Online Searching in English as a Foreign Language

Gyöngyi Rózsa
Budapest University of Technology
and Economics
Műegyetem rkp. 3.
1111 Budapest, Hungary
+36 308595248
rozsa.gyongyi.1990@gmail.com

Anita Komlodi
University of Maryland, Baltimore
County
1000 Hilltop Circle
Baltimore, MD 21250, USA
+1 410 455 3212
komlodi@umbc.edu

Peng Chu
University of Maryland, Baltimore
County
1000 Hilltop Circle
Baltimore, MD 21250, USA
+1 410 455 3212
pengchu1@umbc.edu

ABSTRACT

Online searching is a central element of internet users' information behaviors. Searching is usually executed in a user's native language, but searching in English as a foreign language is often necessitated by the lack of content in languages that are underrepresented in Web content. This paper reports results from a study of searching in English as a foreign language and aims at understanding this particular group of users' behaviors. Searchers whose native language is not English may have to resort to queries in English in support of their information needs due to the lack or low quality of the web content in their own language. However, when searching for information in a foreign language, users face a unique set of challenges that are not present for native language searching. We studied this problem through qualitative research methods and report results from focus groups in this paper. The results reported in this paper describe typical problems foreign language searchers face, the differences in information-seeking behavior in English and in the participants' native language, and advice and ideas shared by the focus group participants about how to search effectively and efficiently in English.

Categories and Subject Descriptors

H.1.2 User/Machine Systems; Human factors; Human information processing

General Terms

Design, Human Factors.

Keywords

Information-seeking behavior; Search; Foreign Language

1. INTRODUCTION

Important inequalities in the availability of content in users' first languages exist on the Internet. While native English speakers make up 26.8% of all Internet users (Internet World Statistics, 2012), English language content makes up 56.6% of all content on the Web (W3Techs, 2012). In contrast, Chinese first language speakers make up 24.2% of Web users (Internet World Statistics, 2012), while Chinese language content accounts for 4.5% of Web

content (W3Techs, 2012). This scarcity of content in other languages requires non-native English users to search in English to find content that is not available in their language, or content that is of better quality in English than in their native language.

When searching in a foreign language for content that was created in an unfamiliar culture, searchers face a myriad of problems: they may speak the language but not grasp the slight variations in meaning that will change their search results drastically; they may not be aware of trustworthy sources; and they may face unfamiliar information architecture and design styles in websites.

This paper presents results from a study of searching in English as a foreign language. The results reported in this paper were generated from focus group interviews. The analysis of the recordings and transcripts of focus groups was completed to identify the typical problems faced by foreign language searchers, the differences in information-seeking behavior in English and in the participants' own native language, and the advice and ideas shared by the focus group participants about searching in English.

2. RELATED RESEARCH

It has been established by many study findings (e.g. Kralisch and Mandl, 2006; Berendt and Kralisch, 2009; Steichen et al., 2014) that performing online searching in foreign languages is an inseparable part of internet users' information behaviors if the content in the user's native language is limited. The following three themes are especially central and have been studied in second language searching:

(1) Reasons for searching in a foreign language: Content in a specific language
One of the most important reasons is the limited availability and perceived low quality of content in underrepresented languages (Kralisch and Mandl, 2006; Aula and Kellar, 2009). Kralisch and Mandl (2006) found that the number of web-hosts per language is determined by the number of native speaker internet users of that language. Thus, there are fewer websites authored in languages spoken by smaller groups of people. Thus, English dominates Web content and speakers of other languages will, by necessity search in English and other well represented languages. Berendt and Kralisch (2009) even argue that underrepresentation of non-English languages is further strengthened by users' tendency to accept information in English, which in turn strengthens this reason.

(2) The role of language proficiency
Searchers' language skills impact their language selection, usage, searching performance and preference (Berendt and Kralisch, 2009). Marlow et al. (2008) also found a strong impact of language proficiency on how users utilized query formulation support tools. Steichen et al. (2014) argued that polyglots would

use any language to search as long as they possess sufficient proficiency in it.

(3) Other contextual factors

A series of contextual factors were analyzed and proven to make a difference for multilingual searcher language selection and usage: domain knowledge (Berendt and Kralisch, 2009), field of knowledge (Clough and Eleta, 2010), usage purpose, and topic domain (Steichen, 2014).

