Investigating the combination of structural and textual information about multimedia retrieval

Sana FAKHFAKH
Computer Science Department
Laboratory MIRACL, Institute of Computer Science and Multimedia
University of Sfax, Tunisia

Mohamed TMAR
Computer Science Department
Laboratory MIRACL, Institute of Computer Science and Multimedia
University of Sfax, Tunisia

Walid MAHDI
Computer Science Department
Laboratory MIRACL, Institute of Computer Science and Multimedia
University of Sfax, Tunisia

***Abstract*—The expansion of structured information in different applications introduces a new ambiguity in multimedia retrieval in semi-structured documents. We investigate in this paper the combination of textual and structural context for 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.**

Given a textual query, our model lets us combine scores obtained from each sources of evidence and return a list of relevant retrieved multimedia element. Experimental evaluation is carried out using the INEX Ad Hoc Task 2007 and the Image CLEF Wikipedia Retrieval Task 2010. The results show that combination of scores of textual modality and structural modality significantly improves compared results of using a single modality.

Keywords—Geometric distance; multimedia retrieval; element; structure; document modeling

I. INTRODUCTION

This paper falls under the context of multimedia retrieval in XML documents. 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][2]. We can cite for example image retrieval that exploits visual features (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 [3] [4]. 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.

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.

Let us take for example an image media. If we exploit the image context which is composed by description of its contents such as its title, name, descriptive texts which surround it, title of XML document ... In following figure, we present document extracted from "WIKIPEDIA" encyclopedia describing lion. We notice the existing simultaneous textual and multimedia information. For image retrieval from time after "Pleistocene", we extract information from the textual description, not from title (figure 1).

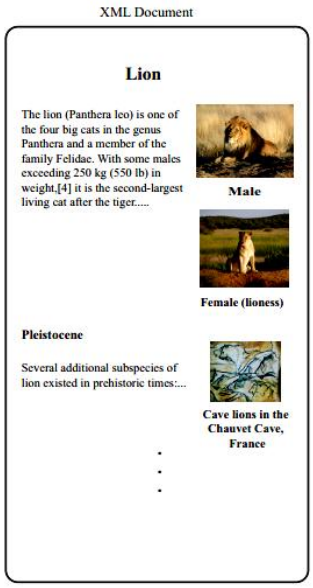

Fig. 1. Example of a multimedia object context.

II. Related Works

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], and 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. We begin with an overview of existing work in multimedia retrieval. Then we turn to the presentation of our approach while 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. Finally, we present the results of applying our method on two universal bases "INEX 2007" and "ImageCLEF 2010".

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].

Several works deal XML document as a flat source of information and ignore the structure of XML documents. In this context, [6] say: "Ignore the document structure is to ignore its semantics". 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 [7][8].

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, [9] proposed a method which combines structure of XML document (XPath) with the use of links (XLink). This method consists to divide XML document into regions. Each region represents an 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, [10] have used the addition of horizontal structure to the notion of hierarchy. [10] 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 [11]. This work was improved by proposing a hybrid approach that combines structure with using of links that is consider as semantic links [12]. This method above consists 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 defines 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.

III. From XML Element To Geometric Characteristic

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. Replying 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 a 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 [15].

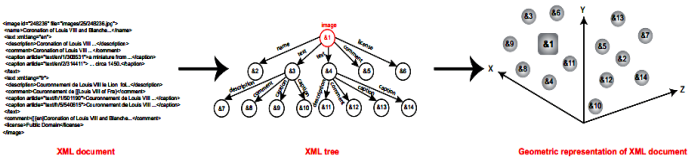

Fig. 2. The steps of passing an XML document to geometric representation.

Figure 2 shows the steps of passing an XML document to a geometric representation of the XML elements in a Euclidean space. The first step consists 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 $\subset E^2$, (*(p,q)* $\in R$ if p is the parent of q) is a set of relations satisfying:

$$\exists! \boldsymbol{r} \in \boldsymbol{E}, \forall \boldsymbol{q} \in \boldsymbol{E}, (\boldsymbol{r}, \boldsymbol{q}) \in \boldsymbol{R} \quad \textbf{(1)}$$

With *r* is the root of the tree.

