Some experiments on the WikipediaMM 2008 task:
Evaluating the impact of image names in context-based retrieval

Mouna Torjmen, Karen Pinel-Sauvagnat, Mohand Boughanem

SIG-IRIT-Toulouse-France

{Mouna.Torjmen , Karen.Sauvagnat , Mohand.Boughanem }@irit.fr

Abstract

The goal of our participation in the WikipediaMM task of CLEF 2008 was to study the use of the name of images in a context-based retrieval approach. We evaluated this factor in three manners. The first one consists of using image names explicitly: we computed a similarity score between the query and the name of images using the vector space model. The second one consists of combining results obtained using the textual content of documents and results obtained using the first method. Finally, in our last approach, image names are used less explicitly. We proposed to use all the textual content of documents, but we increased the weight of terms in the image name. Results show that the "*image name*" is an intersting factor. Even if the image names are used as an additional source of evidence, they allow to better performance. Moreover, we conclude thet using the image name implicitly performs better results than using it explicitly.

Categories and Subject Descriptors

H.3 [**Information Storage and Retrieval**]: H.3.3 Information Search and Retrieval

General Terms

Measurement, Performance, Experimentation

Keywords

Contextual image retrieval, image name, textual content, implicit use, explicit use

1   Introduction

Image retrieval is divided into two main approaches: content-based and context-based image retrieval, each one has some advantages and disadvantages.

Our approach belongs to context-based image retrieval. In this paper, we study the impact of using the name of images in the retrieval process.

We evaluated three algorithms that take into account the name of images in retrieval. In a first approach, we calculate a score for each image using its name. This score in combined in a second approach with the score of the document based on the textual content. In a third approach, we evaluate the impact of the image name implicitly: in fact, we increase the weight of terms in the

image name. Consequently, documents having a relevant image name will be top ranked.

The rest of the paper is organized as follows: section 2 describes the goal of our participation in the WikipediaMM task of Clef 2008. Section 3 presents our approaches. In section 4, we detail our empirical evaluation of the proposed approaches and discuss our results. Finally, section 5 concludes with possible directions for future work.

2 Motivation

In previous work on the WikipediaMM collection, we evaluated the use of ancestors, brothers and child nodes to estimate the relevance scores of images [3]. We also evaluated the impact of using classification scores in [5].

This year, the aim of our participation in WikipediaClef 2008 is to study the impact of the image names in image retrieval. Many works [2, 1] and many web search engines as Google Images use the name of images as a contextual element of the image itself to participate in its relevance estimation. However, to our knowledge, there is no real and explicit study concerning this factor. Our intuition behind this study is that the name of an image, if it is significant, describes well the image content, and consequently, it plays an important role in estimating the image relevance. We only discuss here the case where the name of the image is significant, i.e is composed of real terms. For example, 'Barcelona.jpg' is a significant image name, whereas this is not the case for 'DSCN0075.jpg'.

