A Digital Library Framework for Reusing e-Learning Video Documents*

Paolo Bolettieri, Fabrizio Falchi, Claudio Gennaro, and Fausto Rabitti

ISTI-CNR, via G. Moruzzi 1, 56124 Pisa, Italy
paolo.bolettieri,fabrizio.falchi,claudio.gennaro,
fausto.rabitti@isti.cnr.it

Abstract. The objective of this paper is to demonstrate the reuse of digital content, as video documents or PowerPoint presentations, by exploiting existing technologies for automatic extraction of metadata (OCR, speech recognition, cut detection, MPEG-7 visual descriptors, etc.). The multimedia documents and the extracted metadata are then indexed and managed by the Multimedia Content Management System (MCMS) MILOS, specifically developed to support design and effective implementation of digital library applications. As a result, the indexed digital material can be retrieved by means of content based retrieval on the text extracted and on the MPEG-7 visual descriptors (via similarity search), assisting the user of the e-Learning Library (student or teacher) to retrieve the items not only on the basic bibliographic metadata (title, author, etc.).

Keywords: MPEG-7, LOM, Metadata, Automatic Extraction, Multimedia Content Management System, Similarity Search, User Interface.

1 Introduction

In this paper we present the architecture of a Digital Library for enabling the reuse of learning documents. The Digital Library is based on MILOS, a general purpose Multimedia Content Management System created to support design and effective implementation of digital library applications. MILOS supports the storage and content based retrieval of any multimedia documents whose descriptions are provided by using arbitrary metadata models represented in XML. We present the architecture and the functions of MILOS, a Repository System intended to efficiently support the distributed storage and retrieval of Multimedia Learning Objects, developed by the ISTI-CNR laboratory in the context of the VICE italian project.

VICE is a three-year project, started in 2003, financed by the Italian Ministry of Education, University and Research (MIUR). The objective of the project is

* This work was partially supported by the VICE project (Virtual Communities for Education), funded by the Italian government.

E. Duval, R. Klamma, and M. Wolpers (Eds.): EC-TEL 2007, LNCS 4753, pp. 444–449, 2007.

to enable high quality and effective distance learning in a cost-effective manner, supporting, in an integrated fashion, teaching/learning activities organized by an authority (e.g., be an academic institution, an enterprise, an education provider, etc.) and self-learning (based on self-identified needs and goals), in the context of working activities. In this research we try to apply digital library techniques to support the management, retrieval and reuse of Learning Objects, i.e. collection of content/activities, that can be composed according to different needs and different goals.

The activity carried out from ISTI concerns the implementation of the prototype of repository system for multimedia LOs making advantage of the Multimedia Content Management System MILOS discussed below.

The main contribution of this paper is to show how the combination of the MILOS system and of state of the art tools for automatically extracting metadata from digital content is useful in enabling the reusing of digital material (such as videos, PowerPoint® presentations, etc.) in the domain e-Learning.

In this experimentation we have used LOM and MPEG-7 as metadata standards for the repository, and have proposed the specific use of a XML database combined with an access structure for similarity search for searching and retrieving the stored LOs. In particular our we have concentrated on the generation of "video-centric" LOs based on the analysis of some university lessons of the *Nettuno* [1] consortium, and of some PowerPoint documents taken from the web. To each digital items is associated a LOM descriptors created by hand (in XML), and an MPEG-7 description extracted automatically.

This paper is organized as follows: Section 2 presents the architecture of the MILOS MCMS. Then in Section 3 we present the metadata management by showing the model adopted and the tools exploited. Section 4 gives an overview of the search and browsing Web interface provided with reposting of the VICE project. Finally, Section 5 summarizes our contribution.

2 Repository System Architecture

In this section, we shortly describe the architecture of MILOS, the Repository System for Learning Objects which constitutes the main contribution of the ISTI-CNR Unit within the VICE project. MILOS is designed to support the storage and retrieval of multimedia Learning Objects (LO).

MILOS is a Multimedia Content Management System with a number of characteristics that make it particularly suitable for the development of Digital Library applications. MILOS is based on powerful multimedia database, able to guarantee advanced features for the persistence, search, and retrieval of Learning Objects written as XML documents and described using W3C XML schema [5]. Since the managed document are in XML format, it is possible to integrate heterogeneous XML descriptions such as LOM (The IEEE Learning Object Metadata (LOM) standard [4]) and MPEG-7 [2] metadata standards, since they are fully supported by the XML schemas. In particular, in the context of project VICE, LOM has be used to describe LOs, and MPEG-7 has be exploited for

enriching multimedia components of the LOs. Multimedia components of LOs can be images, videos, PowerPoint presentation, etc. MILOS is based on a three-tier architecture (see Section 2), and the search functionality exported by the services of business logic can be easily adapted on the basis of the XML–Schema of the managed documents.

