A contribution to multimedia document modeling and organizing

Ikram Amous, Anis Jedidi, Florence Sèdes

A Contribution to Multimedia Document Modeling and Organizing

Ikram Amous, Anis Jedidi, and Florence Sèdes

Abstract. This paper presents a solution to resolve the problem of multimedia documents collection reorganizing. This solution is based on a documentary warehouse enriched by metadata (for each media type) elicited, modeled and structured in XML meta-documents. To homogenize these meta-document representation, we based our annotation on a document indexing and segmentation process.
The warehouse thus created is seen as the hyperbase to which the user will apply personalization and querying mechanisms. The personalization enables dynamic re-structuring and re-construction of documents answering to the user queries. This approach is based on the OOHDM methodology extension with the use of the metadata.

1 Introduction

Building personalized web application is a challenging task. For that, we use a documentary warehouse to propose a local structure to documents in order to be able to reorganize them dynamically.

The approach we propose relies a modeling extension based on a set of textual and media-based metadata (specific each media type). The metadata extraction is carried out by specific functions for each media. In our laboratory (IRIT), libraries are developed and are dedicated to each media [9][10]. Once the metadata extracted, they are structured in XML documents called "meta-document", employed to enrich the documentary warehouse thus created [4].

To create dynamic documents, we propose to use the hypermedia design application methodology OOHDM (Oriented Object Hypermedia Methodology Design) [15]. So, we propose to extend this methodology by the introduction of metadata into documents. To dynamically re-organise the documents, we use the abstract view (built from the diagrams of context navigation) defined using ADVs (Abstract Data View).

In this paper, we will present the metadata modeling already detailed in [3] and the use of OOHDM methodology in order to create dynamic views of documents answering to the user's needs expressed through queries. In section 2, we present our motivation. In section 3, we present metadata likely to be extracted and modeled. In section 4, we present our proposition which consists in extending OOHDM by introducing metadata and we conclude in section 5.

.

2 Motivations

The lack of knowledge on multimedia document structure prevents from easily locating and re-covering them [1]. These problems led to several works proposing metadata to improve document consultation. Among these works [7] presents a metadata classification and an identification for the multimedia documents. [2] proposes an oriented object model for spatio-temporal representation of multimedia documents. Other approaches such as MPEG7 [14] aim at offering a complete set of tools for audio-visual description, creating descriptions on the documents and giving access to the contents. All these works propose useful sets of metadata for document annotation but do not take into account the legacy databases nor elicited and modeled non-textual metadata (instantiated by an audio or image value). To cope with these insufficiencies, we propose on the one hand, the introduction and modeling of non-textual metadata [3] on addition to the existing textual metadata.

On the other hand, we propose to create dynamic view of documents. Several works propose methods to dynamically restructure documents. [5] proposes the Araneus System which extracts information and integrates them into views of databases. [8] proposes the WebML language implemented to design complex web sites. Its proposals include HDM, RMM, OOHDM [15] and Araneus.

All these works propose useful concepts to the document collection design but does not consider the elicited metadata. So, we propose to integrate into the generated dynamic view of documents, metadata elicited in the first part of this paper.

3 Metadata

The need of powerful solution for quickly and efficiently identifying various types of audiovisual content of interest to the user, using also non text-based technologies, directly follows from the urge to efficiently use the available multimedia content and the difficulty of doing so. To this end, MPEG initiated a new work item, formally called “Multimedia Content Description Interface”, known as MPEG-7 [14]. MPEG-7 specifies a standard to describe various media types information, including still pictures, video, speech, etc. These descriptions enable the retrieval and the filtering of the next audio-visual data with a predefined structure, but do not provide any representation of "legacy databases".

In order to integrate these applications, we propose a set of elicited metadata in meta-document. Extracting metadata is based on a document indexing and segmentation process elaborated medium by medium. In this section, we present briefly the metadata modeling for each media type. The detail of each one is present in [3].

3.1 Text Media

We propose the annotation of this media by a set of metadata extracted by different tools like Exrep (Extended Regular Expression Processor) [9] (IRIT). These metadata are structured in a meta-document having the root tag "text_file", enriched by the generic attributes "name", "language", "size" and "type".

.