As the element type where we can find image names can change from a collection to another, we identify the names of images in documents using their extension. Any token that ends by an image extension as ".jpeg" and ".gif" for example is considered as an image name (see figure 1). In the indexing phase, we thus index image names in a separate index.
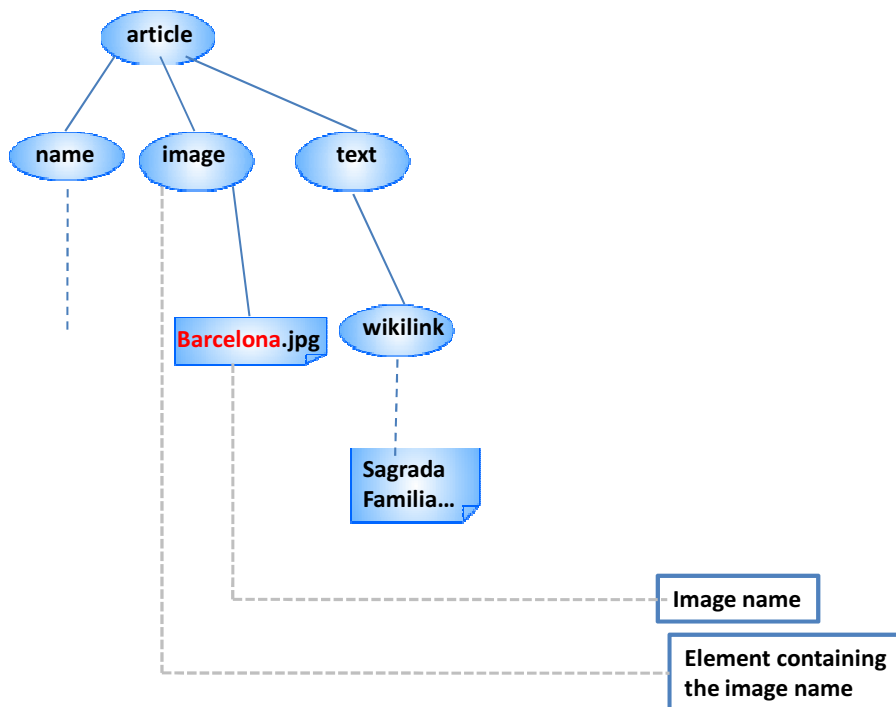

Figure 1: Example of an image name identification

3 Retrieval approaches

3.1 Baseline model: the XFIRM model

We use the XFIRM XML search engine [4] as a baseline model for retrieval. However, as the document structure in the collection is not complex (the average document depth is low), we use a simplified version of the model.

The model is based on a relevance propagation method. During query processing, relevance scores are computed at leaf nodes level and then at inner nodes level thanks to a propagation of leaf nodes scores through the document tree. An ordered list of subtrees is then returned to the user.

Let $q = t_1, \ldots, t_n$ be a query composed of n terms. Relevance values of leaf nodes are computed thanks to a similarity function $RSV(q, ln)$.

$$RSV(q, ln) = \sum_{i=1}^{n} w_i^q * w_i^{ln}, \quad where \quad w_i^q = tf_i^q * idf_i \quad and \quad w_i^{ln} = tf_i^{ln} * idf_i \qquad (1)$$

Where w_i^q and w_i^{ln} are the weights of term i in query q and leaf node ln respectively. tf_i^q and tf_i^{ln} are the frequency of i in q and ln respectively, $idf_i = log(|D|/(|di| + 1)) + 1$, with $|D|$ the total number of documents in the collection, and $|di|$ the number of documents containing i.

Each node in the document tree is then assigned a relevance score which is function of the relevance scores of the leaf nodes it contains.

$$r_n = |L_n^r|. \sum_{ln_k \in L_n} \alpha^{dist(n, ln_k) - 1} \qquad (2)$$

$dist(n, ln_k)$ is the distance between node n and leaf node ln_k in the document tree, i.e. the number of arcs that are necessary to join n and ln_k, and $\alpha \in]0..1]$ allows to adapt the importance of the $dist$ parameter. $|L_n^r|$ is the number of leaf nodes being descendant of n and having a non-zero relevance value (according to equation 1).

As the aim of the task is to return the whole documents (more precisely the file names of documents containing relevant images), we take $\alpha = 1$ to ensure that best ranked elements will be *article* elements.

3.2 Algorithms used to evaluate the impact of image names

3.2.1 Using only the image name terms

To explicitly use the terms composing the image name and study their importance, we first propose to only use the image name terms to retrieve relevant images. We compute a score for each image using the vector space model. We evaluated three similarity measures.

Let us consider a vector space of N dimensions, with N the number of terms in the collection, a query q, and an image name $ImName_i$.