The system is based on a three–tier architecture and composed of five main logical components: *Interface Logic, Automatic Metadata Integrator, Repository Service Logic, LO Database, and Metadata Database*. The Interface Logic includes components that allow users to interact with the system on the web, via normal browsers. The Automatic Metadata Integrator analyzes multimedia part of the LOs, to automatically extract metadata, integrating it to the metadata produced during the authoring phase. The Repository Service Logic manages accesses to data stored in the LO repository and metadata database, on behalf of the other two components. All the components communicate by means of protocols for distributed systems integration (e.g. SOAP). Further details about MILOS can be found in [3].

3 Metadata Management

As explained earlier Multimedia Metadata can be automatically generated using specific processors (e.g., OCR, speech recognizer, cut detector, etc.). The typical LO ingestion workflow is the following (see Figure 1):

1. When a new Raw Media Element is inserted, the phase of Automatic Metadata Integration starts. It extracts some multimedia features (such as scenes, OCR, etc) and transform them in MPEG-7 format.
2. The Raw Media Element is stored in the Large Object DB including its keyframes in case of audiovisual content.
3. The LOM description is created by editing the LOM metadata using a standard Metadata Editor.
4. The LOM description and the MPEG-7 description are associated and stored by means of the Repository Service Logic.

Metadata Representation and Extraction. The metadata generated by the Automatic Metadata Integrator component are represented in MPEG-7. For each Raw Media Element we generate exactly one MPEG-7 description in XML format. The Automatic Metadata Integrator is organized in plug-ins each of devoted to the automatic extraction of metadata of a specific type of Raw Media Element. In this way we guarantee the maximal flexibility and extensibility of the repository. In our implementation we have incorporated plug-ins for the metadata generation from video and PowerPoint documents. Each LO of the repository is composed of a LOM description that contains the educational metadata and a MPEG-7 description that describes the content of the raw media element associated. The two descriptions are integrated using URI link from the LOM description to the MPEG-7 description.

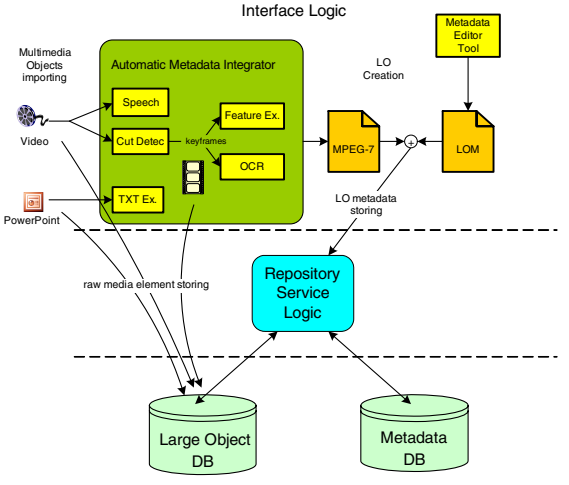

Fig. 1. Learning Obect ingestion workflow

Automatic Metadata Integration of Videos. The MPEG-7 description for e-Learning Video Documents is composed of several MPEG-7 descriptors. We have used the *CreationInformation* descriptor for expressing the common bibliographic metadata (such as, Title, Abstract, Location, Creation Data, etc.), and the *MediaDecomposition* descriptor for expressing the Video Transcript (by means of the AudioSegment descriptor) and the video decomposition in scenes and key–frames (by means of the VideoSegment descriptor). Inside the VideoSegment descriptor we have included the text extracted by the Video OCR component and the Visual Descriptors related to the key-frames. *MetaExtractor* is the tool that includes a set of modules for automatically generating MPEG-7 metadata from video lessons in MPEG-1/2 format. The tool provides the following functionalities:

Scene Detection: This component is used for segmenting video sequences by automatically locating boundaries of shots scene transition effects.
Visual Feature Extraction: This component extracts five MPEG-7 Visual Descriptors (SC, CL, CS, EH, and HT) from each key–frame of the scene detected by the *Scene Detection* component.
Video OCR: It detects, extracts and recognizes the texts contained in the video to enable text-based retrieval from spoken language documents.
Video Transcript: It generates transcript to enable text-based retrieval from spoken language documents.

Automatic Metadata Integration of PowerPoint Presentations. The extraction of metadata from the PowerPoint presentations is performed by extracting the title and the text contained in the slides. This content is organized by means of the MPEG-7, which is able to describe the decomposition of the presentation in slides. We use a free Java tool for automatic extracting the text content from the PowerPoint slides. The text content is the used for creating the MPEG-7 output.

Automatic Image Processing. Feature extraction techniques and automatic generation of MPEG-7 data Feature extraction was performed employing an application we built upon the MPEG-7 experimentation model of MPEG-7 Part 6: Reference Software. The software can extract all MPEG-7 VisualDescriptors. For the VICE repository we extract 5 MPEG-7 descriptors.in an image).