These study findings shed light on multilingual searchers' behaviors. However, the major body of current research usually focuses on "why" of this behavior and the impact of isolated user characteristics, rather than on studying the "how" of the process. In other words, research in this area is lacking data on how searchers actually search in English as a foreign language, the challenges they face, and the potential implications and solutions to improve the situation. This paper is motivated by the findings of previous research that indicate that challenges exist in foreign language searching (Chu et al., 2012) but takes the goals of those studies one step further by identifying the nature of the behavior and the challenges faced by foreign language searchers.

3. METHODOLOGY

Qualitative methods were selected in order to truly understand the roots of searching problems and not simply to observe their occurrence. Qualitative methods allow the researchers to immerse themselves in the experience of the participants and get a uniquely intimate understanding of the participants' perceptions and feelings. The results presented in this study are derived from focus groups conducted with searchers.

Four focus groups were conducted with five Hungarian college students in each group. The focus groups were conducted in Budapest, Hungary. Hungarian graduate students trained in focus group interviewing techniques led the discussions. Thus, the researchers and the participants were not only speakers of the same language, but they were also students of the same university. The interviewers were selected in order to ensure more openness and better rapport between the participants and the researchers, as well as provide more accurate cultural interpretation of the data. Students were invited to participate in the study for two reasons. First, they all spoke English at a conversational level, as this was a requirement for their studies. Second, they regularly searched in English both to support their studies and their non-academic interests. These two criteria qualified them for the purposes of this study.

Of our twenty participants, seven were male and 13 female. Their average age was 23.05 years (SD=1.3). Participants filled out a demographic questionnaire including information about their cultural background, computer, and searching expertise. Prior to the meeting they were also asked to list five topics they regularly searched in English and in Hungarian including information about where and how they searched for information about these topics, how often they searched, and why they chose that particular language. These topics were later discussed in the focus groups discussion. Questions posed to the participants in the focus groups concerned participants' English language search behavior, challenges they faced, and advice and recommendations for other searchers and systems designers. Specifically, first the previously listed topics were discussed followed by questions:

1. What search engines or websites do you use when searching and browsing in English?

2. How does searching/browsing in English differ from searching/browsing in Hungarian in your experience?

3. What problems do you find when searching in English?

4. How can websites and search engines help searching in English?

5. What advice would you give to Hungarian students who speak English but they have just started searching for information in English? (e.g. incoming first year university students)

The focus groups were voice recorded and transcribed. The transcriptions were in Hungarian and the coding process was executed in Hungarian. Two of the authors are native Hungarian speakers and carried out this analysis. The transcripts were coded through an inductive and iterative process. In the initial open coding phase, concepts describing language selection, search strategies, challenges, and ideas for system support were identified. Next, axial coding took place to group the concepts and create relationships between them. The method of constant comparison was used as new interviews were added and the more recently emerging concepts and categories were compared to those previously identified. (Corbin et al., 2008; Merriam, 2009). The results were then translated into English along with representative quotes for each concept and category.

4. RESULTS

First we will present the challenges faces by foreign language searchers, as described in the focus groups. These challenges were either directly reported by participants or addressed in advice they would to new foreign-language searchers searching in English.

4.1 The role of native and foreign languages in searching

Participants in our research combined searching in their native language and English as it suited their needs. Searches in the two languages were complementary and often supported the searchers' interpretation of the information found in the other language. Participants decided to search in English to find content that did not exist in Hungarian. While in some cases they searched for topics where they used English from the start (such as foreign music or movies, international news, and travel information) they often mentioned switching to English after searching in Hungarian. Even if they planned to search in English, they often started in Hungarian to get an overview of a topic in their native language, especially in the case of academic searches:

"Depending on the subject, I choose to use English to search, but first I start searching in Hungarian. I check what is available."

In other cases they did not set out to search in English at the beginning of their search sessions, but resorted to it only as complementary to their native language searching. If they could not find enough content in Hungarian, they switched to English. As such, queries in Hungarian and English often followed one another:

"I alternate queries in the two languages. Once I've searched in Hungarian thoroughly and I couldn't find any information or I found only inadequate information, I continue searching in English."

