

# A Study of Users' Image Seeking Behaviour in FlickLing

Evgenia Vassilakaki, Frances Johnson, Richard 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 a known, non-annotated image in the FlickLing search interface provided by iCLEF2008 track. The main focus of our study was threefold: a) to identify the reasons that determined users' choice of a specific interface mode, b) to examine whether users were thinking about languages when searching for images and to what extent and c) to examine if used, how helpful the translations proved to be for finding the images. This study used questionnaires, retrospective thinking aloud, observation and interviews to meet its research questions.

**Keywords:** Multilingual Information Retrieval, User Behaviour, User Image Seeking Behaviour, Flickr, FlickLing, iCLEF.

## 1 Introduction

Cross Language Evaluation Forum (CLEF) is an annual evaluation campaign that aims to promote the development of monolingual and multilingual information retrieval systems for European Languages. The 2008 iCLEF track focuses both on acquiring a large set of search session logs for the participants to mine and on allowing participants to perform their own interactive experiments with the FlickLing interface provided and adopting the task predefined by the Organizers [1]. The aim of this study is to explore users' image seeking behaviour when searching and retrieving known, non-annotated images across languages in FlickLing. In particular, the research questions that will be addressed in this paper are: a) identify the reasons that determined our users' choice of a specific interface (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.

The remainder of this paper is structured as follows: details concerning the different methods that we used in assembling the data and the way that the study was carried out are given in section 2. We provide an analysis of our findings and a discussion of them in sections 3 and 4 respectively. Finally, we conclude summarizing the different image seeking behaviours that our users exhibited while using FlickLing in section 5.

## 2 Method

In this section, further details about the test object, the users, the task, the methods employed are presented.

**Test Object.** Our main reason of interest in participating in the iCLEF2008 Flickr challenge was to investigate the behaviour of users when asked to search and retrieve a known, non-annotated image across languages.

**Users.** The study was carried out with a sample of 10 users, three male and seven female, ranging in age from 20 to 40. In particular, seven were research postgraduate students, one taught postgraduate student, one lecturer and one member of MMU administrative staff. In addition, four of the users were English native speakers, two Greek, one German, one Spanish, one Arabic and one Luganda. Moreover, one of the users was monolingual, four stated knowledge of a language other than their native and five were multilingual.

The users were also asked to state their level of comprehension for the languages used in FlickLing but also any other additional language. In particular, from the six non English native speakers, two stated an Excellent knowledge of English, three Very Good and one Basic. In regards of German, three of the nine non-German native speakers stated a Basic knowledge of German. Four out of ten users stated knowledge of French, three of whom Basic and one Good. Concerning Italian two out of ten users stated knowledge of Italian language, Basic and Good respectively. Two out of ten stated a Basic knowledge of Dutch and finally, three out of nine non Spanish native speakers stated a Basic knowledge of Spanish.

All ten users have searched in the past for an image on the web. In particular, four stated that they “rarely” have, three “sometimes”, two “very often” and one “often”. In addition, nine out of ten stated that they have searched for an image on the web in a language other than their native and only one had not. In addition to the users’ previous knowledge and experience with Flickr, nine out of ten users answered this question. Three of whom stated that “Yes” they have used Flickr in the past.

**Task.** Our users were asked to search for the first three (3) given images after login using all the features of the FlickLing interface. The users did not know in advance in which of the six languages (English, German, Dutch, Spanish, French, Italian) the image was described enforcing them to use both monolingual and multilingual modes to find the given image. The images presented to users were not controlled but given randomly from a set of 100 stored in the FlickLing database.

**Retrospective Thinking Aloud.** Retrospective thinking aloud is a widely used method for usability testing of software and interfaces. Its basic principle is to ask from potential users to complete a certain task with the testing object in question and to describe their thoughts and actions afterwards on the basis of a video recording their task performance [2]. This method focuses on peoples’ cognitive processes after having completed a specific task. It is a method that

enables the users and not the experts to point out the problems concerning the test object in the usability test. In this context, we used a premiere screen recorder in order to capture the users' search sessions in individual videos and a digital recorder for the retrospective thinking aloud.

**Observation.** The observation method was used to form specific questions regarding preselected research areas of the test object in an attempt to shed light on specific behaviours of the users on specific occasions. A form was created to assist the work of the facilitator at focusing on specific areas of interest and at the same time reflecting on users' behaviour.

