

Chinese Speech Information Retrieval for Questions on Mobile Phone Operation

Kai Ishikawa¹, Susumu Akamine¹, and Ken Hanazawa¹

¹ Media and Information Research Laboratories,
NEC Corporation
1753, Shimonumabe, Nakahara-ku, Kawasaki, Kanagawa 211-8666, Japan
{k-ishikawa@dq, s-akamine@ak, k-hanazawa@cq}.jp.nec.com

Abstract. We propose an information retrieval method based on question identification. We have implemented and integrated a prototype Chinese speech IR system for a commercial PDA with this method for questions concerning mobile phone operations. A preliminary evaluation showed that our method is effective. In the evaluation, our method outperformed a similar document search method by ten points or more in both its precision and recall for top-ranked candidates.

Keywords: Information Retrieval, Chinese Speech Recognition, Question Identification, Question and Answering, FAQ, User's Manual.

1 Introduction

As IT devices are becoming more complex and multifunctional, users are finding it hard to learn how to properly use all the available functions. Information retrieval is a useful technology for users to find the answers to their questions at anytime. Ikeda et al. applied a conventional similar document search function to a user's manual document (Ikeda et al. 2004). However, compared to the Web or other open domain retrieval tasks, a manual/FAQ search is a very domain-specific task, where the terms used in documents and queries are limited, and this causes high overlapping among the terms of different queries. A limited domain would adversely influence the similar document search function, because it evaluates the score of a document by counting the query terms in the document. To overcome this problem, Lytinen et al. proposed a FAQ retrieval method with question types, which are currently used by several question-answering systems competing in the TREC Q&A track (Abney et al. 2000, Cardie et al. 2000, and Harabagiu et al. 2000), and this method has shown improvements in the performance (Lytinen et al. 2002). This method can improve the precision of FAQ candidates by taking into consideration the similarities in a query's question type and a FAQ's question type. However, a question type is not a sufficiently detailed and reliable information to be used in the manual/FAQ retrieval system on a mobile device. Therefore, our idea is to solve the manual/FAQ retrieval problem by using an identification approach along with a conventional retrieval approach.

We propose in this paper a hybrid information retrieval method combined with a request identifier based on a syntactic parser. The parser uses "generic" syntactic rules, while requesting "specific" syntactic rules at the same time, and outputs the most plausible request ID according to the disambiguation of the syntactic structure. By introducing the identifier of a user's request within a particular domain, the method can provide highly precise answers. We collected 10,000 request sentences from a mobile phone operation domain, clustered the sentences into groups, and manually assigned a unique request ID to each set. An answer document (or a few answers) for each ID was selected from the corresponding user's manual documents. In our method, we first introduce a request identifier based on syntactic parsing to identify the user's information request, and then combine the identifier with a conventional similar document search function based on character uni-grams and bi-grams. We developed a prototype of a Chinese speech information retrieval system for a Chinese mobile phone user manual on a commercial PDA. We show the efficiency of our method through a preliminary evaluation.

2 Modeling of IR in a mobile phone operation domain

2.1 Problems when applying a conventional similar document search

The following problems will influence the application of only a conventional similar document search to user's manual documents.

(1) Diversity of phrasing to express the same request

A simple request can be expressed by various users in a wide variety of ways. Table 1 shows a variation of the request sentences for a Chinese mobile phone operation, "How to check the weather." With the application of only a similar document search method, this phrase variation will make it hard to adequately estimate the document scores.

(2) Overlapping of characteristic words in different types of requests

Many characteristic words in different types of request sentences will overlap because of the narrow domain. Figure 1 shows an overlapping of query words among different requests. When a similar document search is applied, the correct answer may be buried among other candidates in the search result, because documents with more characteristic words will be ranked higher.

Table 1. Wide variety of phrasing to express equivalent "How to check the weather" requests.

Request ID	Request sentence (indicates characteristic words)
37	/手机/如何/实现/天气/查询/的/功能/
37	/手机/能否/查询/天气预报/
37	/天气/查询/功能/该/如何/实现/
37	/天气预报/查询/步骤/
37	/天气预报/自动/查询/功能/
37	/天气预报/是/如何/查询/的/
37	/天气预报/的/查询/方法/是/什么/

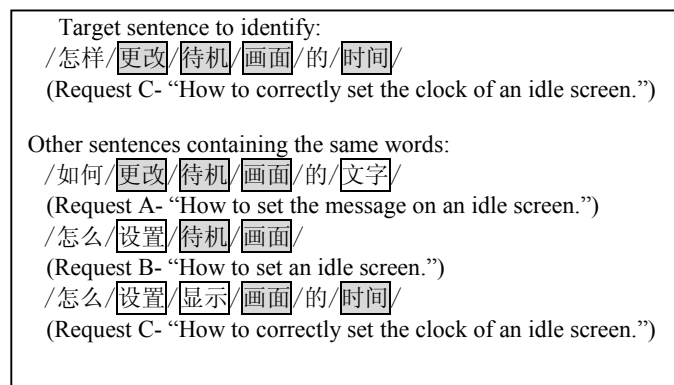


Fig. 1. Overlapping of query words among different request sentences.

