Audio Subtitle Mapping System between Slide and Subtitles by Co-occurrence Graphs on VOD Lecture

Kohei SAKANE[a], Nobuyuki KOBAYASHI[b] , Hiromitsu SHIINA[c*] & Fumio KITGAWA[c]
[a]*Graduate School of Informatics, Okayama University of Science, Japan*
[b]*Faculty of Human Sciences , Sanyo Gakuen University, Japan*
[c]*Faculty of Informatics, Okayama University of Science, Japan*
*shiina@mis.ous.ac.jp

Abstract: Many VOD lectures are played alongside a video, making it difficult to understand some content due to mishearing something or not catching the sound on a slide. Therefore, the ability to display audio subtitles that explain data or paragraphs on slides may improve student understanding of some lectures. This system will generate co-occurrence graphs for each paragraph and subtitle on a slide (Microsoft PowerPoint), and depending on the type of co-occurrence graph, and seek correspondence relations. These correspondence relations are used to support understanding of audio information by converting it to textual data and presenting displays and printouts of explanatory information on subtitles for each part of a slide.

Keywords: VOD Lecture, Co-occurrence Word, Mapping Subtitle to slide

1. Introduction

Presentation of video and slides is a basic function of e-learning systems. E-learning systems are increasingly being provided according to SCROM formats (Advanced Distributed Learning; Fallon & Brown, 2003), like Moodle (Dougiamas). Moreover, the effectiveness of e-learning systems has been developed many researches (Kitagawa & Onishi, 2007; Matsumoto, Miyauchi, & Koga, 2010). Moodle is a management system for e-learning systems, and they are continuing to develop a variety of features for it. However, its main purpose is the development of systems and lecture management functions.

In the form of studies related to research concerning lecture content, systems have been developed to conduct summaries of streaming data utilizing colors, caption information, and animation information(Watanabe, Wu, & Yokota, 2010), but the information utilized is unsuitable for use in ordinary lectures. Speech retrieval systems based on techniques utilizing the degree of relevance of subtitles (Itou, Fujii, & Ishikawa, 2001; Mouri, Funabuki, & Nakanishi, 2008), as well as video segment inference systems utilizing techniques that apply statistical processing based on frequency information (Kobayashi, Koyama, Shiina, & Kitagawa, 2011), are under development. Moreover, research on helping students understand content is being conducted with respect to assignment submission management systems (Seki, Matsui, & Okamoto, 2008) as well as writing assignment problems. However, these systems are intended to improve post-lecture understanding. In addition to these, systems development is also being conducted with respect to authoring systems for instructors.

One of the advantages for students taking courses in e-learning systems is the ability to pursue learning in a free environment over a network. For many VOD lectures, content is considered difficult to grasp because the viewer does not know what he or she has missed or how the slides correspond to what is said, as the lectures are simply played back in time with the video. It is thought that development has been slow on systems for making it as easy as possible to understand VOD lectures while watching them, by making lectures more comprehensible using information provided via e-learning systems.
Therefore, the ability to display subtitles that explain data or paragraphs on slides may improve student understanding of some lectures. This system will generate co-occurrence graphs for each paragraph and subtitle on a slide (Microsoft PowerPoint, later “PPT”), and depending on the type of co-occurrence graph, seek correspondence relations. These correspondence relations are used to support

understanding of audio information by converting it to textual data and presenting displays and printouts of explanatory information on subtitles for each part of a slide.

2. Audio Subtitle Mapping System Development for VOD Lecture

The system created in this study is being developed as an add-on to an e-Learning lecture system through a VOD system that uses a credit transfer system in an education consortium composed of six universities that centers on okayama university science in Japan.

The system retrieves slide data and audio data from content in the e-learning system, and after converting the audio data into text, associates utterance times and executes a similarity analysis of the slide and subtitles. The system is structured to present video, subtitles, and slide data of VOD lectures stored in a database. Videos display as is, and slides switch in sync with change times while subtitles matching slide text boxes are simultaneously displayed.

