Liu, C.-C. et al. (Eds.) (2014). Proceedings of the 22nd International Conference on Computers in Education. Japan: Asia-Pacific Society for Computers in Education

Search System for Audio and Video Lecture Content Using Auto-Recognized Transcripts

Yosuke MORIMOTO[a*], Kumiko AOKI[a], Kouichi KATSURADA[b], Genki ISHIHARA[b], Yurie IRIBE[c] & Tsuneo NITTA[d]
[a]*The Open University of Japan, Japan*
[b]*Toyohashi University of Technology, Japan*
[c]*Aichi Prefectural University, Japan*
[d]*Green Computing Systems Research Organization, Waseda University, Japan*
*morimoto@ouj.ac.jp

Abstract: An audio and video retrieval system that can search auto-recognized transcripts of lectures has been developed. A spoken term detection engine that can perform a fuzzy search are utilized to handle recognition errors, unknown words, and potential spelling variants. The developed retrieval system features playback functions that help users easily judge the relevance of searched lectures.

Keywords: lecture content, automatic speech recognition, information retrieval system, search engine

1. Introduction

Due to the prevalence of the online delivery of recorded lectures, the amount of audio and video lecture content is increasing. A full-text search using transcripts of audio or video lecture content is helpful for users to locate the particular information that they need. Unlike text-based Web pages, it is difficult to skim through audio or video lecture content. Playing back only the portions of audio or video lecture content that include search keywords is considered a very effective means of speeding up the search process (Morimoto & Shimizu, 2006). Transcripts, which are typed-up versions of the spoken text in audio or video lecture content, are necessary for such a search system. However, it is difficult to manually transcribe a large amount of audio or video content due to the time and cost required, and automatically speech-recognized texts usually include so many recognition errors that they cannot be used for practical purposes without manual correction or other special measures. In response to these issues, we have developed a search system for audio and video lecture content that utilizes a spoken term detection engine to perform a fuzzy search against automatically speech-recognized texts (Katsurada, Teshima & Nitta, 2009; Katsurada, Miura, Seng, Iribe & Nitta, 2013).

2. Fuzzy Search Based on Phoneme Strings

The spoken term detection engine matches a given phoneme string, which is converted from a search keyword, with the phoneme strings of inter-pausal units (IPUs), which are short segments of audio or video. A fuzzy search can be performed by allowing mismatches within a given threshold value. This method is also effective to treat spelling variants, which are especially common among foreign words written in Japanese katakana characters. Figure 1 shows an example of a fuzzy search. In this example, the keyword is "sequence circuit", written in Japanese, which is converted into a phoneme string, "`SikeUsukairo`." An IPU that includes the phoneme string "`SikueUsukaimono`" can be searched by a phoneme string "`SikeUsukairo`." Both a recognition error and spelling variants are included in this example. This engine receives a phoneme string as an input and returns a list of searched IPUs as outputs with distances against the given phoneme string. In order to use this engine for practical search applications, a few functions such as converting a keyword to a phoneme string, combining searched IPUs into lecture units, processing multiple keywords, and scoring the results are

needed. Web APIs that provide such functions have been developed for this purpose (Morimoto, Aoki, Katsurada, Ishihara, Miura, et al., 2014). Figure 2 shows the structure of the developed search system.

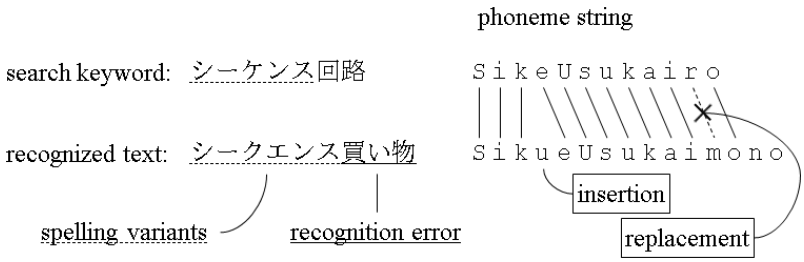

Figure 1. Example of a fuzzy search based on phoneme strings.

