# Providing multilingual access to FLICKR for Arabic users

**Paul Clough**
University of Sheffield
Sheffield, UK
p.d.clough@shef.ac.uk

**Azzah Al-Maskari**
University of Sheffield
Sheffield, UK
lip05aaa@shef.ac.uk

**Kareem Darwish**
IBM
Cairo, Egypt
darwishk@eg.ibm.com

**Abstract**

In this paper we describe our submission for iCLEF2006: an interface that allows users to search FLICKR in Arabic for images with captions in a range of languages. We discuss the results gained from a user experiment following directives given by iCLEF, including an analysis of how successful search tasks were found to be. To enable the searching of multilingual image annotations we use English as an interlingua. An Arabic-English dictionary is used for initial query translation, and then Babelfish is used to translate between English and French, German, Italian, Dutch and Spanish. Users are able to modify the English version of the query if they have the necessary language skills to do so. We have chosen to experiment with Arabic retrieval from FLICKR due to the growing numbers of online Middle Eastern users, the limited numbers of interactive Arabic user studies for cross-language IR to date, and the availability of resources to undertake a user study.

## 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

CLEF, iCLEF, Flickr, multilingual image search, user studies, Arabic retrieval

# 1 Introduction

FLICKR[1] is a large-scale, web-based image database based on a large social network of online users. The application is used to manage and share personal (and increasingly more commercial) photographs and currently contains over five million freely accessible images. These are available via the web, updated daily by a large number of users and accessible to all web users for free. The photos have multilingual annotations generated by authors using freely-chosen keywords (known as a folksonomy). Similar systems are also emerging for collections of personal videos[2] (e.g. youtube.com and CastPost).

[1] http://flickr.com
[2] http://www.techcrunch.com/2005/11/06/the-flickrs-of-video/

Table 1: World Internet usage and population statistic (2006).

| World Regions | Population % of World | Internet Usage % of World | Usage Growth 2000-2005 |
|---|---|---|---|
| Asia | 56.4% | 35.6% | 218.7% |
| Africa | 14.1% | 2.3% | 423.9% |
| Europe | 12.4% | 28.5% | 177.5% |
| Latin America / Caribbean | 8.5% | 7.8% | 342.5% |
| North America | 5.1% | 22.2% | 110.3% |
| Middle East | 2.9% | 1.8% | 454.2% |
| Oceania / Australia | 0.5% | 1.7% | 134.6% |
| Total | 100.0% | | |

Table 2: Proportion of Internet users per language (2004, 2006).

| Language | 2004 | 2006 |
|---|---|---|
| English | 35.80% | 30.6% |
| Chinese | 14.10% | 13.0% |
| Spanish | 9.00% | 7.9% |
| Japanese | 9.00% | 8.5% |
| German | 7.30% | 5.6% |
| Others | 6.50% | – |
| Korean | 4.10% | 3.3% |
| French | 3.80% | 4% |
| Portuguese | 3.50% | 3.2% |
| Italian | 3.30% | 2.8% |
| Dutch | 1.90% | – |
| Arabic | 1.70% | 2.9% |

Despite the popularity of FLICKR on a global basis, to-date there has been little empirical investigation regarding multilingual access to FLICKR. A common remark of CLIR is why users would want to retrieve documents that they (presumably) cannot read. Of course, in the case of image retrieval the motivation for CLIR is much stronger: for many search tasks users are able to judge the relevance of images without the need of additional text, thereby eliminating the need for translation of search results. This linguistic-neutrality of images makes text-based image retrieval an ideal application for CLIR.

For our submission to iCLEF 2006, we wanted to experiment with providing an interface which would enable users to query FLICKR in Arabic. We selected Arabic because of resources available to us locally, the limited number of interactive Arabic user studies so far in cross-language retrieval, the growing number of online Middle Eastern users and the limited availability of online material in Arabic. Tables 1 and 2 provide some figures of online users population per language. According to these statistics it could be estimated that an English speaker has, in general, access to more than 50% more digital material than any other speaker (in particular than the Arabic speaker). According to Reuters and ABC Science Online[3] there are currently only 100 million web pages in Arabic, 0.2% of the total worldwide.

