Supporting Teaching/Learning with Automatically Generated Quiz System

Kousuke Mouri
Tokyo University of Agriculture and Technology, Japan
mourikousuke@gmail.com

Noriko Uosaki
Osaka University, Japan
n.uosaki@gmail.com

Mohammad Hasnine
Kyoto University, Japan
hasnine.mohammadnehal.5z@kyoto-u.ac.jp

Atsushi Shimada
Kyushu University, Japan
atsushi@ait.kyushu-u.ac.jp

Chengjiu Yin
Kobe University, Japan
yin@lion.kobe-u.ac.jp

Keiichi Kaneko
Tokyo University of Agriculture and Technology, Japan
k1kaneko@cc.tuat.ac.jp

Hiroaki Ogata
Kyoto University
Japan
hiroaki.ogata@gmail.com

Abstract: This paper describes an automatic quiz generation system to support memory retention utilizing digital textbook logs as well as to reduce teachers' burden to create quizzes. An experimental was conducted to examine whether our automatically generated quiz system more effective in learners' learning achievement than teacher-created quizzes. Twelve international students participated in the evaluation experiment. The result showed that our proposed system and teacher-created quizzes were equivalently effective. A correlation analysis was conducted in order to find the correlations among learning achievement, the number of quizzes and each variable in a questionnaire. It was found that the number of the quizzes and their learning achievement had strong positive correlation.

Introduction

In recent years, digital textbooks have been drawing attention from educational researchers all over the world. Previous researches have reported that the benefits and effects of introducing digital textbooks to schools (Nakajima, 2013; Shin, 2012). However, little attention has been paid to analyzing the digital textbook logs to facilitate teaching and learning. Our research project has focused on analyzing and visualizing the collected digital textbook logs to facilitate teaching and learning (Mouri & Ogata, 2015; Shimada, Mouri & Ogata, 2017).

We have developed an automatic quiz generation system to support memory retention utilizing digital textbook logs. The concept of the quiz generation is based on Ebbinghaus and retrieval failure theory to make learners sustain their memory. Besides, it takes a lot of time for teachers to create quizzes manually, while our system automatically creates adaptive quizzes by analyzing learners' digital textbook logs. Therefore it is expected to reduce teachers' burden. In order to measure the effect of our automatic generation quizzes compared with teacher-created quizzes, an experiment was conducted.

Literature Review

Researches related to memory retention

Human memory has been classified into two groups: short-term memory and long-term memory (Tulving 1972). According to (Carter 1975), note-taking was effective regarding memory retention for a short-term. However, in order to retain memory it is necessary for learners to convert their short-term memory into long-term memory.

It is reported that the learning only by highlighting the textbooks was not effective unless doing repeated learning (Johnson & Wen, 1976). Repeated learning helps learners retain their long-term memory. It is often argued that memory retention declines in time. Ebbinghaus developed a forgetting curve based on the collected data from his evaluation experiment. The forgetting curve showed that learners tend to forget a half of what they have learned after 1 day. On the other hand, Pimsleur (1967) proposed a memory schedule which defines the length of recall interval is 5 times. It is important to consider recall interval for providing quizzes based on results of the previous

studies.

When learners highlight keywords in digital textbooks, they store not only their spelling and meaning, but also the contexts before and after them. When learners try to recall the highlighted keywords, one or more of these cues will help them remember better. Several researchers reported that the retrieval cues help learners recall their past knowledge (Tulving, Endel, & Osler, 1968; Li et al. 2013, Mouri et al, 2013). With these reviews, we considered learners' recall interval for generating appropriate quizzes and an effective learning method using retrieval cues.

Efficacy of Quizzes

Generally, human memory has been classified into two groups: short-term memory and long-term memory (Tulving, 1972). According to Carter et al. (1975), they reported that note-taking was effective regarding memory retention for a short-term. However, significant co-relation between note-taking and long-term memory retention was not found. It is necessary to convert short-term memory into long-term memory.

When learners highlight keywords in digital textbooks, they store not only their spelling and meaning, but also the contexts before and after them. When learners try to recall the highlighted keywords, one or more of these cues will help them remember better. Several researchers reported that the retrieval cues help learners recall their past knowledge (Tulving, Endel, & Osler, 1968; Li et al. 2013, Mouri et al, 2013). With these reviews, we considered learners' recall interval for generating appropriate quizzes and an effective learning method using retrieval cues. Researchers have focused on the quizzes to provide repeated learning. According to Karpicke et al. (2007), they reported that repeated learning enhanced learning effectiveness. Therefore, many quiz systems have been proposed to enforce learners' memory retention..

