XIRS: an XML-based Image Retrieval System

G. N. FANZOU TCHUISSANG [1], XU DE[2], WANG N.[3]
School of Computer & Information Technology
Beijing Jiaotong University
P.O. Box: 100044 Beijing
P.R. CHINA

FRANÇOIS SIEWE
Team Research Group
De Montfort University
The Gateway
Leicester LE1 9BH
UNITED KINGDOM

Abstract: -This paper presents a formalization of an image retrieval system based on a notion of similarity between images in a multimedia database (namely *XML-Enabled Database*) and where a user request can be an image file or a keyword. The CBIR (Content Based Image Retrieval) system and the current search engines (e.g. Google, Yahoo….) make image search possible only when the query is a keyword. This type of search is limited because keywords are not expressive enough to describe all important characteristics of an image. For example, an exact match request cannot be formulated in such systems. Thus, we propose a search system in which a request might be an image file or a keyword. The MPEG-7 standard is used for describing an image as an XML document. A similarity distance between images is defined which is used to compare the request image with the images of a database. We also propose an algorithm to calculate a similarity distance between two XML nodes with a given precision 'k' (*k is defined by the user: he can fix 'k' at 100% for the exact match retrieval of features*) so as to be able to provide accurate information in response to a user request. The statistics show that our system is more efficient than leading content based image retrieval systems such as ERIC7 and current search engines.

Key-Words: - XML, Image, Multimedia databases, MPEG-7, similarity search

1 Introduction

This paper falls in the field of information retrieval, in particular the search of images in a database when the request is an image or a keyword. The purpose of the search process is to obtain user needed information from a database by comparing the user's requirements with available information in the database. This comparison is carried out by a System of Search for Information (SSI) [2], which is a set of programs with the goal to return to the user the maximum relevant documents available that meet his needs. These needs are translated in a structured way by the user in the form of requests. The concept of relevance being difficult to automate, the goal of the SSI is then to make as accurate as possible the correspondence between the system relevance and the user relevance.

The CBIR (Content Based Image Retrieval) system and the current search engine (CSE) (Google, Yahoo…) make image search possible only when the query is a keyword. This type of search is limited because these keywords are not expressive enough to describe all important characteristics of an image. To resolve this problem, ERIC7 [3] which is a CBIR system compatible with the MPEG-7 [4] Multimedia standard proposed to the user to search images by features. Hence, in ERIC7 the user can choose between 15 features by navigating within XML files using a tool that generates UML diagrams. However, ERIC7 is limited because the user should be an expert in search for images to recognize these features. He should also be able to read and understand XML files and UML diagrams. The search for images present by ERIC7 is then tedious. We also observe that an exact match request cannot be formulated in such systems.

In this work, we propose a search system in which a request might be an image file or a keyword. We describe an image as a XML document using MPEG-7[4] standard. We have defined a similarity distance between images which is used to compare the features of a request image to those of the images stored in a database. We also propose an algorithm to calculate a similarity distance between two XML nodes with a given precision 'k' (*k is defined by the user: he can fix 'k' at 100% for the exact match retrieval of features*) so as to be able to provide accurate information in response to a user request. The statistics show that our system is more efficient than leading content based image retrieval systems such as ERIC7 and the current search engines.

This paper is organized in the following way: Section 2 presents an outline of MPEG-7; Section 3 describes the XIRS system and Section 4 is devoted to the implementation and the discussion. Section 5 concludes the paper and outlines future work.

2 MPEG-7

The ISO's subcommittees SC29, WGll, MPEG (Moving Picture Experts Group), published in February 2002 another standard called "Multimedia Content Description Interface" (in short 'MPEG-7'). The goal of MPEG-7 is to enable fast and effective search and filtering of multimedia content. MPEG-7 is a standardization of XML metadata structures called Descriptors (D) and Description Schemes (DS), which are used to describe and annotate multimedia information [4].

The Ds and DSs are defined using the MPEG-7 Description Definition Language (DDL), which is based on the XML Schema Language. Many technologies still need to be developed around the MPEG-7 for extracting, searching and querying multimedia databases, which involves similarity matching with fuzzy constraints including features, content and semantics [6].

3 XIRS (XML Image Retrieval System)

XIRS is a set of 3 components: the XIRS Mediator, the interrogation module, and the XIRS Server. Starting from the feature extraction and annotation process of a multimedia asset, the XML documents are generated and stored in a repository. One can distinguish two scenarios: *pull* and *push*.

