Building Contextual Implicit Links for Image Retrieval

Hatem Aouadi, Mouna Torjmen Khemakhem and Maher Ben Jemaa
National School of Engineering of Sfax, University of Sfax, Route de la Soukra km 4 - 3038 Sfax, Tunisia

Keywords: Context-based Image Retrieval, Implicit Links, Link Analysis, LDA.

Abstract: In context-based image retrieval, the textual information surrounding the image plays a main role in image retrieval. Although text-based approaches outperform content-based retrieval approaches, they can fail when query keywords are not matching the document content. Therefore, using only keywords in the retrieval process is not sufficient to have good results. To improve the retrieval accuracy, researchers proposed to enhance search accuracy by exploiting other contextual information such as hyperlinks that reflect a topical similarity between documents. However, hyperlinks are usually sparse and do not guarantee document content similarity (advertising and navigational hyperlinks). In addition, there are many missed links between similar documents (only few semantic links are created manually). In this paper, we propose to automatically create *implicit links* between images through computing the semantic similarity between the textual information surrounding those images. We studied the effectiveness of the links generated automatically in the image retrieval process. Results showed that combining different textual representations of the image is more suitable for linking similar images.

1 INTRODUCTION

The amount of images uploaded to the Web each day is growing exponentially. Consequently, the task of finding relevant images in response to a user request becomes increasingly difficult. As a solution, content based image retrieval (CBIR) techniques were proposed. However, despite the progress in this field, these techniques are still producing very poor results due to the lack of semantics in the visual features extracted from the image (Datta et al., 2008) (Zhou et al., 2017). Alternatively, researchers are oriented to the use of textual information surrounding the images (Datta et al., 2008) (Alzu'bi et al., 2015). Thus, most current image retrieval systems are text based. However, this technique also has its limitations. First, the information may not describe the real content of the image. In addition, even if the textual information is relevant to the image, it may not contain the query keywords. Hence, different approaches has been proposed to improve the effectiveness of text-based image retrieval systems. We are interested in this paper to the use of the linkage information for ranking search results.

In Web link analysis domain, the quality of a Web page is generally measured according to the quality of its neighbours. Thus, the environment of a Web page provides an important indicator for its relevance. However, hyperlink may not reflect the content similarity between the interconnected web pages since the existence of spam links (arbitrary created links, navigation links, advertising links , etc.). As a solution, we proposed to automatically create *implicit links* between multimedia documents based on the textual content similarity.

Another problem was encountered when processing the documents that have more than one image: the relevance scores are calculated at the document level and do not reflect the individual relevance of the images in this document. To overcome this problem, we proposed to segment pages into regions according to the image position. Implicit links are then created between the image regions instead of whole documents.

The originality of our work in this paper can be summarized by the automatic creation of implicit links between regions of images in order to (1) use only semantic links in the retrieval process, and (2) cover all semantic similarities between documents (hypertext links cover only some similar documents). To create implicit links, we compute the topical similarity between the extracted regions using LDA topic model. Our proposition are evaluated using the Wikipedia collection of ImageCLEF 2011.

This paper is organized as follows. Section 2 presents some related works. We detailed our approach

Aouadi, H., Torjmen Khemakhem, M. and Ben Jemaa, M.
Building Contextual Implicit Links for Image Retrieval.
DOI: 10.5220/0006808400810091
In *Proceedings of the 20th International Conference on Enterprise Information Systems (ICEIS 2018)*, pages 81-91
ISBN: 978-989-758-298-1

for implicit link construction in section 3. The experimental results are given in section 4. Finally, we conclude this work in section 5.

2 RELATED WORKS

The basic idea behind the use of links in information retrieval is motivated by the intuition that these links are not random and reflect a kind of resemblance between the pages. However, hyperlinks are not always indicators of content similarity. The authors of a web page can put arbitrary links to some pages that are not related to the subject of their page. Thus, hyperlinks can be created for navigation and structuring the site or for advertising purposes. These *spam* links decrease the quality of retrieval accuracy when they are used. Moreover, similar pages are not always linked to each other (generally the author of a web page creates only few links to other pages). In addition, many document collections have no links or have a very weak link structure. These problems present a major obstacle for all web search algorithms that use links in the retrieval process. As a solutions, many works proposed to automatically interlinking the documents by creating *implicit links* between them. We present in the following the use of implicit links in textual and multimedia retrieval.

2.1 Implicit Links in Textual Information Retrieval

Implicit links have been used in different areas including the ranking of search results, document classification and clustering.

Xue et al. (Xue et al., 2003) proposed to explore the user log to make implicit links between documents and then applied a modified PageRank algorithm for small web search. Despite the improved performance of the research, this method can not be applied to large collections because it is intended for a small search on the Web. Another work proposed by Kurland and Lee (Kurland and Lee, 2005)(Kurland and Lee, 2006) used the language models to generate document cluster relationships. The application of Hits (Kleinberg, 1998) and PageRank (Brin and Page, 1998) algorithms in the constructed graph of relationships improved search precision. Nevertheless, these methods had not been compared with the use of explicit links. Xu and Ma (Xu and Ma, 2006) proposed to construct an implicit graph and combined it with the hypertext graph to improve the search performance. Experiments showed the effectiveness of the proposed approach compared to the PageRank algorithm applied in the hyperlink graph. This approach is evaluated using a collection of forums that contains several noisy hyperlinks.

