Multimedia Information Retrieval from Recorded Presentations

Wolfgang Hürst, Rainer Müller, Christoph Mayer
Institut für Informatik, Universität Freiburg, D-79110 Freiburg, Germany
{huerst,rmueller,cmayer}@informatik.uni-freiburg.de

1 Abstract

In presentation recording special effort is usually put into the automation of the production process, that is in automatically creating high quality data files without much or any need for manual recording and post-editing [5]. With the advent of such systems and their usage in classroom teaching, at conferences, etc., there is an increasing need for techniques and abilities which enable users to search in those documents and to localize some specific information. In this paper we describe how we integrated information retrieval techniques into the Authoring on the Fly (AOF) system, an approach for automatic presentation recording. We have chosen the AOF system for two reasons. On the one hand, it is a well-established way for presentation recording, used by various universities and institutions. On the other hand it is general enough to illustrate typical problems and challenges a developer is facing when designing a system for information retrieval from multimedia data streams which occur in the presentation recording scenario.

2 AOF as an Example for Presentation Recording

With the Authoring on the Fly (AOF) approach multimedia presentations are given with an electronic whiteboard program which is used to present material such as slides and annotations as well as external applications (e.g., computer animations or MPEG movies). During a presentation all data streams, that is, the audio, the slides and whiteboard annotations, the video of the lecturer, as well as the commands controlling the external applications, are automatically recorded. Hence, the produced file is a multimedia file which consists of several data streams of various media types. AOF offers a flexible and convenient way for the synchronized replay of these data streams (see [4] for a detailed description of the synchronization model used in the AOF system). The main features separating AOF from other approaches for presentation recording are that the produced multimedia file offers high quality with a reasonable amount of data and that navigation in and access to the recorded data is supported in a convenient way. The latter feature is a result of the used synchronization model [4] which realizes random access into the recorded file during replay, i.e., any data stream, independent of its media type, can be accessed at any position in real-time. Two other features, which can be realized because of the random accessibility, are random visible scrolling and unrestricted cross-referencing. Both are important for information retrieval and will be introduced in Section 3.2. For a detailed description of AOF and a comparison with other approaches for presentation recording we refer to [5].

3 Retrieving Recorded Presentations

The process of finding information in (multimedia) data is usually a two-tiered approach. First, a user enters a query and the computer presents some results, which were identified as relevant to that query. Second, those (potential) results are browsed and scanned by the user to find the needed information. While being necessary for any kind of retrieval, the second step becomes particular important with multimedia data (compare [6] or [2] respectively).

3.1 Query Processing

In general, the information a user is looking for can be located in any of the available media streams. However, not all streams have to be indexed and searched to achieve reasonable retrieval results. This is true because of the redundancy, which exists between the different media types. In our system we can use all streams interchangeably for indexing because a (time-)relationship is implicitly defined through the way how the media streams are synchronized.

The main stream we use for indexing is the one containing the slides and the whiteboard annotations. The reason is that this stream is relatively easy to index because of the existence of reliable text retrieval techniques. Additionally, it leads to reasonable results because the slides usually contain a representative overview of the part of the talk which is under current focus of the presenter. Slide transitions can be used for segmentation of a document, since a single slide usually represents a self-contained content unit. This way, (long) lectures are segmented into shorter, self-contained parts, which will be presented to the user as results to a particular query. Cases where some information is split over two or more consecutive slides, are not that critical, because of the way we present the results to the user (compare Section 3.2).

After extracting the words from the slides we produce an index using standard text indexing techniques. If the
slides are stored in a format that does not allow for direct access of the data (e.g., GIF images) we use commercial OCR software to get the text from the images. Application specific techniques can be used to extract text from external applications as well. For indexing we use standard techniques, such as stop word lists and word stemming.

Weighting techniques based on the distribution of index terms within a document, such as the well-known Inverse Document Frequency \( IDF \), turned out not to be useful for ranking single slides. The reason is that slides usually do not contain continuous text, but only descriptions of the according topic using keywords. This fact influences the distribution of these terms in the whole document significantly. Therefore, we introduced several heuristics for the ranking of the slides, instead of relying on statistical values. For example, an index term occurring on one slide is considered more relevant and thus given a higher weight, if the slide is shown to the audience for a longer time than another one which contains this index term as well (In [3] the same heuristic is used to identify important slides while creating a summary of a video recording of a presentation). Additionally, an index term is weighted according to its position and appearance. For example, it is assumed to be more relevant, if it appears in the headline or if it has a color or style (italic, bold, etc.) which makes it stand out from its surroundings.

3.2 User Filtering
After placing a query the user generally has to filter the retrieved results. This filtering process is important in multimedia information retrieval and crucial to the success and usability of any system. Even if this task has to be done by him/herself, a good system can and must offer facilities and features to support the user in information filtering. This can be done, for example, by the way the user interface is designed and by offering convenient navigation and browsing techniques.