For example, Flanagan et al. (2013) proposed an automated method to generate fill-in-the-blank and multiple-choice quizzes from learners' writing. Zeng et al. (2013). proposed an automated method to generate fill-in-the-blank quizzes based on keywords that appear in the same sentences. Li et al. (2013). proposed an automated method to generate yes/no and multiple-choice quizzes based on what they have learned in the ubiquitous learning system called SCROLL. It is necessary to consider the automatic method to provide adaptive quizzes. With these reviews, we proposed an automatic quiz generation method using digital textbook logs.

Previous Studies

To collect digital textbook logs, we developed a digital textbook system called AETEL (Kiyota et al. 2015a, 2015b). Fig. 1 shows the course directories created by a teacher. Teachers create contents using PowerPoint or Keynote prior to their class and use them in their class. The uploaded contents are converted to EPUB format.

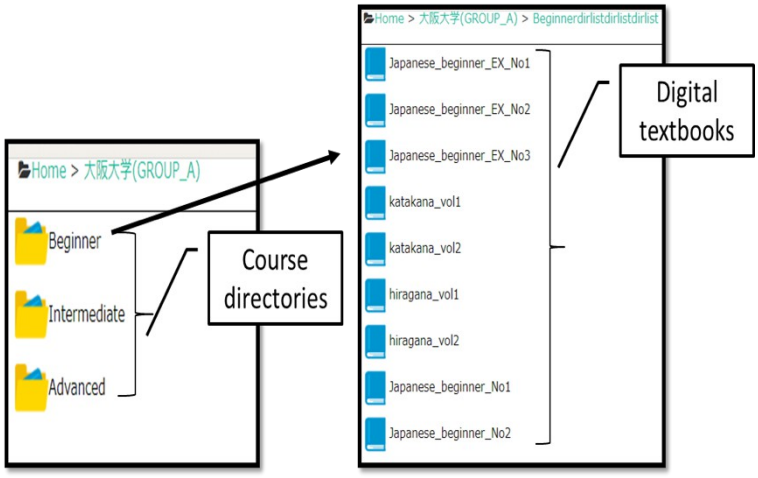

Figure 1. a selection interface of digital textbooks

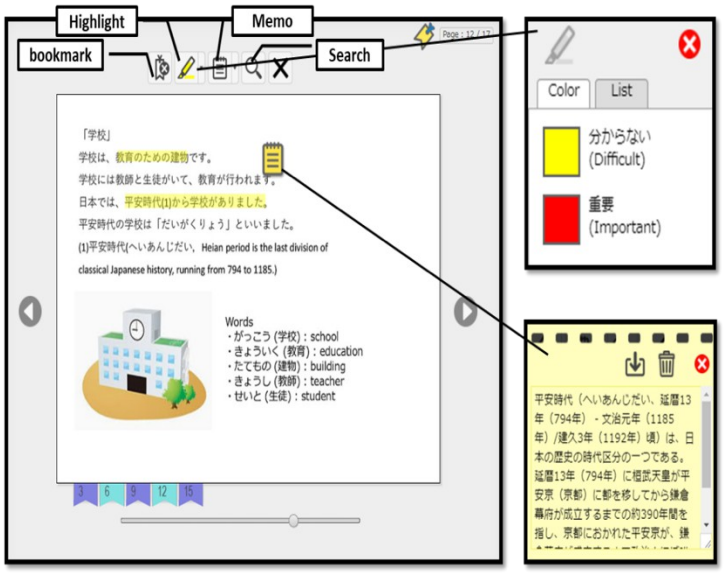

Figure 2. a digital textbook viewer

Fig. 2 shows an interface of digital textbook viewer. Learners can read the digital textbooks on their web browser anytime and anywhere. The functions of the digital textbook viewer consist of four components: bookmark, highlight, memo and search. By clicking the bookmark button in a page, the page is saved as their favorites page, so that learners can find the page quickly and easily afterward. When learners want to learn important words in the digital textbooks, they can highlight them with red color. When learners want to learn new words in the digital textbooks, they can highlight them with yellow color. By using different colors, we can analyze their intentions of highlights. When learners click the memo button in a digital textbook, they can write a memo concerning the target words as shown in Fig. 2 (Right-bottom). In addition, they can search the page number corresponding to the target word in the digital textbook by clicking the search button.