For document clustering, Zhang at al. (Zhang et al., 2008) defined an implicit link as co-authorship link. They also used explicit links composed of citation links and hyperlinks, and pseudo links such as content similarity links. The experimental results showed that linkage is quite effective in improving content-based document clustering.

In the document classification area, Shen et al. (Shen et al., 2006) compared the use of explicit links represented by the hyperlinks and implicit links generated from the query logs. In their study, they demonstrated that implicit links can improve the classification performance compared to the explicit links. Query logs are also used by Belmouhcine and Benkhalifa (Belmouhcine and Benkhalifa, 2015) to create implicit links between web pages for the purpose of web page classification. Experimental results with two subsets of the Open Directory Project (ODP) have shown that this representation based on implicit links provides better classification results.

In the biomedical domain, Lin (Lin, 2008) applies Hits and PageRank algorithms on implicit links (content similarity links) analysis in the context of the PubMed search engine. He demonstrated that it is possible to exploit networks of content similarity links, generated automatically, for document retrieval. Thus, the combination of scores generated by link analysis algorithms and the text-based retrieval baseline improved the precision.

2.2 Implicit Links in Multimedia Information Retrieval

In order to improve the accuracy of image retrieval, several research projects have proposed to build implicit links between images by mainly using visual content. Thanks to these visual links, a visual graph is constracted and then analyzed to calculate the relevance scores of the images. To analyse the constructed graph, random walk method (such as PageRank) has been widely adopted (Jing and Baluja, 2008a) (Jing and Baluja, 2008b) (Zhou et al., 2009) (Zhang et al., 2016) (Wang et al., 2016).

Xie et al. (Xie et al., 2014) proposed to construct an off-line visual graph by taking each image as a query and make a link with the k top returned images. HITS algorithm is then applied on the set of images returned at query time. In the same way, Liu et al. (Liu et al., 2017) followed the same offline step, and at query time, they merge the different graphs obtained using different descriptors. Then, they applied a local ranking algorithm on the resulted graph. Zhang et al. (Zhang et al., 2012) proposed a query-specific fusion approach based on graph , where multiple lists of search results from different visual cues were merged and clustered by link analysis on a merged graph. Wang et al. (Wang et al., 2012) have incorporated several visual features in a graph-based learning algorithm for images retrieval. The creation of implicit links using the textual information is taken into account for the first time by Khasanova et al. (Khasanova et al., 2016) who have built a multilayer graph where each layer represents a modality (textual, visual, etc.). The constructed graph is undirected, where each node is connected only with its k-nearest neighbours (in terms of Euclidean distance). Then, they applied a random walk on the multilayer graph by making transitions between the different layers. The proposed solution achieves good image retrieval performance compared to the state-of-the-art methods. The authors firmly believe that flexible structures like graphs offer promising solutions to capture the underlying geometry of multi-view data.

Other works have been done as part of MediaEval (Liu et al., 2007) (Hsu et al., 2007) (Eskevich et al., 2013) (Chen et al., 2014) (Bhatt et al., 2014) and TRECVID (Simon et al., 2015) evaluation campaigns for video hyperlinking. Chen et al. (Chen et al., 2014) concluded that textual features work better in this task, whereas visual features by themselves can not predict reliable hyperlinks. Nevertheless, they suggest that the use of visual features to re-rank the results of text-based retrieval can improve the performance.

In conclusion, the majority of works use the visual content of the images to create implicit links between them. However, the unresolved problems associated with this modality make this type of links ineffective and unprofitable. For this reason, we propose in this paper to use textual information to automatically create links between images and to explore them in a retrieval process in order to improve the retrieval accuracy.

3 LDA BASED IMPLICIT LINK CREATION

Links between similar documents are more efficient than links between independent pages. However, in the context of the Web, semantically similar pages are not always linked to each other, hence the need to automatically create implicit links between them.

We propose in this paper to automatically create links between similar images through the calculation of the semantic similarity of the textual information surrounding these images. The similarity can be calculated using the vector representation of the texts. However, the textual information in a multimedia document may contain some details that are not related to the image. Thus, an image usually represents an illustration for the overall subject of the document. For example, a page talking about the animal "*lion*" will contain probably images of *lion*. However, words such as "*forest*", "*meat*", "*water*", etc. will be present frequently. If we use the textual representation to calculate the similarity with other images, we can obtain images that are assumed to be similar but do not represent the image of a *lion*. Hence, word level document similarity can be easily spammed when the same words are used in documents with different topics. For this reason, we propose to model documents in a more generalized form.

Latent Semantic Analysis (LSA) (Deerwester et al., 1990) was initially proposed as a topic based method for modelling words semantic. Basically, LSA finds a small representation of documents and words by applying the truncated singular value decomposition (SVD) for the document-term matrix. An improvement of this model has been proposed with Probabilistic and Latent Semantic Analysis (PLSA) (T. Hofmann, 1999) which uses a probabilistic method instead of using matrices. Then, the PLSA model has been generalized to Latent Dirichlet Allocation (LDA) model (Blei et al., 2003). For this reason, we choose the LDA topic model to model documents. Indeed, this model allows to give an overview for the documents in the form of topic distributions, which allows to filter out secondary and noisy information.

In the following, we describe briefly the LDA algorithm, then we describe our method to segment a document to regions according to the image position. After that, a detailed description of our method to create implicit links is done. And finally, we present the application of link analysis algorithms on the created links.

