Automatic generation of tourism quiz using blogs

Zeng, Jun
Graduate School of Information Science and Electrical Engineering, Kyushu University

Sakai, Toshihiko
Graduate School of Information Science and Electrical Engineering, Kyushu University

Yin, Chengjiu
Research Institute for Information Technology, Kyushu University

Suzuki, Takahiko
Research Institute for Information Technology, Kyushu University

他

出版情報：Artificial Life and Robotics. 17 (3-4), pp.412-416, 2013-01-01
バージョン：
権利関係：

Automatic Generation of Tourism Quiz using Blogs

Jun Zeng[1], Toshihiko Sakai[1], Chengjiu Yin[2], Takahiko Suzuki[2], Sachio Hirokawa[2]

[1] Graduate School of Information Science, Kyushu University, Fukuoka 812-8581, Japan
[2]Research Institute for Information Technology, Kyushu University, Fukuoka 812-8581, Japan
(Tel: 81-92-642-2301, Fax: 81-92-642-2294)

[1](zeng.j.000, 2IE11060Y)@ s.kyushu-u.ac.jp
[2](yin, suzuki, hirokawa)@ cc.kyushu-u.ac.jp

Abstract: The one-way information provision can not impress the listeners well. As an efficient way, Question & Answer can help information provider understand the response of listeners. However, it is not easy for everyone to set a suitable question. In this paper, we propose an automatic quiz generation system using tourism blogs. The system can generate the quiz by extracting feature words of the topic keyword from the blogs. Our purpose is to help the tourism information providers to advertise their tourism events in an interactive way, in order to impress the tourists. By comparing with other method of quiz generation, we demonstrate that our method is more suitable for information provision.

Keywords: Quiz Generation, Tourism Blog, Feature Word, Information Provision.

1 INTRODUCTION

The traditional method of information provision is a one-way method from information providers to information recipients. There is no interaction between providers and recipients, so it is an inefficient method to provide information. You can image that when you are attending a philosophy class and listening to the virtually unintelligible theories, you may feel nothing but lethargic. But when the teacher says, “Who can answer this question?”, you may feel fresh instantly. Because you may want to know what the question is, and whether you will be asked to answer the question. If you are the one to answer the question unfortunately, you will be impressed by the question. Therefore, as an efficient way, Question & Answer (Q&A) can help information to provider understand the response of listeners. That is because Q&A changes the one-way into two-ways method. On the one hand, listeners not only just listen but also comprehend the content of information. On the other hand, the provider can also know how much the listeners have understood. As a typical form of Q&A, multiple choice questions require a short response time and with the possibility of covering a broad set of topics.

However, in order to set a question, user must quite understand the knowledge of the topic. Meanwhile, user has to read up a lot of literature to ensure the correction of answers. For helping user to set a multiple choice question easily, we propose a novel method to generate multiple choice questions automatically. In this paper, we mainly consider the automatic generation of quiz related to tourism event. Our purpose is to help the tourism information providers such as travel sites, tour guides and some other tour organizations to advertise their tourism events in an interactive way, in order to impress the tourists. The authors of this paper have been doing many researches on data mining of tourism data [1] [2]. In this paper we adopt tourism blogs as the experiment data. That is because the role of blogs became increasingly main stream in recent years. We collect the official blogs of Kyushu Tourism Promotion Organization[1] to generate quiz. Also, we develop a Automatic Quiz Generation System (AQGS) for realize the automatic generation of tourism quiz by using blogs.

[1]http://www.welcomekyushu.jp/

2 RELATED WORKS

There is a large body of related work on automatic generation of multiple choice questions.

Tonoike et al [3] proposed an approach of answer validation based on the strengths of lexical association between the keywords extracted from a question sentence and each answer candidate. Tsumori and Kaijiri [4] generated the questions adapted to students' understanding based on the similarity of each word. Mitkov et al [5] distinguished the kind of problem sentences, and analyzed the structures (SVO, SVC, etc.) of the sentences. Moreover, the words that have the similar meaning with the correct answer in Wordnet are chosen as the distracters. Hoshino and Nakakawa [6] made a machine learn based on Naive Bayes and K-Nearest Neibors by using a collection of TOEIC question as training data. The blank in question was determined from the training result. The answers were random and chose from the same sentence.

