Eurovision – an image-based CLIR system

Mark Sanderson and Paul Clough

University of Sheffield, UK

m.sanderson@shef.ac.uk, p.clough@dcs.shef.ac.uk

Abstract

This short position paper outlines the aims of the Eurovision project, which will explore cross language retrieval of images via their text captions. We propose that CLEF, as a part of its future works, examine retrieval of multimedia data. By describing the areas to be explored by Eurovision, the paper shall show how multimedia retrieval across languages promotes exploration of interaction, requires potentially more sophisticated language processing, and ultimately stimulates research in an area of CLIR where there appears to be commercial potential. Finally, we propose that CLEF use the Eurovision image collection that is currently being built.

Introduction

A common question asked of someone conducting cross language information retrieval (CLIR) is "why would anyone wish to retrieve documents they presumably can't read?". Over the years from early CLIR research, answers to this question have been improved and refined with progressively better responses being developed. The original replies postulated users that were readers not writers of a language or had access to costly translation resources, which needed to be fed limited numbers of documents. It would appear that these are the users CLEF and iCLEF currently have in mind when designing retrieval experiments. While undoubtedly, such users exist, in our limited experience of the two cross language projects being run in Sheffield, there are other, potentially larger, user groups with alternate needs that require the attention of the cross language community.

There are those users who are fluent in more than one language. For them, it is simply annoying that in order to search across all the languages they know, queries to existing multi-language text searching tools (e.g. web search engines) are entered multiple times, once for each language known to the users. Such people were found to be common in the user needs and up-coming testing work of the Clarity project (Hansen, 2002). Those with similar bi- or tri-lingual skills constituted the large majority of the population used in the testing phases of the early cross language project, MULINEX (Capstick, 2000). How well this class of user is served by the current research in interactive CLEF is not clear. What is also not clear is how significant this population is; they may not constitute the majority of *people* overall, but they may constitute the majority of *users* who wish to access text-based CLIR.

However, although we consider searching systems for users with strong language skills an important future area for CLEF, they are not the focus of this paper. Instead, it is our contention that there are areas of cross language retrieval where there will be users wishing to search in languages they have no knowledge of. One area of CLIR research that has received almost no attention is retrieving from collections where text is used only to describe the collection objects, and the objects themselves are understandable by anyone regardless of their language skills. One such collection is a picture archive where each image is described by a text caption. Here a wide range of people may wish to retrieve from the collection regardless of the language they speak. For any vendor of an image library, use of CLIR offers the opportunity of broadly expanding the range of potential searchers of their library.

Retrieval from such an archive presents a number of challenges and opportunities. The challenges come from matching queries to the typically short descriptions associated with each image. The opportunities derive from the unusual situation for CLIR systems of users being able easily to judge the relevance of retrieved objects and to feed this information back into the retrieval system. Use of this information may enhance the user's current query, improve image representation, or provide information to the translation system on how to better translate a similar query at a later date. The research outlined in this document will build and test an image CLIR system. Starting with a description of the retrieval problem, the paper outlines how such a form of CLIR is likely to foster more interaction with users and, therefore, allows a better study of it in the context of cross language. A review of the few pieces of past work in this area are next discussed followed by a presentation of the implications of interaction in a cross language environment. Finally, an image retrieval task for future CLEF is proposed.

Past work

Because there is almost no work investigating cross language image retrieval, this Section will first review the two component research areas separately: image retrieval by associated text and cross language IR followed by an examination of the proposals and occasional attempts at performing image CLIR.

Retrieval of images by text queries matched against associated text has long been researched. As part of his PhD investigating multimedia retrieval, Dunlop examined such an approach to image retrieval (1993). The ideas in this work were later extended to a study of image retrieval from art gallery Web sites by Harmandas et al (1997), who showed that associated text was well suited for retrieval over a range of query types. At the same conference, research was presented on the successful use of a thesaurus to expand the text of image captions (Aslandogan et al 1997). More recently, research in combining content-based image retrieval with caption text has been explored in the work of Chen et al (1999).

Most CLIR research has concentrated on locating and exploiting translation resources. Successful methods centre on use of either bilingual dictionaries, machine translation systems, or parallel corpora. With reasonably accurate translation, effective cross language retrieval is possible as confirmed by TREC and CLEF (thoroughly reviewed in Gollins, 2000). While there are still research questions in translation for CLIR, there is increasing interest on the user's role in CLIR: in query formulation; and on the ability of systems to produce translated surrogates of retrieved documents. Resnik reported his early work on so-called *gists* of documents (1997), however evaluation of the system was such that it is hard to speculate on users' abilities to judge relevance from the gists. The CLEF evaluation plans this year to feature an interactive evaluation where the user's ability to judge the relevance of document surrogates will be measured.