3.1 LDA Technique

Blei et al. (Blei et al., 2003) have proposed LDA topic model that can reduce the representation of documents as a mixture of latent topics. The model generates automatic topical summaries in terms of discrete probability distributions on words for each topic, and infers further discrete distributions by document on topics.

LDA assumes that all documents are probabilistically generated from a shared set of K common topics, where each topic is a multinomial distribution over the vocabulary (noted by β). The generation of a document is done according to the following generative process :

(1) For each topic
    (a) draw a distribution over words $\varphi \sim Dir(\beta)$
(2) For each document
    (a) Chose $\theta_d \sim Dir(\alpha)$
    (b) For each word
        (i) generate topic $z \sim Mult(\theta)$
        (ii) generate term $w \sim Mult(\varphi)$.

We apply LDA to the textual information representing the images. The outputs of this model are then used to create implicit links between the images. More precisely, we use the topic distributions generated by LDA to compute the similarity between image representations and therefore create implicit links between them.

3.2 Textual Representations of Images

In text-based image retrieval, the basic idea consists in considering the document as an atomic unit and all its textual information is treated in a similar way. Therefore, for a given query, a relevance score is calculated for the whole document and then assigned for all its images. According to this process, all images in a document will have the same relevance score even if they have different relevance levels, or some of them are not sufficiently relevant to the query. This major weakness has led us to the idea of segmenting multimedia documents into image regions. In this way, it would be possible to differentiate images of the same document by approximating the relevance degree of each image separately.

The best textual description of the image is the associated metadata (called in our work IMD: Image Meta Data) because it is the most specific information for the image. However, metadata usually contains few terms or can be missed sometimes. For these reasons, we propose to consider other additional sources of information to describe the image. More precisely, we propose to divide the content of the document into two descriptions for each image. The first description is the container region of the image obtained after segmentation of the document. We call this description ”Specific Image Description” (SID). The second description of this image is the rest of the document (without SID) called ”Generic Image Description” (GID). If a document contains more than one image, the specific description of an image belongs to the generic description of other images and vice versa. Figure 1 shows the different descriptions of an image (*img1*).

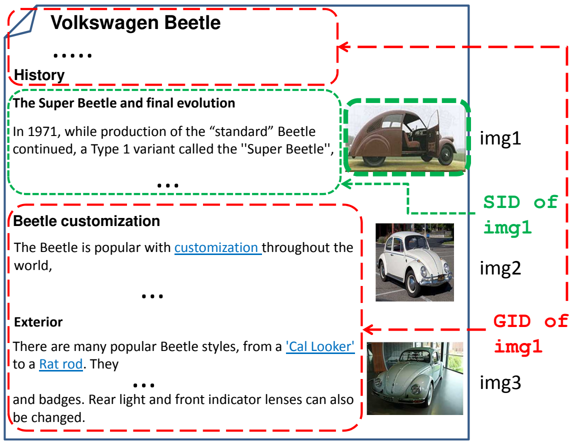

Figure 1: Example of specific (SID) and generic (GID) representations of an image.

In this example, the SID description of the image *img1* is the container paragraph and its GID description is all the textual content of the document except that paragraph. The segmentation of the document into paragraphs could be done easily for web documents thanks to the use of tags. In HTML documents for example, the use of title tags (<H1>, <H2>, etc.) makes it possible to segment the document into paragraphs. Wikipedia documents are also easy to be segmented thanks to their specific tags: the == tag for a first level paragraph (equivalent of the <H1> tag in HTML), the === tag for a second level paragraph, etc. We propose to define the specific description of the image as the smallest paragraph granularity containing that image.

To conclude, each image in the collection is represented using three descriptions: (1) image metadata (IMD); (2) the paragraph containing the image as specific description (SID) and (3) the document without the paragraph containing the image as generic description (GID).

3.3 Implicit Link Creation

In this section, we describe the proposed approach for creating contextual links between images based on the LDA topic model. Many steps are needed to create the links between each pair of images. The link creation process is applied separately for each image representation. Figure 2 presents an overview of the link creation process using generic image descriptions (GID).

We distinguish three steps which are detailed in the following.

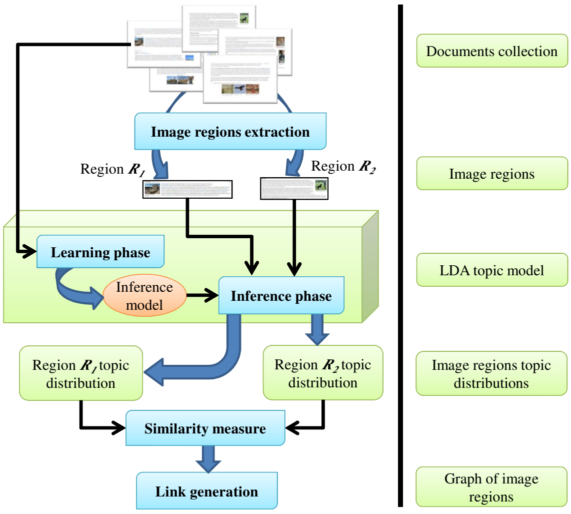

Figure 2: Overview of LDA based link creation between two images using GID.

