

Implementing an Educational Digital Video Library Using MPEG-4, SMIL and Web Technologies

Marcelo Milrad, Philipp Rossmannith and Mario Scholz

Center for Learning and Knowledge Technologies (CeLeKT)

School of Mathematics and Systems Engineering

Växjö University, 351 95, Växjö, Sweden

marcelo.milrad@msi.vxu.se

philipp.rossmanith@msi.vxu.se

mario.scholz@msi.vxu.se

ABSTRACT

This paper describes the results of our efforts with regard to the design and implementation of an educational digital video library using MPEG-4 and the Synchronized Multimedia Integration Language (SMIL). The aim of our work is to integrate MPEG-4 encoding, full text indexing, high-resolution streaming, and SMIL, not only for delivering on-line digital video, but also for enabling content-based search for particular segments of a video clip stored in a repository of educational digital videos. One of the main purposes of the system is to provide new functionalities and solutions, which are not offered in conventional video libraries without online distribution facilities. Our system allows teachers, students and other users from 145 schools in our region, quick and easy access to a digital video repository via the Internet. They are able to store, search and retrieve catalogued streaming digital video content to be used for educational purposes.

Keywords

Educational digital video libraries, MPEG-4, SMIL, digital video retrieval

Introduction

In the past decade, the Internet has spawned many innovations and services that stem from its interactive character. There are numerous indications that the ongoing process of adding mobility to interactivity will transform the role of the Internet and pave the way for yet another set of innovations and services. The XML-based Synchronized Multimedia Integration Language (SMIL), for instance, is devised for the distribution of sophisticated multimedia content in a variety of devices, ranging from stand-alone computers to cellular phones (Bulterman & Rutledge, 2004). Diverse multimedia applications have flourished with recent advances in hardware and network technology, the proliferation of inexpensive video-capture devices, and widespread adoption of the worldwide web.

Video content can significantly enhance the learning and communication experience. When properly linked to text, charts and images, video provides the realism, interest and detail not available in other media (Jonassen et al., 1999).

All these new forms of interactive multimedia and communication offer new possibilities as to the way we learn, think, and communicate. Even if the Internet and other related technologies provide easy access to many resources in the form of static or dynamic web pages, it is undoubtedly more difficult to access high quality videos or film clips on the web. To our knowledge, there are not many databases of educational digital videos that can be accessed on computers via the Internet. Thus, our work is an attempt to tackle the problem of web-based video retrieval to be used for educational purposes.

The purpose of our project is to develop an inexpensive, efficient, and easily accessible on-line digital video library of educational videos. The system provides teachers and students in 145 local schools with on-line access to a video repository made available and maintained by Audio-Visual Media Center (AV-Media), a regional educational centre. Teachers and students are now able to store, search and retrieve catalogued streaming content, and stream specific video segments.

Using a combination of MPEG-4 encoding, SMIL and underlying metadata descriptions, the resulting system allows the semantic search of video content whilst adapting dynamically to the client's bandwidth. This enables users to view video material adapted to their individual needs, in a format adapted to their particular environment and connectivity. We provide different encoding in different qualities to support wide variety of clients. MPEG-4 is used to encode multimedia content in order to offer a better quality at the same bit rate. SMIL is used because

it enables random access to different points in the timeline of video content. We consider this solution superior to approaches utilizing segmentation since in those cases users have only specific, previously determined access points.

In the following sections, we will describe in more details the problems we are trying to overcome, the rationale of our design and our technological approach. We will conclude by describing the architecture of our system and the outcome of our work. These results are based on the efforts we have conducted in this project during the last two years.

Motivation and Rationale

AV-Media is a regional educational center located in the province of Kronoberg, in southeast Sweden. AV-Media has a wide collection of books, films, videos (VHS), CDs, and recently DVDs. The main task of this organization is to give access and distribute all these different educational type of media to the 145 schools in the region. In particular, AV-Media offers a big collection of VHS films consisting of more than 6000 titles, many of those produced by the Swedish Educational TV. Educational experts review each video to ensure their suitability for educational purposes before the center purchases them. The videos are then archived and the related information about each film is stored in a database. The material can be ordered and distributed to the different schools in a number of different ways.