However, all of these researches paid more attention to the difficulty of the questions. That is because these questions were generated for aims of education. In this

paper, we focus on the interest of questions instead. Our purpose is make the listeners remember the information easily, not try to test the level of intelligence of listeners.

3 AUTOMATIC GENERATION OF QUIZ

3.1 Creation of Index

Before developing quiz generation system, we need to create an index. Index is design to optimize speed and performance in finding relevant documents for a search query. In this paper, we will not only build an index of full-text but also an index of each sentence.

For building a sentence index, we need to create a sentence-based frequency file. First of all, we divide every article into sentences. Next, we use "ChaSen" [7], which is a morphological parser for the Japanese language, to analyze morpheme of every word. After creating the frequency file, we use "GETA" [8], which is a generic engine for transposable association, to build an index of each sentence.

3.2 Extraction of Feature words

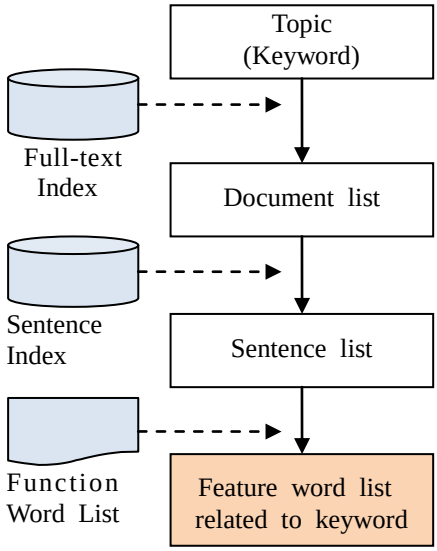


Fig. 1. Extraction of feature words

Typically, a multiple choice quiz contains two parts: a question and a group of suggested answers. In this paper, we generate the answers from a feature word list related to the user's keyword. A feature word is a word that appears in the same sentences with the user's keyword. In this section, we will introduce the extraction of feature words.

Fig. 1 shows the outline of the process for extracting the feature words. Firstly, the user must enter a topic (a keyword) of quiz. When user enters a keyword, the system searches all of the documents by using the full-text index, and generates a list of documents in which the keyword appears. Then, the system finds out all the sentences containing the keyword from the document list by using the sentence index. Finally, system extracts feature word list from the sentence list.

Table 1. Function words in Japanese

Category	Description and Examples
Pronoun	A pronoun is a pro-form that substitutes for a noun (or noun phrase) e.g. 私(I), 彼(he),これ(This), ここ(here)
Conjunction	A conjunction is a word that connects words, phrases, clauses, and sentences. e.g. しかし(but), また(or), そして(and)
Joshi	In Japanese grammar, particles are called joshi. e.g. に(in), で(at), は, を
Auxiliary verb	An auxiliary verb is a verb that gives further semantic or syntactic information about a main or full verb. e.g. できる(can), ない(not)

However, there are many function words in the feature word list. Function words express grammatical relationships with other words within a sentence. They have little lexical meaning or have ambiguous meaning. It is necessary to get rid of the function words from feature word list. Table 1 shows the category of the function words in Japanese. Fortunately, there are not many function words in Japanese, so we can manually create a function word file in advance. By using this file, we get rid of all the function words from feature word list effectively.

3.3 Generation of answers

The answers of a multiple choice quiz always contain a correct answer and a group of distracters. We consider the user's keyword as the topic of the quiz, so we can choose most relevant feature words as the answers. To some extent these feature words can describe the keyword.

In order to find out the most relevant feature words, we need to rank the feature words. In this paper, we use SMART (System for the Mechanical Analysis and Retrieval of Text) [9] to calculate the similarity between feature word and keyword. In other words, every feature word w_i will have $SMART(k, w_i)$, where $SMART(k, w_i)$ is the similarity between keyword k and w_i. If word w_i have a higher $SMART(k, w_i)$, it means that w_i is more similar with k.

Fortunately, GETA supports the computation of SMART, so we can get *SMART(k, w_i)* of all the feature words as soon as we get the feature word list. According to *SMART(k, w_i)*, we sort the feature word list. If a word get a higher *SMART(k, w_i)*, it will appear at a higher rank. Finally, we choose the top 1 feature word as the correct answer, top 2 ~ 5 feature words as the distracters.