3.3.1 Step 1: Topic Distribution Generation

For the SID (resp. IMD) representation, we construct a SID (resp. IMD) collection containing the SID (resp. IMD) of all images. Then, the LDA topic model is applied to the textual content of both images to create links using the whole SID (resp. IMD) collection. More specifically, a learning phase is performed in order to estimate the topic distributions for each SID (resp. IMD) of both images.

For the GID representation, an additional inference phase is performed after the learning phase to compute the topic distributions for each GID. In this type of representation, we kept the original collection and did not create a separately GID collection. In fact, for GID representations, we propose to use the whole documents in the learning phase instead of using only GID descriptions. This decision aims at avoiding the use of redundant information: two GID representations of two images belonging to the same document will contain a lot of redundancy. To better explain this problem, we consider the example in Figure 1. If we create a GID collection, the GID representation of the image *img1* is composed of parag. 1, parag. 3, and parag. 4, while the GID representation of the image *img2* is composed of parag. 1, parag. 2, and parag. 4. We note here that parag. 1 and parag. 4 will be used twice in the learning model and will therefore affect the quality of the topic distributions.

After applying the LDA model on each description collection, two probability distributions are generated: a document-topic distribution that represents the proportions of topics in each image representation; and a term-topic distribution that represents the weights of the terms in each topic. Figure 3 shows an example of document-topic distribution for image representations. The first table represents the topics (T_j) with their corresponding terms. The second table represents the topic distributions for each image representation(Rep_{img_i}).

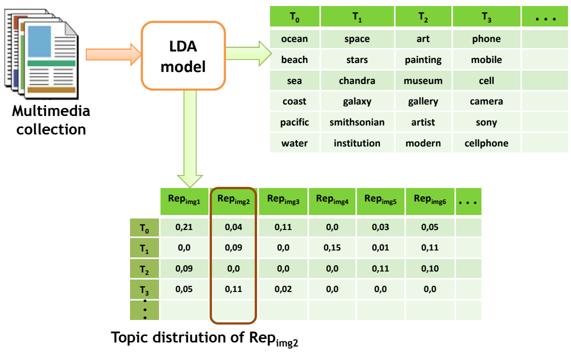

Figure 3: An example of document-topic distribution for image representations.

The best topics for each image representation are assigned the higher scores. For example, the best topic in Rep_{img_2} is T_3 (score = 0,11). To fix the number of topics that will be considered for each representation, a simple way is to fix a static number (for example, each representation will be assigned the top 10 topics). However, the total number of topics varies from one representation to another (we only consider topics with positive scores). Therefore, we propose to set a percentage of the top topics to be used for each image representation (for example, each representation will be presented by 10% of the most relevant topics). Best percentage values are obtained by experiments and detailed in table 1.

Once the topic distributions are calculated for the representations (SID, IMD, or GID) of two images, a link weight must be calculated using a similarity measure between the topic distributions of both images.

3.3.2 Step 2: Similarity Measure Between Two Images

After performing the LDA process, each image representation is defined by a topic distribution vector as shown in Figure 3. In order to create and weight the links between images, we propose to apply a similarity measure on the two by two vectors. In the information retrieval literature, the most commonly used similarity measure between two vectors is the cosine measure (Salton, 1989) (Li and Han, 2013) (Mikawa et al., 2011). We therefore propose to use the cosine measure in our work as follows:

$$cos_{sim}(\overrightarrow{Rep_{img_1}}, \overrightarrow{Rep_{img_2}}) = \frac{\overrightarrow{Rep_{img_1}}.\overrightarrow{Rep_{img_2}}}{||\overrightarrow{Rep_{img_1}}||||\overrightarrow{Rep_{img_2}}||} \quad (1)$$

where $\overrightarrow{Rep_{img}}$ denotes the topic distribution vector of the image representation Rep_{img} which could be GID, SID or IMD.

In our work, we propose to improve the classical cosine measure by including the number of common topics between the two image representations. The intuition of this proposition is that the more the two representations of images have common topics, the more similar they are. For example, if there are two image representations with only one common topic but with a high probability score, and two other image representations with many common topics but low probability scores, the cosine measure will favour the first two images. From another point of view, the number of common topics between two image representations could be very high with low topic distributions. In this case, the link weight will be very high, although there is no high semantic similarity between the two topic distributions. To overcome this situation, we propose to compute the number of common topics using only the most important topics for each image representation. The percentage of the most significant topics noted X% is fixed with the experiments described later. The new equation becomes:

$$sim(\overrightarrow{Rep_{img_1}}, \overrightarrow{Rep_{img_2}}) = cos_{sim}(\overrightarrow{Rep_{img_1}}, \overrightarrow{Rep_{img_2}}) \times |commonTopics(X\%\overrightarrow{Rep_{img_1}}, X\%\overrightarrow{Rep_{img_2}})| \quad (2)$$

With $\overrightarrow{Rep_{img_1}}$ (respectively $\overrightarrow{Rep_{img_2}}$) is the topic distribution of the representation Rep_{img_1} (resp. Rep_{img_2}) and $X\%\overrightarrow{Rep_{img_1}}$ (resp. $X\%\overrightarrow{Rep_{img_1}}$) is the X% of the most relevant topics according to their probability scores for Rep_{img_1} (resp. Rep_{img_2}).

Finally, after calculating the similarity scores between the images, a threshold is applied to reduce the number of implicit links between images. This similarity threshold is set to 0.1 by experiments.

