Indexing and Retrieval of Multimodal Lecture Recordings from Open Repositories for Personalized Access in Modern Learning Settings

Christian Gütl
Institute for Information Systems and Computer Media (IICM), Graz University of Technology, Austria
School of Information Systems, Curtin University of Technology, Perth, Australia
Infodelio Information Systems and GÜTL IT Research & Consulting, Austria
cguetl@iicm.edu and cguetl@acm.org

Vanessa Chang
School of Information Systems, Curtin University of Technology, Perth, Australia
Vanessa.Chang@cbs.curtin.edu.au

Ângelo Manuel Franco
Faculty of Mathematics, University of Madeira, Portugal
spinolaangelo@gmail.com

Paulo Nazareno Maia Sampaio
Faculty of Mathematics, University of Madeira, Portugal
psampaio@uma.pt

Frank Kappe
Institute for Information Systems and Computer Media (IICM), Graz University of Technology, Austria
frank.kappe@iicm.tu-graz.ac.at

Abstract: An increasing number of lecture recordings are available to complement face-to face and the more conventional content-based e-learning approaches. These recordings provide additional channels for remote students and time-independent access to the lectures. Many universities offer even complete series of recordings of hundreds of courses which are available for public access and this service provide added value for users outside the university. These lecture recordings show the use of a great variety of media or modalities (such as video, audio, lecture media, presentation behavior) and formats. Insofar, none of the existing systems and services have sufficient retrieval functionality or support appropriate interfaces to enable searching for lecture recordings over several repositories. This situation has motivated us to initiate research on a lecture recording indexing and retrieval system for knowledge transfer and learning activities in various settings. This system is built on our former experiences and prototypes developed within the MISTRAL research project. In this paper we outline requirements for an enhanced lecture recording retrieval system, introduce our solution and prototype, and discuss the initial results and findings.

Introduction and Motivation

The rapid evolution of communication technologies, advanced data networks and simplified processes of multimedia information processing have greatly impacted learning experiences. Multimedia resources have become increasingly important in the learning process. Not only are the multimedia resources included in short training videos and interactive learning objects but there are also infotainment resources as well as asynchronous and synchronous communication channels (Rajkumar, Gütl, & Ramadoss, 2008; Gütl, 2008). Since more appropriate technology has become available for recording, processing and distributing of multimodal or multimedia sources, lecture recordings have become important educational media types which have raised increasing interest within the last few years. Further information of multimedia resources and their application in e-learning can also be found in literature such as (Huang, Eze, & Webster, 2006).

For more than ten years, various lecture and presentation recording systems have been researched and developed, such as *Classroom 2000* (Abowd et al., 1996) and *Authoring on the Fly* (Bacher, Müller, Ottmann, &

Will, 1997), *Camstasia Studio* (Camstasia Studio, 2008), *E-Chalk* (E-Chalk, 2008), *Lecturnity* (Lecturnity, 2008), *MIT Lecture Browser* (MITNEWS, 2007), *TeleTechingTool* (Ziewer & Seidl, 2002), and *virtPresenter* (Mertens, Ketterl, & Vornberger, 2007). Consequently, an increasing number of lecture recordings are available in order to complement face-to face and the more conventional content-based e-learning approaches but also to provide additional channels for remote students and time-independent access to the lectures. A majority of educational institutions provide only restricted access to their students. However, some universities offer a complete series of recordings of courses which are available for public access. Courses such as *MIT Open Courseware* (MITOCW), *Carnegie Mellon University Open Learning Initiative* (OLI) and Berkeley courses on YouTube (BERKELY, 2007) are available for public access. These open accessible lecture recordings not only support the institutions' own students but they also provide valuable support to other students. These recordings can be used in various learning settings such as self-directed learning, vocational training and life-long learning activities and can also help other teachers or communities in developing countries.