In the *pull* scenario, a user submits queries to the system. In the *push* scenario, the system selects a set of results satisfying the user query constraints (Fig. 1).

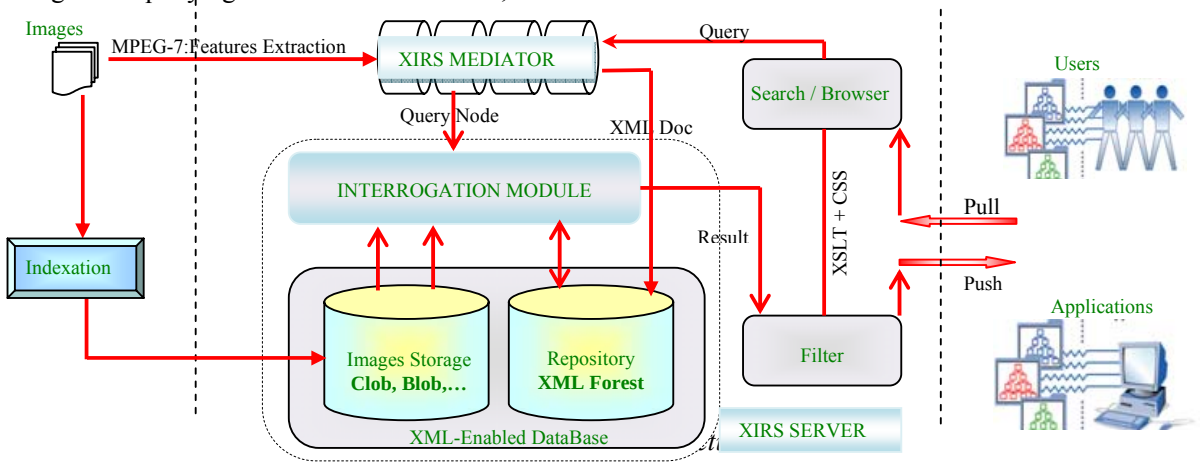

Fig. 2 *XIRS Architecture*

3.1 XIRS Mediator

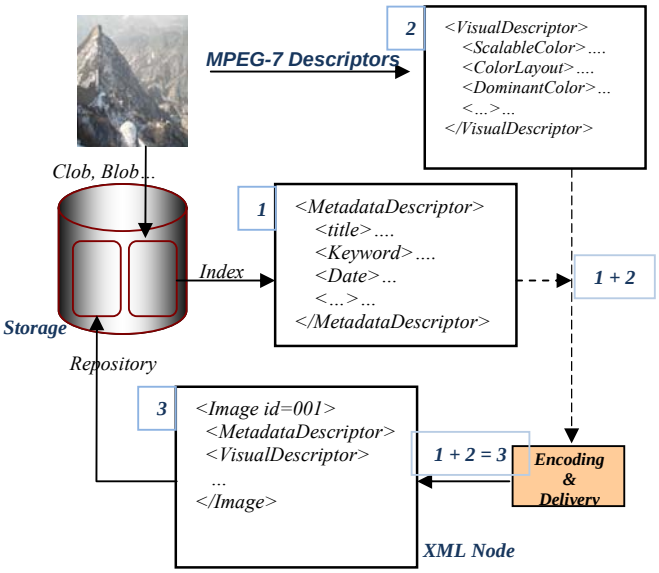

Fig. 2 *XIRS Mediator Scope*

An image is represented as a set of descriptors *(features)* which are structured as XML nodes and stored in a XML document (Fig. 2).

The image and the XML document will then be stored in the Database (*XML - Enabled Database for example*). The XML document used by our system is obtained by combining a part of the MPEG-7 document *(VisualDescriptors)* and some other information coming from the tables of the database where the images are stored *(MetadataDescriptors)*.

In the presence of an image, the XIRS Mediator extracts two description levels which interest us:

- «Visual Descriptors» extracted from the image by MPEG-7,
- «Metadata descriptors»: Our XML document is completed with some information describing the semantic and contents *(free keywords, its author, its size, and its creation date...)*.

We present below the DTD of the XML documents constructed by XIRS Mediator.