2.2 Our approach to identifying user requests

We wanted a model that would allow us to identify a user's request with high accuracy based on a variety of query sentences concerning mobile phone operating procedures. Here, we explain the problem of identifying a request through the example request sentence for identification shown in Figure 1. This sentence corresponds to Request C, "How to correctly set the clock of an idle screen," and has four characteristic words (the characters enclosed in a box). We also have sentences corresponding to

Requests A and B among the various request sentences that share the same characteristic words as Request C. Note that the last request sentence in the example corresponds to Request C, although some of its characteristic words appear as synonyms in the former sentence for Request C. Because of the expression variation in the query sentences, even sentences having the same request ID do not always share characteristic words.

To deal with this type of problem, an identifier should be able to accept a variety of request sentences with a low identification error rate, regardless of how the sentences are expressed. To enable this, we have focused on an identification approach based on syntactic parsing. Figure 2 shows an outline of our approach. First, we apply “generic” syntactic rules, which are independent of the expressions in each request sentence, and obtain the syntactic structure for the requested sentence (the upper structure of Figure 2). Then, we focus on the VP phrase structure that contains the characteristic words of the sentence, and prepare the “specific” syntactic rules shown in Figure 3, which are the specific rules independently prepared for parsing characteristic expressions in each request sentence. By using the “specific” rules along with the “generic” rules in the syntactic parsing process, we obtain a new syntactic structure (at the bottom of Figure 2). In addition, from the initial symbol S_C of the structure, we obtain an identification result that shows that the example request sentence corresponds to Request C.

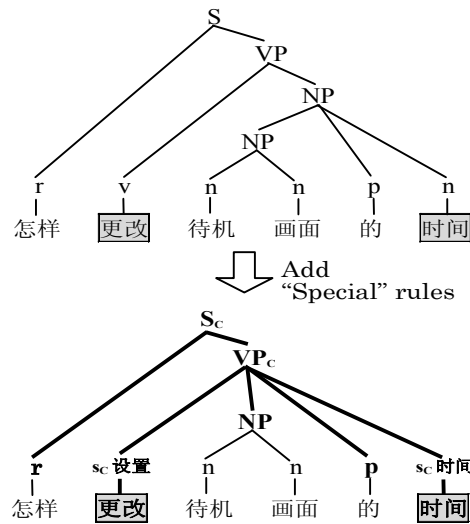


Fig. 2. Identification method based on syntactic structure.

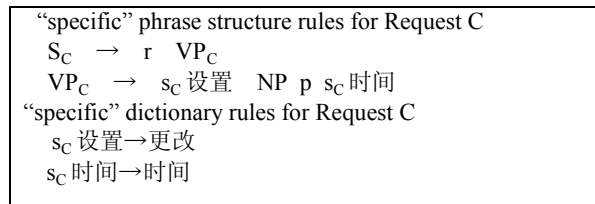


Fig. 3. Phrase structure and dictionary rules for Request C.

3 Our retrieval method

Our approach allows us to identify user requests with high accuracy, but lacks robustness when applied to a new request sentence. Therefore, we created a hybrid information retrieval method that combines our syntactic-parsing-based request identifier and a character-unigram/bi-gram-based similar document search method, which is robust enough for a wide variety of request sentences.