Automatic Quiz Generation System

Overview

The purpose of the automatic quiz generation system is to provide adaptive quizzes in accordance with the highlighted logs for repeated learning. If the highlighted words are too many, there is a possibility that it might generate too many quizzes. Therefore, it is necessary to consider how to provide quizzes effectively. To provide more appropriate quizzes, we hypothesize that the words that many learners highlighted could be regarded as more important.

To examine the above hypothesis, this study first summarized highlighted logs corresponding to each learner in a digital textbook as a matrix. We define the rows as each learner and the columns as each keyword. Secondly, this study determined the priority of the quizzes using the matrix. Finally, the quizzes are provided with hints based on retrieval cues theory.

Recall Interval

In order to overcome the problem of forgetting, it is necessary to consider a proper strategy to retain our memory as discussed in Section 2. This study adopts a recall method proposed by Pimsleur (1967), which states that it is efficient for learners to recall if the interval is 5 times of the previous interval. The recall interval is given the following formula:

This study sets the first quiz recall interval () as 5 hours. It means that the quiz is available 5 hours later after a learner highlighted a keyword in a digital textbook. If the learner answered the quiz incorrectly, the next interval will be reset as the first interval. Next, this study defines the following formula in order to determine the priority of quizzes in the appropriate recall interval.

shows a matrix of the highlighted keywords corresponding to each learner in a digital textbook. The purpose of this formula is to provide the quizzes giving the words which the more learners highlighted the more priority.

Hint interface based on retrieval cues

To assist their recall process more effectively, we implemented a hint function. Fig. 3 shows the hint function. Fig. 3(left) shows the image before clicking the hint button and Fig. 3(right) shows the image after clicking the hint button. By clicking the hint button, learners can answer it while recalling the spelling of their learned words in the blank with the hint. It is expected that the hint function facilitate learners to recall them.

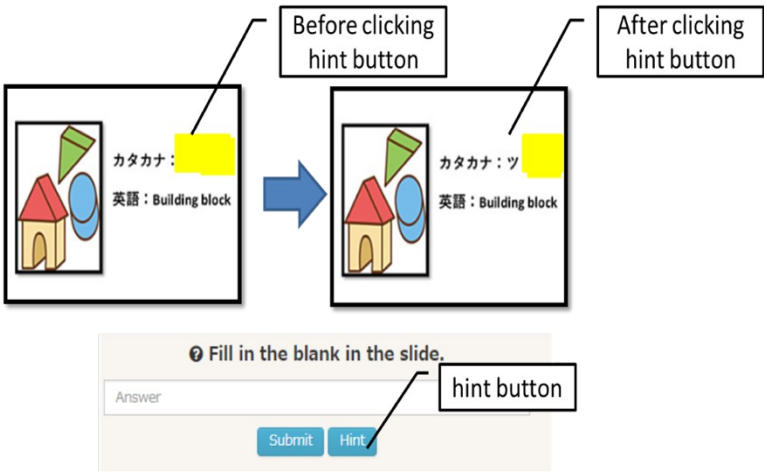

Figure 3. a hint interface

Quiz interface

Fig. 4 shows the quiz interface.