Table 3 indicates the distribution of users by the major script languages where Arabic is 8.93% of the total. Research commissioned from a Dubai-based Internet researcher [4] shows the number of Arabic speaking Internet users could jump to 43 million in 2008 from 16 million in 2004. It also shows that 65% of Arabic Internet users in 2005 could not read English, which accounts for 70% of the material on the Internet. Hence number of Arabic online population is increasing faster than

[3] http://www.abc.net.au/news/newsitems/200604/s1624108.htm (April 26, 2006)
[4] http://www.abc.net.au/science/news/stories/s1623945.htm (April 26, 2006)

Table 3: Distribution of User Population by Major Script Categories (2005).

| Script | Latin | Cyrillic | Arabic | Hanzi | Indic | Other* |
|---|---|---|---|---|---|---|
| Num. users (mil.) | 2,238 | 451 | 462 | 1,085 | 807 | 129 |
| Total % | 43.3% | 8.7% | 8.9% | 21.0% | 15.6% | 2.5% |

*Others include Greek, Georgian, Armenian, Amharic, Dhivehi, Hebrew, etc.

Table 4: Estimation of language distribution of FLICKR.

| | |
|---|---|
| English | 16,593,773 |
| Dutch | 1,367,326 |
| German | 1,396,963 |
| Spanish | 1,002,055 |
| Italian | 665,025 |
| French | 160,104 |
| Arabic | 37 |

online Arabic documents. There is no doubt that a system designed for Arabic speakers (but able to search English documents besides other languages) would open new possibilities both in terms of the quantity of topics covered and the quality of the items retrieved.

FLICKR is one such online resource which, in addition, allows users to add annotations in any language they want. To estimate the degree of multilingual content in FLICKR, we translated 15 of the all-time most popular tags with the highest interestingness rating[5] - wedding, birthday, family, friends, party vacation/holiday, flowers, travel, trip, food, cat, dog, baby, music, and summer - into French, Spanish, German, Italian, Dutch and Arabic[6]. We then ran these queries on FLICKR to find the number of images per language (results for these queries are shown in Table 4). It is clear that most images are tagged in English and almost none in Arabic. For Arabic users wanting browse or find images with FLICKR then unless English or another language is used, results will be limited.

The remainder of the paper describes the system developed for Arabic users, the experiments, results and conclusions. Three main aims of our work were: (1) to analyse the tasks offered by iCLEF, (2) to analyse our initial interface: query translation and use of English as an interlingua, and (3) to observe the searching behaviour of users.

## 2 The system: FLICKRArabic

We developed an Ajax-based online application to provide query translation to the FLICKR API[7] called FLICKRArabic (shown in Fig. 1). The system focuses on translating user's queries from Arabic into English (interlingua) and then translating from here into French, Spanish, German, Italian or Dutch. This is necessary because many translation resources only offer Arabic to English. The English translation is shown to users who can modify the query if they have sufficient language skills. In the case of polysemous Arabic queries, translations for all senses are displayed to the user and they are able to ignore or select the correct translations. Translation between Arabic-English is performed using a bilingual dictionary (described in Section 2.1). Translation between the English and other languages was performed by creating a wrapper for the online MT system Babelfish. This meant that users had little control over translation thereafter. Design and implementation of the system is described further in Section 2.2.

[5]http://flickr.com/photos/tags/

[6]We selected these languages as all (excluding Arabic) were suggested by the example iCLEF system.

[7]http://www.flickr.com/services/api/


Figure 1: FLICKRArabic interface.

## 2.1 Language resources

Two translation resources have been used to create the application. The first is an Arabic-English bilingual dictionary, the second is the online MT tool Babelfish[8] (this is accessed through a custom-built wrapper). To create the bilingual dictionary, two bilingual term lists were constructed using two Web-based machine translation systems: Tarjim[9] and Al-Misbar[10]. In each case, we submitted sets of isolated English words found in a 200 MB collection of Los Angeles Times news stories for translation from English into Arabic [3]. Each system returned at most one translation for each submitted word. All in all, the combined bilingual term lists contained 225,057 unique entries. In preprocessing the Arabic text, all diacritics and kashidas, which are character elongations, were removed, letter normalization was employed to normalize the letters *ya* and *alef maqsoura* to *ya* and all the variants of *alef* and *hamza*, namely *alef*, *alef hamza*, *alef maad*, *hamza*, *waw hamza*, and *ya hamza*, to *alef*, and lastly all words were stemmed using Al-stem[11].

