

Content Classification and Context-Based Retrieval System for E-Learning

Ankush Mittal

Department of Electronics & Computer Engineering, Indian Institute of Technology, Roorkee, India
ankumfec@iitr.ernet.in

Pagalthivarthi V. Krishnan

Department of Applied Mechanics, Indian Institute of Technology, New Delhi, India
kvp@am.iitd.ernet.in

Edward Altman

Institute for Infocomm Research, Heng Mui Keng Terrace, Singapore
ed.altman@acm.org

ABSTRACT

A recent focus in web based learning systems has been the development of reusable learning materials that can be delivered as personalized courses depending of a number of factors such as the user's background, his/her learning preferences, current knowledge based on previous assessments, or previous browsing patterns. The student is often confronted with complex information mining tasks in which the semantics of individual sources require a deeper modelling than is offered by current learning systems. Most authored content exist in the form of videos, audio, slides, text, and simulations. In the absence of suitable annotations, the conversion of such materials for on-line distribution, presentation, and personalization has proven to be difficult. Based on our experiences with Open Courseware (OCW) and Singapore-MIT Alliance (SMA) video database, this paper presents a personalized delivery system that uses a domain ontology and pedagogical models to compose course materials in response to a user's query. We also present several important E-learning applications emerging from the framework.

Keywords

Semantic analysis, Multimedia features, Video indexing, State diagram, Contextual retrieval, User model

Introduction

E-Learning is rapidly changing the way that universities and corporations offer education and training. In recent years, the acquisition and distribution of rich media content has been largely automated, however research challenges still remain for the dynamic creation of media productions for the end user experience (Kinshuk and Lin, 2004). Prerequisites for reusing prepared learning materials typically involve finding relevant documents and context based retrieval of content elements which we will refer to as lecture fragments (Vercoustre and McLean, 2005). The following issues become crucial in reusing materials in context based learning:

- (a) Finding relevant document sources within the context of recent topics learnt and of the nature of audience.
- (b) Selecting more specific parts of documents that could be reused, based on the pedagogical semantics of definitions, examples, graphics, tables, and images.
- (c) Defining the sequence in which document elements for selected concepts should be accessed or presented
- (d) Defining the curriculum planning that would fit with the pedagogic approaches, and that will hopefully adapt to the actual learner.

A generic approach to handling such issues is to define reusable chunks of documents that can be retrieved, adapted, and assembled in a coherent way for a given educational purpose (Fundulaki et al., 2001). Unfortunately, the way fragments are described and used is very much system and application dependent. Therefore it cannot be reused by another system for another learning experience on the same topic but with a different objective, or a different instructional method. Most often the fragments have to be written from scratch with the particular application in mind. In this paper we address the issue of defining and automatically classifying the semantic fragments.

Much of the e-learning materials that have been created in recent years are in raw form as audio, video, slides, text and simulations. Manually annotating this content with semantic labels is a laborious and error prone task. Semi-automatic tools are therefore sought that can perform analysis on these materials and provide semantic descriptions. In this paper, we present a framework to analyze information in varied resources and discuss how fragments can be contextually (re)used for personalized learning. We show how the efficient retrieval in

complex domains can be done by a two stage process of assigning semantic labels to media content and resolving user queries through mediation between pedagogical models and domain models.

Importance of Context-Based Retrieval (CBR) system for e-learning

A key problem exists in bridging the semantic gap between raw video and high level information required by students. For example, what sort of information can we extract from raw video, audio, and slides and how to extract them? What can we understand from the information? How to put the information into appropriate form so that it can be customized for use by other applications? Referring to a course on computational algorithms in the computer science domain, a student reviewing for an exam poses the query “What is the relationship between Dynamic Programming and Greedy Algorithms?”. A straightforward technique involving keyword search is not effective as the terms dynamic programming and greedy algorithms may not appear in the same local context. In addition, there is no such index available in video database for providing answers to semantic queries.