**Interviews.** The last part of the study consisted of small scale individual interviews with every user after the completion of the retrospective thinking aloud. The interviews lasted no more than 10 minutes for every user. The questions asked varied according to user's answers to the questionnaire, search session, retrospective thinking aloud and the notes gathered throughout the study. The main goal of these questions was to clarify specific actions of the user's image seeking behaviour during the search session and expressions that the user used to describe what he/she was doing.

**Experimental Procedure.** The study was carried out in 10 individual sessions, at the same lab and each lasted from one to two hours approximately. During each session, users were given written instructions about the experimental procedure and the task itself. After that, users were asked to fill in the questionnaire on background knowledge, login and start completing the task (search for the first three images) while screen recording software was taping the computer screen. Having done that, users were asked to watch the recorded session and describe what they were doing and thinking in retrospect. Finally, a semi-structured interview lasting no more than 10 minutes was carried out with each user.

### 3 Findings

The analysis of data gathered focused on and will be presented according to the study's three research questions (RQ).

**1st RQ: Identify the reasons that determined each time users' choice of a specific interface (monolingual/ multilingual).** Out of the ten users, two used the monolingual interface and the rest switched between interfaces. The reasons our ten users gave for their behaviour in the thinking aloud process and interviews are stated below.

*Only Monolingual Interface:* Two out of ten users did not use at all the multilingual interface, even though they were given images to search in a language unknown to them. The first user, an English native speaker with basic knowledge in French when informed by the system that the image was annotated in French, stated: "because I am not good in French...I realized that I am never going to find it...So, I decided to give up". When asked why the subject didn't use the

multilingual interface, the subject answered: *“Because I did not trust my abilities with other languages, to be able to put the decent search words in...Because I did not know the keywords to search in other languages”*. As a final remark, the subject added: *“I was not confident with the languages”*.

The second user, a Luganda native speaker with no knowledge of French, stated: *“I went for the hint and it said that the image is described in French... Well, I thought, I do not speak French, I can’t understand that”*. When asked why the subject did not use the multilingual interface, answered: *“If I knew how to use another language, then I could use the multilingual and access the same image in another language”*. When asked how the subject was planning to cope with the problem of searching a French annotated image on monolingual interface by using English keywords, the subject stated: *“I thought that the image was not available and all images should be described in English [as well]. So, I thought that it was inaccessible...that I could not get it”*. These two users were not feeling confident of their language skills and they were not used to searching for images in languages other than English. These users would use the multilingual interface only if they could speak the language in which they were searching and that one would be other than English.

*Switching between Monolingual & Multilingual Interface:* The remaining eight users switched between monolingual and multilingual interfaces in order to complete the given task. A variety of reasons to justify these actions were reported by the users during the retrospective thinking aloud process and interviews. In particular, users identified the following reasons why: *“In order to increase or decrease the number of results, depending on the results that I had on the beginning of my search”*, *“because I assumed that it would give me the highest possible number of relevant results in relation to my query”*, *“I am trying to find the right combination of keywords”*, *“if you know where the picture was from, or if you know the place then you can like recognize the language in which you can type in”*, *“I was looking to isolate words and translate them”*, *“Simple because I wasn’t getting any of the results that I wanted”*, *“I tried to increase my chances of getting the image...I am widening my possibilities”*, *“I am just trying out the system”*, *“So, it was not there [monolingual English], I guess it was in other language”* and lastly: *“For me the problem was more kind of how to find where the image was from”*.

Also, hints played a significant part in users’ choice of an interface. As stated by the users: *“I switched to monolingual because the hint told me that the image was described in English”* and *“I went to ask for a hint on language just in case because that seemed to save me lots of time”*.

Two users, both English native speakers, stayed on multilingual interface though after taking the hint, they both knew that their image was described in English. When asked why they haven’t switched to monolingual, they stated consecutively: *“Because I did not think that would make any difference, because I was assuming that it is in English as well”* and *“Well, because I was there. I did not realize that...I thought, to be honest, I thought, it’s not going to make that much difference really”*.

There were also some cases that although users were seemingly using a specific interface (monolingual or multilingual) they stated during retrospective thinking aloud process and confirmed afterward with the interviews that: *"I was not paying attention to the fact that it was multilingual. Maybe, I forgot about that and left it as it was"* and *"I was so focused on trying to see how to describe the image that I was not paying attention to the interface"*.