From the institutions' and teachers' point of view, the lecture recordings which are time independent provide additional transmission channel and they enable remote students or remote guest lecturers to integrate in the learning setting. The free access to other institutions' lecture recordings can also support a plan for a new curriculum, prepare courses and lectures, and to get other viewpoints of specific subjects. From the students' point of view, lecture recording of courses they are enrolled in enable the students to "consume" lectures they have missed, repeat difficult parts of the lectures and to prepare for examination. Available lecture recordings from other institutions can also help students to get complementary viewpoints of the same topic. The students can also use the recordings as complementary learning content especially if the course is presented by a highly specialized expert. Further details on teachers' and students' viewpoints can be found elsewhere in Spinola (2008).

The above mentioned lecture recording systems provide a great variety of media types and modalities, which include one or several *audio and video streams* from the presenter, presentation media (e.g. presentation slides, white-board tools or other computer-based application), and back channels (e.g. audience or remote participants), *speech-to-text* transcripts or extraction, *face recognition* as well as *face expression and gesture recognition*, *other modalities* from interaction with presentation media, *online and offline annotations* from students, teachers and communities, related documents (e.g. presentation slides, lecture notes, exercises, online tests, and other background knowledge), or further human-based or computer-based *semantic annotations and enrichment* (e.g. concept extraction, topic extraction, presentation type). Our literature review has shown, that (1) lecture recording systems focus on different media according to their main purpose and objective, (2) accessible lecture recordings are available in various media formats and practically none of the systems provide all of the above mentioned media types and modalities, and (3) practically none of the existing systems and services have sufficient retrieval functionality and they do not support appropriate interfaces to enable searching of several repositories.

The findings stated above have motivated us to initiate research on a lecture recording indexing and retrieval system for knowledge transfer and learning activities in various settings, such as in higher education, vocational training and life-long learning. This research is built on our experiences and prototypes developed within the MISTRAL research project (Gütl, 2008) which has focused on an enhanced multimodal information system for meeting scenarios. The remainder of this paper is organized as follows: Chapter 2 outlines requirements for an enhanced lecture recording indexing and retrieval system, followed by a brief introduction into the MISTRAL System in Chapter 3. Based on that the adaptation for the application of lecture recordings is discussed in Chapter 4, and finally lessons learned are given in Chapter 5.

Requirements for an Enhanced Lecture Recording Indexing and Retrieval System

Within our research initiative on lecture recording systems, Spinola (2008) discusses in detail requirements for an enhanced lecture or presentation recording system both for teachers and students. These requirements can be summarized as follows:

- Manage and index lecture recordings as well as integrate relevant documents such as presentation slides, lecture notes, curricula and other background knowledge
- Support various source formats and sets of media, fetch from various repositories, pre-process, convert and semantically enrich these media types to further build on unified internal representations and support the retrieval process and access to the data
- Retrieve and access lecture recordings in various granularity (e.g. entire lecture, part of the lecture segmented by topics or subtopics or content based)

- Provide access to various end devices (e.g. personal computer, PDAs and smart phones) in different modalities (e.g. video streams, only audio stream or text transcripts) and network infrastructure (e.g. broadband networks or wireless networks)

The MISTRAL System at a Glance

The MISTRAL system focused on technology-based methods for semantic annotation, extraction, indexing, retrieval and visualization of multimodal and multimedia data stream in the meeting application domain (García-Barrios, & Gütl, 2006). According to the CAMIS model (Conceptual Architecture of Multimodal Information Systems) the conceptual units of such systems are: (1) *Capturing* (handles the provision of proper data streams from diverse sources for the further process chain), (2) *Abstraction* (deals with data processing and information extraction and may range from simple to more complex tasks; e.g. format conversion, compression or summarization, and information extraction at multiple levels of abstraction), (3) *Fusion* (merges and combines information from the unimodal data sources which may be performed on different levels of abstraction), (4) *Storage* (handles the persistent internal representation of the unimodal data streams and the extracted information but also manage the access and delivery of the data in a trustful and secure manner), (5) *Retrieval* (supports the process of finding relevant information or delivering useful data, manages browsing and searching in various modalities on different semantic levels and structure composites), and (6) *Presentation* (manages the combined and synchronized presentation of multimodal output data for information consumption). (Gütl, 2008b) The MISTRAL system is designed to deal with all aforementioned conceptual units in the meeting domain.

