SPEECH@FIT LECTURE BROWSER

Igor Szöke, J. "Honza" Černocký, M. Fapšo, J. Žižka

Brno University of Technology, Speech@FIT
Brno, Czech Republic

szoke@fit.vutbr.cz

ABSTRACT

This paper describes an innovative web-based browser used for video recordings of lectures that is built on speech and image processing technologies. The aim of this project is to simplify the access to information that is spread across video recordings. This is mainly achieved by coupling to the speech search engine and due to a possibility to quickly navigate through an automatically generated list of slides presented. The reader is briefly acquainted with the technological background of the browser; the emphasis is laid on the use of the browser from the user point of view.

Index Terms— multimedia browser, e-learning, search in speech

1. INTRODUCTION

Over the last few years, audiovisual recording of lectures and seminars has seen an enormous growth across many academic institutions throughout the whole world. The recorded material is often made available on-line on an Internet website, usually accompanied by meta-data, such as keywords, categories, title, author or links. These help users find the lecture they need. However, since lecture recordings are usually long in duration (typically over an hour or two), and often contain a lot of interesting information, a possibility to search inside video is a crucial and desired feature. Otherwise, users can spend a lot of time watching or listening to something that is not much important until they get to really useful parts. In any case, a browser capable of searching in audio/video may dramatically increase the accessibility of recorded lectures, seminars or meetings. Up to now, a considerable number of lecture browsers have been developed, however, only a few support searching in audio/video data. Our browser enables users to search for what was spoken during all lectures or the search can be restricted to a specific one. Each lecture can be accompanied by a smart list of automatically synchronized slides and other useful related information.

Our goals were:

- to create a lecture browser with an intuitive and easy-to-use web-based interface.
- to make the browser work in a standard web-browser, without a need to install any special executables and libraries.

2. SEARCH IN SPEECH

Navigating in speech brings the most significant speed-up in accessing lecture videos. In our system, speech is first off-line transcribed to word lattices by a large vocabulary continuous speech recognition (LVCSR) system, and indexed. The browser then connects to the Brno University of Technology (BUT) speech indexing and search engine [1]. Based on lattices at the output of LVCSR, a forward index is created, which is then converted to an inverted index similarly to standard text-search applications. The search allows for estimation of confidence based on comparison of the actual and the best path in the searched lattice. Although BUT is active in word/sub-word approach to search in speech eliminating the out-of-vocabulary word (OOV) problem [2], the current version of the system supports search in word lattices only.

.

3. IMPLEMENTATION

Our web-based lecture browser is powered by Lighttpd[1] that, in connection with a module for HTTP Flash video streaming, enables users to play the video starting from any key frame. Web pages are dynamically generated using PHP, JavaScript and AJAX. For the sake of easy navigation, lectures are grouped into categories and subcategories based on the course, term and year. Besides that, a breadcrumb trail is provided at the top of each page to help users keep track of their location.

Each page has the same header and footer, only the middle part differs. Based on that, there are three main page types - search results page, category page and lecture page. The search results page is accessed via the input field allowing the user to search in audio data and shows all occurrences matching the search query across all lectures. Each category

[1] http://www.lighttpd.net

page displays a list of assigned lectures along with their basic information. Clicking on any lecture name or image brings the user to the lecture page used for the lecture playback. This page consists of various components organized in boxes that each can be collapsed/expanded by clicking anywhere on its title bar.

4. BROWSER COMPONENTS

Each component is designed to perform different functions and provide users with a piece of information related to a lecture.

4.1. Video Playback

This component enables users to watch lecture video recordings. To have our browser widely accessible on many platforms, we decided to use a Flash-based solution for video playback and chose the popular and flexible JW FLV Media Player[2]. It lets users switch to the full-screen mode and supports subtitles through the use of a special plug-in that can handle W3C Time-text[3] XML files.

4.2. Slides

This component displays slides as they are presented. The data is stored in an XML file that can be automatically pregenerated by our slides synchronization tool[4]. Slides are usually shown in the video recording along with the lecturer. However, in many cases, those containing small font text are difficult or even impossible to read due to low resolution and video compression. To reduce this problem, this component can display sharp images of slides matching those shown in the video. It is divided into three sections. The upper one shows the current slide image. A click on it pops up a new window showing the image in its maximum resolution. The middle one is used for a table that contains a clickable list of all slides (thumbnails). This enables the user to easily browse through the lecture and quickly access the desired slide position in the video. Multiple occurrences of slides in the lecture are marked. This reflects situations when the lecturer returns to a previously presented slide. Finally, the lower section may provide links to presentation materials, mostly PDF or PowerPoint files.

4.3. Search in Audio

The actual search is performed by our speech search engine that can run as a non-blocking server either on the same or a different computer. This component provides the user with an input field and a search button. A search query (a keyword or a sequence of keywords) is sent to the server to be processed. As long as there are some occurrences, the server returns results formatted in XML. Otherwise, it responds with an error message. The communication between the client and the server is done through AJAX. An important advantage of AJAX is that it is asynchronous, so data is transferred and processed in the background without interfering with the existing page. This allows users to continue watching the lecture. Once the results are obtained from the server, they are displayed along with their confidence scores. Results with the highest confidence are shown first. Each row represents one occurrence matching the search query that can also be accompanied by the surrounding transcript segment. The user can begin playing the video from any occurrence. Search queries might be logged for later analysis. This can help extend the recognition vocabulary if the word is undefined.

4.4. Speech Transcript

In our browser, this component displays a written form of what was spoken. Transcription is the 1-best output from our speech recognizer. The output is divided into short textual segments. Each contains its starting point in the lecture timeline and a corresponding transcript part. In the lecture playback mode, the segments are highlighted as they are played. The current segment can also appear in the video as a caption. By clicking on a table row, users can begin playing the video from the starting point of the assigned segment.

5. CONCLUSION

We have developed a browser that significantly helps users navigate in lecture recordings. This is achieved not only by coupling to the speech search engine, but also by providing a navigation list of all slides presented. Furthermore, an intuitive web interface and hierarchical lecture categorization ensure that the desired lecture is quickly found. Using a Flash player to play video recordings, the browser runs on many computers with different operating systems. A demo is available at http://www.superlectures.com.

6. REFERENCES

[1] Černocký, J., et al.: Search in Speech for Public Security and Defense. In Proc. IEEE Workshop on Signal Processing Applications for Public Security and Forensics, Washington D.C. (2007).

[2] Szöke, I., Burget, L., Černocký, J., Fapšo, M.: Sub-word modeling of out of vocabulary words in spoken term detection. In Proc. 2008 IEEE Workshop on Spoken Language Technology, Goa, IN (2008), pp. 273–276.

[2] http://www.longtailvideo.com/players/jw-flv-player/

[3] http://www.w3.org/AudioVideo/TT/

[4] see detailed transcription at http://www.superlectures.ccm/about.php