The direct application of text mining tools or natural language processing tools on an e-learning text database would not, in general, yield a meaningful index. This is because the traditional techniques of keyword identification and hot spotting of concepts do not work well when the query is mostly semantic in nature. Additionally, the target keyword or abstract concept is likely to occur many times within a course, thus contextual knowledge is required to refine access to the information. Furthermore, student learning is impeded by the lack of a video index that currently makes the tasks of browsing and retrieval highly inefficient. The construction of a video index is a tedious task when done manually. Content based solutions are available for other media intensive domains including sports and news, but have not yet been systematically explored for educational videos (Idris and Panchanathan, 1997; Woudstra et al., 1998; Mittal and Cheong, 2003). The key contributions of this approach are:

1. Classification of semantic level events from the event flow of the lecture video.
2. Use of a rule based system to conduct inference and discover relations in the space of potential presentations.
3. Formation of a base for providing personalization tools for various users.

The paper also shows how the material once developed can be reused in context of the type of user and learning mode of the user. Using our technique, we are able to separate the lecture videos into several component states and personalize the video presentation from these states. For our experiments, we used a corpus of 26 lecture videos from the Singapore-MIT Alliance along with the associated PowerPoint slides.

Organization of the paper

This paper describes an automatic methodology for the indexing of the lecture videos. The second section described the distance learning paradigm, and the problems faced. The third section discusses the formulation and analysis of a state model for lectures. In the fourth section, video indexing features are discussed. The section titled of Lecture Video Indexing elaborates on the mapping of low-level features to lecture semantics. Finally, we discuss the experimental results, several applications and significance of taking this approach, as well as examine the future direction of this research.

Distance learning Paradigm

Issues in Designing CBR for Distance Learning

Indexing in the present context means labelling the content into semantically meaningful units corresponding to the different topics and ideas that a lecturer has introduced (Semple et al., 2000). Extracting the content of the lecture allows students to identify the underlying structure of the lecture and easily access the parts in which they have the greatest interest. In traditional books and textual documents, the organization of the learning material is decided by the author and the learner is expected to read the document linearly, although nothing prevents him to jump to the conclusions first or to skip a section if he is already familiar with the concepts. The flexible nature of hypertexts and on-line materials offers new opportunities and challenges for learning support that can guide the learner in a more personalized way. In particular, when the content is split into smaller units, the learning system is expected to provide some guidance as to which part to read next based on prior knowledge of the user and nature of the user.

Related Work

Using video for educational purposes is a topic that has been addressed at least since the 1970s (Chambers and Specher, 1980). Recently the focus of the research has been to maximize the utilization of educational video resources which have accumulated over a period of time. Ip and Chan (1998) use the lecture notes along with Optical Character Recognition (OCR) techniques to synchronize the video with the text. A hierarchical index is formed by analyzing the original lecture text to extract different levels of headings. An underlying assumption is made that the slides are organized as a hierarchy of topics, which is not always the case. Many slides may have titles which are in no way related to the previous slide.

Bibiloni and Galli (2003) proposed a system using a human intermediary (the teacher) as an interpreter to manually index the video. Hwang et al. (1997) propose a hypervideo editor tool to allow the instructor to mark various portions of the class video and create the corresponding hyperlinks and multimedia features to facilitate the students' access to these pre-recorded sequences through a web browser. This scheme also requires a human intermediary and thus is not generalized.

The recently developed COVA system (Cha and Chung, 2001) offers browsing and querying in a lecture database; however, it constructs the lecture index using a digital text book and neglects other sources of information such as audio or PowerPoint slides.

Temporal state model for lectures

The integration of information contained in e-learning materials depends upon the creation of a unifying index that can be applied across information sources. In content based retrieval systems, it is often convenient to create a *state model* in which the nodes represent semantically meaningful states and the links between nodes represent the transition probabilities between states. In the case of educational video information systems the state model for the lecture is composed of states that represent the pedagogical style of teaching. For the purpose of illustrating the concept, let us consider computer science courses, especially theoretical ones like the Introduction to Algorithms. In this case, each lecture can be said to contain one or more topics. Each topic contains zero or more of the following pedagogical elements:

- Introduction – general overview of the topic.
- Definitions & Theorems – formal statement of core elements of the topic.
- Theory - derivations with equations and diagrams.
- Discussions - examples with equations and diagrams.
- Review – repetition of key ideas.
- Question and Answer – dialogue session with the students.
- Sub-Topic – branch to a related topic.

A simple state model for video based lectures can be represented as shown in the Figure 1. Machine learning techniques are used to construct a state model consisting of 8 different states linked by maximal probability edges. Each edge from a given node represents the probabilistic transition to another state. For example, from state Topic, the next state is Definition with probability 0.91 and the next state is Discussion with probability 0.09. The edge labels in Figure1 show the transition probabilities using the corpus of SMA lectures as a training set.

