"Interactive" Undergraduate Students: UNIPD at iCLEF 2008

Giorgio Maria Di Nunzio

Department of Information Engineering – University of Padua
Via Gradenigo, 6/a – 35131 Padova – Italy
dinunzio@dei.unipd.it

Abstract. This is the first year of participation of the University of Padua to the interactive CLEF track. A group of students of Linguistics of the Faculty of Humanities were asked to participate in the experiment. An analysis of the questionnaires together with some log analysis is carried out with the aim of studying: the interaction of the user with a cross-lingual system, the solutions they find for a given task, and the tools that a system should provide in order to assist the user in the task.

1 Introduction

The CLEF interactive track (iCLEF)[1] has been conducted since 2001 in the context of the Cross Language Evaluation Forum (CLEF)[2] with the aim of studying the interaction with a multilingual information retrieval system from a user point of view. Since 2005, the iCLEF has shifted the focus from the search for textual documents to the search of images [1]. This year, the iCLEF 2008 [2] is focused on the known-item image retrieval based on the Flickr[3] database of images using the Flickling[4] the search interface.

The University of Padua (UNIPD) participated in this track for the first time. In order to have a large number of users, students of Linguistics of the Faculty of Humanities were asked to participate in the game. Participation was not mandatory; nevertheless, an incentive was given in order to convince them to play: extra points to an exam as a reward of their effort. The possibility to have these students was important for the aim of this study since these are users who use different languages every day.

The paper is organized as follows: Section 2 presents briefly the Flickling system and how the scores are computed; Section 3 describe the group of students which have been asked to participate in the experiment. In Section 4 the analysis of the questionnaires gathered on-line and off-line is presented. Section 5 shows the analysis carried out on the logs of the Flickling system. Final remarks and comments are presented in 6.

[1] http://nlp.uned.es/iCLEF/
[2] http://www.clef-campaign.org/
[3] http://www.flickr.com/
[4] http://soporte1.lsi.uned.es/flickling/

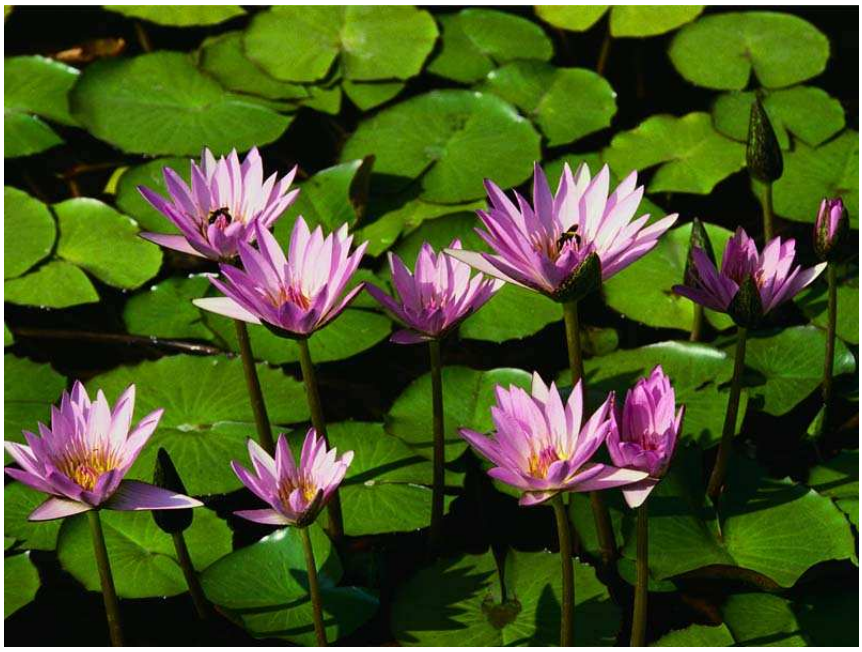

Fig. 1: An example of a possible image given by the Flickling system.

2 Flickling: the Game

In this section, we want to summarize the main part of the Flickling [3] system in order to help the reader to understand the analysis of Section 4 and Section 5.

The search interface provided by iCLEF organizers is a basic cross-language retrieval system for the Flickr image database, presented as an online game: the user is given an image, and he must find it again without any a priori knowledge of the language (one or more) in which the image is annotated. For example, in Figure 1 we see a picture of flowers, water lilies. The system allows you to make a text search like "flowers water lily", and the system automatically retrieves those pictures which have those description tags. If the image cannot be found, the user can reformulate the query in the same language, in a different language, or use the cross-lingual interface of the system. If the user finds the image, 25 points are earned otherwise the user can give up the search and go to the next image. The user can also ask for hints for the search, each hint costs 5 points; at most 5 hints can be asked (with 1 hint the score goes down to 20, with 2 hints to 15, and so on).

At the end of each search, a questionnaire is shown to the user to ask him how easy/hard it was to find (or not find) the image.