3.1 Request identification based on syntactic parsing

We created a hybrid information retrieval method that combines our syntactic-parsing-based request identifier and a character-sequence-based similar document search method, which is robust enough for a wide variety of request sentences. As explained regarding the request identification approach, we introduce “specific” rules for each request ID, which have a request ID number as a feature value for their non-terminal symbols, and apply these rules, along with the “generic” rules, to obtain the syntactic structure of an input request sentence. The dictionary rules and the phrase structure rules applied in our method are represented formally in Fig. 4. Here, V_T is a member of the terminal symbols referring to words, V_C is a member of the non-terminal symbols referring to a part-of-speech, V_X is a member of the non-terminal symbols referring to a phrase structure, V_{C_i} is a member of the non-terminal symbols referring to a “specific” synonym class belonging to a request ID = i ($i = 1, \dots, N$), where i is a member of the request IDs and V_{X_i} is a member of the non-terminal symbols referring to a “specific” phrase structure of the request ID = i . By applying these rules in the parsing process, we obtain a syntactic structure for the input request sentence. Here, the request ID of the input sentence can be known from feature i of the initial symbol σ ($\in V_X, V_{X_1}, \dots, V_{X_N}$) of the syntactic structure. In this paper, we use a probabilistic context-free grammar (PCFG) in the parsing process to obtain the most plausible syntactic structure. Here, for convenience, we use the same frequency for all the rules while obtaining a probabilistic model.

3.2 Introducing a similar document search method

We use a similar document search based on character unigrams and bi-grams. The information retrieval score ($Score(D, Q)$) of document D for request sentence Q is calculated by the following Okapi BM25 formula (Robertson et al. 1995).

$$Score(D, Q) = \sum_{T \in Q} w^{(1)} \cdot \frac{(k_1 + 1) \cdot tf}{K + tf} \cdot \frac{(k_3 + 1) \cdot qtf}{k_3 + qtf},$$

$$w^{(1)} = \log \frac{N - n + 0.5}{n + 0.5}, K = k_1 \cdot \left((1 - b) + b \cdot \frac{dl}{avdl} \right)$$

where, T refers to any character sequence in the request sentence Q , N refers to the total number of documents, n refers to the number of documents containing T , tf refers to the frequency of T in document D , qtf refers to the frequency of T in Q , dl refers to the total number of character sequences in D , and $avdl$ refers to the average number (for all the documents) of character sequences in a document. In addition, k_1 , k_3 , and b refer to constant values, which we set as $k_1=1$, $k_3=1000$, and $b=1$.

“generic” dictionary rules $A \rightarrow B \quad A \in V_C, \quad B \in V_T$ “generic” phrase structure rules $A \rightarrow \alpha \quad A \in V_X, \quad \alpha \in (V_C \cup V_X)^+$ “specific” dictionary rules for request ID= i $A \rightarrow B \quad A \in V_{C_i}, \quad B \in V_T$ “specific” phrase structure rules for request ID= i $A \rightarrow \alpha \quad A \in V_{X_i}, \quad \alpha \in (V_T \cup V_C \cup V_{C_i} \cup V_X \cup V_{X_i})^+$

Fig. 4. Rules introduced in our request identification approach.

Table 2. Example titles of user’s manual documents.

Document title	English translation
待机时间与通话时间	Battery time in standby mode and talk mode
检查电池	Checking battery level
插入和取出 miniSD 卡	Inserting and removing miniSD card
开机	Turning power on
接听一个来电	Receiving phone call

3.3 Fusion of retrieval results

The mapping table from request IDs to their answer documents is manually prepared in advance. The fusion process combines the results of the request identification and the similar document search and generates the ranking of documents as the final retrieval result. In this process, the answer document(s) for the identified request ID is placed at the top of the ranking, and then the documents obtained by the similar document search are placed in the order of their score, followed by the previously placed documents at the top. When the identification process generates no answer candidate, the final retrieval result consists of the documents from only the similar document search.

4 Prototype system

We implemented our proposed method and integrated a speech IR system on a commercial PDA (OS: PocketPC2003, CPU: Intel PXA255 400 MHz, memory: 64 MB) by combining it with a speech recognition module. As a retrieval target, we extracted 94 documents (Table 2. From the user’s manual of a Chinese mobile phone N840). In addition, we defined 55 independent requests regarding a generic mobile phone’s operations, and prepared a question-answer correspondence between the 55 request IDs and the 94 answer documents. Figure 5 shows an outline of the system configuration. The user inputs a request sentence in Chinese regarding a mobile phone operation. Then, the speech recognition module outputs the recognition results for the inputted speech in a word sequence. The request identification module identifies the request ID (55 requests in total) of the input request, and the similar document search module generates similar document candidates and their scores. Finally, the retrieval results fusion module generates a document ranking as the final result from the identification and similar document search results (Fig. 6).

The speech recognition module processes the user’s Chinese speech input using large vocabulary (30 K words) continuous speech recognition. The Chinese acoustic model developed for the Mandarin language is used (Hanazawa et al. 2006). For the language model, we use a word-based N-gram model developed from the 10,000 request sentences we collected. Out of the available sentences, we selected 4,846 sentences corresponding to the 55 request IDs as a training set used for writing the rules of the identifier. All the “generic” dictionary and phrase structure rules, and the “specific” dictionary and phrase structure rules were created manually based on the training set.