Each text file is composed of documentary granules (indexed by key word) identified by a structure recognition process. The documentary granule can be chapter, section, paragraph, … We propose to identify by "Text_Unit" (TU) any documentary granule chosen by the user. All these metadata are structured in a meta-document modeled as follows:

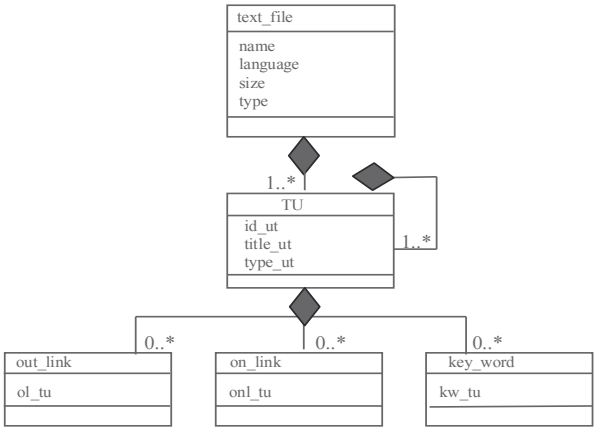

Fig. 1. Text model

3.2 Audio Media

The representation and the description of the audio documents require procedures and techniques developed within the framework of several researches in segmentation, indexing and extraction :

Indexing processes using for example speech recognition enable to elaborate indexing motives (recognized key words, orthography...).

Segmentation processes based on a time-localized abrupt changes in significative parameters of sound have to be detected and classified as diagnostic cues for understanding change in the content of the sonic flow (word, music, noise) [12].

Extraction signal tools allow the annotation of various sound by amplitude, width, speed... like the Transcriber tool [6].

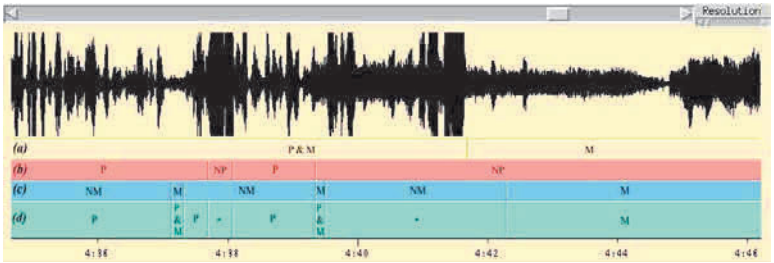

Fig. 2. The Transcriber segmentation

These process results are integrated to annotate the audio documents by metadata. They are structured in a meta-document having the root tag "audio_file" modeled as follows:

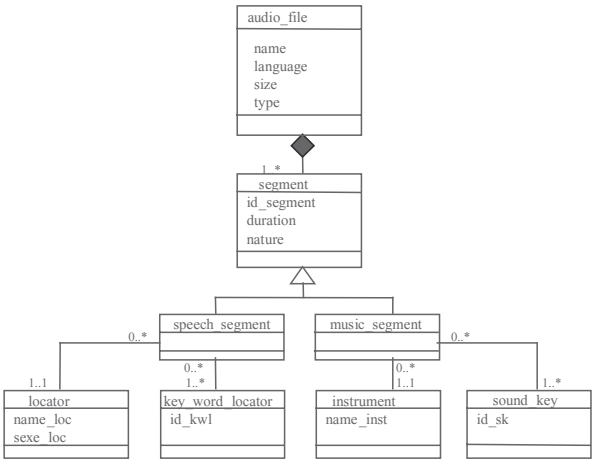

Fig. 3. Audio model

3.3 Image Media

For the image media, we propose textual and graphical metadata, extracted by means of image analysis and recognition techniques [10] (IRIT). The segmentation/generation process is similar to that of digital documents in which one segment characterizes a chunk of text, an image, a page, ... and for which the Optical Character Recognition (OCR) enables to develop motives for indexing segments. The image segmentation process attempts to identify regions to which are associated features such as color, texture [11]. The process results consist in segmenting an image into several regions to annotate the image media by metadata. All these metadata are structured in a meta-document modeled as follows:

