

SINAI at ImageCLEF Wikipedia Retrieval task 2011: testing combined systems

Miguel Ángel García-Cumbreras¹, Manuel Carlos Díaz-Galiano¹, L. Alfonso Ureña-López¹, Javier Arias-Buendía¹

¹ University of Jaén, SINAI Group,
Campus las Lagunillas, A3, 23071 Jaén, Spain
{magc, mcdiaz, laurena, jarias}@ujaen.es

Abstract. Several researches demonstrate that the use and integration of various knowledge sources improves the quality and efficiency of information systems. This paper presents the system developed by the SINAI research group at the ImageCLEF Wikipedia Retrieval task. Using only the English text associated with each image or its translation from French or German, our system applies several combinations of the textual tags and the combination of information retrieval (IR) systems. The influence of machine translation, the use of the different annotation tags and the retrieval results from different IRs systems are the main aims of this work. The obtained results show that the election of the annotation tags and the IR system have influence in the results, and some fusions of the retrieved lists can improve the performance.

Keywords: Text Retrieval, Machine Translation, Indexing, Retrieved Lists Combination, IRs Fusion

1 Introduction

The Wikipedia Retrieval task is an ad-hoc image retrieval task. The aim of ImageCLEF Wikipedia Retrieval task is to investigate retrieval approaches with a large and heterogeneous collection of images and their annotations, extracted from Wikipedia¹ [1].

In 2011 the collection is the same used in the 2010 ImageCLEF Wikipedia task [2], and it is composed by 237,434 Wikipedia images with unstructured and noisy annotations in English, French and German.

Given an English query as an image, sample images and annotations, our text-based system uses the annotations to find relevant images from the Wikipedia image collection. Several combinations of the results obtained with the annotation labels were tested and the fusion of the results obtained with two IRs systems.

The remainder of the paper is organized as follows: in the second section the system is described. In Section 3, the experiments and results are presented. Finally, conclusions and further work are shown in Section 4.

¹ Available at <http://www.wikipedia.org/>

2 Description of the system

An image from the Wikipedia collection contains some data, such as the name or the file of this image and annotations data. These annotations contain three optional tags (description, comment and caption) in different languages (English, French and German) also optional. Fig. 1 shows an example of an image with metadata from the collection.

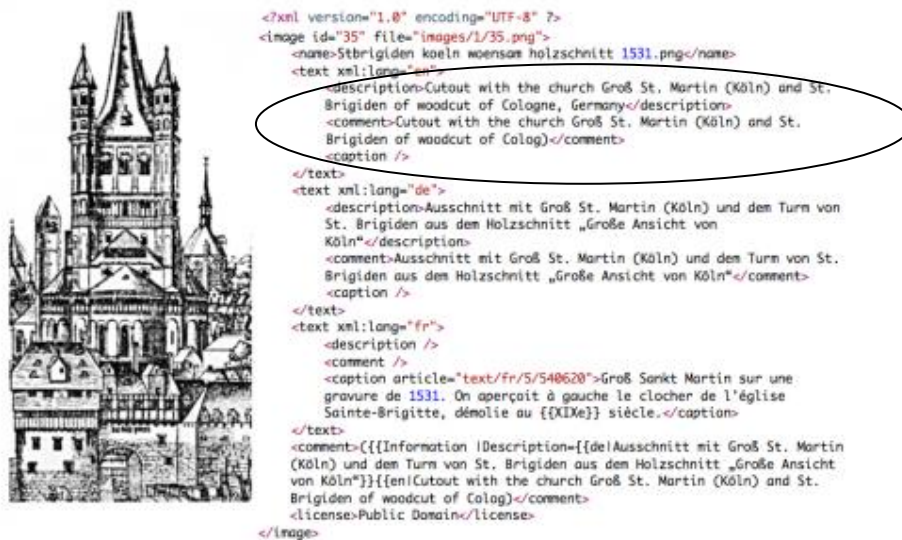


Fig. 1. Example of image from the Wikipedia collection and its metadata.

The following Fig. 2 shows the general architecture of our system. The collection modules (on the left of the figure) work with the collection of annotations. The first module translates non English data into English, as it has been explained previously. The second module processes the English data, applying stopwords removal and stemming. Caption and description tags are extracted and different combinations are created. Then, the IR systems Lucene and Lemur create different indexes with the combinations of tags.

The queries module (on the right of the figure) works with the queries. The first module preprocesses the query (stopwords removal and stemming), and then the information is run against the indexes, obtaining different lists of retrieved documents. In some cases a fusion module is applied to join the lists from the IR systems.