3 Flickling: The Gathering (of UNIPD Students)

The users involved are students from the Faculty of Humanities of the University of Padua, of the course of "Linguistics and Modern Cultures" and "Languages for Cultural Mediation". During the first year of their career, students have a class of "Informatica Generale", a basic course on Computers and Computer Science, and in the context of this class they were asked to participate in this game. They were free to participate and interact as long as they wished; however, an incentive (extra points to the final exam) was put in order to persuade them to use some more of their spare time. There were also a prize for to the student with the highest score. Given this particular situation the students were asked not to cheat and follow this simple rule:

- for the first game, they had to register under the group of "University of Padua - Linguistics";
- if they wanted to play again and improve their score, they had to register under the group of "University of Padua 2 - Linguistics".

Therefore, results of this second group are highly biased by the fact that these students had already played and knew many of the keyword already used to find the pictures. This group will not be considered in the analysis.

The number of students of this course was around 250, students who regularly attended the lessons were around 120. At the end the number of students who participated in the experiment was 60 which was surprusingly high. Consider also that the students are not familiar with search engines, and only two of them knew Flickr before the game started, just a reminder that Flickr is much less known than some other Web entertainments (i.e. YouTube, Facebook).

3.1 Language skills

This group of students who participated in the experiment were interesting to study since their linguistic skills. They are likely good at the evaluation of the quality of the translations and the suggestions of the possible translations of the words to describe the picture. Moreover, these students had a different levels of preparation on different languages. It was not possible to track all the levels of knowledge for each student, however we can roughly divide the students in the following overlapping groups:

- the main mother tongue language is Italian;
- the majority of students study English and/or Spanish;
- German, French and Portuguese are usually the second, or third, language chosen for studies;
- a minority of students study eastern country languages, such as Russian, Greek, or Slavic languages.

It is also important to underline that within the students who participated there were foreign students, not Italians, which complete sample of skills among the students.

4 Questionnaire Analysis

During the Flickling game, there are questionnaires that users have to fill-in. Questionnaires are shown:

- at the end of the search of each image. There are two types of questionnaire: the found image questionnaire (when the image is found), and the give up questionnaire (when the user decides to skip the image because it was not possible to find it);
- after a certain number of images: the overall questionnaire, which asks general questions about the whole cross-lingual task, the interface, and possible improvements.

4.1 A Questionnaire for each Image

There are two types of questionnaiers which are shown at the end of the search for an image: the image found questionnaire, and the give up questionnaire. There are six possible answers for the first questionnaire, and five possible answers for the second one. The analysis of these two questionnaire aims to provide insights about the differences between the group of UNIPD and the other users.

Found Image Questionnaire For the found image questionnaire, in the logs there are 1,607 records for UNIPD and 1,993 records for the others. In the following tables we maintain the same enumeration given to the possible answers (from zero to six). The distribution of the answers are shown in Table 1 and Figure 2. The distribution is similar for both groups, however there are some interesting points to highlight.

The possible answers to the question "What problems did you encounter while searching for this image?" are:

zero: It was easy: this answer got the highest value for both groups; in the case of UNIPD the value reaches 55% (which means that in 55% of the cases in which the image was found, it was an easy task), while for the others is about 36%.

one: It was hard because of the size of the image set: 10% of the answers of UNIPD confirm this, while it is 20% for the others.

two: It was hard because the translations were bad: this answer gets 6% to 10% according to the UNIPD group and the others.

three: It was difficult to describe the image: this is the second highest answer in terms of positive response (users who said this claim was true). 21% of UNIPD and 17% of the others tell that even though the image was found, the search was hard because the descritpion was complicated.

four: It was hard because I didn't't know the language in which the image was annotated: 6% of the answers of UNIPD consider this fact as true, while it is 16% for the others.

five: It was hard because of the number of potential target languages: around 2% and 3% for the two groups.

	zero	one	two	three	four	five	six
UNIPD	887	172	102	345	98	31	28
Others	723	418	215	456	331	59	168

Table 1: Found image questionnaire: number of answers for each question. The total number of questionnaires compiled by UNIPD is 1,607 while by the other groups id 1,993.

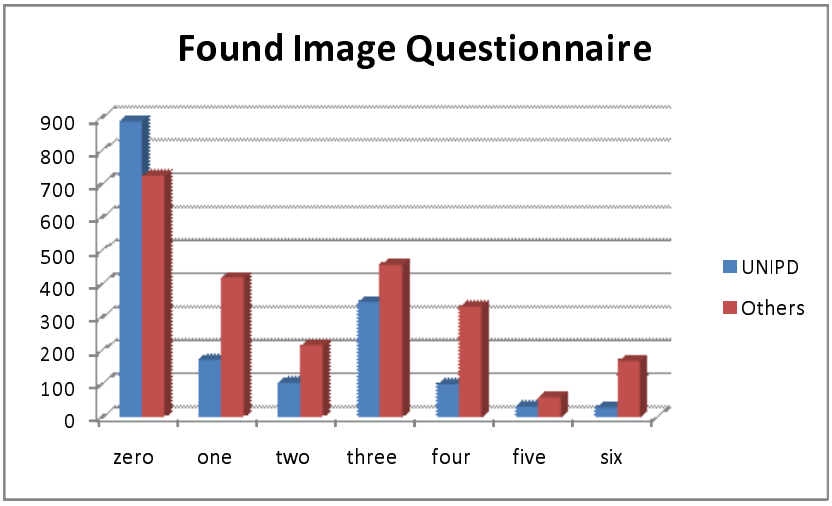

Fig. 2: Found Image Questionnaire. Histogram based on the data of Table 1.

six: It was hard because I needed to translate the query: 2% for UNIPD while it goes up to 10% for the others.