3.3.3 Step 3: Link Direction Estimation

After constructing the implicit links between the image regions, we can use the link analysis algorithms of the literature to compute the relevance of the nodes in the constructed graph given a query. However, these algorithms that originally designed for web search assume that the links are directed, i.e. the link has a starting node and a one-way ending node. In our case the implicit links obtained by similarity calculation are bidirectional. Indeed, when we say that a node *A* is similar to a node *B*, the node *B* is also similar to *A* with the same degree of similarity. In this case, if we want to apply the HITS algorithm for example, the hub and authority scores for a given node will be the same because the number of incoming and outgoing links of these nodes will be the same. Thus, the HITS algorithm will not work properly.

To determine the direction of links, we rely on the following intuition: when two representations have some common topics, the region containing more information about these topics (high probability) is suitable to be the destination of the link. For this, we propose to calculate a direction score according to the percentage of information shared between the two representations. In other terms, we propose to determine how much the information of the representation of the image 1 (Rep_{img_1}) is presented in the representation of the image 2 (Rep_{img_2}). The following formula is used:

$$Score_{Direction}(Rep_{img1} \rightarrow Rep_{img2}) = \frac{\overrightarrow{Rep_{img_1}} \cdot \overrightarrow{Rep_{img_2}}}{\overrightarrow{Rep_{img_1}}} \quad (3)$$

with $Rep_{img1} \rightarrow Rep_{img2}$ means that the link is from Rep_{img_1} to Rep_{img_2}.

Note that we consider in Equation 3 only the common topics among the X% most important topics in the Rep_{img_1} and Rep_{img_2} representations and not the whole topic distributions, as explained in the previous subsection.

Based on the intuition that the link should start from the general image representation to the more specific image representation, the direction of the implicit link between two similar representations Rep_{img_1} and Rep_{img_2} can thus be defined as follows:

- If $Score_{Direction}(Rep_{img1} \rightarrow Rep_{img2}) = Score_{Direction}(Rep_{img2} \rightarrow Rep_{img1})$, both documents have almost the same amount of information about the shared content. In this case, two links are created: one from Rep_{img_1} to Rep_{img_2} and the other in the opposite direction;
- If $Score_{Direction}(Rep_{img1} \rightarrow Rep_{img2}) < Score_{Direction}(Rep_{img2} \rightarrow Rep_{img1})$, the link should be directed from Rep_{img_1} to Rep_{img_2}. In fact, the representation Rep_{img_1} contains more information about the shared content than Rep_{img_2}. This implies that Rep_{img_1} is more general, and the representation Rep_{img_2} describes a specific part of Rep_{img_1};
- If $Score_{Direction}(Rep_{img1} \rightarrow Rep_{img2}) > Score_{Direction}(Rep_{img2} \rightarrow Rep_{img1})$, the link

should be directed from Rep_{img_2} to Rep_{img_1}.

Figure 4 shows an example of determining the link direction.

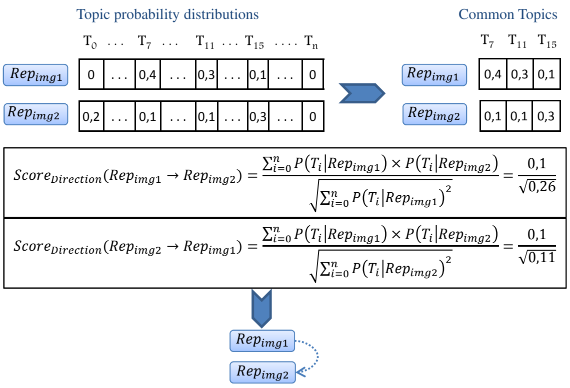

Figure 4: Example of computing the link direction.

These two image representations share three topics: T_7, T_{11} and T_{15}. By computing the direction scores between the two representations, we obtain $Score_{Direction}(Rep_{img2} \rightarrow Rep_{img1}) > Score_{Direction}(Rep_{img1} \rightarrow Rep_{img2})$. Consequently, the link is directed from img_1 to img_2.

3.4 Implicit Link Analysis

Once the implicit links are created between the images for each type of image representation (IMD, SID, GID), a link analysis algorithm such as HITS and PageRank or a social network analysis algorithm such as Betweeness, Closeness and Degree; could be applied. Concerning the link analysis algorithms, we propose to use the HITS algorithm because it is query dependent, i.e. it is done for each new query, which allows it to be always close to the subject of this query. For the social network analysis algorithms, we propose to use the most used ones in the literature namely the Degree and Betweenness (Brandes, 2001) centralities.

A score based on the links for each image representation is thus calculated. Therefore, each image will have three scores based on the links: the SID link score, the GID link score, and the IMD link score. Finally, to obtain a single score based on links, we propose to combine the three links based scores as follows:

$$LinkScore(img_i) = \alpha * LinkScore(IMD_{img_i}) + \beta * LinkScore(SID_{img_i}) + \gamma LinkScore(IMD_{GID_i}) \quad (4)$$

where α, β and γ are parameters used to adjust the importance of each representation in the computing of the final link score for an image. Their sum equals to 1.

4 EVALUATION

We develop our experiments to evaluate the performance of the proposed approach. First of all, we present the evaluation protocol and the parameter settings for our work. Then we provide comparative results of different link analysis algorithms.

