3 results]
[1 relevant result - previously retrieved]

Search 8: Library & Information Services databases
Expand “user(w)studies” [no results]

Search 9: Library & Information Services databases
Expand “user(w)behavior” [no results]

Search 10: Library & Information Services databases
Expand “searching”
E9 = searching behavior [119 results]
s e9 and (multilanguage OR multilingual? OR MLIA OR CLIR OR cross-
language OR cross(w)language) [2 results]
[no relevant results]

s e9 and multi(w)language [no results]
s e9 and language [5 results]
[no relevant results]

These queries retrieved the bulk of the documents I provided to Dr. Wang, though it was still less than I would have liked. In all, 34 relevant documents (including a handful of possibly relevant documents) were retrieved, consisting mainly of scholarly articles, conference proceedings, one news item and one web resource.

Presentation of Results

The user report was delivered as a Microsoft Word document, as requested by the user. Results were cited using APA citation style. Though abstracts were not requested by the user, I included them with citations in order to provide more detailed information than what a citation could give. I also included hyperlinks for convenient retrieval.

I included results I considered relevant (by reading the abstracts) and a few possibly relevant documents (since Dr. Wang indicated he would allow some non-relevant articles). I colored the “possibly relevant” articles in a blue font to distinguish them from the relevant articles. A couple of otherwise relevant articles were omitted from the report as they were co-authored by Dr. Wang (these were useful as sources for additional search terms).

At the end of the report, I included two appendices: one detailing the concept outline and another detailing the queries and their effectiveness. Though this would probably be more useful to a lay person than a professional, it may still provide additional information.

User Evaluation

Since my user had a great deal of knowledge of the research topic, a lengthy reference interview was not necessary. The search task was very specific and clearly stated. Communication was via email messages in which results were given and questions asked. Some of these were not answered. Understanding that Dr. Wang was probably pressed for time and could not review everything immediately, I took it upon myself to research terminology, using keywords and index terms from search results and information science resources to identify possible search terms.

Review of Process

Considering my inability to locate relevant news articles and multimedia documents, I should have been more clear about my need for help with this. Also, I did not realize until late that, although the search topic was for documents within the past 40 years, no relevant document was dated earlier than 1995. A possible explanation for this is that the area of computer science has evolved rapidly; older documents may be indexed with dated terminology that is no longer used. I should have asked if there were older terms I could use or researched these on my own.

I also could have asked what has already been searched; I didn't ask this mainly because it was stated that research had been done on this topic two years ago, but not much since. However, it may have helped a great deal to find older articles.

I did not ask how many documents were expected as it was stated the user wanted all relevant documents. However, I should have asked anyway as I thought there should have been many more than I actually found.

A surprising thing I discovered was that natural language searching in Revision 3 seemed to turn up more relevant results than carefully constructed Boolean queries, even in sophisticated databases where such queries are considered essential for locating information. I had not anticipated this and it caused me to wonder if I was overthinking the search task.

I used a wide variety of sources in order to ensure high recall. Some overlap was expected; I found myself going back and forth to make sure I had not already retrieved a document from a previous database. This became easier over time as I grew more familiar with the content. The databases/search engines I chose were based on personal experience (Google, Google News, Google Scholar, Google Videos, Yahoo!, YouTube, LISTA, UB Libraries), suggestions by the user (Dialog, ACM Digital Library, NY Times, Science Magazine), and databases introduced during the course of LIS 566 Digital Information Retrieval (Infomine, Web of Science, LexisNexis). I decided on these due to breadth of content, relationship to information science, and authority.

An essential part of the search strategy evolution was the technique of “pearl-growing” as it helped me identify useful search terms over the course of the project. Synonyms of these terms, combined with “OR” Boolean statements, helped to produce high recall in the results. With further searching, I have no doubt more terms could be identified.

Overall, I enjoyed the challenge this task presented. I imagine more time would need to be devoted to it in order to be able to retrieve all the relevant documents. I have offered my assistance if Dr. Wang wishes to continue the project.