Teachers in need of a particular educational video call to AV-Media's booking unit. Teachers also have the option to search a database containing information about all available titles through AV-Media's web site. Teachers calling the center can ask for advice on the type of the video and its content. They can also come to the center, preview the video, and get professional support. The center also provides service cars that deliver videos at different schools. The schedule of this service is available online. It is easy to see that present video distribution involves many people and is very expensive. Thus, there is a need to improve the way the educational material is stored and distributed.

New advances in digital video techniques and broad band distribution channels make it possible to explore new ways of creating, processing and distributing educational video material to schools. It is now possible to use existing open standards to compress (Sikora, 1997), play back, index and annotate (Manjunath et al., 2002) and distribute multimedia stream data. Our efforts are primarily motivated by the need to provide access to digital video segments to a wide variety of users, to allow them to look for particular sequences and to improve the way this educational material is distributed. Thus, one of the main objectives of the project is to create a repository of streaming videos for K-12 teachers and students.

Our work also focuses on organizing and indexing videos. Teachers and students are now able to search for, retrieve, manage, and share digital video for use in the classroom. The system is accessible through a web interface via internet. It contributes to the development of the educational community in the region of Kronoberg, by providing online access to new resources and tools for the classroom, thus eliminating some of the barriers of time and distance as described above.

Related Work

Researchers have now realized that while an enormous amount of unstructured video data exists, and its use as a data source in many fields has greatly increased, there are several difficulties involved in its manipulation and retrieval. MPEG-4 is a relatively new, open standard for compression and delivery of high quality audio-visual multimedia applications that addresses scene content as a set of audio-visual objects (Sikora, 1997).

There are two main approaches for digital video retrieval. The first, content-based video retrieval (Marchand-Maillet, 2000), deals with low-level features of content - such as color histograms, motion, texture and shape. This approach uses automatic means to extract content features, but is not on a semantic level. The second technique tries to enable users to search by semantic concepts. The advantage is that this is much closer to the way users think of video (Vailaya, et al., 2001), but in order to achieve satisfying results manual creation of metadata is needed for indexing; a subjective and time consuming, thus expensive process.

Despite the increasing amount of research in the domain of image recognition (Martinez & Serra, 2000; Mojsilovic & Rogowitz, 2001) the results lagged behind expectations. Thus, recent research suggests that the

combination of the approaches described above will generate better results (Li et al., 2003). The recent metadata standard MPEG-7 (Manjunath et al., 2002) also targets both high- and low-level metadata. Extensive research has also been carried out by IBM's Cue Video project to study various aspects of segmentation, automated video indexing (including audio segmentation and speech recognition), browsing by generating compact video previews (including storyboard, animation), slide show of key frames, and retrieval and time scale modification for fast video browsing in the application domain video for training and education (Amir et al., 2001).

Takeshi and colleagues (2002) elaborate upon work related to designing mobile streaming media using Content Distribution Network (CDN), a scheme which pushes multimedia content to the Internet and enhances streaming media quality for mobile clients while utilizing network resources effectively and supporting client mobility in an integrated and practical way applying segmentation, request routing, pre-fetch control, and session handoff. Perhaps, the closest effort related to our work in this project regarding streaming and indexing, has been carried out by Hunter & Little (2001). In their work, they used a combination of high level and low level indexing for composite mixed-media digital objects and MPEG-1 for video streaming. In the coming sections, we describe in details the rationale of our design and technological approach with regard to the educational digital video library we have implemented.

