

promoting access to White Rose research papers



Universities of Leeds, Sheffield and York
<http://eprints.whiterose.ac.uk/>

This is an author produced version of a paper published in **Evaluation of Multilingual and Multi-modal Information Retrieval**.

White Rose Research Online URL for this paper:
<http://eprints.whiterose.ac.uk/78505>

Published paper

Karlgren, J., Gonzalo, J. and Clough, P. (2006) *iCLEF 2006 Overview: Searching the Flickr WWW Photo-Sharing Repository*. In: Peters, C., Clough, P., Gey, F.C., Karlgren, J., Magnini, B., Oard, D.W., Rijke, M.D. and Stempfhuber, M., (eds.) *Evaluation of Multilingual and Multi-modal Information Retrieval*. 7th Workshop of the Cross-Language Evaluation Forum, 20th - 22nd September 2006, Alicante, Spain. Springer Berlin Heidelberg, 186 - 194.
http://dx.doi.org/10.1007/978-3-540-74999-8_27

iCLEF 2006 Overview: Searching the Flickr WWW photo-sharing repository

Julio Gonzalo	Jussi Karlgren	Paul Clough
UNED	SICS	University of Sheffield
Madrid, Spain	Kista, Sweden	Sheffield, UK
julio@lsi.uned.es	jussi@sics.se	p.d.clough@sheffield.ac.uk

Abstract

This paper summarizes the task design for iCLEF 2006 (the CLEF interactive track). Compared to previous years, we have proposed a radically new task: searching images in a naturally multilingual database, Flickr, which has millions of photographs shared by people all over the planet, tagged and described in a wide variety of languages. Participants are expected to build a multilingual search front-end to Flickr (using Flickr's search API) and study the behaviour of the users for a given set of searching tasks. The emphasis is put on studying the process, rather than evaluating its outcome.

Despite the emphasis placed on offering a much simpler experimental setup than in previous years, the number of participating groups to iCLEF was an all-time low. Three groups submitted papers: UNED, a joint submission from the University of Sheffield and IBM, and a submission from SICS. However, despite this limited participation, we are able to offer some insights into cross-language searching in Flickr and these will be discussed at the CLEF workshop, together with a discussion on how participation better could be encouraged for future evaluation cycles.

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, online photo sharing, multilingual image search, user studies, evaluation, evaluation campaigns

1 Introduction

iCLEF (the CLEF interactive track) has been devoted, since 2001, to study Cross-Language Retrieval from a user-centered perspective. The aim has always been to investigate real-life cross-language searching problems in a realistic scenario, and to obtain indications on how best to aid users in solving them. iCLEF experiments have investigated the problems of foreign-language text retrieval, question answering and image retrieval, including aspects such as query formulation, translation and refinement, and document selection. The focus has always been on improving

the outcome of the process in terms of a classic notion of relevance, and the target collection has always consisted of news texts in languages foreign to the user. Finally, the task has always involved the comparison of a reference system with a contrastive system, combining users, topics and systems with a latin-square design to detect system effects and filter out other effects.

Although iCLEF in only a few years of activity has established the largest collected body of knowledge on the topic of interactive Cross-Language Retrieval, the experimental setup has proven limited in certain respects:

- The search task itself was unrealistic. News collections are comparable across languages, and most of the pertinent information tends to be available in the user's native language; why would a user search for this information in an unknown language? Translating this problem to the WWW, it would be like asking a Spanish speaker to find information about the singer Norah Jones in English, in spite of the fact that there are over 150,000 pages in Spanish about her¹.
- The target notion of "relevance" does not cover all aspects that make an interactive search session successful.
- The latin-square design imposed heavy constraints on the experiments, making them costly and with a limited validity (the number of users was necessarily limited, and statistically significant differences were hard to obtain).

In order to overcome these limitations, we have decided to propose a new, pilot framework for iCLEF which has two essential features:

- We have chosen Flickr (the popular photo sharing service) as the target collection. Flickr is a large-scale, web-based image database serving a large social network of WWW users. It has the potential to offer both challenging and realistic multilingual search tasks for interactive experiments.
- We want to use the iCLEF track to explore alternative evaluation methodologies for interactive information access. For this reason, we have decided to fix the search tasks, but to keep the evaluation methodology open. This allows each participant to contribute with their own ideas about how to study interactive issues in cross-lingual information access.

The remainder of the paper describes the track guidelines in detail, and summarizes the experiments conducted by the research groups that submitted results.