Give Up Questionnaire For the give up questionnaire, there are 479 records for UNIPD and 516 records for the others. Like the found image questionnaire, the same enumeration of the questionnaire is mainained to present the results (answers from zero to five). The distribution of the answers are shown in Table 2 and Figure 3.

The possible answers to the question "Why are you giving up on this image?" are:

zero: There are too many images for my search: this answer was higher for UNIPD with 42%, while the others reached 30%.

one: The translations provided by the system are not right: 8% of the answers of UNIPD confirm this, 13% for the others.

two: I can't find suitable keywords for this image: this is the highest reponse from the group of the others with 52% of positive answers, while it is about 38% for UNIPD.

	zero	one	two	three	four	five
UNIPD	204	38	184	16	59	20
Others	156	66	266	33	86	43

Table 2: Give up questionnaire: number of answers for each question. The total number of questionnaires compiled by UNIPD is 479 while by the other groups id 516.

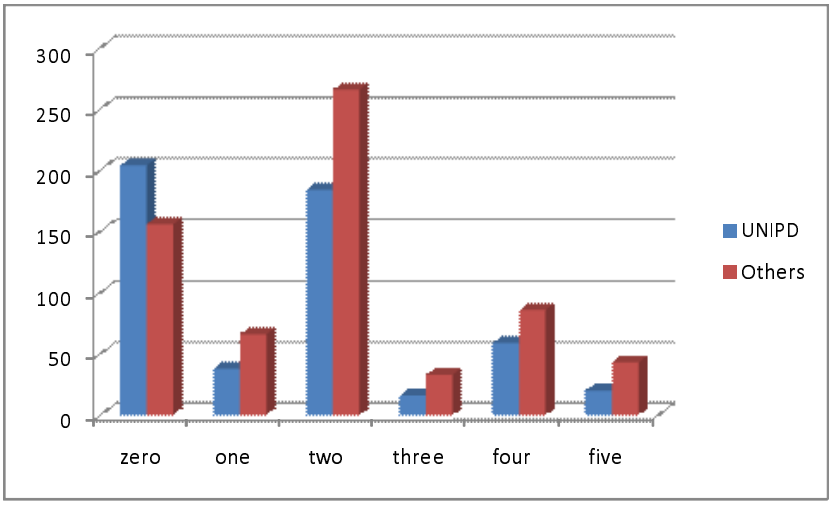

Fig. 3: Give up Questionnaire. Histogram based on the data of Table 2.

three: I have difficulties with the search interface: 3% to 6% according to UNIPD and the others respectively.

four: I just don't know what else to do: 12% (UNIPD) to 17% (others) stop the search for this reaason.

five: Other (please, comment below): around 4% and 8% of images where user gave up were commented.

4.2 Overall Questionnaire

The overall questionnaire consists of 27 single-choice answers plus 2 open questions. The analysis presented here compares the answers of the UNIPD group to the answer of the rest of the participants who filled-in the questionnaires. For UNIPD, the questionnaires were gathered both during the on-line game and off-line during the final exams, one month after the end of the game. In particular:

- 27 students filled in the on-line questionnaire;
- 17 students filled in the questionnaire on paper during the exams.

It has to be noted that the students who filled in the quetionnaire on paper are not all the same students of the on-line questionnaire. It was not possible

	frequently	sometimes	rarely	never
UNIPD	11	24	8	1
Ohters	9	17	8	2

Table 3: Question 1A.

	frequently	sometimes	rarely	never
UNIPD	2	24	15	3
Ohters	10	8	14	4

Table 4: Question 2A.

to know exactly who did what (some students even forgot if they had done this questionnaire), but we can rouglhy say that the overlap between the two groups is less than 10 people (which means that less than 10 people filled in both questionnaires). The number of quesionnares filled in by other users is 36.

In the following, we analyze the answers for each questions and carry out a t-test (significance level at 5%) between the two groups UNIPD and the rest of the participants (remind that the group "University of Padua 2 - Linguistics" is not taken into account). Questions are identified with a number and a letter.

1A: Do you need to search information in foreign languages in your daily life? Table 3 shows the answers for the two groups. The use of foreign languages in daily is important for both groups, only a small part (between 20% and 25%) considers the use of other languages for searching information as less important.

The t-test confirms that the answers of the two groups are not different.

