The New Navigators: from Professionals to Patients
R. Baud et al. (Eds.)
IOS Press, 2003

Indexing method of digital audiovisual medical resources with semantic Web integration

Marc Cuggia[a], Fleur Mougin[a], Pierre Le Beux[a].

[a]Laboratoire d'Informatique Médicale, University of Rennes1, France
{marc.cuggia,fleur.mougin.pierre.lebeux}@univ-rennes1.fr

Abstract

Digitalization of audio-visual resources combined with the performances of the networks offer many possibilities which are the subject of intensive work in the scientific and industrial sectors. Indexing such resources is a major challenge. Recently, the Motion Pictures Expert Group (MPEG) has been developing MPEG-7, a standard for describing multimedia content. The goal of this standard is to develop a rich set of standardized tools to enable fast efficient retrieval from digital archives or filtering audiovisual broadcasts on the internet. How this kind of technologies could be used in the medical context? In this paper, we propose a simpler indexing system, based on Dublin Core standard and compliant to MPEG-7. We use MeSH and UMLS to introduce conceptual navigation. We also present a video-platform with enables to encode and give access to audio-visual resources in streaming mode.

Keywords:
Audiovisual indexing, medical informatics, semantic web, MPEG-7, e-learning.

1 Introduction

The development of digital video-based communication systems is heavily influenced by capabilities of the new digital technology[1]. High-density storage and ubiquitous networking have enabled a class of video-based applications that were not previously viable. Owing to progress of High quality video compression standards (e.g. MPEG-2 or MPEG-4); it becomes finally possible on Internet, to watch great quality video. In the medical field, video productions have grown exponentially over the past several years. The video content is quite diversified and generally produced in clinical activity context. For example, it is the case of endoscopies, Ultra Sound exams, or video-surgery. Medical teaching constitutes also another significant source of video production.
In the future, this production will certainly increase. Several factors can explain this perspective. For example, PACS deployment within health settings and the necessity to store medicals files for a long period. The fact that generally these videos are stored on analogical video tapes and are not indexed represents an obstacle for their exploitation.

For a long time, broadcasting content has been confronted with audio-visual indexing problem. This one has been studied by research and led to several industrial applications [2, 3].
Are these solutions adapted to the medical field? What are the constraints related to this domain? In this paper, we propose an indexing method for medical audio-visual resources which should be both robust and compatible with future standards.

2 Background and objectives

2.1 Indexing problems

Indexing is the process which represents data resulting from the analysis of a document (or a part of it) with elements of a documentary language.

2.1.1 Based content characteristic extraction

Video indexing takes fundamental principles of still image indexing theories. A still image is characterised by three types of attributes. First, primitive's attributes intended to describe basic characteristics like colour, form, texture, spatial or temporal frequencies components of the image. The second type of attribute (called "logical attributes") relates to objects descriptions (e.g.: hand, head, heart) and logical links with the primitive attributes (e.g. heart ⇔ red, head ⇔ oval). Lastly, the content of a picture (or its parts) is described by the semantic attributes. Attributes extraction of picture can be automatic, semi-automatic or manual. Primitive attributes can be easily extracted by using simple's processes of images analysis. On the other hand, the logical attributes are much more difficult to extract. They generally require semi-automatic processes. An expert has to correct and validate what the computer found. Automatic extraction of semantic information from still images and the analysis of their content are far from being currently usable. Even if this field of research is very active, this type of extraction is made generally manually. Software can however facilitate input and especially classification of semantic informations.

Compared to still images, videos present supplementary difficulties due to their dynamic dimension (temporal and spatial dimension). Because of these new constraints, the semantic description becomes all the more difficult and requires human interpretation. Original processes were used to automate the semantic extraction like voice recognition of sound-track or optical character recognition (OCR) of video's captions i.e. text contained in videos.

2.1.2 Indexing systems

Once the extraction of attributes is realized, the second important stage is to structure them to obtain an optimal indexation. Some experiences were led in medical domains but they do not integrate the necessity of adopting semantic web standards[4].