Figure 1 outlines the MISTRAL architecture at a glance and shows the relations to the CAMIS units. On the outer left side of the diagram, the unimodal units for *Audio*, *Text*, *Video* and other environmental *Sensors* deal with capturing and abstraction within the CAMIS model. Based on the unimodal data processing, the *Multimodal Merging* unit combines extracted data on various abstraction levels based on semantic, spatial and temporal characteristics. Further information enrichment and contradiction checks on extracted information are performed based on domain knowledge within the *Semantic Enrichment* unit. Both the Multimodal Merging unit and the Semantic Enrichment unit address fusion tasks of the CAMIS model. The captured data from different modalities together with metadata as well as extracted and derived information on various semantic levels are managed and made accessible by the *Data Repository* unit. These units stated so far assemble the *MISTRAL Core System*. Different *Semantic Applications* can make use of the functionality of the core system and the great variety of available data on different semantic levels. Based on concrete application scenarios different aspects of retrieval and presentation related to the CAMIS model are addressed. (Gütl, 2008b)

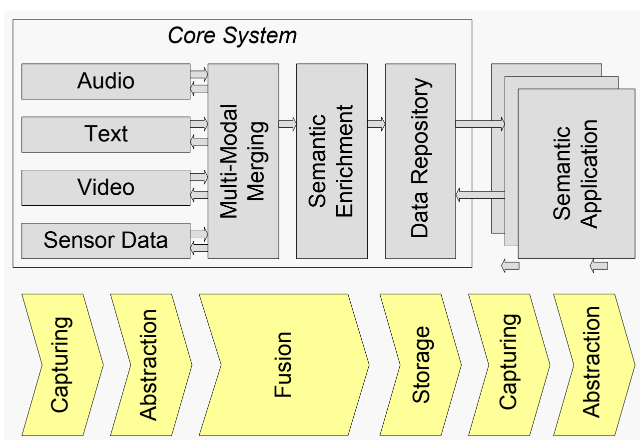

Figure 1: MISTRAL overall architecture and its mapping to the CAMIS units

Development of a Lecture Recording Indexing and Retrieval System

Based on the requirements stated above, we have adapted the MISTRAL prototype for the meeting application scenario in order to handle lecture recordings appropriately. For the first experiments and assessing the prototype, we have used and adopted lecture recordings from MIT Open Courseware (http://ocw.mit.edu/OcwWeb/web/home/home/index.htm) and from the Faculty of Arts and Sciences, Harvard University (http://www.fas.harvard.edu/). Given our main interest is on personalized retrieval and access of lecture recordings for learning purposes and the main focus is on the indexing and retrieval part, an understanding of the development of the core system and findings are outlined briefly in the following paragraph.