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE Images [
<!ELEMENT Image (MetadataDescriptor, VisualDescriptor)>
 <!ATTLIST image id CDATA #REQUIRED>
 <!ELEMENT MetadataDescriptor
(ContentDescriptor,SemanticDescriptor)>
 <!ELEMENT ContentDescriptor
(keyword*,identifier,date,link,size)>
 <!ELEMENT keyword    (#PCDATA)>
 <!ELEMENT identifier    (#PCDATA)>
 <!ELEMENT date    (#PCDATA)>
 <!ELEMENT link    (#PCDATA)>
 <!ELEMENT size    (#PCDATA)>
 <!ELEMENT SemanticDescriptor (title*)>
 <!ELEMENT title    (#PCDATA)>
 <!ELEMENT VisualDescriptor
(DayNight,Orientation,ShotType,IntExt,ScalableColor,ColorLay
out, DominantColor)>
 <!ELEMENT DayNight (#PCDATA)>
 <!ELEMENT Orientation (#PCDATA)>
 <!ELEMENT ShotType (#PCDATA)>
 <!ELEMENT IntExt (#PCDATA)>
 <!ELEMENT ScalableColor (Coefficient*)>
 <!ATTLIST ScalableColor NumberOfCoefficients CDATA
#REQUIRED>
 <!ELEMENT Coefficient (#PCDATA)>
 <!ELEMENT ColorLayout (Ycoeff,CbCoeff,CrCoeff)>
 <!ATTLIST ColorLayout NumOfYCoef CDATA
#REQUIRED>
 <!ELEMENT Ycoeff (YDCCoef,YACCoef)>
 <!ELEMENT YDCCoef (#PCDATA)>
 <!ELEMENT YACCoef (#PCDATA)>
 <!ELEMENT CbCoeff (CbDCCoef,CbACCoef)>
 <!ELEMENT CbDCCoef (#PCDATA)>
 <!ELEMENT CbACCoef (#PCDATA)>
 <!ELEMENT CrCoeff (CrDCCoef,CrACCoef)>
  <!ELEMENT CrDCCoef (#PCDATA)>
  <!ELEMENT CrACCoef (#PCDATA)>
  <!ELEMENT DominantColor
(ColorSpaceType,SpatialCoherency, Percentage,
ColorValueIndex, ColorVariance)>
  <!ATTLIST DominantColor size CDATA #REQUIRED>
 <!ELEMENT ColorSpaceType (#PCDATA)>
 <!ELEMENT SpatialCoherency (#PCDATA)>
 <!ELEMENT Percentage (#PCDATA)>
 <!ELEMENT ColorValueIndex (#PCDATA)>
 <!ELEMENT ColorVariance (#PCDATA)>
]>
```

Example of XML Document

For an example, let us concentrate on the *image of the gate of Beijing Jiaotong University.*