Features and Functions of the System

As indicated earlier, our project aims to provide a streaming environment that offers every school a simple, fast and easy online access to streaming media allowing also semantic, high-level search for content. We are using streaming and archiving techniques in a system that adjusts and adapts itself to the available client bandwidth dynamically (using SMIL). The system we developed is able to identify the available bandwidth of different clients. It can then adapt to changes in network performance and client characteristics: each video clip is encoded in several different qualities. Based on the connectivity to the client side, the server chooses the most appropriate encoding to deliver the desired video clip. This latest feature opens up the possibility of "ubiquitous" distribution through access to the content even from mobile devices supporting MPEG-4.

Our system also includes administrative functions enabling users to upload, categorize, index and annotate the required material. All videos to be streamed are converted to DVD (Digital Versatile Disc) from the original analog and digital sources. The compression process creates several media files in MPEG-4 encoding. Textual metadata is associated with temporal "segments", i.e., a sequence of the video contains a defined start and end time. At the current stage, this process has nothing to do with automatic video segmentation; content producers/AV-media personnel decide and enter the start and end time and the associated metadata manually. The person who is uploading the digital movie to the system is responsible for generating the metadata. The metadata entails content description and time stamps marking the beginning and end of different segments of the movie. How to set the time stamps is decided by the person uploading the film. Each segment should be set such that it is logically cohesive and it contains a number of attributes describing the different topics associated with the chosen segment that will be indexed.

Attaching metadata on a segment level allows users to go directly to segments containing relevant material. For this purpose, full-text search facilities are provided. Furthermore, the material is sorted into hierarchical categories. Users can browse the library by exploring the classification hierarchy and viewing selected videos. They can search the database by specifying a category and typing a keyword. Thus, users are offered direct search and browsing interfaces (Hearst et al., 2002). It should be noticed that the present metadata schema attached to the segments has been customized to the current application.

The system has several advantages. It avoids distributing unnecessary video data by the adaptation to client bandwidth. This also increases usability, as clients with lower bandwidth have less waiting time while still serving high quality to broadband users. Since human beings are not involved in the mechanics of multimedia distribution, further efficiency and cost-effectiveness is achieved. It offers a great deal of flexibility in the storage, distribution and retrieval of videos. The system runs on relatively inexpensive hardware and software. However, we want to point out that despite progress in the areas of retrieval and distribution, the manual generation of metadata - the content description and time stamps that are central to the functionality of the system - remains a considerable drawback.

System Architecture, Technological Aspects and Implementation

Our system is running on a Linux server. It consists of a video encoding processor, a streaming server, a web server and a Database (MySQL). An overview of the system's architecture is illustrated in figure 1. Video encoding is done at the client side. Users uploading material need to have the proper encoding software installed. The files are then uploaded to an ftp server. An Apache web server hosts our web-based search interface and an interface for upload and initial annotation of submitted material. The MySQL database stores metadata and indexes for each media object. Finally, a Darwin streaming server delivers the MPEG-4 videos. For viewing, clients must have a web browser and QuickTime stand-alone viewer installed.

