Audiovisual Document Modeling

By Metadata

Hela Elleuch
Ameni Yengui
Mahmoud Neji
ABSTRACT:
Information query in multimedia resources is based on processes facilitating access to information regardless of their heterogeneity. These resources are the subject of pedagogic information processing by video conferencing which constitutes an interesting area in our daily lives.

Some work has studied the generic modeling of multimedia documents by metadata. Other studies have relied on models that aim to bring the total composition of each document. Thus, many works were based on the annotation of documents.

Modelling of multimedia documents was a focus for several authors in the literature. In this paper, we study a modelling tool for video conferences in medicine that decomposes them into different media such as text, image and audio. For each medium, we define the necessary metadata so that a lay person who inquires about details of a medical video conference can have the information needed through our tool.

**Keywords:** Modelling, Multimedia Document, Metadata, Video Conferencing, Media.

INTRODUCTION
A video conference is a form of teleconference, i.e. a conference that uses telecommunication networks. Therefore, it can be considered as a support to "mediated" communication, that is to say, communication operated by telecommunication". The term "mediated" means the transmission of communication by means of a medium [01]. In this paper, we will deal with the modelling of medical video conferencing.

The investigation of multimedia resources, our field of study, is gaining interest given the importance of images, text and sound in multiple applications. These resources to be treated are heterogeneous in terms of semantics, structure (structured, semi-structured or unstructured), content (text, audio, image, video) and formats (RTF, XML, MP3, SMIL, MPEG-7).

Given this heterogeneity, the user is unable to control all the tools of pattern recognition, indexing and image processing. He should seek those needs using the form of documents and not the exact criteria of their contents. The description of these requirements will be more "semantic" (relatively speaking) than formal. It considers the document as a whole and not as an aggregate media. Hence emerges the need to propose new models of representation and description that homogenize the structure of these media and their content. These contents are used as a support to query mechanisms that are better adapted to the user’s abilities.

We will decompose a video conference according to types of media, and for each type, we will define the necessary metadata. Metadata is used to define all technical and descriptive information added to the documents for better qualifications. This is to describe or define a given data using a different data to the document since metadata is information that makes up the document.

Our research problem is summarized by the following three questions:
- How to standardize the descriptive structures of documents issuing from different media?
- What are the "metadata" that help define the content and semantics of each type of medium?
- How can this metadata help them to query a multimedia document without reference to its neither actual content nor return of empty answers?

I. STATE OF THE ART
Information today comes in many forms. Language, encoding, and the type of information are becoming increasingly heterogeneous. The textual information, regardless of the language of writing, is certainly the most common. But with the rise of Internet and multimedia tools, the language of writing is not the only conveyer of knowledge.

Modelling of multimedia documents was the centre of interest of several authors in the literature. Some works have studied the generic modelling of multimedia documents by metadata. They propose an approach of the homogenization of representation structures of documents to facilitate their final treatment without the use of the content of multimedia themselves. They proposed the structuring of the metadata in XML documents called "meta-documents". They represent an additional structure in relation to any logical or physical structures written by the authors of the documents. They extended the meta-materials by integrating semantic descriptors defined by the user’s needs and based on
spatial and temporal relations in the query of multimedia documents. In his works, Jdidi proposed a tool for the graphical formulation of queries, X Query, using metadata by integrating the spatio-temporal relations between these metadata [02].

To present such a modelling of multimedia documents, it is important to link the metadata. There are four ways to make a connection between a document and the metadata associated with it:
- by direct insertion into the document: the metadata is contained in the document.
- by support: metadata is external but conveyed by the document and referenced by its URL,
- by Href link to the document: external metadata that can be conveyed separately from the document and referenced with a specific URL,
- by wrapping: the metadata set includes the resource.

Other studies have relied on models that aim to translate the total composition of each document on the basis of both their structures and their contents. They proposed a model that can handle two levels of heterogeneity documentary document: intra-and inter-document.

For the heterogeneous intra-document, a dichotomy between the structures and contents of documents is used. It is about translating separately the logical organisation of media and the content of each. For each document, it will be possible to link as many structures as required (nature of the documents, user’s point of view, etc.) [03].

The heterogeneity of inter-document defines a document’s class. Each class will be presented by a structure called a generic structure. Thus, many works were based on the annotation of documents. Annotating is to associate additional data to a document or part of existing document. This very general definition masks a wide variety of annotation process which is distinguished by:
- implementation automatically or manually,
- the language: natural language in a formal language,
- the goal of annotation: comments, information, qualification, filtering, indexing,
- the nature of the annotation: objective (adding metadata standard) or subjective (expression of a view).