4 Web Search and Browsing Interface

The whole retrieval interface layout consists of four parts: (i) a *query frame*, in which the user can formulate fielded and full–text queries (top-left frame), (ii) an *hitlist frame*, in which the ranked list of matching items with some basic metadata (title, type, etc.) is displayed (bottom-left frame), a (iii) *LOM view frame*, where the whole metadata set of the LOM description for the selected item is displayed (top-right frame), and a (iv) *raw media element frame*, where the details of the raw media element associated with LOM are displayed (bottom-right frame). Through the menu "metadata" the fielded search form allows us to select which metadata model (LOM or MPEG-7) we have to use for the query search (Figure 2, left side). Selecting a specific model, the fields of the form on which to perform the search are automatically restored on the basis the metadata model selected. In particular, selecting the MPEG-7 model we can make searches on the OCR of the keyframes of the videos, on the transcripts of the spoken of the videos and on the textual content of the PowerPoint presentations slides.

The full–text interface contains only a simple input box, allowing us to submit queries at the same time on the entire metadata database independently from the model type. The two frames on the right side allow us to visualize an item of the list returned by the query (in the hitlist frame). By selecting an item from the hitlist the top frame show the content of its LOM metadata and the bottom frame the content of the raw media element. In particular, if the retrieved MPEG-7 is associated with a video, the raw media element frame displays the keyframes of the scene identified in the video and the complete transcription of speech (see

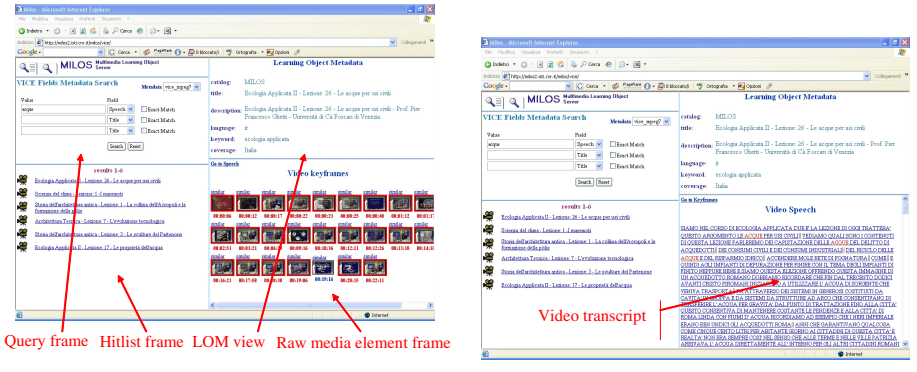

Fig. 2. Visualization of the video keyframes (left) and of the video transcript (right)

Figures 2, right side). From here, by making click on a keyframe or a phrase of the transcription, it is possible to play the video from the time corresponding to the scene or spoken phrase. By selecting the link "similar", present on top of each keyframe, it is possible to perform a search for similarity over all the keyframes of all the videos of the repository If the MPEG-7 description is associated to a PowerPoint document, the raw media element frame executes the presentation starting from the slide that matched the search.

5 Conclusions

Although from the theoretical point of view the idea of using automatic tools for the extraction and the enhancement of metadata in the field of the digital libraries is not at all new, it finds it hard to be used in the real world. The reason may be due both to the high cost of these tools or simply to the fact that people do not give sufficient confidence in their results. Moreover, sometimes digital libraries and metadata are seen by the user with suspicious eyes. We argue that the use of automatic tools is the only way to convince people of the importance of metadata and indexing techniques. This is demonstrated by the success of tools as *Desktop Search*: nobody is willing to install a digital library on his or her own personal computer for searching personal documents manually filled with metadata.

With this article we want to demonstrate, instead, that these technologies are truly useful. Also because some of them are available free of charge. We propose the use of a content management system based on a XML search engine and we experimented it on a dataset of documents belonging to the domain of the e–Learning. We showed that with a minimal cost in terms of time spent by the cataloguers (who have just to add the LOM descriptions) it is possible to reuse audiovisual and PowerPoint documents facilitating their utilization. We believe that the proposed approach can also be applied to other domains of digital library beyond the one of the e–Learning. To see a demo of the web search interface of VICE, go to the Web site `http://milos2.isti.cnr.it/milos/vice/`.

References

1. Nettuno, `http://www.uninettuno.it`
2. Mpeg requirements group, mpeg-7 overview, Doc. ISO/IEC JTC1/SC29/WG11N5525 (2003)
3. Amato, G., Gennaro, C., Savino, P., Rabitti, F.: Milos: a multimedia content management system for digital library applications. In: Heery, R., Lyon, L. (eds.) ECDL 2004. LNCS, vol. 3232, pp. 14–25. Springer, Heidelberg (2004)
4. Committee, I.L.T.S.: Ieee 1484.12.1-2002, draft standard for learning object metadata (2002), `http://ltsc.ieee.org/wg12/index.html`
5. Consortium, W.W.W.: XML path language (XPath), version 1.0, W3C. Recommendation (November 1999)