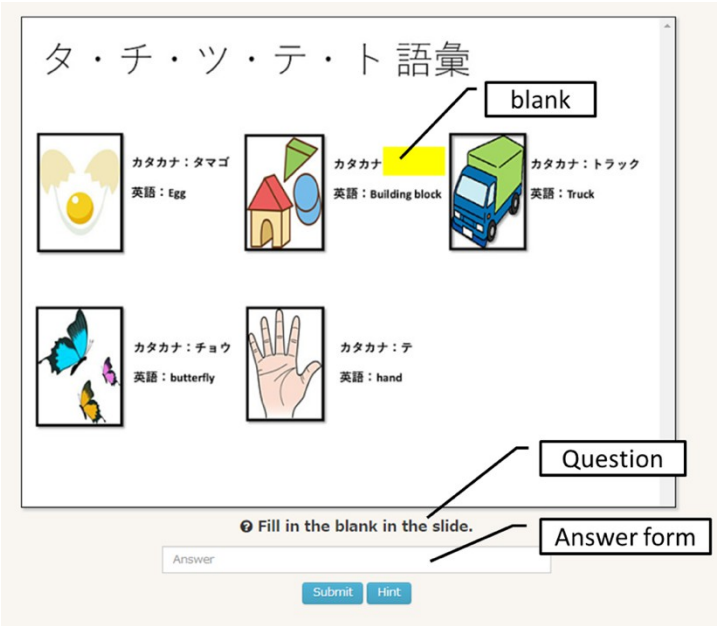

Figure 4. quiz interface

The quiz system makes the the highlighted words blank. This function allows learners to recall their highlighted words, which leads to support their memory retention.

Experimental Design

The purpose of this experiment was to explore the effect of our proposed system compared with teacher-created quizzes. It is one of the objectives of this study to reduce teachers' burden in creating quizzes. Therefore, a comparison was made between the learners who used our proposed quiz system and those who took the teacher-created quizzes.

Participants

Twelve international students who were studying Japanese language at the University of Osaka in Japan participated in the evaluation experiment. They were from China, German, Russian, Swedish, France, Finland, British and Macau.

Instruments

The instruments adopted in this study were pre-test, post-test, digital textbook logs and the questionnaires for testing user acceptance. The pre-test and post-test consist of 40-items vocabularies. The participant selected the test-level and textbook-level according to the results of their Japanese language level check created by CIEE (Center for International Education and Exchange), Osaka University. The questionnaire was created based on Technology Acceptance Model (TAM). TAM was proposed by Davis (1993) to diagnose system design problems. He claimed that Perceived Ease of Use (PEU) and Perceived Usefulness (PU) are two fundamental factors of user acceptance technology. PU means that the degree to which a person believes that using a particular system would enhance his or

her job performance. PEU means that the degree to which a person believes that using a particular technology would be free from effort. In addition to these factors, Lee et al. (2003) extended to Perceived Enjoyment (PE) and User Intention (UI). PE means that the degree to which users prefer to use a particular system. UI means that the degree to which users want to use a particular system in the future.

Procedure

Fig. 5 shows the experimental procedure.

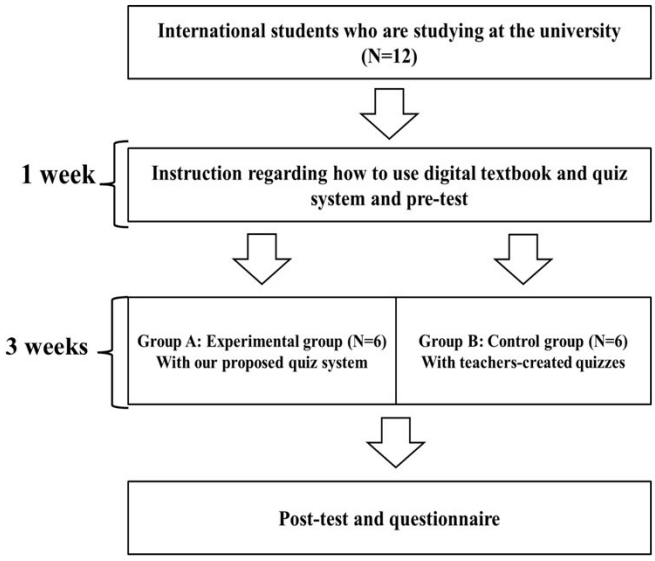

Figure 5. the experimental procedure

The teacher uploaded digital textbooks to the server before her class. First, she explained how to use the digital textbook and quiz system since it was their first time to use them. After that, the participants took the pre-test.

The participants were divided into the experimental group and the control group based on the result of the pre-test. Tab. 1 shows the result of the pre-test in each group. The means and standard deviations were 32.6 and 6.15 for the experimental group, 31.1 and 5.58 for the control group. The t-test shows that there was not significant difference between the two groups ($t=0.44$, $p>0.05$).

The participants of both groups learned vocabularies in the digital textbooks during the evaluation phase for three weeks. The experimental group reviewed the vocabularies taking quizzes created by our proposed system, while the control group reviewed the vocabularies taking teacher-created quizzes. After the evaluation experiment, the participants took the post-test and answered five point-scale questionnaires for testing user acceptance of the digital textbook system.