## 2.2 Interface design and functionality

The interaction design for this system has emerged from an interactive evaluation design process. A user-centered approach was implemented: five Arabic potential users were involved during the stages of designing and building of this system following the advice of (Hackos and Redish, 1998). Pre-evaluation allowed us to explore different aspects of interactive cross-language searching. During the pilot session, users were observed and questioned about their cross-language actions (e.g. editing the translation of the Arabic query and flipping through the results of other languages). Previous research which has considered user interaction in the formulation of multilingual queries includes the Keizai system [7], ARCTOS [8], MULINEX [2], WTB (Web site Term Browser) [6],

[8] http://babelfish.altavista.com
[9] http://tarjim.ajeeb.com, Sakhr Technologies, Cairo, Egypt www.sakhr.com
[10] http://www.almisbar.com, ATA Software Technology Limited, North Brentford Middlesex, UK
[11] http://www.glue.umd.edu/ kareem/research/

MIRACLE [4], EUROVISION ?? and CLARITY [9].

The basic functionality of the system is as follows:

- Users can search FLICKR using initial English or Arabic queries
- If searching in English, the system calls the FLICKR API and displays results
- If writing in Arabic:
  - The query is first converted from UTF8 to CP1256 (using *iconv -f utf8 -t cp1256*) and stemmed (using the *stem_cp1256.pl* Perl program)
  - The query is then translated into English using the Arabic-English dictionary
  - The system returns all dictionary matches (senses and synonyms) for each translated query
  - Users can modify the English translation by deleting unwanted terms or adding their own
  - The English query is then used for searching FLICKR using the API[12]
  - The user can view photos annotated in Arabic only (for comparison with other languages)
- Results for each language are displayed in separate tabs with the total number of images found displayed (when each language selected)
- The user can view photos with annotations in any one of the five languages: French, Spanish, German, Italian and Dutch[13]
- Users can select the following search options:
  - Display 10, 20, 50, 100 or 200 images per page
  - Sort images by a relevance score or interestingness
  - Search all annotation text (titles, tags and descriptions) or tags only (all query terms or any)

We displayed the English translation of the query to users and allowed them to edit this during the search to determine whether this particular group of users would find this helpful (given that previous research had shown this to be the case for most bilingual users). We also showed results for English to the users first (left-hand tab) because during initial testing we found that most users could make use of images with English annotations. The ordering of languages thereafter was arbitrary. Results for Arabic were also provided to enable users to compare results from searching FLICKR using purely Arabic.

# 3 The experiment

To obtain feedback on the implemented system, we recruited 11 native Arabic users to carry out the tasks specified by iCLEF [5]. The subjects represent a certain group of Arabic users: undergraduate and postgraduate students with good English language skills[14]. The mean age of the 11 users was 28 years old, and 85% stated they typically searched the Web using English. They also had the following characteristics: 82% used the Internet several times a week, all had a great deal of experience with point-and-click interfaces, 46% searched for images very often and 46% of those people often found what they were searching for.

[12] Only photos posted before 1/6/2006 were returned

[13] If the query is not translated, it remains in English and results for this are returned to the user

[14] We are currently planning experiments with monoglot users: those who can only make use of Arabic

Table 5: Search results for task 1 (parliament).

| User | total found | correct | unique | recall | precision | inside | outside |
|---|---|---|---|---|---|---|---|
| 1* | – | – | – | – | – | – | – |
| 2 | 20 | 11 | 3 | 0.27 | 0.55 | 9 | 11 |
| 3 | 11 | 8 | 7 | 0.88 | 0.73 | 4 | 7 |
| 4 | 20 | 17 | 11 | 0.65 | 0.73 | 4 | 13 |
| 5 | 13 | 12 | 8 | 0.67 | 0.92 | 4 | 9 |
| 6 | 12 | 10 | 9 | 0.90 | 0.83 | 3 | 9 |
| 7 | 12 | 12 | 8 | 0.67 | 1.00 | 5 | 9 |
| 8 | 8 | 7 | 6 | 0.86 | 0.88 | 5 | 3 |
| 9 | 12 | 12 | 9 | 0.75 | 1.00 | 0 | 12 |
| 10 | 13 | 10 | 7 | 0.70 | 0.77 | 4 | 9 |
| 11 | 10 | 9 | 7 | 0.78 | 0.90 | 3 | 7 |
| Total | 131 | 108 | 75 | 0.69 | 0.84 | 41 | 80 |

*The system did not function correctly for this user during this task.