The core system has been extended in order to enable fetching and pre-processing of lecture recordings from free accessible repositories by developing a tailored spider component based on the xFIND gatherer (Gütl, 2002). Figure 2 depicts the overall situation composed of available lecture recording systems (either real life streaming or accessible by a repository), their connectivity via the Internet and the core system for processing and managing the lecture recordings for the indexing and retrieval process. One of the challenging issues in this context is to find appropriate content repositories and related documents linked to presentations and recordings. Our first approach is built on manually configuring the resource fetcher. Once the data are locally cached, the core components of the system can further pre-process and process the data accordingly. Audio and video data can be processed and converted into the real media format and related documents can be transformed in an internal data format for further processing. Furthermore, if text transcripts are not available from the repositories, a speech-to-text process creates the transcript. In case of too much ambient and background noise, most important words and concepts are extracted to support the retrieval process. Furthermore, presenters' behavior (gestures and facial expressions) as well as interaction with the presentation media can be processed by the sensor unit. However, provided presentation recordings lack information on these modalities. The information extracted so far is further merged by the multi-modal merging unit and annotated with further information by the semantic enhancement unit. This process steps enables to link related content (e.g. lecture notes or background knowledge) with scenes of lecture recordings or to segment the recording on a semantic level. At this stage we have faced the problem, that appropriate domain knowledge on the course subject and meta-knowledge on lecture activities must be modeled, which is time consuming and needs to be adopted for each of the subject domains. Overall, we believe that the MISTRAL approach is flexible enough to handle appropriately lecture recordings too. However, referring to the identified issues and challenges above, there is much space for further research and improvements.

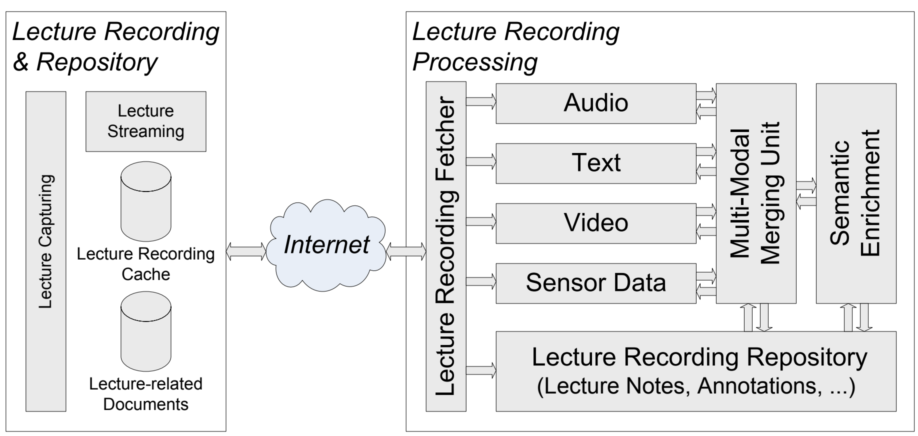

Figure 2: Conceptual distributed Architecture of Lecture Recording Repository and Processing for retrieval and access Purposes.

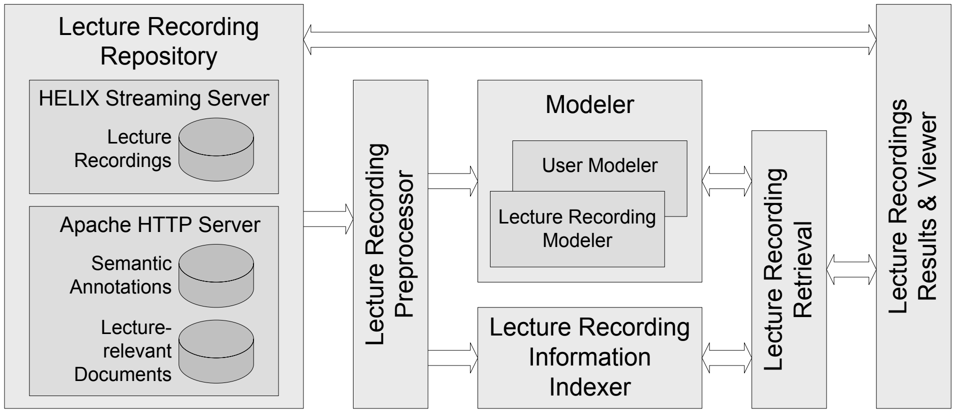

Figure 3: Simplified Architecture of Presentation Recording Retrieval and Visualization.