After searching in English, they sometimes switched back to Hungarian to double-check their interpretation of information:

"If we've found something in English maybe afterwards we search for it in Hungarian as well."

"Wikipedia is a good example, after searching for the content in English we usually check the Hungarian article in order to have an overview."

This complementary interaction of the two languages shows that while they do not play the same role, results from each language contribute something unique to the users' understanding of an information problem. Search engines allow for this integrated interaction to a certain extent, but better support could more closely integrate searching in two languages.

4.2 Stress and attention in foreign-language searching

Participants often described themes of attention, cognitive load, and stress associated with foreign-language searching. In general, they did not enjoy the experience of searching in a foreign language and they felt they were unsure of what they were doing:

"I didn't like searching in English for a long time. I know that I understand the content in English but it bothered me. Especially when I search for an unknown topic… it makes me unsure…"

4.3 Short queries, large result sets

4.3.1 Constructing queries

Participants in all four focus groups reported that they used only a few words in English as query terms, whereas in Hungarian, they were more likely to use longer phrases and even sentences. They used fewer words and simpler expressions for searching in English, as these were easier to create than longer, more complex queries.

"In English, we'd rather use keywords. If I enter a whole sentence I may make grammatical or other mistakes so I prefer entering an important keyword and I hope I will find what I seek. In Hungarian, I have a greater chance to find specifically what I am thinking of."

"It's easier to express problems in our native language, maybe this is the reason why we like to use whole sentences."

General, one-word queries also mask searchers' lack of specific vocabulary and insecurities and unawareness about slight variations in meaning. Our participants reported using the most general term describing a concept as they struggled with translating domain-specific terminology:

"There are a wide variety of Hungarian expressions for different meanings which are supposed to exist in English as well but we don't know them."

"Google translate shows 2-3 expressions but I can't decide which one is the best, most commonly used."

To address the problem of synonyms, three of the four groups reported "testing synonyms". This strategy involves finding synonyms in dictionaries or resulting documents and querying each synonym separately to better understand slight variations in their meanings through the result sets. Our participants often tried several synonyms before arriving at the appropriate query.

"In English, I always open a dictionary, usually a monolingual one, in order to see synonyms too because I often don't know which word will result more websites… so I try several words."

Other pointers to address problems with vocabulary included the use of printed literature related to the academic searches to learn technical terms.

"They should hold printed literature in their hands to see technical terms."

Another tool to help solve problems with a lack of advanced language skills was the use of Google Translate and other translation tools. Our participants suggested searchers use Google Translate as it is easily accessible and usually offers several alternatives. At the same time, they recommended searchers should use dictionaries and other translators as well when searching for important information to supplement and double check Google Translate's recommendations.

"Don't exclusively trust in Google Translate. If they search for special topics it should be translated by more translators."

"Don't take translations by Google Translate for granted."

Our searchers also commented that Google's recommendations make it possible to enter query expressions with mistakes so searchers do not have to worry about grammatical mistakes and misspelling in English.

"Don't be afraid of grammatical mistakes and misspelling. Google will provide results websites even if you swap letters because it recognizes what you've wanted to enter. It's not necessary to enter grammatically correct phrases in the search engine only what we can type by heart depending on our English knowledge."

4.3.2 Visual search

Participants felt searching for and scanning images was sometimes more useful than searching for text, as it was faster for them to skim the images than opening each resulting page. They often arrived at the appropriate website after a related image or chart has been found.

"Images can lead us to websites where I find much more information than if I simply enter keywords." (Product design)

Thus, the visual information leads searchers to textual information that helps them create new queries.

4.3.3 Result sets

These short, broad, and often vague query expressions resulted in large and imprecise result sets, as opposed to the more specific and more relevant results in Hungarian. Our participants found these large result sets overwhelming. Several participants mentioned that especially in the early phases of their English language searching, they often felt intimidated by the large amount of information returned to them in English.

"It takes so much time… I have to open several pages to find what I really need."

Further complicating this problem was the quality of the results. Participants often reported dissatisfaction with the quality of the information returned. The low quality of the result sets may have been caused by low quality of the queries. Finding relevant information in a large amount of results returned was often a daunting task.

"It didn't show the important websites, there were rather plenty of useless results."

