

A study of Impact of Integration Structural Context in Multimedia Retrieval: Application on Image Media

Sana FAKHFAKH, Mohamed TMAR, and Walid MAHDI

Laboratory MIRACL, Institute of Computer Science and Multimedia of Sfax, Sfax University, Tunisia
sanafakhfakh@yahoo.fr, mohamed.tmar@isimsf.rnu.tn, walid.mahdi@isimsf.rnu.tn
<http://www.miracl.rnu.tn>

Abstract. The expansion in volume of information organized according to a structure in multiple applications introduces a new equivocal in multimedia retrieval in semi-structured documents. We study in this paper the impact of introduce a structural context on multimedia retrieval in XML document thus we present a indexing model which combines textual and structural information. We propose a geometric method who use implicitly of textual and structural context of XML elements and we are particularly interested by improve the effectiveness of various structural factors for multimedia retrieval. Using a geometric metric, we can represent structural information in XML document with a vector for each element. Experimental evaluation is carried out using the INEX Ad Hoc Task 2007, ImageCLEF Wikipedia Retrieval Task 2010 and ImageCLEF Plant Task 2014 in the framework of our participation in CLEF 2014 campaign. The results show that integration of structural context significantly improves compared results of using a single textual context. Our proposed method perform as compared to other systems evaluated on two coollection INEX 2007 and ImageCLEF 2010.

Keywords: Structural context, Textual context, Approximative resolution, XML element, Image retrieval

1 Introduction

The joint evolution of user needs and electronic semi-structured documents constantly raises new challenges in the Information Retrieval (IR) field. The need with this kind of information is justified by quick change of scopes of application which use structural documents (format HTML or XML) what imposes new challenges in the field of search for information. Indeed, nowadays XML document passed a simple tool for exchanging data to a new storage medium. XML document includes textual element and multimedia element such as image, audio and video. These elements are organized according to structure which includes information notably although there is not only one manner to organize

contents. However, the choice of structure depends greatly on the context of use of the textual contents. Mainly in the literature, there are two main classes of approaches in the field of multimedia retrieval: retrieval methods based on multimedia content (MR-content) and multimedia methods to retrieval based on context (MR-Context).

The approaches of the multimedia retrieval based on content use specific features of low level according to type of media [1]. We can cite for example image retrieval that exploits visual features (the color, texture, forms ...). These methods have proven effective with media "image" in well defined fields such as medical field this is due to requirement for thorough knowledge of distinctive media. This type of research can be applied to only one type of media in system due to lack of semantic representation in media content.

The approaches of the multimedia retrieval based on context do not depend on type of media in question [2] [3]. Indeed, these methods rely on information surrounding the multimedia element representing its semantic description. Multimedia retrieval based on textual context is most used, although the structural context remains an obvious source which plays a part paramount in understanding of structured documents.

Multimedia retrieval based on textual context is most used, although the structural context remains an obvious source which plays a part paramount in understanding of structured documents. In this paper, we are interested in Context-based MIR techniques, and more precisely in MIR based on textual and structural context in XML documents. Image context is composed all textual information surrounding the image. For retrieve image presented in Figure 1, we can use text surrounding image such as document title, image name, image caption, etc

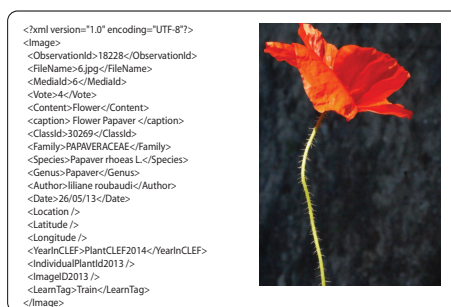


Fig. 1. Example of a multimedia element context.

The textual context remains insufficient in most of time. In this context, [4] say: "Ignore the document structure is to ignore its semantics". There are other sources of evidence that were used as visual descriptors, information from link around the image, structure of XML document. Indeed, We focus on XML

documents don't have a homogeneous structure. What makes the structure as new source of evidence.