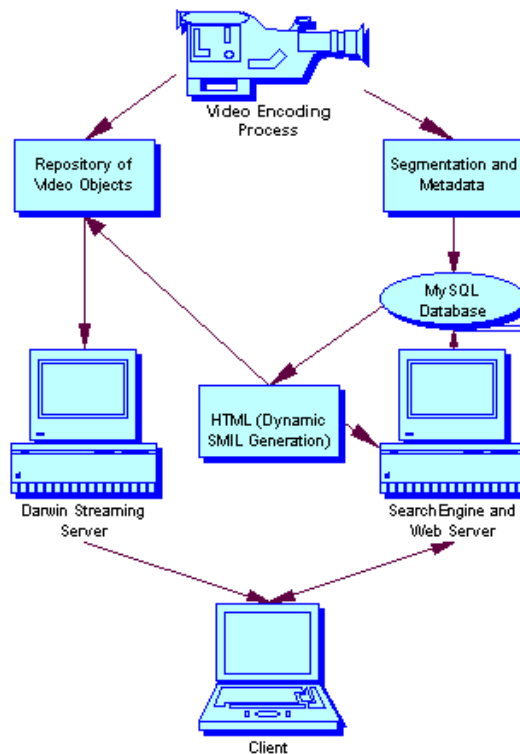


Figure 1. The system's architecture

The interface for video playback is a SMIL file dynamically created using PHP. While investigating for the most appropriate solution to this issue, we found out a number of problems related to the compatibility between SMIL implementations and supported functionality within standard media players on the market. Only QuickTime allowed us to offer a functional *seek feature* via SMIL in a supported media format like MPEG-4 and MOV (QuickTime associated format). The implementation of these ideas is illustrated in figure 2, as presented below. SMIL is seen as being superior to other approaches since it allows the system to access video material at any point in the stream. Other solutions that use a segmentation approach are seen as less flexible, and therefore less desirable, since all access points in the timeline of the video are previously determined.

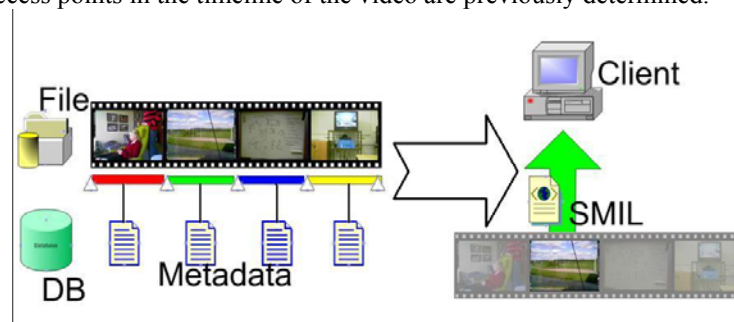


Figure 2. Using SMIL for presenting the different segments of a movie file

The structure of the metadata associated to each file is very simple. Figure 3 illustrates a screenshot of the administration interface (in Swedish) for adding and editing metadata. It can be seen that a film has a title; a category and language type. A film can be segmented into different chapters that can be described by free text, keywords and start and end time of the segment in which events to the associated keywords will appear. This latter feature (start and end time) is used for generating the dynamic SMIL files as described above. We use a simple keyword-based approach where we assign keywords to whole movies or movie segments. In addition, administrators can specify the language of movies and assign them to nodes in a tree-like taxonomy. The taxonomy can be extended manually when needed. The generated metadata are stored in a SQL database, from which they are extracted with PHP queries when needed.

It is clear that this simple approach limits interoperability and integration with external systems, which could be achieved by utilizing standards such as LOM (IEEE LTSC, 2002), RDF (Miller et al., 2004) and MPEG-7 (Manjunath et al., 2002) for content description or at least committing to an accepted ontology, as it was illustrated by Ronchetti and Saini (2004). The current implementation of our system does not support interoperability with other digital media libraries. However, based on existing software functionality implemented in the system, we could easily adapt it to produce XML style metadata files satisfying the RDF standard. These XML files could contain the location of the clip with all its related information such as language, length and content. Depending on the requirements of other existing systems with which we want to connect to, a RDF query interface needs to be developed.