Group	N	Mean	SD	t	p
Experimental	6	32.6	6.15	0.44	p>0.05
Control	6	31.1	5.58		

Table 1. The result of the pre-test (full mark: 40)

Results

Learning achievement

Tab. 2 shows the result of the post-test. The t-test shows that there was no significant difference between them (t=1.34, p>0.05). It means that there was no significant difference in the participants' learning achievement when learning with two different approaches.

Group	N	Mean	SD	t	p
Experimental	6	35.8	2.91		
Control	6	32.6	6.26	1.34.	p>0.05

Table 2. The result of the post-test(full mark: 40)

Analysis of digital textbook logs

To examine the number of the quizzes by our proposed system, we compared the number of the quizzes taken by the experimental group with that of the control group. The t-test shows that there was no significant difference between the two groups (t=0.09, p>0.05). Therefore, there was no significant difference as for the number of the quizzes between the two different approaches.

Group	N	Mean	SD	t	p
Experimental	6	98.5	64.8	0.09	p>0.05
Control	6	94.1	103.1		

Table 3. The result of the number of the played quizzes

Questionnaire for testing user acceptance

Tab. 4 shows the result of the five-point-scale questionnaire. PE showed that the lowest mean score, which was 2.97 when the participants were asked whether they enjoyed using the digital textbook system. PEU shows that the highest mean score, which was 3.92 when the participants were asked whether it was easy to use the digital textbook system.

Variable	Question	Mean	SD
PE	I like to use highlight, memo, bookmark function to learn.	3.23	0.97
	I enjoy using the digital textbook system.	2.97	0.91
PEU	It is easy to remember how to use the functions of the digital textbook system	3.92	0.72
	The learning contents of the digital textbooks is clear and understandable	3.84	0.76
PU	I feel using the digital textbook system enhances my learning effectiveness	3.07	1.14
	I feel using the digital textbook system makes learning more efficient	3.15	0.94
UI	I think that I want to use the digital textbook system in the future.	3.07	0.91

Table 4. The result of TAM questionnaire

Correlation analysis

In order to find the correlation among the learning achievement, the number of quizzes and each variable in TAM, a correlation analysis was conducted. Tab. 5 shows the result of the correlation analysis.

With regard to PE, the result showed no correlation with Post-test, while it showed a positive correlation with UI and PU. With regard to PEU, a strong positive correlation was found with PU. With regard to PU and UI, the result showed a positive correlation with PE. With regard to Post-test, the result showed a strong positive correlation.

	PE	PEU	PU	UI	Post-test	Quizzes
PE	1.000					
PEU	0.274	1.000				
PU	0.422	0.704	1.000			
UI	0.614	0.351	0.471	1.000		
Post-test	0.051	0.124	0.524	0.325	1.000	
Quizzes	-0.365	0.029	0.242	-0.108	0.744	1.000

Table 5. The result of the correlation analysis

With regard to PE, the result showed no correlation with Post-test, while it showed a positive correlation with UI and PU. With regard to PEU, a strong positive correlation was found with PU. With regard to PU and UI, the result showed a positive correlation with PE. With regard to Post-test, the result showed a strong positive correlation with PU.

Discussion and Conclusion

According to the learning achievement in the evaluation experiment, there was no significant difference between our proposed system and teacher-created quizzes. When comparing the difference from pre-test to post-test in each group, the experimental group increased by 3.2 point, while the control group increased by 1.5 point. Therefore, the experimental group slightly increased the score than the control one. In addition, this study analyzed the number of the quizzes in each group. As a result, the number of quizzes taken by the both groups were not significantly different. Although a significant difference was not found, at least we can safely say that our automatic quiz system and teacher-created quizzes were equivalent in terms of effectiveness.

According to the questionnaire in the evaluation experiment, the PEU factor shows that the highest mean score. It means that it was easy to use the digital textbook system. Moreover, the correlation was found among learning achievement, the number of quizzes and each variable in TAM using a correlation analysis. As a result, this study found two strong positive correlations: PU and PEU and Quizzes and Post-test. From the correlation between the number of quizzes and the post-test result, we conclude that learning based on the quizzes enhanced their learning achievement.