In this article, we focus on techniques for multimedia retrieval based on textual and structural context in XML documents. This type of document includes textual information and structural constraints. So, XML document cannot be effectively exploited by classical techniques of IR, which regard document as a plane source of information. The implicit incorporation of multimedia elements in XML documents requires the exploitation of textual context for multimedia retrieval. However, the textual context remains insufficient in most of time. The idea is to calculate the relevancy score of media element based on information from the textual and structural context to answer a specific information needs of user, expressed as query composed of set of keywords. And seeking the most appropriate manner to combine two sources of evidence: text and structure. Our main inspiration is to use the structure to involve each textual information depending on its position in XML document, that is textual information that gives the best possible description of multimedia element. In our work, we will be interested by media "image". Most existing work in this area uses the information from textual description of image. There are other sources of evidence that were used as visual descriptors, information from link around the image [5], structure of XML document. To resolve difficulties in multimedia retrieval field, you must define adequate source of evidence for representation a multimedia element and defining appropriate indexing model. In this context, we present our structural indexing system combining conceptual information for semi-structured documents dedicated to approximate retrieval data. In Section 3, we presents the details of our proposed method where detailing the preprocessing, extraction of textual and structural and phase calculation relevance of multimedia element in information a better response to needs expressed by user and we describe our structural indexing system combining conceptual information for semi-structured documents dedicated to approximate retrieval data. In Section 5, we presents the results of her applying on three data sets "INEX 2007", "ImageCLEF 2010" and "ImageCLEF 2014". The last section provides our conclusions and future works.

2 RELATED WORKS

The advent of structured documents has caused new problems in information retrieval world, and more specifically in multimedia elements retrieval. These problems are strongly related to nature of these documents that provide the structure as a new source of evidence. Thus, nowadays, XML documents include multimedia elements of different types (audio, video and image) implicitly embedded in the textual elements. These multimedia elements (such as physical objects) do not contain enough information to be able to answer a given query. Therefore, the calculation of relevance score of multimedia element must be linked to textual and structural information provided by other nodes XML [5]. Indeed, XML document is used to describe a set of data by a structure that

provides a semantic lexicon. Thus, it facilitates the presentation of information in terms of interpretation and exploitation. Replying to this need, new works appear in the field of multimedia retrieval that takes in account the structure as source of relevant information. Existing work in structured retrieval of multimedia elements is decomposed in two classes.

The first class includes some works which proceed to adopt some traditional technical of retrieval information as language model. In this context, the team *CWI/UTwente* performs a step of filtering results to keep the fragments containing at least one multimedia element [6][7].

The second class includes the specific work to be structured multimedia retrieval. This class uses the structure as a source of evidence in the process of selection of multimedia elements. As first step, [8] proposed a method which combines structure of XML document (XPath) with the use of links (XLink). This method consist to divide XML document into regions. Each region represent a area of ancestors of the multimedia element. His score is calculated in function of the scores of each region. This method exploits vertical structure only. In a second time, [9] have used the addition of horizontal structure to the notion of hierarchy. [9] use a method called "CBA" (Children, Brothers, Ancestors), which takes into consideration the information carried by the children , brothers and fathers nodes for calculate the relevance of multimedia elements. The authors propose an alternative method "OntologyLike" which is based on the identification of XML document to ontology. To calculate the similarity between nodes the authors use similarity measures that are mainly based on the number of edges to calculate the distance between nodes.

There are other approaches to multimedia retrieval are based on exploitation of links in XML document [10]. This work was improved by proposing a hybrid approach that combines structure with using of links who is consider as semantic links [11]. This method above to divide the document into regions according the hierarchical structure and the location of image in document. This factor plays a role in the weighting of links for compute the score of image.

In this paper, we propose a new metric for multimedia retrieval in XML documents which involves the use of geometric distances to calculate the relevance of each node from the multimedia node. This method consists of placing the nodes of XML document in Euclidean space and define each node by a vector of coordinates to calculate then the distance between each pair of nodes. This distance will play a beneficial role in to calculate the score of multimedia element.