Subjects were asked to perform 3 tasks: (1) a classical ad-hoc task: "find as many European parliament buildings as possible, pictures from the assembly hall as well as from the ouside" (parliament); (2) a creative instance-finding task: "find five illustrations to the text - the story of saffron - the goal being to find five distinct instances of information described in the narrative: saffron, flower, saffron thread, picking the thread/flower, powder, dishes with saffron (saffron); and (3) a visually orientated or known-item task: given a picture, find the name of the beach on which the crab is resting (crab). More details of the tasks can be found in [5].

In these experiments, users first completed a preliminary questionnaire, then spent 20 minutes on each task. Tasks were assigned randomly to users to reduce the effects of task bias on the results (e.g. user 1 performed tasks 2-3-1, user 2 performed 3-1-2 etc). We also asked subjects to perform a final search where they were able to search for images on any topic. Finally, we asked users to complete a questionnaire to establish their overall satisfaction with and impressions of the system. During the experiment, we recorded some attributes of the task such as time taken and queries input, as well as taking notes of the user's searching behaviour during each task.

# 4 Results and observations

## 4.1 User effectiveness

We first discuss how well users were able to perform the tasks[15]. Almost all users (10/11) were able to perform task 3 (crab) successfully[16]. Table 5 shows the results for task 1 (parliament). For this task, from the total number of images found, we judged those which we deemed to be *correct* and of these we counted the number of pictures of different European parliament buildings (*unique*). To compute recall we divided unique by correct; for precision we computed correct divided by total. We also divided the total number of pictures found between those of the inside of the building versus the outside. Across all users we obtained a recall of 0.69 and precision of 0.84. This varied between users with some scoring higher recall (e.g. user 3) and others higher precision (e.g. users 7 and 9). In general, almost double the number of total images found were of the outside of buildings.

Table 6 shows the results for task 2 (saffron). In this task users were asked to find 5 images to illustrate each part (or instance) to the story of saffron. Users were given one point for retrieving each instance (*counted*) and no credit for repeated instances (i.e. no additional credit for selecting two images of the same aspect of the story). Results show the number of images found for each aspect/instance and the precision (counted divided by total). Overall precision was 0.70 for this

[15]The system effectiveness and the correlation between user and system effectiveness is explored in [1]

[16]Despite users not being familiar with German, they were able to recognise the name of the beach.

Table 6: Search results for task 2 (saffron).

| User | flower | thread | food | powder | picking | total | counted | precision |
|---|---|---|---|---|---|---|---|---|
| 1 | 3 | 2 | – | – | – | 5 | 2 | 0.40 |
| 2 | 1 | 1 | 2 | – | 1 | 5 | 4 | 0.80 |
| 3 | 1 | 1 | 1 | 1 | – | 4 | 4 | 1.00 |
| 4 | 1 | 2 | 2 | | – | 5 | 3 | 0.60 |
| 5 | 3 | – | – | – | – | 3 | 1 | 0.33 |
| 6 | 2 | 1 | 1 | – | 1 | 5 | 4 | 0.80 |
| 7 | 1 | 1 | 1 | 1 | 1 | 5 | 5 | 1.00 |
| 8 | 2 | 1 | 1 | – | 1 | 4 | 5 | 0.80 |
| 9 | 1 | 1 | 1 | 1 | 1 | 5 | 5 | 1.00 |
| 10 | – | 4 | 1 | – | – | 5 | 2 | 0.40 |
| 11 | 2 | 1 | 2 | – | – | 5 | 3 | 0.60 |
| Total | 18 | 17 | 11 | 2 | 4 | | | 0.70 |

Table 7: User relevance per language.

| | Task 1 (parliament) | | | Task 2 (saffron) | | | Task 3 (crab) | | |
|---|---|---|---|---|---|---|---|---|---|
| Language | Highly | Partially | Not | Highly | Partially | Not | Highly | Partially | Not |
| English | 9 | 2 | 0 | 7 | 4 | 0 | 3 | 5 | 3 |
| French | 4 | 5 | 2 | 7 | 4 | 0 | 2 | 7 | 2 |
| Spanish | 5 | 4 | 2 | 7 | 4 | 0 | 4 | 6 | 1 |
| German | 4 | 5 | 2 | 8 | 3 | 0 | 2 | 6 | 3 |
| Italian | 5 | 4 | 2 | 6 | 5 | 0 | 3 | 3 | 5 |
| Dutch | 2 | 5 | 4 | 3 | 5 | 3 | 3 | 3 | 5 |
| Arabic | 0 | 0 | 11 | 0 | 0 | 11 | 5 | 3 | 3 |

task and again, precision varied between users as some were good at instance-finding (e.g. users 3 and 9); others were less successful (e.g. users 5 and 11).