To counteract this problem, a few of our participants mentioned trying to create more specific queries in English, however, found this task daunting.

"I try to narrow the resulting pages down by entering specific queries because I struggle more with finding information in

English context. So I try to narrow it down so that I have to read less to get the information."

When giving advice to new searchers, one of the most often voiced recommendations was to limit the size of the result sets. Our participants advised new searchers to start searching for broad topics with general query expressions but then narrowing it down to avoid overwhelming amounts of information. Thus, our participants felt that limiting the results set and making it more specific would help searchers process information more efficiently.

"One should start searching in a broad topic; it is not worth to enter a lot of words as query because some results might not appear."

"They should paraphrase it with more words so that narrow the results down."

Participants encouraged the use of operators and the use of help systems about those operators. While they may not have often used operators themselves, they felt these would be helpful. The use of operators can help searchers form more sophisticated queries and narrow search result sets.

"One should search for an infographic containing all search aids."

They also recommended using format-specific searches as they can also narrow the resulting pages down.

"Google has thematic search options. It's possible to search for standards, books… Many people don't know about these but it's worth using them."

Thus our participants struggled with forming queries and tended to use shorter, more general queries, which resulted in large result sets of varying quality hits. More support for forming queries would provide tremendous help for second language searcher. When examining and interpreting the results language proficiency problems created more challenges, we will describe these in the next section.

4.4 Reading search results

As discussed above, our participants often described being faced with overwhelmingly large result sets of varying levels of relevance. When reading through selected results searchers were faced with the challenges of reading and understanding information in a foreign language. Not surprisingly, it usually took them longer and they found it more difficult to process text in English:

"In a long text, I can spot the keywords and go through it easier in Hungarian but in English, it takes more time to understand the gist of texts."

In order to compensate for the difficulty of reading in a foreign language, our participants reported using various strategies. These strategies helped them evaluate the relevance and the credibility of the sources. One strategy used was to focus on the main topic of search within the text usually by finding the keywords in the document so that it is not necessary to read the full text of a webpage. While scanning web pages is typical behavior of native language users as well (Krug, 2014), the foreign language searchers in our study emphasized the usefulness of this strategy for foreign language documents. Participants advised skimming text rather than reading the content of websites line by line. To

make it easier, searchers should use 'Ctrl+F keyboard shortcut to easily locate keywords.

"I often use the 'Ctrl+F' keyboard shortcut to search for a given word on a website. In Hungarian, it's much easier to look through the full text and I notice immediately which part I should read; but in English it helps."

"They should search with Ctrl+F as well so that it's visible where and how many times keywords appear in the text."

Another strategy some participants described was to choose one specific website and continue searching there in order to avoid facing the information overload of the result set:

"When I search in English, first I start on Google but after that I try to narrow it down because it results in a wide range of information. There is a huge amount of information available in English and I rather find one website on Google which is related to the topic and I search on that website again, maybe there is a better chance for finding it this way."

Foreign language searchers in our study were limited by their language skills and often entered short and vague queries which resulted in large and low quality result sets. To tackle the challenge of finding relevant information in these sets, they skimmed the text for keywords or selected just a few sites to look at. These challenges in understand the quality and relevance of the results can hinder effective and efficient information gathering.

4.5 Judging credibility

A third strategy for interpreting results was to focus on the visual design of the site as a first step, instead of reading the text right away. According to the focus group participants, the quality of the web design was more important on the English-language websites as they decided about the reliability of the website based on the layout and overall impression of the website. They described the quality of the visual design as determined by images, the navigation menu bar, and whether the design seemed professional or a simple template. The quality of the images had a strong impact on participants' impressions of websites as well. Thus, the role of images again proved more important in foreign language pages than in native language sites where they would judge quality by reading the text.

"In English, first I always look at the layout, label design and I seek keywords. In Hungarian, I decide based on the quality of the writing, I rather read an article or at least a few sentences to see how the author of the website formulates the words."

When asked more specifically about what to look for, one participant clarified by adding:

"Pictures, menu bar, design."

Showing that this challenge is a significant barrier, our participants mentioned several strategies to identify high quality websites in their advice to novice searchers. In the case of academic research, first it was recommended searchers should search in databases or on websites which contain trustworthy peer-reviewed work.