Both Oard (1997) and Jones (2001) have discussed cross language retrieval of images, but neither reported any work in the area. However, two examples of cross language image retrieval exist. The IR Game system built at Tampere University (Sormunen, 1998) offers Finish/English cross language retrieval from an image archive. Images are ranked using a best match search, but no form of query or image caption expansion is used. The European Visual Archive (EVA, http://www.eva-eu.org/) offers English/Dutch/German cross language searching of 17,000 historical photographs indexed by a standard set of 6,000 controlled terms. Searching is restricted to Boolean searching. The two works differ greatly from that proposed here as no evaluation of the retrieval effectiveness has been undertaken, searching is limited through lack of expansion or offering only Boolean searching, and relevance feedback is not supported.

How do multimedia ideas impact on CLEF?

Image retrieval is, we believe, worth examining in the context of cross language, as it presents a series of challenges that have not be explored as much in document-based CLIR.

1. The captions of images are generally shorter than documents, it may be necessary, in the context of CLIR, to examine query expansion. Methods of translations used in document CLIR may implicitly assume a large body of text to match on exists and if only captions are being retrieved, translation may not be as successful. In monolingual retrieval, it has been found that processes such as stemming have a much greater impact on retrieval effectiveness, such methods may need to be more broadly applied in multimedia CLIR.

2. Queries to image libraries often contain verbs (e.g. "lions sleeping", "children running", etc.), document retrieval queries typically hold only nouns. Such differences in grammatical category may impact on translation methods also.

3. The key difference in image retrieval is the likelihood of much more user interaction in the retrieval system. It is relatively trivial for users to judge the relevance of retrieved images and relevance feedback is much more likely to be conducted by users of such a system. As soon as users mark documents as relevant, the CL system has a query composed of both source and target language terms, thereby potentially improving the quality of retrieval being conducted. Such feedback information can also be used long term in a series of ways.

   a. The documents marked as relevant by users will contain certain of the users' terms translated in a particular way. The translation system can observe the feedback being provided by users and apply this information, long-term, to guide its means of translating the same terms for later user, a form of translation memory.

   b. It is anticipated that further novel studies can be conducted. When translating, a major source of error comes from the ambiguity of words: for example, the Spanish word "banca" is

generally translated to "bank" but can also be translated to "stand", "stall", "bench", "pull", or "influence". The CLIR translation process is unlikely to know what sense a query word is being used in and will often have to translate to all possible senses. When users mark retrieved images as being relevant to their query, the system can examine how query words were translated into the matching target language words of the image caption. One can reasonably assume that if the image is marked as relevant for a particular query, the query's translation was correct for that image. Consequently, this information can be stored to enable better translations should the users' query, or variations of it, be re-submitted again. The utility of user feedback information in the long-term improvement of the translation process is an important future study stemming from this work.

c. One further use of relevance feedback is perhaps the most intriguing, as it may be possible to use such information to automatically disambiguate (i.e. identify the sense of) the words of image captions. The method can best be illustrated with an example. As above, the Spanish word "banca" is translated into (amongst others) "bank", which is itself ambiguous: potentially referring to the edge of a river; a banked corner, as well as an economic institution. It is common for translation resources to identify the sense of the translated word: e.g. identifying the economic sense of "bank" as a translation of "banca". This information is often discarded by CLIR systems, however, it can be exploited if relevance feedback takes place. To illustrate, if an image retrieved in response the query "banca" was marked as relevant by a user and within the image's caption was the English word "bank", one could reasonably conclude that this use of "bank" refers to the economic sense[1]. Therefore, for this occurrence of this word, its sense has been accurately determined. Use of relevance feedback in this way constitutes a novel approach to disambiguation. As Smeaton and Quigley (1996) demonstrated, if the words of image captions are disambiguated (in their case, manually), expansion of the caption text using thesauri relations is more accurate and retrieval effectiveness subsequently improves. A study of the accuracy and coverage of a method exploiting user feedback for image caption disambiguation would be another future direction for the image CLIR work.

4. Commercial take up of CLIR technology appears to have been relatively slow. It is our contention that in the field of multimedia retrieval this will not be the case. Stock photography archives generate revenue through page impressions (e.g. adverts) and direct sales. The revenue earned is proportional to the number of users accessing the collection. If one can increase the number of users, revenue will grow. One solution is to make images retrievable in other languages, thereby, allowing access to a wider range of markets. Those markets are projected to grow significantly: non-English speakers being the majority of Web users by 2005[2]. Many existing collections are captioned in English only, however: an examination of the Yahoo stock photography section shows that less than 5% of listed sites offer searching in languages other than English. Translation of existing catalogues is expensive and most likely would be only done for common languages. As stock is added, further translation would have to be performed. Installing a cross language component into image searching systems is an attractive cost effective solution to the problem of providing cross language search.

