UAIC at iCLEF 2009: Analysis of Logs of Multilingual Image Searches in Flickr

Florin-Tudor Cristea, Vlad Alexa, Adrian Iftene

UAIC: Faculty of Computer Science, "Alexandru Ioan Cuza" University, Romania
{florin.cristea, vlad.alexa, adiftene}@info.uaic.ro

Abstract. In this paper, we summarize our analysis of the large log of multilingual image searches in Flickr provided to iCLEF 2009 participants, focusing on the UAIC group.

Categories and Subject Descriptors

H.3 [**Information Storage and Retrieval**]: H.3.1 Content Analysis and Indexing; H.3.3 Information Search and Retrieval; H.4 [**Information Systems Applications**]: H.4.m Miscellaneous

General Terms

Interactive Information Retrieval, Cross-language Information Retrieval

Keywords: iCLEF, Flickr, Log Analysis, Multilingual Image Search, User Studies

1 Introduction

iCLEF[1] is the interactive track of CLEF started from 2006, an annual evaluation exercise for Multilingual Information Access systems. From year to year iCLEF has changed collections in order to explore user behavior in scenarios where the necessity for cross-language search come more naturally for the average user. The main question is how best to assist users when searching information written in unknown languages, rather than how best an algorithm can find information written in languages different from the query language.

In iCLEF task the user receive an image (not annotated) from Flickr[2] and the goal is to find the image again from Flickr using *flickling*[3], the search engine provided by iCLEF organizers. The user does not know in advance in which languages the image

---

[1] iCLEF: http://nlp.uned.es/iCLEF/

[2] Flickr is available at http://www.flickr.com.

[3] Flicking interface: http://cabrillo.lsi.uned.es/flickling

is annotated. Therefore, searching in multiple languages is essential to get optimal results.

The task is organized as an online game: the more images you find, the higher you will be ranked. In case of ties, the ranking will also depend on precision (number of images found / number of images attempted) and search time spent in finding the images (less time is better).

The structure of the rest of the paper is as follows: Section 2 describes our group statistics; Section 3 describes issues regarding cross lingual facilities and about post image search questionnaires. In last Section we summarize our participation in the track and give some conclusions about the experience.

2 Log Processing and Characterization of the Search Sessions

In order to make good use of the raw data provided by the *flickling* logs, two processing steps were applied. Thus, the flat files were passed through a Python-based script that cleaned up any possible format abnormalities, followed by the actual import in a SQLite database[4]. From this point, statistical research resumed to a series of standard SQL queries.

For the UAIC group there was a total of **1532** assigned images -- from which the **31** existing users found **1236** and gave up on 296 (though only 39 times explicitly) - therefore 80.67% of the referred challenges were solved.

2.1 Assigned Images

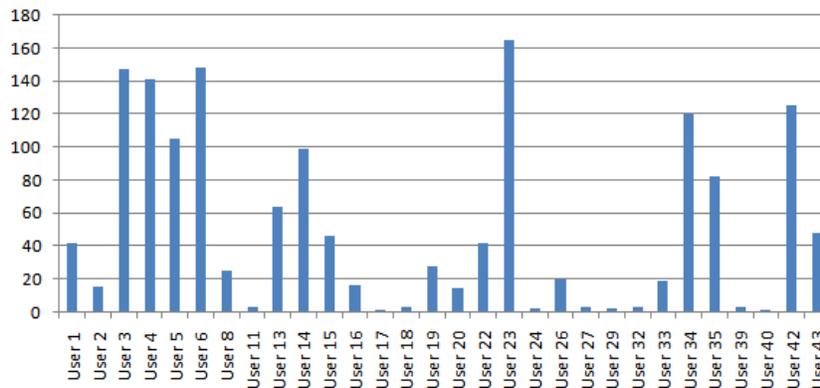A comparison of user performance yields the subsequent graphic:

Figure 1: Assigned Images per User

---

[4] Database structure: 1 table (*flickling*) consisting of 7 columns (*recordTimestamp*, *userId*, *sessionId*, *methodName*, *methodDetails*, *originalTokens*, *translatedTokens*), all of type TEXT.
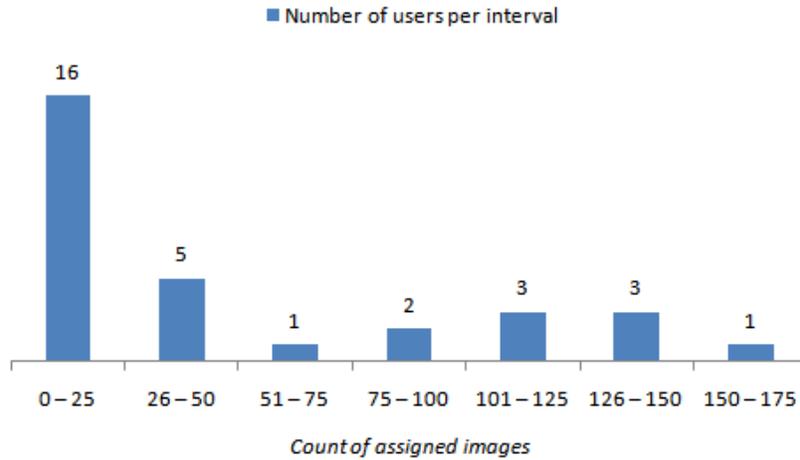
Figure 2: Assigned Images Discretization

While there are present several over-achievers, the average count points to a much lower value, suggesting that some users merely experimented with the system before abandoning it.

2.2 Found and Give-Ups Images

In what concerns *found images*, the global count adds up to **1236** items. The following situation is obtained via analyzing the group's success rate:
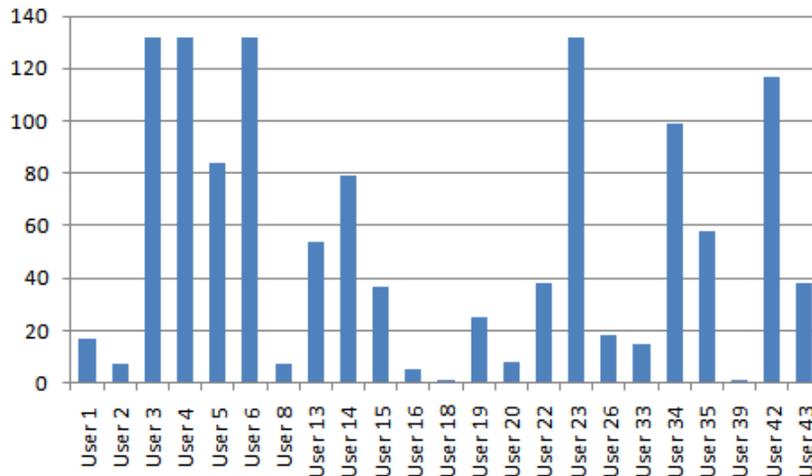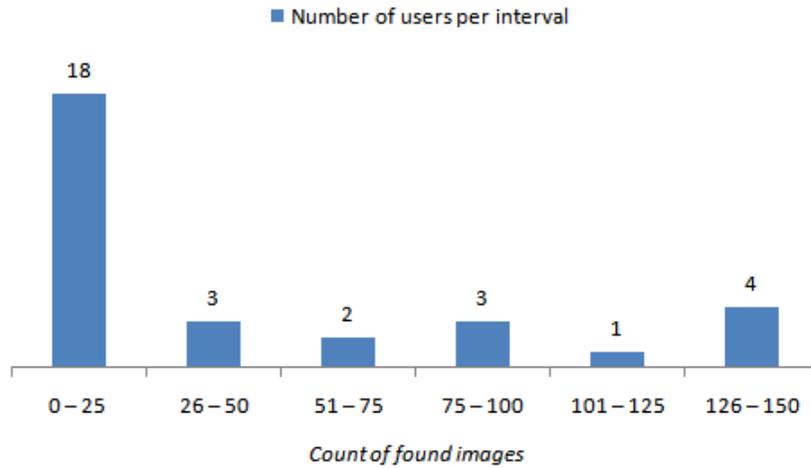

Figure 3: Found Images per User

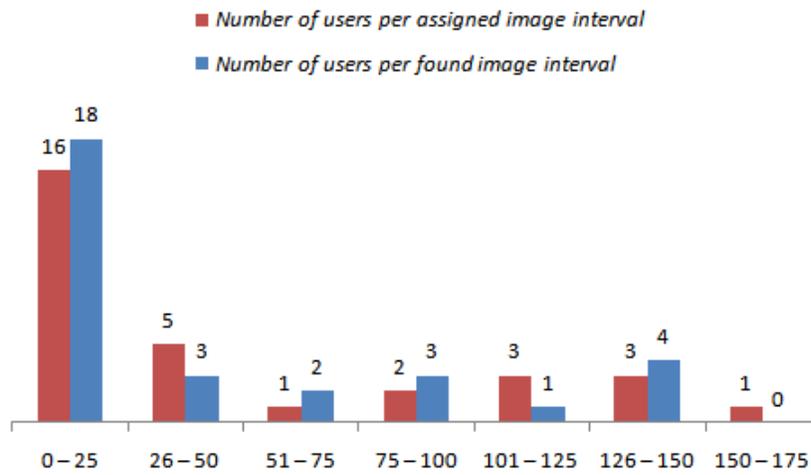**Figure 4**: Found Images Discretization

Figure 5: Assigned vs. Found

Again, the numbers are pushed up by a few ambitious players, though at least in some intervals an increase in interest can be observed.

