

Users' Perceptions of Searching in FlickLing

Evgenia Vassilakaki, Frances Johnson, R.J. Hartley, and David Randall

Dept. Information & Communications, MMU
{evgenia.vassilakaki}@student.mmu.ac.uk,
{f.johnson,r.j.hartley,d.randall}@mmu.ac.uk
<http://www.hlss.mmu.ac.uk/infocomms/>

Abstract. This study aims to explore users' image seeking behaviour when searching for known, non-annotated images across languages in FlickLing. In particular, the investigation reported in this paper aims to demonstrate the value in focusing on user's trust and confidence in the exploration of seeking behaviour to reveal users' perceptions of the tasks involved when searching across languages. The task assigned to our 24 users was to search for three specific images described in Dutch, German and Spanish consecutively. In this context, four different methods, both qualitative and quantitative (questionnaires, observation, retrospective thinking aloud and interviews) were employed. Preliminary finding regarding users' switching between the two modes, users' thinking of languages and their answers to the two questionnaires on trust and confidence are only reported here. It appears from our preliminary findings that there is variation in our users perceptions of searching for images across languages and their approach to using translations and that this occurs regardless of the amount or type of help or guidance given.

Key words: Multilingual Information Retrieval, User Behaviour, User Image Seeking Behaviour, Flickr, FlickLing, iCLEF

1 Introduction

The development of efficient and effective Cross Language Information Retrieval (CLIR) systems depends not only on the implementation of the various techniques for organizing linguistic resources but also on users' adaptation and usage. The interactive CLEF (iCLEF) track in Cross Language Evaluation Forum (CLEF) provides the participants with the necessary environment to run interactive experiments and focus on users' image seeking behaviour in multilingual environments. This is our second year participating in iCLEF track focusing on users' behaviour while searching across languages in FlickLing.

In the iCLEF2008 paper [4], the aim of the study was to explore users' image seeking behaviour when searching for the three known, non-annotated images across languages in FlickLing. In particular, three research questions were addressed: a) identify the reasons that determined our users' choice of a specific mode (monolingual/ multilingual), b) examine if and/or to what extent users

were thinking about languages when searching and retrieving images and c) examine if and/or to what extent users were paying attention to translations when searching and retrieving images. Four different methods, both qualitative and quantitative were adopted to answer these questions. A questionnaire was used to collect data on the users' knowledge of foreign languages, previous experience in searching for images, in searching across languages and previous experience with Flickr. An observation sheet was designed to help the facilitator take notes and to record comments and remarks made by the users on specific elements of their search behaviour. These notes further aided the formulation of questions during the interviews. In addition, retrospective thinking aloud was implemented to shed light on the users' actions and thoughts. Screen capture software was used to record a search session and this was then played back to the user asking them to describe what they were doing and what they were thinking. Finally, semi-structured interviews were carried out with each participant focusing on further explaining their specific actions, comments and remarks stated during the whole study. For further information on the way these methods were employed see [4], [5].

The iCLEF2008 small study [4] showed that less than half of our ten users appeared to consider identification of the language important in retrieving the images. The majority either lacked confidence in using the different languages or they were so focused on retrieving the images and completing the task that they were not thinking about languages. Only four out of the ten users tried to identify the language of the images and interact with the translations. These users judged the translations as poor because either they were not retrieved, they were not corresponding to users' search terms or they were retrieving irrelevant results. As a consequence, users were losing interest and trust in translations resulting in no usage of them or not paying attention to them.

The iCLEF2008 study also revealed some insight into users' trust in searching in a multilingual environment such as FlickLing and a possible relationship between trust and the use of translations, to search for the images in different languages. In addition, it provided some indication that users' confidence has an impact when searching for images across languages and that confidence may or may not be affected by users' knowledge of foreign languages. In this context, a follow up study was designed for iCLEF2009 drawing from the lessons learned from the iCLEF2008 paper [4] and focusing on users' trust and confidence as revealed and/ or formulated in multilingual environments. As a consequence, the iCLEF2009 paper again aims to explore users' image seeking behaviour when searching known, non-annotated images across languages. In particular, the investigation reported in this paper aims to demonstrate the value in focusing on user's trust and confidence in the investigation of seeking behaviour to reveal users' perceptions of the tasks involved when searching across languages.

The remainder of this paper is structured as follows: a description of the methods employed and the way the study was carried out are outlined in section 2. An initial analysis of the findings is given in sections 3. Finally, we conclude by summarizing the current indications of users' behaviour in section 4.

2 Method

In this section, further details on the test object, the task, the users and the methods employed are illustrated.

Test Object: This study used the FlickLing multilingual search engine for images as developed and provided by iCLEF organizers [4]. The same set up of the interface was kept for this year's participants to carry out their interactive studies.

Task: As in the iCLEF2008 paper, our users were asked to search for the first three known, non annotated images given after login in FlickLing. The users did not know in advance in which of the six languages supported by FlickLing (English, German, Dutch, French, Italian and Spanish) the three images were described. As a result, users had to use all the features and different modes (monolingual/ multilingual) provided by FlickLing to search the three images across languages.