4.1 Evaluation Protocol

We evaluate the effectiveness of implicit and explicit links in image retrieval domain. For this end, we use a Wikipedia collection provided by The ImageCLEF (The CLEF Cross Language Image Retrieval Track) 2011 for the Wikipedia retrieval task. The collection consists of 125 827 documents in three languages, containing 237 434 images. A set of 50 queries is also provided to perform the retrieval accuracy evaluation. We are interested in this paper only in documents written in English where the number is 42 774. However, our approach can be applied to any language and any type of document.

In order to evaluate properly our proposition, we construct a new base of assessments composed only of images belonging to the English documents. Due to computing complexity and time-space costs, we decide to run a textual search and then apply the link analysis for only the first 1000 returned results. To generate the initial textual results, we have used the Lucene[1] search library.

Early precision (P@10, P@20, ...) is important in a Web search context, since users in general examine relatively few results. Finally, the MAP measure is used to evaluate the global systems effectiveness.

4.2 Parameter Settings

4.2.1 LDA Parameters

The mallet library[2] is used to generate the LDA topical representation of documents. LDA parameters are fixed to the most common values used in the literature: $\alpha = 50/K$ where K is the number of topics, and $\beta = 0.01$.

To set the best value of K, we propose to apply and evaluate the LDA model in image retrieval instead of the link creation for time and memory reasons. Thus, the documents and the queries are presented by topics and a matching function is used to retrieve the relevant images. We carried out several experiments with different K values between 100 and 2000. Figure 5 shows the variation of MAP according to K.

[1] http://lucene.apache.org/
[2] http://mallet.cs.umass.edu/

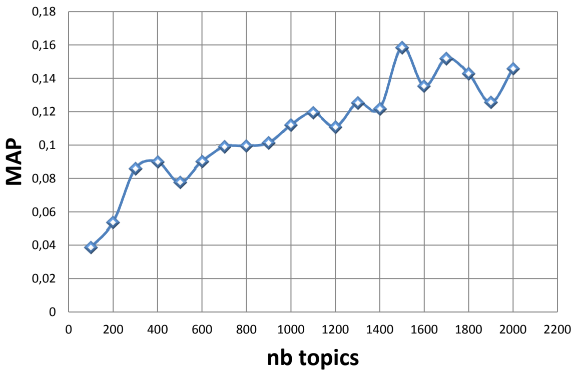

Figure 5: Determining the best number of topics. K.

We note that with small K values, we obtained bad MAP values. This means that the collection covers several topics and that it is not easy to classify documents into few topics. As shown in Figure 5, the best K value is 1500. Thus, we set $K = 1500$ topics in link building experiments.

4.2.2 Similarity Measure Evaluation

In this experiment, our purpose is to evaluate our proposed measure for similarity scores between images used to create links between them. More precisely, we aim to fix the best ratio of top topics according to the topic distribution of both images. Experiments of building and analysing implicit links between images are costly in terms of time and they need to fix several settings. For thus, we propose to evaluate the efficiency of the proposed measure in the image retrieval process. Indeed, we apply this similarity measure to compute relevance scores of images given a query using the Image CLEF Wikipedia collection 2011. In this experiment, the multimedia specificity is not taken into account: the relevance score is computed for the whole document, and then assigned for all its own images. In the retrieval process, the following equation is used:

$$RSV(D,Q) = cos_{sim}(Doc,Q) \times \\ |commonTopics(X\%\vec{D}, X\%\vec{Q})| \quad (5)$$

where $X\%$ is the X percent of the top topics representing the document or the query.

Table 1 presents some experiments of varying the X parameter of equation 5 in addition to the baseline run obtained by using only the original cosine measure (Cos run).

Comparing the cosine measure and our proposed measure according to the MAP metric, we note that better results are obtained when X is between 5% and 60% (MAP $>$ 0.1387). Moreover, the best MAP and P@5 values are obtained with X=10% and the best P@10 is obtained when X=15%. This means that the 10% of the top topic representing documents and queries are the most significant information. Thanks to our measure, the retrieval accuracy is improved by 26.31% according to MAP measure, 43.58% according to P@5 measure and 12.64% according to P@10 measure. These improvements prove that the use of the top common topics between the query and the document is a good relevance indicator.

Table 1: Determining the best percentage of the top topics.

	MAP	P@5	P@10
Cos	0,138	0,156	0,174
X=1%	0,12	0,144	0,158
X=5%	0,157	0,188	0,204
X=10%	**0,172**	**0,224**	0,196
X=15%	0,159	0,204	**0,204**
X=20%	0,152	0,18	0,182
X=30%	0,151	0,188	0,158
X=40%	0,147	0,212	0,192
X=50%	0,142	0,212	0,172
X=60%	0,139	0,208	0,186
X=70%	0,133	0,176	0,162
X=80%	0,135	0,168	0,174
X=90%	0,126	0,156	0,156
X=100%	0,127	0,14	0,152

4.3 Experimental Comparison of Different Link Analysis Algorithms

The aims of this experiment are twofold: (1) evaluation of the separate and the combined use of the different image representations (SID, GID and IMD); (2) comparison between three link analysis algorithms applied in our work.

The combination between the three image representations is based on a simple average of the three scores without taking into account the optimal settings. However, it is possible to run some experiments to set the optimal combination values. The combination equation is:

$$S_{lien}(img_i) = 1/3 * S_{lien}(IMD_{img_i}) + \\ 1/3 * S_{lien}(SID_{img_i}) + 1/3 * S_{lien}(GID_{img_i}) \quad (6)$$

Table 2 depicts overall results, where AverComb is the run obtained by averaging the scores of the three image representations.