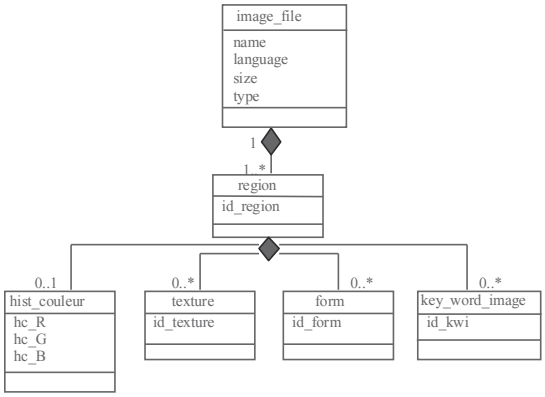

Fig. 4. Image model

3.4 Video Media

Video is composed of audio-visual information. Providing content based access to video data is essential for the successful integration of video into computers. Organizing video for content based access requires the use of video metadata. All these metadata are identified by various segmentation and indexing techniques [13]. The indexing results are applied to annotate video media in a meta-document having the root tag "video_file". All these metadata are structured in a meta-document modeled as follows:

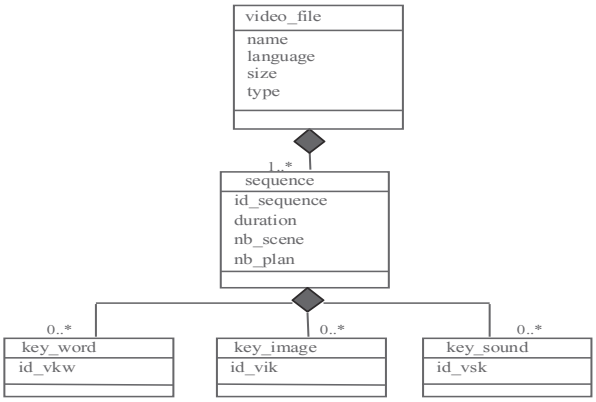

Fig. 5. Video model

3.5 The Basic Meta-model

Our first objective is the annotation of different media documents, by an homogenate structure. For that, we propose a meta-model designed to fit to the different meta-documents of all media types. The meta-model is as follows:

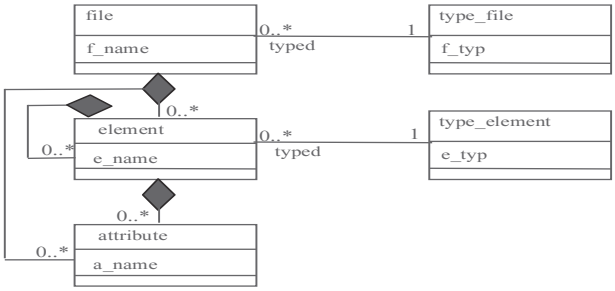

Fig. 6. The meta-model

Our second objective is to restructure the documents according to the user's needs. We propose to extend the OOHDM concepts and to combine them with the previous approach of metadata modeling.

4 Dynamic Views of Documents with OOHDM

OOHDM [15] proposes concepts and mechanisms to structure the access to a web sites. It is a model-based approach for building hypermedia applications. It consists of four different steps namely conceptual design, navigational design, abstract interface design and implementation.

We present then our proposition which consists in extending OOHDM by introducing metadata in all its steps and also in the document contents. We illustrate this methodology on an application example about the laboratory activities.

4.1 Conceptual Design

During the conceptual design, a model of the application domain is built, using object-oriented modeling principles. For each domain class, we add a class meta [4] containing the extracted metadata and including all the "meta-document" of each data type. Each class meta is composed of zero or more files (files containing metadata of an image media, files containing metadata of a video media, etc.). The conceptual model of our application is as follows:

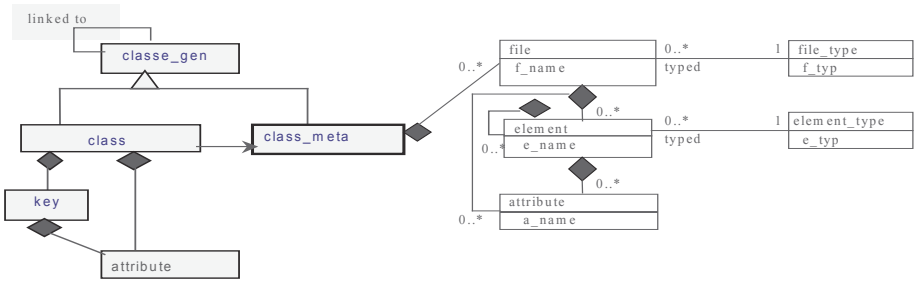

Fig. 7. The conceptual meta-model

4.2 Navigational Design

The navigational design is expressed in two diagrams: the navigational classes diagram and the navigational contexts diagram.