**2nd RQ: to explore if and/or to what extent languages were forming the image seeking behaviour of our users.** Two users out of ten used only the monolingual interface searching in English, although they knew that the images may not be described in English. From their reasons stated for this decision it appears that knowledge of and/or confidence in a language other than their native language is a determining factor. When asked if the subject was thinking about languages while searching, the Luganda native speaker answered: *"No. I did not...because when I am searching for images on the Internet, I normally get them in English because I imagine that...I guess it's a little bit of arrogance, I speak English and I imagine that images...That if you put them in Internet, they should have English tags"*.

The other eight users who were switching between monolingual and multilingual interfaces, can be divided in two groups: a) those who were thinking about languages and b) those for whom languages were not a variable when performing the given task. In particular, four of them stated that: *"Now, I made the relationship of country, Florida...I write them [keywords] in English"*, *"I had the feeling that the building which I recognized, was described in German"*, *"It was not within my results, so, I guessed that it is in other language"*, *"Because by looking at the tortoise had written on it...it was written in English. So, I assumed that it would be in English...And I was also thinking at this time, I wonder if it is English or not...because the child got a little blue and red hat and I was thinking, maybe the child is French...Yes, I changed into multilingual because I think that maybe it is French, with the outside possibilities that it might be Italian"* and lastly *"Well, that's probable a bit Anglo-centrism. You know, well, it is a picture in England"*.

On the other hand, the remaining four users when asked if they were thinking about languages during the task, they said that: *"To be honest, I was not thinking about languages...I did not consider it a variable that influences my results"*, *"I did not bother about languages...I did not really think about them...I did not focus on languages while performing my searches. Maybe, because I am not used to, is not widely used or maybe I am not using languages when retrieving information on the web"*, *"I was not taking languages under consideration when searching for the images"* and *"For me it was not a question of language...In my mind language was a very small factor in there [FlickLing]. It did not really play any important role"*.

**3rd RQ: to examine the use of translations and the influence of translations on the users' information seeking behaviour.** We are obliged at this point to exclude the two users who used only the monolingual interface and

the four users who used the multilingual interface but with no thought to the translations. The remaining four users tried both the monolingual and multilingual interfaces driven by the need to identify the language of the image and the appropriate keywords to retrieve the given images. In particular the four users, when asked if they were paying attention to the translations, stated: *“Yes, but it did not translate anything”*, *“Yes, I did use them”*, *“Yes, at this point I am trying to figure out how this translation thing works”* and *“Yes, I was paying attention to the translations”*.

In addition, when users asked if they could judge the translations that were given to them, users answered: *“Overall, I had the feeling that the translations of the system were not that good...I switched to monolingual because the translations were not doing anything”*, *“I would trust the system to give me the right translations...I would have to for languages unknown to me”*, *“Because I went for the languages that I had a vague idea about and it did not tell me something that I did not really know”* and last *“I was not satisfied with the German translations because I can understand German...it’s not the right word in German for a man. So, it should have been something else...in Dutch I don’t know what the translation is, so, I had to accept it, whatever it is...Yes, I was satisfied [Dutch translations] because the computer knows the Dutch language better than I do... maybe that’s not the best translation, so, I just had to accept it. There was anything that I could do about it really”*.

Finally, when the users were asked if the translations were helpful in terms of actually contributing to the retrieval of the image, users stated: *“Ok, I have got the translations but they are not doing anything to me...at the end, I totally disregard the translations”*, *“I think that I stop searching for translations, when I stop having much confidence that it was bringing me the right translations...”*, *“The words that I was trying to isolate like particular words like London, the different translations there were not coming up...and what it was saying, like gigante in Italian for giant, it told something that I already knew. So it was not isolating the words in the way that I wanted it to. It was just telling me the adjectives were, which I did not really need”* and finally *“At the end I was not paying attention to the translations, I was purely interested in finding the image as quickly as possible because once more I did not think the translations would necessarily help me”*.

