ABSTRACT
This demonstration illustrates different ways to support users dealing with recorded live presentations in order to improve the usability of the corresponding documents. It highlights different problems in this context and presents solutions and alternative approaches for both, multimedia indexing and query processing as well as user interface issues in order to support users who are skimming or browsing such documents in search for information.

Categories and Subject Descriptors
H.3.3 [Information Storage and Retrieval]: Information search and retrieval – retrieval models, search process, selection process.

General Terms

Keywords
Presentation Recording, Multimedia Indexing and Retrieval, Skimming and Browsing of Continuous Multimedia Data

1. INTRODUCTION
Automatic presentation recording, that is, capturing of a lecture or live presentation in order to automatically produce multimedia documents that can be used for (e-)learning, has been a research focus for a couple of years now. Projects such as Classroom 2000 [1], the Cornell Lecture Browser [2], or the AOF system [3] have developed prototypes and tools for this task and evaluated them and their usage in various real-world situations. Some of these systems have even become commercial products that are now routinely used in classrooms and lecture halls. Originally, research in those projects has mainly focused on the two questions; how the recording process can be automated and how a high quality of the resulting documents can be assured. Now that the automatic capturing problem is basically solved and presentation recording is becoming more and more a routine task, research interests in context with lecture recording have shifted to other problems, mainly in context with further processing, storage, and organisation of the produced files. One important aspect, which is the focus of this demonstration, is the question – how a detailed indexing and retrieval, as well as an easy and flexible access to the files and comfortable replay functionality can and should be realised.

Our experience with presentation recording and the usage of the resulting files in real-world situations identified a strong need for such features. For example, students usually do not replay the recorded lectures as a whole but only selectively access particular parts of interest (compare the study presented in [4]). The produced documents are rather long (often more than one hour) and the information a user is looking for might be distributed over several parts of the file. In addition, multimedia documents are generally not as easy to skim and browse by humans as static information such as text or a set of still images. All of these arguments result in a strong need for detailed indexing and reliable query processing, as well as comfortable skimming and browsing functionality. However, most of them are basically true for all kinds of long, unstructured, continuous multimedia documents that are used in an interactive setting where a user is not just passively consuming the content but really “working” with the data. Therefore, a lot of research has already been done in this context. However, both, the task, as well as the special nature of the documents used here pose several different and additional challenges where new approaches have to be found or existing ones have to be adapted to the corresponding circumstances. Therefore, this demonstration presents a system illustrating these problems and introduces the different approaches in order to solve them. A search engine that introduces alternative indexing and query processing is presented (Section 2) as well as various ways to support a user when skimming the data in search for relevant information (Section 3).

2. INDEXING AND QUERY PROCESSING
Many common approaches for indexing and retrieval of text as well as multimedia data have limitations in context with recorded lectures. When dealing with a recorded live presentation, the two most important media streams in context with indexing and query processing are the slides stream containing the slides and annotations of the presentation, and the audio stream containing the voice of the lecturer.

2.1 Indexing of the slides stream
Generally, information from the slides and annotation stream can be retrieved by just using the text from the slides to create a searchable index. However, most traditional approaches for text indexing are based on the two assumptions that a term has a higher relevance if it appears more often in a particular document than another term with less relevance (mathematically modeled
through the so called *Term Frequency, TF*) and if it appears in rather few documents of the whole collection (mathematically modeled through the so called *Inverse Document Frequency, IDF*). Such approaches that depend on the overall term distribution in a document collection can be critical if those documents contain mainly keywords instead of continuous sentences. In favour of a short and compact presentation, it is rather uncommon to use single terms much more than once when preparing slides (even if they have a high relevance in a particular context) but presenters often use other features instead to express relevance of single terms, such as a different colour or font style. Therefore, we developed an alternative approach for searching and ranking where different heuristics are used that consider these additional features. In addition, initial experiments with a machine learning-based approach have proven to increase retrieval performance significantly and will be presented in this demo as well.

