Using MPEG-7 for Automatic Annotation of Audiovisual Content in eLearning Digital Libraries *

Giuseppe Amato Paolo Bolettieri Franca Debole Fabrizio Falchi
Claudio Gennaro Fausto Rabitti

ISTI - CNR
Pisa, Italy
{giuseppe.amato,paolo.bolettieri,franca.debole,fabrizio.falchi,claudio.gennaro,fausto.rabitti}@isti.cnr.it

Abstract

In this paper we present a prototype system to enrich audiovisual contents with annotations, which exploits existing technologies for automatic extraction of metadata (such as OCR, speech recognition, cut detection, visual descriptors, etc.). The prototype relies on a metadata model that unifies MPEG-7 and LOM descriptions to edit and enrich audiovisual contents, and it 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. 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.).

1 Introduction

Although over the past few years several efforts have been devoted to develop metadata specifications for audiovisual content, the problem of creating and editing metadata descriptions has not received as much attention. The main reason for content providers creating useful metadata descriptions is not in the critical path of content creation and for managers of digital content repositories, this task is simply too expensive and time–consuming.

Ever since search engines have been available, creating metadata for text content has become more and more insignificant (and boring). The general opinion is that adding metadata to electronic text content is useless if information retrieval techniques can be adopted [7]. For audiovisual content the reality is completely different. Audiovisual content is particularly important in many different application sectors, such as TV broadcasting, professional applications (e.g. medicine, journalism, advertising, education, etc.), movie production, historical video archives. Furthermore, most of the video material produced is extremely difficult to access, due to several limitations: e.g., video documents are extremely large in size, so that archiving and transmission are expensive. Classical bibliographic metadata such as title, creator, creation date, etc., are insufficient for retrieval purposes and video document's content, even if extremely rich, is difficult to extract in order to support an effective content-based retrieval.

However, although automatic generated metadata and annotations are often error-prone, these tools are practically the only way to make the audiovisual content retrievable and accessible. Unless we are willing to spend a large amount of time to manually annotate the digital documents. For audiovisual material such as news, documentaries, distant learning courses, the speech content is very significant and correlated with the visual part. In these scenarios also text extraction from the visual part can be very relevant. Still, imagine a video lesson where a teacher speaks and a slide presentation is shown besides or in the foreground. Text identification and recognition (OCR) tool can be useful in retrieval the text in the slides. Even if affected by errors the text using these tools is still valuable. Some empirical evidence of this aspect is presented in [11]. These experiments illustrated the degradation of information retrieval at different levels of speech recognition accuracy and showed that word error rates up to 25% did not significantly impact information retrieval and error rates of 50% still pro-

*This work was partially supported by the VICE project (Virtual Communities for Education) and by the NeP4B project (Networked Peers for Business), funded by the Italian government.

vided 85 to 95% of the recall and precision relative to fully accurate transcripts in the same retrieval system. Image retrieval of visual content based on similarity search paradigm is another important way to retrieve audiovisual content. It is well known that similarity search based on low level features (e.g., the ones provided by MEPG-7) has strong limitations in retrieving relevant images in large collections of data containing videos or images belonging to numerous completely different themes (as for example the Web). Nevertheless when the collection is homogenous similarity search can be very useful. For instance, if the images are keyframes extracted from video scenes, the search can retrieve similar scenes even if they are very distant in time.

In this paper, we present the architecture and the functions of MILOS, a Digital Library 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.

The main contribution of this paper is to show the architecture of a Digital Library for enabling the reusing of multimedia documents (such as videos, PowerPoint® presentations, etc.) in a e-Learning context. The reuse of Learning Objects is based on automatically extracted descriptors carrying a semantic meaning for the professional that uses these Learning Objects to prepare new interactive multimedia lectures. The importance of this approach is implied by the expected economic advantages. In fact, the opportunity of reusing existing digital material by enriching the basic multimedia content with metadata, allows us to save the cost of creating new content, and, on the other hand, the cost of annotating it. This last point is important since, the alternative approach of manually annotating the content is time- and cost- consuming, and, thus, makes the system hard to scale.

This paper is organized as follows: Section 2 outlines the context of the VICE Project. Section 3 presents the architecture of the MILOS Multimedia Content Management System (MCMS). Then in Section 4 we present the metadata management by showing the model adopted and the tools exploited. Section 5 gives an overview of the search and browsing Web interface provided with reposting of the VICE project. Finally, Section 6 summarizes our contribution.

2 Context of the VICE 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 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. A Learning Object (LO) is any resource that can be used for an educational purpose. Within VICE we consider all kinds of possible LOs, from low-end technology (e.g. a web page or a PowerPoint presentation), to high-end technology (e.g. MPEG-4 interactive objects). LOs can be atomic, or structured (consisting of other LOs). They can have (semantic) relationships to other LOs (e.g. "x" is a prerequisite for "y", or "x" is an exemplification of "y"). LOs can be consumed individually, or can be combined to an organized course; a course can be (simply) a linear combination of LOs, or, in a more sophisticated manner, a workflow combining LOs in a complex manner.

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

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, 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.

3 Repository System Architecture

In this section, we describe the target 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 [9]. Since the managed documents are in XML format, it is possible to integrate heterogeneous XML descriptions such as LOM (The IEEE Learning Object Metadata (LOM) standard [8]) and MPEG-7 [3] metadata standards, since they are fully supported by the XML schemas. In particular, in the context

of project VICE, LOM will be used to describe LOs, and MPEG-7 will 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 3), 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.