For the UAIC group there were a total of **39** photos effectively given up on.

2.3 Dictionary Adds

Our users made good use of the personal dictionary feature, with a total of **1349** additions. The underneath figure illustrates specific usage:
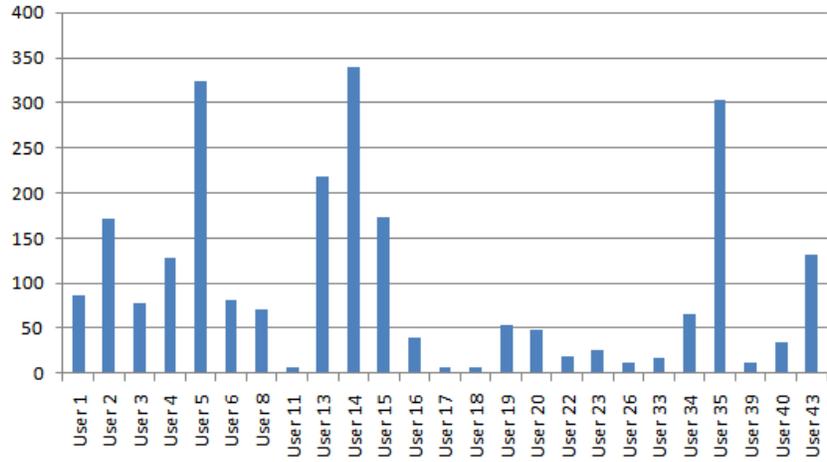
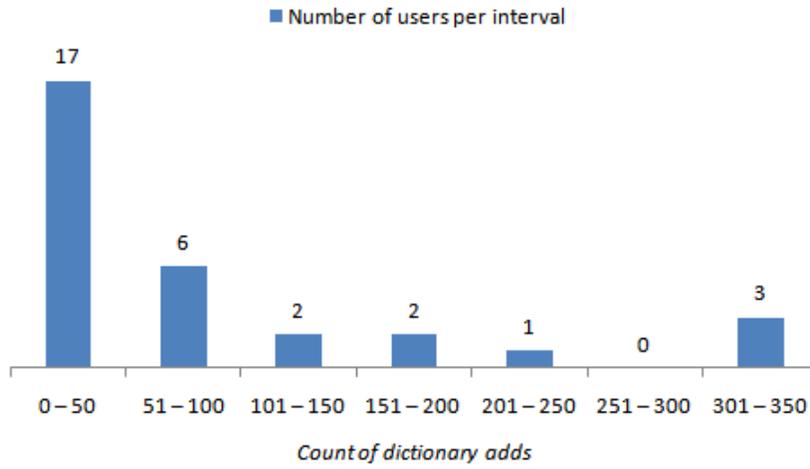

Figure 6: Dictionary Adds per User

Figure 7: Dictionary Adds Discretization

No direct connection between usage of this feature and higher scores can be made (in fact the top achievers seem to have rarely used it at all).

2.4 Distinct Searches

During the competition our users launched a number of **6591** distinct searches (counting **5442** unique terms).
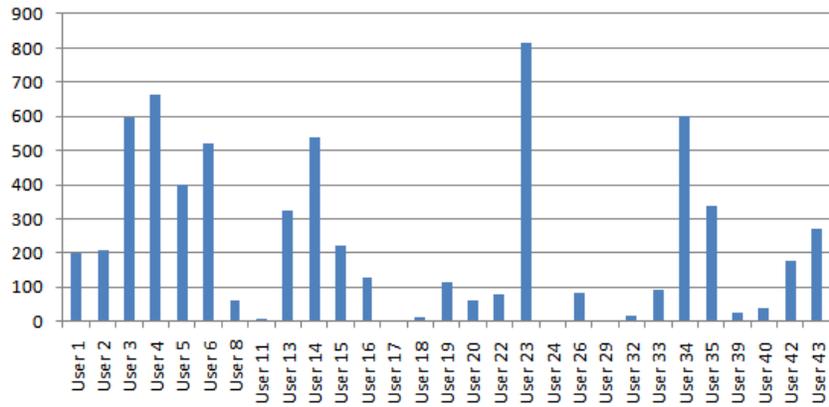