3.4 Generation of questions

As we mentioned above, before we extract the feature words, we have get all the sentences that contain the keyword. We will generate questions from the sentence list.

Table 2. An example of quiz

Question	(　)は、これを防ぐために、若者達が喊声をあげて、観音堂内の鬼の面箱を封じ直すという勇壮な裸絵巻。 現在では、「童子の舞」が(　)の中心となっていて、童子の舞や所作は一見の価値があります。
Answers	行事、イベント、公園、祭り、神社

Firstly, in order to avoid that the distracters appear in the questions, we need to get rid of the sentences that contain the distracters. And then we randomly choose a sentence from the sentence list as the question. Meanwhile, the correct answer will be replaced by brackets. Table 2 shows an example of quiz, where the red answer is the correct answer.

4 AUTOMATIC QUIZ GENERATION SYSTEM

4.1 Design goal

As we mentioned above, the goal of Automatic Quiz Generation System (AQGS) is to help the tourism information providers such as travel sites, tour guides and some other tour organizations to generate quiz easily. These quizzes can help them to advertise their tourism events in an interactive way, in order to impress and attract the tourists.

4.2 Outline of AQGS

We develop the AQGS based on the proposed method with PERL language. User can open the CGI file of AQGS by using any kind of browser. Fig. 2 shows the interface of AQGS. There is only a "topic textbox" and a "search button". When user enters a topic keyword and clicks the search button, the quiz will be displayed below.

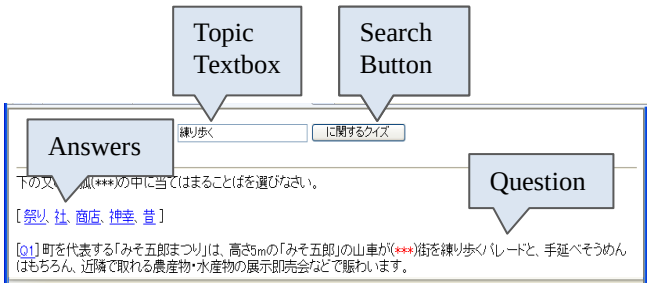


Fig. 2. The interface of AQGS

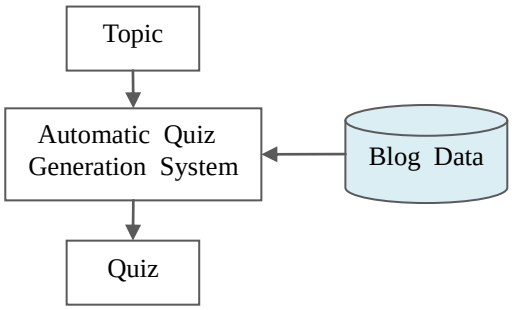


Fig. 3. The outline of AQGS

Fig. 3 shows the outline of AQGS. User only needs to enter a topic keyword. The system will automatically generate quizzes about the topic. It is not necessary for user to understand the topic well.

5 EXPERIMENT AND EVALUATION

5.1 Experiment Data

We collected the official blogs of Kyushu Tourism Promotion Organization to generate quiz. We focused on the events in the region, so we collected 906 events, at the first stage. However, some articles did not contain the date of event. So, we chose 316 articles that contain the date of the events. Some events are performed every year. So, we identified the articles of the same event in different years. As the result, we obtained 312 articles by which we generate quiz.

5.2 Methodology and procedure

To evaluate the usability of AQGS, we introduce a usability experiment. We developed two quiz generation system, one is proposed system by using feature word (F-System), and the other system generates the quiz by using Wordnet [5] (W-System). 5 participants joined in the experiment. Before the experiment, we did not introduce participants the purpose of this experiment. We design the task of usability experiment as follows:

(1) We generate the quizzes by F-System or W-System randomly. And the topic of quiz will also be determined randomly.

(2) We ask participants to answer these quizzes. They do not know whether the quizzes are generated by F-System or W-System.

(3) Participants evaluate each quiz they answered by 4 criterions, every criterion has 4 degrees. Table 3 shows the criterions and their degrees.