Table 7 shows the number of users who judged results for each language and task as highly relevant, partially relevant or not relevant. It would appear that users found relevant images with annotations in most of the languages, except Arabic. Users commented that they were dissapointed with the Arabic results. This would suggest that multilingual access to FLICKR could improve retrieval. For task 2 (saffron), most users found relevant images in Italian which is likely due to the narrative mentioning Italy.

Table 8 shows the user's satisfaction with the accuracy and coverage of the search results. Overall it would appear that users were very satisfied with the accuracy of search results from FLICKRArabic for tasks 1 and 3, but less satisfied with the accuracy of results for task 2. Similarly, users appear in general satisfied with the coverage of results (again less so for task 2). We asked users whether they felt accuracy or coverage was more important for the tasks 1 and 2. In task 1 (parliament), 7 users favoured accuracy and 4 coverage; in task 2 (saffron) 8 users favoured accuracy and 3 coverage. Overall, it would appear that users would prefer more accurate results than more results and this is likely to reflect the precision-orientated nature of the tasks.

Table 8: User's satisfaction with accuracy and coverage.

| | Accuracy | | | Coverage | | |
|---|---|---|---|---|---|---|
| Task | Highly | Partially | Not | Highly | Partially | Not |
| 1 (parliament) | 10 | 1 | 0 | 8 | 2 | 1 |
| 2 (saffron) | 5 | 6 | 0 | 5 | 6 | 0 |
| 3 (crab) | 10 | 1 | 0 | 9 | 2 | 0 |

Table 9: User-rated usefulness of image attributes.

| | Task 1 (parliament) | | | Task 2 (saffron) | | | Task 3 (crab) | | |
|---|---|---|---|---|---|---|---|---|---|
| | V. imp. | Imp. | Unimp. | V. imp. | Imp. | Unimp. | V. imp. | Imp. | Unimp. |
| Image only | 8 | 3 | 0 | 8 | 2 | 1 | 11 | 0 | 0 |
| Image and caption | 10 | 0 | 1 | 8 | 3 | 0 | 5 | 6 | 0 |
| Comments | 7 | 3 | 1 | 5 | 5 | 1 | 1 | 6 | 4 |
| Foreground details | 7 | 1 | 3 | 4 | 5 | 2 | 6 | 3 | 2 |
| Background details | 1 | 3 | 7 | 1 | 3 | 7 | 6 | 4 | 1 |
| Previous knowledge | 2 | 2 | 7 | 5 | 4 | 2 | 2 | 1 | 8 |

We asked users about the usefulness of the results for each task and overall 70% of users were very satisfied with the results (30% partially satisfied). Table 9 indicates how user rate the importance of factors in helping to determine the usefulness of the images for each search task. Users rated these as very important (v. imp.), important (imp.) and not important (unimp).For task 1, users found textual information very important in addition to the image itself. We expected this as users need to check the annotations to determine whether a parliament building is European or not. In task 2, users found the image and caption to be the most important for determining relevance. Users were able to identify possible pictures of saffron, but needed the captions to confirm their decision. In task 3, as expected, users found the visual content of the photos most useful. This reflects the fact that this task is more visual in nature. Users also found foreground and background text useful to determine the beach where the crab was placed.

## 4.2 User's searching behaviour

Tables 10 to 12 provide details of a number of observed characteristics recorded during user's searching. This includes the number of images the user selected to view, the queries submitted (the initial query was in Arabic) and the languages viewed for the search results. The following obeservations regarding searching behaviour were observed: users typically viewed initial results in English before trying other languages. This is because they were able to read annotations in English. Two main strategies for searching prevailed: some users input fewer queries and looked through many pages of results; others input many queries and if no relevant found in the first page of results reformulated the query. For the search results, some users would systematically look through results for each language from left to right; others would start with the languages which returned the least number of results (testing each language first). Most users selected 100 images at a time to view in the search results suggesting they are able (and willing) to view a large number of thumbnails.