2A: Do you often use image search facilities? Table 4 shows the answers for the two groups. In this case, we can split the answer in two: people who use and those who do not use search facilities. Among those people who use seach facilities, UNIPD students claim to use them from time to time, while the others say that the use is frequent. This last result is probably biased by the fact that the majority of the positive answers were given by the dzoom group, a group of photography fans.

The t-test confirms that on average the answers of the two groups are not different.

3: The search task you performed was: Table 5 shows the answers for the two groups divided by subquestions. The possible subquestions were: clear (3A), easy (3B), familiar (3C), interesting (3D), relevant to you (3E). We can draw some general results which are common for both groups: the search task was clear, not so easy (half of the people agree and the other disagree on that). The search task is not so familiar for the UNIPD group while more familiar to

3A	strongly agree	agree	disagree	strongly disagree
UNIPD	4	35	4	1
Ohters	7	22	6	1
3B	strongly agree	agree	disagree	strongly disagree
UNIPD	3	19	21	1
Ohters	1	14	15	6
3C	strongly agree	agree	disagree	strongly disagree
UNIPD	4	21	18	1
Ohters	1	24	9	2
3D*	strongly agree	agree	disagree	strongly disagree
UNIPD	17	24	2	1
Ohters	9	17	8	2
3E	strongly agree	agree	disagree	strongly disagree
UNIPD	5	25	11	3
Ohters	2	12	18	4

Table 5: Question 3A (clear), 3B (easy), 3C (familiar), 3D (interesting), and 3E (relevant to you). The star sign "*" indicates a significant statistical difference between the two groups.

	strongly agree	agree	disagree	strongly disagree
UNIPD	9	28	6	1
Ohters	6	17	12	1

Table 6: Question 4A

the others group. It is more interesting for the UNIPD than the others, with a significant statistical difference here (p-value = 0.028). The search task results relevant for half of the user and less relevant for the other half.

4A: Did you find multilingual search capabilities useful to find images in Flickr? Table 6 shows the answers for the two groups. There is a general agreement on the usefulness on multilingual search capabilities. A bit less for the group of the others.

The t-test confirms that on average the answers of the two groups are not different.

5A: Would you now prefer to use a multilingual search facility for your own image searches? Table 7 shows the answers for the two groups. The answers are similar to question 4A: UNIPD users are more willing to use multilingual search facilities respect to other users.

The t-test confirms that on average the answers of the two groups are not different.

	strongly agree	agree	disagree	strongly disagree
UNIPD	6	32	5	1
Ohters	9	17	10	0

Table 7: Question 5A

6A	strongly agree	agree	disagree	strongly disagree
UNIPD	7	25	9	3
Ohters	12	13	7	4
6B	strongly agree	agree	disagree	strongly disagree
UNIPD	13	27	3	1
Ohters	15	14	5	2
6C	strongly agree	agree	disagree	strongly disagree
UNIPD	6	26	11	1
Ohters	14	15	5	2
6D	strongly agree	agree	disagree	strongly disagree
UNIPD	11	22	10	1
Ohters	16	14	5	1

Table 8: Question 6A, 6B, 6C, and 6D.

6: Which, in your opinion, are the most challenging aspects of the task? Table 8 shows the answers for the two groups divided by subquestions. Question 6 contains interesting subquestions which need to be analyzed one-by-one.

6A: Finding the correct terms to express an image in my own native language: there is a strong agreement on the fact that it is difficult to find the correct terms even in the native language of the user.

6B: Selecting/finding appropriate translations for the terms in my query: like question 6A, there is a general consensus on how hard it is to find good translations for the term in the query.

6C: Handling multiple target languages at the same time: in this case it seems that the UNIPD users are more willing and able to manage different languages than the other group. This makes sense since this group is made by students who study languages. The output of the t-test confirms that on average the answers of the two groups are not different; however, there is a tendency (p-value = 0.097) to have a different answer between the two groups.

6D: Finding the target image in very large sets of results: like question 6A and 6B, in general finding images given a large set is a difficult and challenging task.

7: Which interface facilities did you find most useful? Table 9 shows the answers for the two groups divided by subquestions. Each subquestion is analyzed individually. All the subquestions show no statistically significant difference between UNIPD and the others.

7A	strongly agree	agree	disagree	strongly disagree
UNIPD	13	22	8	1
Ohters	7	17	11	1
7B	strongly agree	agree	disagree	strongly disagree
UNIPD	5	28	9	2
Ohters	4	21	9	2
7C	strongly agree	agree	disagree	strongly disagree
UNIPD	10	18	10	6
Ohters	7	13	10	6
7D	strongly agree	agree	disagree	strongly disagree
UNIPD	4	22	13	2
Ohters	7	17	8	4

Table 9: Question 7A, 7B, 7C, and 7D.

7A: The automatic translation of query terms: general agreement on the fact the automatic translation is a positive facility of the interface, a little bit more useful (strongly agree) for UNIPD than the others.