The state model implicitly encodes information about the temporal relationships between events. For instance, it is clear from Figure 1 that the introduction of a topic is never followed by a *theorem* without giving a *definition*. The state model when supplemented with our indexing techniques (discussed in later sections) provides useful information regarding the possible progression of topics in the lecture.

The semantic analysis of raw video consists of four steps;

1. Extract low and mid level features. Examples of low level features are color, motion, and italicized text. Some mid-level features are zoom-in and increased hand movement of lecturer.
2. Classify the feature vectors from the lecture into a finite set of states. The resultant states correspond to salient events in the lecture and are assigned semantically meaningful labels, such as Definitions, Emphasis, Topic Change, Q&A, and Review.
3. Apply contextual information to the sequence of states to determine higher level semantic events, such as *defining a new term*, *reviewing a topic*, or *engaging in off-topic discussion*.

4. Apply a set of high level constraints to the sequences of semantic events to improve the consistency of the final labelling.

Semantic analysis of the video begins with the extraction of salient features. Features are interpretation independent characteristics that are computationally derived from the media. Examples are pitch and noise for audio and color histogram and shape for images. Quite a number of features have been identified and many feature detection algorithms already exist (Gonzalez and Woods, 1992; Gudivada and Raghavan, 1995).

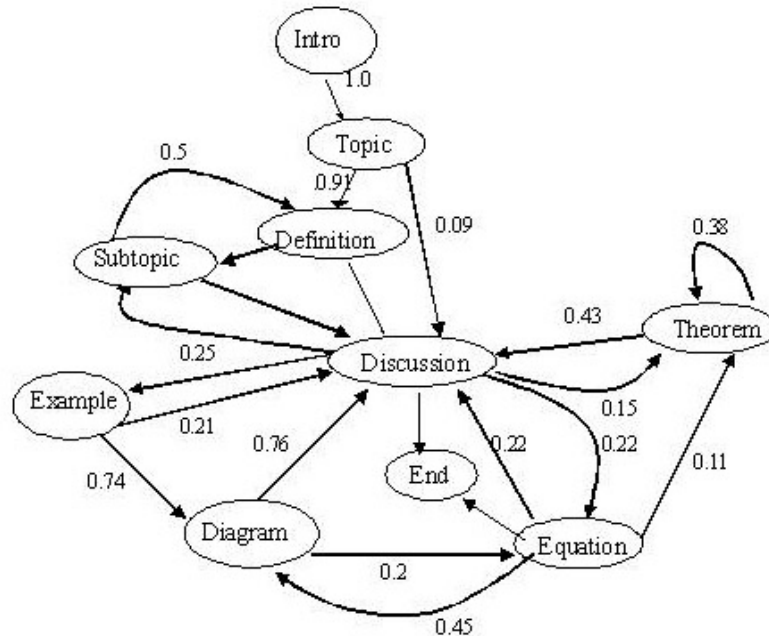


Figure 1. State Diagram of Lectures - Each state follows the probabilistic edges to go to another state

Video Indexing Features

The evaluation of algorithms used in video segmentation engines and the detailed mechanisms for feature extraction are beyond the scope of this paper. Rather we will present a list of the most useful features for audio, video, and text that are used in the video indexing arena. We then discuss some indexing techniques that are based on these features.

Audio features

There are many features that can be used to characterize audio signals. Volume is a reliable indicator for detecting silence, which may help to segment an audio sequence and to determine event boundaries. The temporal variation in volume can reflect the scene content. For example, a sudden increase in volume may indicate of the transition to a new topic. Spoken word rate and recurrence of a particular word is an indicator of the scope for a topic or cluster of words within a discussion (Witbrock and Hauptmann, 1997).

Video Features

A great amount of research has gone into summarizing and reviewing various features useful for the video segmentation (Wang et al., 2000). Some of the most common features used in video analysis are discussed below. The color histogram, which represents the color distribution in an image, is one of the most widely used color features. If the difference between the two histograms is above the threshold, a boundary shot is assumed. Motion is an important attribute of the video. Motion information can be generated by block matching or optical flow techniques (Akutsu et al., 1992). Motion features such as motion field, motion histogram, or global motion parameters can be extracted from motion vectors. High level features that reflect the camera motions such as panning, zooming and tilting can also be extracted (Rui et al., 1999). Motion features such as hand velocity, hand position, black board motion, pointing gestures, etc. inherently store much information.

