Evaluating some contextual factors for image retrieval *ReDCAD participation at ImageCLEFWikipedia 2011*

Hatem Awadi, Mouna Torjmen Khemakhem, and Maher Ben Jemaa

Research unit on Development and Control of Distributed Applications (ReDCAD),
Department of Computer Science and Applied Mathematics,
National School of Engineers of Sfax, University of Sfax
{awadi.hatem,torjmen.mouna,maher.benjemaa}@gmail.com
http://www.redcad.org

Abstract. Our participation in the ImageCLEF Wikipedia retrieval task aims to study the efficiency of using two contextual factors in image retrieval: metadata which contains specific information about images, and textual content which contains general information about images. For this aim, the Lucene library is used for indexing and searching. We propose also to combine both factors using two different methods: one based on simple linear function and one based on scores comparison. In addition, a comparison between monolingual and multilingual image retrieval using queries in a single language (English) and queries in different language is done.
Results show that the use of textual content is more useful then the use of metadata and the combination of both factors further improves results. In addition, the use of all provided languages exceeds over the use of only English langage.

Keywords: context-based image retrieval, textual content, metadata

1 Introduction

This is our first participation in imageClef Wikipedia retrieval task. We aim in this work to evaluate the impact of some contextual factors in image retrieval using the Wikipedia collection. More precisely, we used the Lucene Search engine[1] to calculate an image relevance score for each contextual factor: textual content and metadata. The difference between these two factors is that the metadata contains the description of the image, so the most specific information. On the contrary, the textual content can be the same for two or more images, so it contains general information about the image.

[1] http://lucene.apache.org/

Consequently, comparing the use of specific information and general information to represent images returns to compare the use of metadata and textual content to compute the image score.

To well evaluate these two factors, a combination between them should be done. We propose so to use either a classical linear combination, or a comparison-based approach.

In addition, we would like to study the impact of the monolingual and multilingual image retrieval by using queries in a single language (English) and queries in different language. When using an English query, only the English documents are used, and when using a query in different languages (English, French and Dutch), the three translations of the query will be used in a single query, and the whole collection (English and/or French and/or Dutch documents) will be used.

The remainder of this paper is organized as follows. Section 2 describes our retrieval model. Section 3 details the runs and discuss the results. Finally, a conclusion and future work are done in section 4.

2 Retrieval model

2.1 Textual model

To calculate a relevance score for images using the metadata or the textual content, we used the Lucene library for indexing and searching.

Let q be a query, t a term of q , d_j a given document and $im_{i,j}$ is an image belonging to the document d_j. The relevance score of an image equals the relevance score of the document containing the image. It is calculated according to the following formula provided by the Lucene search engine [1]:

$$S(im_{i,j}, q) = coord(q, d_j) * queryNorm(q) * \sum (tf(t \in d_j) * idf(t)^2 * boost(t.field \in d_j) * lengthNorm(t.field \in d_j)) \quad (1)$$

where :

- $coord(q, d)$: Coordination factor, based on the number of query terms the document contains. The coordination factor gives a boost to documents that contain more of the search terms than other documents.
- $queryNorm(q)$: Normalization value for a query, given the sum of the squared weights of each of the query terms.
- $tf(t \in d_j)$: Term frequency factor for the term t in the document d_i: how many times the term t occurs in the document.
- $idf(t)$: Inverse document frequency of the term: a measure of how "unique" the term is. Very common terms have a low idf; very rare terms have a high idf.
- $boost(t.field \in d)$: Field and document boost, as set during indexing. It can be used to statically boost certain fields and certain documents over others.

– $lengthNorm(t.field \in d)$: Normalization value of a field, given the number of terms within the field. This value is computed during indexing and stored in the index norms. Shorter fields (fewer tokens) get a bigger boost from this factor.

2.2 Combination model

In order to combine both scores of textual content and metadata, we propose to use either a simple linear combination or a comparison-based combination.

Linear combination

To calculate a linear combination score of each image $S_{LC}(im)$, we used the following equation:

$$S_{LC}(im) = \alpha \times S_{txt}(im) + (1 - \alpha) \times S_{md}(im) \quad (2)$$

where $S_{txt}(im)$ is the image score according to the textual content and $S_{md}(im)$ is image score according to the metadata.

Comparison-based combination

Since both scores of metadata (specific information) and textual content (generic information) are computed by the same equation and the same way, we propose here to use for each image only the best factor, which gives the best representation for the image. To achieve this idea, we propose to use the maximum score of both factors after normalization, as follows:

$$S_{CC}(im) = Max(S_{txt}(im), S_{md}(im)) \quad (3)$$

Where $S_{CC}(im)$ is the final score of the image after combination.

3 Runs and Results

3.1 Textual model results

In this section, we present the results obtained by our approach when using the metadata and textual content. Table 1 presents our official runs.