7B: The possibility of improving the translations chosen by the system: agreement (less strong than question 7A) on the possibility to improve the translations offered by the system.

7C: The additional query terms suggested by the system (You might also want to try with...): in this case the answers are spread all over the four possible answers. There is a small preference on this facility; nevertheless it is a feature that may not be useful in general.

7D: The assistant to select new query terms from the set of results: the majority of users agree on the positive aspect of this feature of the interface. However, a big percentage of users (about 35% in both cases) do not find the assistant useful.

8: Which interface facilities did you miss? Table 10 shows the answers for the two groups divided by subquestions. Each subquestion is analyzed individually. All the subquestions, except for 8E, show no statistically significant difference between UNIPD and the others.

8A: Detection and translation of multiword expressions: in this case users are split in two groups: people who like this feature and people who do not like it. Users of the UNIPD group are evenly distributed, while there is a small positive preference for the others.

8B: Bilingual dictionaries with a better coverage: there is a strong positive response regarding the use of bilingual dictionaries. In both cases a number of users expressed a strong preference on this item.

8C: A system able to select the translations for my query terms better: in this case, the group of the others prefer more this type of selection of the translation of terms compared to UNIPD.

8A	strongly agree	agree	disagree	strongly disagree
UNIPD	3	20	19	2
Ohters	3	18	9	1
8B	strongly agree	agree	disagree	strongly disagree
UNIPD	11	22	9	2
Ohters	13	15	7	1
8C	strongly agree	agree	disagree	strongly disagree
UNIPD	6	25	12	1
Ohters	10	20	5	1
8D	strongly agree	agree	disagree	strongly disagree
UNIPD	5	24	15	0
Ohters	9	18	7	2
8E*	strongly agree	agree	disagree	strongly disagree
UNIPD	7	22	15	0
Ohters	18	13	4	1
8F	strongly agree	agree	disagree	strongly disagree
UNIPD	3	22	17	2
Ohters	9	16	8	3
8G	strongly agree	agree	disagree	strongly disagree
UNIPD	4	24	15	1
Ohters	10	15	10	1

Table 10: Question 8A, 8B, 8C, 8D, 8E, 8F, and 8G. The star sign "*" indicates a significant statistical difference between the two groups.

8D: More support to decide what the possible translations mean and therefore which ones are really appropriate: same result of question 8A: users split in two groups, with a small preference for this feature for the group of the others.

8E: The possibility to search according to visual features of the images (search images that look like this, search only B/W images, search only for dark images, etc.): at first glance, results may be similar to question 8A and 8D. However, the statistical test shows that there is a significant difference in the answers of the two groups. Therefore, it is less important for the group of UNIPD to use visual features for the search of images.

8F: The classification of search results in different tabs according to the image caption language(s): again, a situation similar to question 8A and 8D. More positive the answer of the others, more evenly distributed amonf the group of UNIPD.

8G: An advanced search mode giving more control on how Flickr is queried: the same consideration of question 8A, 8D, and 8F.

9: How did you select/find the best translations for your query terms? Table 11 shows the answers for the two groups.

9A: Using my knowledge of target languages whenever possible: all the users use mainly their knowledge to translate the query in the target language. There is no distinction between UNIPD and the others.

9A	frequently	sometimes	rarely	never
UNIPD	28	12	4	0
Ohters	19	14	3	0
9B	frequently	sometimes	rarely	never
UNIPD	4	15	10	15
Ohters	10	9	7	10
9C	frequently	sometimes	rarely	never
UNIPD	5	9	15	15
Ohters	2	14	10	10

Table 11: Question 9A, 9B, and 9C

9B: Using additional dictionaries and other online sources: in this case the answers are spread all over the possible choices. However, it is less frequent for the users of UNIPD compared to the others.

9C: I did not pay attention to the translations, I just trusted the system: interesting to note, the user do not trust the translation of the system. It is more true for the group of UNIPD compared to the others.

5 Log Analysis

The logs made available for studying the actions of each user and were released as a text file. Each row of the file contains either an action of the user or an action of the logging system. The log goes from April 24th 2008 until June 16th 2008 for a total of 1,483,806 recorded actions. For the purpose of the analyisis and for a more convenient management, this file was loaded into a table of a PostgreSQL[5] database. The records were also cleaned, in the sense that some of the actions recorded were not useful for the analysis. The following actions were removed:

- register, remindPwd;
- login, login2, logout;
- getUserInfo;
- getTargetImg, all except for "new ActiveSearch created photoid xxxxxx" which starts a new search;
- pauseTime, playTime;
- some other actions like: search (using Flickr's API), clsearch (query not found in DB, let's use Flickr's API), log (click on target image) and few others.

At the end the log was reduced to 1,139,339 records. A view on the main table was made to focus only on the actions of the group of UNIPD (432,813 actions).