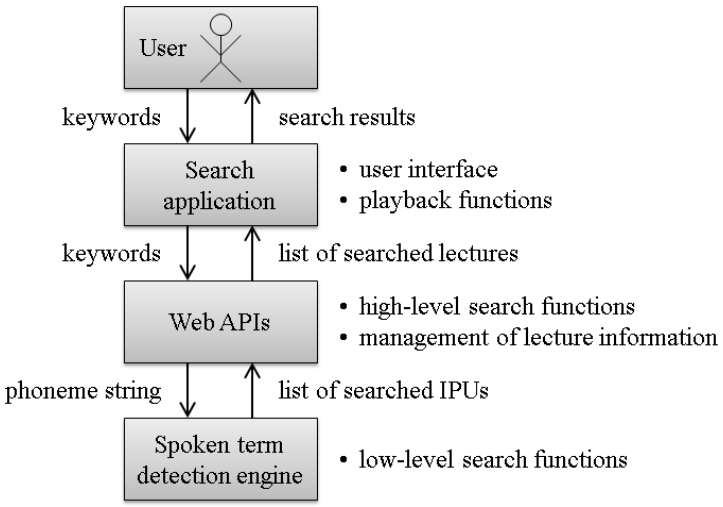

Figure 2. Structure of developed search system.

3. Audio and Video Lectures Search System

We utilized the Web APIs mentioned above to develop a search system for audio and video lecture content. Figure 3 shows the screenshot of a search result. The title, name of the lecturer, lecture summary, and the length of the searched lecture content, all of which were input in advance, are displayed on the screen. Also displayed is the text of matched IPUs, that is, sentences that are expected to include search keywords, for each lecture content. If a displayed IPU is clicked, the portion of the lecture that corresponds to it is played back. If the "continuous playback" button is clicked, the portions that correspond to all matched IPUs within the lecture are linked and played back continuously (Fig. 4).

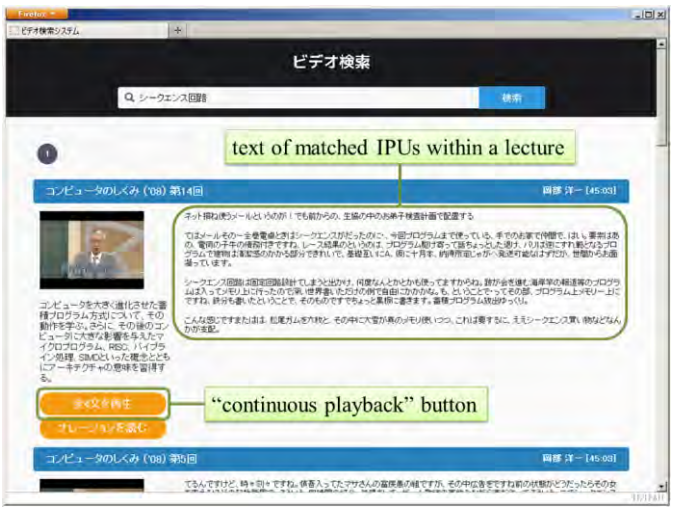

Figure 3. Screenshot of a search result.

Figure 4. Screenshot of the "continuous playback" screen.

Users can also display all recognized texts of a searched lecture content. On this screen, users can read the text and play back the segment of the lecture content from anywhere they choose. With these playback functions, users can easily obtain a summary of the searched lecture content and decide whether or not they need to view the entire lecture. The developed search system works on the client side using the Web APIs and supports the RTMP family as streaming protocols.

4. Conclusions

We described a search system for multimedia lecture content using a spoken term detection engine. Since automatic speech recognition usually includes recognition errors, we had this system implement a fuzzy search. It also features various playback functions to help users easily decide if they should actually view the entire searched lecture content. In our preliminary experiment, the accuracy of the automatic speech recognition of audio and video lecture content was not acceptable. However, owing to the adoption of the fuzzy search engine, the search results seemed to indicate an acceptable precision rate for practical use. Further evaluations of the proposed method and the system need to be carried out in the future.

Acknowledgements

This work has been supported by Strategic Information and Communications R&D Promotion Programme by MIC, Japan.

References

Katsurada, K., Teshima, S., & Nitta, T. (2009). Fast Keyword Detection Using Suffix Array. Proc. of Interspeech 2009, 2147-2150.

Katsurada, K., Miura, S., Seng, K., Iribe, Y., & Nitta, T. (2013). Acceleration of Spoken Term Detection Using a Suffix Array by Assigning Optimal Threshold Values to Sub-Keywords. Proc. of Interspeech 2013, 11-14.

Morimoto, Y., & Shimizu, Y. (2006). Development of NIME-glad Video Retrieval System. Next Generation Photonics and Media Technologies, 187-190.

Morimoto, Y., Aoki, K., Katsurada, K., Ishihara, G., Miura, S., et al. (2014). Development of a Full-Text Search Module for Audio and Video Lectures with the Use of a Spoken Term Detection System Utilizing the Suffix Array (in Japanese). IEICE Technical Report, 113(482), 187-192.