Content-Based Search and Annotations in Multimedia Digital Libraries

J. Aníbal Arias
Laboratory of Interactive and Cooperative Technologies, CENTIA
Universidad de las Américas, Puebla
Cholula, Puebla 72820 México
anibal@ieee.org

J. Alfredo Sánchez
Libraries Division and CENTIA
Universidad de las Américas, Puebla
Cholula, Puebla 72820 México
alfredo@mail.udlap.mx

Abstract

This paper describes a solution for the organization and management of multimedia collections in digital libraries. The Video U-DL-A (VUDLA) system allows for storage, indexing and annotation of multimedia documents in such a way that text- and image-based queries can be issued in order to retrieve specific scenes from digital video collections. Technologies such as image and speech processing, video streaming, multimedia databases, information retrieval and graphical user interfaces are integrated to produce a novel multimedia, multimodal environment which re-valuates text as an important knowledge transmission medium. We have developed a fully operational testbed to explore multimedia data properties and organization possibilities as well as a wide range of practical applications.

1. Introduction

Internet information is much more than just text: increasingly, multimedia resources are becoming available through global networks including documents with images, audio and video. At the same time, fast data access from any place at any time is motivating developments in several directions to organize huge multimedia information repositories and to make them accessible to users in a simple and effective way [15]. The challenge we are facing is to move ahead from a text-based information retrieval paradigm that has been derived from and applied to conventional databases onto new retrieval alternatives that include audio and video as data sources, and generally developing mechanisms for handling data formats with temporal properties.

In this paper we describe the design and implementation of VUDLA, a system component that endows a digital library with capabilities for managing multimedia collections. In particular, VUDLA introduces a data model for digital video as well as content-based search and annotation mechanisms. In order to extend the digital library architecture, VUDLA integrates diverse technologies such as speech and image processing, video streaming, multimedia databases and information retrieval, thus producing "knowledge centers" in which authoring, reading and viewing multimedia documents occur seamlessly along with activities such as searching, annotating and linking for all available media.

The remainder of the paper is organized as follows: Section 2 provides some additional background on the context of the project. Next, we describe the design of a video digital library and discuss the main aspects of its prototypical implementation in Sections 3 and 4, respectively. We refer to salient related work in Section 5 and, finally, we present some preliminary results and conclusions in Section 6.

2. Context

2.1 Multimedia objects in databases

For some time now, commercial database management systems have been incorporating technologies for multimedia management extending regular data models [17]. In limited ways, commercial products have proposed storage and temporal functions as well as annotating services for multimedia materials. Exploiting or customizing this functionality is not trivial, so usually projects supporting multimedia technology are based on commercial products extended with proprietary software. Moreover, projects involving multimedia management usually need to cope with growing hardware and software requirements [6].

2.2 Streaming media

Streaming media technology continuously sends data from a server to a client. Depending on available bandwidth, the server may adjust the transfer rate allowing for real-time visualization on the client side, which in this way has minimal memory requirements.

2.3 Content-based search

Traditional functions over multimedia utilize what may be referred to as their syntactical properties: moving forward and rewinding functions manipulate the temporal structure of these data types [3]. Nevertheless, particular event search or automatic summarizing requires semantic document analysis. Currently, semantic analysis is performed by humans in tasks such as interactive CD creation, but for digital libraries the amounts of data to be processed makes it necessary to define automatic

techniques for generating content representation of multimedia information. Metadata generation for multimedia data can be accomplished using speech recognition [31] and image processing as indexing technologies, whereas information retrieval models may be used to organize metadata and facilitate searching through the resulting index space.

2.4 Annotations

Enabling text and graphic annotations in video materials is important in order to analyze and share views on the contents presented in multimedia format. Simone [24] observes that writing has allowed for more articulate, refined and complex forms of expression, whereas knowledge expressed or acquired from audiovisual media tends to be less articulate and less subtle. Some of the advantages of the cognitive processes associated with reading as compared to new media are:

Pacing. In general, the pace of reading is determined by the reader ("pulling" text as desired), whereas the pace of viewing a video is decided by its author, who "pushes" images at will towards the viewer.

Corrigibility. A reader may stop at any point in time to reflect on the text just read; a viewer cannot do this easily.

Encyclopedic references. As a result of its user-controlled pace, reading allows users to stop and use complementary sources, whereas this cannot be done easily when viewing without disrupting the intended rhythm of the materials.

Citability. A text that has been read can be easily cited or even quoted. What has been viewed does not exhibit this property.

Simone explains that nowadays reading is clearly losing adherents mainly due to the effortless way in which audiovisual information is acquired ("the effort of reading cannot compete with the ease of viewing"). Linking, annotating and searching in VUDLA as an effort to combine the *sequential intelligence* derived from reading with the *simultaneous intelligence* fostered by viewing.

3. Adding video to a digital library

We have extended the architecture of a digital library to include multimedia collections and services. Our work is framed by an initiative we termed University Digital Libraries for All (U-DL-A). U-DL-A has produced a highly distributed digital library that now comprises a wide range of collections, services and user interfaces. Collections include theses and dissertations, university publications and historic archives. Services vary from information retrieval methods to agent services and interoperability mechanisms. Finally, user interfaces include personal and group spaces, visualization aids and user agents (publications concerning this work can be found at [9]). We describe the design of VUDLA next.

3.1 Digital library architecture

The general architecture of VUDLA, including the new components designed to support multimedia functionality, is illustrated in Figure 1. The architecture is an evolution of the one proposed at the inception of our digital libraries program [19] As can be observed, this is a layered, extensible, client-server architecture comprising collections, data management and modeling facilities as well as a variety of services on the server side, and various user interfaces and work environments on the client side. In Figure 1, the major components that needed to be developed to extend the existing digital library (DL) architecture are presented in a darker shade, whereas components that were adapted or used directly by VUDLA (as described in the next section) are presented in a lighter shade.

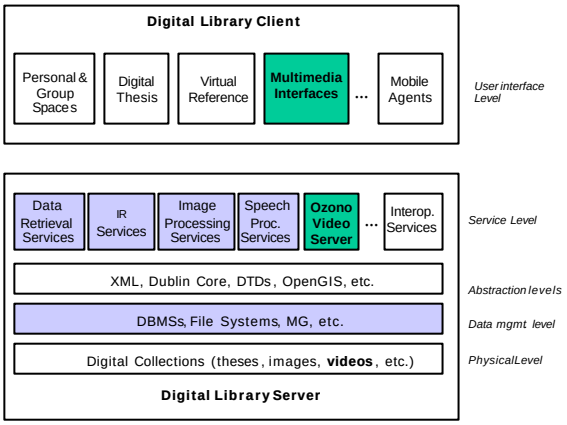

Figure 1. General DL architecture with video management capabilities.

The storage and retrieval of digital video constitute a first step towards the exploration of video in a digital library. Visualizing, controlling, annotating, querying, and linking video segments to and from other resources are among the desirable capabilities. Thus, VUDLA has been designed to include the following functionality:

- Interfaces that allow diverse users to maintain multimedia collections as well as to visualize, search, broadcast (or multi-cast), annotate, link from and refer to any portion of their contents.
- A video streaming server to provide an efficient video-on-demand retrieval mechanism.
- A speech recognition service to generate textual transcriptions of the audio track that accompanies the video (word spotting). These metadata can be used later on when searching the video collections for the occurrence of specific utterances.

- An image processing service to create vector representations from the key video frames based on their color and texture contents. This is also helpful metadata when searching the video collections for the occurrence of images or frames with certain characteristics.
- An information retrieval server to provide rich access means to the metadata generated by the speech and image processing components. Various information retrieval models should be available so the user may be presented with multiple paths to explore the multimedia collections.

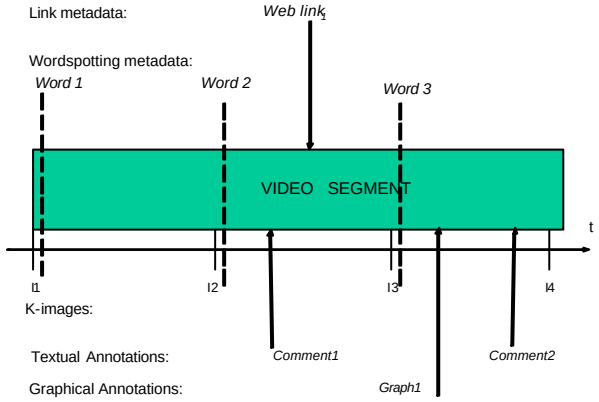

Figure 2. Metadata layers of a video segment.

3.2 Digital Video Data Modeling

There has been a significant amount of work in the area of digital video data modeling that can be applicable when defining the structure of a multimedia collection. Approaches vary in such aspects as the levels of granularity (from frames to segments) or content considerations (from purely physical features to more elaborate models that involve the notions of scenes, subjects or hyper-linking provisions). Some relevant work in this area can be found, for example, in [27] [8] [10]. Progress in modeling and manipulating multimedia collections has permeated to research and commercial database systems. In the domain-independent data model that we opted for in our DL design, multiple metadata layers can be associated with any instant of a given video segment so applications can be developed to manipulate multimedia collections in flexible and varied manners. Figure 2 exemplifies a video segment with which five layers of metadata have been associated: three words have been spotted by a speech processing service; four images have been taken from video frames and processed as color and texture vectors; a website is associated with the second word; two textual annotations have been typed and should be displayed at specific times. In addition, before the second textual comment is presented, a graph should be displayed.

4. The VUDLA prototype

4.1 Supporting Technologies

As noted previously, we conceive of the construction of VUDLA as the result of integrating various technologies. Our prototype has resulted from integrating the following:

- A graphical user interface (GUI) implementing most of the desired functionality. We relied for this on Java and its specialized classes for graphical user interfaces (Swing), Internet communication (RTP), database access (JDBC) and multimedia management (JMF) [25].
- A speech processing component for generating metadata from the accompanying audio track. Since our materials are mostly in Spanish, we used a speech recognizer based on neural networks that has been trained for Mexican Spanish [23]. This work specializes a speech processing toolkit known as CSLU [26], of which a "shell" version has been used [22]. The corpus used to train the 3-layers neural network was generated in Mexico at the speech recognition lab of the University.
- The Ozono media streaming server [13], a video-on-demand component based on Java JMF. We are using the Solaris-based version 1.1.5 of Ozono, which serves MPEG-1 digital video with guaranteed quality of service at 350 kbps.
- The Hermes Information Retrieval Server, which provides access to several information retrieval models and facilitates access to collections in the library [14]. At present, Hermes implements three models for information retrieval: vector space extended Boolean and latent semantic indexing [2].
- The Informix RDBMS and its multimedia extensions (Video and Image "DataBlades") [12] [30]. These modules provide data types and SQL extensions that allow for video management and image processing [4] using the data model described in the previous section.

We describe the operation and properties of VUDLA next.

4.2 Adding and managing multimedia content

VUDLA offers user interfaces so multimedia content can be easily added to the collections. Every time a video document is added, the speech recognition and image processing systems generate the metadata that will be used to support video querying and viewing. All the words

recognized by the speech processing system are organized as a metadata stratum, whereas image processing generates color and texture indexes from key images taken from video frames.

An interface is also available that allows specialized users to create annotations associated with stored multimedia objects. Users may add textual or graphical annotations or specify web links that will be attached to materials at specific points in time. These analytic and supplementary metadata are also organized as metadata strata.

The information retrieval component takes audio transcriptions resulting from speech processing and text annotations provided by users. Lemmatization and stopword elimination processes are applied to create indexes used for query execution.

Figure 3. Client user interface of VUDLA.

4.3 VUDLA Interfaces

The main user interface of VUDLA is illustrated in Figure 3. The three tabs at the top allow users to switch from navigating the collections to searching their contents or to receiving a multicast video for one-way videoconferencing. Though the latter is a useful feature, we focus our discussion on the first two modes (navigation and search), as they present very similar interfaces and demonstrate our approach.

The figure shows the results of searching the video collection for utterances that are related to "libraries" ("bibliotecas" in Spanish). The metadata of video segments that matched the queries, as well as the relevant time periods, are returned by the information retrieval service and listed in the top left portion of the interface. In the figure, the user has selected one of the segments and is viewing its contents, as supplied by the Ozono media server, on the bottom left panel. Naturally, the user may easily move back and forward in the video segment or control the audio volume by using the buttons and sliders displayed in this part of the interface. On the panels to the right of the main interface, three major metadata strata are displayed as they appear in the timeline when the video is played. In the figure, a textual annotation is being displayed when the video reaches a specific instant. Similarly, a reference to a website and a related image are shown. The user may choose to stop the video and check the annotations or visit the referenced websites or just continue to examine the video segment. The numbers to the right of the annotations provide handles to multiple associated metadata so they can be revisited by the user at will. The text search module uses the temporal DBMS component to manipulate time pointers to video segments. These functions allow VUDLA to establish relationships among the different metadata layers. For example, the *Overlap* function reports if two segments have a time point in common.

Web links are presented as URLs and their contents can be displayed in a web browser, according to user preferences. Moreover, when materials of the collection are visualized, all web links associated to a video document are active and can be selected through the visual component when the current displaying time matches the "validity" time of a link attached to that material. This is possible because our data model includes a metadata stratum for web links referring to video objects.

For image searching, the interface allows the user to load a target image from the local file system. This image's color and texture components are compared with those derived from the key images stored at the corresponding metadata stratum. In our current implementation, when the comparison results in a similarity measure of 95% or higher in at least one of the two criteria, the video segment associated to the corresponding key image is incorporated to the result set. The level of similarity should be tuned according to application and user requirements.

5. Preliminary results

The VUDLA prototype has been undergoing tests in two major areas: functionality and usability. In the first area, the main challenge is the integration of diverse technologies and get them to work seamlessly. In the second area, we aim at demonstrating that VUDLA enables the use of digital video in new, much richer ways when compared to its analog equivalent, and that users are able to perform tasks involving multimedia resources not only without the

limitations referred to in Section 2, but with the added value of functionality made possible only by the digital nature of the collections at hand.

5.1 Speech recognition

Tests were carried out to determine the reliability of the speech recognition system. Using a general purpose corpus of 300 voices sampled at 8 kHz in a free-noise environment, a threshold was established to make decisions from the ranking that the system gives to the words of the vocabulary. We defined a vocabulary for each video in the repository by choosing a set of words that might identify the topics in each video.

In general, with our current settings, 60% of the words in the audio track were correctly ranked (detected as part or not of the audio track) by the system and can be used for content-based queries. When analyzing recognition problems, we found that most of them were related to substitution problems due to the fact that the recognizer assigns a high ranking to words that are phonetically similar to those in the vocabulary. If more precision is needed, the speech processing component allows for finer precision tuning.

5.2 Video Streaming

Communication with the Ozono media streaming server allows VUDLA users to navigate through the video collections and within each video segment in a transparent fashion. Video is delivered in response to requests submitted from the graphical user interface and can be displayed according to specified parameters (e.g. starting at a particular point in time).

5.3 Information Retrieval

This is one of the most robust system components. Textual audio transcriptions, user annotations and other metadata are used extensively in the application of the available information retrieval models. Metadata and rankings returned by the Hermes server are displayed has allowed users to identify video segments that best fit their needs.

5.4 Image Processing

Image characteristics are highly variable and factors such as lighting, focus, or subject position or motion greatly influence any color or texture analysis. In our current implementation, color and texture comparisons have been useful only when images used in queries come directly from (or are very similar) to frames in video sequences in the database.

5.5 Usability

From the most preliminary interface designs, usability has been the most important area to evaluate [16]. Around ten important usability problems were initially detected and corrected, some of them requiring important development time but substantially improving the system's user interface. The current implementation of VUDLA has been in use for one semester. Our video collection is mostly academic, completely in Spanish, and it includes interviews with researchers from various areas, lectures and software presentations. Its use has been restricted to our research center and a few volunteer users, but is now open for beta testing at http://ict.udlap.mx/people/anibal/vudla/vudla.html.

What is most important for our research interests, we have been able to observe users interact with the new multimedia collections as we expected. It is useful to revisit the relevant media traits introduced in Section 2.4, which guided our design:

Pace. In VUDLA, the interaction with resources in all available media is completely under the user's control. The user determines the pace at which temporal media are "played".

Corrigibility. Users of VUDLA may stop at any point in time to reflect on resources being examined, regardless of their type. If audio or video are momentarily suspended, other activities may take place and streaming will continue when the user so decides.

Encyclopedic references. Not only can complementary sources be used while examining multimedia resources in VUDLA, but those resources may be associated directly with a section of a document and available immediately for perusal.

Citability. Content-based search over digital video using text or images, as well as the possibility to point to any instant within a video segment make multimedia collections in VUDLA citable and quotable.

VUDLA redefines the concepts of reading and viewing. In the digital library, reading may be an isolated activity, but the reader may also choose to be aware of the presence of other readers and decide to interact with them by taking advantage of the library's collaboration facilities. Though videoconferencing has been mentioned as one of the features of VUDLA, other communication and awareness facilities are available in the overarching digital library (as described, for example in [20] and [21]).

6. Related work

Given its integrative nature, our project builds upon a significant number of related efforts. Work on multimedia has been undertaken from several perspectives and much progress has been done in areas such as compression, transmission and storage techniques. Semantic analysis and indexing make it possible to perform content-based search and summarization in multimedia documents.

Informedia [29][28] is one of the most complex projects in the area of multimedia information management. Its first

stage also focused on the integration of speech, image and language processing for creating digital video libraries. About 1500 hours of TV news have been segmented into some 40,000 independent segments and used to explore issues in storage, retrieval and applications of digital video. A second stage of the project is focusing on summarizing and visualizing video information as well as on spatial and temporal analysis for query processing. The important ramifications and accomplishments of the Informedia initiative are difficult to assess.

The QBIC (query by image content) [5] system was developed to retrieve images based on their visual characteristics (color, texture and form). The user is able to specify search parameters in queries. For video management, the system has functions for scene detection, histogram change analysis, and camera movements.

In order to search in multimedia documents where most of the information is in audio format, researchers of the project VideoMail [11] designed a speech recognition system based in Hidden Markov Models and a text retrieval engine. The user interface presented multimedia information to the user even if all the background processing was text-based.

An interesting development in multimodal interfaces is the Music Library [32], a system with tune recognition functionality. Acoustic input mechanisms allow users to sing or whistle a melody to search through more than 10,000 traditional songs.

In [18] the author attempts to describe human behavior in a video sequence. Based upon silhouette and facial expression recognition, an estimation model is constructed to establish probabilities of what the person in the video is doing.

Our project has been strongly influenced by Informedia and other projects. We also aim at integrating technologies and generally exploring digital video as a medium with properties that are different from text, still images and analog video. However, we have emphasized issues that have not appeared in other projects or have received only marginal attention from other research groups. From technical and cultural viewpoints, we are interested in developing video corpora in Spanish and exploring mechanisms for retrieval that are particularly appropriate for our collections. From a more philosophical perspective, we are interested in the impact of the deployment of multimedia collections in the context of communities with a strong, reading-based research tradition. Our proposal ultimately aims at contributing to make digital libraries a catalyst for the transformation of work practices in knowledge-intensive activities.

7. Ongoing work

VUDLA is a relatively young project and can use that fact to its advantage, as much of the ongoing work in the area (e.g., projects referred to in the previous section) can be thought of as an opportunity for improvement. For instance, visual information retrieval does not yet produce impressive results in VUDLA. In that respect, new processing techniques are being studied and simulated, included contour and form detection. In that direction, the Video-Cuebik system [7] appears as a promising alternative to consider.

Some of the existing functionality in our prototypical implementation has ample room for improvement. For example, we plan to provide various formatting options for textual annotations that we think will improve text readability. We will also implement *frame hotspots*, which should map specific frame regions to annotations or web links. We also have started work to explore heuristics and formal models for relationships among video metadata [1].

Other scheduled evolutions are the development of a mechanism involved that allows for annotations to be kept personal, not shared for dealing with an ever-growing number of annotations and exploring new video data formats as MPEG-4, MPEG-7 and SMIL 2.0.

As mentioned previously, VUDLA was originally developed using the Informix RDBMS and its extensions for multimedia management. This made our prototype strongly dependent on proprietary data types and functions. In order to generalize the operation of the system so any RDBMS can be used, we are in the process to migrate our current implementation to MySQL. The process required the implementation of functions at the various architectural levels. About 80% of the functionality available in the Informix-based version is now operational with MySQL.

Overall, we think of VUDLA as a testbed to explore the potential of multimedia digital libraries. Systematic user involvement and observation, corpus construction, and technology integration are driving issues in our agenda.

8. Conclusions

Multimedia collections are increasingly popular and in many ways becoming a fundamental means for knowledge dissemination. New digital video collections are created every day and occupy huge amounts of disk space. Novel digital libraries integrate technologies and provide interfaces for seamless media integration that foster rich interactions between users and media and among media themselves. We have described VUDLA, a research and development project that integrates image and speech processing with database and information retrieval techniques to produce a DL with desirable media properties.

One of the concerns regarding the prevalence of video and the abandonment of reading is that video appears as a regression from the structured, sequential reasoning promoted by reading to more primitive, sensorial forms of intelligence. In VUDLA multimedia resources are experienced by the user's senses but associative thinking is also promoted. While still some of the mind structuring properties of knowledge construction based primarily on reading may be at risk, there sure are new, more powerful cognitive processes triggered by multimedia DLs. Our work makes a contribution to the study of the properties and means for organization, manipulation and presentation of multimedia collections in digital libraries.

9. Acknowledgments

We would like to express our appreciation to Hugo López and Arturo Lazo of our Videoconferencing Department for their support in the construction of the Ozono Media Streaming Server. We also thank our colleagues in the Tlatoa Speech Processing Lab, whose support allowed us to use the CSLU Toolkit. We appreciate the colaboration of Fernanda Maldonado in the design and implementation of the Hermes IR Server.

This work has received financial support from the National Council of Science and Technology (Conacyt Projects No. 35804-A and G33009-A).

10. References

[1] Ahanger, G., Little, T.D.C. Data semantics for improving retrieval performance of digital news video systems. Database semantics. Kluwer Academic Publisher, 1999, 47-64.

[2] Baeza, R. and Ribeiro-Neto, B.. Modern Information Retrieval. Addison Wesley/ACM, 1999.

[3] Bernard, M., Dubois F. An Agent-based Architecture for Content-Based Multimedia Browsing. Intelligent Multimedia Information Retrieval. The MIT Press. Bedford, Mass., 1997, 281-294.

[4] Del Bimbo, A. Visual Information Retrieval. Morgan Kauffman Publishers, 1999.

[5] Flickner, M., Sawhney, H., Niblack, W. Query by Image and Video Content: The QBIC System. IEEE Computer, September 1995.

[6] Grosky, I. W. Managing Multimedia Information in Database Systems. Communications of the ACM, December 1997, 72-80.

[7] Hauptmann, A., Papernick, N. Video-Cuebik: Adapting Image Search to Video Shots. In Proceedings ACM JCDL'02 (Portland, Oregon, July 13-17), 2002, 156-157.

[8] Hjesvold, R., Midstraum, R. Modeling and querying video data. Proceedings of the 20th VLDB Conference, 1994, 686-694.

[9] ICT. Laboratory of Interactive and Cooperative Technologies. Universidad de las Américas, Puebla, México. 2002. http://ict.udlap.mx.

[10] Jiang, J., Elmagarmid, A. K. Spatial and temporal content-based access to hypervideo databases. The VLDB Journal 7, 4, 1998, 226-238.

[11] Jones, G. and Foote, J. Video Mail Retrieval using voice: An Overview of the Stage 2 System. In Proceedings of the MIRO workshop, University of Glasgow, 1995.

[12] Informix. Informix Video Foundation DataBlade Module User's guide. Informix, 1998.

