T. Hirashima et al. (Eds.) (2011). Proceedings of the 19th International Conference on Computers in Education. Chiang Mai, Thailand: Asia-Pacific Society for Computers in Education

A novel approach for enhancing student reading comprehension by activating prior knowledge

Jun Ming Chen[a*], Meng Chang Chen[b], Yeali S. Sun[a] & Ying Ying Chen[a]
[a] *Department of Information Management, National Taiwan University, Taiwan*
[b] *Institute of Information Science, Academia Sinica, Taiwan*
*d96725001@ntu.edu.tw

Abstract: Reading is an important skill for gaining knowledge and discovering new information in a global society. Reading competence also strongly influences a person's learning ability. However, in building better reading skills, one of the major difficulties is an absence of background knowledge to help learners read and understand material. Background knowledge helps learners make connections to construct clues within the text and determine the meaning of new vocabulary or sentences. Often, learners lack a mechanism to help them construct prior knowledge, preventing them from fully understanding their reading. To cope with these problems, this study adopts one of the most popular Web 2.0 techniques—social tagging—to help learners both read and understand English articles. We test our approach using a tag-based collaborative reading learning system. Our conclusions demonstrate the approach's effectiveness, and reveal areas for future research.

Keywords: Social tagging, Reading comprehension, Information retrieval

Introduction

Prior knowledge plays an important role in reading comprehension. Effective reading comprehension requires the integrated interaction of derived text information and pre-existing reader knowledge [5, 12], especially with learners of foreign languages such as English. The Scholastic Achievement Test (SAT), which claims to predict how well students will do in college, is very dependent on prior knowledge [2]. Studies have found that strong prior knowledge of subject material enables students to attain higher comprehension, performance, and motivation. This further suggests it is important to assist students in obtaining relevant prior knowledge, as this can enable them to engage meaningfully with the learning materials [10].

However, despite the value of prior information, Taiwanese senior high schools have largely focused on skills development rather than expanding children's knowledge of the world, such that reading comprehension and prior knowledge instruction are still a challenge in English as a foreign language (EFL) classes in Taiwan. Because of Taiwan's exam-oriented education, students spend most of their time preparing for tests, and rarely have enough time to acquire knowledge from extensive personal reading or living experiences. This leads to poor levels of reading comprehension among students, such that even average students are unable to read and fully understand material [2].

To cope with the problem, researchers and educators continue to seek new teaching methods. One such method includes using Web 2.0 tools to develop adaptive learning environments. To help EFL learners improve their English reading comprehension, recently researches have investigated the use of tagging mechanisms in e-learning systems that consider learner reading status and browsing behaviors [3, 4]. Although this approach can significantly enhance student reading comprehension and assist teacher assessment of literacy, there are a few problems to be aware of. One such problem is students who lack sufficient background knowledge and activities. Therefore, in this paper, we use social tagging services that allow users to annotate various online resources (materials) with freely chosen tags. The tagging certain activities can help students summarize new ideas and quickly grasp the structure and concepts of English articles. Moreover, these tags are also designed to enhance critical thinking skills by directing students to evaluate and then support or oppose different viewpoints on their readings. They not only facilitate the users in finding and organizing online resources, but also provide meaningful collaborative semantic data which can potentially be exploited by recommender systems [7].

Meanwhile, designing prior knowledge learning environments that help promote critical thinking through article construction can activate a learner’s existing schema and help them realize new information from articles more easily. Such background information may even help learners find clues for identifying the meanings of new vocabulary or sentence patterns [9, 11]. The most important value of this social tagging system, however, is the promise it shows for dramatically improving student reading comprehension.

This paper outlines our experiences with applying social tagging within an e-learning system to help students increase their reading comprehension by activating prior knowledge, summarizing new ideas and quickly grasping the structure and concepts of English articles. We then test novel approach and evaluate student learning performance, through a collaborative reading learning system [6].