In Table 12 the list of the participants is shown, ordered by the number of users per group (last column), with the respective total score and the average score per user. The UNIPD group had the highest total score, and one of

[5] `http://www.postgresql.org/`

the highest average scores per user. In the following sections, the scores of the UNIPD users are studied in order to understand whether there are different in the strategies among users, how many hints have been requested, how many times a cross-lingual search has been performed and so on.

5.1 University of Padua - Linguistics

The number of students of the University of Padua who participated in the experiment was 60, the largest group of the iCLEF 2008. Since students were asked to enjoy the image search and not to reach a specific score, there were students who gave it a try and left after few images (sometimes only one) while others tried to finish all the set of available images, and played more games under the second group "University of Padua 2 - Linguistics".

Table 13 shows the top scorers of UNIPD. For each user the table lists the number of images viewed, the images found and the images skipped. Only ten people finished the whole set of images available, 103. In the following subsections, we try to understand if there is any correlation among the scores and the strategies.

Found or Skipped? In Figure 4 scores with respect of the number of images viewed and found are plotted. There is an obvious positive correlation between the scores and the number of images (the more images, the higher score), however there are differences which can be underlined, for example the scores versus the number of images found are more scattered and some differences among top scorers can be appreciated. This plot also tells that when a comparable number of images are found among different users, the fact that there may be differences in scores is that hints are used more frequently from one user than another. For example, user_07 found all the 103 with a score of 1,280 while user_50 the top scorer found only 94 images. This means that user_07 asked many hints which penalized his score.

Hints and Clues In Figure 5 the highest scores of UNIPD and the other best participants are shown with respect to the average number of hints asked per image. This plot shows that the best participants, in terms of scores, used on average about 2 hints per image. Now, it is important to understand what the first hint is: when you ask for the first hint, the system tells you in what language the image you are searching for is described. This means that, on average users needed to know in advance the language of the description of the image before finding it.

Monolingual or Multilingual? In Table 14 and Figure 6 the average of monolingual and cross-lingual searches are shown for the top scorers. It is not easy to find regular patterns in the behavior of the users. On average, the top scorers use from 5 to 6 monolingual searches per image, and from 6 to 7 cross-lingual

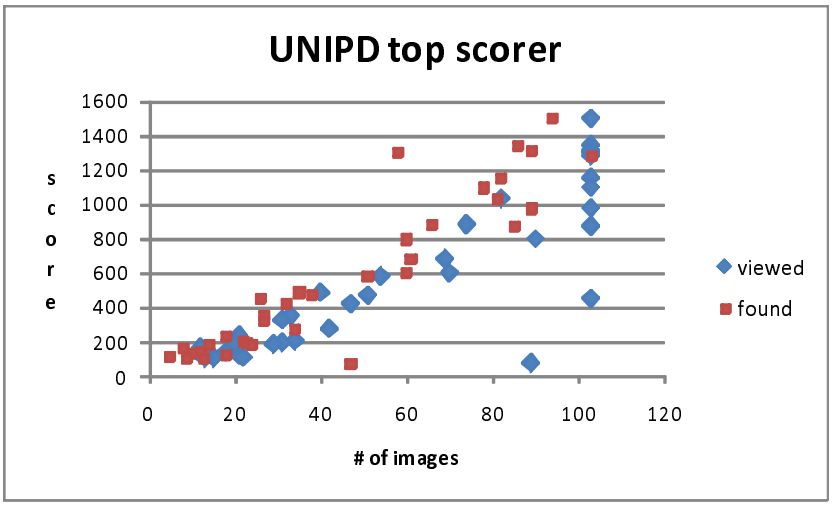

Fig. 4: Scores with respect to number of images viewed and found.

searches per image; however, "average" users are not common. In fact, it is more frequent the situation where a user prefers either to search in one language or to do a cross-lingual search. Performances, in terms of scores, seem not to be affected by the strategy chosen: top scorers can use a mixed strategy (user_50 and user_57), prefer a monolingual strategy (user_01 or user_07), or prefer a cross-lingual strategy (user_51 or user_52).

6 Comments and Final Remarks

In this section, we try to summarize the main points highlighted in this work and give a critical analysis.

From the point of view of the aim of the game, finding the image, the hardest obstacle was probably the size of the set of images retrieved. In both cases, image found or image skipped, a large number of users claimed that it was hard to find the image because there were too many images retrieved. However, from the direct interaction with the students and from some comments written in the questionnaires there were many cases in which the set of retrieved images contained the same "object" of the picture but not the exact picture. In real cases, you probably want to look for some image, not one in particular. The extra effort, which in our opinion is not realistic, that iCLEF participants has to do should be taken into consideration when doing the analysis of the data.

