

An Interactive Framework for Document Retrieval and Presentation with Question-Answering Function in Restricted Domain

Teruhisa Misu and Tatsuya Kawahara

School of Informatics, Kyoto University
Kyoto 606-8501, Japan
misu@ar.media.kyoto-u.ac.jp

Abstract

We propose a speech-based interactive guidance system based on document retrieval and presentation. In conventional audio guidance systems, such as those deployed in museums, the information flow is one-way and the content is fixed. To make the guidance interactive, we prepare two modes, a user-initiative retrieval/QA mode (pull-mode) and a system-initiative recommendation mode (push-mode), and switch between them according to the user's state. In the user-initiative retrieval/QA mode, the user can ask questions about specific facts in the documents in addition to general queries. In the system-initiative recommendation mode, the system actively provides the information the user would be interested in. We implemented a navigation system containing Kyoto city information. The effectiveness of the proposed techniques was confirmed through a field trial by a number of real novice users.

1 Introduction

Most of the conventional information retrieval systems assume that a display is available as an output device, and thus a list of relevant documents can be presented. However, this is not always the case when only speech interface is available, for example, audio guidance systems. Considering user's easiness of comprehension, we need to limit the amount of content presented. However, simply summarizing the retrieved document may cause the loss of an important portion of information that the user wanted to know or may have been interested in. Actually, in the conventional audio guidance systems deployed in museums and at sightseeing spots, users cannot ask questions concerning missed portions. Therefore, we propose a more interactive scheme that incorporates a question-answering (QA) technique to follow up the initial query, enabling random access to any part of the document.

Now, there are some problems with QA in such situations. One important issue is contextual analysis. During a dialogue session, users tend to make questions that include anaphoric expressions. In these cases, it is impossible to extract

the correct answer using only the current question. For example, “When was it built?” makes no sense being used by itself. In many conventional database query tasks, this problem is solved by using the task domain knowledge, such as the semantic slots of the backend database [1, 2]. Whereas the majority of conventional QA tasks, such as TREC QA Track [3], have dealt with independent questions that have respective answers for each, there have been only a few works that have addressed successive questions [4]. However, they have basically hand-crafted questions rather than collecting real dialogues. In this work, we address the QA task in a real interactive guidance system using a topic tracking mechanism.

In addition, we introduce our system-initiative information recommendation function. In spoken dialogue systems, users often have difficulty making queries because they are unsure of what information the system possesses. Moreover, the system-initiative guidance is also useful in navigating users in the tasks without definite goals, such as sightseeing guidance. In order to make the guidance interactive, we propose the application of a QA technique to generate system-initiative recommendations.

Based on the above concepts, we have designed and implemented an interactive guidance system of “Dialogue Navigator for Kyoto City”, and conducted a field trial for about three months. The key evaluation results of the QA function are presented in this paper.

2 System Framework

The proposed guidance system has two modes, a user-initiative retrieval/QA mode (pull-mode) and a system-initiative recommendation mode (push-mode), and switches between them according to the user’s state. When a query or a question is posed by a user, the system switches to the retrieval/QA mode and generates a respective response. When the system detects the silence of the user, it switches to the system-initiative recommendation mode and presents information related to the current topic. The flow of this process is shown in Figure 1.

We adopted a sightseeing guidance of Kyoto city for our target domain. The knowledge bases (KB) of this domain are Wikipedia¹ documents concerning Kyoto and the official tourist information of Kyoto city. Table 1 lists the size of these KBs.

3 User-initiative Information Retrieval and Question-Answering

The user utterances are classified into two categories. One is an information query, such as “Please explain about the Golden Pavilion”. For such queries, the system retrieves the appropriate information from the KB by section unit, and

¹ <http://wikipedia.org/>

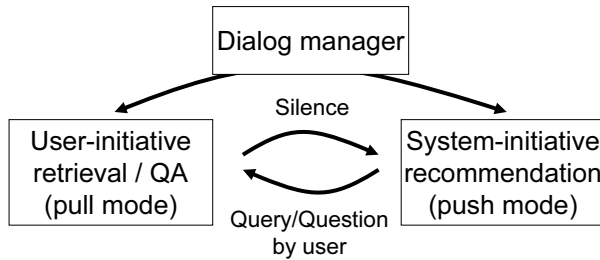


Fig. 1. System overview

Table 1. Knowledge base (KB) specifications

	# documents	# sections	# words
Wikipedia	269	678	150K
Tourist information	541	541	70K
Total	810	1,219	220K

the section of the document with the highest matching score is presented to the user. The other is a question, such as “When was it built?”. The system extracts the sentence from the KB that includes the answer to the question and presents it to the user. This procedure is shown in Figure 2.