1. TAK: A Tag-based Prior Knowledge Recommendation Approach

To help EFL learners improve their English reading comprehension, this study proposes an automatic personalized prior knowledge network, based on collaborative filtering by article and learner similarities, called Tag-based Prior Knowledge recommendation (TAK). TAK supports the rich prior knowledge of EFL learners and enhances their ability to correctly guess the meaning of unknown subject matter, especially science materials. Fig. 1 illustrates the framework of our approach, which consists of three parts: data preprocessing, topic filtering and personalized background knowledge discovery.

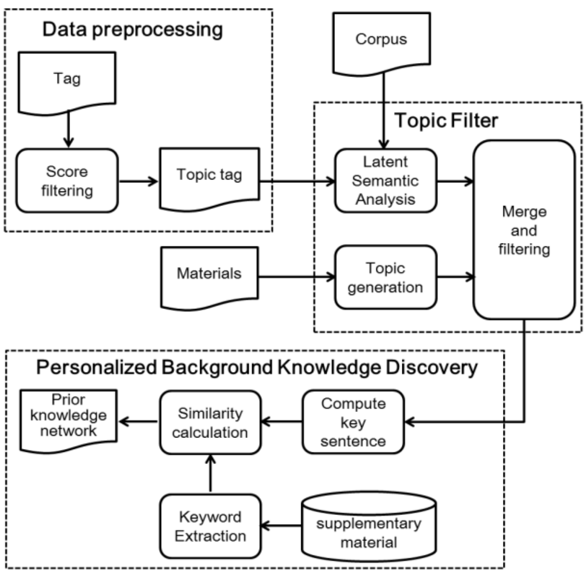

Figure 1: framework for the support approach

T. Hirashima et al. (Eds.) (2011). Proceedings of the 19th International Conference on Computers in Education. Chiang Mai, Thailand: Asia-Pacific Society for Computers in Education

1.1 Data Preprocessing

In order to diminish the impact of data sparseness, this study uses several preprocessing techniques for information retrieval, including Porter stemming and stop word. After the pre-processing of documents, tokenizing (dividing sentences into fragments), stemming (reducing irregular verbs to their base form) and stop-word removal (removing vague, non-descriptive wording), a number of different types of input data are generated for scoring functions on various data sparseness processes. A tag scoring mechanism is also constructed to filter out irrelevant student tags from useful ones.

The following section outlines the use of tags for articles and how they are filtered. In our experiment, each student reads an article and takes an exam to evaluate their comprehension. We assume the students' scores approximate their level of comprehension, and thus our tag score mechanism is a combination of a student's exam score with their tagging preferences for the same article. First, we use the standard error of measurement (SEM) to recalculate the scores so that these scores better mirror a student's comprehension, relative to their peers. To calculate the SEM, the confidence coefficient is first determined as follows:

$$r_{xx} = \frac{K}{K-1}\left(1 - \frac{M(K-M)}{K\sigma_t^2}\right)$$

where K is the number of questions in the article for which we want to filter useful tags(topic tags), M is the average exam score of every student who takes the exam, and $\sigma_t{}^2$ is the variance of every student who takes the exam. When we know the confidence coefficient r_{xx} of the exam, the standard error of measurement is calculated as:

$$SEM = S \times \sqrt{1 - r_{xx}}$$

In this formula, S is the standard deviation of the exam. Next, we use the standard error of measurement to recalculate the exam score. We give the tag score to evaluate the quality of the tag. The tag score that the student annotates can be represented as follows:

$$t_{xkj} = SEM_k \times \left(\frac{score_{xk}}{score_{avg.k}}\right) + score_{xk}$$

where t_{xkj} is the score of the j_{th} tag in the k_{th} exam that the x_{th} student annotates; D_k is the k_{th} exam of the article, $k = 1\sim o$; U_x is the x_{th} student, $x=1\sim n$; t_j is the j_{th} tag, j=1~m; $score_{xk}$ is the score of the k_{th} exam that the x_{th} student receives; and $score_{avg.k}$ is the average score of the k_{th} exam.