```
<?xml version="1.0" encoding="UTF-8"?>
<Images>
 <image  id=001>
  <MetadataDescriptor>
    <ContentDescriptor>
       <Identifier>bjtu001</Identifier>
       <Keyword> Jiaotong University </Keyword>
       <Keyword>north Gate </Keyword>
       <link>...\bjtu001.jpg </link>
       <size>6k </size>
       <date>02/06/2007</date>
     <ContentDescriptor>
     <SemanticDescriptor >
        <title>Beijing Jiaotong University north  gates</title>
     </SemanticDescriptor >
  …
    </MetadataDescription>
   <VisualDescriptor>
     <DayNight>day</DayNight>
     <Orientation>vertical</Orientation>
     <ShotType>general</ShotType>
     <IntExt>out</IntExt>
     <ScalableColor numberOfCoefficients="63">
          <Coefficients> 1 1 0 1 1 0 0 1 1 1 1 0 0 0 0 1 1 1 0 0 0 0
        0 1 1 1 0 1 1 0 0 1 0 0 <Coefficients>
     </ScalableColor>
    <ColorLayout numOfYCoef="64">
     <Ycoeff>
      <YDCCoef>13</YDCCoeff>
      <YACCoef> 27 23 2 16 10 14 16 9 9 17 14 13 16 16 16 16
       14 15 17 17 16 16 17 16 14 15 </YACCoef>
     </Ycoeff>
    …
    </ColorLayout>
    <DominantColor size="8">
    <ColorSpace type="RGB"/>
    <SpatialCoherency>0.3722258333336</SpatialCoherency>
    <Values>
     <Percentage>0.0838</Percentage>
     <ColorVariance>23161.6189638.56.291</ColorVariance>
    </Values>
   </DominantColor>
  </VisualDescriptor>
 </image> /* End of first image: Image of id '001' */
</Images>
```

The Color Layout and the Dominant color are low level colors. *The IntExt* indicates if the image were taken outside, in the nature, or inside; *The DayNight* indicates if the image were taken during the day or during the night; *The ShotType* characterizes the framing of the characters of the image and *The Orientation* are high level Colors. For each one of these features, a similarity distance is defined which makes it possible to measure the similarity of two images.

Once XIRS Mediator described an image in XML node, the node is categorized (*to prevent too bulky XML documents*) and stored in XML documents of the collection. The role of XIRS Mediator is thus to define an image in XML and vice versa. The reverse way is done easily by using the node: <link>...\bjtu001.jpg</link>

3.2 Interrogation Module

The data model of the XIRS interrogation module is a simplification of XPath data model presented in [1], where a structured document is a tree, composed of simple nodes, sheet nodes and attributes. *A node can be a document, an element, a text, a namespace, an instruction or a comment.* Two cases of request arise.

3.2.1 The request is a keyword

A request is a conjunction of sub-requests. We have the following illustration:

Query → sub-request AND sub-request | sub-request OR sub-request | NOT sub-request.

The Boolean model introduced in [2] defined the similarity between an image I and a request Q as:

$$d(Q,I)=\begin{cases}1 \text{ if } I \in \text{the set described by the request Q}\\ 0 \text{ otherwise}\end{cases}$$

3.2.2 The request is an image

The comparison between an image and a request amounts calculating a score. The image relevance with respect to the request is calculated by a similarity function noted d*(Q, I)*, where Q is the request image and I is an image of the Database. It thus leads to calculate a similarity distance between two XML nodes. We will used the following notations : $I = (I_1, I_2, \ldots, I_m)$ for an Image set and $T = (t_1, t_2, \ldots, t_n)$ for a keyword set. We describe the image I_j as a vector : $\vec{I}_j = (w_{1,i}, w_{2,i}, \ldots, w_{j,i}, \ldots, w_{n,i})$ *where* $w_{i,j} \in \{0, 1\}$ *is the term-weighting.* Let f_i denote the function that returns the associated weight of the term t_i : $f_i(\vec{I}_j) = w_{i,j}$.

The XML node produced by the XIRS Mediator and corresponding to the request image is regarded as a block of requests (*like a system of equation with several unknown factors*), in which each sub-node *(features)* is seen as a request. It is thus a question of reassuring when one has a node coming from a XML document of the Database that both sub-nodes are similar.

