A Following Method of Annotations on Updated Contents and Its Evaluation

Hisayoshi Kunimune[1], Kenzou Yokoyama[2], Takeshi Takizawa[2], and Yasushi Fuwa[2]

[1] Faculty of Engineering, Shinshu University
[2] Graduate School of Science and Technology, Shinshu University
4-17-1 Wakasato, Nagano City, Nagano 380–8553 Japan
kunimune@cs.shinshu-u.ac.jp

Abstract. We have already developed an annotation sharing system for web-based learning materials. This system allows learners to write annotations such as markers and memorandums directly on materials and to share these annotations with lecturers and learners. If necessary, web-based materials can often be updated by authors; however, the annotations on the material are not in the proper position after the update. In the present study, we propose a method to follow the proper position of annotations in updated materials, and conduct experiments to evaluate the method. The following paper describes the proposed method and the experimental evaluation.

Keywords: annotation, e-learning, updated contents, following method.

1 Introduction

Web-based training courses are in widespread use by educational facilities and companies. These courses provide materials as web pages, and these web-based materials have the following useful features compared to paper-based materials:

- Multimedia contents such as movies, audios, and images can be included.
- The contents can be revised by lecturers at anytime.

Despite the popularity of web-based courses, learners still like to write annotation texts and marks on paper-based materials to aid their understanding[6]. However, if learners print these web-based materials to write annotations, they lose many of the benefits of web-based materials as mentioned above. That is, the printed materials do not include any multimedia contents and become out of date, whereas the web-based materials can be updated and revised at anytime. Many learners studying self-paced distance e-learning courses have problems related to printed web-based materials[3].

To solve these problems, several systems that allow the user to incorporate annotations in the web-based contents and to share these annotations have been proposed[1,2]. Learners can write annotations directly on web-based materials

J.D. Velásquez et al. (Eds.): KES 2009, Part II, LNAI 5712, pp. 555–563, 2009.

with these systems and browse the materials with multimedia contents. However, lecturers update the materials, and annotated sentences or words are moved and not displayed in the proper positions.

We proposed an annotation sharing system called *Writable Web* [3,4], which is implemented as a web server application. This system allows users to write annotations —marks on text, memos on text and images, and freehand drawings on images— on web-based materials and share them with lecturers and learners. Moreover, we propose a method to follow the proper positions of annotations, after the materials have been updated, using natural language processing techniques.

The present study examines a method to follow proper positions of annotations on updated materials, and describes the evaluation of the method according to experimental results.

2 Overview of Writable Web

Writable Web has the following features:

- Works on common web browsers and does not require the installation of any special software to run.
- Offers writing marks and memos on texts (Fig. 1) and freehand drawings and memos on figures (Fig. 2).
- Annotations are in proper positions after materials are updated.
- Supports online asynchronous discussion with shared annotations.

The architecture of the *Writable Web* system is illustrated in Fig. 3. The system works as a server-side web application between the web browser of a user and web servers that provide web-based learning materials. *Writable Web* works on commonly used web browsers, and the annotations that are written by each user are stored in the same database on the server, allowing users to share their annotations.

