Pattern-based image retrieval with constraints and preferences on ImageCLEF 2004*

Maximiliano Saiz-Noeda, José Luis Vicedo and Rubén Izquierdo

Departamento de Lenguajes y Sistemas Informáticos.
University of Alicante. Spain
{max,vicedo,ruben}@dlsi.ua.es

Abstract. This paper presents the approach used by the University of Alicante in the ImageCLEF 2004 adhoc retrieval task. This task is performed by multilingual search requests (topics) against an historic photographic collection in which images are accompanied with English captions.

This approach uses these captions to make the retrieval and is based in a set constraints and preferences that will allow the rejection or the scoring of the images for the retrieval task. The constraints are implemented through a set of co-occurrence patterns based on regular expressions and the approach is extended in one of the experiments with the use of WordNet synonym relations.

1 Introduction

Bilingual ad hoc retrieval is one of the tasks defined within the ImageCLEF 2004 campaign [1] as part of the Cross Language Evaluation Forum (2004). The objective of this task, celebrated since last 2003 campaign [2], is to retrieve relevant photographic documents belonging to a historic photographic collection in which images are accompanied with English captions. These photographs integrate the *St Andrews photographic archive* consisting of 28,133 (aproximately 10% of the total) photographs from *St Andrews University Library photographic collection* [3].

The method followed to retrieve the relevant images is based on three experiments where a set of preferences and constraints are applied. The constraints, based on a set of co-occurrence patterns will reject the potential incompatible (non-relevant) images related to the question. Preferences will score the images in order to give a list according to their relevance degree. Furthermore, a Wordnet-based query expansion is tested.

This is the first time that the University of Alicante participates in this specific task and the main objective in the starting premise is to make a simple and low cost approach just to test its possibilities.

* This work has been partially supported by the Spanish Government (CICYT) with grant TIC2003-07158-C04-01.

Next sections describe specific characteristics of the dataset, relevant for the retrieval process, just as the strategy set up by the University of Alicante's team in order participate in the forum. Finally, some evaluation results will be discussed and some of the future intervention lines in order to improve the system and provide better results will be presented.

2 Photographic dataset

As mentioned, the photographic dataset used for the Image Clef 2004 evaluation is a collection of 28,133 historical images from *St Andrews University Library photographic collection*. Photos are primarily historic in nature from areas in and around Scotland; although pictures of other locations also exist.

All images have an accompanying textual description consisting of a set of fields. In this apporach, we have used a file containing all image captions in TREC-style format as detailed below:

```
<DOC>
 <DOCNO>stand03_2096/stand03_10695.txt</DOCNO>
 <HEADLINE>Departed glories - Falls of Cruachan Station above Loch
 Awe on the Oban line.</HEADLINE>
 <TEXT>
  <RECORD_ID>HMBR-.000273</RECORD_ID>
  <SHTITLE>Falls of Cruachan Station.</SHTITLE>
  <DESCRIPTION>Sheltie dog by single track railway below embankment,
  with wooden ticket office, and signals; gnarled trees lining
  banks.</DESCRIPTION>
  <DATE>ca.1990</DATE>
  <PHOTOGRAPHER>Hamish Macmillan Brown</PHOTOGRAPHER>
  <LOCATION>Argyllshire, Scotland</LOCATION>
  <NOTES>HMBR-273 pc/ADD: The photographer's pet Shetland collie
  dog, 'Storm'.</NOTES>
  <CATEGORIES>[tigers],[Fife all views],[gamekeepers],[identified
  male],[dress - national],[dogs]</CATEGORIES>
  <SMALL_IMG>stand03_2096/stand03_10695.jpg</SMALL_IMG>
  <LARGE_IMG>stand03_2096/stand03_10695_big.jpg</LARGE_IMG>
 </TEXT>
</DOC>
```

The 28,133 captions consist of 44,085 terms and 1,348,474 word occurrences; the maximum caption length is 316 words, but on average 48 words in length. All captions are written in British English, although the language also contains colloquial expressions. Approximately 81% of captions contain text in all fields, the rest generally without the description field. In most cases the image description is a grammatical sentence of around 15 words. The majority of images (82%) are in black and white, although colour images are also present in the collection.

The type of information that people typically look for in this collection include the following: Social history, e.g. old towns and villages, children at play and work. Environmental concerns, e.g. lanscapes and wild plants. History of photography, e.g. particular photographers. Architecture, e.g. specific or general places or buildings. Golf, e.g. individual golfers or tournaments. Events, e.g. historic, war related. Transport, e.g. general or specific roads, bridges etc. Ships and shipping, e.g. particular vessels or fishermen.

Although all these fields can be used individually or collectively to facilitate image retrieval, in this approach only a few of them have been used, in particular, fields related to the photographer, location and date, apart from the headline, have been selected for the retrieval.

3 Technique description

As it is the first time this group participates in this task, we decided to make a naive approach with the smallest possible quantity of resources and time-consuming. So, this technique does not use any kind of indexing, dictionary or entity recognition and it makes use of a single POS tagging. Nevertheless, within the three developed experiments, improvements of the method includes the use of co-occurrence patterns and WordNet for the query expansion.

