Spoken Document Retrieval by contents complement and keyword expansion using subordinate concept for NTCIR-SpokenDoc

Noboru Kanedera
Ishikawa National College of Technology
Tsubata, Ishikawa, 920-0392, Japan
kane@ishikawa-nct.ac.jp

ABSTRACT
In this paper, we report our experiments at NTCIR-10 IR for Spoken Documents (SpokenDoc) task. We participated SCR subtask of SpokenDoc. The keyword expansion using the subordinate concept and dictionary improved the mean average precision (MAP) from 0.320 to 0.324 for the lecture retrieval task. For the passage retrieval task, all of contents complement, keyword expansion, and subword were used. The subword was effective because a retrieving keyword was not contained in target in many cases. Moreover, it was found that a beginning subtopic is useful as topic information in the contents complement.

Categories and Subject Descriptors
H3.3 [Information Storage and Retrieval]: Information Search and Retrieval

General Terms
Algorithms, Experimentation, Performance

Keywords
NTCIR-10, Spoken Document Retrieval, Team Name: [INCT]

1. INTRODUCTION
The method of retrieving a subtopic of a lecture video is examined. When the subtopic of the lecture video is retrieved, we can use slide information [1] etc. However, only speech information is used in this study, because the lecture using blackboards that doesn't use the slide is targeted.

When speech recognition is used, measures against the speech recognition errors are important. The method of using the subword[2][3][4], a statistical translation technique[5] etc. have been proposed. The methods based on query expansion using related web documents [6][7] and corpus[8] have been also proposed.

In this paper, we report on the result of investigating which relationship is important among hypernym and hyponym relationships in retrieval keyword expansion. Moreover, we report the effect of the keyword expansion which used the subordinate concept and a dictionary, and the contents complement for spoken document retrieval for SCR lecture retrieval task and SCR passage retrieval task[15].

2. Spoken document retrieval method
2.1 Retrieval keyword expansion
In the text retrieval, the keyword expansion might be used. This study investigates how the retrieval keyword expansion is effective for the transcription text of lecture speech and the text by automatic speech recognition.

When the method of retrieval keyword expansion is classified in the source, the method of using the dictionary, the method of using information such as Web, and the method of using together, etc. are devised. A general dictionary doesn't often contain the technical term though it contains a lot of general terms. Oppositely, the possibility including the technical term is high, though the method of using information such as Web might not contain a general term.

The broader concept and the subordinate concept etc. are considered when classifying it by the concept. We investigated the method to use Japanese Wordnet[9] which can retrieve these concepts.

Figure 1 shows the method of retrieving the subtopic of lecture video. The index (TF-IDF) of each subtopic of the lecture video is obtained beforehand. When retrieving it, the specified keywords by a user are expanded to the association key words such as broader concepts, subordinate concepts, and simultaneous appearance words, etc. using the dictionary[10] and Wordnet. Next, the retrieval vector is generated.

The constant value α is set to the retrieval vector element corresponding to the retrieval keywords. The value $(1-\alpha)/$ (the number of associated keywords) is set to the retrieval vector element corresponding to the association keywords. The retrieval is done by comparing the retrieval vector with the TF-IDF vector of each subtopic.

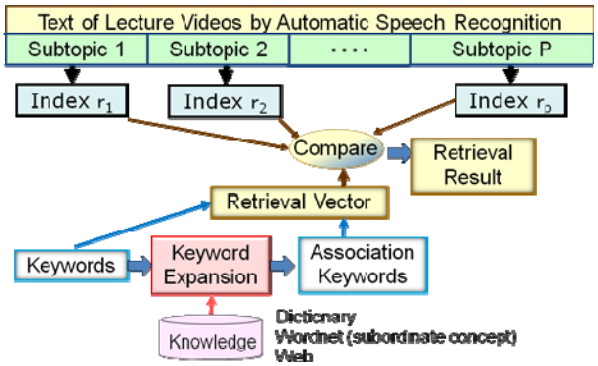

Figure 1. A lecture subtopic retrieval method by keyword expansion.