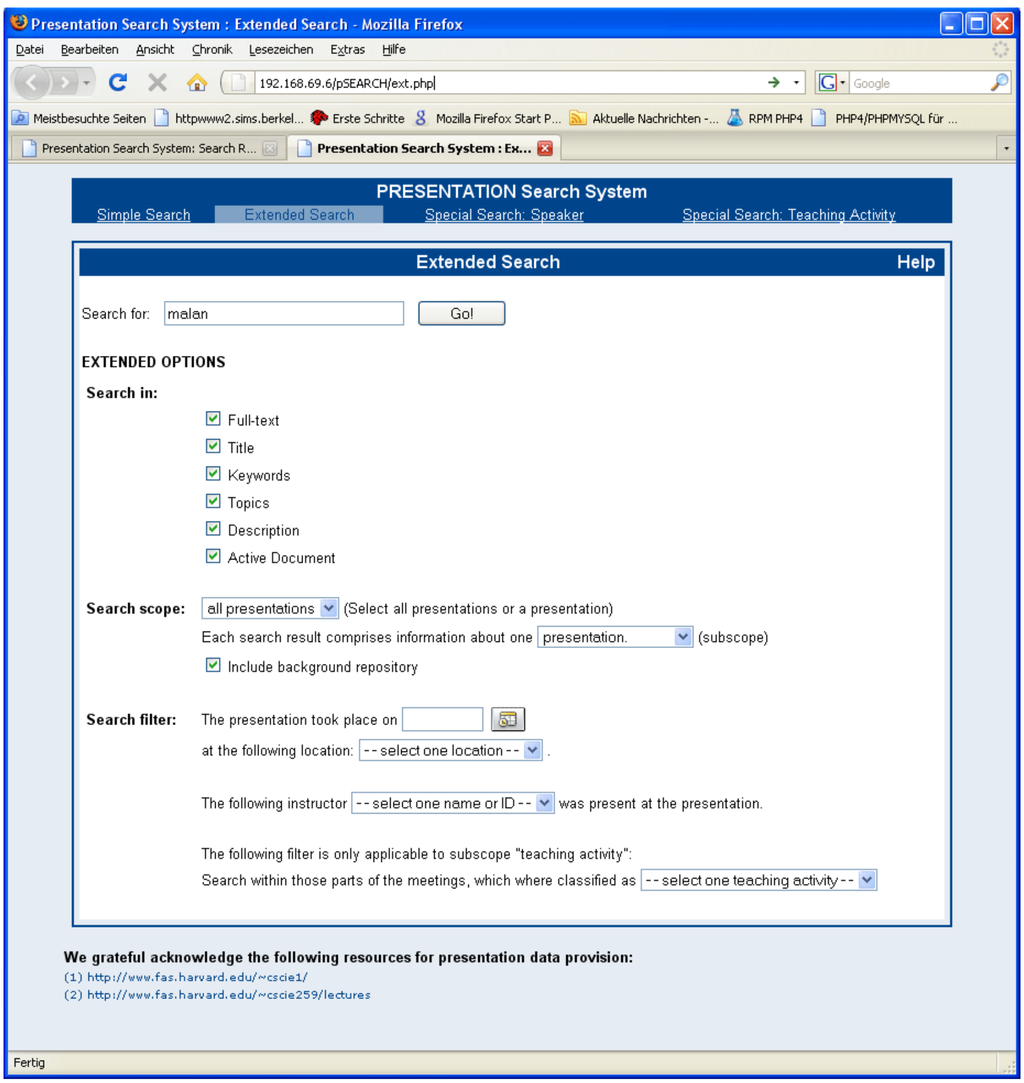

Figure 4: Extended Search Form.

Figure 5: Search Result Presentation.