Text Features

In the distance learning paradigm, text is one of the most important features that has still not been extensively researched and utilized. Ip and Chan (1998) propose text assisted video content extraction, but only to synchronize the video with the text. Text in the form of PowerPoint slides, which generally is the case with the educational videos inherently stores a great deal of information as we shall see with the SMA lecture corpus.

Lecture video Indexing

We introduce a general framework for video indexing systems in this section. The first step is to extract features and information from the raw data, which in this case are the video and the lecture notes. The most important and basic step in a video indexing engine is to extract the right features and then combine these to get the most efficient indexes.

Deriving semantics from low-level features

The mid level video features such as camera zoom or switching between the speaker and audience when used alone can not be reliably associated with topic change. However, when used to supplement other features they may provide important discriminatory information. The audio features such as detecting a silence may help to determine the beginning and end of salient moments in the lecture but certainly not the occurrence of a topic change. It does help to identify a question and answer session with the back and forth switching of audio from teacher to student. But these features are not sufficient to extract the different topics and their interrelationships as presented in the lecture. The potentially richest source of structural information is the black board activity, which in turn is represented in the lecture notes. Thus a proper analysis of the lecture notes, that is, the PowerPoint slides along with the properties discussed below can indeed be used to identify the lecture structure.

The PowerPoint slides that serve as lecture notes inherently store important information regarding the lecture which is still largely untapped. The text formatting information in the form of font size, shape, color, and boldness in itself reveals important aspects of the lecture. The state model of the educational videos discussed earlier includes four basic categories that also apply to slides, namely: Definitions and Theorems, Examples, Proofs, and Formulae. Analysis of the SMA lecture corpus prior to the computer based indexing identified a set of formatting rules in the PowerPoint slides such as: all the important words (keywords with definition) are always red and in italics; the special names are always in quotes; the slides having an example always have a word "example" in it; the questions and FAQs have a question mark in the particular slide; the common names are always in square brackets, etc. Some video feature such as, camera zoom in or zoom out to the blackboard or to the audience also specify transition of the lecture from one state to another say from Diagram to Discussion state. These rules may be specific for the SMA courseware, but the broader picture says that we can similarly define a set of rules existing in a large corpus of distance learning coursewares. This is then synchronized with the video to get the exact clip. The rules for indexing the slides in the above mentioned four categories can be summarized in the following categories.

Category 1: Definitions / Theorems

The keywords or defined terms are always red and in italics.. The word *definition* or *theorem* may be present and the string queried has to be definitely found in the slide.

Category 2: Examples

The course under consideration for the Introduction to Algorithms has an associated image file for all the examples to represent special graphics or equations. The presence of the text pattern for *examples* or *examples:* along with the string queried is mandatory for a slide to qualify as one containing examples. When analyzing the text, the context of the current slide is related to the previous slides. Thus the context of the particular example is linked to the contents above it and the topic in currently being discussed.

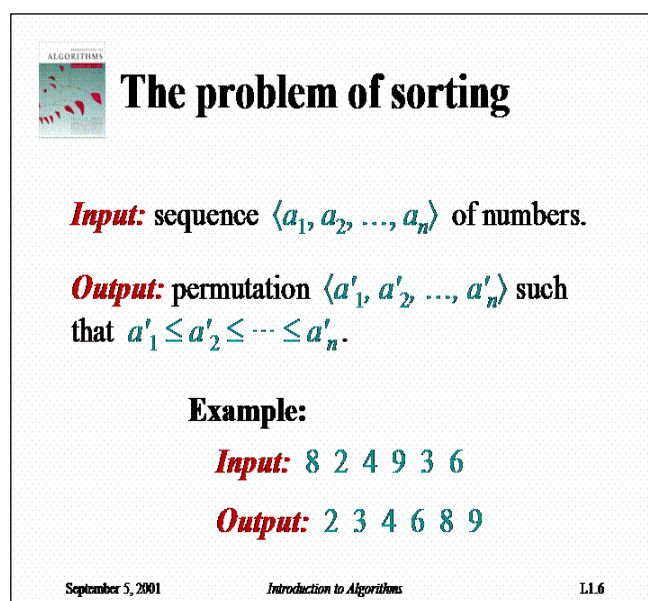
Category 3: Proof

The word *proof* along with the string queried is assumed to be present in the slides having relevant information associated with the query. This assumption is a generalized one and can be used for all distance courseware.