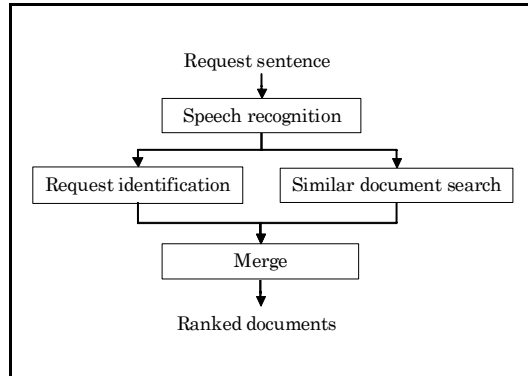


Fig. 5. System configuration.

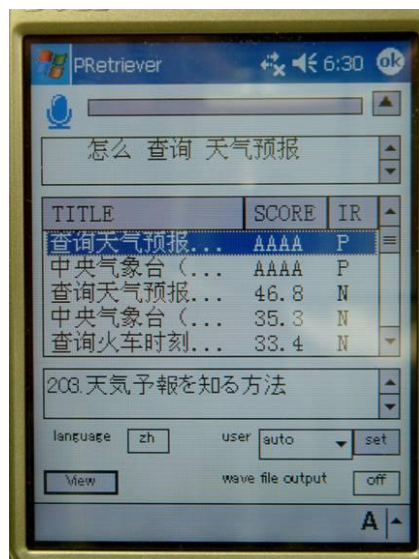


Fig. 6. Photograph of the prototype's screen.

5 Evaluation

5.1 Evaluation set

We created 101 request sentences for the evaluation set, corresponding to the 55 request IDs, which had no overlap with the training set for the rule writing. In addition, we prepared speech data for the evaluation set recorded by five native Chinese speakers in standard spoken language. The retrieval target was a set of 94 documents extracted from the text of a user's manual for a Chinese mobile phone.

Table 3. Request identification performance for transcript.

	Training set (4846)	Evaluation set (101)
Precision (%)	97.9	80.6
Recall (%)	60.6	35.6

5.2 Evaluation of request identification performance

The request identification performance shown in Table 3 is for the training set and the evaluation set. Here, “precision” is the ratio of the number of correctly identified sentences to the number of sentences output as identification results, and “recall” is the ratio of the number of sentences output as identification results to the total number of sentences in the data set. The results of an “evaluation set” are the evaluation results with the input of text transcribed from spoken requests.

5.3 Comparative evaluation

To evaluate the effectiveness of the proposed method, we compared its recall and precision performances with the transcription and speech inputs to the performance of a similar document search. The results from the similar document search and those of the proposed method are shown in Figures 7 and 8, respectively (The horizontal axis shows the number of top-ranked candidates evaluated in the retrieval results.). We also compared the performance of both methods when the speech input data was used (Figures 9 and 10, respectively). The results with the speech input data show the average values of speech data for the five speakers.

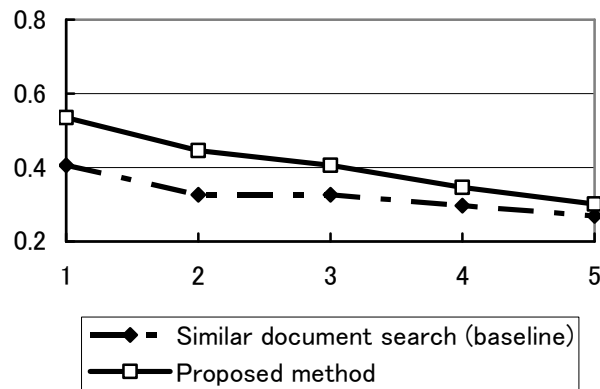


Fig. 7. Precision of baseline method and proposed method with text input.

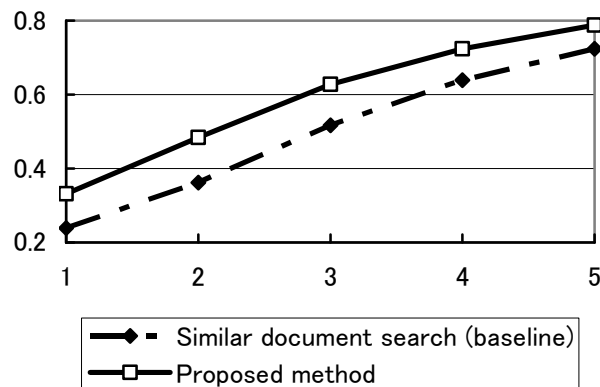


Fig. 8. Recall of baseline method and proposed method with text input.