By further focusing on the lecture recording retrieval and access part of the system, the architecture and components have also been adapted accordingly to the above mentioned requirements. This part of the system is based on the open source retrieval system xFIND (Gütl, 2002; García-Barrios, & Gütl, 2006). Figure 3 shows the simplified architecture for presentation recording retrieval and visualization purposes. In order to further process and deliver presentation recordings in a common media format, we have decided to cache all presentation information in our *Lecture Recording Repository*. To support streaming of the media and enable users to access particular scenes, all recordings have been converted to real video and real audio streams. In addition all presentation relevant documents (such as presentation slides and lecture notes) as well as extracted and enriched annotations (both by humans and computer tools) are cached in the repository. A *Lecture Recording Preprocessor*, a very specific instance of the pre-processor component built on the xFIND gatherer, is used to prepare the data (recordings, related documents as well as annotations) to feed the modeler and the indexer. The *Lecture Modeler* component holds information about the lectures (such as title, starting time, duration) and links the information with user information about the lecturer. The *Indexer* component takes into account extracted information from the different media streams, metadata and modeler information and builds indices which enable to search within the composed information. The *Retrieval* component handles user queries, transforms the requests in order to get search results from the indexer but also from the modeler component. The components are built on an adapted xFIND Indexer and xFIND Broker (Gütl, 2002) in order to deal with the very specific metadata fields for the indexing as well as for the retrieval process. The *Lecture Recording Result and Presentation Viewer* component provides the interface to the users by applying a flexible template engine. It enables users to search, retrieve appropriate search results and access to the presentation recordings in various granularities, e.g. to the entire presentation or to part of the presentation. Our developed prototype provides three search forms, e.g. a single search form with pre-defined search parameters, an extended search form with high flexibility for compiling tailored search queries (see Figure 4) and a special search form supporting specific search scenarios. Figure 5 depicts an example of search results on the granularity of entire lecture recordings. According to the search scenario, more fine grained result presentations are available. The search result listings do not only show some metadata and text snippets from the presentation recording but thumbnails from the presentation scene, from related presentation media and other related documents will also be shown. Lastly, Figure 6 depicts the presentation recording viewer, which sonorously provides presentation screens and related presentation media according to the selected search result.

In order to link lecture recording information properly with personalized learning activities, we have applied our former developed Dynamic E-Learning Knowledge Repository (DEKOR) which aims at the provision of a set of dynamically compiled information based on the concept level in the context of learning activities. This approach enables teachers and students to define relevant concepts for specific learning tasks and link the concepts with search queries for dynamic content delivery by using the lecture recording information retrieval system or other retrieval system. Further details can be found in (Gütl, 2007).