Category 4: Formulae

Slides containing embedded formulae can be easily identified through the identification of special symbols used to represent the mathematical expressions. Queries for mathematical expressions can be simply resolved by converting the query expression into a string of characters then performing pattern matching.

Sometimes a slide may contain only examples without any reference to the topic, as they might be in continuation with the previous slides. In such cases the system looks for the keywords queried in the previous slides and thereby checks for the presence of context in which the examples are given.



The problem of sorting

Input: sequence $\langle a_1, a_2, \dots, a_n \rangle$ of numbers.

Output: permutation $\langle a'_1, a'_2, \dots, a'_n \rangle$ such that $a'_1 \leq a'_2 \leq \dots \leq a'_n$.

Example:

Input: 8 2 4 9 3 6

Output: 2 3 4 6 8 9

September 5, 2001 Introduction to Algorithms L1.6

Figure 2: An example of the definition of a concept. Note that there are no words related to ‘definition’ such as define, etc. in the slide

It is well known that video features taken in isolation are highly ambiguous. The video feature for *zoom out* may indicate either discussion in the class, the presentation of a definition, or simply a technique to reduce the tedium of the video. Similarly, we find that the video feature *zoom in* may indicate the occurrence of Topic, Example, Diagram, Theorem, or Equation states.

After the entire lecture has been classified, the labelled metadata can be used to perform multiple tasks. The first one is searching in context. Several automatic frameworks exist for searching in context. For an example, see (Mittal and Altman, 2003). Here we employ a simple contextual searching algorithm. To enable searching in context, we need to manually enter the topic names for each video clip associated with a significant pedagogical event previously identified by the application of the classification rule set. Once the topic names have been keyed into the topic lists, we can then perform contextual search just by searching for all occurrences of the queried subject and returning the results.

This method is accurate because under our definition of the topic state, all subject matter which is important enough to be explained separately is classified as a topic or a subtopic. For example, when the term *quicksort* is mentioned under the divide-and-conquer method, our system classifies *quicksort* as a subtopic. Again, when *insertion sort* is compared with *quicksort*, it classifies *insertion sort* as another subtopic. As a result, the topic list is comprehensive in covering all material which is of importance. Hence, we are able to retrieve all instances of a particular query by searching through the topic list.

Synchronization of the video and the retrieved slide

Analysis of the blackboard for characters or the speech for words is performed to find a cluster of words which can be matched with the corresponding cluster of words in the PowerPoint slide. Initial experiments using Optical Character Recognition techniques for blackboard text showed that the efficiency is quite low and it is highly dependent on the lecturer's handwriting. Recent experiments with the speech recognition system seem to be more promising and can be used for alignment as well as keyword detection (Witbrock and Hauptmann, 1997). For effective video search, one needs to know exactly when each word in the transcript is spoken. A speech recognition system can provide this information. Thus, synchronization of the video and text can be achieved. The key idea is to create a system which automatically segments the educational videos which the students can then use to explore the desired sections of the lectures without going through the linear search. Thereby saving time and effort required of the student.

Experimental Results and applications

We tested our method on 26 lecture videos from the Singapore-MIT Alliance course SMA5503. The semi-automatic classification results are tabulated in Table 1

Table 1: Experimental Results in Confusion Matrix. The high value at the diagonal entries denotes the high accuracy in detection of that state

Actual State	Detected State							
	Intro.	Topic	Defn.	Discn.	Theorem	Example	Eq.	Diag.
Introduction	100							
Topic		90		5.5			2.25	2.25
Definition		20	80					
Discussion		7		86	3.5		3.5	
Theorem		6.25		6.25	87.5			
Example		8.5		8.5		83		
Equation		13		25			62	
Diagram			7.7					92.3

Overall, our method has an accuracy of 85.1% in detecting the correct state. The personalization rules being dependent on the first algorithm also have an accuracy of 85.1%. The contextual searching algorithm is solely dependent on the correct classification of the Topic State and, therefore, has an accuracy of 90%. In figure 3, we present some possible fields that must be stored with each fragment. The utility of some of these are obvious while the others are used in the following applications:

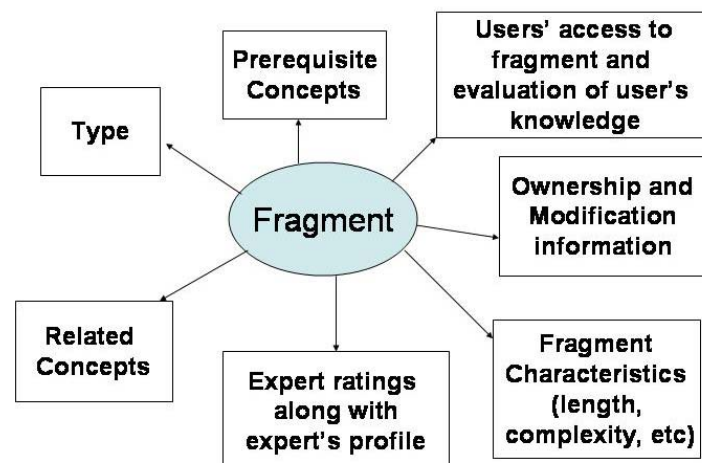


Figure 3: The fields stored with a fragment. The fields are updated when the fragment is created, accessed or rated

Personalization

A lot of research has been dedicated to develop flexible learning material that can deliver personalised courses depending of a number of factors such as the user's learning preferences, his current knowledge based on previous assessments or previous browsing in the material (Sampson et al., 2002). We are able to create several broad categories into which we can segment our target audience. This way we can ensure that the content for an absolute beginner is not the same for that of a student preparing for his exam. The students interested in this courseware can be divided into three broad categories on the basis of their requirements to review the lecture as follows; (a) A student may be viewing the lecture for the first time, (b) a student maybe reviewing it to brush up on concepts, or (c) a student may be reviewing it for the preparation of examination. Amongst these students some may prefer to view only the video part where, there is an example or a definition for a particular topic that is, he/she would like to view the lecture from the perspective of a particular topic. Some of the students may be interested in reviewing only the proofs in order to prepare for the examinations. The student is required to give a keyword for his search, using this keyword the search is performed and the search results are categorized under the four mentioned categories and presented to the student. Depending on the requirement and the available result set the student selects whether he wishes to review definitions, examples, formulae, proofs, or he wishes to review all one by one. The appropriate video along with the slides under the specified category is then provided to the student. If the information is found more than one time in the lecture, the system identifies these parts of the lecture as correlated to the searched topic and thus presents these topics under the head of related topics category to user.

An interesting presentation style would be to consider the user model, his learning objective and contextual information. Table 2 presents a set of rules that could be useful in personalizing the content. The fields in Table 2 with each fragments help determine the prerequisite concepts, type, etc.

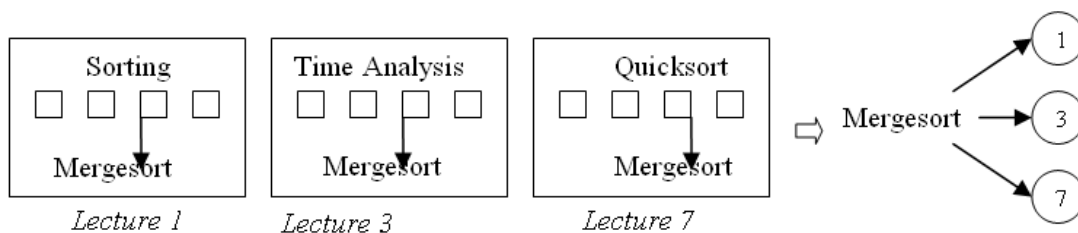


Figure 4: Search for “Merge Sort”. The occurrence of merge sort in three different lectures is determined. Appropriate context would then be used to find out whether the reference is a discussion on sorting, time analysis, or comparison with Quicksort

Summarization

There are many occasions where students are interested in obtaining a summary of the lecture. Summarization of the lecture should be based on the semantics of the lecture video. This can be done in several ways depending on the requirements of the student. Consider a simple and yet very important application of summarization where a student wishes to decide upon registering for the next semester course based on the summary of the course. Just as a movie trailer effectively portrays the type of movie (say violence, action, suspense, etc.), the summarization of the course should be representative of the degree of difficulty, level of mathematics, lecturer's abilities, etc. The flexible selection of content is easily accomplished since metadata for the characteristics of each fragment is stored in a database. The fragment whose characteristics are most common to those of other fragments can be included in the summary. In addition, an important consideration is also to include fragment of each type (such as question-answer session, lectures, etc.).