4.2.1 Navigational Classes Diagram

The navigable objects of an hypermedia application are defined by a navigational diagram classes which reflect a selected view around the application domain. The predefined types of navigational classes are organized into navigational contexts.

We represent in this diagram only the important elements and also the ‘Anchor’ attributes representing links between classes.

4.2.2 Navigational Contexts Diagram

In OOHDM the main structuring primitive of a navigational schema is the notion of navigational context. This latter is a set of nodes, links, context classes and other (nested) navigational contexts induced from navigational classes. An index or a guided tour can define a navigational context. The principal point to build these

contexts is the node which introduces the ‘Anchor’ attribute, whose behavior definition defines the navigation semantics [15]. We propose to extend the OOHDM methodology by introducing metadata in the conceptual schema and also in the navigational context. Thus, navigation (context) can be activated by metadata.

4.3 Abstract Interface Design

Once the navigational structure has been defined, it must be made perceptible to the user through the application's interface, done by defining an abstract interface model. In OOHDM we use the ADV design approach for describing the user interface of an hypermedia application. In our approach, the generated metadata will be taken into account in the ADVs and can be even visualized in the new dynamically created documents [4].

4.4 Experiments

To explain our generic approach, we use an example of a documentary warehouse representing publication of our team. To answer the query "display the publications interested to ‘OOHDM’ by key word" with XML and by a dynamic way, we propose to use the view concept. Each publication answering the query can be represented in one or two view(s) enriched by metadata. To navigate the publications or views composing the publications, we create links by means of Xlink [16] and Xpointer [17].

Restructuring publications is based on two principles:

Display the selected data and metadata of each publication answering to the query [4]. Each publication can be represented in one or two views. In the first case, the view contains the data and metadata selected by the user. In the last case, the first view contains the selected data whereas the second view contains the selected metadata.

Display for each publication, only the documentary units answering to the query (‘OOHDM’ in the example) and their metadata. Each publication can be presented in one or two views. In the first case, the view contains the documentary units answering to the user query and their metadata. In the last case, the first view contains the documentary units whereas the second view contains the metadata of the documentary units displayed in the first view.

We illustrate these query types by examples elaborated by OWS (Oracle Web Server).

4.4.1 First Principle

To have an answer to the query presented above, we propose to display the data (title, subject, type, etc.) and metadata (key word, inlinks, onlinks, image forms, etc.) of each publication answering to the user queries (’OOHDM’ in the example).

Fig. 8 shows a Web site view presented with the WebML language [8]. The hypertext is composed of four pages. Each page presents a set of information to be displayed.

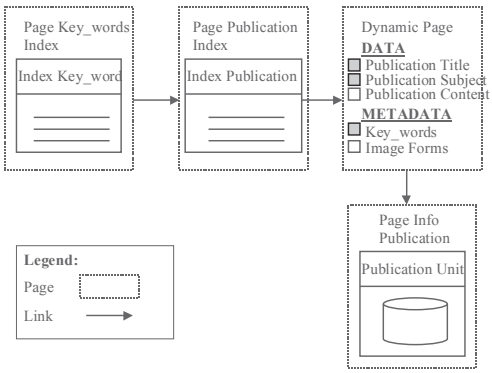

Fig. 8. A Web site view with the graphical WebML notation

The page "Page Key_words Index" presents an index on the key_words used in all publications. The page "Page Publication" presents a set of Publications using the key word selected in the page "Page Key_words Index". To reconstitute each publication, we use a "Dynamic Page". This last allows to the user to select all information what he want to display.

Example: if the user select to display in a view the publication data like title and subject and the publication metadata like key_words, the view is then presented as follows:

```
<?xml version="1.0"?>
<VIEW id="view1">
     <DATA>
            <title> title 1 </title>
            <subject> subject 1 </subject>
     </DATA>
     <METADATA>
            <key_word kw_tu_1= "word1"  kw_tu_2= "word2" />
     </METADATA>
     <A xml:link="simple"
     href="http://www.irit.fr/proto.affich_pub?numpub=2"> next publication </A>
</VIEW>
```

Gathering data and metadata, the content of each views of documents is different from one to another, according to the user needs expressed into his queries and his selection.

We can use in this case the Xlink links to browse the documents created dynamically.