Another hard point was the difficulty in describing the image. Finding suitable keywords is indeed a hard task. If you want to see the problem from the other point of view, the tags which describe the image may be inappropriate for the same reason. As a result, it is difficult to get a good match between the

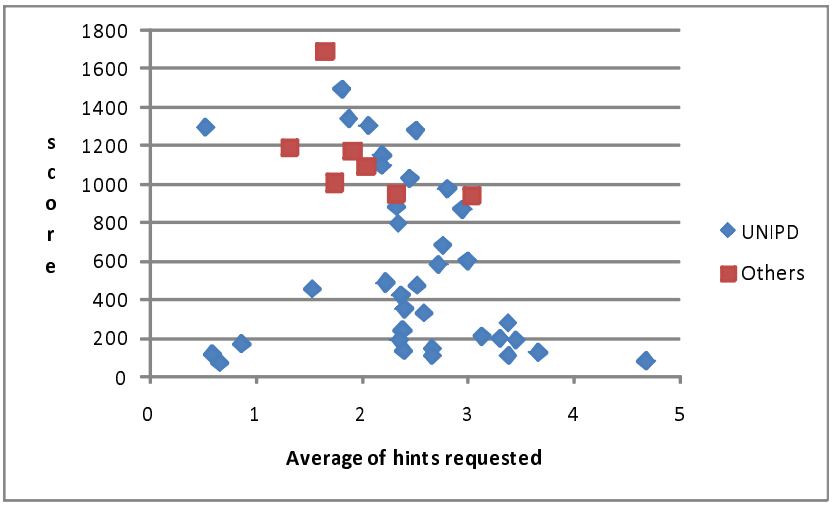

Fig. 5: Scores with respect of the number of hints requested on average.

words used to describe the image and the words used to query the database. A possible solution could be adding the possibility to search according to visual features of the images. However, the answers in the questionnaires were not so positive about this tool.

Users in general may find difficult to describe the image because the language in which is described is not known. As one could expect, this problem is less evident for the UNIPD students. This problem goes together with question 8 of the overall questionnaire (which interface facilities did you miss?). There is a strong positive response regarding the use of bilingual dictionaries with a better coverage, and a system able to give good suggestions for translating the keywords.

We also saw that there is not a strategy that outperforms the others. Using more monolingual searches than multilingual, a mix of the two, or prefer multilingual searches does not influence the final score. It would be interesting to study how user reformulate queries and if the reformulation changes from one strategy to another. This was not part of the analysis and is currently future work.

One final comment is about the time for each search. Unfortunately, the calculation of the time was not accurate enough to do this type of analysis, During the observation of the students of UNIPD, the feeling is that users spend much more time in the search compared to a similar realistic situation. We tried to simulate a "real user scenario" with this idea in mind: a user does not spend more than two or three minutes per image and can ask at most one hint, and user should not be influenced by the final score. This user, the author of the paper himself, is actually user_01 shown in all the previous tables and plots. This strategy easily brings to a low precision, many images are skipped, but in

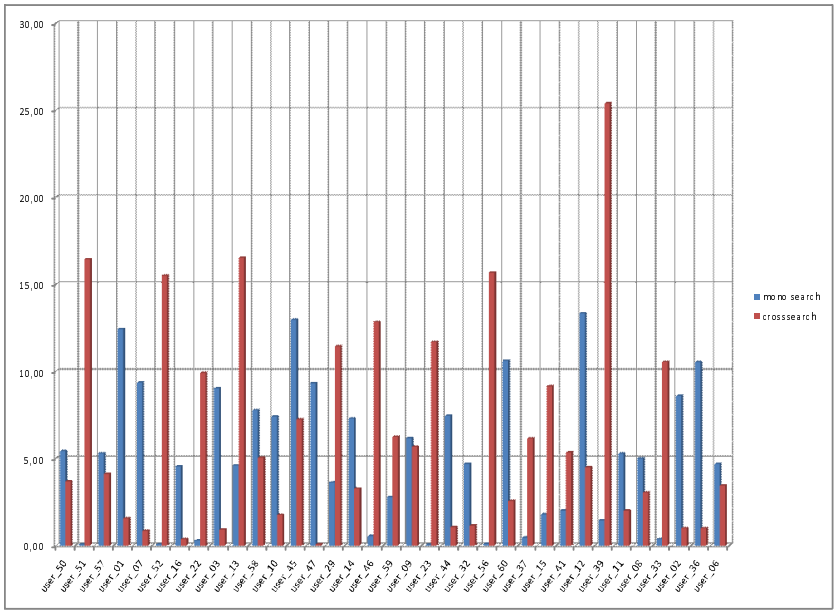

Fig. 6: Average numbers of monolingual and cross-lingual searches per image. Histogram based on the data of Table 14.

a real scenario the same user would have been satisfied by the search because usually a similar image (to the given image) is found. The time spent for each image is very low (probably the lowest compared to the other users), but in this case there is a bias to take into account when looking at the scores: the expertise in using search engines.

In conclusion, the experience of iCLEF and Flickling was exciting, especially for the students who really liked the game. It was also positive because it showed them that it is possible to have tools like multilingual search engines. Many students considered for the first time to investigate the possibility of doing cross-lingual searches for their studies.

Acknowledgements

The work reported has been partially supported by the TrebleCLEF Coordination Action, as part of the Seventh Framework Programme of the European Commission, Theme ICT-1-4-1 Digital libraries and technology-enhanced learning (Contract 215231).