3 Proposed Approach

We focus on techniques for multimedia retrieval based on textual and structural context in XML documents. XML documents cannot be effectively exploited by classical techniques of IR, which regard document as a bog of words. Therefore, the calculation of relevance score of multimedia element must be linked to textual and structural information provided by other nodes XML [5]. Thus, it facilitates the presentation of information in terms of interpretation and exploitation. Re-

plying to this need, we propose a new method in the field of multimedia retrieval that takes into account the structure as a source of evidence and its impact on search performance. We present a new source of evidence dedicated to multimedia retrieval based on the intuition that each textual node contains information that describes semantically a multimedia element. And the participation of each text node in the score of a multimedia element varies with its position in there XML document. To compute the geometric distance, we initially place the nodes of each XML document in an Euclidean space to calculate the coordinates of each node by algorithm 1. Then, we compute the score of a multimedia element depending on the distance between each textual node. Figure 2 shows the steps

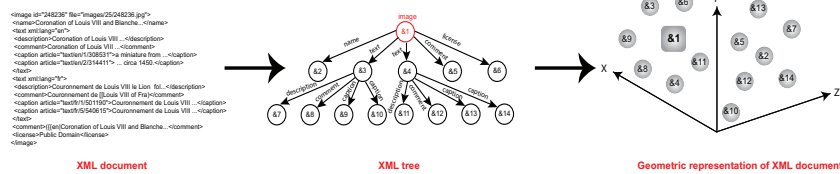


Fig. 2. The steps of passing an XML document to geometric representation.

of passing an XML document to a geometric representation of the XML elements in a Euclidean space. The first step consist to present a XML document as XML tree to take into account XML document properties. An XML tree is described by a set of relationships between nodes. Formally an XML tree is a pair $A = (E, R)$ where E is a set of XML elements and $R \subset E^2$, $((p, q) \in R$ if p is the parent of q) is a set of relations satisfying:

$$\exists! r \in E, \forall q \in E - \{r\}, (r, q) \in R \tag{1}$$

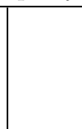
With r is the root of the tree.

$$\forall p \in E - \{r\}, \exists! q \in E, (p, q) \in R \tag{2}$$

Each node has a parent except the root r . In second step, we will spend to presentation of XML tree in a geometric representation. This step is mainly based on equalities extraction in XML tree according to our proposed hypotheses. The XML tree representation allowed us to unveil certain relationships of neighboring, brotherhood and offspring. Indeed, the distance d which separate two or more brothers with their common ancestors iteratively is the same. And brothers of the same hierarchical level are equidistant. These distances are defined according to the relationship of contiguity and semantic similarity between nodes. These distances are not quantized but will be extracted in function of the position of each textual node in XML tree. All these properties result in: For all $q_i = (x_{i1}, x_{i2} \dots x_{im})$ and $q_j = (x_{j1}, x_{j2} \dots x_{jm})$ where Q is a set of vectors in \mathbb{R}^m .

- In the same hierarchy, if there are more than two brothers then their adjacent nodes are equidistant:

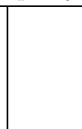
property 1



$$\forall q_i, q_j, q_k \in Q, \text{ if } A_1(q_i) = A_1(q_j) = A_1(q_k) \\ d(q_i, q_j) = d(q_i, q_k)$$

- The distance between any node and its descendants is the same:

property 2



$$\forall q_i, q_j, q \in Q, n \in \mathbb{N}, A_n(q_i) = A_n(q_j) = q \\ d(q_i, q) = d(q_j, q)$$

With $\forall n \in \mathbb{N}^*$, we define function A_n by: $\forall q \in E$,

$$A_n(q) = \begin{cases} \{q\} & \text{if } n = 0 \\ A_{n-1}(p) & \text{if } \exists p \in E, (p, q) \in R \text{ and } n > 0 \\ \emptyset & \text{else} \end{cases}$$

