Vocabulary Building Support System by Converting Web Pages

Taro Yabuki, Hiroshi Sakuta and Hiroki Yoshioka
Department of Integrated Information and Technology,
College of Science and Engineering, Aoyama Gakuin University
{yabuki,sakuta}@it.aoyama.ac.jp, yoshioka@idea.it.aoyama.ac.jp

Abstract

An English self-learning system is proposed for effective vocabulary building. The system is based on the annotation technology of Web pages. Learners can provide quiz pages from active Web pages like Wikipedia. They read texts and fill the blanks on quiz pages. The system gives their feedback to the results immediately and records them for organizing their learning history to make their individual learning guideline and new quiz pages for them. This system works on the proxy server and can be setup easily by users. The features are that learners can make vocabulary building exercises by quiz pages automatically which include practical usage of the target words, and that learning processes are recorded to propose learners' better step-up. This technology of converting Web pages is not restricted in English learning, but specified subjects can be applied.

1. Introduction

International communication in education is based on the common language, English. Even in Chinese character culture countries of the East Asia area, each nation's people have to depend on English to communicate with the others, like other area of the world. Japanese average people can only guess meanings of text from Chinese characters in Chinese documents. They cannot understand oral conversations by Chinese language even one word. Learning English is better solution to communicate each other for the nations.

English language education in Japan has been kept several hours a week for 6 years from junior hi-school to graduation of hi-school. Besides, in many cases, college students have to study more specified English course. Generally, the learners' vocabulary building is not effective through their school learning, because there are few chances for practical use. While global culture mixing through the Internet and video

media makes young pupils and students to access many English words, their perception may not reach to deeper meaning.
We propose a effective solution for English vocabulary building for English self-learner as a foreign language

1.1 Background in Japan

Learning Management Systems are applied in many schools and colleges in Japan. The learning system supplying problems for the vocabulary power improvement is, for example, working in English Language Education. It may, however, offer almost memorizing problems of separated word from English texts. Therefore, modern methods of natural learning in vocabulary building are not applied.
Even if effective educational support system is introduced, time for preparation is needed as well as the case of study by past "Paper and pencil" for making the problem. If the problem making work can be automated, such a load is sure to be reduced greatly. In this report, a system is proposed that automatically generates the making up problem for the English word acquisition on the Web page written in English as an example of automating the problem making work. This system is achieved by applying technology that rewrites the Web page proposed by Watanabe[1]. Effective study is expected to be done by recording user's behavior on the rewritten Web page, and making the learning scheme based on the record.

1.2 Purpose

The learning system that supports the vocabulary power improvement of English is developed. The system can provide a learning method that memorizes the word while reading English texts. English Web page in the Internet is converted into a fill-in-the-blank question, and apply to that. The load of the problem generation and teachers' load are reduced so that the system may do making and grading the problem by the automatic operation.