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

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 $q_i = (x_{i1}, x_{i2} \cdots x_{im})$ and $q_j = (x_{j1}, x_{j2} \cdots x_{jm})$ 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

$$\forall q_i, q_j, q_k \in Q, if\ A_1(q_i) = A_1(q_j) = A_1(q_k)$$

$$d(q_i, q_j) = d(q_j, q_k)$$

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

Property 2

$$\forall q_i, q_j, q \in Q, \quad n \in \mathbb{N},$$

$$if\ A_n(q_i) = A_n(q_j) = q$$

$$d(q_i, q) = d(q_j, q)$$

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

$$\forall q \in E,$$

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

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

$$d(n_1, n_2) = d(n_1, n_3)$$
$$d(n_1, n_2) = d(n_1, n_4)$$
$$d(n_1, n_7) = d(n_1, n_8)$$
$$d(n_1, n_7) = d(n_1, n_9)$$

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 followed by 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:

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

Algorithm 1 Resolution algorithm approximate nonlinear system of equations

Require: $(Q = (q_1, q_2 \cdots q_{|Q|}), R)$: an XML tree as $q_i = (q_{i1}, q_{i2} \cdots q_{im}) \ \ \forall i \in [1, |Q|]$

m:dimension

for $(i,j) \in [1,Q]^2$ **do**
$q_{ij} \leftarrow random\ value$
end for

$Q_1 \leftarrow (q_1, q_2 \cdots q_{|Q|})$

Repeat
$P \leftarrow Q_1$
for $(i,j) \in [1,Q]^2$ **do**
$Q_2 \leftarrow (q_1, q_2 \cdots q_{i-1}, q_i + d_j(1), q_{i+1} \cdots q_{|Q|})$
$Q_3 \leftarrow (q_1, q_2 \cdots q_{i-1}, q_i + d_j(\varepsilon), q_{i+1} \cdots q_{|Q|})$
$Q_4 \leftarrow (q_1, q_2 \cdots q_{i-1}, q_i + d_j(1-\varepsilon), q_{i+1} \cdots q_{|Q|})$
$t \leftarrow 0$

While *error(Q1)> error(Q2)> error(Q3)> error(Q4)* **do**
$Q_4 = (q_1, q_2 \cdots q_{i-1}, q_i + 2^t d_j(1), q_{i+1} \cdots q_{|Q|})$
$t = t + 1$
end while

$t \leftarrow 0$

While *error(Q1)< error(Q2)<error(Q3)< error(Q4)* **do**
$Q_1 = (q_1, q_2 \cdots q_{i-1}, q_i - 2^t d_j(1), q_{i+1} \cdots q_{|Q|})$
$t = t + 1$
end while

While $|error(Q1) - error(Q2)| > \varepsilon$ **do**
$Q_5 \leftarrow \dfrac{Q_1 + Q_2}{2}$
let $Q_5 = (p_1, p_2 \cdots p_{|Q|})$

if $error(p_1, p_2 \cdots p_{i-1}, p_i - d_j(\varepsilon), p_{i+1} \cdots p_{|Q|}) > error(p_1, p_2 \cdots p_{i-1}, p_i + d_j(\varepsilon), p_{i+1} \cdots p_{|Q|})$ **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 $\forall v \in \mathbb{R}, D_j = (d_1, d_2 \cdots d_m)$ is such as:

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

A. INDEXING SYSTEM

We propose an indexing system ***MXS-index*** composed by two parties: party of textual indexing and party of structural indexing. In first party, our approach uses NLP (Natural Language Processing) techniques to extract the candidate XML nodes of the resulting indexing. The weight of these nodes is depending on the frequency of each of these terms and the number of elements in the corpus according to the number of elements containing the term. In Second party, we built structural index using information extract from XML tree and geometric metric.

Each XML node will be presented by a characteristic vector (figure 3). We start by extract geometric proprieties. And we compute coordinates of each XML nodes. This party is accompanied by generating XML data model which processes ancestor, descendant and proximity relationships (figure 4).