"In the case of academic research I often search on Google only for the second time. I don't trust in all information available on the Internet."

"I would recommend specific websites because Google is not appropriate for searching for specific content."

This step is often problematic since these searchers do not have an understanding of the types of website genres and culturally-specific markers of quality in a foreign information space.

Another piece of advice for judging credibility was to check and compare multiple sources of information.

"They should find at least one other source in any case."

"They shouldn't accept the first result; it has to be verified."

5. SYSTEM SUPPORT FOR FOREIGN-LANGUAGE SEARCHING

Participants also provided suggestions about how search engines and websites could provide help in searching in foreign languages.

5.1 Filtering by document genre

Our participants felt that it would be useful if they could filter the results by sources/genres of documents, such as only academic websites. When discussing this suggestion they were striving to identify the quality of websites from the result sets by filtering on these source/genre markers. They felt that certain genres of websites would ensure high quality information.

"On Google, it's not possible specifically searching for professional websites."

"It's difficult to find technical nature information because it results forums instead of professional sites."

5.2 Search Operators

Participants found search operators really useful, however they felt they did not know how to use them. Especially during searching in foreign language, it would be convenient to see the most important operators on the search screen and to be able to easily add them to queries. Our participants knew very few operators and they were not familiar with the actual functions of operators, but felt these would be useful to create better queries.

"You can search on a site, these [operators] are not emphasized only you can find them on the umpteenth page of Help. I don't know them by heart, it would be better to make them more visible/accessible."

"They could put a little guide on Google about how operators can be used because nobody knows about it."

"They could write tips for searching on Google for instance in drop-down box."

"I want a filter which shows only those resulting pages where the words of query terms are in the same sentence."

Although, one of the participants thought in the case of searching in English it's not essential to use operators.

"I use operators when I search in Hungarian but I think operators are not essential to search in English."

5.3 Icon to reflect reliability of websites

Our participants would also find it useful if on SERPs there were icons next to each website to show how reliable it is. Our participants do not have much information about foreign websites, for instance they do not know which website is the official source of the information needed. They felt they were not well equipped to judge the quality of websites.

"Especially in the case of searching in English, we don't know the foreign websites as the Hungarian ones and we have no idea how reliable they are. I don't know if Google has or not but websites could have a small icon with ratings."

"There is a plenty of virus which I can get easier on English websites as I don't know the websites, whereas I can eliminate this problem on the Hungarian ones."

5.4 Related searches and search history

There is a need for showing related searches on SERPs which would help when searchers are hindered by language limitations. Our participants reported problems with creating alternative query expressions when the ones they entered had not returned the expected results.

"Related searches, I don't know if it's available at regular searches but it's really useful when I search for images…. Sometimes I get stuck not because I misspell but because I don't know what other words could be used to search for it."
"It works very well on YouTube because it knows what kind of videos I search for and it shows the related ones. Yet Google doesn't show the similar topics, only specifically what I searched."

Some participants would also like to see a brief search history on the SERPs because they find it difficult to remember all of the previous query terms especially in foreign languages. By seeing the last 5-10 queries, searchers could recall the best alternatives or find a different query expression.

5.5 Automatic searches in foreign languages

Cross language information retrieval (CLIR) systems were also raised as possible solution. In CLIR systems searchers can enter queries in their own native language and the search engine would seek in English. This way they could skip one step – using translators. However, they only wanted queries translated, but not results.

"Integrated translator so that we don't need to switch web search engine and translator all the time. So I enter queries in Hungarian and it translates immediately."

Furthermore, in the case of searching for images, some of our participants expressed a need for search engines to search for the keywords in all languages as an option.

"It could search for images in other languages; because I enter it in Hungarian but for example in English or in Russian there will be six hundred times more results… I don't say to do it automatically – because it won't be good – but an option would be great."

5.6 Translation of websites

Participants of the focus groups considered website translator features useful, however, we must mention that the English-Hungarian translation is far from perfect.

"The translator could be improved. I don't want English websites to be translated because it annoys me that it doesn't translate well and sometimes it's really incorrect."

5.7 Site map

Site maps support foreign language searchers to find the information on the websites. Site maps can orient searchers to the structure of websites, which is especially helpful when searchers

may not be familiar with the web design standards of a certain culture.