As opposed to last year's study, the three images were pre-selected and presented to all 24 users so that the seeking behaviour and the thoughts recorded for each user referred to the experience of searching for the same three images. The three images were described in Dutch, German and Spanish consecutively. Drawing from [4] during the selection process all English described images were excluded as it was found that the majority of the users used English keywords to search for the given images and only a landmark or a piece of writing would trigger them to search across languages before going for the hints. As a result, no English images were given in an attempt to encourage users to think about languages and interact with the translations.

In addition, the images selected had both a 'heading' and at least three 'tags' describing them in order to provide our users with as many entry points to these images as possible. This emerged from the iCFLEF2008 study [4] as users complained when the images lacked either a heading or tags making their task of finding it more difficult. Furthermore, all three images included a distinctive landmark or a piece of writing that hinted at the language of the image. Again the 2008 study highlighted the users behaviour in using details in the image to identify the language of the image. Furthermore, there was no order of easy, medium and hard to find images adopted and applied in both studies since there was not such a distinction in FlickLing.

In this context, it should be also noted that during this year's study, words like task, or study were carefully avoided both from the written instructions read to the users and during the process of recruitment. Again drawing from [4] it was found that the users were obsessed with finding the images and completing the task "successfully" and that this competitive behaviour may have been influenced by the request for users to take part in the FlickLing task.

User Sample: This study used a sample of twenty four (24) users, eighteen (18) female and six (6) male ranging in age from 18 to 32 and only one stated above 32. In addition, eleven out of the twenty four users were 1st year undergraduate students, five 2nd year undergraduate students and the remaining eight postgraduate students all studying at Manchester Metropolitan University

(MMU). Furthermore, twenty two users were English Native speakers and the remaining two were Greek and Finnish native speakers. Eight were monolingual, two bilinguals and sixteen users stated knowledge of one or more foreign languages (this includes the two bilingual users who stated knowledge of additional foreign languages). In particular, the eight monolingual users were English native speakers, one of the bilinguals had Bengali as a native language, with an excellent knowledge of English and a basic knowledge of French and Spanish and the second user stated Urdu and English as native languages, a basic knowledge for German, Dutch, Italian and Spanish, a good knowledge for French and an excellent knowledge of Hindi. Finally, from the fourteen remaining users the two non English native speakers stated an excellent knowledge of English, seven users stated a basic and one very good knowledge of German, seven users stated a basic, one good, one very good and one excellent knowledge of French, two users stated a basic and one excellent knowledge of Italian, one stated a basic knowledge of Dutch and finally five users stated a basic knowledge of Spanish (see Table 1).

Language	Basic	Good	Very Good	Excellent
English	0	0	0	2
German	7	1	0	0
French	7	1	1	1
Italian	2	0	0	1
Dutch	1	0	0	0
Spanish	5	0	0	0

Table 1. Users' Knowledge of non-native Languages

All the twenty four users stated their experience in searching for images. In particular, five of them search for images very often, eight of them often, eight of them sometimes and the remaining three users rarely. In addition, eight out of the twenty four users have searched on the web for an image in a language other than their native. These users employ mainly their own language skills (29%), an online (24%) and hard copy dictionary (24%) consecutively, an online translator (13%) and the help of a friend (10%). The main reason for doing so was for school projects and assignments. The remaining sixteen users who are searching images only in their native language explained that they “never needed to”.

In regards to previous experience with Flickr, eighteen out of the twenty four have never used Flickr and explained “never heard of it”, “never needed to”, “I have not got around to using it” etc. Only six users have used Flickr in the past mainly for searching images (57%), sharing images (29%) and “been directed to it by Google” (14%).

Instructions: In iCLEF2008 study, the same written instructions were read to all users yet it was observed that the majority of users would use the Help instructions and ask for additional clarifications both regarding the task and the various features of FlickLing. In the context of trying to investigate to what

extent instructions may affect or not users' trust and confidence in FlickrLing and again retain an element of control in order to compare the findings, three different texts of instructions were written and used in the iCLEF2009 study. In particular, the users were divided in three different groups upon arrival. Each group was provided with different level of instructions of the given task going from 'no' to 'detailed' information regarding both the task and the functionality of FlickrLing's features. In particular, the information given to the first group was kept on a 'need to know basis' (see Appendix: Task Instructions). Users were given the minimum information necessary to understand and complete the task assigned. No explanation was given of the different fields and features employed in FlickrLing or how they could register and login. On the contrary, users were left alone to discover the multilingual environment and form their search strategies to retrieve the three given images.

The second group of users beside the general instructions concerning the task were also informed about the languages that FlickrLing supports and in which the three images can be described (see Appendix: Task Instructions). In addition, further information concerning the features of FlickrLing were briefly mentioned before describing them after login. This was a way of preparing (even unconsciously) the users for what was to be described thoroughly afterwards, on first login. Users were also instructed about the registration process but the information given was kept to what each field stands for rather than how it functions. In this context, information about the fields and features of FlickrLing was again provided after login but it was limited to what each field stands for rather than disclosing their functionality. This element was left to users to explore on their own while interacting with the interface.