References

1. Artiles, J., Barker, E., Clough, P., Gonzalo, J., Karlgren, J., Peinado, V.: Large-scale interactive evaluation of multilingual information access systems - the iCLEF

flickr challenge. In: Workshop on Novel Methodologies for Evaluation in Information Retrieval (30th European Conference for Information Retrieval (ECIR 2008)), Glasgow (March 2008)
2. Gonzalo, J., Clough, P., Karlgren, J.: Overview of iCLEF 2008: search log analysis for multilingual image retrieval. In Borri, F., Peters, C., eds.: Cross Language Evaluation Forum (CLEF 2008) Workshop Notes, Aarhus (September 2008) To appear
3. Peinado, V., Artiles, J., Gonzalo, J., Barker, E., Lpez-Ostenero, F.: Flickling: a multilingual search interface for flickr. In Borri, F., Peters, C., eds.: Cross Language Evaluation Forum (CLEF 2008) Workshop Notes, Aarhus (September 2008) To appear

Participant	Points	Average	Participants
Other Users	4910	51	95
University of Padua - Linguistics	20465	341	60
dZoom	7370	184	40
University of Padua 2 - Linguistics	5330	444	12
UNED LSI	2120	176	12
PLN	1835	183	10
UNED IA	2310	330	7
University of Sheffield	870	145	6
Manchester Metropolitan University	2250	450	5
CWI	335	67	5
Manchester Metropolitan U.	185	37	5
UNED DIA	470	117	4
UNED SCC	945	315	3
UNED ISSI	0	0	3
Yahoo	1060	530	2
Hildesheim University	215	107	2
Chemnitz	215	107	2
XEROX	100	50	2
U. Padua	0	0	2
UNED others	0	0	2
U. Twente	1095	1095	1
IST-KolKata	925	925	1
LabTL	630	630	1
APL	165	165	1
Hagen	85	85	1
UNED NLP	50	50	1
NII	25	25	1
SINTEF	25	25	1
MIRACLE	20	20	1
NTU	0	0	1
TKK	0	0	1
AIFB	0	0	1
DFKI	0	0	1
Hildesheim	0	0	1
KAIST	0	0	1
North Texas University	0	0	1
REINA	0	0	1
SIG	0	0	1
TextMESS	0	0	1
U. Tehran-1	0	0	1

Table 12: Total score for each participating group, the average score per user, the number of participants per group.

userid	viewed	found	skipped	pending	total score
user_50	103	94	9	0	1495
user_51	103	86	17	0	1340
user_57	103	89	14	0	1305
user_01	103	58	45	0	1295
user_07	103	103	0	0	1280
user_52	103	82	20	1	1150
user_16	103	78	25	0	1095
user_22	82	81	0	1	1030
user_03	103	89	14	0	975
user_13	74	66	7	1	880
user_58	103	85	18	0	870
user_10	90	60	29	1	795
user_45	69	61	7	1	680
user_47	70	60	9	1	600
user_29	54	51	2	1	580
user_14	40	35	4	1	485
user_46	51	38	12	1	470
user_59	103	26	77	0	450
user_09	47	32	14	1	420
user_23	33	27	5	1	350
user_44	31	27	3	1	325
user_32	42	34	7	1	275
user_56	21	18	2	1	235
user_60	34	22	12	0	205
user_37	31	23	7	1	195
user_15	29	24	4	1	185
user_41	20	14	5	1	185
user_12	12	8	3	1	165
user_39	18	12	5	1	140
user_11	14	10	3	1	130
user_08	21	18	2	1	120
user_33	22	5	16	1	110
user_02	15	13	1	1	105
user_36	13	9	3	1	105
user_06	89	47	41	1	75

Table 13: UNIPD scores per user, number of images viewed, number of images found and skipped. Image pending means that user has not finished the search.

userid	mono search	cross search
user_50	5.43	3.69
user_51	0.07	16.44
user_57	5.29	4.12
user_01	12.44	1.56
user_07	9.37	0.84
user_52	0.08	15.50
user_16	4.54	0.37
user_22	0.29	9.91
user_03	9.03	0.92
user_13	4.61	16.53
user_58	7.78	5.05
user_10	7.41	1.77
user_45	12.97	7.25
user_47	9.31	0.07
user_29	3.63	11.44
user_14	7.30	3.28
user_46	0.55	12.84
user_59	2.79	6.25
user_09	6.17	5.68
user_23	0.06	11.70
user_44	7.45	1.06
user_32	4.69	1.14
user_56	0.10	15.67
user_60	10.62	2.56
user_37	0.45	6.16
user_15	1.79	9.14
user_41	2.00	5.35
user_12	13.33	4.50
user_39	1.44	25.39
user_11	5.29	2.00
user_08	5.00	3.05
user_33	0.36	10.55
user_02	8.60	1.00
user_36	10.54	1.00
user_06	4.70	3.45

Table 14: UNIPD average monolingual and cross lingual searches.