Finally, we sum each student's score of the same tag. This score represents the weight of the tag in the article, which is shown below:

$$T_{kj} = \sum_{x=1}^{n} t_{xkj}$$

1.2 Topic Filter

This study uses three methods to glean important clues from an article: extending the topic tag, generating the topic of an article, and combining the topic tag and topic to identify the article's meaning.

A topic tag is used as part of the score mechanism to glean useful tags that represent the article's meaning. However, because polysemy and synonym are major problems of language research, in order to decrease the influence of the above problem we extend the topic tag. For extending the topic tag and improving the precision of the analysis results, we use Latent Semantic Analysis (LSA) to find synonyms for the topic tag. Before using the tag to identify the main point of an article, we use LSA to calculate the similarity between sentences and group similar sentences into clusters and we name these cluster "block." We not only group the sentence, but also use LSA to calculate the similarity between words and blocks, and then selects the words of highest value to form the subtopic term of the block. . These blocks serve as subtopics for the article, some of which signify main points of the article while others are less important. In order to filter out less important subtopics, we consider three factors: article position, shared similarity with the main topic(the similarity between main topic and the subtopic), and tag overlap. Article position is defined as follows:

$$BS_{ih} = \frac{(|P_g| - L(s_i) + 1)}{|P_g|}$$

In this formula, BS_{ih} is the position value of the h_{th} sentence in the i_{th} block, $h=1\sim z$, $i=1\sim n$; P_g is paragraph, $g=1\sim q$; S_i is the sentence of the i_{th} block; $L(S_i)$ is the position of sentence S of the i_{th} block within the paragraph. After we know the position values of sentences in each block, they are summed to determine the position value of the block.

$$BP_i = \sum_{h=1}^{z} BS_{ih}$$

where BP_i is the weight of i_{th} block, $i=1\sim n$.
The shared similarity of the main topic is computed according to the following formula:

$$B_i = BR_{iy} - \max(BR_{i(i-1)}, BR_{i(i+1)})$$

where $BR_{i(i-1)}$ is the similarity between i_{th} block and $(i\text{-}1)_{th}$ block; $BR_{i(i+1)}$ is the similarity between i_{th} block and $(i+1)_{th}$ block; BR_{iy} is the similarity between i_{th} block and the main topic block. The shared similarity of the main topic can let us know whether the subtopic is a branch of the main topic, or just represents the context of other subtopics. If the value is negative, then we know this subtopic is dissimilar with the main topic, and thus is not a suitable branch of the main topic of the article. The weight of the block is then comprised of the summed positional value of the block, the similarity of the main topic, and tag overlap.

$$B_i' = T_{ij} + BP_i + B_i$$

Where B_i' is the weight of i_{th} block; T_{ij} is the subtopic term of i_{th} block overlapping with a topic tag from the article. If they overlap, then $T_{ij} = T_{kj}$; otherwise $T_{ij} = 0$.

1.3 Personalized Background Knowledge Discovery

Before identifying the level of a student's background knowledge, we recommend personal topics to each student. When students read an article, they affix tags to it, which helps determine what the students focused on while reading. We compare these tags with the topic tag that we found via our tag scoring mechanism. Based on how well their tags match our topic tags, we will recommend the topic tag to them. Every student has his or her own recommended topic tags.

To find suitable background knowledge for a student, we recommend article topics by examining every key sentence within a topic and recommend topic tags to find sentences with similarities. We then present these sentences along with the article to the student. This process is shown below,

$$Sim_{ki}(S,T) = \sum_{j=1}^{m} \frac{T_{kij}}{\sum_{j=1}^{m} T_{kij}} \times W_{sj}$$

Where S is the set of sentences; T is the set of tags; T_{kij} is the weight of the j_{th} tag of i_{th} block in the k_{th} article, $j=1\sim m$; W_{sj} is a binary value. When the j_{th} tag exists in the sentence S, then $W_{sj} = 1$; otherwise $W_{sj} = 0$.