By comparing the MAP values of the different image representations without combination, we note that the GID run gives the best results. This means

Table 2: A comparison of different link analysis algorithms according to different image representations.

	IMD	**SID**	**GID**	**AverComb**
	Degree Centrality			
P@5	0,088	0,092	0,128	**0,148**
P@10	0,104	0,072	0,108	**0,136**
MAP	0,058	0,064	**0,1**	0,094
	HITS			
P@5	0,104	0,052	0,08	**0,108**
P@10	0,096	0,052	**0,106**	0,104
MAP	0,056	0,059	0,074	**0,081**
	Betweeness Centrality			
P@5	0,056	0,088	0,08	**0,104**
P@10	0,048	0,066	0,08	**0,084**
MAP	0,028	0,046	0,061	**0,062**

that the generic information is the best source of evidence to represent images in this work. This interpretation could be explained by the specific/generic vocabulary notion: if the query vocabulary is generic, it is better to represent the image by generic information, and if the query vocabulary is specific, it is better to represent the image by specific information. We note that a query is called specific if the results represent the same object (for example, "London Bridge") and is generic if the results represent many objects (for example, "skyscraper building tall towers").

To validate this interpretation, we have computed the number of specific and generic queries in the ImageCLEF Wikipedia collection 2011 and we found that 72% of queries are generic and 28% are specific. Thus, it is not surprising that generic descriptions outperform specific descriptions.

Another interpretation could be drawn from the results: combining the three image representations by averaging their scores improves in general the results. This confirms our assumption that the use of the three sources of evidence is very useful. However, by analysing query by query, we have observed that GID representations give the best results for 40% of queries, SID and IMD gives the best results for 30% of queries for each one. So, it will be interesting to train a query classifier in order to choose the best representation to use for each query according to its class: generic or specific. We recall that the combination settings are not optimized. In our work, we decided to use a basic linear combination which is the score average in order to validate our approach without an optimal tuning of parameters. Further experiments are needed to set the best combination factors.

Finally, by comparing the use of the different link analysis algorithms, we note that the Degree Centrality gives the best performance. We could argue this result by the following reasons: Betweenness algorithm assumes that the graph should be undirected, which is not the case in our work. Moreover, the HITS algorithm idea is to select the K top ranked documents according to the query and then extend this initial set root by other documents. This basic idea is not respected in our work as we use the 1000 top documents ranked by a textual model without extension.

5 CONCLUSIONS

In this paper, we presented our contributions in context based image retrieval by building implicit links between images and exploring them in the image retrieval process. We first defined three types of textual representations for each image: (1) Specific Image Description (SID); (2) Generic Image Description (GID) and (3) the Image MetaData (IMD). Thereafter, we proposed a method to build LDA based links for each representation. Consequently, we obtained three types of links: SID links, GID links and IMD links. When comparing the different representations separately, the experiments chowed that GID based links gives the best results. This may be due to the queries types where most of them are generic. Nevertheless, combining the different representations enhance the results.

In future work, we plan to compare our approach of building implicit links with other works such as the use of hyperlinks and some state of the art implicit links. We also plan to investigate the combination of explicit and implicit links in the image retrieval. On one hand, explicit hyperlinks are more semantic than implicit links if they are informational hyperlinks as they are created manually. On the other hand, it is not possible to link manually all similar information in the collection. So it is interesting to build automatically implicit links between images to take into account all possible similarities. Moreover, it is interesting to learn a classifier to tune automatically the best combination parameters of the image representations according to the query type: generic or specific. If the query vocabulary is generic, it is better to give more importance to the generic description than the specific one, and if the query vocabulary is specific, it is better to give more importance to the specific description than the generic one.

Another issue to investigate is the multi-modality image retrieval. The combination of content and context based image retrieval has shown its effectiveness in several works. Therefore, we plan to integrate visual features in the construction of implicit links. It is possible for example to build visual based links and explore them in the image retrieval process.

REFERENCES

Alzu'bi, A., Amira, A., and Ramzan, N. (2015). Semantic content-based image retrieval: A comprehensive study. *Journal of Visual Communication and Image Representation*, 32:20–54.

Belmouhcine, A. and Benkhalifa, M. (2015). Implicit links based web page representation for web page classification. In *Proceedings of the 5th International Conference on Web Intelligence, Mining and Semantics*, page 12. ACM.

Bhatt, C., Pappas, N., Habibi, M., and Popescu-Belis, A. (2014). Multimodal reranking of content-based recommendations for hyperlinking video snippets. In *Proceedings of International Conference on Multimedia Retrieval*, page 225. ACM.

Blei, D., Ng, A., and Jordan, M. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022.

Brandes, U. (2001). A faster algorithm for betweenness centrality. *Journal of mathematical sociology*, 25(2):163–177.

Brin, S. and Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. In *Proceedings of the 7th international conference on World Wide Web (WWW)*, pages 107–117, Brisbane, Australia.

Chen, S., Eskevich, M., Jones, G. J., and OConnor, N. E. (2014). An investigation into feature effectiveness for multimedia hyperlinking. In *International Conference on Multimedia Modeling*, pages 251–262. Springer.

Datta, R., Joshi, D., Li, J., and Wang, J. Z. (2008). Image retrieval: Ideas, influences, and trends of the new age. *ACM Computing Surveys (Csur)*, 40(2):5.