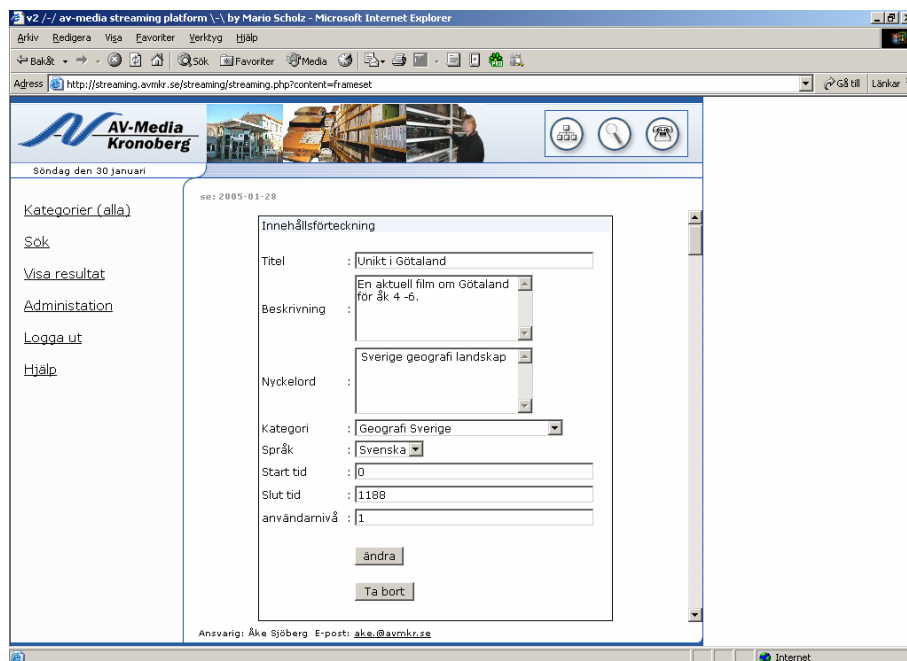


Figure 3. The administration interface for editing metadata

However, this issue regarding interoperability was never defined as an intended functionality of the system. The main design criteria specified by people at AV-Media were ease of use, simplicity, and functionality. In this particular case, the web interface is sufficient, and there is no need for automatized access via other channels. Besides, this is not desired by AV-Media and content producers. When it comes to the technical implementation of the system, we tried to rely on open-source applications to keep costs low, as this aspect has been defined as one of the desired features of the system. At the time of designing the system, we did not find an open source MPEG-4 encoder that satisfied our expectations. Hence, for our initial implementation we used a commercial product (Sorenson Squeeze, 2005). We currently also achieve encoding not only with Sorenson. We are using an open source solution called MPEG4IP (MPEG4IP, 2005). MPEG4IP works stable and is used for encoding digital films at schools. This solution has been even used for a few live broadcasts without prior recording with satisfactory results.

The MPEG-4 video compression standard supports various bit rates. A single stream can serve several mediums with multiple bit rates, which MPEG-4 supports in the range of 20 Kbps to 6 Mbps. However, we had experienced some problems delivering different encoding bandwidth from a single file. Thus, our system builds on replicating multimedia files in different qualities. In addition to platform independence, MPEG-4 video

compression provides high-resolution images. It supports larger resolutions close to TV-quality (VGA 640 X 480). Much of the content to be delivered by the system has been originally recorded for TV (720 X 576, PAL). If the material contains subtitles or other text, they become hard to read when encoding in lower resolutions. Thus, VGA is supported and is the preferred option for content delivery (see figure 4). While a smaller resolution has to suffice when content is delivered at low bandwidth rates, good quality, large screen resolution greatly enhances the user experience and, given a choice, the user is attracted to a device with a larger resolution.