If a feature of an image is indexed by t_j and if $t_j < t_k$ then it is also indexed by t_k. Therefore, one can extend the vector I_i so that: $\forall j,k \in [1;n]$, $w_{k,i} = 1$ if $w_{j,i} = 1$ and $t_j < t_k$, otherwise $w_{k,i} = 0$. The usual similarity measure in the vectorial model [2] is the cosinus.

$$\cos(I_k; I_l) = \frac{\vec{I}_k * \vec{I}_l}{\left|\vec{I}_k\right| * \left|\vec{I}_l\right|} = \frac{\sum_{j=1}^{n} w_{j,k} \times w_{j,l}}{\sqrt{\sum_{j=1}^{n} w^2_{j,k}} \times \sqrt{\sum_{j=1}^{n} w^2_{j,l}}}$$

Hence, $\mathbf{d(I_k; I_l) = (1 - \cos(I_k; I_l))}$ is a similarity distance between two images I_k and I_l.

The following grammar gives a complete description of the request language used. The axiom of the grammar is **Query**, Non-terminal symbols are in **bold**, *terminal* symbols (Tokens) are in *italic* and the production rules are described below:

Query → **r1** | **r2**
r1 → **ExpressionA ExpressionB**
ExpressionA → *keyword* **SuiteExpressionA** | (*keyword*) **SuiteExpressionA**
SuiteExpressionA → **ExpressionA** | ε
ExpressionB → **BooleenOperator r1** | ε
BooleenOperator → *O*R | *AND* | *NOT* | ε

r2 → **ExpressionStructure SuiteExpressionStructure**
ExpressionStructure → *elementName[* **Condition** *]*
Condition → *@attributName = keyword* | **r1** | ε
SuiteExpressionStructure → B**ooleenOperator** E**xpressionStructure** | ε

Caption:
ε denotes an empty string
keyword: terminal symbols representing a keyword
elementName: terminal symbols representing a name of tag
attributName: terminal symbols representing a name of attribute

3.3 XIRS Zone Server

3.3.1 XIRS principle: Search for images by similarity

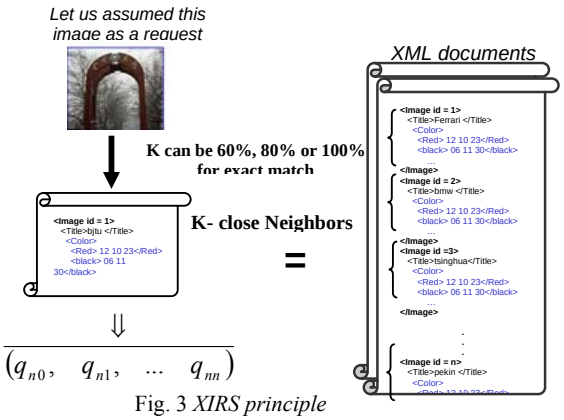

Fig. 3 *XIRS principle*

The image request is a node; it is a question of returning all the nodes of the XML documents of the collection which are similar to the request node according to a precision ''k''. A *similarity distance* 'd' between two nodes is defined by:

$$d : N \to D$$

$$\begin{pmatrix} s_{n0} \\ s_{n1} \\ \ldots \\ s_{nn} \end{pmatrix} \mapsto d \begin{pmatrix} s_{n0} \\ s_{n1} \\ \ldots \\ s_{nn} \end{pmatrix} \qquad d \begin{pmatrix} s_{n0} \\ s_{n1} \\ \ldots \\ s_{nn} \end{pmatrix} = \begin{pmatrix} w_{0q} \\ w_{1q} \\ \ldots \\ w_{nq} \end{pmatrix}$$

Where s_{n0}, s_{n1}, … , s_{nn} are variables *(sub-nodes representing the features of the image)* coming from XML documents of our Database, N is a set of Nodes and D is a set of distances. The description of the image request (Fig.3) being a XML node, q_{n0}, q_{n1}, … , q_{nn} are fixed and are query sub-nodes; w_{0q}, w_{1q}, … , w_{nq} are weight (*similarity distance between features*) associated to the sub-node s_{nl} compared to the request q_{nl} with $l \in [0,n]$, 'l' is the number of sub-node of a node.

3.2.3 Construction of S: set of results

Definition 1: Two XML nodes are k-similar if 'k' percent of their sub-nodes (features) are identical.

Definition 2: A node belongs to S if and only if this node is K-similar to the node describes by the request image, *ie: if* $\mathbf{AVG}(w_{0q}, w_{1q}, \ldots, w_{nq}) \geq k$.

- **Construction of S**

For Each node Sn **of** an XML document of the Database

If $\frac{w_{0q} + w_{1q} + \ldots + w_{nq}}{n+1} \geq k$ **Then**

$S \leftarrow S + s_n$

Else

$S \leftarrow S + \{\ \}$;

take another node

EndIf

End For

w_{iq} *is similarity distance between features*

- **Calculation of** w_{iq}

If $tag(ss_{ni}, qs_{ni}) = true$ **then** Keyword= content-$q_{ni}^{(1)}$

/ it's currently necessary to calculate the various weights between the contents of the tags of* ss_n *and that of* qs_n **/*

if $d_B(ss_n, qs_n)=1$ **then** $w_{iq}=100$;

else

$$w_{iq} = \frac{\sqrt{\sum_{j=1}^{n} w^2_{j,k}} \times \sqrt{\sum_{j=1}^{n} w^2_{j,l}} - \sum_{j=1}^{n} w_{j,k} \times w_{j,l}}{\sqrt{\sum_{j=1}^{n} w^2_{j,k}} \times \sqrt{\sum_{j=1}^{n} w^2_{j,l}}} \times 100$$

else *take another sub node*

Here ss_n is a sub-node *(sub feature)* and qs_n is a query sub-node, $d_B(ss_n, qs_n)$ is *the Boolean model distance.*