From these relationships, we can generate system of equations taking into account for kinship relationships nodes based on hierarchy and adjacency. These relationships are decryed by equalities in this order (these equations are only examples)(Figure 2):

$$\begin{aligned} d(n_1, n_2) &= d(n_1, n_3) \\ d(n_1, n_2) &= d(n_1, n_4) \\ d(n_1, n_7) &= d(n_1, n_8) \\ d(n_1, n_7) &= d(n_1, n_9) \end{aligned}$$

These distances are defined according to the relationship of contiguity and semantic similarity between nodes. They are not quantized but will be extracted in function of the position of each textual node in the XML tree. The resulting system is nonlinear, its resolution requires the use of an approximate resolution iteratively method where we used iterative solution method (see Algorithm 1). The process begins by assigning to each XML node a random vector. It Tries to improve the coordinate values of each node according to an error value (the sum of the squared deviations). At each iteration, the coordinates are improved together with the minimization of this error. The algorithm stops when the error reaches its minimum value (no improvement is possible). Let Q the set of vectors obtained at a given iteration during the running of the algorithm, the error is defined by:

$$\begin{aligned} error(Q) &= \sum_{\substack{q_i, q_j, q_k \in Q, \\ A_1(q_i) = A_1(q_j) = A_1(q_k)}} (d(q_i, q_j) - d(q_i, q_k))^2 \\ &+ \sum_{\substack{q_i, q_j, q \in Q, n \in \mathbb{N}, \\ A_n(q_i) = A_n(q_j) = q}} (d(q_i, q) - d(q_j, q))^2 \end{aligned}$$

Algorithm 1 Resolution algorithm approximate nonlinear system of equations

Require: $(Q = (q_1, q_2 \dots q_{|Q|}), R)$:an XML tree as $q_i = (q_{i1}, q_{i2} \dots q_{im}) \quad \forall i \in [1, |Q|]$
 m :dimension
for $(i, j) \in [1, |Q|]^2$ **do**
 $q_{ij} \leftarrow$ random value
end for
 $Q_1 \leftarrow (q_1, q_2 \dots q_{|Q|})$
repeat
 $P \leftarrow Q_1$
 for $(i, j) \in [1, |Q|]^2$ **do**
 $Q_2 \leftarrow (q_1, q_2 \dots q_{i-1}, q_i + d_j(1), q_{i+1} \dots q_{|Q|})$
 $Q_3 \leftarrow (q_1, q_2 \dots q_{i-1}, q_i + d_j(\varepsilon), q_{i+1} \dots q_{|Q|})$
 $Q_4 \leftarrow (q_1, q_2 \dots q_{i-1}, q_i + d_j(1 - \varepsilon), q_{i+1} \dots q_{|Q|})$
 $t \leftarrow 0$
 while $error(Q_1) > error(Q_2) > error(Q_3) > error(Q_4)$ **do**
 $Q_4 = (q_1, q_2 \dots q_{i-1}, q_i + 2^t d_j(1), q_{i+1} \dots q_{|Q|})$
 $t = t + 1$
 end while
 $t \leftarrow 0$
 while $error(Q_1) < error(Q_2) < error(Q_3) < error(Q_4)$ **do**
 $Q_1 = (q_1, q_2 \dots q_{i-1}, q_i - 2^t d_j(1), q_{i+1} \dots q_{|Q|})$
 $t = t + 1$
 end while
 while $|error(Q_1) - error(Q_2)| > \varepsilon$ **do**
 $Q_5 \leftarrow \frac{Q_1 + Q_2}{2}$
 let $Q_5 = (p_1, p_2 \dots p_{|Q|})$
 if $error(p_1, p_2 \dots p_{i-1}, p_i - d_j(\varepsilon), p_{i+1} \dots p_{|Q|}) > error(p_1, p_2 \dots p_{i-1}, p_i + d_j(\varepsilon), p_{i+1} \dots p_{|Q|})$ **then**
 $Q_1 \leftarrow Q_5$
 else
 $Q_2 \leftarrow Q_5$
 end if
 end while
 end for
until $P = Q_1$