As an addition to standard VOD lectures, this system displays video of a lecture with subtitles on the upper left of the screen (Figure 1). In addition, it displays slide on the upper right side and mapping subtitle to text box of PPT.

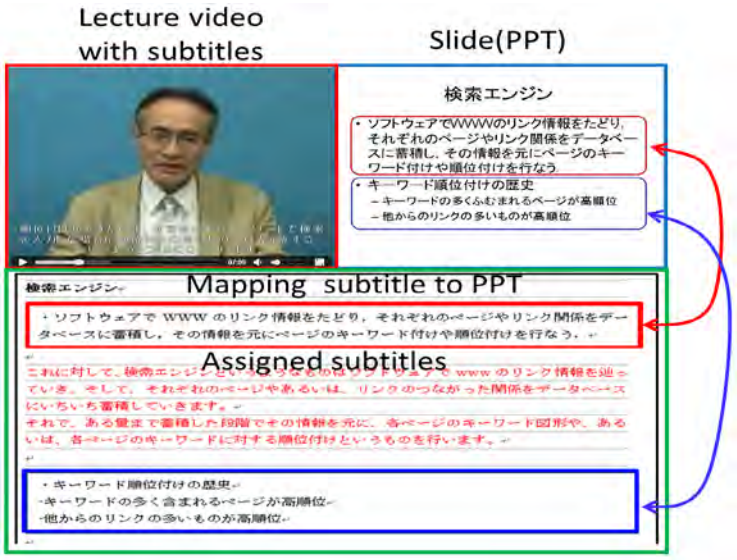

Figure 1. Screen of audio subtitle mapping system.

3. Mapping Subtitles in PPT

3.1 Co-occurrence Words and Co-occurrence Graphs

Co-occurrence words are generally considered words that appear alongside one another in the same text. A co-occurrence graph is generated using each page of a PPT and the subtitles corresponding to each PPT page as targets.

(1) Extracting PPT Co-occurrence Graphs
All descriptors in the PPT are divided into individual text boxes(Figure 2(a)), the retrieved text is partitioned into morphemes using morphological analysis(Figure 2(b)) (Kudo, Yamamoto & Matsumoto, 2004), and a co-occurrence graph(Figure 2(c)) is generated.

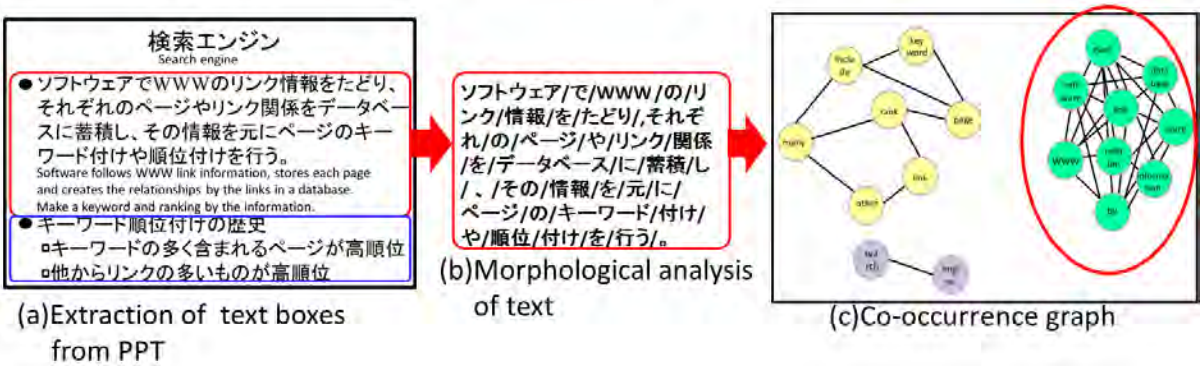

Figure 2. Generation Process of co-occurrence Graph by extracting PPT.