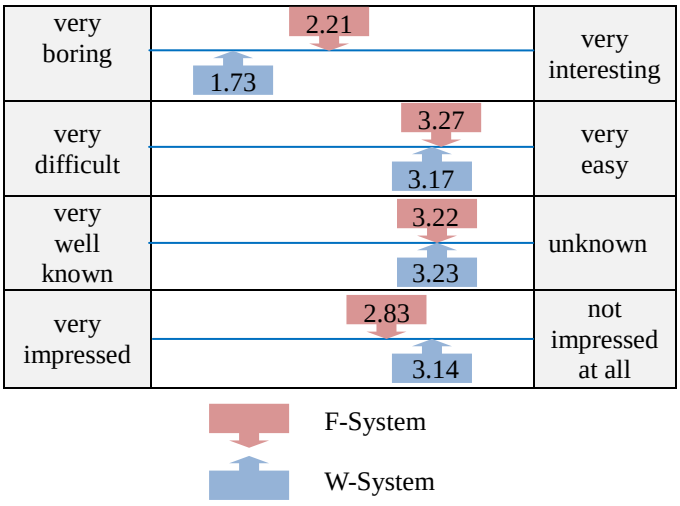
Table 3. Criterions of evaluation

Criterion	Degree
Interest	1 very boring, 2 boring, 3 Interesting, 4 very interesting
Difficulty	1 very difficult, 2 difficult, 3 easy, 4 very easy
Acquaintance	1 very well known, 2 well known, 3 known, 4 unknown
Impression	1 very impressed, 2 impressed, 3 unimpressed, 4 not impressed at all

5.3 Results and Analysis

Finally, we get 200 evaluation data: F-System 100 and W-System 100. We calculate the average degree of each criterion. Table 4 shows the result of evaluation, where the numbers in the arrow present the average degree of each criterion.

Table 4. Result of evaluation



Although “Difficulty” and “Acquaintance” are not the most important criterions, the result shows that quizzes generated by two systems have the same difficulty and acquaintance. It proves that the quizzes are comparable.

The result also shows that the quizzes generated by F-System are considered more interesting than those of W-System. Because the distracters generated by W-System are chosen from similar words of Wordnet, such as ship, boat, vessel and watercraft. Participants consider that it is boring to tell the difference from the words. After all, the quizzes are not generated for intelligence test. Meanwhile, Participants consider the quizzes generated by F-System are more impressed than those of W-System. The result demonstrates that our method is more suitable for information provision instead of intelligence test.

6 CONCLUSION AND FUTURE WORK

In this paper, for providing an interactive way for the tourism information providers, we proposed a novel method to generate the tourism quiz automatically by extracting the feature words of topic keyword. We also developed a Automatic Quiz Generation System. By comparing with other method of quiz generation, we ensure that our method is more suitable for information provision instead of intelligence test. Because the quizzes are interesting, as well as they can impress the users.

In the future, we are planning to build a difficulty and interest model to modulate the quizzes automatically.

REFERENCES

[1] Jun Zeng, Sachio Hirokawa. Component Search Engine using Tree-View Interface for Tourist Blogs, 3rd International Conference on Awareness Science and Technology, Dalian, China.pp.327-331. 2011.

[2] Chengjiu Yin, Tetsuya Nakatoh, Sachio Hirokawa, Xiaobin Wu, Jun Zeng,A proposal of search engine “XYZ” for tourism events,JCAI2010 IEEE, pp.178-181, 2010.

[3] Masatsugu Tonoike,Satoshi Sato,Takehito Utsuro,Answer Validation by Keyword Association,Vol.2004 No.47, 2004-NL-161,pp.53-60, 2004.

[4] Shinichi Tsumori, Kenji Kaijiri, A Proposal for the Automatic Generation of Multiple-Choice Problems Adapted to Students' Understanding, Japanese Society for Information and Systems in Education,pp.26,2009.

[5] Mitkov, R., Ha, L.A., Karamanis, N.A computer-aided environment for generating multiple-choice test items Natural Language Engineering 12 (2), pp.177-194, 2006.

[6] Hoshino, A. and Nakagawa, H. A real-time multiple-choice question generation for language testing: a preliminary study, Proceedings of the second workshop on Building Educational Applications Using NLP, pp.17-20. 2005.

[7] http://chasen-legacy.sourceforge.jp/

[8] http://geta.ex.nii.ac.jp/

[9] G. Salton, The SMART Retrieval System—Experiments in Automatic Document Processing, Prentice-Hall, Inc. 1971.