A key characteristic of the MILOS system is its **flexibility** in managing different types of data and metadata and its independence from the specific format used to represent them. This implies that the application developer is not required to specify the details of the storage strategies used and the details of the access methods to be adopted; he/she only needs to specify the characteristics of the data and metadata and the functionality that are required, such as the requirement of storing high resolution photos, and supporting their efficient access based on the combination of metadata attributes and physical characteristics of photos. The flexibility of the MILOS system is also related to the possibility of developing end-user applications which are independent from the modality used to store data and search them. In particular, it is possible to store data described with a specific metadata model and to search them by using a different model.

Another key characteristic of the MILOS system is the **efficiency** in storing and searching multimedia objects. This requires a system which is (a) scalable with respect to the size of the archive and to the number of users accessing the application, and (b) efficient in processing complex queries over metadata values and data object's content.

The system is based on a three–tier architecture (see Figure 1) and composed of five main logical components: *Interface Logic, Automatic Metadata Integrator, Repository Service Logic, LO Database, and Metadata Database*. Figure 1 also illustrates the relationship between the LO Authoring Tools (box on top-left) and the LO Repository Tools (box on top-right). 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. In the following, we will give a more detailed description of each of these components. All the components will communicate by means of protocols for distributed systems integration (e.g. SOAP).

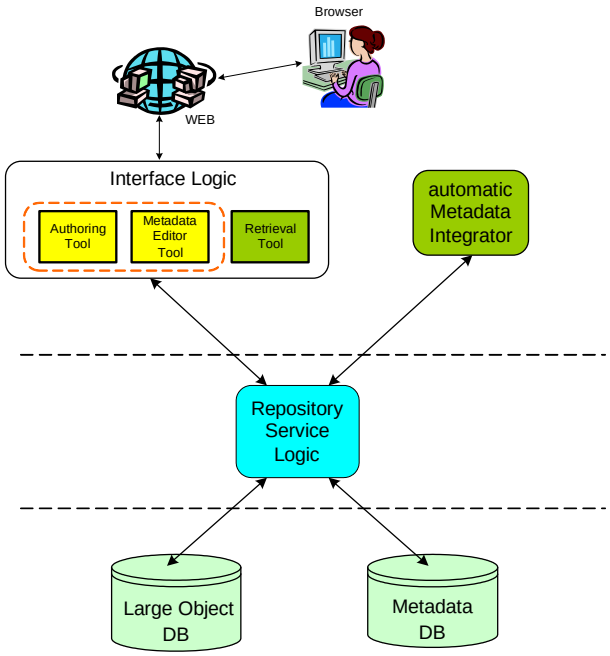

Figure 1. Architecture of Repository Systems for Learning Objects

3.1 Interface Logic

The interface logic includes modules, related to the activities that can be carried out by users of the system. Users typically interact with the system through normal web browsers. The metadata editor tool, which is part of the authoring tools, allows users to manually edit and review metadata associated with LO documents. The user can either edit automatically generated metadata, as for instance scene boundaries, or he/she can add additional metadata manually. The retrieval tool is used to search the system LOs. Various possibilities are offered by this interface: users can retrieve documents by performing full–text retrieval on the transcript or descriptions associated with LOs, selecting specific fields of the metadata structure, similarity search on multimedia content, or combination of them.

3.2 Automatic Metadata Integrator

This component integrates metadata generated during authoring of LO with automatically extracted information. It is composed of different modules, each one dedicated to a different automatic processing technique. For example, the OCR module recognizes textual video captions. The speech recognition module is able to generate a transcript in correspondence of an audio or audio/video document. The gener-