(2) Extracting Subtitle Co-occurrence Graphs

Subtitle co-occurrence graphs are generated by partitioning morphemes in subtitles for individual PPT pages. Then, a graph is generated by linking words that have co-occurrence relations using node and word proximities. Since the word count of a co-occurrence graph of the subtitles on each PPT page can be high, making it more complicated, the word count is refined. Refinement is executed using strength of co-occurrence between words by Jaccard similarity coefficient (Jaccard, 1921; Romesburg, 2004;Tan, Steinbach, & Kumar, 2005), and it eliminates edges of co-occurrence graphs. Finally words with less sum of Jaccard similarity coefficient are also eliminated.

For example, the relations between subtitles and co-occurrence graphs in Figure 3 are used the subtitles from section 2, page 6 of the VOD lecture “Database”. The left side table is list of subtitles, it is divide into each PPT pages. The right side is the co-occurrence graph which is assigned the refined co-occurrence graph to subtitles.

A part of co-occurrence graph which is created from subtitle is eliminated by Jaccard similarity coefficient, some subtitle is also eliminated in mapping system. We are considering that eliminated subtitles are not important. In case of co-occurrence graph is created from some subtitles, these subtitles assign to co-occurrence graph. Then mapping system assign these subtitle to text block of PPT.

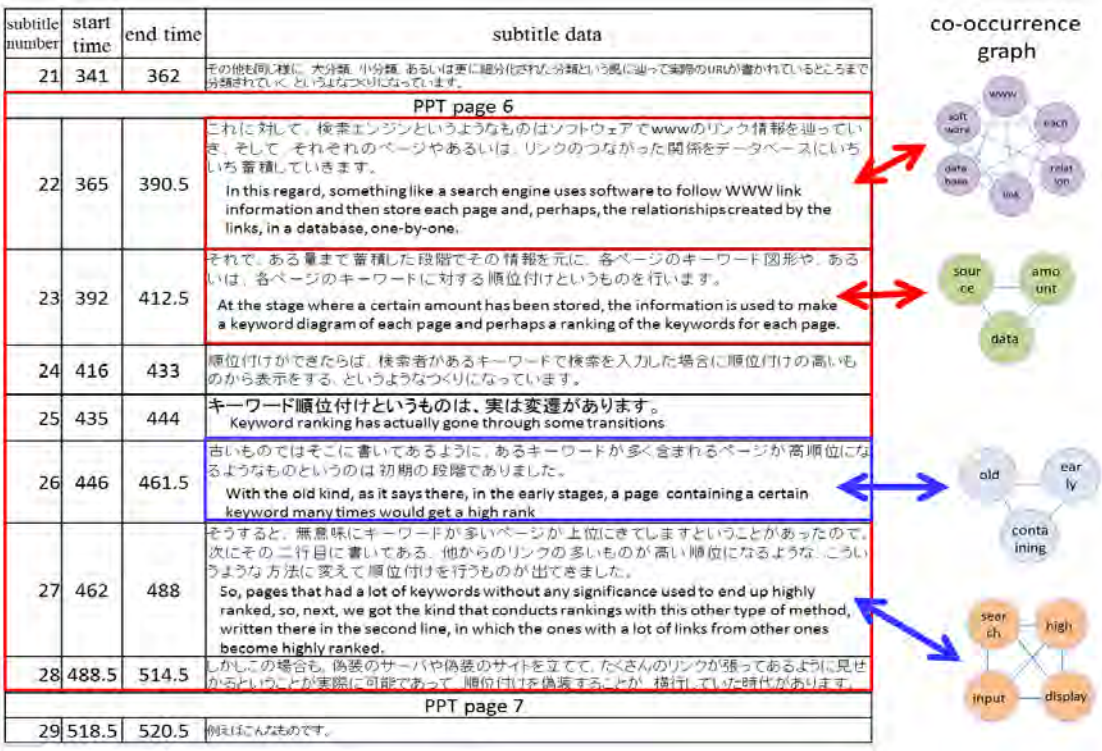

Figure 3. Mapping subtitles to refined co-occurrence graphs.