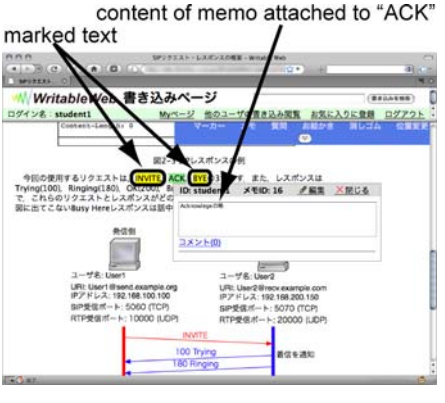

Fig. 1. Marks and memo on text

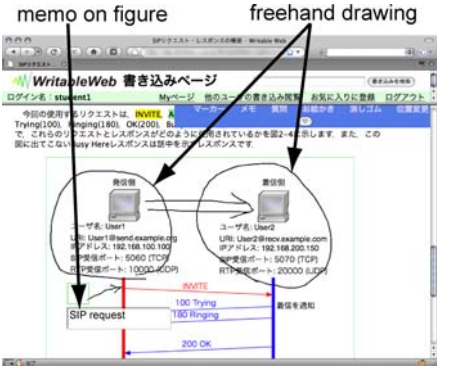

Fig. 2. Freehand drawings and memo on figure

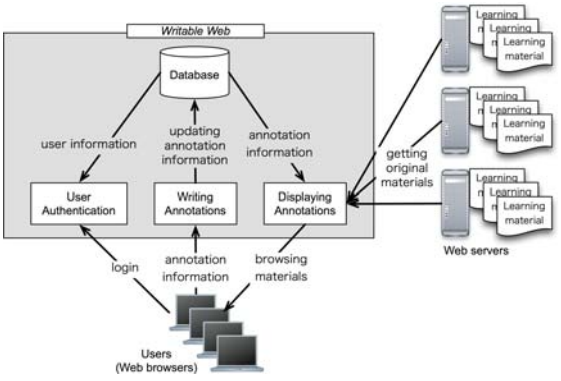

Fig. 3. Architecture of *Writable Web*

When a user writes an annotation on a page, *Writable Web* stores information which specifies the position, the type, the owner and the content of the annotation in a database. When users browse that page, the system inserts HTML tags to draw annotations into the original material only on memory, and does not make any changes to the original material itself. Moreover, the system stores the HTML file of the original material as a cache, and uses the cache, instead of the original material, when a user browses the same material, and will use the cache to detect for any updates.

3 Following Method of Annotation Positions on Updated Contents

Lecturers sometimes update contents to refine their materials, and annotations written before updating are unintentionally misplaced. Then, some paragraphs, sentences or letters are added, deleted or moved by the lecturer, and the positions of annotated texts or words are changed. *Writable Web* stores only the physical position of annotations, which consists of the start and the end of annotations, such as the paragraph number on a web page and the letter number in a paragraph.

In the present study, we propose a method to estimate the proper position of paragraphs and letters that have moved following the update of web content. *Writable Web* detects the updates and automatically revises stored positions of annotations to the proper positions using this method.

The proposed method estimates the proper positions of annotations using information within the text as mentioned below.

3.1 Information to Estimate Proper Position

Information to estimate the proper position of an annotation consists of following elements in addition to the physical position of the annotation (L). N_b and N_a will be empty if there is no noun before and after an annotation in a paragraph.

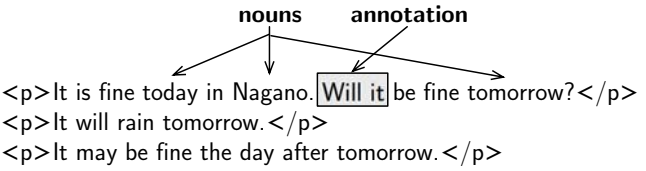

Fig. 4. Example of information from annotation

$$\begin{cases} S & : \text{an annotated string} \\ N_b & : \text{the sequence of nouns before the annotation in a paragraph} \\ N_a & : \text{the sequence of nouns after the annotation in a paragraph.} \end{cases}$$

An example of the information from an annotation is shown in Fig. 4. In this example, each element of information is as follows:

$$\begin{cases} L & = \{\{1, 29\}, \{1, 35\}\} \\ & \quad (\{\{\text{start paragraph, start letter}\}, \{\text{end paragraph, end letter}\}\}) \\ S & = \text{"Will it"} \\ N_b & = \{\text{"today"}, \text{"Nagano"}\} \\ N_a & = \{\text{"tomorrow"}\}. \end{cases}$$

3.2 Estimating Proper Paragraph

The method estimates the proper paragraph in updated material as the first step in estimating the proper position of an annotation. Let $p_1, p_2, \cdots, p_l$ denote the string of each paragraph, where the updated material has l paragraphs. Let P' denote the paragraph most similar to the annotated paragraph before the update (P) as the proper paragraph. If the degree of similarity is lower than threshold, the method discontinues estimating the proper position.