## 4 Discussion

The evaluation of CLIR effectiveness often does not involve the end user largely because initial hypotheses often exclude their experiences. It follows that experimental success is not success in the users eyes. Much of what we set out to do was to assess the difference between experimental assumptions and a user perspective rather than provide a test of “success”. On the one hand, it may be assumed that since the translation is automated the user has no role to play or possibly that the user has no interest in the translation, providing the system is effective. On the other hand, the non trivial challenges posed in the effort in designing realistic task scenarios, recruiting participants, analyzing large amounts

of data to obtain user assessments or to observe search behaviour can be prohibitive. However, we take the view of Petrelli [3] that effective system design must be in accordance with the end users' needs and to best assist users involved in cross-language information retrieval we need to understand their behaviours and the search problems they face. Petrelli's study of users involved in CLIR presented a number of interesting findings. In particular, the users preferred the interface which hid the translation and that language knowledge and sight of the translation affected search behaviour.

Our study based on retrospective thinking aloud revealed a complex picture of the influence (or not) of language skills and confidence therein and of perceptions of the role of the multilingual interface, language and translations in image retrieval. Most revealing and of potential interest to future study of users of CLIR is the finding that less than half of our users appeared to consider identification of the language to be essential in retrieving the image. The majority either lacked confidence in using different languages or were so focused on finding the given images and completing the task that were not thinking at all about languages. Indicative of this was the comment *"...completing the task successfully. What was success for me? That you find the image. In any way I possible could. I was not focusing on translations...I thought my task is to find that image and I will do whatever I could to find it"*. Only four users attempted to identify the language of the images from its context (or from the Hint feature) and use it to their benefit. Of those for which languages played a significant role in the process of identifying keywords to search for the images, the translations were judged to be poor as either the translations were not coming up, were not corresponding to the users' keywords or were judged to be resulting in the retrieval of 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. Some were treating the multilingual interface as a translator, trying to isolate specific words, translate them on the multilingual interface and use the translations to retrieve the images on the monolingual interface. Another user stated that: *"I think I just saw it as a translation tool and not as an integrated translation thing that already was retrieving images. I did not really use it in this way because in my mind, it was only translating my keywords"*.

One of the initial aims of this study was to look in greater detail at how working with the translations affected search behaviour with regards to the actual search terms entered by users. Unfortunately, this study could not reach a conclusion because only four out of ten users used the translations and this was in a way not anticipated. Our study did reveal many reasons for non use of the multilingual interface, ranging from a lack of confidence in languages to a lack of trust in the system translations to a disregard for the need to search in other languages. However a further factor which may have influenced little use of translations is the interface design in presenting to the user how the multilingual interface worked or how the translations could be used to benefit the search. The feedback we obtained from the users suggested a variety of reasons for using the multilingual interface other than to address a recognized need to

search for the image in another language. This may suggest that the purpose of the multilingual interface was not clear to the users as does the observation that it was used as a translator tool to run both the search terms and the translations in the search box.

On the whole, it would appear that users were so focused on completing the task, “obsessed” (as a user stated) of finding the images that even from the beginning of the task, they were not thinking really which interface they are going to use and for what reason. Even users who were concerned about languages, at the end of the task, also admitted that they were not paying much attention to the interfaces because they thought that it was not making any difference.

## 5 Conclusion

This study aimed at investigating the users’ image seeking behaviour when retrieving a known, non-annotated image in Flickling. In particular, we identified the reasons why two of our users were choosing to search only on the monolingual interface and the eight switching between interfaces. We demonstrated that only four users were thinking about languages when trying to retrieve the given images while the rest of our users were more preoccupied with finding the images and completing “successfully” the task. Consequently, we showed that only these four users were paying attention to translations provided by the system. These stated that translations were not helpful or they were not making much difference in finding the given images since the results were irrelevant to what they were looking for.

This small study has also shown that if we are to ask whether a CLIR system should display query translations or not, then the answer is no. Our users were either not interested in the translations or found them to be poor. However taking the findings to such conclusion would be foolhardy given the complexity of the activity highlighted in the users’ comments that they were so engaged in finding the image that language or translations played little or no part. Rather than reaching firm conclusions, this small study has suggested the need for more research into users’ search behaviour with translations (and in image retrieval) if we are to design CLIR systems which will not place additional or unnecessary cognitive demands on the user and will support effective search behaviour and performance.

## References

1. Gonzalo, J., Clough, P., Karlgren, J.: Overview of iCLEF 2008: Search Log Analysis for Multilingual Image Retrieval. In: Peters, C., et al. (eds.) CLEF 2008. LNCS, vol. 5706, pp. 227–235. Springer, Heidelberg (2009)



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)