3.2 Co-occurrence Graph Corresponding Ratios

Based on a search of a corresponding count for the words in a subtitle co-occurrence graph and the words in a PPT co-occurrence graph, a corresponding ratio for the PPT co-occurrence graph and subtitle co-occurrence graph is generated with the following method.

Let C_{PPT} be the number of co-occurrence words in PPT, C_{Sub} be the number of co-occurrence word in subtitles, and $CP(G_{PPT}, G_{Sub})$ be corresponding ratio of G_{PPT} and G_{Sub}.

$$CP(G_{PPT}, G_{Sub}) = \frac{|C_{PPT} \cap C_{Sub}| \times 2}{|C_{PPT}| + |C_{Sub}|}.$$

In the PPT and its subtitles from section 2, page 6 of the VOD lecture "Database" , for example, when the word set for the co-occurrence graph created from the PPT C_{PPT} = {software, database, WWW, storage, data, relation, source, each, execute}, and the word sets for three types of co-occurrence graphs G_{Sub1} = {relation, WWW, software, follow, each, database}, C_{Sub2} = {amount, data, source}, C_{Sub3} = {search, high, input, display} the agreement ratios between co-occurrence graphs are $CP(G_{PPT}, G_{Sub1}) = 5 \times 2 / (9+6) = 2/3 = 0.6667$, $CP(G_{PPT}, G_{Sub2}) = 2 \times 2 / (9+3) = 1/3 = 0.3333$ and $CP(G_{PPT}, G_{Sub3}) = 0 \times 2 / (9+4) = 0/13 = 0.0000$. Figure 4 shows relation between co-occurrence graphs of PPT and subtitles.

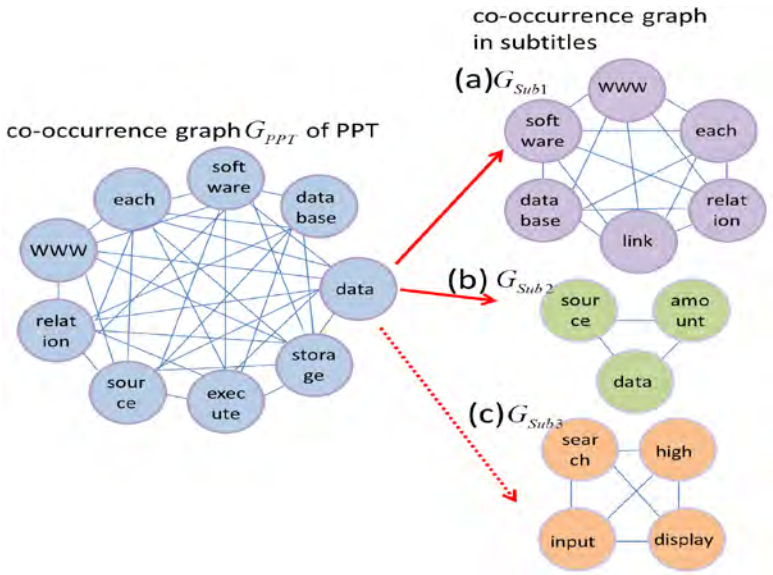

Figure 4. Corresponding ratio between PPT and subtitles

Summary which is not assign to text block of PPT is created. The ratio of summary original subtitles is defined the number of subtitle of summary divided by the number of original subtitles.

3.3 PPT and Subtitle Co-occurrence Words

Lastly, subtitle co-occurrence graphs with higher thresholds are correlated with PPT co-occurrence graphs, and subtitle co-occurrence graphs with lower thresholds are eliminated.

In the PPT and its subtitles from section 2, page 6 of the VOD lecture "Database", for example, while the agreement ratios between co-occurrence graphs are 0.6667, 0.3333, 0.0000, co-occurrence graphs with an agreement ratio with a threshold lower than 0.2 are removed from correspondence relations with PPT. An example of a PPT and subtitle co-occurrence graph is shown in Figure 5. In Figure 5, subtitle co-occurrence graphs corresponding to PPT co-occurrence graphs are encircled by the same color.