The necessity to provide detailed and often highly structured descriptions of multimedia content data has led to MPEG-7 Standard definition[5]. MPEG-7 is an ISO/IEC standard which aims to normalise a core of audio-visual features, called Descriptors (D), and structure of descriptors and their relationships, call Description Schemes (DS). An enormous potential interest exists in an international standard format for the storage and exchange of descriptions that could ensure interoperability between video archive operators, perennial relevance of the metadata and a wider diffusion of the data between the professional and general public.

2.2 Semantic descriptions

Whatever the indexing system is, instantiations of semantic descriptors must be based on concepts and their conceptual relations. Without the use of specific ontology, the range of semantic description does not exceed the document itself. In other words, the use of domain's ontology (e.g. UMLS for medical field) guaranties "semantic continuance"[1] between documents description.

3 Objectives

One of the most important goals in the MVU[6] is the sharing of the medical knowledge. This knowledge is included into what we named a resource and all the resources are indexed and available on web servers. The Resources Servers allows acquiring, managing and storing the medical multi-media resources (texts, graphs, images, sound and video documents) according to descriptive, semantic and pathological criteria. Thus, these supports will be usable for initial or continuous teaching and particularly in remote teaching. The content indexing should allow the optimization of the access to relevant information in the fields of medical knowledge.
In this context, our objective was to enable a video indexing system which takes account into all of these constraints. Unfortunately, up to now, tools allowing the use of MPEG-7 were not available. Therefore we had to adopt a less elaborated solution but immediately usable.

4 Methods and materials

4.1 The video's server platform

In context of the French project of Medical Virtual University (UMVF), we produced a streaming-video platform (fig. 1) which enables the broadcasting of medical videos in several formats (QuickTime, MPEG4, MPEG2, MPEG1, Real) to medical students.
We could test and validate our indexing system on a medical video corpus.

4.2 Software and language choice

Since it was intended to function on Internet, we chose Open Source Web technologies. Thus, our indexing system is developed in PHP language. It uses XML to insure the best interoperability and facilitate database exchanges. Descriptions of video are stored in a MySQL database. Web pages are managed by Apache server. The whole of the system works on Linux Mandrake 9.0.

4.3 The choice of Standards

As we already said, we could not use the whole MPEG-7 technology. We started with an already existing solution i.e. Dublin Core (DC)[7]. DC was designed specifically for generating metadata to facilitate the search of Web resources through the use of 15 core elements and specific sub-elements (qualifiers).
Hunter[8] has specified a mapping between MPEG-7 and DC descriptors. Thus, it is possible to express video's description with DC grammar and conserve compatibility with MPEG-7 standard.

4.4 Semantic aspects

With DC, Semantic content can be described with "SUBJECT" qualifier. DC authorizes the employment of MeSH scheme with this element, thus it is possible to use concepts of UMLS but restricted to the MeSH thesaurus.

4.5 Coding assistant tools

Descriptions of video are entered through a Web interface. The search of MESH terms is facilitated by the use of NomIndex[9] tool's functionalities. This allows users to input a natural language expression and get the corresponding MeSH terms. NomIndex can also extract MeSH terms from the web pages which contain the video resource. This functionality is very interesting because video resources are usually embedded in html and these pages contain semantic textual information.

5 Results

We have implemented a Web-server architecture with intuitive interface for users to add or to update video indexation.

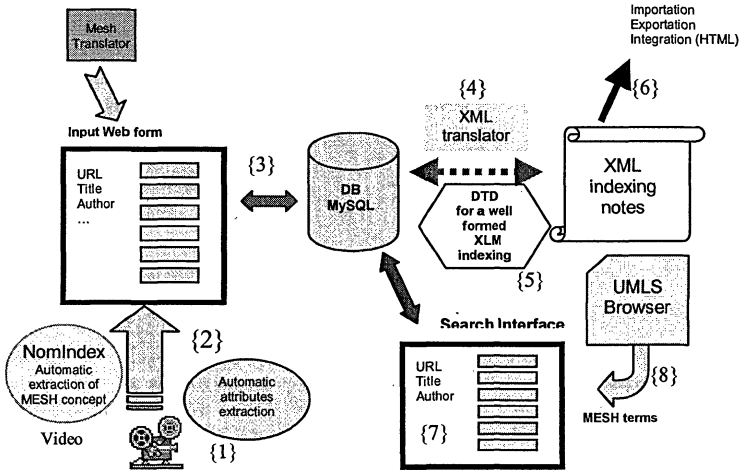

Fig 1 : Audio-visual Web indexing system

The number n {} are referring to figure 1.