2 The target collection: Flickr

The majority of Web image search is text-based, and the success of such approaches often depends on reliably identifying relevant text associated with a particular image. Flickr is an online tool for managing and sharing personal photographs, and currently contains over thirty million freely accessible images². These are updated daily by a large number of registered users and available to all web users.

2.1 Photographs in the collection

Flickr provides both private and public image storage, and photos which are shared (around 5 million) can be protected under a Creative Commons (CC) licensing agreement (an alternative to full copyright). Images from a wide variety of topics can be accessed through Flickr, including people, places, landscapes, objects, animals, events, etc. This makes the collection a rich resource for image retrieval research.

¹google.com results as of 11 August 2006.

²As of August, 2006.

There were two possibilities to use the collection: reaching an agreement with Flickr to get a subset of their database, or simply using Flickr's public API to interact with the full database. The first option is more attractive from the point of view of system design, because it is possible to obtain collection statistics to enhance the search (for instance, tf-idf weights, term suggestions and term translations adapted to the collection) and because it gives total control on the search mechanism. A crucial advantage of having the collection locally stored is enabling the possibility of doing content-based retrieval.

The second option, by contrast, by design reflects the dynamic nature of open databases and allows users to interact with a larger and more current target database - which is more realistic and thus preferable. While recall-oriented success measures are ill defined, collection frequencies more complex to calculate, and reproducibility of results can be called into question, this option relieves organisers and participants alike from cumbersome administration and distribution of test collections. In any case, until an agreement with Flickr is reached, using Flickr's API to access the full Flickr database was the only choice available for this first pilot task.

As of October 2005, 1.2 million Flickr users added around 200,000 images a day to the collection³. In order to keep the collection somewhat more constant across the experiments, we decided to restrict the experiments to images uploaded before 21 June 2006 (immediately before iCLEF experiments began).

2.2 Annotations

In Flickr, photos are annotated by authors with freely chosen keywords ("tags") in a naturally multilingual manner: most authors use keywords in their native language; some combine more than one language. User tags may describe anything related to the picture, including themes, places, colours, textures and even technical aspects on how the photograph was taken (camera, lens, etc.). Some tags become naturally "standardized" among subsets of Flickr users, in a typical process of so-called "folksonomies" [5].

In addition, photographs have titles, descriptions, collaborative annotations, and comments in many languages. Figure 1 provides an example photo with multilingual annotations.

Photos can also be submitted to online discussion groups. This provides additional metadata to the image which can also be used for retrieval. An explore utility provided by Flickr makes use of this user-generated data (plus other information such as viewing statistics and ratings) to define an "interestingness"⁴ view of images.

2.3 Flickr's search API

While Flickr has many modes in which users can explore the photo collection, its search capabilities are rather simplistic. One can choose between searching tags only, or full text search (title, description and tags). In the tag searching mode there are two options: conjunctive (all keywords) or disjunctive (any keyword). In the full text search, only the conjunctive mode is provided⁵. This is a serious restriction when performing cross-language searches that had to be taken into account by participants in the task.

During the experiments, the API has offered fast and stable response times.

3 The task

Images are naturally language independent and often successfully retrieved with associated texts. This has been explored as part of ImageCLEF (Clough et al, 2005) for areas such as information access to medical images and historic photographs. This makes images specially attractive as a scenario where cross-language search arises more naturally.

³<http://www.wired.com/news/ebiz/0,1272,68654,00.html>

⁴<http://www.flickr.com/explore/interesting>.

⁵At the time of writing this overview, Flickr has started providing full boolean search, which can be exploited in future experiments.

outono/inverno



começando a série "velha, mas enxuta" uma da BR 116 num dia assim como o de hoje frio e chuvoso.

vontade de fazer esse mesmo percurso com mais calma dia desses, hein Hartmann?

grande abraço e bom feriado,

750 x 1125

de volta ao mundo dos interestingness!

Uploaded on June 15, 2006
by [corbata1982](#)

**corbata1982's
photostream**

This photo also belongs to:

menu (Set)

tus faves (Set)

mis faves (Set)

" La tengo y te reto "
(Pool)

" TE DESAFIO " (Pool)

1 million photos (Pool)

Tags

- BR116
- estrada
- way
- via
- caminho
- árvores
- trees
- tree
- arbol
- arboles
- arvore
- chuva
- rain
- lluvia
- frio

Figure 1: An example Flickr image, with title, description, classified in three sets (user-defined) and three pools (community shared), and annotated with more than 15 English, Spanish and Portuguese tags.