Table 1. Impact of textual content and metadata

Runs	MAP	P10	P20	Rprec	bpref
redcad01tx	0.1335	0.3160	0.2760	0.2181	0.1735
redcad02tx	**0.2306**	**0.3700**	**0.3060**	**0.2862**	**0.2326**
redcad01md	0.1578	0.3140	0.2420	0.2052	0.1730
redcad02md	0.1896	0.3400	0.2920	0.2386	0.2016

- **redcad01tx**: This run used only the English version of textual content and queries. In fact, a score was calculated for each document using the formula 1, and then, this score was attributed to all images in this document.
- **redcad02tx**: This run used all provided textual content and all query languages. More precisely, we have concatenated the queries provided in three language in a single query, and then performed the search in the textual content in different languages.
- **redcad01md**: This run used only the English version of the metadata and the queries. All extracted fields are used with the same importance i.e. the boost factor in formula 1 was fixed to 1.
- **redcad02md**: This run used the metadata with all queries language as in the case of *redcad02tx* approach. All fields are used with the same importance.

According to the different metrics, results show that the use of a query composed of three languages (redcad02tx and redcad02md) is more efficient than the use of a single language query (redcad01tx and redcad01md). This result is expected given that the images are described or belonging on documents in different languages. In fact, it is possible that some relevant images are described in a language and the query is in another language, then the image cannot be retrieved.

By comparing the use of textual content and the metadata, we note that results obtained by using textual content outperform results obtained by using the metadata. A possible reason is that some relevant images have a short or no associated metadata.

3.2 Combination model results

In our additional runs, we combined the best results obtained by using the textual content and the metadata. In these experiments, the whole collection and only one query composed of the three queries were used.

Table 2 presents obtained results. Our official runs are in grayed boxes.

According to the different metrics, the combination of both scores of textual content and metadata improves the retrieval accuracy. More precisely, best results are obtained with $\alpha \in [0.4..0.6]$. Consequently, we can conclude that both factors are complementary and support the image retrieval.

Concerning the comparison-based combination, table 2 presents obtained results. We note here that using a comparison-based combination improves slightly results of a classical linear combination.

3.3 Discussion

Thanks to previous experiments, we can conclude that using generic information (contextual content) is more significant than using specific information (metadata) about images. However, combining both factors improves results, so they are both important to determine image relevance. In fact, both factors are complementary.

Table 2. Classical linear combination of textual content and metadata

Runs	α	**MAP**	**P10**	**P20**	**Rprec**	**bpref**
redcad02md	0	0.1896	0.3400	0.2920	0.2386	0.2016
redcadComb1	0.1	0.2092	0.3660	0.2960	0.2609	0.2177
redcadComb2	0.2	0.2244	0.3880	0.3140	0.2823	0.2379
redcadComb3	0.3	0.2355	0.4220	0.3210	0.2953	0.2520
redcadComb4	0.4	0.2412	**0.4380**	0.3280	0.2965	0.2571
redcadComb5	0.5	**0.2468**	0.4200	0.3300	**0.2978**	**0.2584**
redcadComb6	0.6	0.2467	0.4160	**0.3380**	0.2940	0.2579
redcadComb7	0.7	0.2416	0.4180	0.3340	0.2888	0.2531
redcadComb8	0.8	0.2355	0.4100	0.3260	0.2951	0.2511
redcadComb9	0.9	0.2318	0.4060	0.3240	0.2929	0.2498
redcad02txt	1	0.2306	0.3700	0.3060	0.2862	0.2326

Table 3. Comparison-based combination of textual context and metadata

Run	**MAP**	**P10**	**P20**	**Rprec**	**bpref**
redcadCBC	0.2537	0.3920	0.3310	0.3008	0.2494

To compare fairly our runs to official ones in the Wikipedia retrieval, we take into account only official runs using textual information without relevance feedback or query expansion. Table 4 shows official results ranked by MAP:

Table 4. Results of approaches using only simple textual information

Rank	**Participant**	**MAP**
1	UNED	0.3044
2	DEMIR	0.2432
3	ReDCAD	0.2306
4	SZTAKI	0.2167
5	DBISForMaT	0.2096
6	SINAI	0.2068

According to this table, we have obtained third rank, but if we include our best run which is obtained by comparison-based combination (MAP=0.2537), our rank becomes the second.

4 Conclusion and future work

In this paper, we addressed the context-based image retrieval approach. We studied and evaluated the impact of using two contextual factors: textual content and metadata in image retrieval. Results in the Wikipedia Clef 2011 collection

showed that the textual content which contains general information about images is more significant then the metadata which contains specific information, but the combination of both factors is encouraged and consequently the two factors are complementary.

In addition, comparing the use of monolingual and multilingual retrieval according to the query language, we can conclude that the use of a query composed of different language is more interesting than the use of queries in a single language.

In future work, we plan to extract and use semantics of image contextual factors in image retrieval and to add a content based image retrieval in our system.

References

1. Michael, M., Erik, H., Otis, G. : Lucene in Action 2nd Edition. Manning Publications Co, United States of America (2010)