The users in the third group were given the same instructions with the second group with the difference that the functionality of the features in FlickrLing was also disclosed (see Appendix: Task Instructions). In particular, the languages supported in FlickrLing and in which the images could be searched across and instructions concerning the registration process and how the various fields function were thoroughly described to these users. In addition, the fields found in FlickrLing and their functionality were also presented and thoroughly explained to the users in a more technical way.

This group was given all the information available on the task and the FlickrLing interface, information focusing on both what is in there and how it functions. These were explained in order to release users from the burden of exploring it on their own. In addition, it was done to meet the purpose of providing different level of information between the three groups.

The assignment of users to each group was done based on users' knowledge of foreign languages. In particular, based on the number of foreign languages of which they have some knowledge, the users were assigned to a specific group. This was done in the context of equally assigning users with no knowledge of foreign languages, knowledge of only one foreign language and knowledge of more than one foreign languages across the three groups. In this way, homogeneity of the

user sample distributed to the three groups in regards of knowledge of languages was enabled.

Data Collection: The same four methods, both qualitative and quantitative (questionnaire, observation, retrospective thinking aloud and interviews) were used for both studies. The adoption of the same methods was agreed on the basis that these methods enabled us to draw insights on different aspects of users' characteristics and/or search behaviours. In addition, the defined order of their implementation enabled us to gradually gain a better and a more comprehensive insight into users' actions and thoughts and formulation of relevant questions. Moreover, the way of combining the chosen methods contributed to offset their weaknesses and draw on their strengths enabling a better and more comprehensive understanding of users' image seeking behaviour in multilingual environments [5]. In particular, further description of the differences and similarities between the way the four methods were designed and applied for both studies is given below.

Questionnaires: In the iCLEF2008 study one questionnaire was adopted and implemented inquiring on users' native language, level of comprehension for a series of foreign languages, previous experience in searching for images, in searching across languages and in using Flickr and finally user demographics. This questionnaire was also adopted in the iCLEF2009 study because it successfully enabled the definition of the user sample characteristics and users' previous experience in each of the sections of interest and the formulation of relevant questions during interviews [5]. In addition, one more question was included in this questionnaire as a supplement regarding the means users employ to search across languages. This question was derived from users' comments during retrospective thinking of the iCLEF2008 study when they were describing the way in which they search across languages.

In addition to the first questionnaire, two more questionnaires were designed and implemented inquiring on users' trust and confidence which also emerged from [4]. The second questionnaire consisted of a series of statements reflecting on users' trust and confidence in general when they are searching for images on the web. A predefined six points scale (Don't Know, Not at all, A little, So-So, Very Much So, Totally) was provided for users to choose for each statement. These statements did not exceed ten and they were implemented as a second part of the first questionnaire. This was done both for presentation reasons and in an attempt to relief users from the extra hassle of having to complete two different questionnaires on arrival.

The third questionnaire had the same format as the second questionnaire. It consisted of a series of statements with a predefined scale of six points answers reflecting this time users' trust and confidence in FlickLing and its various features (monolingual/ multilingual mode, translations, hints etc). These statements did not exceed twelve in number. These two questionnaires were used to formulate an idea of each users' trust and confidence both in general and in particular in FlickLing and formulate relevant questions during the interviews. Finally, all three questionnaires were employed to gather information on specific

aspects of this study. At the same time, they all contributed to the same goal to explore users' image seeking behaviour in multilingual environments.

Observation: This study adopted again the same observation sheet as designed for the iCLEF2008 study [4]. It was kept the same because it proved valuable for making notes on specific actions of users while performing the task and formulating relevant questions during interviews. In addition, it enabled the facilitator to note down specific user's expressions while performing the task and use those to engage users in conversation shading light on the specific actions and thoughts of users reflecting on the reasons why [5].

Retrospective Thinking Aloud: This study also employed the retrospective thinking aloud method to reflect on users thoughts and action when searching for images across languages. This method belongs to the Think Aloud Protocols and it is a widely used method for usability testing of software and interfaces [2]. It was adopted and applied in the same way and using the same screen capture software as in the iCLEF2008 study [4]. The same method was adopted because during the iCLEF2008 study it enabled the facilitator to gain insights on users' information seeking behaviours, identifying the reasons why users were behaving in this way and asking relevant questions during interviews [5].

Interviews: The same method of interview was also adopted in this study as in the iCLEF2008 study [4] because it contributed to further clarifying users' comments and sayings during retrospective thinking aloud, facilitator's notes on the observation sheet and users' answers in the questionnaires. These interviews were again semi-structured and kept as brief as possible because users were already tired because of the retrospective thinking aloud and they were not willing to spend more time on responding to additional questions [5].