Because the students have their own recommended topic tag, the key sentence will be different for each student. When the key sentences of every student are located, we calculate the similarities between these key sentences and the sentences from the background knowledge articles. This helps identify the most proper articles for helping improve background knowledge.

2. Experiment Design

To evaluate the efficacy of our approach, an experiment was conducted from March 2011 to May 2011 on reading activity at a senior high school in Taiwan. This section outlines the details of our experiment.

First, 60 participants were divided into two groups, an experimental group and a control, and were taught how to use the on-line learning platform. Then, we adopted a pre-test and post-test experimental design that employs before-and-after surveys to demonstrate the usefulness of our approach among participants. Each group was given a pre-test evaluation and a post-test over one month. All other conditions, such as the selected reading material, were kept identical between the two groups.

We select some reading comprehension examination questions from College Entrance Examination Center (CEEC) and simulated exam. The reading articles are all belong to science field like biology and global warming because in the past study, science article differ from social or history article, the science concept are more difficult to ground in everyday experiences so the importance of prior knowledge is more significant in the science field than other field.

After the experiment, a post-test was used to test for differences in improvement in reading comprehension ability between the control and experimental group. The experimental group answered a questionnaire that used a five-point Likert Scale to evaluate the level of agreement with the learning model and the usage of learning systems.

Based on our novel approach, we implemented a tag-based prior knowledge recommendation and learning system (TAK) on a Windows 2003 server. This system provides both reading and guided interfaces in a tag-based learning environment.

As shown in Fig. 2 the student interface consists of four operation areas: (1) article content located in the left side of the window; (2) An input and input hints area located in the lower-middle side of the window, where students can construct meaningful words or phrases to represent the article's ideas; (3) a reading quiz located in the lower-right side of the window; (4) lastly the "network of prior knowledge" and "personal structure annotations" functions are located in the upper-middle and upper-right sides, respectively. The system processing flow is follows: (1) User selected articles by drop down menu and then system shows the context of the articles; (2) Meanwhile, the functions (the network of prior knowledge and personal structure annotations) showed in screen that article content is highlighted when students select the "personal structure annotations" function. This highlighted information provides a quick but useful snapshot of an article's major themes. Furthermore, "the network of prior knowledge" provides the adaptive and necessary background knowledge and personal topic construction when students select any node in the prior knowledge network in the middle side of the window (3) The student can utilize the input area to create a list of tags. This tag list provides a quick but useful snapshot of a article's major themes or ideas (4) Lastly, a reading quiz located in the lower-right side of the window, which helps teachers evaluate student comprehension of article content.