This study developed an automatic generated quiz system to reduce teachers' burden in creating quizzes and examined the learning effect of our proposed system. The comparison was made with the teacher-created quizzes. In this study, our proposed system was able to enhance the learners' learning achievement with the increase of the number of quizzes they took. Our proposed quiz system will be evaluated repeatedly. We will consider to apply to other domains such as math, physics and programming education with a sufficient number of participants.

Acknowledgments

This part of the research was supported by Grants-in-Aid for Scientific Research No.17K12947, No.18K02820, 18H04125, and No.16H06304 from the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan. We would like to thank Prof. Bipin Indurkhya for proof reading.

References

Carter, J. F. Van Matre & Nicholas, H. (1975). Note taking versus note having. *Journal. of Educational Psychology*, 67(6), pp.900-904

Davis, F. D. (1993). User acceptance of information technology: System characteristics, user perceptions, and behavioral impacts, *International Journal of Man-Machine Studies*, 38(3), pp.475-487.

Ebbinghaus, H (1885). *Classics in the History of Psychology* (Translated by Henry A. Ruger & Clara E. Bussenius (1913) Originally published in New York by Teachers College, Columbia University. http://psychclassics.yorku.ca/Ebbinghaus/index.htm.

Flanagan, B., Yin, C., Hirokawa, S., Hashimato, K. & Tabata, Y. (2013). An Automated Method to Generate e-Learning Quizzes from Online Language Learner Writing, *International Journal of Distance Education Technologies*, 11(4), pp.63-80.

Johnson, D. & Wen, S. (1976). Effects of correct and extraneous markings under time limits on reading comprehension, *Psychology in the Schools*, 13, pp.454-456.

Karpicke, J. and Blunt, J. (2007). Repeated retrieval during learning is the key to long-term retention, *Journal of Memory and Language*, 57(2), pp.151-162.

Kiyota, M., Mouri, K. & Ogata, H. (2015a). Seamless Learning System Based on e-Book, *Proceedings of the 23nd International Conference on Computers in Education*, pp.448-452.

Kiyota, M., Mouri, K. & Ogata, H. (2015b). A Proposal of e-Book Based Seamless Learning System, *Workshop Proccdings of the 23nd International Conference on Computers in Education*, pp.611-616.

Lee, Y., Kozar, K. A., & Larsen, K. (2003). The technology acceptance model: Past, present, and future. *Communications of the Association for Information Systems*, 12(50), pp.752–780.

Li, M., Ogata, H., Hou, B., Uosaki, N. and Mouri, K. (2013). Context-aware and Personalization Method in Ubiquitous Learning Log System, *Journal of Educational Technology & Society*, 16 (3), pp.362–373.

Mouri, K., Ogata, H., Li, M., Hou, B., Uosaki, N. & Liu, S. (2013). Learning Log Navigator: Supporting Task-based Learning Using Ubiquitous Learning Logs, *Journal of Research and Practice on Technology Enhanced Learning*, 8(1), pp.117
128.

Mouri, K. & Ogata, H. (2015). Ubiquitous Learning Analytics in the Real-world Language Learning, *Smart Learning Environments*, 2(15), pp.1-18.

Nakajima, T., Shinohara, S., & Tamura, Y. (2013). Typical functions of e-Textbook, implementation, and compatibility verification with use of ePub3 materials. *Procedia Computer Science*, 22, pp.1344–1353.

Pimsleur, P. (1967). A memory schedule. *The Modern Language Journal*, 51(2), pp.73–75. http://psychclassics.yorku.ca/Ebbinghaus/index.htm.

Shimada, A., Mouri, K. & Ogata, H. (2017). Real-time Learning Analytics of e-Book Operation Logs for On-site Lecture Support, *The 17th IEEE International Conference on Advanced Learning Technologies*, pp.274-275.

Shin, J. H. (2012). Analysis on the digital textbook's different effectiveness by characteristics of learner *International Journal of Education and Learning*, 1(2), pp.23–38.

Tulving, E., & Osler, S. (1968). Effectiveness of retrieval cues in memory for words. Journal of experimental psychology, 77(4), pp.593–601.

Tulving, E. (1972). Episodic and semantic memory. In Tulving, E., & Donaldson, W. (Eds.) (1972). *Organization of Memory*. New York: Academic Press, pp.381–402.

Zeng, J., Sakai, T., Yin, C., Suzuki, T. & Hirokawa, S. (2013). Automatic generation of tourism quiz using blogs, Artifical Life and Robotics, 17(3-4), pp.42-416.