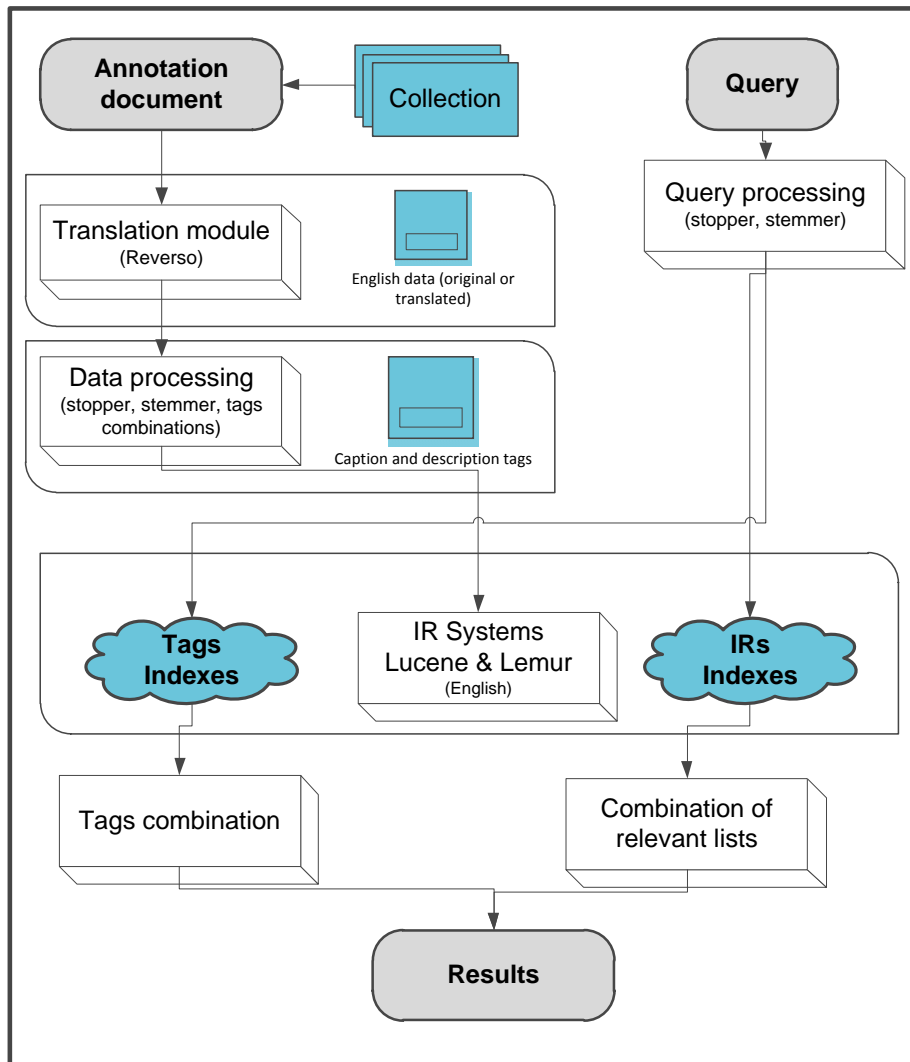


Fig. 2. Architecture of the SINAI system.

Our text-based system uses English as main language and according with the language distribution of the collection, the 237,434 Wikipedia images cannot be used. Some of the do not have annotations or the language is not determined. The following images were used in our experiments:

- English only: 70,127
- English and German: 26,880
- English and French: 20,747

- English, German and French: 22,899
- Translations from
 - o German only: 50,291
 - o French only: 28,461
 - o German and French: 9,646

A total number of 140,653 English annotations were used and 88,398 non-English annotations were translated.

The first module makes the automatic translation for the non-English metadata. For this purpose we used the online machine translator Reverso². Captions and descriptions were translated, and comments were discarded because in previous experiments we concluded that they were not useful. The second module makes a typical text preprocessing (text from caption and description tags). English stopwords removal and the Porter's stemmer [3] were applied.

Then, the third module uses two Information Retrieval (IR) systems, Lemur³ and Lucene⁴. Lucene is a high performance, full-featured text search engine that has shown to be robust in several text retrieval tasks. The Lemur Toolkit is an open-source toolkit designed to facilitate research in language modelling and information retrieval. It supports different automatic indexing strategies and a variety of retrieval models. In both cases the default parameters were used, the Okapi weighting schema for Lemur and automatic feedback was applied.