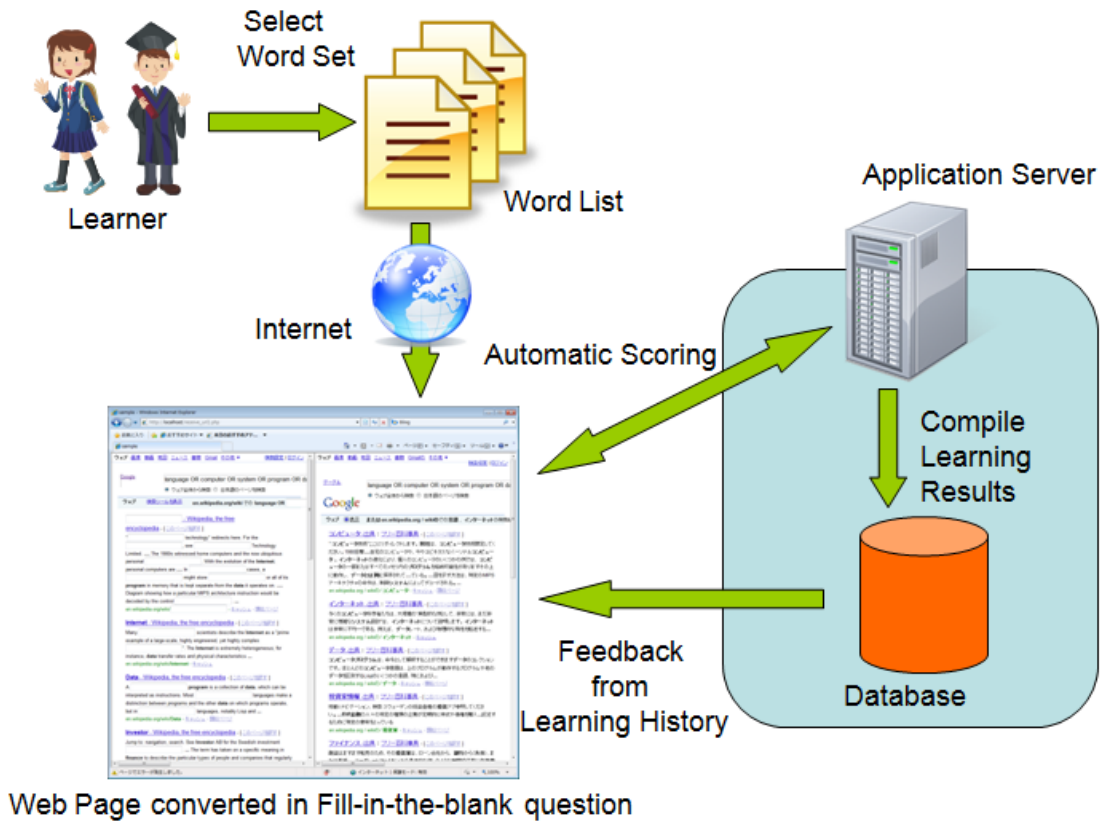

Figure 1. Concept of System Configuration

1. System Outline

2.1 Features

-Conversion Web page to the making up problem

Arbitrary Web page becomes a fill-in-the-blank question by rewriting the key word part to the text box.

-The grading of the fill-in-the-blank question

The grading is done immediately after the input of the answer, and each learner's learning history is recorded as for the result, which word on which page when the preservation learner of the study history notified at once by the learner (The example: Change into red if it is an incorrect answer to blue if the text box is a correct answer) studied

-Recording of each learner's learning context

-Learning plan

To set the problem of the suitable word by the theory of the forgetting curve according to the learning context of the plan learner of the learning scheme based on

the study history at the appropriate time, the learning scheme is planned.

2.2 System configuration

This system consists of the grading server with the conversion server that converts the Web page into the fill-in-the-blank question. When the learner tries to access the Web page selected with the grading server, the Web page is converted into the fill-in-the-blank question with the conversion server. When the learner inputs the word, the content is transmitted to the grading server, and graded at once. The grading result is accumulated in the data base, and influences the selection of the Web page used next.

2.3 Generation of fill-in-the-blank question

The generation of the fill-in-the-blank question is achieved by rewriting key words part on English Web page to text boxes. The key words are selected from the list that the learner or the teacher prepared beforehand by the automatic operation (Details of the selection method are described later). Learning is done by inputting the answer to the corresponding text box. The screen can be divided optionally as shown in Figure 2. Japanese translation is displayed in the right side corresponding to converted original page in the left side as the fill-in-the-blank question. Learners can guess the answers by this translation as a hint.

