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

Noboru Kanedera (Ishikawa National College of Technology)

INTRODUCTION

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 and the contents complement for spoken document retrieval for SCR lecture retrieval task and SCR passage retrieval task.

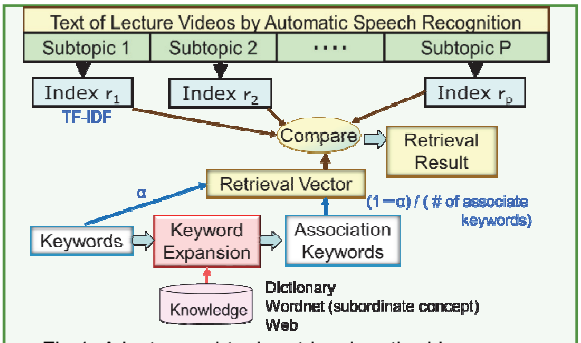

Fig.1 A lecture subtopic retrieval method by keyword expansion.

The specified keywords by a user are expanded to the association key words such as broader concepts and subordinate concepts using Japanese Wordnet. Next, the retrieval vector is generated from keywords and association keywords. The retrieval is done by comparing the retrieval vector with the TF-IDF vector of each subtopic.

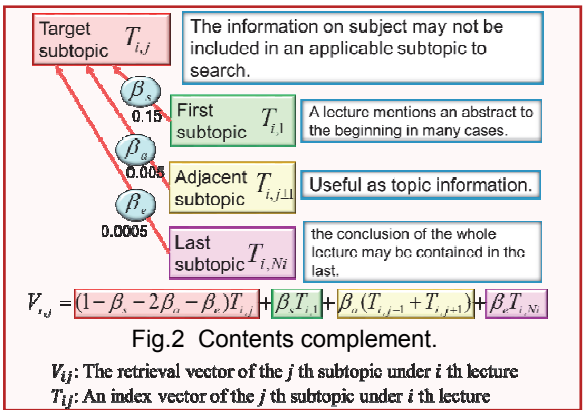

Fig.2 Contents complement.

V_{ij}: The retrieval vector of the j th subtopic under i th lecture
T_{ij}: An index vector of the j th subtopic under i th lecture

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

Table 2. Evaluation results for the lecture retrieval task.

Task	Run	MAP	Method
lecture	1	0.324	keyword expansion (wordnet+NDK)
	2	0.320	keyword expansion (wordnet+NDK) +subword retrieval
	3	0.320	baseline

When the keyword expansion using both NDK basic dictionary 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.

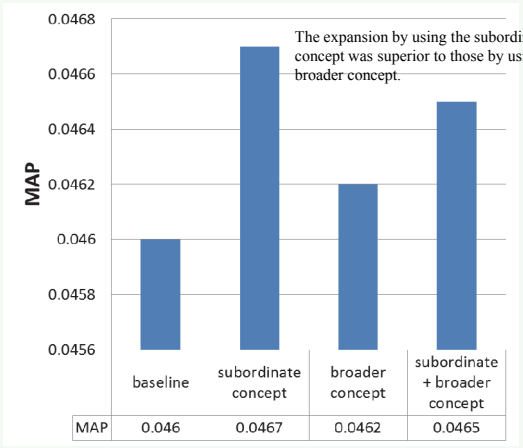

Fig. 3 NTCIR-10 passage retrieval results with keyword expansion by Wordnet to ASR text.

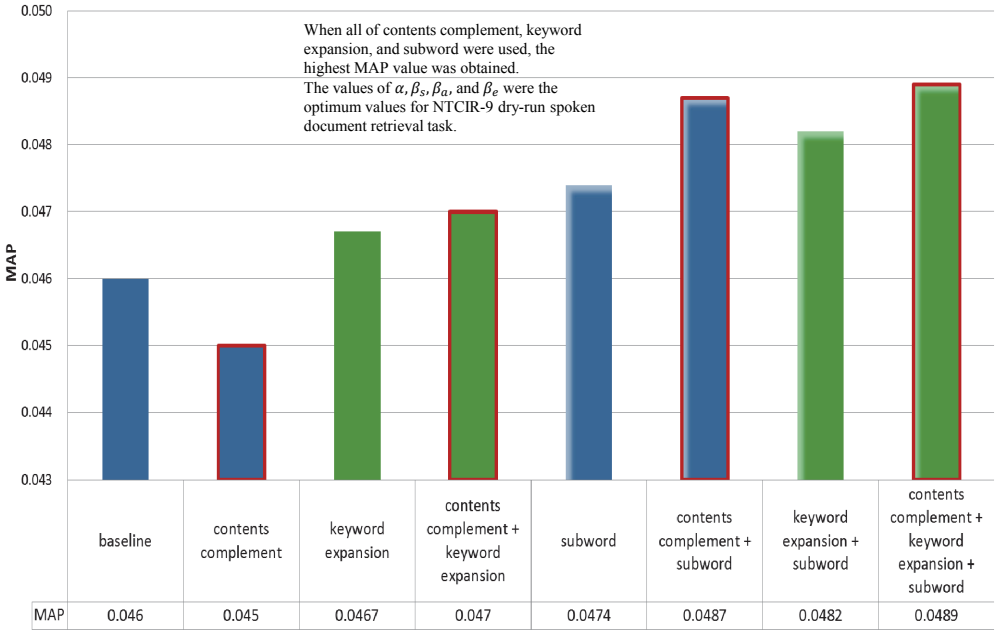

Fig. 4 NTCIR-10 passage retrieval results with contents complement and keyword expansion.

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.