論文 / 著書情報
Article / Book Information

論題(和文)	
Title(English)	Development of a Dialogue System for Web Retrieval
著者(和文)	パスカル アモニック, 岩野 公司, 中川 竜太, 古井 貞熙
Authors(English)	Pascal Hamonic, Koji Iwano, Ryuta Nakagawa, Sadaoki Furui
出典(和文)	日本音響学会2006年春季講演論文集, Vol. , No. , pp. 15-16
Citation(English)	, Vol. , No. , pp. 15-16
発行日 / Pub. date	2006, 3

Development of a dialogue system for Web retrieval *

◎Pascal Hamonic, Koji Iwano, Ryuta Nakagawa, Sadaoki Furui (Tokyo Institute of Technology)

1 Introduction

Recently automatic speech recognition has become a practical technology, and is now used in real-world applications, such as information retrieval. Speech-driven Web retrieval, in which spoken queries are used to retrieve Web pages, has already been explored in the information retrieval community. However, since the retrieval accuracies of such systems are degraded due to recognition errors of input speech, it is crucially important how to reduce such effects. Fujii et al. [1] uses recognized syllable sequences for automatically detected out-of-vocabulary (OOV) word periods, in order to recover the words which are not included in the dictionary. Matsushita et al. [2] uses combination of multiple LVCSR models for increasing recognition performance. In this paper, we investigate a dialogue-based Web retrieval system in which user can select correctly recognized keywords through interactions with the system [3]. Our system can avoid troublesome re-speaking of misrecognized keywords by suggesting possible keywords obtained by an automatic keyword recovering technique. System performance is evaluated by using Japanese queries provided by the speech-driven retrieval subtask in the NTCIR-3 Web task [4, 5].

2 Proposed dialogue-based system

As shown in Fig. 1, an input utterance is transcribed by LVCSR using language and acoustic models provided by the NTCIR-3 Web task [4]. The language model is a word-based trigram model produced for 60,000 high frequency words contained in a 100GB Web collection. The acoustic model was produced using the ASJ speech database having 20,000 sentences uttered by 132 speakers with both genders. A 16-mixture Gaussian distribution triphone HMM was used. The recognizer was built by *Julius* v3.4.

All the nouns, except for words like "記述" or "文書" which are not important for search, are extracted from the transcription as content keywords and presented to the user. The user is requested to select correctly recognized keywords by entering their numbers by keyboard interface. For misrecognized keywords, the user is requested to re-speak the correct keywords. In order to reduce the number of troublesome repetitions, the system automatically suggests possible keywords obtained by a recovering technique described in the next section.

The query corrected in this way is sent to a retriever *akechi* [4] to search 3GB Web collection,

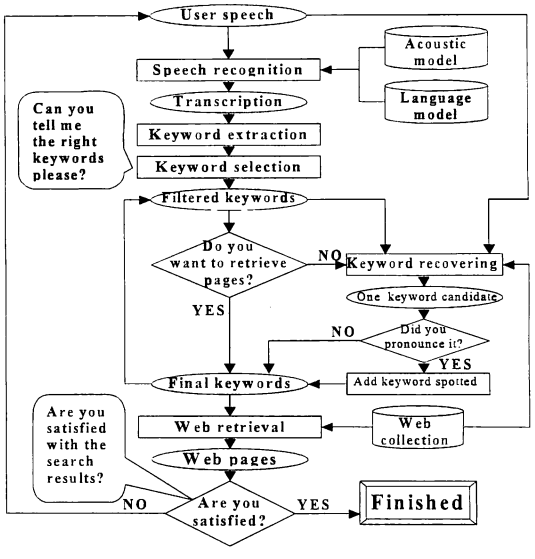

Fig. 1: Global overview of the proposed dialogue-based Web retrieval system.

and 10 documents are retrieved for display. Finally, the system asks the user whether he/she has found proper pages or not. The user answers by entering "y" (yes) or "n" (no). If satisfied (id. "y") the search task is completed, and if unsatisfied (id. "n") the can either be given up or continued by uttering a next query.

3 Keyword recovering

To recover the misrecognized query keywords, the system first constructs a set of candidate keywords. They are obtained from the results of query expansion, namely pseudo relevance feedback (PRF), conducted by a function implemented in the *akechi* retriever. Then the system checks whether each candidate already exists in the first input query or not by a keyword spotting technique.

For the keyword spotting, we first determine what keyword is most likely to have been included in the recognition results of the original utterance using the following grammar G:

$$G = \{all\ syllables\}\ KW\ \{all\ syllables\}$$

The category "*all syllables*" contains all possible Japanese syllables. The category "*KW*" corresponds to a set of 10 keyword candidates obtained by the query expansion. {} denotes zero or more entities.