The way in which users search for images provides an interesting application for user-centered design and evaluation. As an iCLEF task, searching for images from Flickr presents a new multilingual challenge which, to date, has not been explored. Challenges include:

- Different types of associated text, e.g. tags, titles, comments and description fields.
- Collective classification and annotation using freely selected keywords (folksonomies) resulting in non-uniform and subjective categorization of images.
- Fully multilingual image annotation, with all widely-spoken languages represented and mixed-up in the collection.
- Large number of images available on a wide variety of topics from different domains.

Given the multilingual nature of the Flickr annotations, translating the user’s search request would provide the opportunity of increasing the number of images found and make more of the collection accessible to a wider range of users regardless of their language skills. The aim of iCLEF using Flickr will be to determine how cross-language technologies could enhance access, and explore the user interaction resulting from this.

The experiment consists of three search tasks, where users may employ a maximum of twenty minutes per task. We have chosen three tasks of a different nature:

- Topical ad-hoc retrieval: *Find as many European parliament buildings as possible, pictures from the assembly hall as well as from the outside.*
- Creative open-ended retrieval: *Find five illustrations to the article “The story of saffron” (see the text in Figure 2).*
- Visually oriented task: *What is the name of the beach where this crab is resting?* (along with a picture of a crab lying in the sand, see Figure 3). The name of the beach is included in the Flickr description of the photograph, so the task is basically finding the photograph, which is annotated in German – a fact the users is unaware of.

All tasks can benefit from a multilingual search: Flickr has photographs of European parliament buildings described in many languages, photographs about the Abruzzo area and saffron are only annotated in certain languages, and the crab photograph can only be found with German terms. At the same time, the nature of each task is different from the others. The “European parliaments” topic is biased towards recall, and one can expect users to stop searching only when the twenty minutes expire. The text illustration task only demands five photographs, but it is quite open-ended and very much depending on the taste and subjective approach of each user; we expect the search strategies to be more diverse here. Finally, the “find the crab” task is more of a known-item retrieval task, where the image is presumed to be annotated in a foreign language, but the user does not know which one; the need for cross-language search and visual description is more acute here.

Given that part of the experience consisted of proposing new ways of evaluating interactive Cross-Language search, we did not prescribe any fixed procedure or measure for the task. To lower the cost of participation, we provided an Ajax-based basic multilingual interface to Flickr, which every participant could use as a basis to build their systems.

4 Experiments

Fourteen research groups officially signed in for the task, more than in any previous iCLEF edition. However, only the three organizing teams (SICS, U. Sheffield and UNED) submitted results (the worst success rate in iCLEF campaigns). Finding the reasons is left to discussion along the workshop; perhaps the fact that the task is only half-cooked made people feel insecure about

Abruzzo Dishes: The Story of Saffron

Obtained from the dried and powdered stem of the 'crocus sativus' which grows on the Navelli Plain in the Province of L'Aquila, saffron is considered by many to be the single most representative symbol of the traditional products of Abruzzo.

An essential ingredient in Risotto Milanese, the spice also crops up in many other dishes across Italy. For example, the fish soup found in Marche, south of the Monte Conero, contains saffron for its red coloring in place of the more traditional tomato. This coloring property is also widely appreciated in the production of cakes and liqueurs and for centuries by painters in the preparation of dyes. Its additional curative powers have long been known to help digestion, rheumatism and colds.

How a flower of Middle Eastern origin found a home in this unfashionable corner of Italy can be attributed to a priest by the name of Santucci who introduced it to his native home 450 years ago. Following his return from Spain at the height of the Inquisition, his familiarity with Arab-Andalusian tradition convinced him that the cultivation of the plant was possible in the plains of Abruzzo, and so it proved.

Nevertheless, even today the harvesting of saffron is hard and fastidious work with great skill needed to handle the stems without damaging the product inside or allowing contamination from other parts of the plant. The area of cultivation in the region is strictly limited to 8 hectares of land. A sad reduction from the 430 hectares cultivated at the turn of the last century.

Together with the labor intensiveness of the production and the care and patience involved in gathering and drying the flower, the cost of Abruzzese saffron is high relative to its competitors from the Middle East and Iran. Yet all are agreed it possesses superior aromatic qualities and remains the preferred choice in gourmet cooking. It is so good that the saffron from the area is practically all exported.