3.1 Contextual Analysis based on Topic Detection

In dialogue systems, the incorporation of contextual information is important for generating meaningful queries for retrieval. As the deterministic anaphora resolution [5] is not easy and always error-prone, and stochastic matching is used in information retrieval, we adopt a strategy that concatenates the contextual information or keywords from the user’s previous utterances to generate a query. The simplest way is to use all the utterances from the current user. However, this might also add inappropriate context because the topic might have been changed in the session. Therefore, we determine the context length (number of previous utterances) used for the retrieval by tracking the topic of the dialogue.

We use metadata of the KB or a title of the document as a topic. Thus, the topic can be tracked using the current focused documents, which usually correspond to sightseeing spots or Wikipedia entries.

3.2 Document Retrieval

We adopt an orthodox vector space model to calculate the matching score (degree of similarity) between a user query and the document in the KB. That is, the vector of the document is made based on the occurrence counts of nouns in the

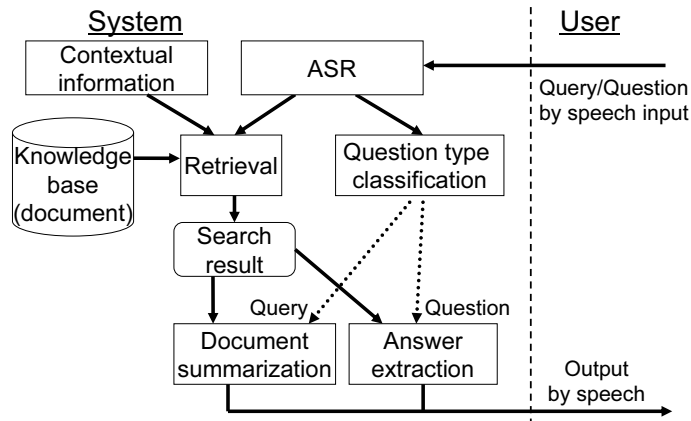


Fig. 2. Overview of document retrieval and QA

document by section unit. The vector for the user query is also made by merging the N-best hypotheses of the automatic speech recognition (ASR) result for the current utterance and previous utterances concerning the current topic as the context. We also use the ASR confidence measure (CM) as a weight for the occurrence count. The matching score is calculated by the product of these two vectors.

To deliver a concise presentation of the retrieved document, a summary is generated by extracting important sentences.

3.3 Answer Extraction

We have implemented the system with a general answer extraction module. A score is calculated using the following features for each named entity (NE) in the retrieved document that matches the question type (who, when, ...).

- Degree of similarity between the user utterance and the document (3.2).
- Number of matched content words in the sentence that includes the NE.
- Number of matched content words included in the clause that depend on/depended by the clause that includes the NE.

The system then selects the NE with the highest score as an answer to the question.

4 System-initiative Recommendation

For interactive information recommendation, we propose the generation of system-initiative questions. They are semi-automatically made from the current document using the QA technique. This is complemented by conventional information recommendation techniques based on the document structure and similarity.

4.1 Generation of System-initiative Questions (Method 1)

This method is intended to successively present more details on the target topic, after the initial summary presentation, because a user may be interested in some information that was not included in the summary. Although it is possible to prompt a user, such as "Would you like more details?", we propose a more interactive method by generating system-initiative questions in order to attract the interest of the user.

A set of possible questions is prepared using the following procedure. It is almost the reverse of the process to find an answer to the user's question.

1. Pick up the NE that may attract the user's interest based on $tf*idf$ criterion.
2. Substitute the NE with the corresponding interrogative.
3. Delete the subordinate clause using a syntactic parser.
4. Transform the sentence into an interrogative form.

Figure 3 shows an example of the transformation of a sentence in the KB into a question using the above-mentioned procedure.