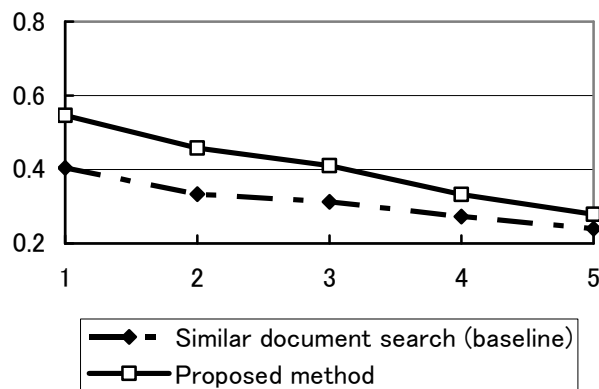


Fig. 9. Precision of baseline method and proposed method with speech input.

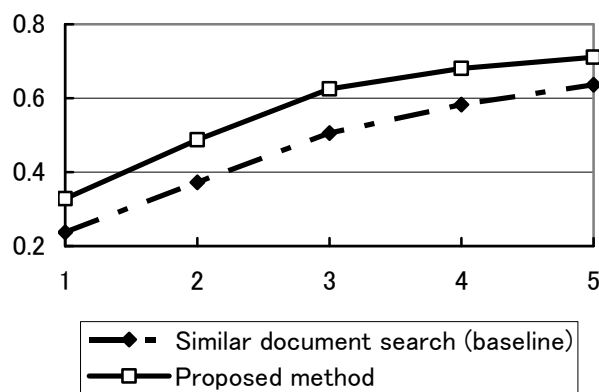


Fig. 10. Recall of baseline method and proposed method with speech input.

6 Discussion

The precision of the request identification with the proposed method, 80.6% in the evaluation set, was much higher than that with the baseline (similar document search) method for the top-ranked candidate. The performance of the proposed method with the speech input was almost the same as with the text input. However, the identification recall with transcription, 35.6% in the evaluation set, was low. This was due to the incomplete rule set, which led to a low recall value of 60.6% even in the training set. With regard to the information retrieval performance, our proposed method outperformed the baseline method by ten points or more in both its precision and recall for top-ranked candidates. The effectiveness of the combination of the proposed identification method with its high precision and the similar document search method is demonstrated by the results. This effectiveness will become clearer when the recall of the identification method is improved through sufficient training of the rule sets.

7 Conclusion

We have proposed an information retrieval method that is highly accurate based on the user request identification. We implemented the proposed method and integrated a speech information retrieval system prototype in a Chinese mobile phone operation domain for a commercially available PDA. To test the effectiveness of our method, we carried out a preliminary evaluation using a Chinese mobile

phone user's manual. The evaluation showed that our method outperformed a similar document search method by ten points or more in both its precision and recall for top-ranked candidates; the high precision of our method was particularly encouraging. We expect further improvement in the next stage, when sufficient syntactic rules are created to fully cover the training data. We also plan to precisely analyze the relation between the speech recognition errors and the retrieval performance.

References

1. Ikeda, T. et al.: Speech-Activated Text Retrieval System for Cellular Phones with Web Browsing Capability. Proceedings of the 19th PACLIC (2005) 265-272.
2. Abney, S. et al.: *Answer extraction*. In *Proceedings of the 6th ANLP* (2000).
3. Cardie, C. et al.: *Examining the role of statistical and linguistic knowledge sources in a general-knowledge question-answering system*. In *Proceedings of the 6th ANLP* (2000).
4. Harabagiu, S. et al.: Falcon: Boosting knowledge for answer engines. In *Proceedings of TREC-9*.
5. Lytinen, S. L. et al.: *The Use of Question Types to Match Questions in FAQFinder*. In *Proceedings of the 2002 AAAI Spring Symposium on Mining Answers from Texts and Knowledge Bases*, p. 46-53.
6. Robertson, S. E. et al.: *Okapi at TREC-3*. Proceedings of TREC-3 (1995).
7. Hanazawa, K. et al.: Development of a Japanese to Chinese interpreter able to work on a mobile terminal. Proceedings of the 68th National Convention of the IPSJ (2006).