Another way of presenting summarization is as follows. Give the contents of the lecture, slide by slide, under the respective slide title. Under each slide title the important points covered in that slide are sited. Using this as a cue the student can link to the appropriate part of the lecture and view the video. Though this in itself will generally be a huge list, it serves the purpose of giving an idea of the contents of lecture as well as relative position. Thus, eliminating not only the linear search but also giving an idea of the breadth of lecture. Secondly, a student may wish to prepare for an examination by just going through definitions and the corresponding examples or by going through definitions and proofs.

Table 2: The adaptive presentation style depending on user model and contextual information

User Type	Contextual Learning model	Desired information	Presentation Style
Biology student	Concept searched as part of another course (biotechnology)	Divide and Conquer	(a) Present definitions & examples (b) Give links to prerequisite concepts (c) Skip theorems and mathematics
Researcher	Terminology/Concept clarification	Strassen's Algorithm	(a) Present definitions, analysis and theorems (b) Relate to places where Strassen's algorithm is used or related, for example, divide-and-conquer
Student beginning the course	Serial coverage	Everything in a lecture	(a) Provide links to skip theorems and analysis (b) Show the general flow of the lectures and the fragment in the entire course to gather synthesis
Student revisiting the course	Exam preparation	Entire course	(a) Relate to discussion and QA session (b) Show related concepts

Retrieving fragments of documents

We are able to efficiently and accurately search in context throughout the video database. For example, by searching for merge sort, we return not only the video clip that teaches merge sort, but also other clips from other lectures where some aspect of merge sort is further explained (see figure 4). In this particular case, merge sort is mentioned in video lecture 1 under the topic Sorting. It is also mentioned again in lecture 3 under Time Analysis, and in lecture 7 where it is compared to quicksort. Hence, when a student uses this system to search for merge sort, he has immediate access to all three related video clips even though they are taught in completely different lectures and different parts of the course. As a result, a student searching for merge sort will get a much clearer idea of how it actually works and all its different aspects.

When users are looking for documented information, expert finding systems can provide useful evidence as to the quality of the information as (Höök et al., 1997) report, saying that a user of a collaborative filtering system may be more interested in what particular experts regard as important information.

Finding experts

Quite often in a collaborative environment, the students wish to know if they could get some directions from someone understanding a particular concept. Matching a learning need to a person that can provide a solution or advice can be supported by finding relevant people based on their expertise as computed by analyzing the fragment of the documents they produce, own, read etc. This is accomplished by keeping track of which user accessed which fragment, along with the assessment by the system on user's understanding of the fragment (through evaluation, FAQs, discussion forum, etc.). This is illustrated in Figure 3.

Conclusions

We have designed a system for indexing videos using audio, video and PowerPoint slides and segmenting them into various lecture components. Personalized documents for use in educational systems enable the presentation of fragments based on the user model and rich semantic descriptions of the fragments. This helps to make the videos more suitable for absorption of the subject matter by the students. While full-text indexed retrieval systems have been proposed earlier, our method is more efficient as it uses all forms of media to segment and index the video. It also allows us to perform efficient contextual presentation with minimum human supervision. The system allows better reuse of the fragments for different purposes. For future work, better rules can be created to handle more diverse categories and to make the personalization tailored to individual needs.

References

Akutsu, A., Tonomura, Y., Ohba, Y., & Hashimoto, H. (1992). Video Indexing using Motion Vectors. *Paper presented at the SPIE Conference on Visual Communication and Image Processing '92*, November, Boston, USA.