Figure 1 shows the process followed by the system. This figure includes three steps related to the three experiments carried out for the evaluation that will be detailed below.

For the application of the basic strategy it is necessary to count on a file with question files and image dataset. As mentioned, the file with the whole set of images in trec format has been used for the retrieval.

Constraints and preferences applied to the retrieval process make use of morphological information. Furthermore, the retrieval process is based on word lemmas. So, a POS tagging of both the question and the image info is necessary. This POS tagging has been made using the TreeTagger analyzer [4]. For the retrieval process itself, a file of stop words have been used in order to eliminate useless words and improve the system speed.

In order to cope with multilingual retrieval aspects, the use of a translator has been planned. In concrete Babelfish [5] translator has been used. This resource has allowed to test the system with topics in German, Chinese, Spanish, French, Dutch, Italian, Japanese and Russian. Following this, all the languages have been equally treated from their translation into English.

According to the information needed for the retrieval, three different experiments have been carried out:

1. Preferences-based retrieval
2. Constraints and Preferences-based retrieval
3. Constraints and Preferences-based retrieval with question expansion

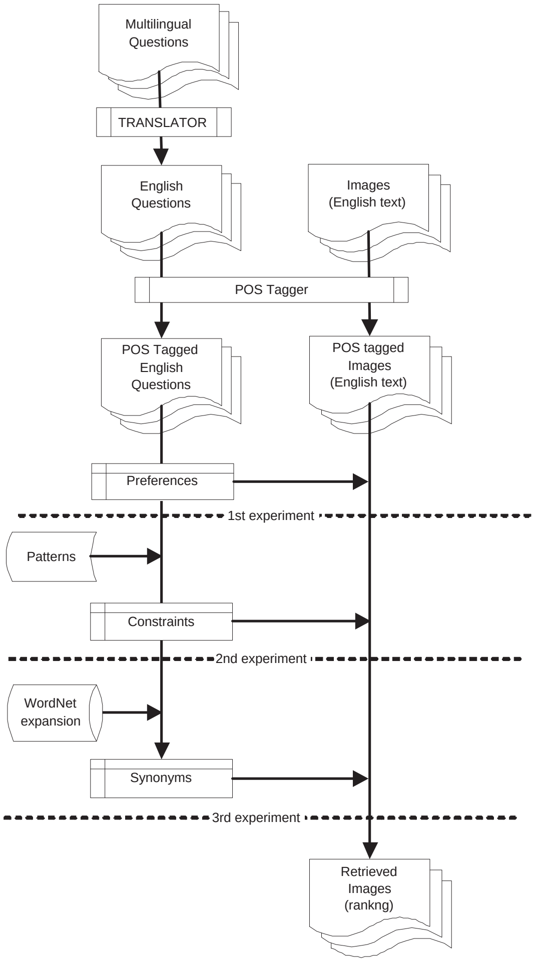

Fig. 1. Image retrieval process and for different experiments

At the beginning of the process, all the images are suggested as solution for each question[1]. From this scoring, some images will be added punctuation according to the experiment development.

3.1 First experiment. Preferences

For the preference applying, a single word matching between the question and the *HEADLINE* field of the image is used. This experiment is used as a baseline

[1] This condition is guided by the idea of giving 1000 images for each question, what constitute a misunderstanding of the evaluation process and will be discussed bellow as an evaluation handicap.

and its main interest, as it will be discussed later, is the way the information added in the other experiments affects to the retrieval process.

For the scoring in this experiment, we have assumed the relevance of proper nouns, nouns and verbs so we have scored following this order when the matching is related to these elements.

Furthermore, if applicable, relations between nouns and verbs with the same lemma are also scored (if we are looking for “golfers swinging their clubs” probably we are interested in a “golfer’s swing”).

It seems almost obvious that a good performance of this technique should be based in a good entity recognition that ensure the correct detection of proper nouns in the question and in the image text info. Probably, as it will be discussed later, counting on a named entities recognizer would improve the overall performance of this experiment is not as good as desired.

3.2 Second experiment. Constraints and preferences. The patterns

This experiment makes use of the previously described preferences and integrates the constraints as a new selection criterion. The main aspect of the constraint is that it should be a rule strong enough to reject an image based on a compatibility guideline. This rule is built through the definition of a set of co-occurrence patterns that establish rejecting rules related to three of the fields contained in the image information: *DATE*, *PHOTOGRAPHER* and *LOCATION*.

These patterns are applied to the question (topic) and generates a XML-style file with the information provided by the patterns. For example, topic:

```
1. Portrait pictures of church ministers by Thomas Rodger
```

is converted into the file:

```
<PREG>
 <PREGNO>1</PREGNO>
 <HEADLINE> Portrait pictures of church ministers by Thomas
  Rodger</HEADLINE>
 <DATE> </DATE>
 <PHOTOGRAPHER> by Thomas Rodger </PHOTOGRAPHER>
 <LOCATION> </LOCATION>
</PREG>
```

where labels $< DATE >$, $< PHOTOGRAPHER >$ and $< LOCATION >$ contain all the information extracted about these information items.