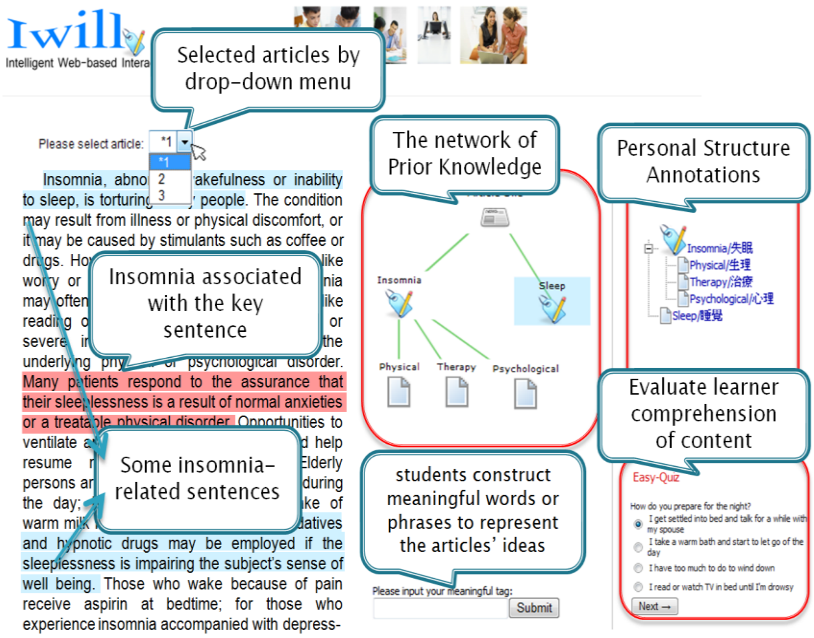

Figure 2: interface of Tag-based Prior Knowledge recommendation (TAK)

3. Experimental Results

3.1 Learning Achievements

Pre-test: all students took a pre-test at the beginning of the reading activity. Table 1 shows the t-test values for the pre-test and post-test results. Here, $|t| = 1.1013 < t_{\alpha}(30) = 1.697$, which implies that the performance of the control and experimental groups in the pre-test is not significantly different. In other words, before performing the experiment, the pre-test revealed that the control and experimental groups demonstrated a similar understanding of the learning topics at an alpha level of 0.05.

Post-test: according to the mean value of the post-test in Table 1, the experimental group performed better than the control. After participating in the learning activities, the experimental group achieved a significant improvement compared to the control group ($t= -2.938$, $p < .01$). The experimental students demonstrated significant improvement in reading comprehension ability by taking advantage of the TAK system and thus enhancing their reading comprehension ability. Lastly, all students were also asked to fill out a questionnaire to understand their learning behavior, system usage, and satisfaction with the system. In the next subsection, the analysis of this survey is discussed.

Table 1: Paired t-test of the pre-test and post-test results

Test	Group	N	Mean	Std. Deviation	Std. Error Mean	t-test
Pretest	Experimental Group	30	52.7778	19.7364	3.6034	t = -1.013
	Control Group	30	47.7778	22.0949	4.0340	p = .319
Posttest	Experimental Group	30	59.0000	23.0247	4.2037	t = -2.938
	Control Group	30	43.3333	27.0164	4.9325	**p = .006****

**p < 0.01

3.2 Questionnaire Analysis

The results of 61 questionnaires are displayed in Table 2, with respondent scores ranging from 1 to 5 (strongly disagree, disagree, neutral, agree, and strongly agree). Each question underwent a discriminate validity test by using factor analysis. The coefficients from the experimental results show that these factors were sufficiently reliable for representing student-tagging behaviors. The major findings are presented as follows:

(1) 72% of students indicated that using tags was easy, and that is was easy to translate the context of the original article into their own words.
(2) 67% of students thought that activating prior knowledge can help students summarize new ideas and quickly grasp the structure and concepts of English articles. Some students indicated that these tags can help them easily realize new information from the article, and that they even used the background knowledge clues to guess the vocabulary or sentence meaning.
(3) 60% of students found that the novel system was easy to use, and only a few did not perceive usability.
(4) 93% of students agreed that learning new words and knowledge from background information is easier and more interesting than the traditional learning method.
(5) Most of the students agreed that the TAK system is capable of helping them easily comprehend the context of reading articles, and can help them improve their reading efficiency.

Table 2: Results of the usability questionnaire for evaluating the purposed system

Questionnaire Item (four factors)	Strongly Disagree (%)	Disagree (%)	Neutral (%)	Agree (%)	Strongly Agree (%)
Usefulness of tagging	-	3.28	24.59	34.43	37.70
Usefulness of the network of prior knowledge clues	-	4.92	27.87	40.98	26.23
TAK is easy-to-use	-	9.84	31.15	36.07	22.95
Usefulness of the TAK system	-	-	40.98	40.98	18.30