Where m is the dimension of the Euclidean space and $\forall v \in \mathbb{R}$, $D_j(v) = (d_1, d_2 \dots d_m)$ is such as:

$$d_k = \begin{cases} 0 & \text{if } k \neq j \\ v & \text{otherwise} \end{cases}$$

3.1 Multimedia Element Representation by Textual and Structural Context

A multimedia element (eg *image*) does not contain textual content. Its score is based on textual nodes in its neighborhood. The transition from the XML tree

structure representation of elements in an Euclidean space, where we exploit the dissimilarity distances separating a multimedia node and other textual nodes, is performed by extracting the equations satisfying the properties defined earlier and the application of algorithm 1. To calculate the distance between a node n and multimedia element H , we will try to use several geometric distances such as Manhattan distance, Euclidean distance and Minkowski distance between their respective feature vectors q_n and q_H described by the following equations:

$$dist_{Manhattan}(n, H) = \sum_{i=1}^m |q_n - q_H| \quad (3)$$

$$dist_{Euclidean}(n, H) = \sqrt{\sum_{i=1}^m (q_n - q_H)^2} \quad (4)$$

$$dist_{Minkowski}(n, H) = \sqrt[p]{\sum_{i=1}^m |q_n - q_H|^p} \quad (5)$$

With m is the dimension of the Euclidean space and $p=1$. q_n is defined by: $q_n=(xn_{i1}, xn_{i2} \dots xn_{im})$ with xn are the vector characteristics of node n . And q_H is defined by: $q_H=(xH_{i1}, xH_{i2} \dots xH_{im})$ with xH represent the coordinates compose the vector characteristics of a node H . We calculate the score for each textual node depending on the frequency of each term (tf) and the number of elements in the corpus according to the number of elements containing the term (idf). A textual node is presented by: $n = (n_1, n_2 \dots n_{|v|})$ where n_i is the weight of the term t_i , v is the set of indexing terms:

$$n_i = tf(t_i, n) \times idf(t_i) \quad (6)$$

With

$$idf(t_i) = \log\left(\frac{N}{N_i}\right) \quad (7)$$

Where N is the total number of XML elements in the corpus, N_i is the number of elements that contain the term t_i and $tf(t_i, n)$ is the frequency of the term t_i in node n . The score of textual node depends on the weight of each indexing term. A query is made by the list $v = (v_1, v_2 \dots v_{|v|})$ where $v_i \in \{0, 1\}$ (0: not exist, 1: exist) according membership t_i at the query. The score of textual node n for the query q is defined by:

$$rsv(q, n) = q \times n^T = \sum_{i=1}^{|V|} q_i \times n_i \quad (8)$$

Where μ is the set of textual elements. The score of multimedia node H is defined by:

$$rsv(q, H) = \sum_{n \in \mu} \frac{rsv(q, n)}{dist(n, H)} \quad (9)$$

With $dist(n, H)$ is the distance (Manhattan distance or Euclidean distance or Minkowski distance) between the feature vectors corresponding to the nodes n and H . This equation leads to assign the importance of contribution of all nodes in computing the score of multimedia element that shows its beneficial impact in multimedia retrieval.

4 Indexing System

We propose a indexing system *MXS – index* composed by two parts: part of textual indexing and part of structural indexing. Our indexing methodology as schematized in Figure 3. The first part consists of four main steps: Pretreatment, term extraction and term weighing using NLP (Natural Language Processing) techniques to extract the candidate XML nodes of the resulting indexing. The first step is to split text into a set of sentences, prune the stop words for each XML node of the corpus and radicalize terms using the algorithm PORTER [12]. The second step is term extraction and the last step is calculating term importance. That is a fundamental step in information retrieval process and it is determined through term frequency (*tf*) and inverse document frequency (*idf*). In Second part, we built structural index using information extract from XML tree and geometric metric. Each XML node will presented by characteristic vector. We start by extract geometric proprieties. And we compute coordinates of each XML nodes. This part is accompanied by generating XML data model which processes ancestor, descendant and proximity relationships (Figure 3). The step of selection of descriptors of each node consists in associating each XML node own these textural and structural descriptors to better combine.