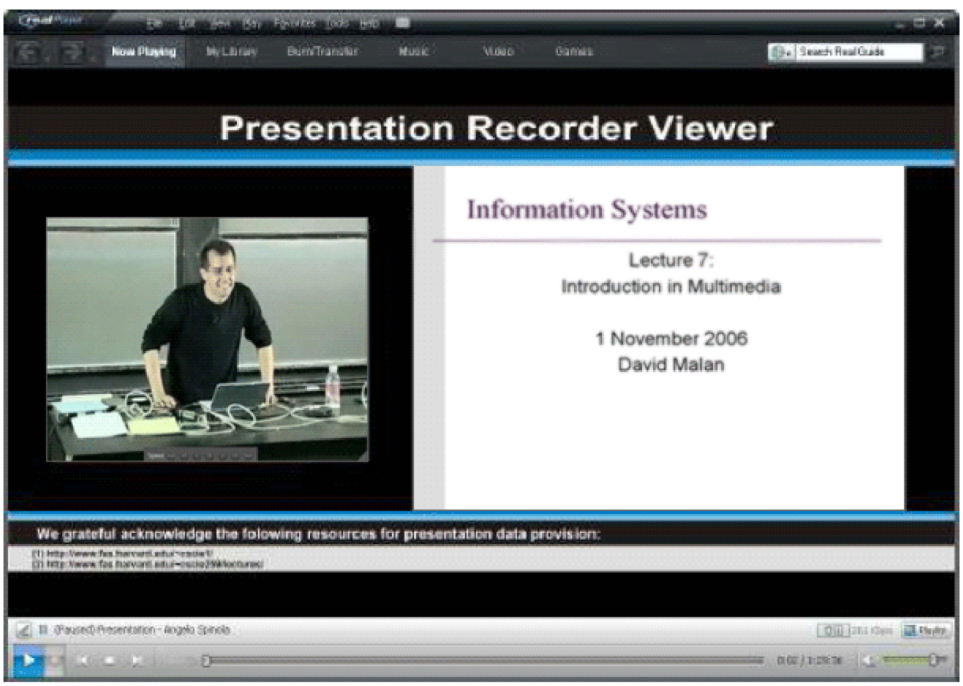

Figure 6: Example of a lecture recording presentation (video stream on the left side and presentation slide on the right side).

Findings and Lessons Learned

From a general perspective MISTRAL's processing, indexing and retrieval components have been adapted with less effort to the specific needs for presentation recordings. The flexible architecture and system design of the MISTRAL has supported to meet the requirements given by the presentation recording scenarios.

From the viewpoint of finding and fetching lecture recordings, the challenging task is not only to find the lecture recording but also to identify related presentation slides, lecture notes, course content and other relevant background information. Repositories are organized differently and relationships between the different media types are hard to find automatically. In particular related media to the recording are not always accessible, they may lack metadata or are provided in different media. To overcome this problem either specific instances of fetcher must be implemented or repositories to provide their data and metadata in a common data structure.

From the data processing and annotation point of view, the most serious issue is that neither sufficient information about the interaction with the presentation media is provided nor can it be automatically extracted. Furthermore, speech-to-text extraction in presentation settings is error prone because of background noise. Both problems have strong impact to indexing and retrieval quality.

From the retrieving and information visualization point of view, we have seen that the provided functionality meets the requirements given by students' and teachers' application scenarios. Search functionalities on different abstraction levels (segments, lecture, and course) or logical levels (lecture, topic, etc.) support both user groups to retrieve relevant learning content.

Conclusions and Future Work

In this paper we have argued that a great variety of accessible lecture recordings and related documents can support learning activities in different learning settings. However, available data in different file formats, in different repositories and appropriate retrieval over multiple repositories are hard to perform. This situation has motivated us to initiate research on a lecture recording information system which is based on the MISTRAL system.

With less effort the processing, indexing and retrieval components have been adapted to the specific needs for presentation recordings. Experiences have also shown that the MISTRAL multimodal and multimedia information system is designed flexible enough that the requirements in the presentation recording scenarios are applicable, and the retrieval functionality meets students' and teachers' requirements.

Altogether, the initial experiences of the lecture recording information system are promising. However, a great deal of further research is still to be conducted. Thus, for future work, we intend to administer research on how to find lecture recording repositories automatically and to link these recordings with other relevant documents. Since speech-to-text transcripts are important for good retrieval performance, we also want to work on enhanced methods taking into account speaker characteristics, knowledge domain of the meeting and information from the presenter's PC. Finally we are also interested in applying the system to real-life scenarios in academic and vocational settings.

Acknowledgements

Parts of project results presented in this work were developed within the MISTRAL research project. MISTRAL is financed by the Austrian Research Promotion Agency (http://www.ffg.at) within the strategic objective FIT-IT (project contract number 809264/9338). The support of all members but in particular of following IICM members is gratefully acknowledged: Dr. Victor Manuel García-Barrios, DI Helmut Mader and DI Martin Ruhmer. The authors also gratefully acknowledges the following resources for open and free access to lecture recordings: Faculty of Arts and Sciences, Harvard University (http://www.fas.harvard.edu/) and from MIT Open Courseware (http://ocw.mit.edu/OcwWeb/web/home/home/index.htm)

References

Abowd, G. D., Atkeson, C. G., Feinstein, A., Hmelo, C. , Kooper, R., Long, S., Sawhney, N., & Tan. M. (1996). Teaching and learning as multimedia authoring: The classroom 2000 project. In Proceedings of the ACM Conference on Multimedia -- Multimedia'96, pages 187-198, 1996.