Figure 8: Distinct Searches per User

2.5 Search Duration

Except for 11 users with average search times of over 30 minutes, the rest of 20 users had closely distributed typical values (figure 9). From our analysis we noticed how users with lower times frequently requested hints in order to identify the target picture, while users with higher time spans used more variants for searches in different languages and in the end obtained the highest scores.
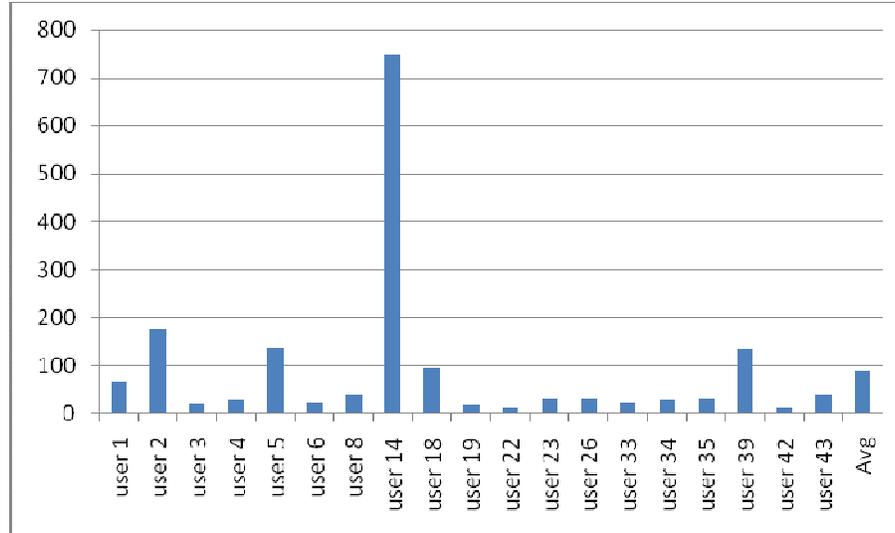
Figure 9: Average of Search Duration per User

2.6 User Highlight

We focus our attention, for this case study, on the 4-th user from the UAIC group. The performances of this user are as follow:
- 132 pictures found;
- 141 pictures assigned;
- 662 queries launched from which 294 where cross-lingual queries;
- 128 adds to the personal dictionary.

The success ratio of user 4 is 93.61%. We chose this user because he is the winner of the 2-nd CLEF Flickr Challenge 2009. Andrei Zapodeanu has obtained the number 1 top score at the Flickr competition.

3 Other issues

3.1 Usage of Specific Cross-Lingual Refinement Facilities

For the UAIC group there were a total of **3016** distinct cross-lingual searches during the competition (counting **2714** unique terms) - 45.75% of all queries.
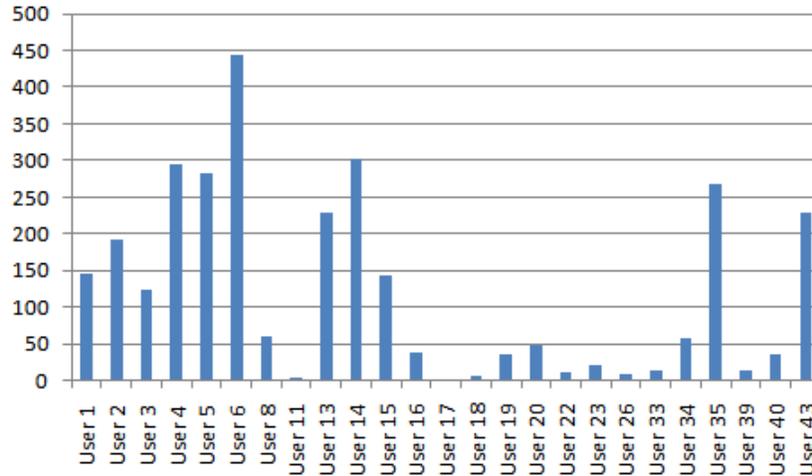
Figure 10: Distinct Cross-lingual Searches per User

3.2 User Questionnaires

After the search process users received two questionnaires: one is presented after each search session (in two forms: one if the search failed and another one if the search succeeded), and another one is presented only once at the end, when the user has performed all search sessions.