As can be seen, multimedia retrieval presents a wide range of opportunities and challenges that are worthy of wide ranging exploration.

Proposal for CLEF

We believe that multimedia cross language retrieval is an important issue to consider in CLEF and through writing this exposition of ideas, we hope that others will be encouraged to focus research in this area. The Eurovision project will, as part of its work, be creating an image test collection. The plan is to use a photography library from St. Andrews University in Scotland, which is composed of several tens of thousands of black and white photographs. Issues of copyright and levels of access to the high resolution digitised images themselves is still being negotiated, but the captions of the collection are (barring some last minute change of mind) going to be made available to anyone who wishes to work with them. A set of queries is currently being developed in Sheffield along with a set of limited relevance judgements. Given the nature of test collection evaluation, where

---

[1] This assumes that the Spanish word "banca" does not have the same ambiguity as its English translation. For almost all words, however, this holds true.

[2] Taken from http://www.nua.ie/surveys/index.cgi?f=VS&art_id=905354950&rel=true

the content of the retrieved document is almost always ignored, we believe that it will be possible to release the Eurovision collection for general use. We would like others in CLEF to utilise our collection and conduct work of their own on this data set. The benefit to Eurovision would be through others providing of a wider range of languages the queries would be translated into and a deeper pool of relevance judgements being created. Obviously, interactive retrieval is restricted by the level of access to the images in the collection. Access to thumbnails is likely (possibly to a restricted group) therefore, interactive image retrieval should also be possible.

It is our hope that the practicalities of this proposal can be discussed at the workshop.

References

Aslandogan. Y.A., Thier. C., Yu, C.T., Zou, J., Rishe, N. (1997) Using semantic contents and WordNet in image retrieval, In the *Proceedings of the 20th ACM SIGIR conference*, pp 286-295

Ballesteros, L., Croft, W.B. (1998): Resolving ambiguity for cross-language retrieval, in *Proceedings of the 21st annual international ACM SIGIR*, pp 64-71

J. Capstick, A. K. Diagne, G. Erbach, H. Uzkoreit, A. Leisenberg M. Leisenberg (2000), A system for supporting cross-lingual information retrieval, Information Processing and Management, 36(2):275–289.

Chen, F., Gargi, U., Niles, L., Schuetze, H. Multi-Modal Browsing of Images in Web Documents, *Proceedings of SPIE Document Recognition and Retrieval VI*, pp. 122-133, 1999

M. D. Dunlop & C. J. van Rijsbergen (1993), Hypermedia and free text retrieval, *Information Processing and Management*, 29(3).

Gollins, T. (2000) Dictionary Based Transitive Cross-Language Information Retrieval Using Lexical Triangulation, Masters dissertation, Information Studies Department, University of Sheffield.

P. Hansen, D. Petrelli, M. Beaulieu, M. Sanderson (2002), User requirements elicitation for cross-language information retrieval, to appear in the *proceedings of the ISIC conference*, 2002.

Harmandas, V., Sanderson, M. Dunlop, M.D. (1997) Image retrieval by hypertext links, In the *Proceedings of the 20th ACM SIGIR conference*, pp 296-303

Jones, G.J.F., New Challenges for Cross-Language Information Retrieval: Multimedia Data and the User experience. In Carol Peters (ed.). *Cross-Language Information Retrieval and Evaluation: Proceedings of the CLEF 2000 Workshop, Lecture Notes in Computer Science 2069*, Springer 2001, pp 71-81.

Oard, D. (1997) Serving Users in Many Languages: Cross-Language Information Retrieval for Digital Libraries In *D-Lib Magazine*, http://www.dlib.org/

Resnik, P. (1997): Evaluating Multilingual Gisting of Web Pages, in *AAAI Spring Symposium on Cross-Language Text and Speech Retrieval Electronic Working Notes*.

Sormunen, E., Laaksonen, J., Keskustalo, H., Kekäläinen, J., Kemppainen, H., Laitinen, H., Pirkola, A., Järvelin, K., (1998) The IR Game - A Tool for Rapid Query Analysis in Cross-Language IR Experiments. *PRICAI '98 Workshop on Cross Language Issues in Artificial Intelligence*. pp. 22-32.