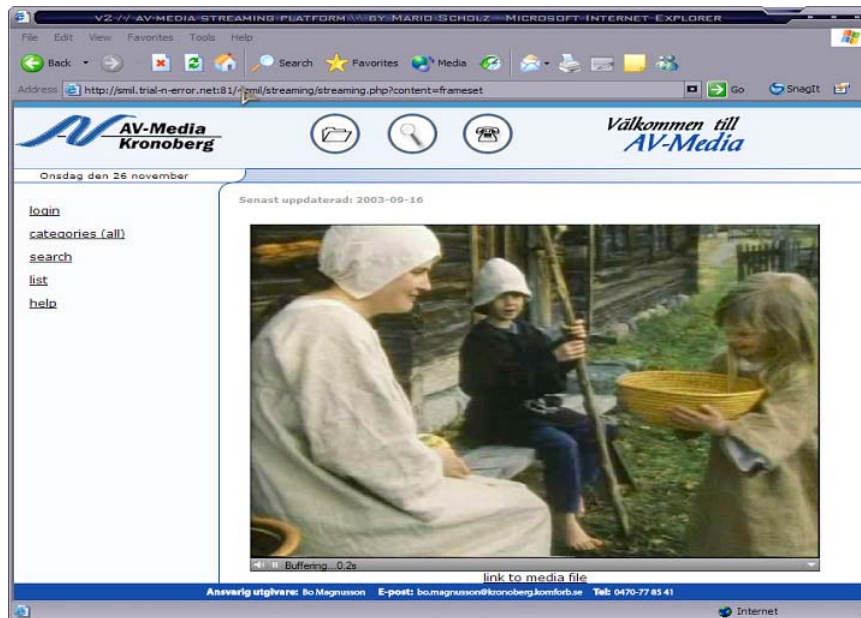


Figure 4. A high resolution for streaming video delivery over the WWW

Conclusions

In this paper we presented the results of our work with regard to the design and implementation of an educational digital video library using MPEG-4 encoding, SMIL and web technologies. We have been able to provide fine-grained, free-text and keyword search and retrieval across different digital video films and clips by appropriately combining complementary metadata derived from the individual digital objects.

The architecture of the systems is now in place, and people at AV-MEDIA have filled the video object repository with a considerable number of titles. One of the major problems we are facing in this respect relates to the issue of how to increase the amount of titles available. In Sweden, we experience some problems with regard to copyright issues for distributing educational video material over electronic networks produced by different content providers. At the moment of writing this paper, there are several hundred titles available through our system.

Due to the reasons described above, users of the system are encouraged to produce their own educational material in order to populate the repository of digital videos. Thus, teachers are contributing to this repository by creating their own educational video material, as an alternative way to enlarge the amount of educational digital videos. These activities are in line with AV-Media current efforts related to training teachers to produce their own educational material using digital video. During the last ten months around 80 teachers have been trained on how to produce digital video material to be used for educational purposes.

In parallel to these activities, our educational video library has been intensively used by more than 100 teachers from the whole region. At present, all schools of our region have access to the system through the internet, so they are able to use the system. The feedback we got from the teachers regarding the quality of service and the response of our system has been satisfactory. Experiences from the teachers using the system show that not only the films offered by AV-MEDIA are of interest for educational use. Also films that have been produced by teachers or students and have been stored in the video repository are highly appreciated. Perhaps this latest aspect is one of the most important issues for the adoption of a digital video library by teachers and students; the fact that they can become content providers and not only consumers of digital media. The implementation of our system allows now schools in our region to share a common database; it contributes to the creation of a stronger

community of educators by providing new resources, in the form of educational digital films, and tools to be used in the classroom. It also enables students and teachers to search for digital videos easily and more effectively. By allowing users to store, retrieve and edit video more flexibly than it has been done before, this technological approach has the potential to significantly improve the economics and logistics of video distribution in educational settings.

One of the main advantages of our approach to web-based video retrieval is the fact that the distribution process can be adapted to the particular environment and connectivity of the user. On the other hand, the main drawback of the system we developed is the manual generation of metadata. This particular activity is a very demanding and time-consuming process. Rossmanith (2003) has recently suggested an innovative approach for the generation of dynamic metadata based on users' feedback. We plan to implement some of these ideas in the near future, in order to allow users to contribute with their metadata to the objects stored in the digital video repository.