[13] López, H., Lazo, A. The Ozono Digital Media Server. Internal Report. Special Projects and Videoconferencing Department. Universidad de las Américas, Puebla, México. 2002. Available from http://www.udlap.mx/~proyesp.

[14] Maldonado, F. Hermes: Servidor y biblioteca de modelos de recuperación de información. B. Eng. Thesis. Universidad de las Américas, Puebla, December 2002.

[15] Maybury, M. T. (Editor). Intelligent Multimedia Information Retrieval. The MIT Press. Bedford, Mass., 1997.

[16] Nielsen, J. and Molich, R. Heuristic evaluation of user interfaces. In Proceedings ACM CHI'90 (Seattle WA, April 1-5), 1990, 249-256.

[17] Ozsu, T. M. Issues in Multimedia Database management. http://www.cs.ualberta.ca/~database/multimedia.

[18] Pentland, A. Perceptual intelligence. Communications of the ACM 43(3), 2000, 35-44.

[19] Sánchez, J. A., Leggett, J. A. Agent services for users of digital libraries. Journal of Networks and Computer Applications 21, 1, 1997, 45-58.

[20] Sánchez, J. A., García, A. J., Proal, C., Fernández, L. 2001. Enabling the collaborative construction and reuse of knowledge through a virtual reference environment. Proceedings of the Seventh International Workshop on Groupware (Darmstadt, Germany, Sept. 6-8). IEEE Computer Society Press, Los Alamitos, Calif. 90-97.

[21] Sánchez, A., Proal, C., Carballo, A., Pérez, D. Personal and Group Spaces: Integrating resources for users of digital libraries. Proceedings of the 4th Workshop on Human Factors in Computing Systems (IHC 2001, Florianópolis, SC, Brazil, Oct. 15-17).

[22] Schalkwyk, J., Colton, D., Fanty, M., The **CSLUsh** toolkit for automatic speech recognition, Technical Report No. CSLU-011-96, Center for Spoken Language Understanding, Oregon Graduate Institute of Science & Technology, 1996.

[23] Serridge, B., Cole, R., Barbosa, A., Munive, N., Vargas, A. Creating a Mexican Spanish version of the CSLU toolkit. Proceedings of the International Conference on Spoken Language Processing (ICSLP), Sydney, Australia, 1998.

[24] Simone, R. *La Terza Fase.* Laterza. Rome, Italy, 2000.

[25] Sun. Java Media Framework API Guide. Sun Microsystems, 1999.

[26] Sutton, S., Cole, R., de Villiers, J., Schalkwyk, J., Vermeulen, P., Macon, M., Yan, Y., Kaiser, E., Rundle, B., Shobaki, K, Hosom, P., Kain, A., Wouters, J., Massaro, M., Cohen, M. Universal Speech Tools: the CSLU Toolkit. In Proceedings of the International Conference on Spoken Language Processing (ICSLP), Sydney, Australia, November 1998, 3221-3224.

[27] Tusch, R. Kosch, H., Böszörményi, L. VIDEX: an integrated generic video indexing approach. Proceedings of the eighth ACM International Conference on Multimedia, 2000, 448-451.

[28] Wactlar, H. D. Informedia - Search and Summarization in the Video Medium. In Proceedings on Imagina 2000 conference (Monaco, January 31-February 2), 1-10.

[29] Wactlar, H. D., Kanade, T., Smith, M. A. Stevens, S. M. Intelligent Access to Digital Video: Informedia Project. IEEE Computer 29, 3 (May), 1996, 46-52.

[30] Westermann, U. and Klas, W. Architecture of a DataBlade Module for the Integrated Management of Multimedia Assets. In Proceedings on Multimedia Intelligent Storage and Retireval Management (MISRM, Orlando, Florida), October 1999.

[31] Witbrock, M. and Hauptmann, A. Speech recognition for a digital video library.http://www.informedia.cs.cmu.edu.

[32] Witten, I., McNab, R., Jones, S. Managing Complexity in a Distributed Digital Library, IEEE Computer 32(2), 1999, 74-79.