Step 1: Video note creation
To index a video {1}, user fills an online Web form containing 15 DC's qualifiers. Some of them could be instanced easily by automatic extraction processes, for instance file's size or format, codec type, bit-rate (i.e. compression rate). Indeed, new video standards like MPEG-2, QuickTime or Windows Media Files supply automatically this kind of information {2}. On the other hand, we have not yet implemented sophisticated processes of video analysis and automatic content extraction.

Step 2: information storage
All informations are stored in a MySQL relational database {3}. Our conceptual data model allows unlimited items description for each resource.

Step 3: Interoperability
Then indexation notes are translated into XML {4} format. To obtain well-formed notes, the validity of XML notes is controlled with a DTD {5}.
XML format facilitates exchanges between databases and allows semantic web integration. By this way, it is easy to import and export informations. Metadata's integration in HEAD section of HTML pages is also facilitated {6}.

Step 4 : Search engine
The search engine allows muticriterion queries {7}. Search can be carried out on the 15 DC's qualifiers.
Semantic search is founded on the MeSH terms. In the present version, user can be helped by the UMLS browser {8} which has been developed in our laboratory. Thus, it is possible to obtain from a natural expression or a UMLS concept[10], all the nearest MeSH concepts.

Conversely, owing to UMLS Browser, it is possible to obtain video sequences which contents are close to a given concept.
For example:
If the query contains "neurological tremors", search engine replies MeSH Term tremor. This term is classified as a sub-element:
Diseases Category? *Nervous System Diseases*? *Neurologic Manifestations*? *Dyskinesias*? ***Tremor***
We obtain child and parent concepts on several levels with respected type of relation:

Parent level:	*Child level:*
... ? INVOLUNTARY MOVEMENT ? CONTROL OF MOVEMENT (+) ? SYSTEME NERVEUX, MALADIES ? DISEASES (MESH CATEGORY) (+) ...	... ? INTENTION TREMOR ? PASSIVE TREMOR ? STATIC TREMOR ? DRUG INDUCED TREMOR ? BENIGN ESSENTIAL TREMOR (+) ? ESSENTIAL AND OTHER SPECIFIED FORMS OF TREMOR ? REST TREMOR ...

Finally, we can also get broader concepts (i.e. CHOREE, MERCURIAL TREMOR) and narrower concept (ARSENICAL TREMOR, COARSE TREMOR) etc...

6 Discussion

6.1 Dublin Core and UMLS limitations

The simplicity of DC presents advantages (i.e. a concise resource description) and disadvantages. The first advantage corresponds to the important liberty of qualifier refinement. Indeed, DC standard offers the possibility of qualifiers declination without insuring a solid structuring.
For example, video shot indexation requires an import number of sub-qualifiers not really structured:

<DC.Type Content="Image.Moving.sequence.scene.shot.frame.object ">
< DC.Description.Text Content="Text ">
< DC.Description.Position Content="Point ">
< DC.Description.Shape Content="Polygon ">
< DC.Description.ShapeContent="Point ">

Thus, DC is better suitable for text indexation. Its major asset is to be adopted by many international organizations. Its implementation is easy and can be realized with semantic web languages like XML, RDF and RDFS. Furthermore MESH is also available in French.

In the current version, UMLS cannot be exploited totally by our system. Indeed, DC authorises only MESH employment for Subject qualifier. Thus, only a part of the conceptual wealth is usable. The lack of French translation of the UMLS is a second problem in our context.