References

- Amir, A., Ashour, G., & Srinivasan, S. (2001). Towards Automatic Real Time Preparation of On-Line Video Proceedings for Conference Talks and Presentations. *Paper presented at the 34th Hawaii International Conference On System Sciences*, January 3-6, 2001, Hawaii, USA.
- Bulterman, D., & Rutledge, L. (2004). *SMIL 2.0: Interactive Multimedia for Web and Mobile Devices*, Berlin, Germany: Springer.
- Hearst, M., Elliot, M., English, J., Sinha, R., Swearinged, K., & Yee, K. (2002). Finding the flow in web site search. *Communications of the ACM*, 45 (9), 42-49.
- Hunter, J., & Little, S. (2001). Building and Indexing a Distributed Multimedia Presentation Archive Using SMIL. *Lecture Notes in Computer Science*, 2163, 415-428.
- IEEE Learning Technology Standards Committee (LTSC) P1484.12 (2002). *Draft Standard for Learning Object Metadata (LOM)*, Retrieved October, 6, 2005, from, <http://ltsc.ieee.org/wg12/index.html>.
- Jonassen, D., Peck, K., & Wilson, B. (1999). *Learning with Technology: A Constructivist Approach*, Upper Saddle River, NJ: Prentice Hall.
- Li, Q., Tang, H., IP, H., & Chan, S. (2003). A web-based video retrieval system: architecture, semantic extraction, and experimental development. In Fuhrt, B., & Marques, O. (Eds.), *Handbook of video databases - design and applications*, Boca Raton, FL: CRC Press, 539-612.
- Manjunath, B. S., Salembier, P., & Sikora, T. (2002). *Introduction to MPEG-7: Multimedia Content Description Interface*, New York: Wiley.
- Marchand-Maillet, S. (2000). Content-based video retrieval: an overview. *Technical Report 00.06, CUI*, University of Geneva, Geneva, Switzerland.
- Martinez, A. M., & Serra, J. R. (2000). A New Approach to Object-related Image Retrieval. *Journal of Visual Languages and Computing*, 11 (3), 345-363.
- Miller, E., Swick, R., & Brickley, D. (2004). *Resource Description Framework (RDF) / W3C Semantic Web Activity*, Retrieved October 25, 2005, from, <http://www.w3.org/RDF/>.
- Mojsilovic, A., & Rogowitz, B. (2001). Capturing image semantics with low-level descriptors. *Paper presented at the International Conference on Image Processing*, October 7-10, 2001, Thessaloniki, Greece.
- MPEG4IP. (2005). MPEG4IP: Open Source, Open Standards, Open Streaming. Retrieved October 6, 2005, from, <http://mpeg4ip.net/>.

Ronchetti, M., & Saini, P. (2004). Knowledge management in an e-learning system. In Kinshuk, Looi, C. T., Sutinen, E., Sampson, D., Aedo, I., Uden, L., & Kähkönen, E. (Eds.), *Proceedings of the 4th IEEE International Conference on Advanced Learning Technologies*, Los Alamitos, CA: IEEE Computer Society Press, 365-369.

Rossmannith, P. (2003). DYMICS: A system for dynamic metadata creation during search. *Unpublished Master of Science Thesis in Computer Science*, School of Mathematics and System Engineering, Växjö University, Sweden.

Sikora, T. (1997). The MPEG-4 Video Standard Verification Model. *IEEE Transactions on Circuits and Systems for Video Technology*, 7 (1), 19-31.

Sorenson (2005). *Sorenson Communications*, Retrieved October 6, 2005, from, <http://www.sorenson.com/>

Yoshimura, T., Yonemoto, Y., Ohya, T., Etoh, M., & Wee, S. (2002). Mobile streaming media CDN enabled by dynamic SMIL. *Proceedings of the 11th international Conference on World Wide Web*, Retrieved October 25, 2005, from, http://portal.acm.org/ft_gateway.cfm?id=511530&type=pdf&coll=GUIDE&dl=GUIDE&CFID=59292773&CFTOKEN=13219801.

Vailaya, A., Figueiredo, M. A. T., Jain, A. K., & Zhang, H. J. (2001). Image Classification for Content-Based Indexing. *IEEE Transactions on Image Processing*, 10 (1), 117-130.