2.2 Contents complement

When searching a lecture etc., the information on subject may not be included in an applicable subtopic to search. Fortunately, a lecture mentions an abstract to the beginning in many cases. Moreover, the conclusion of the whole lecture may be contained in the last. Therefore, it is thought that the information on the subtopic of the beginning of a lecture, the last subtopic, and an adjoining subtopic is useful as topic information. Then, we define the retrieving vector $V_{i,j}$ of the j th subtopic under i th lecture as follows.

$$V_{i,j} = (1-\beta_s-2\beta_a-\beta_e)T_{i,j} + \beta_s T_{i,1} + \beta_a (T_{i,j-1}+T_{i,j+1}) + \beta_e T_{i,Ni} \quad (1)$$

$T_{i,j}$ is a TF-IDF vector of the j th subtopic under i th lecture here. N_i is the number of subtopics of the i th lecture. β_s, β_a, and β_e are the factors corresponding to the first subtopic, an adjacent subtopic, and the last subtopic, respectively.

2.3 Subword Retrieval

The keyword expansion can retrieve the subtopic to which the keyword doesn't appear directly by expanding the specified keyword. On the other hand, the method using the subword can deal with OOV (out of vocabulary)[4]. It is therefore preferable to use both the keyword expansion and the method using the subword.

Figure 2 shows a lecture subtopic retrieval method by subword model. Tri-phone was used as a subword. Confusion matrix between tri-phone models is made. Bhattachayya distance [3] was used for the distance between distributions of tri-phone models. The uttered keywords are converted into the tri-phone string, and collated with the tri-phone strings of subtopics by continuous dynamic programming (DP) using the confusion matrix. As a result of continuous DP, N sections where the spotting distance is small are extracted. In this report, N was assumed to be 500. N sections are sorted in order with small distance. The score of each section is defined as 1/(sorted rank), and it adds to the corresponding subtopic. The subtopic with a large score was assumed to be a subtopic retrieval result.

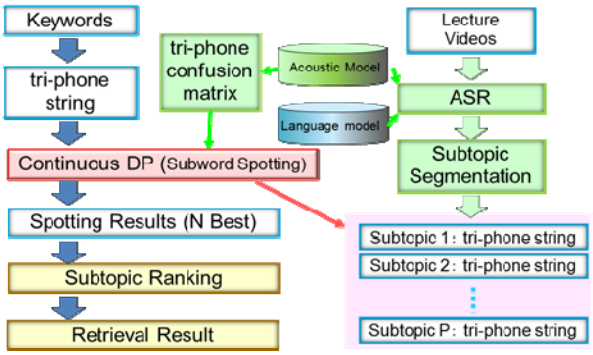

Figure 2. A lecture subtopic retrieval method by subword model.

3. Experiment

3.1 Experimental conditions

Table 1 summarizes the transcriptions used for each run. We used Pivoted Document Length Normalization [14] for the similarity between vectors.

Mean Average Precision (MAP) is used for evaluation measure. For each query topic q, top 1000 documents are evaluated. MAP is calculated as follows.

$$MAP = \frac{1}{|Q|}\sum_{q\in Q} AveP_q \quad (2)$$

where $|Q|$ is the number of all query topics Q. $AveP_q$ is calculated as follows.

$$AveP_q = \frac{1}{|R_q|}\sum_{k=1}^{M}\frac{k}{rank(r_k)} \quad (3)$$

where $|R_q|$ is the total number of correct answer documents. $r_1, r_2, \cdots, r_M$ $(M \le |R_q|)$ are the correct answer documents contained in top 1000 searched documents. $rank(r_k)$ is a ranking of a correct answer document r_k.

Table 1. Summary of the transcriptions used for each run.

task	group	run	transcription	candidate
lecture	INCT	1	REF-WORD	1-best
		2		
		3		
passage	INCT	1	REF-WORD	1-best

3.2 Evaluation results of SCR lecture retrieval task

Table 2 shows evaluation results for the lecture retrieval task. When the keyword expansion using both NDK basic dictionary [10] and the subordinate concept from Wordnet were used, MAP has improved from 0.320 to 0.324. The subword was ineffective. It was considered that the retrieving keyword was contained in the target.