5 Evaluation and Results

We evaluate our system on three databases extracted from three collections : INEX 2007 (Initiative for the Evaluation of XML Retrieval) Ad Hoc task [13], ImageCLEF 2010 Wikipedia image retrieval task [14] and ImageCLEF 2014 Plant task [15]. The first two databases are composed by XML documents extracted from Wikipedia. The latest dataset is collected by scientific community for testing and validation of their approaches (Table 1).

The aim of the experiments in this section is to show the effectiveness of XML structure in multimedia retrieval. For this purpose, we evaluated separately the use of textual context only (**TC**), as well as the combination of the

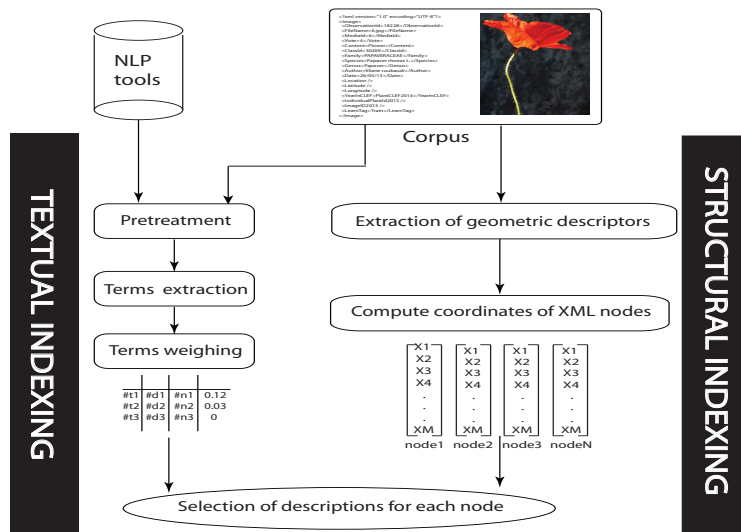


Fig. 3. Architecture of our indexing model *MXS – index*.

Table 1. INEX 2007, ImageCLEF 2010 and ImageCLEF 2014 collections

| Company | INEX 2007 | CLEF 2010 | CLEF 2014 |
|------------------------|-----------------------|---------------------|-----------------|
| Task | Collection XML Ad Hoc | Wikipedia Retrieval | Plant Retrieval |
| Number of XML document | 659388 | 237434 | 47815 |
| Number of image | 246730 | 237434 | 47815 |
| Topics | 19 | 70 | 8163 |

two (TC and TS). For INEX 2007, ImageCLEF 2010 and ImageCLEF 2014 test set, we respectively obtain the following MAP values: 0.2376, 0.1674 and 0.2488 using textual context only (figure 4). We compare between the use of Manhattan distance, Euclidean distance and Minkowski distance. We observed that the difference of results between the three distances is very significant in the INEX 2007 test set, ImageCLEF 2010 test set and ImageCLEF 2014 test set. The Euclidean distance gets a most suitable representation of multimedia element which is none other than the dissimilarity distance between XML nodes. Indeed, the evaluation results show that this distance provides a MAP which is equal to 0.2572 as MAP with using "ImageCLEF 2010" collection. The result has been improved significantly with the "INEX 2007" collection to 0.3102 as MAP relative to Manhattan distance (0.2376 for "INEX 2007" collection, 0.1754 for "ImageCLEF 2010" collection and 0.2276 for "ImageCLEF 2014" collection) and Minkowski distance (0.2876 for "INEX 2007" collection, 0.2245 for "ImageCLEF 2010" collection and 0.3267 for "ImageCLEF 2014" collection). This increase is due to nature of "INEX 2007" collection who includes XML documents with heterogeneous structure. So in "INEX 2007" collection we find documents with