The first similarity measure used is the Cosine Similarity (Equation 3).

$$W_{ImName}(Im) = cos(q, ImName_i) = \frac{\sum_{j=1}^{N} w_j^q w_j^{ImName_i}}{\sqrt{\sum_{j=1}^{N} (w_j^q)^2} \sqrt{\sum_{j=1}^{N} (w_j^{ImName_i})^2}} \qquad (3)$$

The second one is the Dice Coefficient presented in Equation 4.

$$W_{ImName}(Im) = sim(q, ImName_i) = \frac{2 \sum_{j=1}^{N} w_j^q w_j^{ImName_i}}{\sum_{j=1}^{N} (w_j^q)^2 + \sum_{j=1}^{N} (w_j^{ImName_i})^2} \qquad (4)$$

The last one is the Inner Product measure, expressed in Equation 5.

$$W_{ImName}(Im) = sim(q, ImName_i) = \sum_{j=1}^{N} w_j^q w_j^{ImName_i} \qquad (5)$$

In the three formula, $w_q^j \in [0..1]$ is the weight of term t_j in the query, and $w_j^{ImName_i} \in [0..1]$ is the weight of t_j in the image name. These weights are evaluated using a tf*idf formula.

As elements to return must be documents and not images, final results are the documents containing the relevant images. We thus have:
$W_{ImName}(Doc) = W_{ImName}(Im)$
There is no modification in ranking, as each document contains only one image.

3.2.2 Combining image name scores to document scores

To evaluate the textual content of documents, we used the XFIRM system. As results returned by this system are XML elements and not documents, we rerank results according to documents: in fact, for each relevant element, we use its document in the final results. If two elements from the same document are returned, we rank the document according to the element having the best score.

The score of the document obtained thanks to the image name $W_{ImName}(doc)$ can be combined with the score obtained by the XFIRM system $W_{XFIRM}(doc)$. We used the combination function presented in equation 6.

$$W(doc) = \lambda W_{XFIRM}(doc) + (1 - \lambda)W_{ImName}(doc) \qquad (6)$$

where λ is a pivot parameter $\in [0..1]$, $W_{XFIRM}(doc)$ represents the document score obtained using the *XFIRM* system, and $W_{ImName}(doc)$ is the score of the same image obtained by the image name.

3.2.3 Implicit use of the image name terms

We propose in this section to modify the term weighting formula (equation 1) used in the XFIRM model, by increasing the score of terms belonging to the image name. The new formula is as follows:

$$RSV(q, ln) = \sum_{i=1}^{n} w_i^q * w_i^{ln},$$

$$where\ w_i^q = tf_i^q * idf_i \quad and\ w_i^{ln} = \begin{cases} K * tf_i^{ln} * idf_i & if\ t_i\ is\ an\ image\ name\ term \\ tf_i^{ln} * idf_i\ otherwise \end{cases} \qquad (7)$$

$K > 1$ is a constant used to increase the weight of image name terms.

In the example of figure 2, without the implicit specification of the image name, the scores of document 1 and document 2 are equal. When applying our algorithm, document 1 becomes more relevant than document 2. This seems more logic as the image of document 1 concerns the query more than the image of document 2.