Table 2. Evaluation results for the lecture retrieval task.

Group ID	run	MAP	Method
INCT	1	0.324	keyword expansion (wordnet+NDK)
	2	0.320	keyword expansion (wordnet+NDK) +subword retrieval
	3	0.320	baseline

3.3 Evaluation results of SCR passage retrieval task

Table 3 shows evaluation results for the passage retrieval task. All of contents complement, keyword expansion, and subword were used. The subword was effective because a retrieving keyword was not contained in target in many cases.

Figure 3 shows NTCIR-9 dry-run spoken document retrieval results with keyword expansion by Wordnet to ASR text. The expansion by using the subordinate concept was superior to those by using the broader concept. The keyword expansion using the NDK basic dictionary is effective to the same extent as the subordinate concept by Wordnet as shown in Figure 4. When the keyword expansion using both dictionary and Wordnet were used, MAP improved further.

The candidate for a retrieving is the subtopics divided every 30 sentences in 2702 lectures included in the Corpus of Spontaneous Japanese (CSJ) [11]. These lectures are a society lecture or a simulation lecture. The retrieving experiment was conducted to the transcription text obtained by using a word-based automatic speech recognition system for dry-running 39 query offered by NTCIR-9[13] .

Figure 4 shows NTCIR-9 dry-run spoken document retrieval results with contents complement and keyword expansion to ASR text. When all of contents complement and keyword expansion were used, the highest MAP value was obtained. Table 4 shows the optimum value of $\alpha, \beta_s, \beta_a,$ and β_e in contents complement and keyword expansion. Figure 5 shows the effectiveness of the keyword expansion with complement by using the beginning subtopic to ASR text. Figure 6 shows the effectiveness of the complement by β_s , with keyword expansion to ASR text.

Table 3. Evaluation results for the passage retrieval task.

Group ID	run	MAP	Method
INCT	1	0.0452	contents complement +keyword expansion (wordnet+NDK) +subword retrieval

Table 4. The optimum value of α, β_s, β_a and β_e in contents complement and keyword expansion (wordnet+NDK dictionary) for NTCIR-9 dry-run task.

	Optimum value			
	α	β_s	β_a	β_e
transcription text	0.45	0.10	0.03	0.0001
ASR text	0.45	0.15	0.005	0.0005

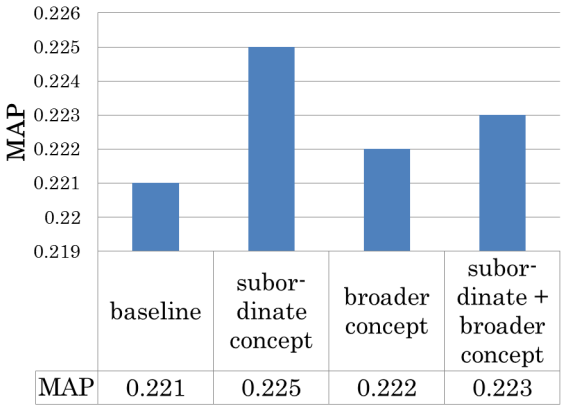

Figure 3. NTCIR-9 dry-run spoken document retrieval results with keyword expansion by Wordnet to ASR text.

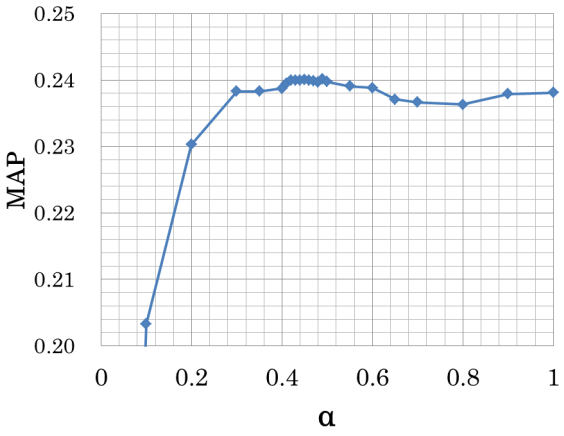

Figure 5. The effectiveness of the keyword expansion (wordnet+NDK dictionary) with complement by using the beginning subtopic to ASR text for NTCIR-9 dry-run task.