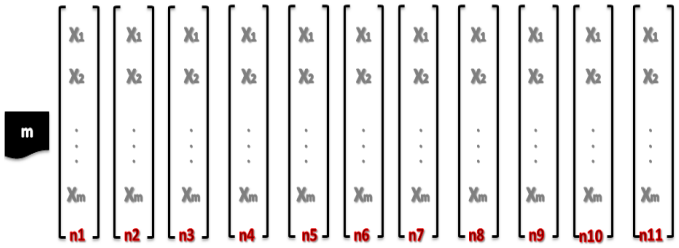

Fig. 3. Geometric characteristic vector of XML node

Figure 5 schematize the process of textual and structural indexing XML documents with our indexing system. Well as the transition of XML document as a tree presentation to geometric presentation in Euclidean space.

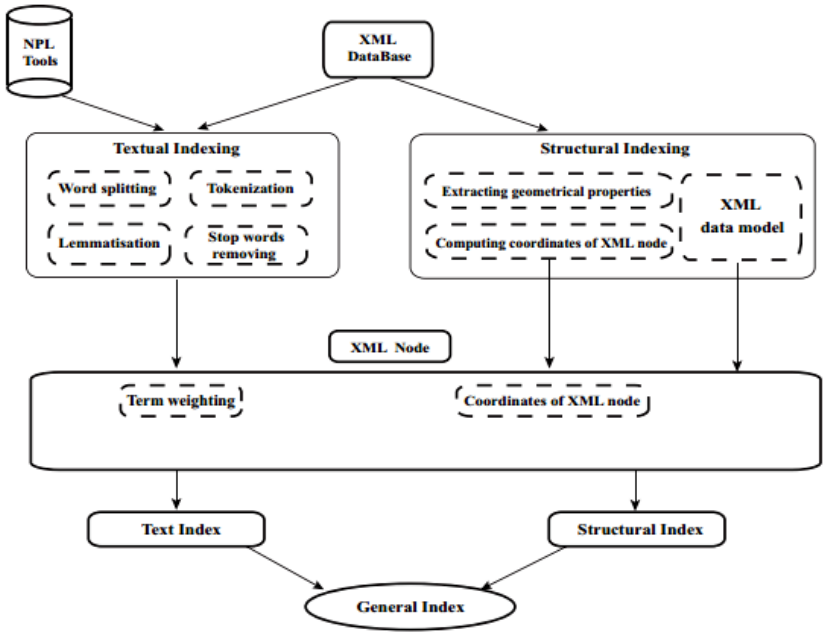

Fig. 4. Architecture of our indexing model MXS – index.

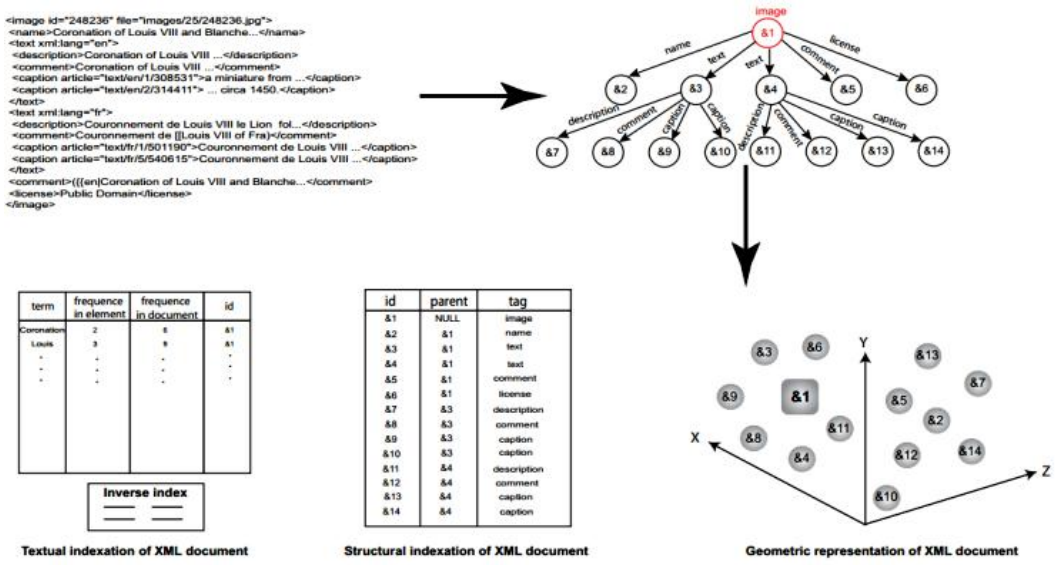

Fig. 5. Treatment process of XML document.