Let us see an example of a publication seen in one view:

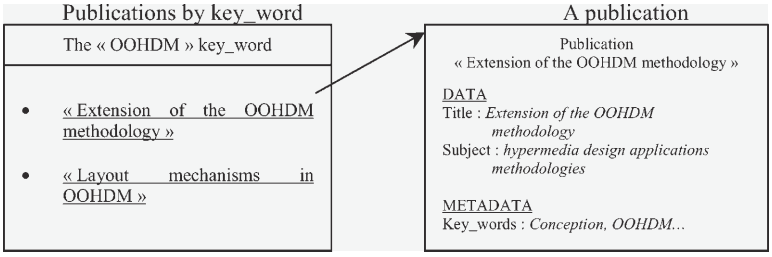

Fig. 9. A publication viewed in one view

4.4.2 Second Principle

We propose an other example to restructure the publications answering the same query. In each publication, we only present the documentary units relevant for the user's query ('OOHDM' in the example) and their metadata. If we represent each publication in two views for example, we represent the documentary units in a view and the metadata in another view.

Example: if we suppose that one of the publications answering to the user needs contains two documentary units interested to 'OOHDM' (the first one is text while the second is an image), this publication can be presented in two views as follows:

The first view representing the documentary units of the publication is presented as follows:

```
<?xml version="1.0"?>
<VIEW id="View1_1">
     <DATA>
          <DU id="1"> ud1
          <xlink:simple
            href='http://www.irit.fr/proto.affich_d_pub?numpub=1#xpointer
          (VIEW/DATA/DU[position()&lt;=2])'/>
          </DU>                    //the first DU which type is text
          <DU id="2"> ud3
          <xlink:simple
          href='http://www.irit.fr/proto.affich_d?numpub=1
            #xpointer(VIEW/DATA/DU[position()&lt;=1])'/>
          </DU>         //the second documentary unit which type is image
     </DATA>
     <A xml:link="simple"
     href="http://www.irit.fr/proto.affich_md?numpub=1"> next page </A>
</VIEW>
```

In this first view, we add Xpointer links to navigate the documentary units.

The second view representing the metadata of the documentary units (DU) is presented as follows:

```
<?xml version="1.0"?>
<VIEW id="View1_2">
     <METADATA>
          <MD-DU id_du="1">    //metadata of the first DU which is text
            <key_word  kw_tu_1= "word1"  kw_tu_2= "word2"/>…
          </MD-DU>
          <MD-DU id_du="2"> //metadata of the second DU unit which is image
            <form  id_form_1= "form1"  id_form_2= "form2"/> …
          </MD-DU>
     </METADATA>
     <A xml:link="simple"
     href="http://www.irit.fr/proto.affich_d?numpub=1"> previous page </A>
</VIEW>
```

Let us see an example of a publication which can be seen in two views (cf. Fig.10)

The difference between this second case and the first one consists in the user choice. The user can select in this second case to display the documentary units interested to 'OOHDM' and not the publication data (like title, subject, etc.). The content of the views in the two cases is completely different.

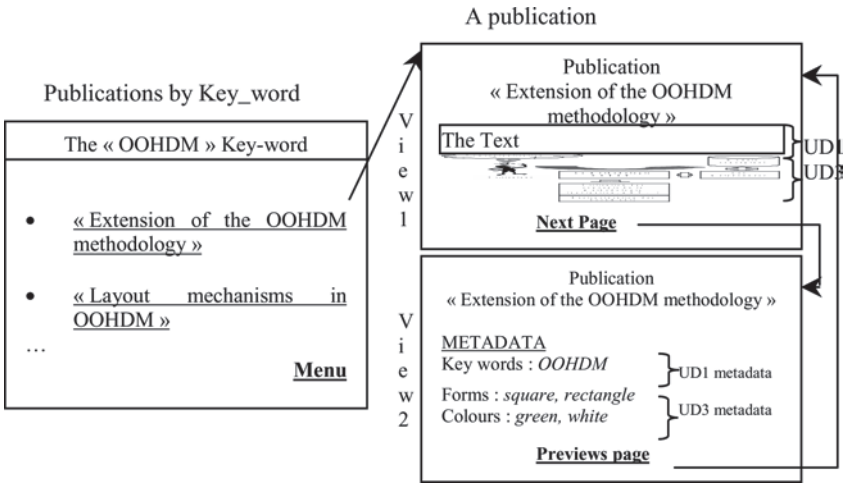

Fig. 10. A publication viewed in two views

5 Conclusion

The semi-structured data context led to introduce the metadata concept, concept that we propose to model and to add in our model to improve search process. The various metadata families process is based on the indexing process and is carried out by extraction (elicitation, synthesis) from the multimedia contents. It can be supplemented by a more or less assisted manual process. The metadata modeling presented in the section 2, is more detailed in [3]. We use this modelisation to enrich the OOHDM methodology in order to build dynamic documents. They are created by the use of views containing the data and metadata answering to the user's requests. Our approach is generic and can be applied on all application type, but to clarify our method, we have chosen a given example.