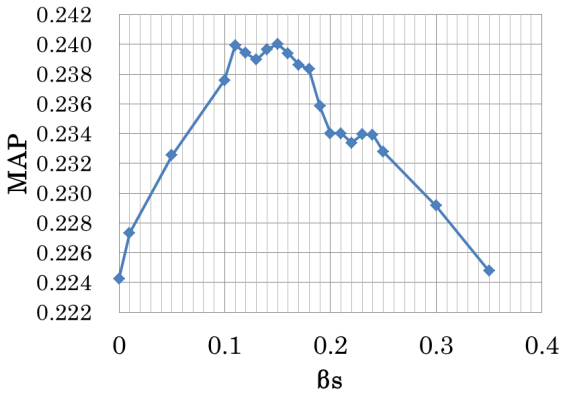

Figure 6. The effectiveness of the complement by using the beginning subtopic, with keyword expansion (wordnet+NDK dictionary) to ASR text for NTCIR-9 dry-run task.

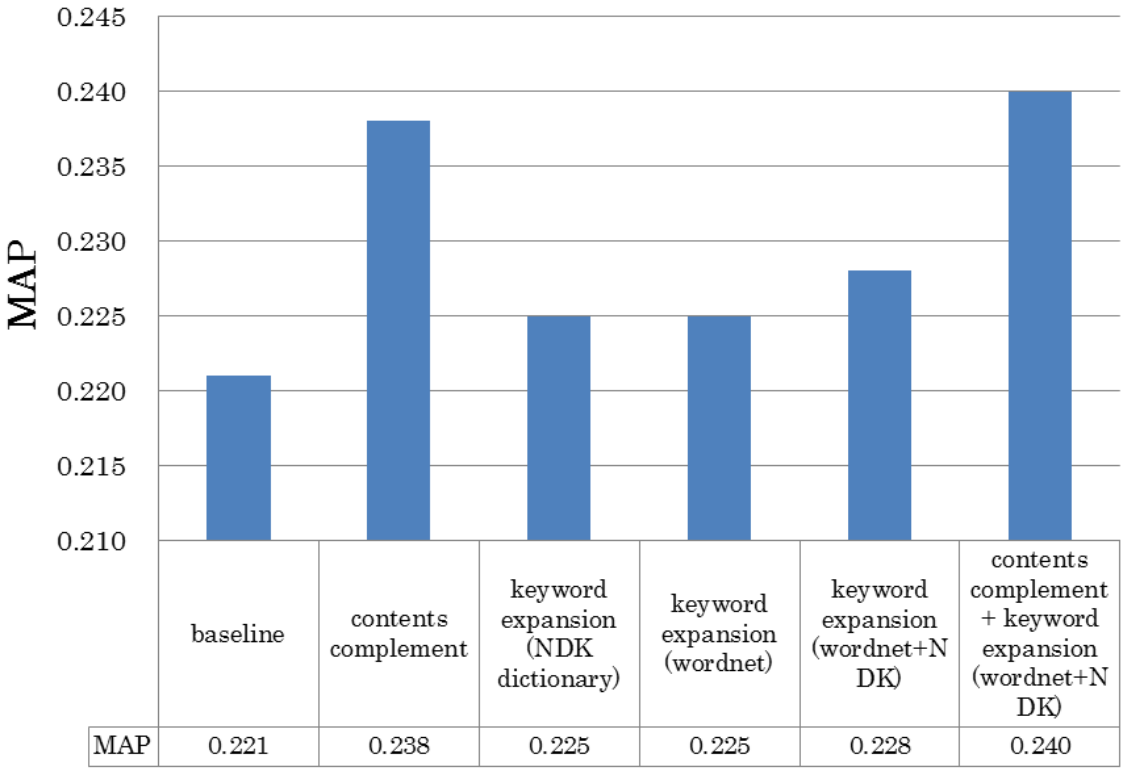

Figure 4. NTCIR-9 dry-run spoken document retrieval results with contents complement and keyword expansion to ASR

4. CONCLUSIONS

The method of spoken document retrieval was examined using the contents complement and keyword expansion. It was found that a beginning subtopic is useful as topic information in the contents complement. The expansion of the retrieval keyword by using the subordinate concept was effective. Moreover, the method using the contents complement and keyword expansion was better than the individual use.