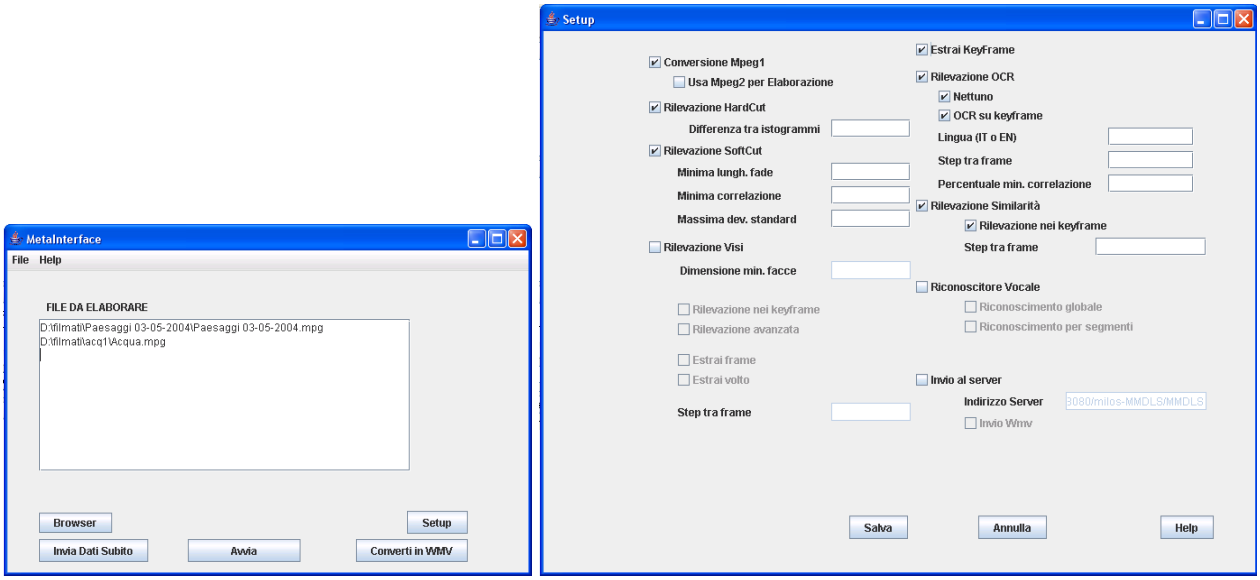

Figure 2. Screenshot of the Automatic Metadata Integrator: GUI for specifying the video files to elaborate (left); Dialog Box for specifying the parameters to be used during the elaboration of the video files (right).

ated transcript is indexed and the corresponding document can be retrieved by performing full–text retrieval. The cut detection module analyzes a video document and automatically identifies scene changes. In this way, metadata can be associated with specific portions of the document, instead of the whole document. The feature extraction module analyzes multimedia document in order to extract properties that can be used to perform similarity retrieval. Typical features extracted are color distribution, texture, edges, shapes, and motion vectors. All these automatic extracted information are then stored in MPEG7 format. Note that in the context of the project we did not develop specific multimedia feature extraction modules, but we integrated available open source products in order to implement the functions of the Automatic Metadata Integrator[1]. Figure 2 shows the screenshot of the Automatic Metadata Integrator GUI.

3.3 Repository Service Logic (RLS)

This component manages the accesses to the underlying databases: the LO database, that physically stores Learning Objects managed by the system, and the metadata database, where all metadata associated with the LOs are stored. It manages query processing by integrating and aligning information stored in the two databases. It performs reconciliation of retrieved data by managing ranking.

[1] `http://jakarta.apache.org/poi/`
`http://www.lienhart.de/Source_Code/source_code.html`

3.4 Metadata Database

This part of the system manages the Metadata associated with the LOs. Both Learning Metadata (i.e., LOM) and Multimedia LO Metadata (i.e., MPEG-7) are represented here in XML format, using appropriate XML Schema definitions. Therefore, the power of XML structuring and interlinking will be used in searching LO through the various metadata associated, exploiting also the interconnections between different metadata, at different levels of detail (e.g. information about the content of a scene of an audio/visual presentation, related to the range of technical requirement for display, and minimal background prerequisite for learners).

For this purpose we have designed and implemented an enhanced native XML database for Digital Library application. Our native XML database, in addition to support XML query language standards such as XQuery and XPath, offers advanced search and indexing functionality on arbitrary XML documents. Consider that new generation metadata standards, such as MPEG-7, include in their description also features automatically extracted from visual documents, such as color histograms, textures, edges, shapes, etc. Thus in addition to index structure that supports high performance search and retrieval on heavily structured XML documents [4], our XML database also provides full–text search, automatic classification [10], and feature similarity search [6] functionality. Specifically, the XML

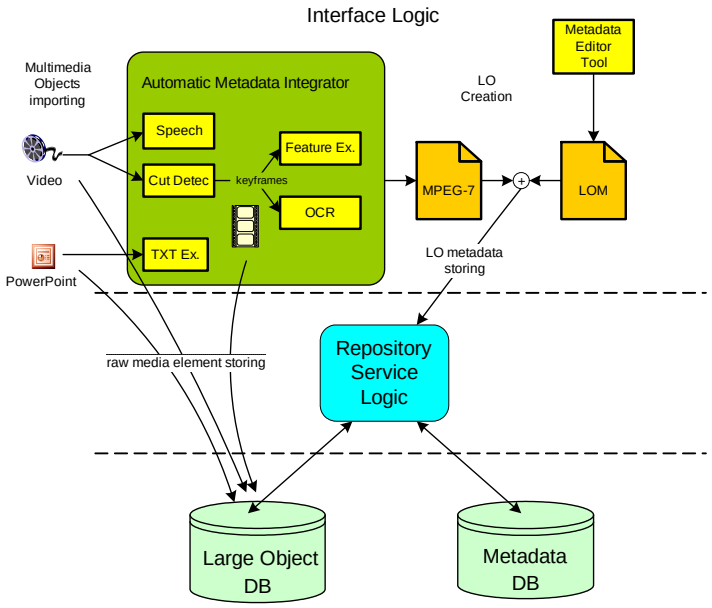

Figure 3. Learning Obect ingestion workflow

database allows the system administrator to associate specific element names of the XML with special indexes. So, for instance, the tag name `<abstract>` can be associated with a full–text index and to an automatic topic classifier that automatically index it with topics chosen from a controlled vocabulary. On the other hand, the MPEG-7 `<VisualDescriptor>` tag might be associated with a similarity search index structure and with an automatic visual content classifier. To deal easily and transparently with these advanced search and indexing functionalities, we have extended the syntax of the basic XQuery language with new operators that deal with approximate match and ranking. For more details about MILOS see [5].

3.5 Large Object Database

The Large Object Database is able to deal with various media formats. We refer to the items contained in this database as *Raw Media Element*. Examples include pdf/word documents, web pages, PowerPoint presentations, audio/video documents, etc. XML structures will be used to describe aggregations, at different levels, of elements included in LOs. It will be possible to access an entire LO or to access its specific parts in a selective way. Basic searching functionalities will be provided as well.

The key idea is that the DL application should deal with documents in a uniform way, independently of the specific strategy used to manage them. Thus, the Large Object Database identifies all documents with an URN and maintains a mapping table to associate URNs with actual storage locations. Applications use the URN to get or store documents from the Large Object DB, which behaves as a gateway to the actual repository that stores the document. The system administrator can define rules that make use of MIME types, to specify how the Large Object DB has to store a document of a specific type. For example, the rule may specify that an MPEG-2 video has to be stored in a tape of a silos, while an image will be stored in an array of disks.

A special care is taken to deal with the actual access protocols offered to retrieve the documents. An application will refer a specific document always using its URN. However, the retrieval of the document should be done using an access protocol compatible with the storage and delivery strategy associated with the document. For instance, when the document is stored in a web server it will be retrieved with an HTTP request. On the other hand, suppose that a video document is served through a commercial video server; in this case the real time streaming of the video will be obtained using RTSP [2]. When an application requires to retrieve a document, the MMS will translate the given URN into a specific handle (for instance an RTSP URL) that the application will use to access the document.

4 Metadata Management

4.1 Introduction

As explained earlier Multimedia Metadata can be automatically generated using specific processors (e.g., OCR,

speech recognizer, cut detector, etc.). Besides the automatic metadata generation, a standard editor for LOM description is placed in. The editor is intended to be used by the manager of the LO repository, who inserts new LO and that specify its metadata. The typical LO ingestion workflow is the following (see Figure 3):

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.

4.2 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. Each description is expressed by an XML item in the *Metadata Database*, which in turn is univocally identified by an URN (Uniform Resource Name). Therefore, the association LOM/MPEG-7 is obtained by specifying the URN of the MPEG-7 description in the technical/location field of the inside the LOM description (see Figure 4).

4.3 MetaExtractor: Automatic Metadata Integration of Video Documents

The MPEG-7 description for e-Learning Video Documents is composed of several MPEG-7 descriptors. Figure 5 shows an example of an instance of MPEG-7 associated

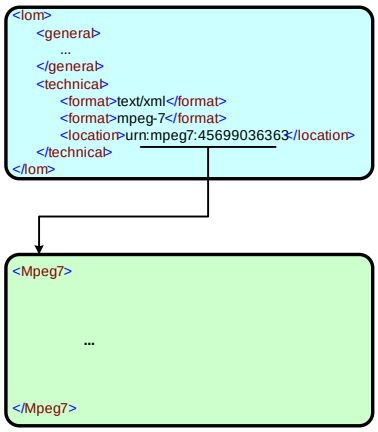

Figure 4. Representation of the link between a LOM description and the MPEG-7 descriptor.

to a video content. 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 keyframes (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 keyframes. MetaExtractor is a tool that comprehends 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. It uses two types of boundaries detection: one for the detection of the hard cuts and one for the detection of the fades and dissolves. In addition it determines the keyframe of each scene used by the Visual Feature component.

Visual Feature Extraction This component extracts five MPEG-7 Visual Descriptors (ScalableColor, ColorLayout, ColorStructure, EdgeHistogram and HomogeneousTexture) from each keyframe 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.

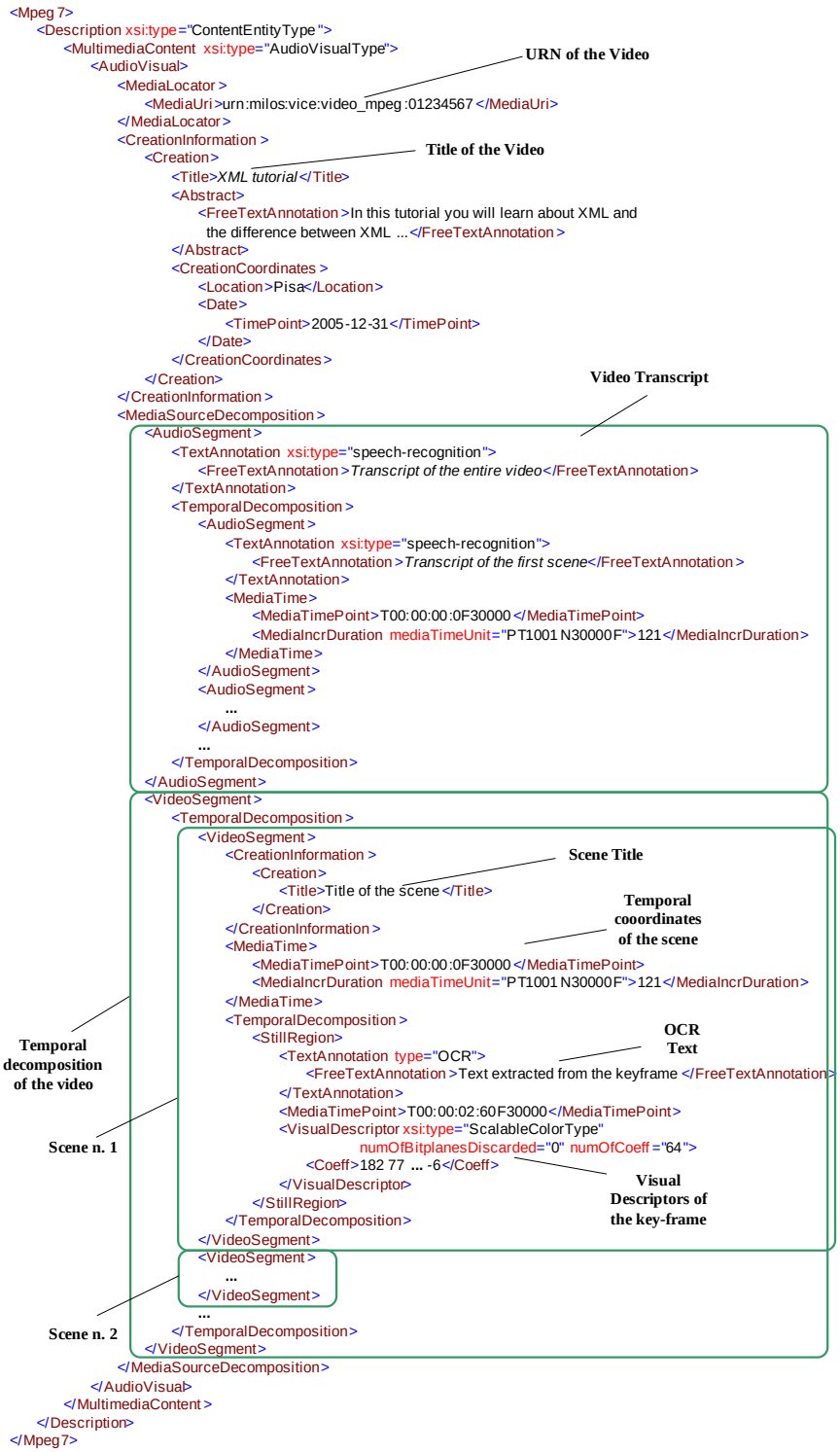

Figure 5. Example of an instance of MPEG-7 associated with a video.

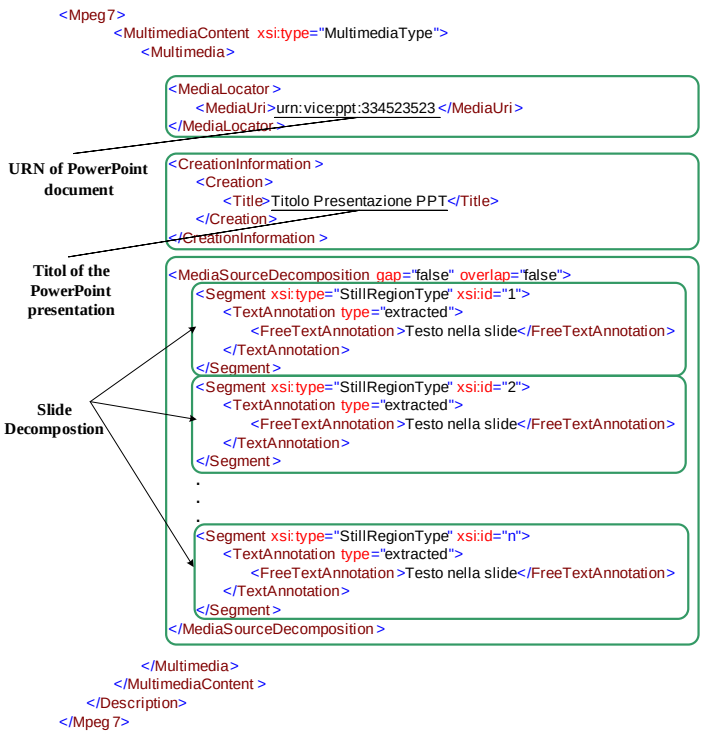

Figure 6. Example of an instance of MPEG-7 associated with a PowerPoint presentation.

4.4 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. Figure 6 shows an example of description in MPEG-7 used in VICE is shown an example of one in order to index documents PowerPoint.

4.5 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 (XM, [13]) of MPEG-7 Part 6: Reference Software. The software can extract all MPEG-7 VisualDescriptors defined in [12]. For the VICE repository we extract 5 MPEG-7 descriptors: ScalableColor (a color Histogram in the HSV Color Space), ColorStructure (captures both color content and information about the spatial arrangement of the colors), ColorLayout (represents the spatial layout of color images), EdgeHistogram (spatial distribution of five types of edges), HomogeneousTexture (characterizes the properties of texture in an image). The result of the extraction process is an XML document like the one in Figure 5 without usage and creation information. The values inside the `<VisualDescriptor>` tags are integer vectors (SC, CS and EH) or more complicated XML subtrees with integers as values (CL and HT).

5 Web Search and Browsing Interface

As explained in the introduction, the main contribution of this paper is to show the architecture of a Digital Library for enabling the reusing of multimedia documents in a e-Learning context. The reuse of Learning Objects is based on automatically extracted descriptors carrying a semantic meaning for the professional that uses these Learning Objects to prepare new interactive multimedia lectures. However, one of the major obstacles to implement this paradigm is the difficulty in mining large datasets of unstructured multimedia documents, such as videos. For this reason, the design of the retrieval interface is of primary importance in this context.

The objective of the retrieval interface is twofold: (1) to

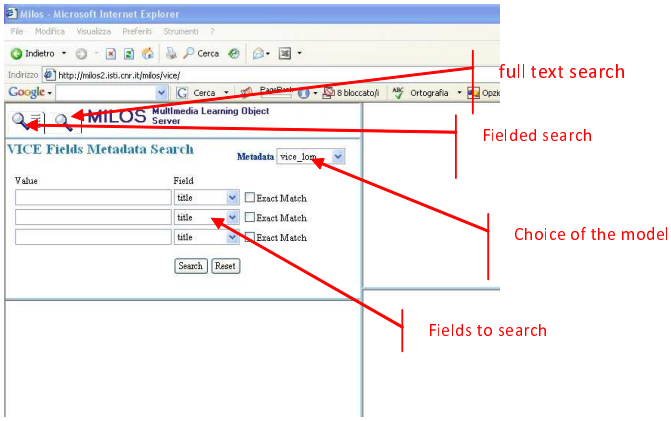

Figure 7. The query frames for the fielded search.

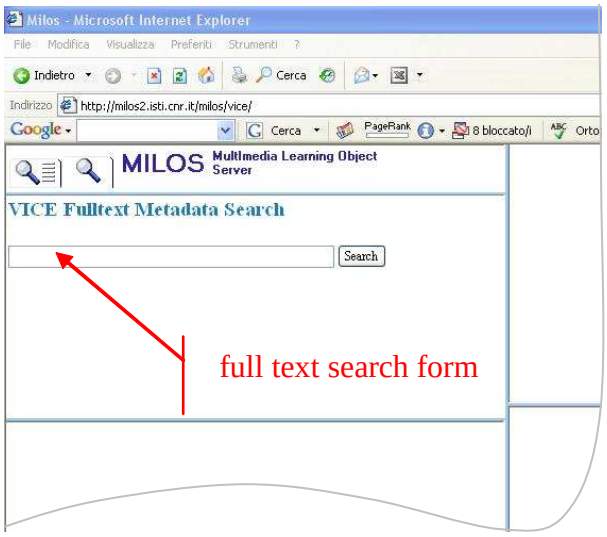

Figure 8. The query frames for the full–text search.

allow users to find lecture documents, such as videos and PowerPoint presentation, etc. within a large repository of documents, and (2) to allow user to navigate the retrieved content. The former problem is approached by allowing users to search on textual content of documents, (such as metadata, video transcript) and to search for similarity on video keyframes. The latter problem is particularly complex when the content is a video. In fact, the lecture videos captured in classrooms usually lack syntactic structure and are highly redundant in visual content. Additionally, the definition of video content differs for different presentation formats. For example, in the lecture videos of blackboard presentations, the visual content refers to the written chalk content on the board [14]. For lecture videos with electronic slide presentations (as in the context of this paper), we try to detect the changes of slides and relate slide content to video segments (OCR), event detection, and audio transcript synchronization. We then present the video as a sequence of keyframes and audio transcript over the entire timeline of the video.

The Retrieval user interface in the VICE prototypes is built up from J2EE Web Application which are displayed in a Web browser. The layout of the retrieval Web pages is controlled by a set of XML files with appropriate style sheets. This concept guarantees a maximal flexibility and extensibility. The whole retrieval interface layout consists of four parts (see Figure 9):

- a *query frame*, in which the user can formulate fielded and full–text queries (top-left frame),
- a *hitlist frame*, in which the ranked list of matching items with some basic metadata (title, type, etc.) is displayed (bottom-left frame),
- a *LOM view frame*, where the whole metadata set of the LOM description for the selected item is displayed (top-right frame),
- and a *raw media element frame*, where the details of the raw media element associated with LOM are displayed (bottom-right frame).

For each of these frames, a separate style sheet is provided, which controls the arrangement and visibility of attributes within this 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 7). 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. Beside every text input box of search of the field, a check button allows us to select the type of search to carry out (exact or approximate). In particular, selecting the MPEG-7 model we can make searches on the OCR of the keyframes of the

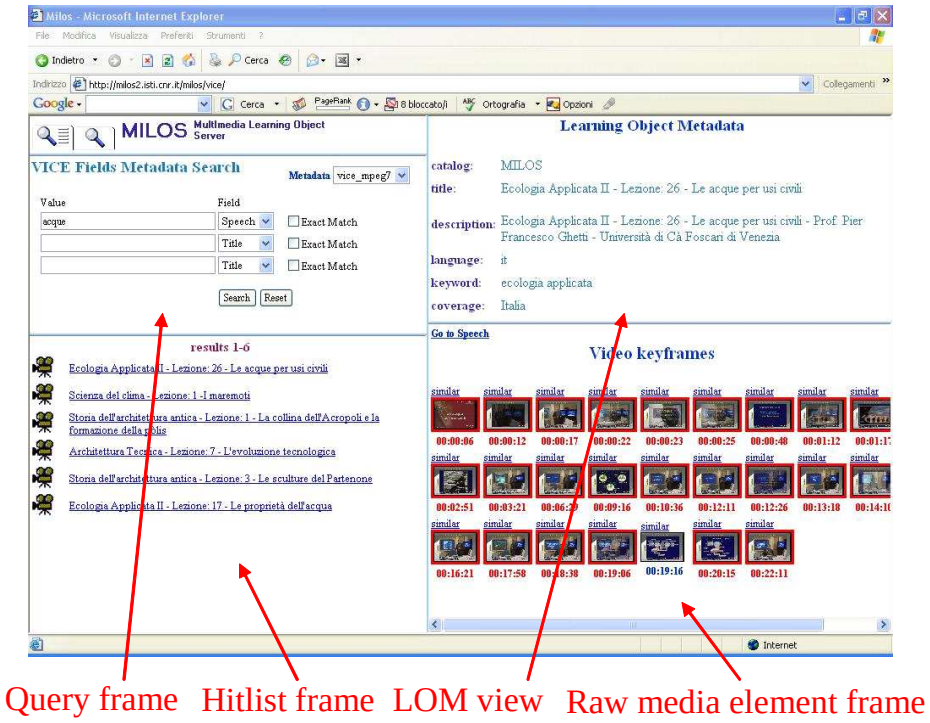

Figure 9. Visualization of the video keyframes.