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., and Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American society for information science*, 41(6):391.

Eskevich, M., Jones, G. J., Aly, R., Ordelman, R. J., Chen, S., Nadeem, D., Guinaudeau, C., Gravier, G., Sébillot, P., Nies, T. D., et al. (2013). Multimedia information seeking through search and hyperlinking. In *Proceedings of the 3rd ACM conference on International conference on multimedia retrieval*, pages 287–294. ACM.

Hsu, W. H., Kennedy, L. S., and Chang, S. H. (2007). Video search reranking through random walk over document-level context graph. In *Proceedings of the 15th ACM international conference on Multimedia*, pages 971–980. ACM.

Jing, Y. and Baluja, S. (2008a). Pagerank for product image search. In *Proceedings of the 17th international conference on World Wide Web*, pages 307–316. ACM.

Jing, Y. and Baluja, S. (2008b). Visualrank: Applying pagerank to large-scale image search. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(11):1877–1890.

Khasanova, R., Dong, X., and Frossard, P. (2016). Multimodal image retrieval with random walk on multilayer graphs. *arXiv preprint arXiv:1607.03406*.

Kleinberg, J. (1998). Authoritative sources in a hyperlinked environment. In *Proc. 9th Annual ACM-SIAM Symposium Discrete Algorithms*, pages 668–677.

Kurland, O. and Lee, L. (2005). Pagerank without hyperlinks: Structural reranking using links induced by language models. In *Proceedings of the Annual International ACM/SIGIR Conference on Research and Development in Information Retrieval*, pages 306–313. ACM.

Kurland, O. and Lee, L. (2006). Respect my authority!: Hits without hyperlinks, utilizing cluster-based language models. In *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 83–90. ACM.

Li, B. and Han, L. (2013). Distance weighted cosine similarity measure for text classification. In *International Conference on Intelligent Data Engineering and Automated Learning*, pages 611–618. Springer.

Lin, J. (2008). Pagerank without hyperlinks: Reranking with pubmed related article networks for biomedical text retrieval. *BMC bioinformatics*, 9(1):270.

Liu, J., Lai, W., Hua, X. S., Huang, Y., and Li, S. (2007). Video search re-ranking via multi-graph propagation. In *Proceedings of the 15th ACM international conference on Multimedia*, pages 208–217. ACM.

Liu, Z., Wang, S., Zheng, L., and Tian, Q. (2017). Robust imagegraph: Rank-level feature fusion for image search. *IEEE Transactions on Image Processing*, 26(7):3128–3141.

Mikawa, K., Ishida, T., and Goto, M. (2011). A proposal of extended cosine measure for distance metric learning in text classification. In *Systems, Man, and Cybernetics (SMC), 2011 IEEE International Conference on*, pages 1741–1746. IEEE.

Salton, G. (1989). Automatic text processing: The transformation, analysis, and retrieval of information by computer. *Reading: Addison-Wesley*.

Shen, D., Sun, J. T., Yang, Q., and Chen, Z. (2006). A comparison of implicit and explicit links for web page classification. In *Proceedings of the 15th international conference on World Wide Web*, pages 643–650. ACM.

Simon, A. R., Sicre, R., Bois, R., Gravier, G., and Sébillot, P. (2015). Irisa at trecvid2015: Leveraging multimodal lda for video hyperlinking. In *TRECVid 2015 Workshop*.

T. Hofmann, T. (1999). Probabilistic latent semantic indexing. In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, pages 50–57. ACM.

Wang, M., Li, H., Tao, D., Lu, K., and Wu, X. (2012). Multimodal graph-based reranking for web image search. *IEEE Transactions on Image Processing*, 21(11):4649–4661.

Wang, X., Zhou, W., Tian, Q., and Li, H. (2016). Adaptively weighted graph fusion for image retrieval. In *Proceedings of the International Conference on Inter-*

net Multimedia Computing and Service, pages 18–21. ACM.

Xie, L., Tian, Q., Zhou, W., and Zhang, B. (2014). Fast and accurate near-duplicate image search with affinity propagation on the imageweb. *Computer Vision and Image Understanding*, 124:31–41.

Xu, G. and Ma, W. Y. (2006). Building implicit links from content for forum search. In *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 300–307. ACM.

Xue, G. R., Zeng, H. J., Chen, Z., Ma, W. Y., Zhang, H. J., and Lu, C. J. (2003). Implicit link analysis for small web search. In *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*, pages 56–63. ACM.

Zhang, S., Yang, M., Cour, T., Yu, K., and Metaxas, D. (2012). Query specific fusion for image retrieval. *Computer Vision–ECCV 2012*, pages 660–673.

Zhang, W., Ngo, C. W., and Cao, X. (2016). Hyperlink-aware object retrieval. *IEEE Transactions on Image Processing*, 25(9):4186–4198.

Zhang, X., Hu, X., and Zhou, X. (2008). A comparative evaluation of different link types on enhancing document clustering. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, pages 555–562. ACM.

Zhou, W., Li, H., and Tian, Q. (2017). Recent advance in content-based image retrieval: A literature survey. *arXiv preprint arXiv:1706.06064*.

Zhou, W., Tian, Q., and Li, H. (2009). Visual block link analysis for image re-ranking. In *Proceedings of the First International Conference on Internet Multimedia Computing and Service*, pages 10–16. ACM.