5. ACKNOWLEDGMENTS

Part of this research was supported by Grant-in-Aid for Scientific Research (No. 23501192) from the Japanese Ministry of Education, Culture, Sports, Science and Technology.

6. REFERENCES

[1] Okamoto, T. et al., "Presentation-Content Retrieval Integrated with the Speech Information," IEICE Trans. Inf. & Syst. (Japanese Edition), J90-D(2), pp.209-222, 2007.

[2] Ng, K. and Zue,V.W., " Subword-based approaches for spoken document retrieval," Speech Communication, 32(3), pp.157-186, 2000.

[3] Iwata, K. et al., "An Investigation of New Subword Models and Subword Phonetic Distance for Vocabulary-free Spoken Document Retrieval System," IPSJ Journal, 48(5), pp.1990-2000, 2007.

[4] Nishizaki, H. and Nakagawa, S., "Robust Spoken Document Retrieval Methods for Mis-Recognition and Out-of-Vocabulary Keywords," IEICE Trans. Inf. & Syst. (Japanese Edition), J86-D-II(10), pp.1369-1381, 2003.

[5] Akiba, T. and Yokota, Y., "Spoken Document Retrieval by Translating Recognition Candidates into Correct Transcriptions," Proc. of the INTERSPEECH, pp.2166-2169, 2008.

[6] Terao, M. et al., "Open-Vocabulary Spoken-Document Retrieval Based on Query Expansion Using Related Web Documents," Proc. INTERSPEECH, pp.2171-2174, 2008.

[7] Uno, Y. et al., "Improvement of the Index using the World Wide Web for Spoken Document Retrieval," Proc. of the Fourth Spoken Document Processing Workshop, Online: http://www.cl.ics.tut.ac.jp/~sdpwg/sdpws2010_proceedings/SDPWS2010-09_uno.pdf, 2010.

[8] Singhal, A. and Pereira, F., "Document expansion for speech retrieval," Proc. of ACM SIGIR'99, pp.34-41.

[9] Japanese Wordnet, Online: http://nlpwww.nict.go.jp/wn-ja/

[10] http://www.ndk.co.jp/dictionary/dic_basic.html

[11] Maekawa, K., Koiso, H., Furui, S. and Isahara, H., "Spontaneous speech corpus of Japanese," in Proc. of the Second International Conference of Language Resources and Evaluation (LREC2000) , pp. 947–952, 2000.

[12] Nanjo, H. and Kawahara, T., "Language model and speaking rate adaptation for spontaneous presentation speech recognition," IEEE Trans. Speech & Audio Process., 12(4), pp. 391-400, 2004.

[13] T. Akiba et al., Journal of Information Society of Japan, Vol.50, No.2, pp.501-513, 2009.2. T. Akiba, K. Aikawa, Y. Itoh, T. Kawahara, H. Nanjyo, H. Nishizaki, N. Yasuda, Y. Yamashita and K. Itoh, "Construction of a Test Collection for Spoken Document Retrieval from Lecture Audio Data," Journal of Information Processing Society of Japan, Vol.50, No.2, pp.501-513, 2009.

[14] A. Singhal et al., "Pivoted Document Length Normalization", Proc. of SIGIR, pp.21-29, 1996.

[15] T. Akiba, H. Nishizaki, K. Aikawa, X. Hu, Y. Itoh, T. Kawahara, S. Nakagawa, H. Nanjo, Y. Yamashita, "Overview of the NTCIR-10 SpokenDoc-2 Task," In Proceedings of the 10th NTCIR Workshop Meeting, 2013.