Bacher, C., Müller, R., Ottmann, T., & Will, M. (1997). Authoring on the Fly A new way of integrating telepresentation and courseware production. Proceedings ICCE . Kuching, Sarawak, Malaysia.

Camstasia Studio. (2008). Last retrieved December 8th, 2008 from Camstasia Studio: http://www.techsmith.com/camtasia.asp

E-Chalk (2008). E-Chalk. Official Web site, last retrieved December 8th, 2008 from http://www.echalk.de/

García-Barrios, V.M., & Gütl, C. (2006). Semantic Applications on MPEG-7 Descriptions of Multi-modal Meeting Corpora: First Results. Special Issue "Multimedia Metadata Community Workshop Results 2005 Dissemination", JUKM (Journal of Universal Knowledge Management), 1, 1, 2006, 45-53.

Gütl, C. (2002). Ansätze zur modernen Wissensauffindung im Internet: Eine Annäherung an das Information Gathering and Organizing System xFIND (Extended Framework for INformation Discovery), PhD Thesis at Graz University of Technology, Austria; pp. 84-97, 2002.

Gütl, C. (2007). Moving Towards a Generic, Service-based Architecture for Flexible Teaching and Learning Activities. In C. Pahl (Ed.) Architecture Solutions for E-Learning Systems (peer-reviewed), Peer-reviewed book chapter, Idea Group Inc., Hershey, USA, pp. 1-24.

Gütl, C. (2008). Enhancements of Meeting Information Management and Application for Knowledge Access and Learning Activities. J.UCS, 14, 10, 1625-1653.

Gütl, C. (2008b). Automatic Extraction, Indexing, Retrieval and Visualization of Multimodal Meeting Recordings for Knowledge Management Activities. In M. Granitzer, M. Lux, and M. Spaniol (Ed.) Multimedia Semantics – The Role of Metadata (peer reviewed), Springer, Berlin / Heidelberg, 239-261.

Huang, W., Eze, E., Webster, D. (2006). Towards integrating semantics of multi-media resources and processes in e-Learning. Multimedia Systems, Springer Berlin / Heidelberg, Vo. 11, No. 3, März 2006, 203-215.

Mertens, R., Ketterl, M. & Vornberger, O. (2007). The virtPresenter lecture recording system: Automated production of web lectures with interactive content overviews. In International Journal of Interactive Technology and Smart Education (ITSE), 4 (1). Troubador publishing, UK. February 2007. S. 55-66.

MITNEWS (2007). MIT develops lecture search engine to aid students. MIT News Office, November 7th, 2007, last retrieved December 8th, 2008 from http://web.mit.edu/newsoffice/2007/lectures-tt1107.html

MITOCW. About OCW, MIT OpenCourseWare site, last retrieved Jan. 20th, 2008 from http://ocw.mit.edu/OcwWeb/web/about/about/index.htm

Lecturnity (2008). Lecturnity – Rapid Authoring Tool. Official Web site, last retrieved December, 2008 from Lecturnity: http://www.lecturnity.de

OLI. Open Learning Initiative. Official Website, Carnegie Mellon University, last retrieved December 8th, 2008 from http://www.cmu.edu/oli/index.shtml

Rajkumar, K.; Gütl, C.; & Ramadoss, B. (2008). Discovering Knowledge from Multi-modal Lecture Recordings. In proceedings of ICDEM 2008, Tiruchirappalli, India, pp. 1-3.

Spinola, A. (2008). Improved Management, Retrieval and Access of Presentation Recordings for Enhanced Learning Processes. Unpublished Diploma Thesis, University of Madeira, Funchal, Portugal, December 2008.

Ziewer, P., & Seidl, H. (2002).Transparent TeleTeaching. In Proceedings of ASCILITE 2002, Auckland, NZ, December 2002.