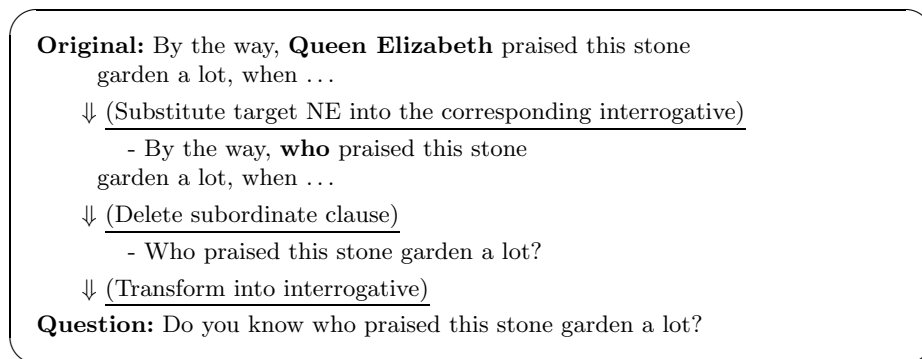


Fig. 3. Example of system-initiative question generation

4.2 Recommendation based on Document Structure and Similarity

We have also implemented following two conventional recommendation techniques based on the document structure and similarity.

– **Recommendation based on document structure (Method 2)**

Wikipedia documents are hierarchically described using the section structure. Thus, another section of the current document can be picked for presentation.

U1: Please explain about the Golden Pavilion.
S1: The Golden Pavilion is one of the buildings at the Rokuon-ji in Kyoto, and is the main attraction of the temple sites. The entire pavilion except the basement is covered with pure gold leaf.
U2: When was it built?
S2: It was originally built in 1397 to serve as a retirement villa for Shogun Ashikaga Yoshimitsu.
(Silence)
S3: Well then, do you know what was awarded to this temple in 1994?
U3: No, please tell me.
S4: It was awarded as listing on the UNESCO World Heritage in 1994.
U4: How can I get there?
.....

Fig. 4. Example dialogue 1

– **Recommendation based on document similarity (Method 3)**

We can select a document that has a large similarity with the current document. This technique is often adopted in the information recommendation of Web sites.

5 System Evaluation

We implemented the guidance system “Dialog Navigator for Kyoto City”. An example dialogue of the system using the QA technique is shown in Figs. 4 and 5. We carried out a field trial at our university museum. Users range in age from children to seniors and apparently have little experience in using spoken dialogue systems. No instructions on the system were given. We collected a total of 2,500 dialogue sessions (20,000 utterances).

5.1 Analysis of User Utterances

We have transcribed 465 dialogue sessions chosen from a particular time period. There were in-domain sessions in 52.7% of these sessions, and 23.4% of these sessions included users’ questions. For the ASR system, a trigram language model was trained using the KB, a dialogue corpus of a different domain, and Web texts that were automatically selected considering topics and styles [6]. The average word accuracy was 70.6%.

We then evaluated the content of the user utterances. There were 2,930 user utterances in the above dialogue sessions. Table 2 lists these their classification. In this table, “Query” denotes the search queries about sightseeing of Kyoto

U1: Tell me about Ryoan-ji.
 S1: Ryoan-ji is a Zen temple located in northwest Kyoto. Belonging to the Myoshinji school of the Rinzai branch of Zen Buddhism, the temple is a UNESCO World Heritage Site.
 (Silence)
 S2: Do you want to know about the stone garden?
 U2: Yes, please.
 S3: This garden is a famous karesansui rock garden. It consists of raked gravel and fifteen moss-covered boulders, which are placed so that, when looking at the garden from any angle, only fourteen of the boulders are visible at one time.
 (Silence)
 S4: Well then, do you know who praised this stone garden a lot?
 U3: No, I don't.
 S5: Queen Elizabeth praised this stone garden a lot on her visit here.
 U4: Who made this temple?

Fig. 5. Example dialogue 2

(excludes questions). “Question” denotes specific questions about Kyoto, such as “When was the Golden Pavilion built?”. “Acceptance/Rejection” is the user’s reply for confirmation or recommendation by the system. “Greeting” denotes a greeting to the agent, such as “hello”. “Out-of-domain” does not belong to any of the above categories. Most of them are non-sensical phrases. “Non-speech” denotes a false detection caused by background noise, which is supposed to be rejected.