Summary which is not assign to text block of PPT is created. The compression ratio of summarization is defined as the ratio of number of sentences in assigned subtitle to the number of sentences in the original subtitles. In an example of the compression ratio of summarization for subtitles from section 2, page 6 of the VOD lecture “Database”, the compression ratio of the subtitles except the page only for title page and figures was: minimum is 0.167, maximum is 0.6 and average is 0.38.

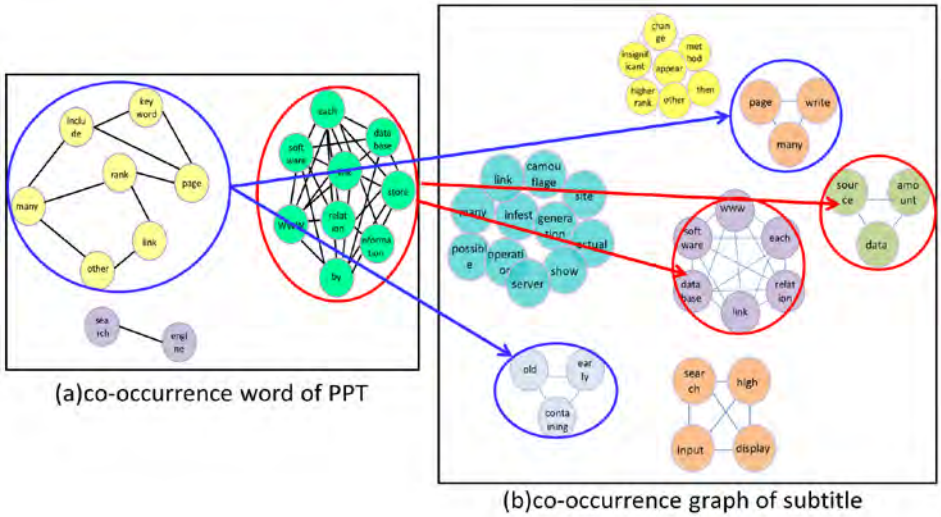

Figure 5. Refined mapping co-occurrence graph of PPT and subtitles.

4. Evaluating Subtitle and PPT Correspondence Relations

Correspondence relations between PPT and subtitles from section 2, page 6 of the VOD lecture “Database” were evaluated by Japanese students. The evaluation method was in two parts. The first was to rate the match level of corresponding subtitles at one of four levels: skewed, slight, frequent, and consistent. The second was to evaluate them as captions by assigning a rating out of five (1, 2, 3, 4, 5). Table 1 is result of evaluations.

Table 1: Evaluation by questionnaires.

PPT page number	Evaluation1		Evaluation2	
	P1	P2	P1	P2
No. 1(title page)	-	-	-	-
No. 2	consistent	5	consistent	5
No. 3	slight	3	slight	3
No. 4	slight	3	slight	4
No. 5	consistent	5	slight	2
No. 6	slight	2	slight	2
No. 7	slight	2	slight	2
No. 8	slight	3	slight	3
No. 9	slight	1	slight	1
No. 10	slight	5	skewed	4
No. 11	skewed	1	skewed	1
No. 12	consistent	5	consistent	5
No. 13(figure page)	-	-	frequent -	-
No. 14	frequent	4	slight	4
No. 15	slight	3	slight	4
No. 16	consistent	5	consistent	5
No. 17	slight	4	consistent	5

Ratings as captions were given according to the degree that the subtitles adhered to the content, with half or greater adherence receiving a rating of three or higher. For correspondence relations in

which the level of corresponding subtitles were judged “slight,” strict refinement when generating a co-occurrence graph was also taken into consideration.

5. Future Works

This system generates co-occurrence graphs from PPT and subtitles in PPT of VOD lecture for the purpose of facilitating student comprehension of lectures. Going forward, this system must be used, with foreign students as a target user ship, in order to evaluate its efficacy. We would also like to create a system that helps students gain a deeper understanding of lectures by generating summaries from PPT-compatible subtitles.