4 Runs and Results

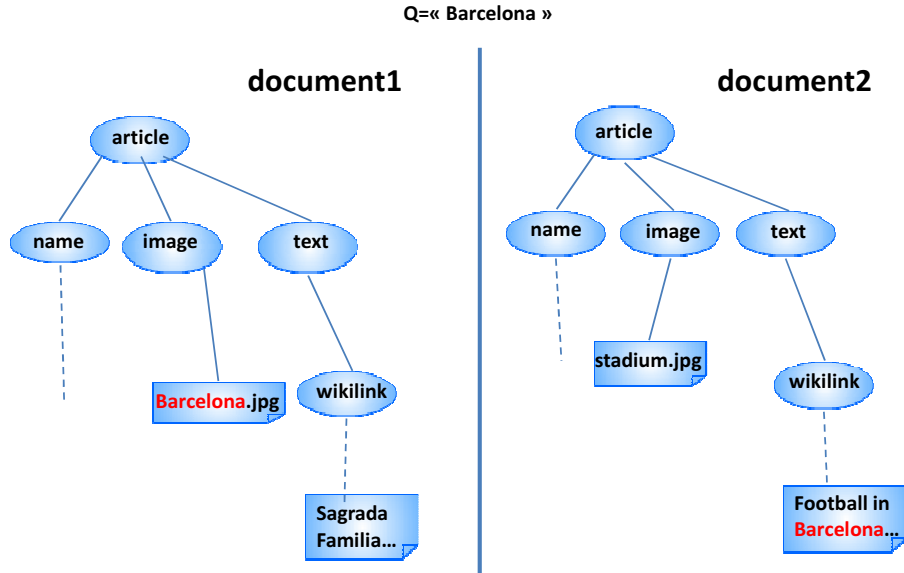In the following, our official runs are in grayed boxes.

Figure 2: Importance of image name terms

Table 1: Explicit use of image name

Runs	Similarity Formula	MAP	P@5	P@10	R-prec	BPREF
SigRunImageName	Cosine	0.0743	**0.1813**	**0.1573**	0.1146	0.0918
SigRunImageNameDice	Dice	0.0416	0.1200	0.1040	0.0745	0.0563
SigRunImageNameInner	Inner Product	**0.0757**	0.1760	0.1493	**0.1238**	**0.1078**

Results obtained using our first algorithm (equation 3, 4 and 5) are presented in table 1.

The Cosine (SigRunImageName) and Inner Product (SigRunImageNameInner) measures give approximately the same results, which are better than those using the Dice coefficient.

Table 2 shows results obtained when combining image scores evaluated using the cosine measure (equation 3) with whole document score (equation 6).

Table 2: ombination of image name results and XFIRM system results using Equation 6

Runs	λ	MAP	P@5	P@10	R-prec	BPREF
SigRunComb0	0	0.0743	0.1813	0.1573	0.1146	0.0918
SigRunComb01	0.1	0.1318	0.2560	0.2147	0.1782	0.1458
SigRunComb02	0.2	0.1380	0.2880	0.2320	0.1857	0.1537
SigRunComb03	0.3	0.1446	0.2987	0.2493	0.1919	0.1609
SigRunComb04	0.4	0.1472	0.3067	0.2600	0.1961	0.1645
SigRunComb05	0.5	0.1537	0.3227	0.2747	0.1979	0.1684
SigRunComb06	0.6	0.1572	**0.3253**	0.2720	0.2043	0.1755
SigRunComb07	0.7	0.1614	0.2327	0.2773	0.2123	0.1817
SigRunComb08	0.8	0.1610	0.3093	0.2773	0.2215	0.1825
SigRunComb09	0.9	**0.1681**	0.3093	0.2800	**0.2270**	**0.1876**
SigRunText	1	0.1652	0.3067	**0.2880**	0.2148	0.1773

Run *SigRunText* only uses the document textual content evaluated with the XFIRM system. The other runs, obtained by combining image score evaluated with the cosine measure and image score evaluated with all textual content of documents, shows that this combination allows to bet-

ter performance. the best run obtained is the one evaluated with $\lambda = 0.9$. This means that the image name can improve results as an additionnel source of evidence but not as a main one: the importance factor of textual content is 0.9 comperatively to 0.1 for the image name.

Runs obtained by implicit use of the image name (equation 7) are detailed in table 3.