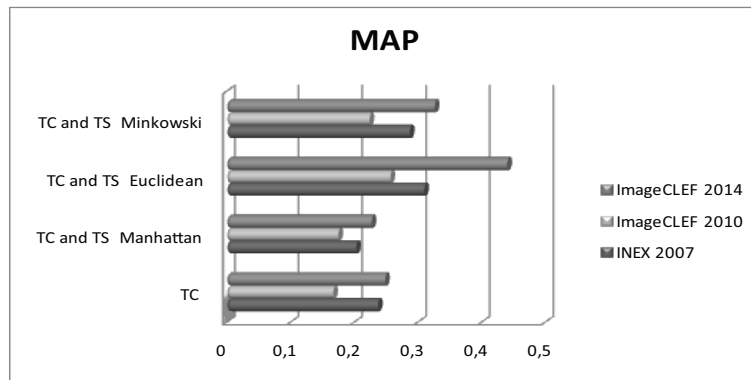


Fig. 4. Results of the impact our approach on INEX 2007, ImageCLEF 2010 and ImageCLEF 2014 based in MAP(Mean Average Precision).

high depth. This factor highlights structural information and amplifies effect textual information based on computed distances. For against, our system is more stable with "ImageCLEF 2010" collection, this is due to rapid convergence of results. With our measure, we have shown that combined use of textual and structural context can properly determine the relevance of multimedia element, and the structure plays a primordial role in multimedia retrieval (Figure4). We can conclude that an using structural information returns better pertinence in case uses multimedia retrieval with using "ImageCLEF 2014" collection with MAP equal to 0.4406. In fact, this collection contains quite specific documents include descriptions of plants which reduces textual concepts; see Figure 4.

After evaluation of our system on the three described collections, we will try to position itself compared to participants in INEX 2007 and ImageCLEF 2010. Our system gives better results using INEX 2007 collection (figure5). Indeed, comparing work proposed by IRIT system, we obtained a MAP equal to 0.31 [13][9]. Our system also gives a better MAP which is equal to 0.25 compared to XRCE system with using textual context with ImageCLEF 2010 collection (figure 6) and he got the better result with textual and visual context a MAP value which is equal to 0.27 [14].

6 CONCLUSION

This approach allowed us to calculate the score of element multimedia according the textual context provided by nodes in proximity and structural context from distance between nodes and multimedia element. This method was evaluated with using of three collections "INEX 2007", "ImageCLEF 2010" and "ImageCLEF 2014". In this work, we studied the impact of textual and structural context on multimedia element retrieval, where the user need can be a multimedia element (text). We plan to investigate the impact of a mixture of text and multimedia element(text+image) with to using visual descriptors.. In

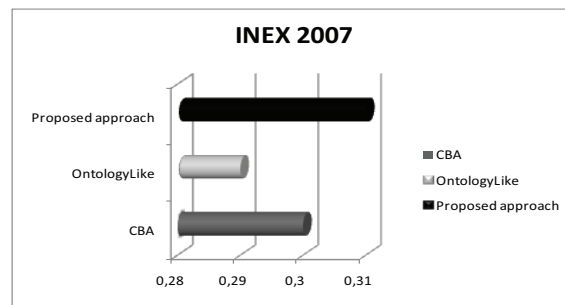


Fig. 5. Compared results on INEX 2007 collection with MAP(Mean Average Precision) metric.

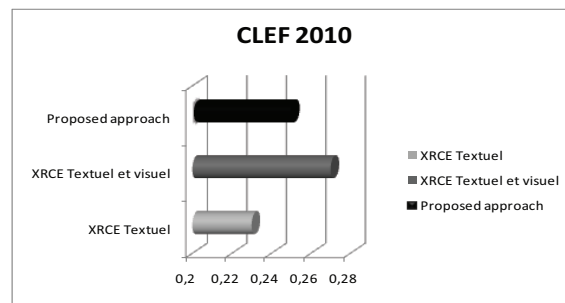


Fig. 6. Compared results on ImageCLEF 2010 collection with MAP(Mean Average Precision) metric.

the future, we want to exploit another factor to calculate the relevance of multimedia element such as the title of image, the weighting of the links in XML document ... As well as another source of evidence as visual descriptors.