2.1 Combining the lists

We have implemented two simple methods to combine the lists of relevant documents retrieved by the IR or IRs systems.

The first method is the combination of the lists for each annotation tag. The baseline experiments were run using only the description and caption tags. The results of these experiments were two lists of relevant documents for each IR system used. In a second step we used a simple technique to merge these lists [5,6]. The method is simple; first, we have normalized the relevant values of each list; then, we applied different weighting percentages to each partial list and finally we have merged them into one final list re-ranking the new relevance obtained.

The different combinations are showed in the following list:

- The use of the description tag only
- 10% of caption and 90% of description
- 20% of caption and 80% of description
-
- 80% of caption and 20% of description
- 90% of caption and 10% of description
- The use of the caption tag only

² Available at <http://www.reverso.net/>

³ Available at <http://www.lemurproject.org/>

⁴ Available at <http://lucene.apache.org/>

All these combinations were indexed with Lucene and Lemur separately and then the queries were run against each index. We made previous experiments with the ImageCLEF Wikipedia Retrieval 2010 framework (non-official results) and the best results were obtained with the combination 80% of caption and 20% of description. The next section shows these experiments and results.

The second method fusion the lists of relevant documents obtained the IRs systems used. Lucene and Lemur results were combined for each query, with different weights for each list, as it was explained in the previous paragraph. Based on the experiments made with the 2010 framework, the results show that Lemur works better than Lucene. The best combined result was obtained with the fusion 90% of Lemur and a 10% of Lucene, although the improvement over the Lemur global result was not significant. The next section shows also these experiments and results.

3 Experiment description and results

This section describes some previous experiments made in the ImageCLEF Wikipedia framework that were the base of the system developed. The previous results are not shown completely due to the amount of them.

As it was described previously, the dataset used contains 140,653 English annotations and 88,398 non-English annotations translated (229,051 in total). The complete collection contains 237,434 Wikipedia images, so we used a 96% of the total collection.

3.1 Previous experiments

Some previous experiments were run with the 2010 framework (the same collection and 70 different queries). These experiments were evaluated using the relevance judgments delivered. The main aims were to test the best combination of tags (caption, description and comments) and the improvement of results with a simple fusion of lists of relevant documents delivered by the IR systems Lemur and Lucene.

The following experiments were made (with Reverso as automatic translator):

- Lucene IR system and different combinations of caption and description tags:
 - o caption tag
 - o description tag
 - o caption + description tags
 - o 10% of caption + 90% of description
 - o
 - o 90% of caption + 10% of description
- Lemur IR system with caption and different weighting functions (Okapi with feedback, kl divergence and kl divergence with feedback).
- Lemur IR system with Okapi and feedback and the same combinations of caption and description tags explained before.

- Lemur IR system with Okapi and feedback and Lucene with different weights for each list of relevant documents (from 10% Lemur and 90% Lucene to 90% Lemur and 10% Lucene).

The following Table 1 shows some relevant results obtained with the 2010 framework:

- Exp1: Lucene without feedback. Caption label.
- Exp2: Lucene without feedback. Description label.
- Exp3: Lucene without feedback. Caption and description labels.
- Exp4: Lucene without feedback. 80% Caption and 20% description labels.
- Exp5: Lemur with Okapi and feedback. Caption label.
- Exp6: Lemur with Okapi and feedback. Description label.
- Exp7: Lemur with Okapi and feedback. Caption and description labels.
- Exp8: Lemur with Okapi and feedback. 80% Caption and 20% description labels.
- Exp9-Best combination: 90% Lemur with Okapi and feedback and 10% Lucene without feedback. Caption and description labels.

All the results are evaluated using the Mean Average Precision (MAP) function. Last column shows the improvement or decrement in percentage of each MAP result over the baseline case with Lucene and Lemur.

Table 1. Some results obtained with the 2010 framework.

Experiment	MAP	% improvement
Exp1	0.1488	100%
Exp2	0.0728	49%
Exp3	0.1422	96%
Exp4	0.1628	109%
Exp5	0.1665	100%
Exp6	0.0698	42%
Exp7	0.1991	120%
Exp8	0.1991	120%
Exp9 BC	0.2003	101%

The best textual run in 2010 was submitted by XRCE with a MAP of 0.2361 [4].

After the analysis of these results we extracted the following conclusions:

- Lemur with feedback works better than Lucene.
- Caption tag works well. Description tag decreases the results. The joined use of caption and description tags improves the baseline results and also a percentage combination of both tags.
- The combination of both IR systems with caption and description tags improves a bit the textual result.