## 4.3 User's comments on the tasks

To determine the sucess of each task, we gathered user's comments on different aspects of the tasks as shown in Tables 13 and 14. Overall (from Table 13) it would appear that tasks 1 and 3 were the clearest, task 1 the easiest, task 3 the most familiar, and tasks 2 and 3 the most interesting for users. Interestingly, the majority of users did not find any of the tasks relevant to them. This is primarily because these topics were not designed specifically for Arabic users who are likely to search for different topics than Europeans and the tasks themselves not being entirely realistic (e.g. searching for a crab on a beach to find a specific location). It was also interesting to find that users were reluctant to search for crab because in Arabic this has another sense: cancer. Also, the query for saffron was alien to most of the male searchers who did not cook and therefore were unsure what saffron was or looked like.

In Table 14 comparing against tasks, users found task 1 the most interesting and easiest, tasks 1 and 3 the most enjoyable and task 1 the most realistic out of the three tasks. Users commented

Table 10: User's searching behaviour for task 1.

| User | # images to view | Queries submitted | Languages viewed |
|---|---|---|---|
| 1 | 100 | european parliament | English |
| 2 | 100 | european parliament building, european parliament, inside parliament | English |
| 3 | 50,100 | european parliament, parliament building, parliament, inside parliament | English |
| 4 | 100 | parliament, european parliament, parliament assembly hall | English, all left-right |
| 5 | 100 | european parliament, parliament France, parliament Berline | English |
| 6 | 100 | european parliament building, european parliament design, architecture european parliament, Scotish parliament, russian parliament | English |
| 7 | 100 | european parliament building, european parliament, assembly happ european parliament | English |
| 8 | 100 | parliament, parliament Madrid, parliament Frace, parliament london | English |
| 9 | 100 | parliament, parliament building, parliament hall | Arabic, Dutch, English |
| 10 | 20, 200 | european parliament | English, French |
| 11 | 100 | parliament, inside parliament | English |

Table 11: User's searching behaviour for task 2.

| User | # images to view | Queries submitted | Languages viewed |
|---|---|---|---|
| 1 | 100 | saffron | English |
| 2 | 100 | saffron, saffron plant, saffron and corcus, saffron | English |
| 3 | 100 | saffron flower, saffron plant, saffron | Italian and Arabic |
| 4 | 100 | saffron, crocus sativious, saffron powder, saffron thread, saffron stem | Arabic, English, all left-right |
| 5 | 100 | saffron | English and Dutch |
| 6 | 100 | saffron, saffron flower, saffron kitchen | English |
| 7 | 100 | the saffron | English, then all left-right |
| 8 | 100 | saffron | English (didn't view Arabic) |
| 9 | 100 | saffron, saffron picking | Italian, English, Arabic |
| 10 | 20, 200 | saffron | French, English, all left-right |
| 11 | 100 | saffron | Arabic, English, all left-right |

Table 12: User's searching behaviour for task 3.

| User | # images to view | Queries submitted | Languages viewed |
|---|---|---|---|
| 1 | 100 | crab | English |
| 2 | 100 | crab beach | English |
| 3 | 200 | crab animal, crab sea, crab | Dutch, English, all left-right |
| 4 | 10,50,100 | crab, crab beach | French, Arabic, English, all left-right |
| 5 | 100 | crab | English, all left-right |
| 6 | 100, 200 | sea crab, sea crab beach, crab | English, all left-right |
| 7 | 100 | crab | German |
| 8 | 100 | crab | French, English, all left-right |
| 9 | 50, 200 | crab sea | English, Dutch, French |
| 10 | 100 | crab | English, French, Spanish, German |
| 11 | 10, 20 | crab | English, French, Arabic, German, Italian |

Table 13: User's assessment of the search tasks (1).

| | Task 1 (parliament) | | | Task 2 (saffron) | | | Task 3 (crab) | | |
|---|---|---|---|---|---|---|---|---|---|
| | Highly | Partially | Not | Highly | Partially | Not | Highly | Partially | Not |
| Clear | 9 | 1 | 1 | 4 | 4 | 3 | 11 | 0 | 0 |
| Easy | 8 | 3 | 0 | 7 | 4 | 0 | 6 | 4 | 1 |
| Familiar | 7 | 3 | 1 | 3 | 8 | 0 | 9 | 2 | 2 |
| Interesting | 5 | 3 | 2 | 6 | 5 | 0 | 7 | 3 | 1 |
| Relevant | 2 | 4 | 5 | 1 | 4 | 6 | 1 | 4 | 6 |