Interface Design: Presenting Retrieved Results
Not only relevance information is important during the filtering process, but the context a retrieved data fragment appears in and its position within the whole document as well. For example, in classroom teaching, new terms are usually defined before they are used. As a consequence, data which was recorded at the beginning of a lecture can be of higher relevance for a user than some other slides (and the data in the according streams) which were shown at the end, even if those were given a higher ranking by the retrieval algorithms. Additionally, our user studies indicated that users do not like to jump permanently between different presentations. Instead of this, they prefer to examine all relevant positions found in one particular lecture, before looking at another presentation, which contains relevant parts as well.

Hence, the ubiquitous approach to present retrieved results as a ranked list is not useful in our scenario and alternative presentation techniques are needed. We decided to present the results in two steps. First a ranked list of documents is shown, each one indicating one whole recorded presentation (see Fig. 1 a). After selecting any of the retrieved presentations (or starting the replay of the file) an overview is presented to the user which shows the title of all slides of the according document (see Fig. 1 b) as well as additional scoring information. This information indicates the relevance of this part of the document for the query placed by the user.

This presentation has several advantages over a ranked list. First, the position of a potential result within the whole document is always clearly shown, an information which can be very important for the filtering process, as described above. Additionally, the context, that is, the data surrounding the retrieved fragment, is not lost. For example, if a topic is discussed on three or four consecutive slides, they might be of more interest to a user than a single slide, even if it is ranked much higher than each of the other slides. Such situations can be identified very easy, since we preserve the original order of the slides when presenting the retrieval results to a user.

Browsing Retrieved Results
Common approaches for browsing multimedia data can be classified into two categories\(^2\). The first one is to give an overview of the data by extracting some parts that are representative for the content. This can be static data, such as keyframes of a video or thumbnails of slides, as well as dynamic data, such as short video or audio clips. The other approach is to use time compression techniques to quickly browse the files. Audio or visual information is replayed at a higher frame rate, thus allowing the user to get an idea of the content even if not being able to understand it completely. The fast forward function of a regular VCR is an example for time compression of a video, while [1] describes a system, where audio files containing speech recordings can be replayed at a higher frame rate for the same purpose. In practice, combinations of both approaches are realized as well.

We use a mix of techniques, which can be classified in both groups. Some outstanding positions within the document are indicated in the viewer once the replay is started.

\(^2\)It should be noted that other classifications exist as well (e.g., see [2] or [3]).
random visible scrolling

(Fig. 2). Those positions, such as the beginning of external applications and transitions from one slide to another (Fig. 2 a, b), allow the user to gather information about the content of the respective data. For example, a user can browse through a recorded presentation just by looking at any slide or external application. This kind of browsing resembles skimming through a book by, for example, just looking at the included images.

Additionally, the title of each slide is already shown in the retrieval interface (compare Fig. 1 b or Fig. 2 c), thus giving the user a first hint about the content of the according data fragment. Those indications represent a link into the document at the respective position. The random accessibility provided by our synchronization model does not only allow the user to start the replay at these links, but to jump between the according positions during replay without any restrictions or delays. We call this feature unrestricted cross-referencing. It offers a convenient way for the user to verify the relevance of the retrieved results shown in Fig. 2 c. This possibility to jump between retrieval results in real-time during replay introduces a clear advantage over traditional approaches for browsing multimedia retrieval results.

Time compression techniques are very helpful to get an overview or to localize a particular position within a single data fragment or the whole file. However, instead of using approaches with a fixed frame rate for replay, we leave it to the user which part he/she wants to browse at which speed. That is, the user can visually scroll through the document at any time and with any speed by using a regular scrollbar (Fig. 2 d). This allows for a much more granular and flexible skimming than with many other approaches that also offer different but fixed frame rates for scrolling. An information a user is looking for can be found much easier this way. For example, the position where a particular handwritten annotation was made can be found very fast and localized accurately, since any visual change in the media streams is indicated immediately. We call this feature random visible scrolling.

4 Summary and Discussion

We illustrated how we integrated information retrieval techniques into an existing system for automatic presentation recording. Our approach offers query processing as well as browsing and skimming functionality. Special focus has been put on the user interface and comfortable navigation mechanisms that support the user in filtering retrieved data.

First experiments with the system have proven to achieve good retrieval results. However, further evaluation has to be done and is the focus of our current work. To do this, we use data which was recorded at our university during one class this term (two lectures each week). We also implemented a tool which enables us to (anonymously) log user behavior during replay of a recorded presentation. By knowing if, when, and how often someone makes use of the features which we provide for browsing, we gain knowledge about user acceptance and usability of these techniques. Even if more data has to be collected to identify reliable user profiles, there is a clear trend in the data we evaluated so far which indicates that users like the functionalities we proposed and are using them intensively.

References


Acknowledgements

This research is supported by the German Research Society (DFG) project Generation of Multimedia Documents on the Fly, No. Or 64/13-1, part of the research initiative V1D2. Additionally, we like to thank Bernd Zupancic who implemented the module for logging user behavior during replay which we use for evaluation.