We analyzed some features of the queries such as the number of words of the query and the relevant documents obtained with each IR system, and some features of the retrieved list of relevant documents such as the RSV (Retrieval Status Value).

After this analysis we concluded that these features cannot determine the use of one or another IR system or the use of each tag or its combination.

3.2 Results with the 2011 framework.

Based on the previous results shown we run six experiments with the 2011 framework to test some features of our system. These six official runs are the following:

- Exp1: Lucene without feedback. Caption label.
- Exp2: Lemur with Okapi and feedback. Caption label.
- Exp3: Lucene without feedback. Caption and description labels.
- Exp4: Lucene without feedback. 80% Caption and 20% description labels.
- Exp5: Lemur with Okapi and feedback. Caption and description labels.
- Exp6: 90% Lemur with Okapi and feedback and 10% Lucene without feedback. Caption and description labels.

Table 2 shows the official runs and results for Image Wikipedia retrieval 2011.

Table 2. Official runs and results for Image Wikipedia retrieval 2011.

Experiment	MAP
Exp1	0.1496
Exp2	0.1732
Exp3	0.1200
Exp4	0.1618
Exp5	0.2052
Exp6	0.2068

The best result was obtained with the combination of IR systems, using caption and description tags but the improvement over the use of Lemur with caption and description tags is not significant. Some conclusions of these official results are the following:

- Lemur improves Lucene in all the experiments.
- Caption tag introduces almost the complete relevant documents, and description tag does not introduce some relevant documents.
- The combination of tags improves the results.
- The fusion of lists of relevant documents retrieved by the IR systems improves only a bit the MAP result.

4 Conclusions and further work

In this paper we present the experiments and results obtained with our system developed for the ImageCLEF Wikipedia retrieval environment. We tested some techniques with the 2010 framework, applying them to the official runs in 2011.

In this participation we have tested some textual features such as different IR systems, the importance of the annotation labels and some simple fusion methods to combine these labels and the lists of relevant documents retrieved.

Our results show that the IR systems do not obtain the same results, some labels are more important than others and the combination of them improves the final results.

As further work we want to analyze in depth the results obtained, queries grouped by topics, by length, etc. We will also test the influence of the translation with different automatic translators. An important future aim will be to train a machine learning system with relevant features that decide what to apply to improve the results (combination of IR systems, external knowledge or the annotations from the different tags).

Acknowledgments. This work has been supported by the Regional Government of Andalucía (Spain) under excellence project GeOasis (P08-41999), the Spanish Government under project TEXT-COOL 2.0 (TIN2009-13391-C04-02) and the local project RFC/UJA2009/12/14.

References

1. Theodora Tsirikla, Adrian Popescu, Jana Kludas: Overview of the wikipedia image retrieval task at ImageCLEF 2011. CLEF 2011 working notes, Amsterdam, The Netherlands, (2011)
2. A. Popescu, T. Tsirikla and J. Kludas Overview of the Wikipedia Retrieval Task at ImageCLEF 2010. In CLEF (Notebook Papers/LABs/Workshops) (2010)
3. M. F. Porter: An algorithm for suffix stripping. In Readings in information retrieval. ISBN 1-55860-454-5; pages 313-316. Morgan Kaufmann Publishers Inc., (1997)
4. Stéphane Clinchant, Gabriela Csurka, Julien Ah-Pine, Guillaume Jacquet, Florent Perronnin, Jorge Sánchez, and Keyvan Minoukadeh: XRCE's Participation in Wikipedia Retrieval, Medical Image Modality Classification and Ad-hoc Retrieval Tasks of ImageCLEF 2010. In CLEF (Notebook Papers/LABs/Workshops) (2010)
5. Manuel Carlos Díaz-Galiano, Miguel Angel García-Cumbreras, María Teresa Martín-Valdivia, and Arturo Montejo-Ráez. Knowledge integration using textual information for improving ImageCLEF collections. In ImageCLEF book. ISBN 978-3-642-15181-1; pages 295-313; DOI: 10.1007/978-3-642-15181-1_16. Springer Berlin Heidelberg, (2010)
6. M.C. Díaz-Galiano, M.T. Martín-Valdivia, A. Montejo-Ráez, and L.A. Ureña-Lopez. Improving Performance of Medical Images Retrieval by Combining Textual and Visual Information. In Artificial Intelligence - Special Session, 2007; MICAI 2007; pages 185-192; DOI: 10.1109/MICAI.2007.12. (2007)