4. Discussion and Conclusions

This study extends the application of social tagging by designing a tag-based prior knowledge recommendation and learning system (TAK) to provide opportunities for students to demonstrate knowledge connection. Meanwhile, new learning unfolds as students attempt to reduce inconsistencies between their existing knowledge structures and new information [1]. The experimental results demonstrate that the proposed system can effectively assist students in enriching prior knowledge and raise students' learning achievement. Despite these encouraging results however, there are still difficulties in creating a quality measurement of social tagging for tag-based learning environments [8]. One major problem is that tags have issues with both sparseness and noise. Here, sparseness refers to problems with users not applying any tags at all to certain sections or web pages, especially those web pages that are common, too new, or uninteresting. Before performing our experiment, this study used several preprocessing techniques to reduce the influence of sparseness, including Porter stemming and stop word [3]. In addition, pre/post-test results show that the control group had worse results on the post test. This may have been due to a lack of general searching proficiency and inductive capacity that kept some students from successfully extending their knowledge, such that students have difficulty in constructing knowledge effectively and enhancing learning achievement. One future solution to this problem is to use our novel approach, as well as makes use of intelligent social tagging technology in order to help guide student knowledge construction. The study also proposes a series of tag implementation guides to ensure that students tag ideas successfully. Further research is needed to investigate this methodological concern and its practical applications.

Acknowledgements

This work was supported by the National Science Council of Taiwan, R.O.C., under contract numbers NSC100-2221-E-001-015-MY3.

References

[1] Ainsworth, S., & Burcham, S. (2007). The impact of text coherence on learning by self-explanation. *Learning and Instruction*, 17, 286–303.

[2] Richard, L. A., & Patricia, M. C. (2001). Schools that Work: Where All Children Read and Write. *Boston: Allyn & Bacon*.

[3] Chen, J. M., Chen, M. C., & Sun, Y. S. (2010). A novel approach for enhancing student reading comprehension and assisting teacher assessment of literacy. *Computers & Education*, 55(3), pp. 1367-1382.

[4] Chen, J. M., Chen, M. C., & Sun, Y. S. (2010). A novel approach for assisting teachers in assessment of student reading ability in web-based learning environment. *International Conference on Computers in Education (ICCE)*, 260-264.

[5] Corredor, J. (2006). General and domain-specific influence of prior knowledge on setting of goals and content use in museum websites. *Computers & Education*, 47(2), 207-221.

[6] Ding, N. (2009). Visualizing the sequential process of knowledge elaboration in computer-supported collaborative problem solving. Computers & Education, 52, 509–519.

[7] Guan, Z., Wang, C., Bu, J., Chen, C., Yang, K., Cai, D., & He, X. (2010). Document Recommendation in Social Tagging Services. *In Proceedings of the 19th international conference on World Wide Web*, pp. 391-400.

[8] Gupta, M.; Li, R.; Yin, Z. & Han, J. (2010). Survey on Social Tagging Techniques. *SIGKDD Explorations*, 12(1), 58-72.

[9] Hakkarainen, K., Lipponen, L., & Jarvela, S. (2002). Epistemology of inquiry and computer-supported collaborative learning. In T. Koschmann, R. Hall, & N. Miyake (Eds.), CSCL 2: Carrying forward the conversation, pp. 129-156. Mahwah, New Jersey: Lawrence Erlbaum.

[10] Lin, Y. C., Lin, Y. T., & Huang, Y. M. (2011). Development of a diagnostic system using a testing-based approach for strengthening student prior knowledge. *Computers & Education*, 57(2), pp. 1557-1570.

[11] So, H. J., & Brush, T. A. (2008). Student perceptions of collaborative learning, social presence and satisfaction in a blended learning environment: relationships and critical factors. *Computers and Education*, 51(1), 318-336.

[12] Wang, T. H., Wang, K. H. & Huang, S. C., & Chen, S. Y. (2004). Web-based Assessment and Test Analyses (WATA) system: Development and evaluation. *Journal of Computer Assisted Learning*, vol. 20, no. 1, pp. 59–71.