References

1. Michael S. Lew. Content-based multimedia information retrieval: State of the art and challenges. *ACM Trans. Multimedia Comput. Commun. Appl.*, 2:1–19, 2006.
2. Haytham Elghazel, Khalid Idrissi, Atilla Baskurt, and Chokri Ben Amar. Approche textuelle pour la recherche d'image. In *3rd International Conference on Sciences of Electronic, Technologies of Information and Telecommunications SETIT 2005*, March 2005.
3. Dian Tjondronegoro, Jinglan Zhang, Jinfeng Gu, Anthony Nguyen, and Shlomo Geva. Integrating text retrieval and image retrieval in xml document searching. In *INEX*, pages 511–524, 2005.
4. T. Schlieder and M. Holger. Querying and ranking xml documents. *Journal of the American Society for Information Science and Technology*, 53:489–503, 2002.
5. Angelos Hliaoutakis, Giannis Varelas, Epimeneidis Voutsakis, Euripides G. M. Petrakis, and Evangelos Milios. Information retrieval by semantic similarity. In *Intern. Journal on Semantic Web and Information Systems (IJSWIS). Special Issue of Multimedia Semantics*, pages 55–73, 2006.

6. Theodora Tsikrika, Pavel Serdyukov, Henning Rode, Thijs Westerveld, Robin Aly, Djoerd Hiemstra, and Arjen P. Vries de. Structured document retrieval, multimedia retrieval, and entity ranking using pf/tijah. In *6th Initiative on the Evaluation of XML Retrieval, INEX 2007*, volume 4862 of *Lecture Notes in Computer Science*, pages 306–320, London, March 2008. Springer Verlag.
7. Thijs Westerveld, Henning Rode, Roel Os van, Djoerd Hiemstra, Georgina Ramirez, Vojkan Mihajlovic, and A.P. Vries de. Evaluating structured information retrieval and multimedia retrieval using pf/tijah. In N. Fuhr, M. Lalmas, and A. Trotman, editors, *Comparative Evaluation of XML Information Retrieval Systems*, volume 4518 of *Lecture Notes in Computer Science*, pages 104–114, Berlin, Germany, 2007. Springer Verlag.
8. Zhigang Kong and Mounia Lalmas. Xml multimedia retrieval. In *SPIRE*, pages 218–223, 2005.
9. Mouna Torjmen, Karen Pinel-Sauvagnat, and Mohand Boughanem. Using textual and structural context for searching multimedia elements. *IJBIDM*, 5(4):323–352, 2010.
10. Hatem Awadi and Mouna Torjmen. Exploitation des liens pour la recherche d’images dans des documents xml. March 18-20 2010.
11. Hatem Aouadi, Mouna Torjmen-Khemakhem, and Maher Ben Jemaa. Combination of document structure and links for multimedia object retrieval. *Journal of Information Science*, 38(5):442–458, October 2012.
12. M. Porter. An algorithm for suffix stripping. *Program*, 14(3):130–137, 1980.
13. Norbert Fuhr, Jaap Kamps, Mounia Lalmas, Saadia Malik, and Andrew Trotman. Overview of the inex 2007 ad hoc track. In *INEX*, pages 1–23, 2007.
14. Adrian Popescu, Theodora Tsikrika, and Jana Kludas. Overview of the wikipedia retrieval task at imageclef 2010. In *CLEF (Notebook Papers/LABs/Workshops)*, 2010.
15. Linda Cappellato, Nicola Ferro, Martin Halvey, and Wessel Kraaij, editors. *Working Notes for CLEF 2014 Conference, Sheffield, UK, September 15-18, 2014*, volume 1180 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2014.