- Bibiloni, A., & Galli, R. (2003) Content Based Retrieval Video System for Educational Purposes. *Proceedings of Eurographics Workshop on Multimedia "Multimedia on the Net, EGMM 1996*, retrieved October 25, 2005 from <http://citeseer.ist.psu.edu/53258.html>.
- Cha, G. H., & Chung, C. W. (2001). Content-Based Lecture Access for Distance Learning. *Paper presented at the International Conference on Multimedia and Expo*, August 22-25, 2001, Tokyo, Japan.
- Chambers, J. A., & Specher, J. W. (1980). Computer Assisted Instruction: Current Trends and Critical Issues. *Communications of the ACM*, 23 (6), 332-342.
- Fundulaki, I., Amann, B., Scholl, M., Beeri, C., & Vercoustre, A.-M. (2001). Mapping XML Fragments to Community Web Ontologies. *Paper presented at the Fourth International Workshop on the Web and Databases, WebDB 2001*, May 24-25, 2001, Santa Barbara, California, USA.
- Gonzalez, R. C., & Woods, R. E. (1992). *Digital Image Processing*, New York: Addison-Wesley.
- Gudivada, V. N., & Raghavan, V. V. (1995). Content Based Image Retrieval Systems. *IEEE Computer*, 28 (9), 18-22.
- Höök, K., Rudström, Å, & Waern, A. (1997). Edited Adaptive Hypermedia: Combining Human and Machine Intelligence to Achieve Filtered Information. *Paper presented at the Flexible Hypertext Workshop held in conjunction with the 8th ACM International Hypertext Conference*, retrieved October 25, 2005 from <http://www.sics.se/~kia/papers/edinfo.html>.
- Hwang, J.-N., Youn, J., Deshpande, S., & Sun, M.-T. (1997). Video Browsing for Course-On-Demand in Distance Learning. *Paper presented at the International Conference on Image Processing (ICIP)*, October 26-29, 1997, Santa Barbara, CA, USA.
- Idris, F., & Panchanathan, S. (1997). Review of Image and Video Indexing Techniques. *Journal of Visual Communication and Image Representation*, 8, 146-166.
- Ip, H. H. S., & Chan, S. L. (1998). Automatic Segmentation and Index Construction for Lecture Video. *Journal of Educational Multimedia and Hypermedia* 7 (1), 91-104.
- Kinshuk, & Lin, T. (2004). Cognitive profiling towards formal adaptive technologies in web-based learning communities, *International Journal of Web Based Communities*, 1 (1), 103-108
- Mittal, A., & Cheong, L.-F. (2003). Framework for Synthesizing Semantic-Level Indices. *Journal of Multimedia Tools Applications*, 20 (2), 135-158.
- Mittal, A., & Altman, E. (2003). Contextual Information Extraction for Video Data. *Paper presented at the 9th International Conference on Multimedia Modeling (MMM)*, January 7-10, 2003, Kamkang, Taiwan.
- Rui, Y., Huang, T. S., & Chang, S. F. (1999). Image Retrieval: Current Technologies, Promising Directions, and Open Issues. *Journal of Visual Communication and Image Representation*, 10, 39-62.
- Sampson, D., Karagiannidis, C., & Kinshuk (2002). Personalised Learning: Educational, Technological and Standardisation Perspective, *Interactive Educational Multimedia*, 4, 24-39.
- Semple, P., Allen, R. B., & Rose, A. (2000). Developing an Educational Multimedia Digital Library: Content Preparation, Indexing, and Usage. *Paper presented at the EdMedia 2000 Conference*, June 26 - July 1, 2000, Montreal, Canada.
- Vercoustre, A., & McLean, A. (2005). Reusing Educational Material for Teaching and Learning: Current Approaches and Directions. *International Journal on E-Learning* 4 (1), 57-68.
- Wang, Y., Liu, Z. & Huang, J. C. (2000). Multimedia Content Analysis using both Audio and Visual Cues. *IEEE Signal Processing Magazine*, 17 (6), 12-36.
- Witbrock, M. J., & Hauptmann, A. G. (1997). Using Words and Phonetic Strings for Efficient Information Retrieval from Imperfectly Transcribed Spoken Documents. *Paper presented at the second ACM International Conference on Digital Libraries*, July, 23-26, 1997, Philadelphia, USA.
- Woudstra, A., Velthausz, D. D., de Poot, H. G. J., Hadidy, F., Jonker, W., Houtsma, M. A. W., Heller, R. G., & Heemskerck, J. (1998). Modeling and Retrieving Audiovisual Information- A Soccer Video Retrieval System. *Paper presented at the 4th International Workshop on Advances in Multimedia Information Systems (MIS'98)*, September 24-26, 1998, Istanbul, Turkey.