6.2 MPEG-7

MPEG-7 corresponds totally to the constraints of video indexing. It allows expression of all attributes in structured arborescence. But contrary to DC standard, some fundamental fields use too many arborescence levels.
(I.e. DC: Subject ⇔ MPEG-7: CreationMetaInformation.Classification.PackagedType)
MPEG-7 is initially destined to cinematographic and broadcasting sectors. It is perfectly adapted to the description of shots, scenes, soundtracks, or storyboards. But these features are not all interesting in medical context. Medical use of MPEG-7 requires special adaptations. Owing to its adaptability, MPEG-7 allows redefinition of description's schemes and marks. Thus, it is possible to define a set of description items destined to specific objectives (i.e. a spatial descriptor for endoscopy describing progression of endoscope tube relative to dental arches).

6.3 Description granularity

Information being used for description must have a granularity adapted to the type of videos. Thus, endoscopy indexing requires high level of description to describe for example mucous membrane colour or certain regions of interest. On the other hand, teaching videos don't require such description level.
Cost of indexing must be considered, especially if this one is carried out manually.

6.4 Indexing automation

Extraction of basic videos characteristics (i.e. length, bit-rate, size) should be easily realized. However, video standards are not all compliant to this kind of extraction. Some of these standards embed these characteristics directly in the video file; others require file loading in memory to make this information accessible.
Extraction methods of high level semantic attributes are still on going research topics and are not currently used. Even if it requires a human intervention, semantic description will be facilitated by the use of indexing assistance tools. These tools will detect changes likely to correspond to different semantic characteristics in the video stream.

6.5 Embedded metadata

Metadata integration in video files constitutes a decisive progress in the indexing audiovisual field. Indeed, previously, indexing notes were separated from the video files.
To reach metadata relating to a video file, it was necessary to be connected to a specific database. Owing to metadata embedding, video files content their own structured metadata and so, their own semantic. Even if currently video standards do not allow such integration, while being compatible with MPEG-7, our system will offer this possibility.

7 Conclusion

Audio-visual indexing poses many fundamental problems. Promising technical solutions (MPEG-4 and MPEG-7) are not today ready to be used but will be probably a great success. It appeared interesting to develop a system which is interoperable and compatible with these new standards. The migration of our indexing system towards MPEG-7 should not pose any particular problem. But from now on, our system allows foreseeing the possibilities of a conceptual indexing based on medical ontologies.

8 References

[1] Dechilly T., B.B., *Une ontologie pour éditer des schémas de description audiovisuels, extension pour l'inférence sur les descriptions.* Actes de IC'00, Toulouse, ed. T. Actes de IC'00. mai 2000.
[2] Ichiro, I., *An automatic Video Indexing Method Based on Shot Classification.* Systems and computers in Japan, 2001. **32**(9): p. 1543-1551.
[3] Jeannin, S., et al., *Motion descriptors for content-based video representation.* Signal Processing: Image Communication, 2000. **16**(1-2): p. 59-85.
[4] Berners-Lee, T. and J. Hendler, *Publishing on the semantic web.* Nature, 2001. **410**(6832): p. 1023-4.
[5] Software-Research-Center, *MPEG-7 Description Tool MovieTool.* 2002.
[6] Le_Beux, P., et al., *The French Virtual Medical University.* Studies in Health Technology and Informatics, 2000. **77**: p. 554-562.
[7] Initiative, T.D.C.M., *Dublin Core Qualifiers.* 2000.
[8] Hunter, J. *MPEG-7 Harmonization with Dublin Core: Current Status and Concerns.* in *3rd MPEG Meeting.* 2000. Beijing,.
[9] Pouliquen, B.D., P;Le Beux,P. *Indexation de textes médicaux par extraction de concepts, et ses utilisations.* in *JADT 2002.* St Malo France.
[10] Zou, H., et al., *Structured Data Management - the Design and Implementation of a Web-based Video Archive Prototype.* Proc AMIA Symp, 2001: p. 786-90.