In general, the same methods were adopted and applied in the same order because they succeeded in gathering and verifying the validity and reliability of the data on users' image seeking behaviour in multilingual environments. The weaknesses posed by every method was also considered and tackled by employing a mixed methods research so that the weaknesses of one approach were offset by the strengths of others [5]. Although the same methods was used and implemented in the same order, minor changes and adjustments to the aim of this year study was considered important in an attempt to focus more on users' trust and confidence when searching for images in multilingual environments.

iCLEF2009 Study's Experimental Procedure: This study consisted of 24 individual sessions which lasted on average one hour and a half each. Each user on arrival was given the two questionnaires in the form of one to fill in. Based on users' number of foreign languages as they would define them in the first questionnaire, the facilitator would assign them a specific group (e.g. 1st, 2nd or 3rd). In addition, a number (01, 02, 03 etc) was also assigned to each user based on users' arrival in order to ensure the anonymity of the data. Once the user had the questionnaire completed, the facilitator read the relevant instructions depending on which group the user was assigned. The users then started the task while the facilitator observed and took notes. After the completion of the task, users were asked to fill in the last questionnaire. Users were then asked, based

on the video capturing their search session, to describe their actions and reflect their thoughts. At the same time, the facilitator took notes of users sayings and scanned through the two last questionnaires focusing on specific patterns regarding users' trust and confidence. Finally, a brief interview followed enabling the facilitator to ask questions based on the notes taken throughout the process to further clarify users' image seeking behaviour.

Data Analysis: The data gathered throughout the study were then analyzed to shed light on the different aspects of users' actions and perceptions. In particular, the questionnaires were processed in order to define the user sample's characteristics (see Table 1) and users' level of trust and confidence both in general (see Table 5) and in FlickLing (see Table 6). In addition, the transcripts based on retrospective thinking aloud and interviews were used to reveal users justifications of their actions and thoughts as well as their answers in the two questionnaires on trust and confidence. Furthermore, the videos based on users' image seeking behaviour were used to identify the specific actions of users and reflect back to users justifications and thoughts recorded during retrospective thinking aloud.

The transcripts based on retrospective thinking aloud only were further analyzed in regards to specific elements such as: a) users' switching between the modes (monolingual/ multilingual) throughout the three images (see Table 3 and 4), b) users' thinking about languages throughout the task (see Table 2), c) what triggered the users to think about the language of each image, d) the reasons why they were giving up on each image, e) why they were using the hints each time for each image and their actions after taking each hint, f) users' comments regarding trust and confidence. In addition, the users' explanations of their answers in the two questionnaires on trust and confidence during interviews were also identified and grouped. In particular, for each user specific comments were identified regarding a) users' trust in general, b) users' trust in FlickLing and c) whether their trust remained the same or changed throughout the task. The same was done for users' comments regarding confidence.

3 Findings

The small scale iCLEF2008 study laid the foundations and revealed the reasons why users were switching between the two modes, the extend to which they were thinking about languages and paying attention to translations consecutively. The current study aims to extend and reveal users' perceptions of the tasks involved when searching across languages based on these behaviours and the expressed justifications. In the context of a preliminary analysis and interpretation of the data collected, we can report the following in regards to users' perceptions when searching for known, non-annotated images across languages. In particular, data collected from the retrospective thinking analysis, the questionnaires on trust and confidence and the videos recorded will be presented in an attempt to identify users' perceptions when searching across languages.

One important element in the process of identifying users' perceptions is the extent to which they were thinking about languages throughout the three images. In particular (see Table 2), on the first image thirteen users did not think about languages before going for the first hint, six of whom were from the first group, three from the second and four from the third group. In addition, five users were thinking about languages from the beginning of the task, one of whom belonged to the first group and two to the second and third consecutively. Furthermore, six users started the task without thinking about languages but at some point before taking the hint started realizing that they needed to search across languages. It is also interesting to note that nineteen out of twenty four users started thinking about languages after taking the hint. Four users misinterpreted the hint on language and used it as a keyword. One user did not use the hints at all.

In terms of the second image, eight users started searching without thinking about languages, five of whom were from the first group, one from the second and two from the third. In addition, ten users were thinking about languages, three of whom from the first and second group consecutively and four from the third. Five users started without thinking about languages but at some point before going for the hint realized that they would need to search across languages, four of whom were from the second and one from the third group. Only one user started with a perception of languages but at some point he realized that he was actually trying to search only in English and he belonged to the third group. On the second image, none of the users misinterpreted the hint and used it as a keyword and again one user did not make use of the hints.

In regards to the last image, ten users again started without thinking about languages, four of whom were from the first, five from the second and one from the third group. In contrast twelve were thinking about languages, four of whom were from the first group, three from the second and five from the third. Two users started thinking about languages after spending sometime searching for the image but before going for the hint, these belonged to the third group. On the third image, again no users misinterpreted the hint and this time four of them did not make use of the hints.

It is interesting to point out that only three users were thinking about languages and trying to search for the three images across languages from the beginning and throughout the task. The rest of the users either did not think about languages throughout the three images and only the hint triggered them to think about it or at some point either on the first or the second image started to understand that they would need to search across languages.

Of particular interest is some of users' remarks regarding the task of searching across languages:

“but I didn't know what else to write or what I should do, because I didn't know any languages and things like searching in other languages”,

“even though I am writing in Dutch it doesn't necessarily mean that it will bring up different images or any more images”,

“but I didn't think that it would affect, especially that it would affect the search that you would use because you would have Dutch as a language to search so, I

just kept on typing English search”,

“ok it’s making me do more than just search, it’s making me use other languages for me to search and then I was happier because I could search a Spanish word which I obviously don’t speak”,

“searching in different languages seems, sounds like it would be much harder than searching in English”,

“I just found that I had to think about a lot of things like you have to find the image and try to translate it as well as find out if it is in English...it took me a while to grasp, what I was meant to be doing for this one, like I realized it had to be in a different language the image but actually I was looking for keywords...”.

Another aspect of users perception of searching across languages is the use of the two modes (monolingual/multilingual) (see Table 3 and Table 4). Four out of twenty four users stayed and used only one mode throughout the three images. In particular, three used only the monolingual mode and one only the multilingual mode throughout the three images. In terms of the rest twenty users, on the first image ten were switching between the two modes, eight used only the monolingual mode and the remaining two only the multilingual. On the second image, sixteen users were again switching between the two modes and four used only the multilingual mode. On the last image, seventeen users were switching between the modes, two were using only the monolingual and one the multilingual mode.

In the context of investigating users’ perceptions of the tasks involved when searching for images across languages, users’ degree of trust in applications and confidence in their searching and language skills were also monitored both with the two questionnaires employed, the users’ comments during retrospective thinking aloud and the individual interviews. In regards to the second questionnaire (see Table 5), the majority of the twenty four users have “Not at all” confidence in their language skills to search for images across languages on the web (S9), trust “A little” online translators to help them translate a text (S3), for languages of which they have some knowledge, they trust “A little” provided translations for searching on the web (S5) and they trust “A little” for use the search terms suggested to them by search engines on the web (S6).

In addition, the majority of the twenty four users trust “So-so”¹ applications on the web to help them do the things they want to do (S1), the results that search engines on the web give (S2), and they have “So-so” confidence in their personal skills and abilities to search for images on the web (S8). Furthermore, the majority of the twenty four users answered equally that for languages they do not understand, they trust “A little” and “So-so” provided translations for searching on the web (S4). The majority again equally stated that they trust “So-so” and “Very much so” Help instructions to understand the use of interfaces (S7). Finally, the majority of users who have experience in searching on the

¹ A predefined six points scale (Don’t know!, Not at all, A little, So-so, Very much so, Totally) was given to users to choose for each statement (see section 2: Questionnaires).

web across languages stated that they have “Very much so” confidence when searching images across languages because of their relevant experience.

In terms of the third questionnaire (see Table 6), the majority of the twenty four users stated they “Don’t know” for the statement “I trusted the help instructions to guide me through FlickLing (S8)”. In addition, the majority had “Not at all” condence in their language skills to search for the images across languages in FlickLing (S11). The majority of users had “A little” condence in their personal skills to search for the images in FlickLing (S10) and in their experience that they were gaining in FlickLing for searching the images across languages (S12). In addition, the majority “So-so” trusted the results that FlickLings both monolingual and multilingual mode gave them (S1 and S2), trusted “So-so” the translations FlickLing gives for languages they do not know (S4) and they are “So-so” condent that FlickLing helped them search across languages for the images (S9). Finally, the majority trusted “Very much so” FlickLing to translate their query (S3), the translations FlickLing gives for languages they understand (S5), FlickLing to suggest search terms (S6) and the hints provided by FlickLing to search for the images (S7).

The second goal of the analysis would be to identify the reasons why users were or were not thinking about languages on each image and throughout the task. In addition, the reasons why the four users used only one mode for searching across languages and why the rest were either switching between the two or using just one mode while searching for the three images are to be further analyzed. In terms of users’ trust and confidence, further analysis is intended to combine the data gathered from questionnaires, retrospective thinking aloud and interviews to shed light on the reasons why users provided the following answers to the 2nd and 3rd Questionnaires and determine whether trust and confidence are factors affecting users’ perceptions of the tasks involved when searching across languages. Finally, to meet our aim to gain insight into users search behaviour and use of translations further analysis will be carried out on data such as users’ comments regarding the three given images, their comments regarding their language skills, and their interpretation and usage of the translation mechanism. It appears from our preliminary findings that there is variation in our users perceptions of searching for images across languages and their approach to using translations and that this occurs regardless of the amount or type of help or guidance given. Further analysis of the data collected on users perceptions of the task and their explanation for their actions may provide further insight into understanding the user in the design and development of multilingual retrieval systems.

4 Conclusion

This study aimed at exploring users’ image seeking behaviour when searching for known, non-annotated images across languages in FlickLing. In particular, this paper aimed to demonstrate the value in focusing on user’s trust and confidence in the investigation of image seeking behaviour to reveal users’ perceptions of

the tasks involved when searching across languages. The task assigned to our 24 users was to search for the three images given to them after first login. These images were described in Dutch, German and Spanish consecutively. In this context, four different methods, both qualitative and quantitative (questionnaires, observation, retrospective thinking aloud and interviews) were employed to meet its research question.

Preliminary findings regarding users' switching between the two modes, users' thinking of languages and their answers to the two questionnaires on trust and confidence were only reported due to the inability to process the large amount of data collected in such a short period of time. Nonetheless, it appears from our preliminary findings that there is variation in our users' perceptions of searching for images across languages and their approach to using translations and that this occurs regardless of the amount or type of help or guidance given. Further analysis of the data will provide a better understanding of users' behaviour and perceptions so as to build cross-language information retrieval systems that better serve users.

References

1. Gonzalo, J., Clough, P., Karlgren, J.: Overview of iCLEF 2008: Search Log Analysis for Multilingual Image Retrieval. In Peters et al (eds): Evaluating Systems for Multilingual and Multimodal Information Access 9th Workshop of the Cross-Language Evaluation Forum, CLEF 2008, Aarhus, Denmark, September 17-19, 2008, Revised Selected Papers, Springer LNCS.
2. Haak, M., de Jong, M., Schellens, P.J.: Retrospective vs. Concurrent Think-aloud Protocols: Testing the Usability of an Online Library Catalogue. *Behaviour & Information Technology* 22, 339–351 (2003).
3. Petrelli, D., Beaulieu, M., Sanderson, M., Demetriou, G., Herring, P., Hansen, P.: Observing Users, Designing Clarity: A Case Study on the User-centered Design of a Cross-language Information Retrieval System. *JASIST* 55, 923–934 (2004).
4. Vassilakaki, E., Johnson, F., Hartley, R.J., Randall, D.: A Study of Users Image Seeking Behaviour in Flickling. In Peters et al (eds): Evaluating Systems for Multilingual and Multimodal Information Access 9th Workshop of the Cross-Language Evaluation Forum, CLEF 2008, Aarhus, Denmark, September 17-19, 2008, Revised Selected Papers, Springer LNCS.
5. Vassilakaki, E., Johnson, F., Hartley, R.J. and Randall D., 2009. Users Image Seeking Behaviour in Multilingual Environments: experience in combining qualitative and quantitative data. In International Conference on Qualitative and Quantitative Methods in Libraries, Chania, Greece, 26-29 May 2009.

Appendix: Task Instructions

This study used three different groups of users. These users received different levels of information about the task to be carried out and the interface to be employed. The full text of the instructions provided to each group of users (1st, 2nd and 3rd) is outlined here. The reader can find a detailed description of these in the relevant chapter.

First Group Task Instructions

- The organizers would like to thank you for sparing some of your time to participate in this study.
- I would like you first to fill in this short questionnaire on basic background knowledge.
- This study uses FlickLing, a multilingual interface for searching images across languages. You are asked to use this interface to search the first three images given to you after login. The images given can be described in any of the following six (6) languages: English, German, Dutch, French, Italian and Spanish. You will not know in advance in which of these languages the images provided are described. Therefore, you need to use the features supported by the FlickLing interface in order to search the three (3) images across languages.
- At this point, the organizers feel the need to remind you that it is the system and not you that is being evaluated and that you should perform as usual, spending as much time as you would have normally spent on searching.
- You will be asked to fill in another short questionnaire after completing the searches for the three images.
- Your search session will be system captured and played back to you. You will then be asked to describe your thoughts and actions while searching for the three images. The facilitator will encourage you to keep on describing your thoughts and actions when going to silence. A digital recorder will be used to tape your description.
- At the end, a short length interview will take place between you and the facilitator that will be also taped.
- I would now ask you to register as a member of the MMU team and start searching for the images.

Second Group Task Instructions

- The organizers would like to thank you for sparing some of your time to participate in this study.
- I would like you first to fill in this short questionnaire on basic background knowledge.
- This study uses FlickLing, a multilingual interface for searching images across languages. You are asked to use this interface to search the first three images given to you after login. The images given can be described in any of the following six (6) languages: English, German, Dutch, French, Italian and Spanish. You will not know in advance in which of these languages the images provided are described. Therefore, you need to use the features (Help instructions, monolingual/multilingual mode, translations, suggested terms, give up button and hints) supported by FlickLing interface in order to search the three (3) images across languages.
- At this point, the organizers feel the need to remind you that it is the system and not you that is being evaluated and that you should perform as usual, spending as much time as you would have normally spent on searching.

- You will be asked to fill in another short questionnaire after completing the searches for the three images.
- Your search session will be system captured and played back to you. You will then be asked to describe your thoughts and actions while searching for the three images. The facilitator will encourage you to keep on describing your thoughts and actions when going to silence. A digital recorder will be used to tape your description.
- At the end, a short length interview will take place between you and the facilitator that will be also taped.
- I would now guide you through registration for login to the FlickLing. You need to click on ‘register’ button. First, you need to type in an ‘email account’ and provide a ‘password’. Choose your ‘native language’ from the menu and the ‘language of the interface’ again from the ones in the relevant menu. Determine your team from the scroll down menu as ‘MMU’ team. Choose from the menu the ‘level of comprehension’ for the six languages supported by FlickLing and then click on ‘ok’.
- You are now ready to use the FlickLing interface and search for the three images. As you can see on your left, you have the first image. You can click on it in order to bring it forward and see it in detail. Two modes are provided, the monolingual and multilingual. On the monolingual, you write in one language and you get results only in the language of your query. On multilingual mode, you write your query in one language and you can determine in which languages of the six supported by FlickLing you can get results. On the right side of the search box, you will get each time the translations. Each time, you can cross out or provide a new translation for each desired language. You can also use the suggested terms provided by the system and the tags found in the images of the results. At any time, you can give up on a photo or make use of the hints provided by FlickLing.
- You can now start searching for the three images.

Third Group Task Instructions

- The organizers would like to thank you for giving some of your time to participate in this study.
- I would like you first to fill in this short questionnaire on basic background knowledge.
- This study uses FlickLing, a multilingual interface for searching images across languages. You are asked to use this interface to search the first three images given to you after login. The images given can be described in any of the following six (6) languages: English, German, Dutch, French, Italian and Spanish. You will not know in advance in which of these languages the images provided are described. Therefore, you need to use the features (Help instructions, monolingual/multilingual mode, translations, suggested terms, give up button and hints) supported by FlickLing interface in order to search the three (3) images across languages.

- At this point, the organizers feel the need to remind you that it is the system and not you that is being evaluated and that you should perform as usual, spending as much time as you would have normally spent on searching.
- You will be asked to fill in another short questionnaire after completing the searches for the three images.
- Your search session will be system captured and played back to you after completing the searches. You will then be asked to describe your thoughts and actions while searching for the three images. The facilitator will encourage you to keep on describing your thoughts and actions when going to silence. A digital recorder will be used to tape your description.
- At the end, a short length interview will take place between you and the facilitator that will be also taped.
- I would now guide you through registration for login to FlickLing. You need to click on 'register' button. First, you need to type in an 'email account' and provide a 'password'. Choose your 'native language' from the menu and the 'language of the interface' again from the ones in the relevant menu. Determine your team from the scroll down menu as 'MMU' team. Choose from the menu the 'level of comprehension' for the six languages supported by FlickLing and then click on 'ok'.
- You are now ready to use the FlickLing interface and search for the three images. As you can see on your left side, you have the first image. You can click on it in order to bring it forward and see it in detail. By clicking on ok button, you go back to your previous screen. Two modes are provided, the monolingual and multilingual. In the monolingual, you search in one language and you get results only in the language of your query. In multilingual mode, you can determine in which language you write the query from the drop down menu and in which languages you get results by crossing out the relevant language buttons. If you do not cross out a language button then you get translations and consecutively results in all six languages. On the right side of the search box, you will get with each query the relevant translations. The system retrieves at the same time your search terms and the provided translations, so there is no need to retype the translations in the search box. Each time, you can cross out or provide a new translation for each language by clicking on the translation. You can use the suggested terms provided by the system after each search underneath your image by clicking on the term and selecting one of the options provided (include only this word, include this word to the rest of the query, include the translation of this word to the query). You can also use the tags found in the images of the results again by clicking on the tag and selecting one of the options provided (include only this word, include this word to the rest of the query, include the translation of this word to the query). The results are presented underneath your image and are listed in pages. The number of the pages varies according to the number of results. At any time, you can give up on a photo by clicking on the give up button found on the bottom side of the image. After clicking on the give up button, you can either make use of the hints provided by FlickLing or decide to skip this image and go for the next

one. The system provides a number of hints, disclosing each time the language in which the image is described and the terms that retrieve this image. You can use as many hints as you wish by repeating the same steps (giving up and going for the hint button). You can now start searching for the three images.

Appendix: Analysis Tables

Group_User ID Images	TAL					
	1st		2nd		3rd	
	BH	AH	BH	AH	BH	AH
G1.01	N	Y	N	Y	N	Y
G1.04	N	Y	N	Y	N	Y
G1.07	N	N	Y	Y	N	Y
G1.10	N	Y	N	Y	Y	Y
G1.13	N	Y	N	Y	N	Y
G1.16	Y	Y	Y	Y	Y	Y
G1.21	N/Y	Y	Y	Y	Y	Y
G1.23	N	N	N	Y	Y	Y
G2.02	N	Y	N	Y	N	Y
G2.05	Y	Y	Y	Y	N	Y
G2.09	N/Y	Y	Y	Y	Y	N/A
G2.11	N	Y	Y	Y	N	Y
G2.14	N	Y	N/Y	Y	N	Y
G2.17	Y	Y	N/Y	Y	Y	Y
G2.19	N/Y	Y	N/Y	Y	Y	Y
G2.22	N/Y	Y	N/Y	N/A	N	N/A
G3.03	Y	Y	Y	Y	Y	Y
G3.06	N	Y	Y	Y	N/Y	Y
G3.08	N	Y	Y	Y	Y	N/A
G3.12	N	N	Y/N	Y	Y	Y
G3.15	N/Y	Y	N	Y	N/Y	Y
G3.18	N/Y	N	N/Y	Y	N	N/A
G3.20	Y	Y	Y	Y	Y	Y
G3.24	N	N/A	N	Y	Y	Y

Table 2. Users' Thinking About Languages

Table 2

TAL= Thinking About Languages

BH= Before taking Hint

AH= After taking Hint

Y= Yes

N= No

N/Y= No at first and then Yes

Y/N= Yes at first and then No

N/A= Non Applicable

Group_User ID	SBM		
	1st	2nd	3rd
Images			
G1.01	N	N	N
G1.04	N	Y	Y
G1.07	N	Y	Y
G1.10	N	Y	Y
G1.13	N	N	N
G1.16	N	Y	Y
G1.21	Y	Y	Y
G1.23	Y	Y	Y
G2.02	Y	N	Y
G2.05	Y	Y	Y
G2.09	N	Y	Y
G2.11	Y	Y	Y
G2.14	Y	Y	Y
G2.17	Y	N	Y
G2.19	N	N	N
G2.22	N	Y	N
G3.03	N	N	Y
G3.06	N	N	N
G3.08	Y	N	N
G3.12	N	Y	Y
G3.15	Y	Y	Y
G3.18	N	Y	N
G3.20	Y	Y	Y
G3.24	N	Y	Y

Table 3. Users' Switching Between Modes

Table 3

SBM= Switching Between Modes

Y= Yes

N= No

Group_User ID	SBM		
	1st	2nd	3rd
Images			
G1.01	OMN	OMN	OMN
G1.04	OMN	M&M	M&M
G1.07	OMN	M&M	M&M
G1.10	OMN	M&M	M&M
G1.13	OMN	OMN	OMN
G1.16	OMN	M&M	M&M
G1.21	M&M	M&M	M&M
G1.23	M&M	M&M	M&M
G2.02	M&M	OML	M&M
G2.05	M&M	M&M	M&M
G2.09	OMN	M&M	M&M
G2.11	M&M	M&M	M&M
G2.14	M&M	M&M	M&M
G2.17	M&M	OML	M&M
G2.19	OMN	OMN	OMN
G2.22	OML	M&M	OMN
G3.03	OML	OML	M&M
G3.06	OML	OML	OML
G3.08	M&M	OML	OML
G3.12	OMN	M&M	M&M
G3.15	M&M	M&M	M&M
G3.18	OMN	M&M	OMN
G3.20	M&M	M&M	M&M
G3.24	OMN	M&M	M&M

Table 4. Users Modes' Usage

Table 4

SBM= Switching Between Modes

M&M= Both Monolingual and Multilingual Modes

OMN= Only Monolingual Mode

OML= Only Multilingual Mode

Answers	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
Don't know!	0	0	2	4	3	0	2	0	1	1
Not at all	0	0	2	1	0	2	2	0	7	0
A little	4	4	9	7	9	8	5	1	5	3
So-so	12	14	6	7	5	7	7	10	4	1
Very much so	7	6	5	5	7	7	7	9	5	5
Totally	1	0	0	0	0	0	1	4	2	0

Table 5. Users' Answers to the 2nd Questionnaire

Table 5

S= Statement

S1= In general, I trust applications on the Web to help me do the things that I want to do.

S2= In general, I trust the results that search engines on the web give me.

S3= In general, I trust online translators to help me translate a text.

S4= In general, for languages I do not understand, I trust provided translations for searching on the web.

S5= In general, for languages of which I have some knowledge, I trust provided translations for searching on the web.

S6= In general, I trust for use the search terms suggested to me by search engines on the web.

S7= In general, I trust Help instructions to understand the use of interfaces.

S8= In general, I have confidence in my personal skills and abilities to search for images on the web.

S9= In general, I have confidence in my language skills to search for images across languages on the web.

S10= In general, I have confidence when searching images across languages because of my relevant experience.

Answers	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12
Don't know!	1	3	2	2	2	1	0	10	2	0	1	3
Not at all	2	2	3	2	0	1	0	2	4	5	12	1
A little	5	4	0	1	4	3	4	4	4	10	4	9
So-so	10	8	9	10	5	8	5	3	7	5	3	6
Very much so	5	7	10	6	9	10	11	5	6	2	4	4
Totally	1	0	0	1	2	1	4	0	0	2	0	1

Table 6. Users' Answers to the 3rd Questionnaire

Table 6

S= Statement

S1= I trusted the results that FlickLing's monolingual mode gave me.

S2= I trusted the results that FlickLing's multilingual mode gave me.

S3= I trusted FlickLing to translate my query.

S4= I trusted the translations FlickLing gives for languages I do not know.

S5= I trusted the translations FlickLing gives for languages I understand.

S6= I trusted FlickLing to suggest search terms.

S7= I trusted the hints provided by FlickLing to search for the images.

S8= I trusted the help instructions to guide me through FlickLing

S9= I am confident that FlickLing helped me search across languages for the images.

S10= I had confidence in my personal skills to search for the images in FlickLing.

S11= I had confidence in my language skills to search for the images across languages in FlickLing.

S12= I had confidence in the experience that I was gaining in FlickLing for searching the images across languages.