Table 14: User's assessment of the search tasks (2).

| | Interesting | | | Easiest | | | Enjoyable | | | Realistic | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Task | Most | Somewhat | Least | Most | Somewhat | Least | Most | Somewhat | Least | Yes | No |
| Task 1 | 6 | 1 | 4 | 7 | 1 | 3 | 4 | 5 | 2 | 6 | 5 |
| Task 2 | 3 | 6 | 2 | 1 | 7 | 3 | 2 | 6 | 3 | 3 | 8 |
| Task 3 | 2 | 4 | 5 | 3 | 4 | 4 | 4 | 3 | 4 | 1 | 10 |

Table 15: User's queries for free-search task.

| User | Query | Language of query | Languages of viewed results |
|---|---|---|---|
| 1 | Damascus | Arabic | Arabic, German, then English |
| 2 | Leptis (a city in Libya) | Arabic (not in dict), then English | Arabic, then English |
| 3 | Shef Uni, Mecca, Jeddah | Arabic | English, Arabic, then right-left |
| 4 | Tower, Skyscraper, Suspension bridge | Arabic | Arabic, English, then left-right |
| 5 | Damascus, Mecca | Arabic | Arabic |
| 6 | Orientalism | Arabic (not in dict), then English | English |
| 7 | Welding, Pyramid | English | English |
| 8 | Mosque | Arabic | English, German, Italian, Dutch, Arabic |
| 9 | Castles in Oman, andalus,Hamraa palace | Arabic | Arabic, then English (few results) |
| 10 | Cats, Muscat | Arabic and English | English, Arabic, French; English, French, left-right |
| 11 | Manchester, England, Libya | English | English |

that task 3 was very unrealistic and not representative of the type of search task they would perform. For iCLEF organisers, this might indicate that concentrating on adhoc search is more likely to represent user's tasks and be less artificial.

## 4.4 Free-search task

If users could search for their own topics, what would they search for? We asked users to submit their own queries to determine the types of topics representative or interesting to this user group. Table 15 shows the queries submitted for each user, the query language used, and the language of the annotations viewed in the results. As expected, most users typed queries more related to their culture (e.g. names of places such as Oman and Leptis and objects such as a mosque) and interests (e.g. welding). Most users searched using Arabic but found results in other languages helpful or useful. For queries containing out-of-vocabulary terms (e.g. Leptis), users had sufficient language skills to search in English.

## 4.5 Overall user comments

Table 16 indicates overall user's comments about the system we implemented. This represents user's satisfaction as recorded by the categories: very satisfied, partially satisfied and not satisfied. User's were very positive about the system and definitely found the provision of multilingual access to be useful to them (the ability to view pictures with annotations in various languages). Most users (9/11) were very satisified with the multilingual search results (compared against using FLICKR as is). However, the majority of these users were not happy with the query translation (7 partially satisfied and 1 not satisified). Indeed we found that because this user group had good English language skills, the translation from Arabic to English was actually an unneccessary step and most preferred to formulate and modify queries in English (10/11 users were willing to modify the English version of the query, and all users would enter the English version of a query term if not in the dictionary and add synonyms).

Most users (8/11) would use the system again and all users would recommend the system to

Table 16: User's overall satisfaction rating of the system.

| | Very | Partially | Not |
|---|---|---|---|
| Overall success | 7 | 4 | 0 |
| Multilingual usefulness | 11 | 0 | 0 |
| Multilingual satisfaction | 9 | 2 | 0 |
| Use system again | 8 | 0 | 3 |
| Recommend system to friend | 11 | 0 | 0 |
| Easiness of use | 11 | 0 | 0 |
| Quality of translation | 3 | 7 | 1 |
| Willingness to modify query | 10 | 0 | 1 |
| Entry of synonyms | 11 | 0 | 0 |

a friend or colleague. All users were very satisfied with the ease with which the system could be used. Many users said they would have viewed most non-English annotations if translations had been provided into English (or Arabic). This suggests some form of document translation could improve the user's search experience.