We propose in this paper with comparison to the paper [4] to use metadata modeling and not only the elicited metadata to build dynamic documents. We propose also to create one or tow views according to the user's needs. The created view contains not only the selected data and metadata but also the documentary units answering the query and respective metadata.

The future prospects of our work concern on the one hand the improvement of metadata modeling by a lower granularity level for all media type. We focus here to the scalable metadata. They can be represented at various levels according to the evolution of the document, the user's needs, the indexing tools. So it is necessary to define and create the scalable operators for a better metadata modeling. For video medium, for example, we can considerate the spatial and temporal features as "inter-media metadata" integrating image and audio media.

On the other hand, we can propose to extend query languages by operator applied for multimedia document querying.

References

1. Abiteboul, S.: Querying semi-structured Data, ICDT'97, Invited Talk, LNCS n°1186, 1997.
2. Adiba, M., Zechinelli-Martini, J.L.: Spatio-temporal Multimedia Presentations as Database Objects, Lecture Notes in Computer System, Vol. 1677, 1999, pp. 974-985.
3. Amous, I., Jedidi, A., Sèdes. F.: Organising and Modeling Metadata for media-based documents, 4th International Conference on Entreprise Information Systems ICEIS 2002, pp. 18-25, 04 - 06 avril 2002, Cuidad Real/Espagne.
4. Amous, I., Chrisment, C., Sèdes. F.: Reengineering the Web sites by using metadata and a methodological approach, 3rd International Conference on Information Integration and Web-based Applications & Services IIWAS 2001, pp. 127-138, 10-12 September 2001, Linz/Autriche.
5. Atzeni, P., Mecca, G., Merialdo, P.: Design and Maintenance of Data-Intensive Web Sites, Proc. EDBT 1998, pp. 436-450.
6. Barras, C., Geoffrois, E., Wu, Z., Liberman, M.: Transcriber: a Free Tool for Segmenting, Labeling and Transcribing Speech, 1st International Conference on Language Resources and Evaluation (LREC), pp. 1373-1376, May 1998.
7. Böhm, K., Rakow, T.C.: Metadata for multimedia documents, SIGMOD-RECORD, Vol. 23, n°4, pp. 21-26, décembre, 1994.
8. Ceri, S., Fraternali, P., Bongio, A.: Web Modeling Language (WebML): a modeling language for designing Web sites, 9th International World Wide Web Conference The Web, Amsterdam, May 15-19, 2000.
9. Lambolez, P.Y., Queille, J.P., Voidrot, J.F., Chrisment, C.: EXREP: a generic rewriting tool for textual information extraction, Ingéniérie des Systèmes d'Information, Vol. 3, n°4, pp. 471-485, 1995.
10. Desachy, J.: Image processing, signal processing, and synthetic aperture radar for remote sensing, SPIE proceedings series, Vol. 3217, September, 1997, London, UK.
11. Gong, Y.: Advancing content-based image retrieval by exploiting image color and regions features, Multimedia Systems, Vol. 7, n°6, pp. 449-457, November 1999.
12. Gauvain, J.L., Lamel, L., Adda, G.: Transcribing Broadcast News for Audio and Video Indexing, Communication of the ACM, Vol. 43, n°2, pp. 64–67, February 2000.
13. Lienhart, R., Effelsberg, W.: Automatic text segmentation and text recognition for video indexing, Multimedia Systems, Vol. 8, n°1, pp. 69-81, January 2000.
14. Mpeg-7.com: a gateway into the world of content management and the interface between industry and the MPEG community. http://www.mpeg-7.com.
15. Schwabe, D., Rossi, G., Barbosa, S. DJ.: Abstraction, composition and layout definition mechanism in OOHDM, 1995. http://www.cs.tufts.edu/~isabel/schwabe/MainPage.html
16. W3C: XML Linking Language (XLink) Version 1.0, 2000. http://www.w3.org/TR/xlink/
17. W3C: XML Pointer Language (XPointer) Version 1.0, 2001. http://www.w3.org/TR/xptr