Annotations can be used as memory for the user, they can also be used in research but also in the description, classification and community work. The annotation of multimedia documents is an extension of the annotation of textual records such as GDA (Global Document Annotation) [04]. They combine several types of annotations, related to images, audio or video. Research on the annotation of paper are mainly focused on the development of guidelines and standards of annotations such as text encoding initiative (TEI) [05], the corpus encoding standard (CES) [06] or Dublin Core (DC) for the annotation of textual records. This type of annotation is to identify the information that makes the context of the document content as intelligible syntactic structure, the structure of relationships. Automatic analysis of the content and meaning of words generates textual annotation. There are several annotation tools and description of documents, among these tools we quote: Annotea [07] which is a system for creating and editing Web documents annotations built on HTTP, XML and RDF.– Annotea provides an interoperable protocol appropriate to the performance in Web browsers that allows users to attach data to Web pages, and see the attached data when they review later the same pages. On the generation of metadata, annotate automatically fills the fields described by Dublin Core but do not realize the extraction of information. However, it proposes an annotation server which makes annotations publicly available.

- OntoMarkup [02] is an ontology-based annotation tool that allows the annotator to incorporate semantic information in the documents from the Web. The semantic annotation tool for marking contains a component-based ontology that allows the user to navigate and mark the appropriate pieces of information, a learning component that learns rules from examples and information extraction component that extracts the objects and the relationships between these objects.

- OntoMat-Annotizer [HAND, 01] is an annotation tool DAML + OIL, easy to use and interactive Web pages.– It has two browsers, one for an ontology and one for HTML. The browser aims to explore ontology of ontologies and shows the HTML for the parts annotated text. OntoMat-Annotizer is based on Java and provides an interface module for extensions. The intended users are individual annotators is people who want to enrich their web pages with metadata DAML. Instead of manually annotating the page with a text editor such as XEmacs, OntoMat allows the annotator to highlight relevant parts of the web page and create new instances by interaction type "drag and drop". This tool uses the model CREAM - an open model for the development of annotation tools based on ontologies.

- OntoAnnotate [08] is a tool for semi-automatic annotation, which brings together the knowledge of documents and Web pages. It creates a database containing metadata and enriching resources Web or Intranets with this metadata. With this tool, you can annotate not only static HTML documents but also Excel or Word documents.

For audiovisual materials, the platform offers Advene a number of modular tools for presenting and manipulating annotations in various ways. The only element always present is the video player, around which the user can choose to open different views: to take notes, bookmark view, video ...

Faced with the limitations of previous work which can be summarized mainly by the absence of some metadata, "document’s author" and "year of publication", that are deemed interesting for a text document in the works of Jdidi [02] and ambiguity in the media in the works of Mbarki [03], we will present, in this paper, a model of metadata associated with a text document,
audio and image taking into account the limitations mentioned above.

Faced with the limitations of previous work which can be summarized mainly by the absence of some metadata, "document’s author" and "year of publication", that are deemed interesting for a text document in the works of Jdidi [02] and ambiguity in the media in the works of Mbarki [03], we will present, in this paper, a model of metadata associated with a text document, audio and image taking into account the limitations mentioned above.

II. PROPOSAL FOR METADATA MODELLING

Our contribution in the modelling of documents is summed up in the integration of a specific family of metadata that describe each type of media (text, image, audio) representing the components of a pedagogic video conference in medicine. We add a dimension of semantic modelling by using a family of generic-level metadata defined in an objective manner by each user. Thus, for each medium, we define a diagram showing the metadata, a definition of the document type (Document Type Definition: DTD) by the language of description of SMIL documents and, finally, the merger of these metadata.

II.1. Metadata Proposed for Text Documents

Our proposal is as follows:

<table>
<thead>
<tr>
<th>Metadata</th>
<th>Attribute</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Document Name</td>
<td>nom_doc</td>
<td>Represents the name of the text file, used for cataloging and indexing in search engines</td>
</tr>
<tr>
<td>Document Size</td>
<td>taille_doc</td>
<td>Refers to the size of the document. It can be used prior to loading or sending the document via the Internet to find the transfer speed according to system capacity available</td>
</tr>
<tr>
<td>Outgoing links</td>
<td>l_sor_doc</td>
<td>This metadata contains a list of links that bind the document to remote resources. These are internal links to the document.</td>
</tr>
<tr>
<td>Incoming links</td>
<td>l_ent_doc</td>
<td>This metadata contains a list of links to documents containing the document question in their outgoing links.</td>
</tr>
<tr>
<td>Keywords</td>
<td>mot_clé_doc</td>
<td>This metadata is the list of words that are considered most relevant document.</td>
</tr>
<tr>
<td>The author of the document</td>
<td></td>
<td>These metadata correspond to the author who prepared the document. This information is necessary to know how important the document is: for example, a document prepared by an expert is more important than a document prepared by a non-expert.</td>
</tr>
<tr>
<td>The year of publication</td>
<td>ann_doc</td>
<td>These metadata correspond to the year of publication of the document. These metadata tell whether the document is recent or old</td>
</tr>
</tbody>
</table>

Table 1. Metadata from a Text Document

This metadata is presented in the following figure

```xml
<smil>
  <Head>
    <Meta name="title" content="the DTD in a text file" />
    <layout type="text/smil-basic"/>
  </Head>
  <body>
    <par>
      <seq>
        <text src="Nom the document "Project value="Rapport end d'étude">
        <text src="Taille document "value="11 pages ">
        <text src="Liens outgoing "value="L1, L2, L3">
    </seq>
  </body>
</smil>
```

Figure 1. Metadata with the Document Text
II.2. Metadata Proposed for Audio

<table>
<thead>
<tr>
<th>Metadata</th>
<th>Attribute</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>A number</td>
<td>num_seg</td>
<td>The number for a given segment</td>
</tr>
<tr>
<td>Nature</td>
<td>nat_seg</td>
<td>A segment may designate a speech, music or other</td>
</tr>
<tr>
<td>A frequency</td>
<td>Fré_seg</td>
<td>frequency is used to search for the frequency band of the document</td>
</tr>
<tr>
<td>A start and end mark</td>
<td>rep_seg</td>
<td>This metadata represents the beginning and end of a sequence</td>
</tr>
</tbody>
</table>

### Table 2. Metadata of an Audio Document

If an audio document is of nature "Word", we add the speaker and key words as metadata.

- "speaker" is the speaker associated with a given segment. This metadata is represented by the attribute loc_seg.
- "Keywords" are the keywords associated with a given segment. This metadata is represented by the attribute mot_seg.

This metadata is presented in the following figure:

**Figure 2. Metadata with the Audio Document**

The DTD describing the metadata:

```xml
<smil>
  <head>
    <Meta name = "title" content = "the DTD of an audio document" />
    <layout type= "audio /smil-baisc ">
    </layout>
  </Head>
  <body>
    <par>
      <seq>
        <audio src: "un number "value="1 ">
        <audio src: "une nature "value="parole ">
        <audio src: "une frequency "value="5000 Hz ">
        <audio src: "un mark start and end "value="05 minutes ">
        <audio src: "un speaker "value="Amin ">
        <audio src: "les keywords of this segment "value="va, travaille, buy ">
      </Seq>
      </Par>
    </Body>
  </Body>
  </Smil>
</smil>
```

Example of a DTD describing an audio nature of "word":

```xml
<smil>
  <head>
    <Meta name = "title" content = "the DTD of an audio document" />
    <layout type= "audio /smil-baisc ">
    </layout>
  </Head>
  <body>
    <par>
      <seq>
        <audio src: "un number "value="2 ">
        <audio src: "une nature "value="parole ">
        <audio src: "une frequency "value="3000 Hz ">
        <audio src: "un mark start and end "value="05 minutes ">
        <audio src: "un speaker "value="Amin ">
        <audio src: "les keywords of this segment "value="va, travaille, buy ">
      </Seq>
      </Par>
    </Body>
  </Body>
  </Smil>
</smil>
```

Example of a DTD describing an audio nature of "word":
II.3. Metadata Proposed for Image Documents

For an image document, we propose the following metadata:

<table>
<thead>
<tr>
<th>Metadata</th>
<th>Attribute</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Image number</td>
<td>Num_img</td>
<td>It is the number associated with an image document.</td>
</tr>
<tr>
<td>Subject of the image</td>
<td>Suj_img</td>
<td>is the subject associated with an image document.</td>
</tr>
<tr>
<td>Shape of the image</td>
<td>Frm_img</td>
<td>It is the form associated with an image document. In fact, each image has a definite form, for example square, rectangle, circle ...</td>
</tr>
<tr>
<td>Image size</td>
<td>Taille_img</td>
<td>It is the size associated with an image</td>
</tr>
<tr>
<td>Lighting level</td>
<td></td>
<td>It is the degradation from the original color to the white</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Metadata</th>
<th>Attribute</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Numéro of l'image</td>
<td>Num_img</td>
<td>It is the number associated with an image document.</td>
</tr>
<tr>
<td>Sujet of l'image</td>
<td>Suj_img</td>
<td>is the subject associated with an image document.</td>
</tr>
<tr>
<td>forme of l'image</td>
<td>Frm_img</td>
<td>It is the form associated with an image document. In fact, each image has a definite form, for example square, rectangle, circle ...</td>
</tr>
<tr>
<td>taille of l'image</td>
<td>Taille_img</td>
<td>It is the size associated with an image</td>
</tr>
<tr>
<td>niveau d'éclairage</td>
<td>Taille_img</td>
<td>It is the degradation from the original color to the white</td>
</tr>
</tbody>
</table>

This metadata is presented in the following figure:

La tuberculose est une maladie infectieuse transmissible et non immunisante, avec des signes cliniques variables. Elle est provoquée par une mycobactérie du complexe tuberculosis correspondant à différents germes.
III.1. Presentation of a Video Conference
This video conference on TB is the beginning of a series of medical video conferences in the year 2008 at the initiative of the Committee for Scientific Cooperation with Vietnam, the Pasteur Institute in Paris and the support of the Agence universitaire de la Francophonie.

By applying our tool in this video conference, we identified the following metadata

III.2. Text Metadata
- Document Name: Introduction to video conference.
- Text size: 4 lines.
- Outgoing Links: Tuberculosis
- Incoming links: Definition
- Key words: tuberculosis, miobactérie, medical videoconferencing.
- Author of the document: Agence universitaire de la Francophonie.
- Year of publication: 2008

III.3. Audio Metadata
- Number of the author: 1
- A nature: speech
- A frequency: 50 MH
- A mark start and end: start: 13h 30, end: 14h

III.4. Image Metadata
- Form of the image: Rectangular
- Image size: height: 10.88 cm, Width: 16.93 cm
- Uniformity: 50%
- Image number: 2
- Subject of image: video conference on TB

III.5. Fusion: Generic Metadata
To solve the problems of overlap, we need: Replace metadata Num_image audio ID = 2 and = 1 by the metadata Numéro_sequence = 2, so our presentation will be as follows:
- Document Name: Introduction to video conference.
- Text size: 4 lines.
- Outbound Links: Tuberculosis
- Incoming links: Definition
- Key words: tuberculosis, miobactérie, medical videoconferencing.
- Author of the document: Agence universitaire de la Francophonie.
- Year of publication: 2008
- Nature of the sequence: word
- Frequency of the sequence: 50 MH
- Finds the start and end of the sequence: start: 13h 30, end: 14h
- Shape of the sequence: rectangular
- Size of the image of the sequence: height: 10.88 cm, Width: 16.93 cm
- Uniformity: 50%
- Sequence number: 2
- Subject of the sequence: video conference on TB

The DTD describing the metadata:

```
<smil>
  <Head>
    <Meta name = "title" content = "Generic DTD" />
    <layout type= "text /smil-baisc ">
  </Layout>
  </Head>
  <body>
    <par>
      <seq>
        <text src: "Nom the document "Introduction to video conference value=" ">
        <text src: "Taille text "4 lines value=" ">
        <text src: "Liens outgoing "value=" tuberculosis ">
        <text src: "Liens incoming "value=" definition ">
        <Text src "Keywords" value = "TB miobactérie, videoconferencing
        Medical ">
        <text src: "Author of the document" value = "Agence universitaire de la Francophonie ">
        <text src: "Année edition "value="2008 ">
        <audio src: "Nature of the sequence" value=" parole">
        <audio src: "Frequency of sequence" value="50 MH ">
        <audio src: "un mark start and end "value="13h 30, end: 14h ">
        <img src "Subject of the sequence" value = "video conference on TB ">
        <img src "forme of séquence" value="rectangulaire ">
        <img src "image size of the sequence" value = "height: 10.88 cm width: 16.93 cm">
      </Seq>
    </Par>
  </Body>
</Smil>
```

CONCLUSION
Modeling of multimedia documents is a center of active interest. We have presented a documents modeling that takes into account each type of media. This model incorporates a dimension of semantic annotation according to the class of generic metadata. It can be defined objectively by each user to allow a non-expert user to interact in a more friendly way.
REFERENCES: