

**Threading Information Pathways  
Through NSDL Video**

**NSF Proposal 85834  
Final Report  
For the period September 2000-2002**

**Carnegie Mellon University  
School of Computer Science  
Pittsburgh, PA 15213-3890**

**Principal Investigator:**     **Howard Wactlar**  
School of Computer Science

**Co-PI's:**                   **Michael Christel**  
                                  **Alexander Hauptmann**  
Computer Science Department

# 1 Overview

The National Science, Mathematics, Engineering, and Technology Education (SMETE) Digital Library (NSDL) is anticipated to evolve into a federated repository of quality teaching and learning SMETE resources appropriate for educators and learners at all levels. This repository will naturally include video resources, considering the capability of the media to engage, inspire or immerse the viewer in the subject matter. However, information discovery and quality assessment with respect to video can be frustrating and error-prone due to insufficient and inexact content-based indexing of video. Also, video's temporal dimension often leads to time-consuming inspections where an hour-long video may take an hour or more to review. Most importantly, the intellectual effort applied in creating a new overlay of annotations and sequences of the video information is lost not only to other patrons, but often to the author as well over time. This project proposes to deliver services to the NSDL enabling it to better create, organize and deliver video to its user communities. These services build from prior success with automated video processing, retrieval and summarization while leveraging from newly established World Wide Web Consortium recommendations and the creation of better dynamic annotation management services.

Web/Internet Site: <http://www.Informedia.cs.cmu.edu/nsdl>

## 2 Activities and Findings

### 2.1 Open Archives Initiative

The Open Archives Initiative (<http://www.openarchives.org>) developed a protocol "Open Archives Initiative Protocol for Metadata Harvesting" (<http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm>) which allows repositories to make their collections available to the world at large. The protocol is html-based and lets others look in and see what information is being published. We made our NSDL video collection available to others via this protocol. Open Archives sites can look at the information that we publish, and Open Archives users and/or web spiders will index our materials. Our site is accessible via <http://infsearch.cs.cmu.edu/cgi-bin/oai.pl> but is only directly useable via protocol-compliant requests (a simple tool like [http://purl.org/net/oai\\_explorer](http://purl.org/net/oai_explorer) can do this.)

We interface our Open Archives data with the protocol using a perl toolkit from Hussein Suleman at Virginia Tech. We developed software to take the requests made by Open Archives users, and query our database to return the correct answers. We registered with the [http://purl.org/net/oai\\_explorer](http://purl.org/net/oai_explorer) and also with the Open Archives. Both do acceptance tests to assure that compliance with the protocol and to handle "odd" cases. We are now an Open Archives repository, with 1075 records that describe segments from 80 separate videos. For each segment (as part of the protocol), we provide a text transcript, keywords, etc., as well as a Windows media url that will allow one to play that segment of the video.

### 2.2 Web formats promoting flexible presentations of video metadata

The Informedia Project is migrating to an XML/XSLT delivery infrastructure, where users via web browsers will be able to query the library and retrieve data in XML format corresponding to the metadata associated with video in the library. Depending on user locale and the rights associated with the video clips, the video itself may also be accessible. The architecture of the Informedia library system is being re-engineered, migrating from a two-tier client-database server

system to a multi-tier client: user interface, presentation transformations (e.g., XSL to convert XML into HTML presentation for web-based client), video library logic (e.g., ideal use of shot images within slide shows and synchronized presentations containing text fragments), and database.

Scalable Vector Graphics (SVG) [refer to <http://www.w3.org/Graphics/SVG>] is a language for describing two-dimensional graphics in XML. Map and “visualization by example (VIBE)” interfaces for video exploration have been built as SVG, generated from the same XML data representing a set of video from the Infromedia library. The use of XML, XSL, XPath, SVG and other W3C recommendations allows client browsers to access the Infromedia library and present multiple summaries of the video. Through client-side transformations of XML into HTML, SVG, or other formats, a user can manipulate summaries, i.e., video collages, dynamically to explore features of interest specific to his or her needs, without burdening the Web server or requiring further download time overhead. See [Christel, Maher and Begun] for details.

### **2.3 Improved Web Accessibility**

A major problem associated with the web client was that loading of the mpeg files was extremely slow, and hence any system using this was impossible to use over a slow connection. Thus, streaming video files were used. These, though, came with their own set of problems. The strength of the Infromedia System comes from the plethora of interactions developed to provide access to the video beneath the interface. Most of these interactions provided a quick way to reach a certain point in the video. Streaming files buffered every time they moved to a new point, making it tedious for people with fast connections. Thus we needed to design the system such that the same interface would be able to provide both mpg and wvx version to the end user.

The concept of usability was integrated into the system from the very beginning and several formal and informal usability tests were done to determine what worked and what did not. Some of the interactions, hitherto hidden, were displayed more prominently to the user. New interactions were added and old and unused interactions removed or replaced, based on user feedback and response.

The files were created so that it would be easy to scale up the system to integrate more videos without too much work. Common interfaces for handling both mpg and streaming video were developed, and an architecture was built that was modular, flexible and scalable. Most of the code made use of XML to store the information returned from the database, making it easy for the information to be presented in a variety of ways. This was crucial to the several interaction techniques that we used in the video player.

Most features from the Visual Basic client regarding video playback were integrated with the web client. These included video controls, result visualization, and shot break image techniques. Each video that was initially presented to the user was a segment of the full video. Video controls allowed the user to view the full video as well as the sibling segments besides being able to control the standard features like volume, position, pause, play, etc. Like the VB Client video player, the web client video player sported a transcript. As the video progressed the corresponding text was highlighted. A major addition was that the users could click anywhere on the transcript and the video would jump to the corresponding point. Furthermore, words in the search query were appropriately highlighted.

The search result was also displayed in other forms. One of the representations allowed the user to repeatedly click on the word and cycle through all its instances in the video. When the user

clicked on the word, the video would start playing from that point onward. Another representation presented the occurrences of the result words in appropriate colors (matching the color highlights in the transcript) in the form of vertical lines on a timeline representing the video. The user could click on these vertical lines to jump to a particular instance of the search word.

Each segment is composed of several key frame images representing a short section of the segment. These key-frame images were integrated with the video player to allow the user to browse through the video visually, sort of like a visual transcript. When the user clicked on the images, the video would start playing from that point onwards. Another feature integrated into the video player allowed the user to annotate the video at desired time points. The user could create notes that could be either public or private and accordingly the notes could be shared. Any content added in the note was immediately searchable, thus enabling them to use notes as bookmarks. The video segments themselves were bookmarkable, though the results content was lost.

## **2.4 Explicit Annotation Mechanisms**

Users may want to mark portions of video for personal use, for passing on to others for a specific purpose, or for better documenting the video presentation for general access. Annotations are one of the three types of metadata discussed by Carl Lagoze in the context of NSDL (the others being "item" and "collection"; see "Presentations by Technical Infrastructure team" at <http://www.smete.org/nsdl/meetings/tech1001/> for the full slide set (URL <http://www.smete.org/nsdl/meetings/tech1001/technicaloverview1001.ppt> ). More information about metadata for NSDL in general is found in the NSDL Metadata Primer by Diane Hillmann at <http://metamanagement.com.nsdl.org/outline.html>. In the video context, support has been added to allow text annotations to be created and targeted toward a specific temporal sequence of video. These annotations are immediately added to the word indexing service for the video collection so that the annotation can be searched by contents, author, or both to find video resources. Annotations can be public or private. For example, you can search for all videos annotated publicly by a particular person, for others' public annotations with the words "Jupiter gravity", or for your own annotations including the word "physics." A teacher might note that a NASA video is of use for a particular purpose, and add the annotation "shows flips in low gravity environment, useful for teaching physics concepts; tie into rotational velocity demo." Another library user searching on "rotational velocity" will then retrieve this NASA video based on the added annotation text. Annotations hence enrich a video library with additional descriptors added by the community of users.

## **3 Contributions to Resources for Research and Education**

The National Science, Mathematics, Engineering, and Technology Digital Library (NSDL) will contain a wealth of data, including video. The services being developed here enable NSDL patrons to efficiently locate video resources and meld them into compositions that effectively support particular objectives. Annotations will capture the investment made in creating new pathways through the video information, enabling continued publication, reuse, and revision of that work for the benefit of the broad NSDL community.