Figure 2. Web Page converted into Fill-in-the-blank question

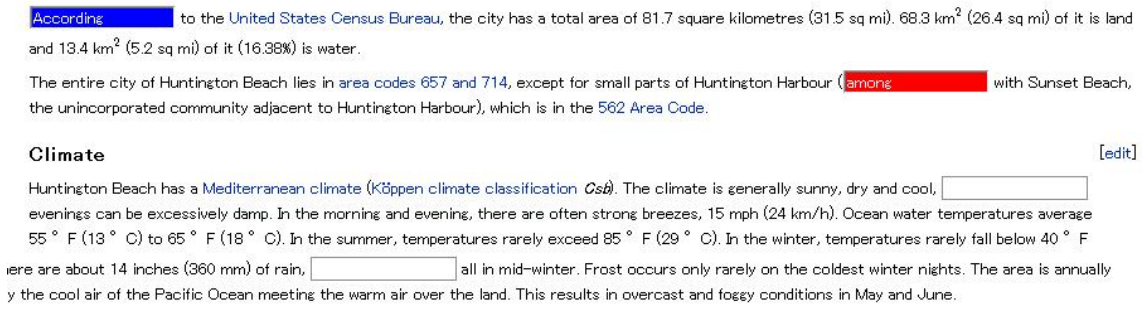

Figure 3. Automatic grading by color to answers

3. Learning steps

3.1 **Rank division of word**

To achieve Auto Select of the key word, this system divides the word in the selected list into five ranks. The rank division is done in consideration of the forgetting curve and the study history concept.

3.2 Forgetting curve

The forgetting curve is that shows the relation of forgetting a mid-term memory and time. According to the theory of the forgetting curve, the interval of first learning and second is 1 week, the one of second and third is 2 weeks, and the one of third and forth is 4 weeks. Then, when the fourth study ends, it becomes a long-term memory. To achieve this, the system divides the word into the following ranks.

-Rank 1: Words never studied
-Rank 2: Word the first study ends
-Rank 3: Word the second study ends
-Rank 4: Word the 3th study in rank ends
-Rank 5: Word the 4th study in rank ends

3.3 Study history

URL, learner ID, word list on the Web page that became it in the fill-in-the-blank question, and the key words are each learners' learning history. The word is distributed to each rank by the study end frequency. It is judged that learning ends when a correct answer is input to the making up part, and one Rank : above. The learning end date of the history is used to decide the study time of each word. The study time is calculated by the theory of the study end date and the forgetting curve.

3.4 Automatic generation function of key word set

When the fill-in-the-blank question is generated, sets of words that become key words from the word into which the rank is divided based on the following priority level (key word set) are first generated automatically.

1. Key word set in all words that reached at study time of words from rank 2 to rank 4
2. Key word set in all words of rank 1
3. Key word set in all words of rank 5

The system will set the problem by the best word by prioritizing the set of the key word of the word that reached at the key word set learning time in all words of rank 5 at the best study time based on the forgetting curve.

3.5 Automatic generation function of page set

When the key word set is decided, the Web page to be converted next is extracted on the page of English version Wikipedia based on the word of the key word set. The extraction is done by searching for the set on the page where all words of the given key word set are contained by using the result of totaling the appearance word and the occurrence count in each page of English version Wikipedia preserved in the data base of the prior preparation.

3.6 Automatic grading

An automatic grading is done by communicating with the grading server at the input of the answer. ID is allotted to the text box in the page when the making up problem is converted, and ID of the text box and the correct answer corresponding to it are recorded in the grading server. Grading is done by comparing correct answers recorded as the answer that the learner input by the Ajax communication based on ID of the text box.

4. Implementation

In the grading server, an automatic grading is done by the Ajax communication. PHP and Relational Database Management System (MySQL) are used for the operation of the data base. The selection of the page that becomes original has been achieved by PHP. Moreover, the conversion of the page in the conversion server is done by PHP.

5. Conclusion

In this report, the system that applied the technology of rewriting the Web page to the English word study was developed. It is thought that this system enables problem

generators' load reduction of incidence to feed back by the automatic operation after the problem is made, graded, and it studies. Moreover, this system provides the function to plan the learning scheme automatically in consideration of the forgetting curve, and, as a result, it is thought that effective study can be supported.

Referances

[1] Takashi WATANABE, Taro YABUKI, and Hiroshi SAKUTA, Development of real time monitoring of Web browsing in classes and learning status. IEICE（Journal of The Institute of Electronics, Information, and Communication Engineers [Educational Engineering]）, Vol. 109, No. 335, pp.37–42, 2009.