# 5 Discussion and Conclusions

The goal of our work was to build and test a simple Arabic query interface to FLICKR enabling users to view images with annotations in a range of languages. To enable Arabic translation into multiple languages, we first translated into English (interlingua) using a bilingual Arabic-English dictionary. From initial user testing, we decided to show users the English translations and allow them to edit as desired. The English version of the query was then translated into other languages as users requested to view results in those languages using the Babelfish MT system.

Overall with the group of users recruited for this experiment, we found that providing Arabic as an initial query language was unneccessary and caused more frustration due to poor translation or query terms not found in the dictionary than it was useful. Users were much happier submitting and reformulating queries in English (particularly for the tasks set which were orientated to Europeans and unlikely to perform well with Arabic searching). Some users did express the need for the ability to search in Arabic in some cases (e.g. when they are unable to formulate a query in English), but this was not the case for most of these tasks. However, users did comment that being able to start the search in Arabic to obtain some terms in English was a useful way to begin their search. Some users also suggested that being able to combine Arabic and English queries would be useful.

Compared to the current FLICKR system, it would seem that being able to submit an English query and translate this into multiple languages is considered as very beneficial to end users. It was particularly apparant with some queries whereby the search results are very much language-dependent and different (e.g. searching for *car* typically produces British-built cars for English and *voiture* produces French-built cars). It would, therefore, seem more important to focus on this part of the system than initial query translation. Users were generally complementary of our system and they were able to carry out the search tasks set with reasonable success: overall precision of 0.84 for task 1, precision of 0.70 for task 2 and 10/11 users completed task 3 successfully.

Further work is planned in the following areas:

- We plan to run the experiment again with users who are less proficient in English and would be less likely to reformulate the English version of the query.
- We aim to improve the query translation by increasing the size of the dictionary, handling the translation of phrases and enabling the user to correct erroneous dictionary entries.
- On presenting results to the user and providing different ranking and clustering strategies to reduce the number of images users much search to find relevant images.

- We also plan to perform simultaneous searches for each language to provide users with a summary of the number of results for each language. This will help those users who typically view languages which exhibit the fewest number of results first.
- Translate the annotations of images returned from FLICKR.

## Acknowledgments

This work has been partially supported by the European Commision, project MultiMatch (FP6-2005-IST-5-033104).

## References

[1] A. Al-Maskari, P. Clough, and M. Sanderson. User's effectiveness and satisfaction for image retrieval. In *Workshop on Information Retrieval, Hildesheim, Germany*, page In Submission, 2006.

[2] J. Capstick, A. Diagne, G. Erbach, H. Uszkoreit, F. Cagno, G. Gadaleta, J. Hernandez, R. Korte, A. Leisenberg, M. Leisenberg, and O. Christ. Mulinex: Multilingual web search and navigation, 1998.

[3] Kareem Darwish and Douglas W. Oard. Clir experiments at maryland for trec 2002: Evidence combination for arabic-english retrieval, 2002.

[4] Bonnie J. Dorr, Daqing He, Jun Luo, Douglas W. Oard, Richard Schwartz, Jianqiang Wang, and David Zajic. iclef 2003 at maryland: Translation selection and document selection. In *Proceedings of Cross-Language Evaluation Forum (CLEF2003) Workshop*, pages 435–449, 2003.

[5] J. Gonzalo, J. Karlgren, and P. Clough. iclef 2006 overview: Searching the flickr www photo-sharing repository. In *Proceedings of Cross-Language Evaluation Forum (CLEF2006) Workshop, Alicante, Spain*, 2006.

[6] A.Pe nas, F. Verdejo, and J. Gonzalo. Website term browser: Overcoming language barriers in text retrieval. *Journal of Intelligent and Fuzzy Systems*, 12(3–4):221–233, 2002.

[7] W. Ogden and J. Cowie. Keizai: An interactive cross-language text retrieval system, 1999.

[8] W. Ogden and M. W. Davis. Improving cross-language text retrieval with human interactions. In *Proceedings of the Hawaii International Conference on System Science (HICSS-33), Vol. 3*, 2000.

[9] Daniela Petrelli, Stephen Levin, Micheline Beaulieu, and Mark Sanderson. Which user interaction for cross-language information retrieval? design issues and reflections. *Journal of the American Society for Information Science and Technology (JASIST)*, 57(5):709–722, 2006.