B. ADDING STRUCTURAL INFORMATION IN THE RELEVANCE SCORE OF MULTIMEDIA ELEMENT

A multimedia element (e.g. 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 a 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 calculate the Euclidean distance between their respective feature vectors q_n and q_H:

$$\boldsymbol{dist(n,H)} = \sqrt{\sum_{i=1}^{m}(\boldsymbol{q_n} - \boldsymbol{q_H})^2} \quad \textbf{(3)}$$

With *m* is the dimension of the Euclidean space. $q_{n\ is}$ defined by: $q_n = (xn_{i1}, xn_{i2} \cdots xn_{im})$ with *xn* are the vector characteristics of node *n*. And q_H is defined by: $q_{Hn} = (xH_{i1}, xH_{i2} \cdots xH_{im})$ 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: $n = (n_1, n_2 \cdots n_{|v|})$ where n_i is the weight of the term t_i, v is the set of indexing terms:

$$\boldsymbol{ni} = \boldsymbol{tf}(\boldsymbol{t_i}, \boldsymbol{n}) \times \boldsymbol{idf}(\boldsymbol{t_i}) \quad (4)$$

With

$$\boldsymbol{idf}(\boldsymbol{t_i}) = \boldsymbol{log}(\frac{N}{N_i}) \quad (5)$$

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 $v = (v_1, v_2 \cdots v_{|v|})$ 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:

$$\boldsymbol{rsv}(\boldsymbol{q}, \boldsymbol{n}) = \boldsymbol{q} \times \boldsymbol{n^T} = \sum_{i=1}^{|V|} \boldsymbol{q_i} \times \boldsymbol{n_i} \quad (6)$$

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

$$\boldsymbol{rsv}(\boldsymbol{q}, \boldsymbol{H}) = \sum_{\boldsymbol{n} \in \mu} \frac{\boldsymbol{rsv}(\boldsymbol{q}, \boldsymbol{n})}{\boldsymbol{dist}(\boldsymbol{n}, \boldsymbol{H})} \quad (7)$$

With $dist(n, H)$ is the distance between 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.

IV. Evaluation And Results

We evaluate our system into two databases extracted from two collections: INEX 2007 (Initiative for the Evaluation of XML Retrieval) Ad Hoc task [13] and ImageCLEF 2010 Wikipedia image retrieval task [14]. These databases are composed by XML documents extracted from Wikipedia (Table I).

TABLE I. INEX 2007 And ImageCLEF 2010 Collections

Company	INEX 2007	ImageCLEF 2010
Task	*Collection XML Ad Hoc*	*Wikipedia Retrieval*
Number of XML document	659388	237434
Number of image	246730	237434
Topics	19	70

We evaluate our method with using only textual context (TC). The XML structure is not taken account. For INEX 2007 and ImageCLEF 2010 test set, we respectively obtain the following MAP values: 0.2376 and 0.1674. In the second time, we use XML structure will determine the image relevance score and will differentiate between images (using textual and structural context TC and TS). The evaluation results show that this method 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. This increase is due to nature of "INEX 2007" collection that includes XML documents with heterogeneous structure.

So in "INEX 2007" collection we find documents with 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 (Figure6).

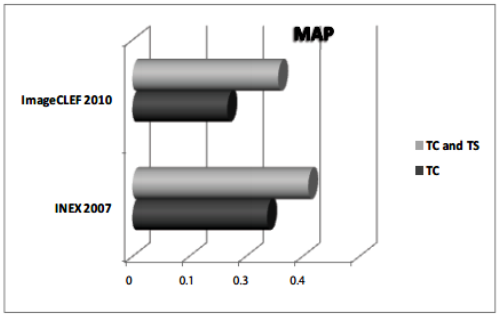

Fig. 6. Results of the impact our approach on INEX 2007 and ImageCLEF 2010 based in MAP(Mean Average Precision).

V. Conclusion

In this paper, we propose a novel approach for multimedia retrieval in XML documents. This method consists to calculate the score of multimedia element according the textual context provided by nodes in proximity and structural context from distance between nodes and multimedia element. Thank to geometric metric, we could assign a weight to each textual node in the XML document. Although all textual parts are useful, they should not all be taken into account with the same importance degree.