Table 3: Implicit use of image name

Runs	K	MAP	P@5	P@10	R-prec	BPREF
RunImage11	1.1	**0.1724**	**0.3227**	**0.2867**	0.2329	**0.1874**
RunImage12	1.2	0.1701	0.3200	0.2787	**0.2271**	0.1845
RunImage13	1.3	0.1721	0.3093	0.2760	0.2267	0.1854
RunImage14	1.4	0.1714	0.3093	0.2720	0.2258	0.1855
RunImage15	1.5	0.1686	0.3040	0.2760	0.2247	0.1858
RunImage16	1.6	0.1681	0.3040	0.2720	0.2252	0.1859
RunImage17	1.7	0.1665	0.3093	0.2733	0.2238	0.1841
RunImage18	1.8	0.1667	0.3067	0.2707	0.2253	0.1841
RunImage19	1.9	0.1658	0.3093	0.2760	0.2261	0.1839
SigRunImage	2	0.1595	0.3147	0.2787	0.2211	0.1798
SigRunImage5	5	0.1481	0.2960	0.2547	0.2130	0.1716
SigRunImage10	10	0.1410	0.2933	0.2573	0.2011	0.1615

Comparing the best run obtained by implicit use of image name ($K = 1.1$, RunImage11) to the run obtained using document textual content (SigRunText), we notice that the MAP measure of the first one is more better.

As a main conclusion about the use of image names in contextual image retrieval, we notice that the image name is a relevant contextual element of image retrieval whatever the way it is introduced (implicitly or explicitly). Despite these good results, we think that results could be more improved if we resolve the following problem: in the used collection, many image names are not well indexed as they concatenate several terms. Let us consider the example of the image name "Alonewitheverybody.jpg", this name is considered as a single term while in reality it is composed of 4 terms "alone", "with", "every" and "body".
More work is thus needed to resolve the image name indexing problems and strongly confirm our conclusions.

Obtained results confirm our intuition that the image name is an important factor in the image retrieval when it is significant. In fact, combining the image name score and the document textual score improves results comperatively to ones obtained by these sources of evidence separately. Whereas, the more interesting result is that using the image name implicitly allows obtaining higher retrieval accuracy in term of MAP measure than using the image name explicitly.

5 Conclusion and future work

In this paper, we presented a study about the impact of the use of image name terms in contextual image retrieval. We evaluated the explicit and the implicit use of these terms. Results in Wikipedia Clef 2008 showed that this factor leads to performance improvement.
Results can be more improved as we have a problem of the indew process of image names (these latter are not well indexed as terms are concatenated).

In future work, we will try to find a solution of the image name indexing problem. We also plan to evaluate our methods on other collections having significant image names.

Our implemented system can only process queries which specify exact images, then our system cannot process concept clauses. Therefore, we should consider how to calculate retrieval status values using concept clauses.

References

[1] Yuksel Alp Aslandogan and Clement T. Yu. Diogenes: a web search agent for person images. In *MULTIMEDIA '00: Proceedings of the eighth ACM international conference on Multimedia*, pages 481–482, New York, NY, USA, 2000. ACM.

[2] Charles Frankel, Michael J Swain, and Vassilis Athitsos. Webseer: An image search engine for the world wide web. Technical Report TR-96-14, 31, 1996.

[3] Lobna Hlaoua, Mouna Torjmen, Karen Pinel-Sauvagnat, and Mohand Boughanem. XFIRM at INEX 2006. Ad-hoc, Relevance Feedback and MultiMedia tracks. In Norbert Fuhr, Mounia Lalmas, and Andrew Trotman, editors, *International Workshop of the Initiative for the Evaluation of XML Retrieval (INEX), Dagstuhl, Allemagne, 18/12/2006-20/12/2006*, volume LNCS 4518, pages 373–386. Springer, mars 2007.

[4] Karen Sauvagnat. *Modle flexible pour la recherche d'information dans des corpus de documents semi-structurs*. PhD thesis, Toulouse : Paul Sabatier University, 2005.

[5] Mouna Torjmen, Karen Pinel-Sauvagnat, and Mohand Boughanem. Mm-xfirm at inex multimedia track 2007. In *Pre-Proceedings of INEX 2007 Workshop, Dagstuhl, Germany*, 2007.