3.2.1 Post-Session Questionnaires

The statistics of post-session questionnaires are presented in Figure 11 (after success search) and Figure 12 (after failed search). In case of success, the task is perceived as easy in 298 cases, and hard in 1171 cases.
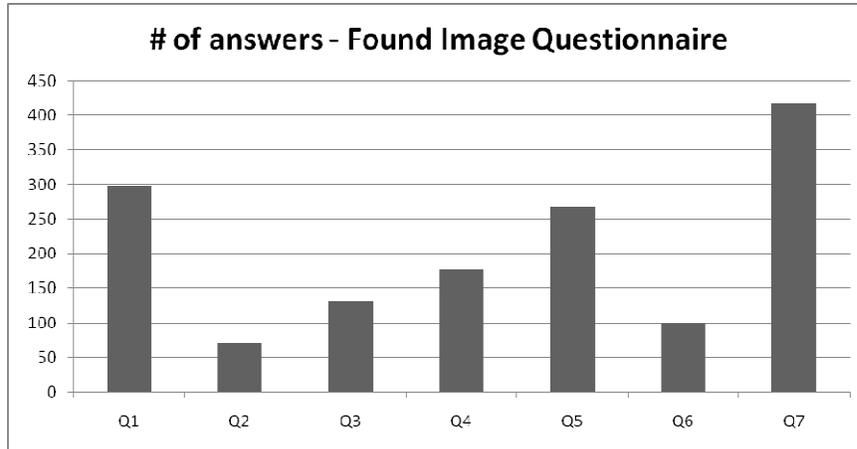
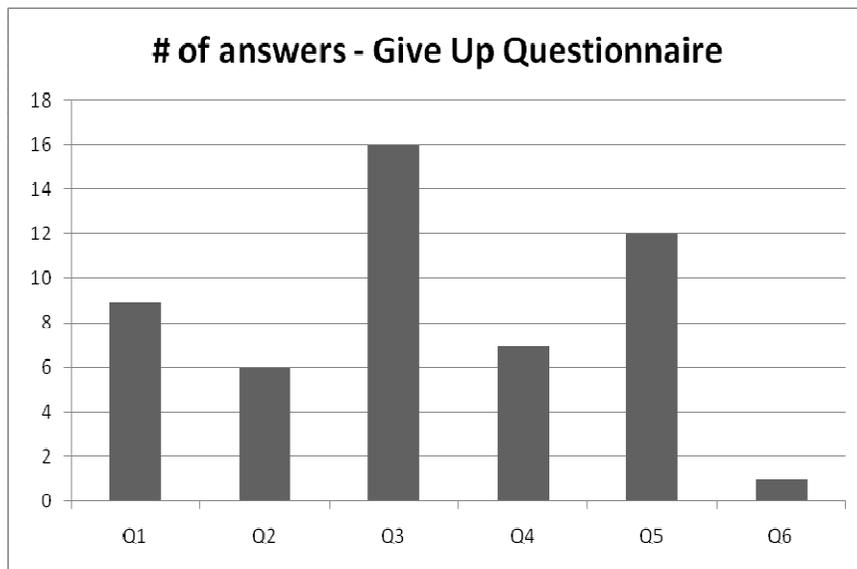**Figure 11**: Post-Image Questionnaires after Finding an Image

Figure 12: Post-Image Questionnaires after Giving up an Image

4 Conclusions

Seemingly, no clear connection between the results of the over-achieving users and their particular actions can be traced. As we mentioned before, at least no correlation exists between the number of dictionary adds and the success rate. Even so, one previously unexplored subject remained, namely social networking between the partici-

pants, how they perhaps exchanged solution ideas. In order to verify this, we focused on the top four participatns, that is to say number users 3, 4, 6 and 23. What we wanted to find out was how many times exactly the same search terms were used by multiple players. Restricting the count to groups of words searched at least twice by at least two distinct users from the four specified above, we obtained a value of 198. Further increasing the limits to a minimum of three search occurrences by no less than three users from the previously enumerated ones, we got a value of 67.

The corresponding amounts for the rest of the population were 292 and 114. Thus by this criterion 40.4% respectively 37.01% of these overlapping searches belong to the over-achieving users, who represent only 12.9% of the entire UAIC group, evidence pointing towards a possible, though limited, collaboration.

Even so it is quite unlikely that this factor proved decisive in the long run, ultimately the users' language skills tilting the balance.

In this paper we tried to show the performances of our group by calculating and creating different reports on each category of the Flicking competition (number of images found vs. number of images assigened, succes rate, the answers given by our users to the specific forms etc.). Also in section 2.6 we presented the statistics of the winner of the competition, Andrei Zapodeanu, member of the UAIC group.

For any other information, such as the complete list of queries, the toolkit used, or any other thing related to this paper, please contact us by e-mail.