5.2 Evaluation of Question-Answering Performance

First, we evaluated the performance of QA in terms of the success rate using 510 questions. Although the QA performance is usually evaluated using the mean reciprocal rank (MRR), we adopted the simple success rate, because it is not possible to audibly present alternative candidates. We regarded QA as successful when the system made an appropriate response to the question. That is, if an answer to the question exists in the KB, we regarded QA as successful when the system presented the answer. On the other hand, if there is no answer in the KB, we regarded QA as successful when the system said so. The QA success rate was 61.4% (64.0% for correct answers existed in the KB, and 46.8% for when they did not).

We then evaluated the effect of the context length (= number of previous utterances) used for the retrieval. This result is shown in Table 3. Without context, the success rate is significantly degraded. However, using all the previous utterances has an adverse effect. The incorporation of appropriate context information by topic tracking effectively improved the performance.

Table 2. Classification of user utterances

Content	Percentage (%)
Query	17.5
Question	8.0
Acceptance/Rejection	23.0
Greeting	23.3
Out-of-domain	15.2
Non-speech	13.1

Table 3. Contextual effect for QA

Use of context	Success rates (%) (correct answer exists, does not exist)
Current topic (proposed)	61.4 (64.0, 46.8)
No context	39.2 (33.3, 72.7)
Previous single utterance	55.9 (56.1, 54.5)
Previous three utterances	57.5 (58.7, 50.6)
All utterances	50.0 (50.3, 48.1)

5.3 Evaluation of System-initiative Recommendation

In order to confirm the effect of the proposed system-initiative question, the system was set to randomly make possible recommendations. The number of recommendations presented by the system during the 427 (hand-annotated subset of the 465 sessions) dialogue sessions was 319 in total. We regarded a recommendation as accepted when the user positively responded² to the proposal given by the system. The acceptance rate of each presentation technique is shown in Table 4. The acceptance rate by the system-initiative question (method 1) was much higher than that of other methods. This result suggests that the recommendations using the question form are more interactive and attractive.

Finally, we evaluated the number of mode shifts between the user-initiative retrieval/QA mode (pull-mode) and the system-initiative recommendation mode (push-mode) (Fig. 1). Among the 427 dialogue sessions, there were 420 mode shifts from the pull-mode to the push-mode, in 185 dialogues, and 297 from the push-mode to pull-mode, in 122 dialogues. This result suggests that many of the users' interests in the target topic were initiated by the recommendations, and continued the dialogue.

6 Conclusion

We have proposed an interactive scheme for information guidance using question-answering techniques. Specifically, we incorporated question-answering techniques

² by human judgment

Table 4. Comparison of recommendation method

Recommendation method	Acceptance rate (%)
Question (Proposed method 1)	74.7
Document structure (Method 2)	51.1
Document similarity (Method 3)	30.8

into both the user-initiative information retrieval and system-initiative information presentation. We have implemented a sightseeing guidance system and evaluated it with respect to the QA-related techniques. We confirmed that the proposed techniques worked well in improving the system performance.

References

1. Bohus, D., Rudnicky, A.I.: RavenClaw: Dialog management using hierarchical task decomposition and an expectation agenda. In: Proc. Eurospeech. (2003)
2. Komatani, K., Kanda, N., Ogata, T., Okuno, H.G.: Contextual constraints based on dialogue models in database search task for spoken dialogue systems. In: Proc. Interspeech. (2005)
3. NIST, DARPA: The twelfth Text REtrieval Conference (TREC 2003). In: NIST Special Publication SP 500–255. (2003)
4. Kato, T., Fukumoto, J., F. Masui, a.N.K.: Are open-domain question answering technologies useful for information access dialogues? – an empirical study and a proposal of a novel challenge. *ACM Trans. of Asian Language Information Processing* **4**(3) (2005) 243 – 262
5. Matsuda, M., Fukumoto, J.: Answering question of iad task using reference resolution of follow-up questions. Proc. the Fifth NTCIR Workshop Meeting on Evaluation of Information Access Technologies (2006) 414–421
6. Misu, T., Kawahara, T.: A bootstrapping approach for developing language model of new spoken dialogue systems by selecting Web texts. In: Proc. Interspeech. (2006) 9–12