Video is a rich medium for communicating visual, time-dependent phenomena, and for providing real-world footage capable of illustrating and motivating math and science concepts. Vast collections of video have captured field studies and experiments, documented discoveries in space and throughout our planet, and recorded events in our world and in micro-environments not accessible to the human eye. However, despite their potential for use in educational settings,

video resources are often discounted or overlooked by science, mathematics, engineering, and technology (SMETE) educators and students. Barriers include:

- Loss of investment made by authors of video compositions who blaze pathways through the video information space; these pathways remain unmapped and hence undiscovered by other information foragers with similar requirements.
- Lack of support to tailor video resources to specific needs.
- Frustration in searching and browsing video, as much time is invested in viewing numerous video clips to gauge their relevance.
- Inability to locate pertinent video material, due to insufficient indexing of its contents.

These barriers can be overcome by capturing and managing the threads of video information access, use, and reuse within the NSDL. Specifically, this work will provide the following services:

- Creation and organization of annotations for video compositions and information pathways, enabling a dynamic information repository where one's diligent work in producing a stellar video lesson plan can be recognized, rewarded, archived, and reused in future overlays of video information.
- Support for composition of video lesson plans and multimedia essays from component clips meeting the time, message, and pedagogical requirements of the NSDL patron.
- Explicit video annotation mechanisms, whereby NSDL patrons can access reviews and other commentary aligned and synchronized with video resources.
- Implicit annotation mechanisms for video, allowing information retrieval schemes with relevance judgments based on access frequency and incorporation of video resources into derivative works.
- Enhanced content-based video search functionality derived from the integration of speech recognition, language processing, and image processing automated techniques.

The project focus on video complements the research of others focusing directly on the text or image domains. Carnegie Mellon University is uniquely positioned to pursue this work, given its continuing legacy of speech recognition, image processing, and information retrieval achievements and the past accomplishments of the Informedia Project.

#### **4 Outreach Activities**

We demonstrated the Informedia system to public broadcasters and university extension programs, discussing the automatic derivation of metadata for video enabling full-content search and retrieval, and the special interface needs for dealing efficiently and effectively with growing amounts of metadata for video collections that are growing as well. Michael Christel was an invited panelist to the "Reforging the Links"; Workshop at the University of Wisconsin Sept. 10-11, funded by the Television Future Fund of the Corporation for Public Broadcasting (<http://reforginglinks.uwex.edu/>). This workshop discussed the general theme of links between public television and the university in the digital age, with emphasis on digital video and digital asset management. There is potential for some educational video produced by public television stations to perhaps be processed through Informedia technology and have its usage tracked as part

of our NSDL video service contract.

Informedia was mentioned in the article "The Technology Review, Emerging Technologies that will Change the World: Data Mining", by M. Mitchell Waldrop, which appeared in MIT's Technology Review, Annual Innovation Issue Vol. 104, No. 1, February 2001. This article discusses rising technology fields, in this case video mining, that the magazine predicts will have a profound impact on the economy and on how we live and work: "Another hot area is video mining: using a combination of speech recognition, image understanding and natural-language processing techniques to open up the world's vast video archives to efficient computer searching. For instance, when Carnegie Mellon University's Informedia II system is given an archive of, say, CNN news clips, it produces a computer-searchable index by automatically dividing each clip into individual scenes accompanied by transcripts and headlines.";

MIT's Technology Review also mentioned Informedia in the July/August 2001 issue, in an article by David Voss entitled "Upstream: Video Searching"; The article addresses some issues that complicate video retrieval in an increasingly cluttered digital space: "At Carnegie Mellon University, researchers are creating a digital library that combines natural-language processing, speech recognition and image analysis. 'The integration of these different technologies is the key,' says Howard D. Wactlar, director of the Informedia Digital Video Library Project at Carnegie Mellon. A prototype captures news broadcasts from around the world and stores them, along with summaries or storyboards. Someone can then type in a question, or just utter the question aloud: 'Tell me about oxygen problems on the Russian space station Mir.' All the relevant news clips are displayed as frame icons you can click on. The system is also incorporating face recognition to make it possible to call up all the clips of a particular person."

A new company (Vivisimo, Inc.) has spun out of the Informedia project and CMU's Computer Science Department, with investment from the Commonwealth of Pennsylvania and a grant from the NSF SBIR program. The company is commercializing "document clustering" technology that: 1) gets inserted in a flow of search results, and 2) groups the search results into spontaneous hierarchical folders that are annotated with concise, informative labels. The mission of the company is to change how search results are delivered on computer screens everywhere, by replacing the current long, tedious lists with well-sorted, hierarchical folders.