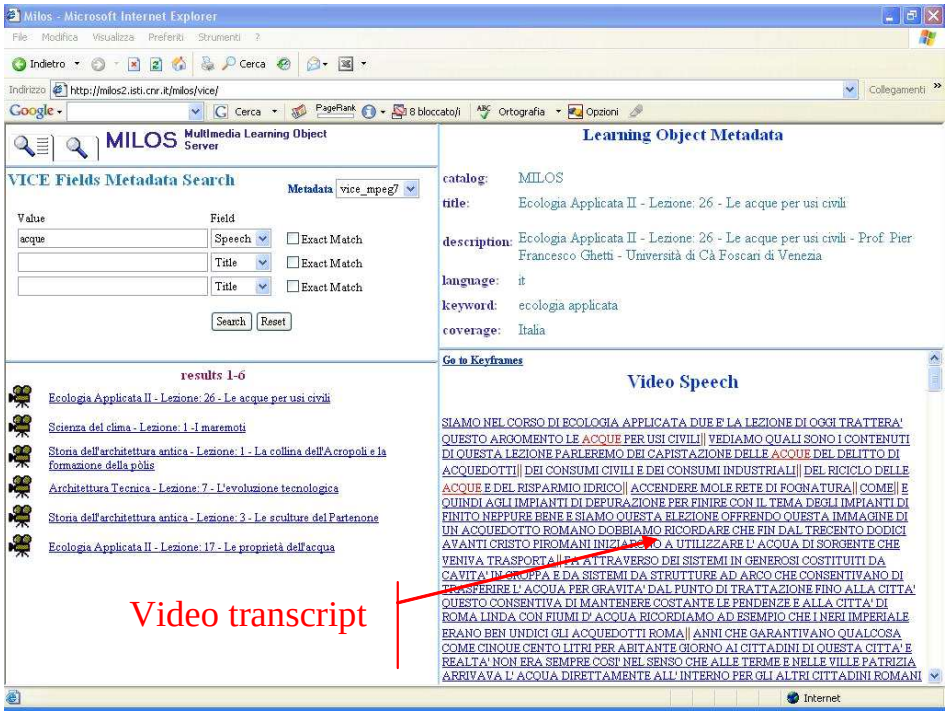

Figure 10. Visualization of the video transcript.

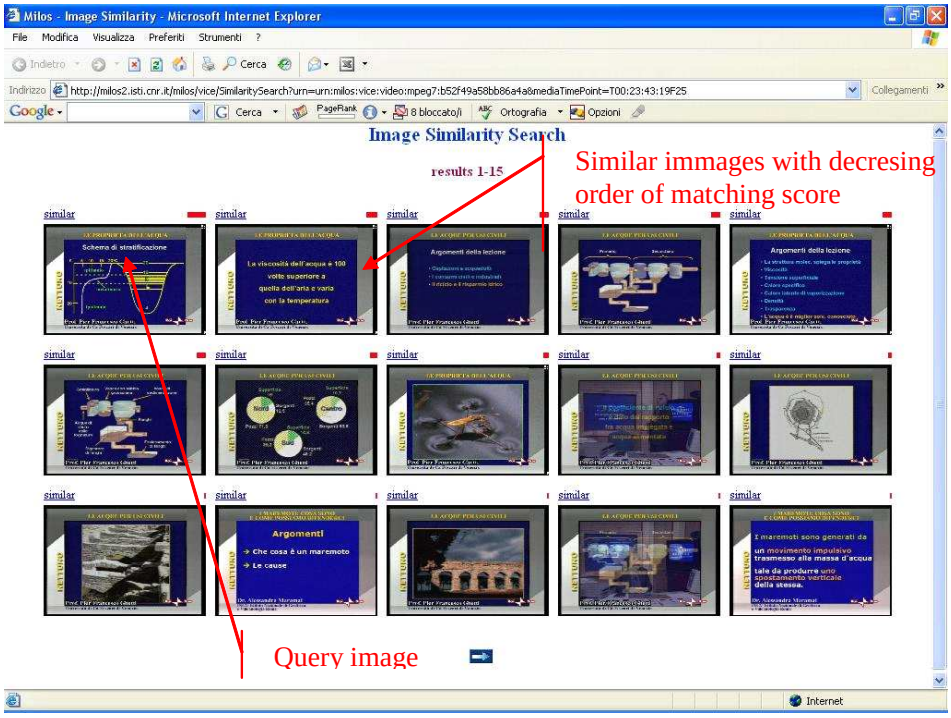

Figure 11. Visualization of the result of a similarity query over the video keyframes.