Frame-averaged likelihood for the most likely keyword *KW0* is computed, and, in order to determine whether the *KW0* is adopted or rejected as a

* Web 検索のための音声対話システムの構築
アモニック パスカル，岩野公司，中川竜太，古井貞熙（東工大）

recovering keyword, the likelihood is compared to a threshold which is decided by preliminary experiments.

4 Experiments

4.1 Conditions

The 47 queries provided by NTCIR-3 were sorted according to their mean average precision (MAP) when inputting correct queries, and the top 15 were selected for evaluation. A system without having the keyword recovering function (system **A**) was prepared as a baseline to compare with the system with the recovering function (system **B**). Twelve subjects, each uttering 5 queries for each system, evaluated both systems. The queries were preliminarily selected randomly for each subject. The evaluation order of both systems was also randomly determined for each subject.

4.2 Evaluation

Mean keyword recall/precision was used for evaluating speech recognition performance, and mean MAP was used for evaluating retrieval performance. The number of speech inputs was also used for evaluation. For subjective evaluation, users were asked to answer questionnaires for comparing both systems. Scores were selected between 1 (bad) and 5 (good) in terms of the following criteria: system general speed, handling easiness, user irritation, and general evaluation.

4.3 Objective evaluation results

Before comparing the performances of the systems **A** and **B**, keyword recall/precision and MAP for the first query input was compared between these systems, and it was found that they were almost the same. This means that the experimental results evaluating the keyword recovering mechanism were meaningful.

Table 2 shows objective evaluation results when users completed all retrieval tasks. Although keyword precision rate was degraded, recall rate was improved by using the keyword recovering technique. Since keyword recall is more relevant to retrieval performance than precision [2], MAP was also increased by using the keyword recovering. The relative improvement of MAP by **B** from **A** was 9.8%. "correct" in the table shows the performance when correct keywords were input by the first query. It can be found that the MAP score of system **B** (0.380) is close to the ideal MAP value (0.390).

System **B** yielded relative reduction of 13% in the average number of speech inputs comparing to system **A** by using the keyword recovering.

4.4 Subjective evaluation results

Figure 2 shows subjective evaluation results. According to the hypothesis testing for the two-sample test at the significance level of 5%, meaningful improvement is observed only for the general evaluation criterion. This means that, in system **B**, although general velocity was decreased and required time to complete each task rose up, users were generally pleased by observing that correct keywords were successfully recovered and presented by the system. In terms of the handling easiness, users did not find any change when they switched from system **A** to **B**. The improvement in the user's irritation criterion was not as significant as we expected.

Table 2: Keyword Recall/Precision and MAP for system **A**/**B**.

System	Recall	Precision	MAP	#speech inputs
A	61.6	87.0	0.347	3.1
B	63.5	84.6	0.380	2.7
correct	100	100	0.390	1.0

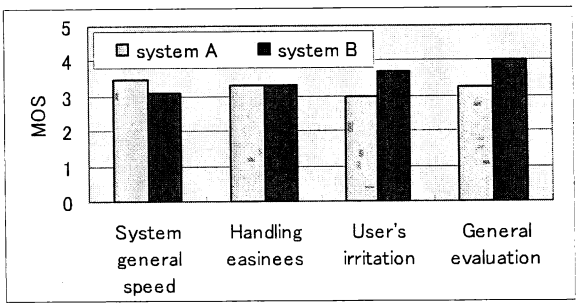

Fig. 2: Subjective evaluation results for system **A**/**B**.

5 Conclusions

This paper investigated a dialogue-based Web retrieval system with a keyword recovering function. Experimental results show that the function is effective for improving retrieval performance and subjective evaluation results.

Future research includes applying parallel computing techniques to the keyword spotting process for improving general speed of the system. We will also try to entirely remove keyboard interface by enabling users to interact with the system exclusively by voice.

Acknowledgments We would like to thank Prof. Atsushi Fujii of University of Tsukuba for providing us with the retriever *akechi*.

References

[1] A. Fujii, et al., "A method for open vocabulary speech-driven text retrieval," *Proc. Empirical Methods in Natural Language Processing*, pp. 188-195, 2002.

[2] M. Matsushita, et al., "Improvement of Keyword Recognition and Extraction for Speech-driven Web Retrieval Task", *Technical Report of IEICE*, pp.13-18, 2004. (Japanese)

[3] D. Kim, et al., "Language models and dialogue strategy for a voice QA system," *Proc. ICA*, vol.5, pp. 3705-3708, 2004.

[4] A. Fujii and K. Itou, "Building a Test Collection for Speech-Driven Web Retrieval," *Proc. Eurospeech*, vol.2, pp. 1153-1156, 2003.

[5] K. Eguchi, et al., "Overview of the Web Retrieval Task at the Third NTCIR Workshop," *Proc. Third NTCIR Workshop*, pp.1-26, 2002.