Anyone interested in buying the end product should note that although sachets of saffron powder can be purchased, the real thing should only be bought as the characteristic dried fine stems.

The saffron is grown in an area comprising the comune of Navelli, Civitaretenga, Camporciano, San Pio delle Camere and Prata D'Ansidonia.

Figure 2: Creative task: Find five illustrations to this text



Figure 3: Visually oriented task: *What is the name of the beach where this crab is resting?*

what or how to measure. We hope that the results obtained by the three organizing groups will encourage a much broader participation next year.

This is a brief summary of the experiments performed at iCLEF 2006:

- UNED [1] measured the attitude of users towards cross-language searching when the search system provides the possibility (as an option) of searching cross-language, using a system which allowed for three search modes: “no translation”, “automatic translation” (the users chooses the source language and the target languages, and the system chooses a translation for every word in the query) and “assisted translation” (like the previous, but now the user can change the translation choices made by the system). Their results over 22 users indicate that users tend to avoid translating their query into unknown languages, even when the results are images that can be judged visually.
- U. Sheffield (and IBM) [2] experimented with providing an Arabic interface to Flickr in an attempt to increase the amount of material accessible to the Arabic online community. An Arabic-English dictionary was used as an initial query translation step, followed by the use of Babelfish to translate between English and French, German, Dutch, Italian and Spanish. Users were able to modify the English translation if they had the necessary language skills. With the user group testing the system (bilingual Arabic-English students), it was found that these users: preferred to query in English (although liked having the option available to them of formulating the initial queries in Arabic), found the system very helpful and easy to use, and overall were able to search effectively in all tasks provided. Users found viewing photos with results in multiple languages more helpful and important than the initial query translation step from Arabic to English. Users needed to view the image annotations for most tasks, but had no problems doing this for the languages available to them. If non-European languages had been included (e.g. Chinese or Japanese) then users would not have been able to use the annotations effectively. An analysis of the tasks was also carried out and included in the results. Overall it was found that these tasks were not well-suited to Arabic users.
- SICS [3] The experiments carried out at SICS this year centered on user satisfaction and user confidence as target measures for information access evaluation. The users were given the tasks, and after some time were given a terminological display of terms they had made use of together with related terms. This enabled them to broaden their queries: success was not measured in retrieval results but in change of self-reported satisfaction and confidence as related to the pick-up of displayed terms by the user[4].

5 Conclusions

Flickr would certainly seem to provide cross-language research with a highly plausible and well-motivated cross-language search scenario, especially given that results are visual in nature. The goal of iCLEF 2006 has been to investigate this particular cross-language scenario with users on a range of search tasks designed to favour multilingual search. While the number of participants has always been low in any organised evaluation events involving users, this year number of submissions cannot be viewed as anything but disappointing: the lowest turnout so far in iCLEF. However, the interest as measured by pre-registration was considerable. This year, the lessons learnt are mainly methodological — can target measures based on other notions than “relevance” give interesting results? The submissions to iCLEF do, however, offer some insights into cross-language searching in Flickr and these will be discussed at the CLEF workshop, together with a discussion on how participation better could be encouraged for future evaluation cycles.

Acknowledgments

This work has been partially supported by the Spanish Government under project R2D2/Syembra (TIC2003-07158-C04-02) and by the European Commission, project MultiMatch (FP6-2005-IST-5-033104) and the DELOS Network of Excellence (FP6-G038-507618).

References

- [1] Javier Artiles, Julio Gonzalo, Fernando López-Ostenero, and Víctor Peinado. Are users willing to search cross-language? an experiment with the flickr image sharing repository. In *This volume*, 2006.
- [2] Paul Clough, Azzah Al-Maskari, and Kareem Darwish. Providing multilingual access to flickr for arabic users. In *This volume*, 2006.
- [3] Jussi Karlgren and Fredrik Olsson. Trusting the results in crosslingual keyword-based image retrieval. In *This volume*, 2006.
- [4] Jussi Karlgren and Magnus Sahlgren. Automatic bilingual lexicon acquisition using random indexing of parallel corpora. *Natural Language Engineering*, 11(3):327–341, 2005.
- [5] Adam Mathes. Folksonomies-cooperative classification and communication through shared metadata. Technical Report LI590CMC, Computer Mediated Communication, Graduate School of Library and Information Science, University of Illinois Urbana-Champaign, 2004.