"Site map helps but depend on the website where they put, create it or not."

6. CONCLUSION

Searchers who formulate queries, process result sets, and read results in a foreign language face a special set of challenges. Short and often crude queries result in large result sets containing items of widely varied relevance. Narrowing result sets through reformulating queries is limited by the language proficiency of searcher, specifically vocabulary and grammar problems. Translation tools, easy access to search operator, related queries and other query recommendation, and search histories can help searchers form better queries. Selecting items from the results lists is difficult without a strong ability to quickly judge the quality and trustworthiness of sites and interpretation of documents. Various information retrieval system features such as the filtering and categorization or visualization of results sets, improved text summarization algorithms, and clear and apparent reliability indicators can help alleviate these problems but are often not readily available in the first-level user interfaces of popular search engines. Future research should examine how systems can support foreign language searchers and which of a myriad of possible system features can in fact enrich the search experience.

7. REFERENCES

[1] Aula, A. and Kellar, M. (2009) Multilingual search strategies, CHI '09 Extended Abstracts on Human Factors in Computing Systems, New York, NY, USA, 3865-3870, DOI: 10.1145/1520340.1520585

[2] Berendt, B. and Kralisch, A. A user-centric approach to identifying best deployment strategies for language tools: the impact of content and access language on Web user behaviour and attitudes. Inf. Retr., 12, 3 (2009), 380-399., DOI: 10.1007/s10791-008-9086-4

[3] Caidi, N., Allard, D., & Quirke, L. 2010. The Information Practices of Immigrants. Annual Review of Information Science and Technology (ARIST), 44: 493-531., DOI: 10.1002/aris.2010.1440440118

[4] Chu, P., Jozsa, E., Komlodi, A., Hercegfi, K. (2012). An exploratory study on search behavior in different languages. Proceedings from the 4th Information Interaction in Context Symposium (IIiX 2012), ACM Press (2012), 318-321., DOI: 10.1145/2362724.2362784

[5] Clough, Paul, and Irene Eleta. "Investigating language skills and field of knowledge on multilingual information access in digital libraries." International Journal of Digital Library Systems (IJDLS) 1.1 (2010): 89-103., DOI: 10.4018/jdls.2010102705

[6] Corbin, Juliet, and Anselm Strauss. 2008. Basics of qualitative research: Techniques and procedures for developing grounded theory. Sage publications.

[7] Internet World Statistics. 2012. Top Ten Languages Used in the Web (Number of Internet Users by Language) Accessed April 2, 102. http://www.internetworldstats.com/stats7.htm

[8] Kralisch, A. and Mandl, T. Barriers to Information Access across Languages on the Internet: Network and Language Effects. In Proceedings of the Proceedings of the 39th Annual Hawaii International Conference on System Sciences - Volume 03 (2006). IEEE Computer Society, DOI: 10.1109/HICSS.2006.71

[9] Marlow, J., Clough, P., Recuero, J., and Artiles, J. (2008) Exploring the Effects of Language Skills on Multilingual Web Search. Proceedings from the IR research, 30th European conference on Advances in information retrieval (ECIR'08), 126-137., DOI: 10.1007/978-3-540-78646-7_14

[10] Merriam, Sharan B. 2009. Qualitative research: A guide to design and implementation. John Wiley & Sons

[11] Oard, D. W. 2009. Multilingual Information Access. In: Encyclopedia of Library and Information Sciences, 3rd Ed., edited by Marcia J. Bates, Editor, and Mary Niles Maack, Associate Editor, Taylor & Francis.

[12] Steichen, B., Ghorab, M. R., O'Connor, A., Lawless, S., and Wade, V. (2014) Towards Personalized Multilingual Information Access - Exploring the Browsing and Search Behavior of Multilingual Users. Proceedings from Conference on User Modelling, Adaptation and Personalization (UMAP 2014), Aalborg, Denmark, July 7-11., DOI: 10.1007/978-3-319-08786-3_39

[13] W3Techs. 2012. World Wide Web Technology Surveys. Usage of content languages for websites December 31, 2012. Accessed April 2, 2012. http://w3techs.com/technologies/overview/content_language/all