(1) fixes the contents of q_{ni} as a keyword i.e. $<q_{ni}>$ content- q_{ni} $</q_{ni}>$

tag(a, b) is a function which returns true when his arguments have a similarity content of tags(*according to the precision k*)

Example: *if k=80%*

tag(<name> fanzous </name>, <names> fanzoug </names>) = = true

d(a, b) is the function which returns the percentage of similarity between the data of a sub-node

Example:

d(<name> fanzous </name>, <names> fanzoug </names>) = = 90%

4 Implementation and discussion

We have used PHP 5.0 to build an interrogation interface (Fig.4). The Database Oracle 8i was used for storage of the images and the XML documents. We used MPEG-7 library to implement XIRS Mediator.

We have created a Database of 1 500 Images. After categorization of images, we obtained 15 XML documents in our collection. The evaluation was conducted on a computer having: 1.6 GHz of processor, 80 Go of hard disk and 512 Mo of RAM. The operating system was Windows XP SP2.

Fig.4 *XIRS interface*

To validate our system, we measured *the precision of retrieval (percentage of similarity between the query and the result (PR))***.** We believe that a better and more accurate measure could be achieved by using this metric.

- **The precision of retrieval (PR)**

To choose an appropriate set of queries for the evaluation, we considered the types of queries used in basic processing operations of search:

- ✓ **Exact match search** (Fig.5 a): when the value of k is equal to 100%, XIRS returns only the XML nodes identical to the XML node of the request image and thus the returning images are the one identical to the image request. In 100 images return by ERIC7, 40 are totally different to the request image depending by the features given by the user. In CSE, 70% of returned images are not similar.
- ✓ **Full text search** (Fig.5 b): the PR of ERIC7 is 88.4% while that of XIRS is 88.3%, due to the database clustering done by ERIC7.
- ✓ **Semantic search** (Fig.5 c): XIRS is about 35 % more efficient than ERIC7, due of the semantic descriptors insert in the XML Nodes by XIRS Mediator.

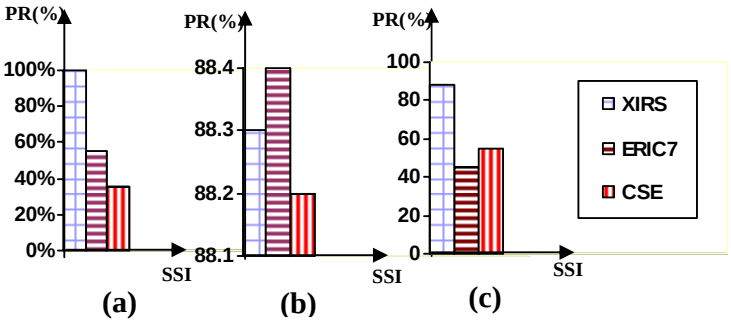

Fig.5 *Experiment results*

- **Applications**

Posting an image for the similarity search in a Database can have an importance in Hospitals to find the diagnosis of the radiographic stereotypes. It can also be used to implement iconic communication systems such as those described in [5].

5 Conclusion

In this paper, we have defined a search system for images when the request is an image or a keyword. The user has the possibility to formulate his requirements in information using a given precision K. The similarity between two images is defined by the similarity between two XML nodes representative the two images. An evaluation of XIRS shows the effectiveness of this system towards the CBIR systems and the Current Search Engines (CSE) (Google, Yahoo…) as for the search for images.

The reformulation of the requests, the consideration of several images like the request (*iconic sentences for example*) and the consideration of the heterogeneous sources of images constitute prospects for the continuation of this work.

References:

[1] S. Boag and al. (Eds), *XQuery 1.0 : An XML Query Language,* W3C Working draft, 2003.

[2] G. N. Fanzou T., *XIRL : XML Information Retrieval* , Mémoire de DEA, University of Yaounde 1, Cameroon, 2006.

[3] L. Gagnon, S. Foucher, V. Gouaillier, *ERIC7: An Experimental Tool for Content-Based Image Encoding and Retrieval under the MPEG-7 Standard*, R&D Department, Computer Research Institute of Montreal, 2004.

[4] H. Kosch, *MPEG-7 and Multimedia Database Systems,* SIGMOD Record, Vol. 31, No. 2, June 2002**.**

[5] N. C. Kuicheu, P. L. Fotso, F. Siewe. *Iconic Communication System by XML Language (SCILX)*. In the proceedings of *the 2007 ACM International Cross-Disciplinary Conference on Web Accessibility*, Banff, Canada, 7-8 May 2007.

[6] J. M. Martínez*, MPEG-7 Overview (version 10)*, Palma de Mallorca, 2004.