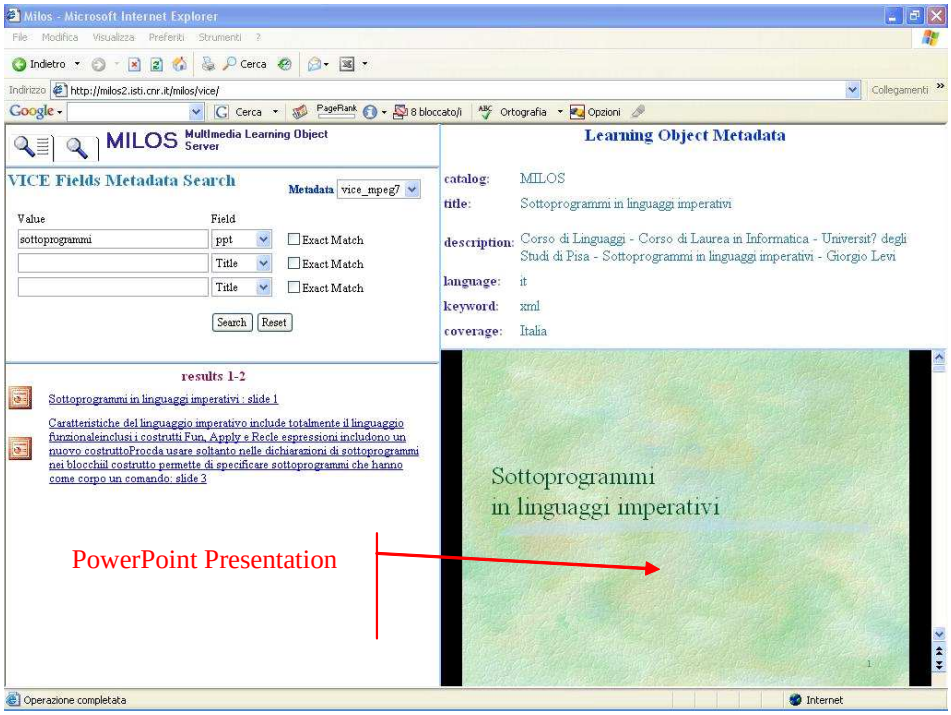

Figure 12. Visualization of the PowerPoint presentation.

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 (see Figure 8). As explained before, the hitlist frame displays all the items that satisfy the query, in a decreasing order of importance. An icon beside the items in the list reveals the type of item found (e.g., LOM description, or MPEG-7 description). In the latter case, since MPEG-7 describes the content of the associated raw media element, the icon represents its type, i.e., video or PowerPoint presentation. The two frames on the right side allow us to visualize an item of the list returned by the query. 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 (see Figure 9) and the complete transcription of speech (see Figure 10). 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. Figure 10 shows the raw media element frame when it visualizes the transcript of a video by highlighting the terms that match the query. 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 (see Figure 11). 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 (see Figure 12).

6 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 proposed the use of a digital library 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 provides annotation as metadata not only for indexing retrieval, but also for further semantic processing. Moreover, it can also be applied to other domains of digital library beyond the one of the e–Learning. A demo of the VICE search interface is available on the web at `http://milos2.isti.cnr.it/milos/vice/`.

References

[1] Nettuno. http://www.uninettuno.it.

[2] Real Time Streaming Protocol. http://www.rtsp.org/ .

[3] Mpeg requirements group, mpeg-7 overview, 2003. Doc. ISO/IEC JTC1/SC29/WG11N5525.

[4] G. Amato, F. Debole, F. Rabitti, and P. Zezula. YAPI: Yet another path index for XML searching. In *ECDL 2003, 7th European Conference on Research and Advanced Technology for Digital Libraries, Trondheim, Norway, August 17-22, 2003*, 2003.

[5] G. Amato, C. Gennaro, P. Savino, and F. Rabitti. Milos: a multimedia content management system for digital library applications. In *Proceedings of the 8th European Conference on Research and Advanced Technology for Digital Libraries (ECDL 2004)*, volume 3232 of *Lecture Notes in Computer Science*, pages 14–25. Springer, September 2004.

[6] C. Böhm, S. Berchtold, and D. Keim. Searching in high-dimensional spaces: Index structures for improving the performance of multimedia databases. *ACM Computing Surveys*, 33(3):322–373, September 2001.

[7] D. C. A. Bulterman. Is it time for a moratorium on metadata? *IEEE MultiMedia*, 11(4):10–17, October 2004.

[8] I. L. T. S. Committee. Ieee 1484.12.1-2002, draft standard for learning object metadata. http://ltsc.ieee.org/wg12/index.html, July 2002.

[9] W. W. W. Consortium. XML path language (XPath), version 1.0, W3C. Recommendation, November 1999.

[10] N. Cristianini and J. Shawe-Taylor. *An Introduction to Support Vector Machines*. Cambidge University Press, 2000.

[11] A. Hauptmann and H. Wactlar. Indexing and search of multimodal information. In *Proceedings of the 1997 International Conference on Acoustics, Speech and Signal Processing (ICASSP-97)*, April 1997.

[12] ISO/IEC. Information technology - Multimedia content description interfaces. Part 3: Visual. 15938-3:2002.

[13] ISO/IEC. Information technology - Multimedia content description interfaces. Part 6: Reference Software. 15938-6:2003.

[14] T. Liu and J. R. Kender. Lecture videos for e-learning: Current research and challenges. In *ISMSE '04: Proceedings of the IEEE Sixth International Symposium on Multimedia Software Engineering*, pages 574–578, Washington, DC, USA, 2004. IEEE Computer Society.