As future works, the subtitles will also possess the ability to add readings of kanji (Chinese characters) to assist foreign students. In addition, research of learning of terminology has been derived (Sun, Kashiwagai, Kang, Kiyomitsu, & Ohtsuki, 2011). As future research, we are considering that extracting data of terminology from actual lecture and developing learning system for terminology using its extract data.

References

Advanced Distributed Learning. *Advanced Distributed Learning Initiative: SCORM 2004 4th Edition.* Retrieved 8 20, 2014, from Advanced Distributed Learning: http://www.adlnet.org/scorm/scorm-2004-4th/

Dougiamas, M. *moodle.* Retrieved from moodle: https://moodle.org

Fallon, C., & Brown, S. (2003). *e-Learning Standards: A Guide to Purchasing, Developing, and Deploying Standards-Conformant E-Learning.* Florida: CRC Press.

Fujita, A., & Ishikawa, T. (2002). Extraction and Organization of Encyclopedic Knowledge Information Using the World Wide Web. *The IEICE Transactions(Japanese Edtion), J89-D*(9), 300-307.

Itou, K., Fujii, A., & Ishikawa, T. (2001). A Lecture-On-demand System using Spoken Document Retrieval. *IEICE Technical Report*, 55-60.

Jaccard, P. (1921). The distribution of the flora in the alpine zone. *New Phytologist*, 37-50.

Kitagawa, F., & Onishi, S. (2007). An Experiment on selective Course with Face-to-Face or/and E-learning, and Students Behavior. *Japan Society of Education Information, 22*(3), 57-66.

Kobayashi, N., Koyama, N., Shiina, H., & Kitagawa, F. (2011). Detecting Movie Segments Using Gaussian Mixture Models for VOD Lectures with. *The Journarl of Information and System in Education, 10*, 39-46.

Kobayashi, N., Shiina, H., & Kitagawa, F. (2009). A Search System by Word Frequency Distribution for a VOD Lecture with Subtitles. *IPSJ SIG Technical Reports, 2009-DBS-149*(7), 1-9.

Kudo, T. Retrieved 8 20, 2014, from MeCab: Yet Another Part-of-Speech and Morphological Analyzer: http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html

Kudo, T., Yamamoto, K., & Matsumoto, Y. (2004). Applying Conditional Random Fields to Japanese Morphological Analysis. *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing (EMNLP-2004)*, 230-237.

Matsumoto, K., Miyauchi, H., & Koga, A. (2010). Adapting the Drill-and-practice-type e-Learning System SANNO KNOWLEDGE FIELD to SCORM Standards and Devloping Its Contents. *The Journal of Information and Systems in Education*, 85-100.

Mouri, A., Funabuki, N., & Nakanishi, T. (2008). A Proposal of a Classification and Link Method of Lecture Scenes for Structured Learning. *IPSJ SIG Technical Reports, 2008-CE-93*(12), 87-98.

RomesburgH. C. (2004). Cluster Analysis for Researcher. NORTH CAROLINA: Lulu. com.

Seki, K., Matsui, T., & Okamoto, T. (2008). An Adaptive Sequencing Method of the Learning Objects for the e-Learning Environment. *The IEICE Transactions D, J86-D1*(5), 330-344.

Sun, Y., Kashiwagai, H., Kang, M., Kiyomitsu, H., & Ohtsuki, K. (2011). A Presenting Method Suited to Learners' Understandings in the Use of IT Dictionary. *Transactions of Japanese Society for Information and Systems in Education, 28*(4), 271-282.

Tan, P., Steinbach, M., & Kumar, V. (2005). *Introduction to Data Mining.* Boston: Addison-Wesley.

Watanabe, Y., Wu, Y., & Yokota, H. (2010). Digesting Online Multimedia Presentation Archives Based on Visual Effects. *Proc. Int. Workshop Streaming Media Delivery and Management Systems, SMDMS*, 477-482.