2.2 Audio-based indexing

In context with audio processing, the most common approach for retrieval is to use automatic speech recognition to produce a textual transcript that can then be used to create a searchable index. Those approaches have proven to work well even with relatively high speech recognition error rates (see [5], for example) because of the high redundancy of the speech signal. However, one big problem with this approach is what is known in the speech recognition community as the *Out-of-vocabulary (OOV) problem*: Common speech recognition systems usually work with a fixed vocabulary and only words from this set can be recognised by the system. This problem is particularly critical in context with retrieval of recorded lectures because of the high frequency of technical terms. To solve it, we present an approach how the words found on the corresponding slides of an audio recording can be used to automatically generate topic related vocabularies in order to improve retrieval performance.

3. ACCESSING, SKIMMING, AND BROWSING THE FILES

This demonstration takes a global view on the overall search process, which is why not only indexing and query processing is considered but multimedia data skimming as well. Comfortable browsing and skimming functionality is an important and essential part of any search process, because search engines usually do not only return relevant data but also documents that have no or limited importance to a user in a particular situation. Reasons for this are limited precision of the indexing and retrieval algorithms as well as ambiguous and vague user queries. Therefore, a good interface design should support a user in filtering relevant from irrelevant documents. Since relevant information might be distributed over several parts of a document, comfortable access, browsing, and navigation has to be supported, as well.

First, we present several ways of how the results of the search engine introduced above can be represented to the users in order to support their relevance decisions. Recorded lectures contain much information that might be useful or important for a user in order to identify relevant parts of the documents. However, it is not clear, which information should be shown and what is the best way to represent it. For example, is it more appropriate to present thumbnails of the slides of the retrieved lectures, which have the advantage of illustrating the overall layout but the disadvantage that the text is often not readable, or should the text form the slides be extracted and presented directly? Is it important to represent some information about the content of the audio signal (e.g., excerpts from the produced transcript which might contain lots of speech recognition errors) as well, or is it sufficient to just use a graphical illustration of the relevant parts of the file?

In addition to these approaches, which are all based on skimming static information that is extracted from the continuous media files, we also present several techniques for time-compressed skimming of the media streams, i.e., the slides and annotation stream, as well as the audio stream. Visual skimming and navigation is supported via a synchronisation model we originally introduced in [6] and which allows for real-time random access of any position of a recorded lecture. With this implementation, a user can visually navigate or scroll through a document at any time, independent of speed and direction using a regular slider or scrollbar, a feature which has proven to be very useful for quick navigation and skimming. For audio skimming, we integrated some techniques which were pioneered by Arons [7] and have been proven to work well in various other applications as well.

4. ACKNOWLEDGMENTS

This work was supported by the German Research Foundation (DFG) as part of the research initiative V3D2 ("Distributed Processing and Delivery of Digital Documents").

5. REFERENCES


Additional comments to the demonstration

I did not include a video of the demonstration since it is rather interactive and therefore hard to illustrate in a video. Instead, I added the following comments and snapshots in order to show the basic idea and planned flow for the demo.

Hardware requirements

The whole demo runs on a laptop. No network connection is needed (If a connection is provided, we could also show an online demo of the system, what would be a nice feature but is not really necessary). Depending on the available space at the conference we might bring or rent an additional monitor.

Outline of the demo

1. Recording of a (very short) live presentation with our presentation recording system
2. Automatic generation of an index of the recorded file (no audio processing, just text/slide-based indexing, therefore very quick, takes only a few seconds)
3. Illustration of retrieval performance with different ranking functions (traditional TF*IDF-based ranking vs. heuristics and machine learning-based ranking with additional (i.e. better) features)
4. Different options for result presentation together with an illustration of the corresponding advantages and disadvantages
5. Different functions for skimming and navigation of the files in order to support search and relevance decisions

Comments:

• **Step 1 and 2** are optional and can be skipped in favour of a shorter demo time. However, they can be useful in order to develop a better understanding of the overall problem. The produced index will be added to a larger index containing several lectures that were recorded at our university. No audio-based indexing will be done at the demo, because it takes too long, but indexing will only be based on the slides used in the presentation. However, we will bring a pre-calculated audio-based index in order to illustrate typical problems with audio indexing, so search will be done based on both, a slides- as well as audio-based index for all files in the database besides the one that is recorded in Step 1.