The patterns are built over regular expressions that allow the extraction of a string contained in any of the mentioned labels. *DATE* constraints treat to reject images not only comparing question and image years, but applying extra information such as months or quantifiers. This way, if the topic asks for “Pictures of Edinburgh Castle taken before 1900”, all the photos taken after 1900 will be discarded. *PHOTOGRAPHER* constraints are based in the whole name of the photographer. *LOCATION* constraints use not only the location itself but also, if applicable, possible relations with other locations (city, country, . . .).

As can be seen, this technique is very general and, therefore, the possibility of errors is high. To soften this error possibility, the strategy has also in mind statistical information from the image corpus. This way, things that can be incorrectly treated by the pattern as photographers or locations are considered as noise and rejected because of their low or null appearance frequency in the corresponding field in the image. In fact, we can use the same pattern for both location and photographer and then decide what to apply depending on the image. For example, according to the image info, a capitalized word after a comma can be consider both a photographer or a location (as shown in topics "Men in military uniform, *George Middlemass Cowie*" and "College or university buildings, *Cambridge*"). After including the extracted string in both fields of the topic generated, the statistical information will determine what is a photographer and what is a location (unless there is a town called George Middlemass or a photographer called Cambridge).

Once the constraint features are determined and included in the topic through their corresponding labels, the system makes a matching task to reject non-compatible images. So, if it is determined that the photographer of the searched pictures is "Thomas Rodger", all the images that don't contain "Thomas Rodger" (or any of its parts) in the *PHOTOGRAPHER* label are rejected.

3.3 Third experiment. Query expansion. Wordnet

The last experiment has been designed to incorporate extra information regarding to the potential synonym relation between terms in query and image. In this case, the system expands the topic with all the synoyms of nouns and verbs contained in WordNet [6].

According to this, in this case, the scoring for each image is increased if not only a lemma of a word in the topic, but its synonyms in WordNet, appear in the image *HEADLINE* text.

Due to there is no lexical disambiguation in the proccess, noun synonyms are best scored than verb synonyms assuming that the former tend to be less generic than the latter. If the synonym is found but with different POS tag, a smaller score is added.

4 Evaluation

Although we knew this is a very general approach to this task, the results obtained after the evaluation of the system are not as successful as desired. At the moment of the writing of this paper we are trying to determine if there is any kind of computing processing mistake that has affected the final scoring. Anyway, there are some considerations extracted from the results.

For the evaluation results, the system was prepared to give always 1000 images as output. This is an error because some (sometimes a lot of) images given by the system are not relevant at all (they have no specific nor score).

Another problematic issue is the way the system score the images. This scoring is also very general and provides many times the same score for a big quantity of images (in fact, all the images can be grouped in four or five different scores). All the images that are equally scored have, for the system, the same order in the final evaluation scored list, but for the evaluation comparing results there are big differences depending on the order provided for the images.

Related to the results of the three experiments, one of the most "eye-catching" thing is that, in general, the preferences-baseline experiment gives the best result or is improved in a very small degree by the rest of experiments. This situation can be put down to the lack of additional information regarding to the entities or proper nouns recognition.

Another interesting thing extracted from the evaluation is that although there is no big differences between the monolingual and the bilingual results, it is clear that automatic translation such as the used one in these experiments incorporates some errors and noise that decrease the system's performance.

Furthermore, basic techniques of lexical disambiguation and restricted-domain ontologies could improve the use of WordNet and give it new use dimensions.

Summarizing, although the results are not very good, the system itself presents a lot of improvement possibilities through the refinement of the scoring system, the addition of new techniques based on entity-recognition, the use of better translators and dictionaries and the incorporation of new semantic and ontological information that enforces WordNet access.

5 Conclusions

In this paper we have described the system carried out by the University of Alicante in the ImageCLEF 2004 adhoc retrieval task. A deep detailed information about the process itself and the strategies and experiments developed for the retrieval task has been given.

The results of the evaluation has been justified and different solutions to improve these results have been outlined in order to define near future steps for getting a better system.

References

1. www: Image CLEF in the Cross Language Evaluation Forum 2004. http://ir.shef.ac.uk/imageclef2004/index.html (2004) Last visited 2-Aug-2004.
2. Clough, P., Sanderson, M.: The CLEF 2003 cross language image retrieval task. In: Working Notes for the CLEF 2003 WorkShop, Trondheim, Norway (2003) 379–388
3. www: St Andrews University Library photographic collection. http://specialcollections.st-and.ac.uk/photcol.htm (2004) Last visited 2-Aug-2004.

4. Schmid, H.: Probabilistic Part-of-Speech Tagging Using Decision Trees. In: International Conference on New Methods in Language Processing, Manchester, UK (1994) 44–49
5. www: BabelFish translator. http://world.altavista.com/ (2004) Last visited 2-Aug-2004.
6. Miller, G.A., Beckwith, R., Fellbaum, C., Gross, D., Miller, K.J.: Five Papers on WordNet. Special Issue of the International Journal of Lexicography **3** (1993) 235–312