## **5 Training and Development**

Based on Informedia DLI-2 work and research directions, updated and delivered "Multimedia", course 20-791, as part of the Carnegie Mellon E-Commerce Master's Program (<http://www.ecom.cmu.edu>), the first E-Commerce degree program in the country. Course Description: Until recent years, most computing tasks dealt with numerical, text, and symbolic data, and Computer Science has emphasized these data types. Digital representations of audio, video, and images are now becoming quite common. With the advent of relatively cheap, large online storage capacities, network transmission speeds and advances in digital compression, comprehensive sources of multiple media (Text, Image, Video and Audio) can be easily stored and made available. Collecting and intelligently integrating these multiple media opens up opportunities for novel business applications. Consequently, an understanding of multimedia is essential for many e-commerce businesses.

This course teaches students to work with multiple media on computers. Students learn the issues involved to capture, process, compress, search, index, store, and retrieve various kinds of continuous media. Projects require work with audio, scanned images, digital video, and other media, all in digital form. Readings and lectures provide a conceptual and technical framework

for multimedia work. After completing this course, students will be able to:

- Appreciate the role of multimedia in E-Commerce
- Understand the concepts underlying multimedia creation, representation and transmission
- Create media for the Web using various software tools to manipulate audio, images and video

## **6 Collaborators or Contacts**

We collaborated with the Indiana Center for Database Systems, in particular the Indiana Telemedicine Incubator (ITI) Project's digital video system for medical education. This EduMed system has a repository of segmented and annotated medical videos which first and second year medical school faculty and students can browse, retrieve, and play. We generated metadata automatically through Informedia processing for five representative videos from this collection, and shared that with the ITI staff. The ITI staff was primarily interested in Informedia speech recognition and text metadata derivation for content-based indexing, along with segmentation issues. This collaboration may result in improved processing for video in the medical education domain.

## **7 Publications**

Christel, M.G., Cubilo, P., Gunaratne, J., Jerome, W., O, E.-J., and Solanki, S., "Evaluating a Digital Video Library Web Interface", (2002). Proceedings of the ACM/IEEE-CS Joint conference on Digital Libraries. Portland, OR, July 2002.

Christel, M. Maher, B. Begun, A., "XSLT for Tailored Access to a Digital Video Library", (2001). Proceedings, Published Collection: Proceedings of the First ACM+IEEE Joint Conference on Digital Libraries, Roanoke, VA, June 2001.

Christel, M. Huang, C., "SVG for Navigating Digital News Video", (2001). Proceedings, Published Collection: Proceedings of the ACM Multimedia 2001, Ottawa, Canada, September 30 - October 5, 2001.

Jin, R. Hauptmann, A., "Title Generation for Machine-translated Documents", (2001). Proceedings, Published Collection: Proceedings of the Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI01), Seattle, WA, August 4 - 10, 2001.

Pan, J. Faloutsos, C., "Videograph: A New Tool for Video Mining and Classification", (2001). Proceedings, Published Collection: Proceedings of the First ACM+IEEE Joint Conference on Digital Libraries, Roanoke, VA, June 2001.

## **8 Presentations**

Oct 2000 Wactlar: "Automated Indexing - New Theory and New Tools in Speech Recognition", Invited talk at FIAT/IFTA World Conference, Vienna, Austria.

Oct 2000 Hauptmann: "Multimedia Information Retrieval from a Digital Video Library", Invited talk, Sixth International Workshop on Multimedia Information Systems, Chicago, Illinois, October 26-28, 2000.

Dec 2000 Hauptmann: "Multimedia Video Mining", Invited workshop speaker, DaimlerChrysler, Ulm, Germany, December 2000.

Feb 2001 Christel: Steven R. Vedro, University of Wisconsin, Telecommunications based Strategies for Education, Economic and Community Development. Research discussion.

Sep 2001 Christel: "Informedia Research Initiatives: What to do with All the Metadata", for the Television Future Fund of the Corporation for Public Broadcasting's "Evolving the Links" workshop on Digital Asset Management, Sept. 10-11, 2001.