• **Step 3**: Online versions of our search engine can be found at [http://ad.informatik.uni-freiburg.de/aofSE](http://ad.informatik.uni-freiburg.de/aofSE) for slide-based retrieval or at [http://ad.informatik.uni-freiburg.de/mmgroup/aofSEaudio](http://ad.informatik.uni-freiburg.de/mmgroup/aofSEaudio) for audio-based retrieval, respectively (see also snapshots on the following pages). Please note, that these are only online-demos which do not represent the current state of the search engine. The offline demo which is presented at the conference integrates both, slide- as well as audio-based retrieval, in a single interface.

• **Step 4**: The following pages show some snapshots of different result presentations for our search engine. The snapshots were taken from two evaluations we did with slide-based or audio-based retrieval, respectively, in order to identify advantages and limitations with the different variants. The results of the first evaluation will be published at this years HCI International 2003 conference, a proposal for publication of the results of the audio-based interfaces’ evaluation is currently under work but not yet submitted anywhere. The demonstration will also show a final, new interface design that considers the outcome of both of these evaluations.

• **Estimated duration**: The demo can be done quite flexible for both, overall duration as well as the presented content. The minimum amount of time needed to present everything should be about 10 minutes, but a more detailed and deeper demonstration of about 30 minutes can be given, as well. Depending on the interest of particular attendants the focus can be more on multimedia indexing (step 2&3) or on user interface aspects (step 4&5) or a general overview of the whole system.

• **User interaction**: The demo can be given in a pure presentation manner but user interactivity is possible and welcome. Please note that all the snapshots on the following pages are taken from lectures given in German. However, we recently recorded two complete courses given at our university by a guest professor from Australia (about parallel or mobile computing, respectively) which were given in English and which are used as database for this demo. Therefore, the attendants can try the search engine and test any kind of result presentation as well as the navigation and skimming functionality offered by the system.
Snapshots

Snapshots of the online version of the search engine (left side: slide-based search, right side: audio-based search)

Entering some search terms brings up a list of relevant lectures.

Selecting an entry from this list brings up detailed information about the corresponding lecture. (In this case, i.e., the online version, all slide titles & some relevance information is shown. The demo that is presented at the conference allows the user to select between different representations of the corresponding lecture content (compare interfaces/snapshots shown on the next page) and integrates slide- and audio-based retrieval in one interface.)

Selecting a lecture starts the player for audio replay together with a whiteboard window showing the used slides and annotations (both streams, audio & whiteboard, are replayed synchronously, annotations are not shown as snapshots but are replayed continuously). Using the scrollbars in each window, the users can visually and acoustically skim through a document.

Result representation for audio-based retrieval
Some snapshots of different versions for the interface design of the search engine (slide- as well as audio-based search, first six or last three snapshots, respectively)

Three interfaces, where the same information (i.e., metainfo about relevant slides & lectures) is shown but arranged in different ways

- A list of relevant slides is shown
- A list of relevant lectures is shown
- Relevant slides are grouped according to the corresponding lectures. (Lectures ranked by relevance)

Three interfaces, where different information about the content of a lecture is shown

- Thumbnails of relevant slides are shown
- Text that is extracted from the slides is shown
- Relevant slides of a lecture are represented along a time line and illustrated using different colours

Interfaces showing information about the content of the audio stream of the recordings

- Relevant slides of a lecture are represented along a time line and illustrated using different colours
- Excerpts of relevant parts of the automatically generated transcript of the audio files are shown
- Automatically generated keywords for each lecture (extracted from the respective audio stream) are shown