Experiments show the interest of our method on INEX 2007 and ImageCLEF 2010 collections. Our work is focused on media image but it can be used with any other media, since the visual context of multimedia objects is not used.

In 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 and the study parameters combination of using of structural, textual and visual context.

References

[1] M. S. Lew, "Content-based multimedia information retrieval: State of the art and challenges", *ACM Trans. Multimedia Comput. Commun. Appl*, vol. 2, pp. 1–19, 2006.

[2] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years", IEEE Trans. Pattern Anal. Mach. Intell., vol. 22, no. 12, pp. 1349–1380, Dec. 2000. [Online]. Available: http://dx.doi.org/10.1109/34.895972

[3] H. Elghazel, K. Idrissi, A. Baskurt, and C. Ben Amar, "Approche textuelle pour la recherche d'image", *in 3rd International Conference on Sciences of Electronic, Technologies of Information and Telecommunications SETIT 2005*, Mar. 2005. [Online]. Available: http://liris.cnrs.fr/publis/?id=2153

[4] D. Tjondronegoro, J. Zhang, J. Gu, A. Nguyen, and S. Geva, "Integrating text retrieval and image retrieval in xml document searching", in *INEX*, 2005, pp. 511–524.

[5] A. Hliaoutakis, G. Varelas, E. Voutsakis, E. G. M. Petrakis, and E. Milios, "Information retrieval by semantic similarity*", in Intern. Journal on Semantic Web and Information Systems (IJSWIS).Special Issue of Multimedia Semantics*, 2006, pp. 55–73.

[6] T. Schlieder and M. Holger, "Querying and ranking xml documents", *Journal of the American Society for Information Science and Technology*, vol. 53, pp. 489–503, 2002.

[7] T. Tsikrika, P. Serdyukov, H. Rode, T. Westerveld, R. Aly, D. Hiemstra, and A. P. V. de, "Structured document retrieval, multimedia retrieval, and entity ranking using pf/tijah", *in 6th Initiative on the Evaluation of XML Retrieval*, INEX 2007, ser. Lecture Notes in Computer Science, vol. 4862. London: Springer Verlag, March 2008, pp. 306–320. [Online]. Available: http://doc.utwente.nl/64734/

[8] T. Westerveld, H. Rode, R. O. van, D. Hiemstra, G. Ramirez, V. Mihajlovic, and A. V. de, "Evaluating structured information retrieval and multimedia retrieval using pf/tijah", *in Comparative Evaluation of XML Information Retrieval Systems*, ser. Lecture Notes in Computer Science, N. Fuhr, M. Lalmas, and A. Trotman, Eds., vol. 4518. Berlin, Germany: Springer Verlag, 2007, pp. 104–114. [Online]. Available: http://doc.utwente.nl/64261/

[9] Z. Kong and M. Lalmas, "Xml multimedia retrieval", *in SPIRE*, 2005, pp. 218–223.

[10] M. Torjmen, K. Pinel-Sauvagnat, and M. Boughanem, "Using textual and structural context for searching multimedia elements", *IJBIDM*, vol. 5, no. 4, pp. 323–352, 2010.

[11] H. Awadi and M. Torjmen, "Exploitation des liens pour la recherche d'images dans des documents xml", *CORIA* ,March 2010.

[12] H. Aouadi, M. Torjmen-Khemakhem, and M. B. Jemaa, "Combination of document structure and links for multimedia object retrieval", *Journal of Information Science*, vol. 38, no. 5, pp. 442–458, October 2012.

[13] N. Fuhr, J. Kamps, M. Lalmas, S. Malik, and A. Trotman, "Overview of the inex 2007 ad hoc track", *in INEX, 2007*, pp. 1–23.

[14] A. Popescu, T. Tsikrika, and J. Kludas, "Overview of the wikipedia retrieval task at imageclef 2010", *in CLEF (Notebook Papers/LABs/Workshops)*, 2010.

[15] S. Fakhfakh, M. Tmar, and W. Mahdi, "A new metric for multimedia retrieval in structured documents", *in ICEIS (2)*, 2013, pp. 240–247.