The degree of similarity between paragraphs used in this method is proposed by Odaka et al. [5], as shown in formula (1), and shows the frequency of appearance of n-grams in two documents from zero (low similarity) to one (high similarity), with n as three in this method. The degree indicates high similarities against changing words, the ending of sentences, and the order of phrases; and Odaka et al. use it to measure the similarities between reports by students.

$$R_j = 1 - \frac{1}{k} \sum_{i=1}^{k} \left\{ \frac{P(X_i) - p_j(X_i)}{P(X_i) + p_j(X_i)} \right\}^2 ,$$

where

$$\begin{aligned} R_j &: \text{the degree of similarity of } P \text{ and } p_j (1 \leq j \leq l) \\ k &: \text{the total number of } n\text{-grams in } P \text{ and } p_j \\ X &: \text{the sequence of } n\text{-grams} \\ P(X_i) &: \text{the frequency of appearance of } X_i \text{ in } P \\ p_j(X_i) &: \text{the frequency of appearance of } X_i \text{ in } p_j. \end{aligned} \tag{1}$$

3.3 Estimating Proper Position in a Paragraph

This method estimates proper position of an annotation in the estimated paragraph (P') as follows:

Step 1. Find the same strings as annotated string (S) from P'. Let $s_1, s_2, \cdots, s_m$ denote the same strings as S in P'.

Step 2. Compare nouns before and after each found string ($s_1, s_2, \cdots, s_m$) with nouns before and after annotated string (N_b and N_a).
Let m_{bj} denote the number of nouns before s_j in the paragraph which matches nouns in N_b, and let m_{aj} denote the number of nouns after s_j in the paragraph which matches nouns in N_a. Let N denote the greater number of nouns before and after S or s_j in the paragraph.
Then, the degree of match (E_j) is defined as shown in formula (2) and has a value between 0 and 1.

Step 3. Decide the position of the string that has the greatest degree of match, as the estimated proper position (L').

$$E_j = 1 - \frac{N - (m_{bj} + m_{aj})}{N}, (1 \leq j \leq m). \tag{2}$$

If there is no string identical to S in P' in Step 1, or the value of the greatest degree of match is lower than threshold, there is a possibility that the annotated string has changed. As a result, the position of a substring, which has the greatest degree of match and the greatest degree of similarity, in the estimated paragraph (P'), is selected as the estimated proper position (L'). However, this method discontinues estimating the proper position when the degree of match or the degree of similarity is lower than threshold.

3.4 Thresholds

This method discontinues estimating the proper position depending on the degree of match or the degree of similarity, as mentioned in sections 3.2 and 3.3. At this point, users of *Writable Web* should manually move annotations to the proper position, and the system offers some functions to support this procedure. We think it is better to stop estimations rather than report incorrect estimations, so the system detects and notifies users of annotations that should be moved manually.

Using a low threshold, this method halts the estimation less frequently, but it incorrectly reports the estimation of annotation positions more frequently.

4 Evaluation

We conducted two experiments to evaluate the method to estimate the proper positions of annotations. The first experiment was conducted with annotations written by learners, and the second experiment was conducted with annotations which were randomly generated according to a trend in the annotations of learners. We confirmed the adequacy of the trend from the annotations of learners prior to beginning the second experiment.

Table 1. Result of following updates with annotations from learners

Estimation result	Number of annotations	Percentage
Proper	358	95.7%
Discontinued	13	3.4%
Incorrect	3	0.9%

Table 2. Result of exploratory experiment

Estimation result	Number of annotations	Percentage
Proper	381	93.2%
Discontinued	25	6.1%
Incorrect	3	0.7%

Table 3. Result of following updates with generated annotations

Estimation result	Number of annotations	Percentage
Proper	1285	61.0%
Discontinued	789	37.5%
Incorrect	32	1.5%

4.1 Following Updates with Annotations from Learners

This experiment was conducted to confirm the precision of the method with annotations written by learners, six undergraduate students and 19 master's students. They learned and wrote 1008 annotations on 11 pages of material using *Writable Web*. The material had been updated, including 14 revisions in one paragraph and the addition of five paragraphs; however, the learners were instructed to use the material prior to the updates, and were unaware of any changes in the material.

The results of this experiment are shown in Table 1. There were 374 annotations, 37.1% of all annotations, which were needed to follow updates. The types of estimation results are estimating the proper position, discontinuing the estimation, and estimating an incorrect position.

These results indicate that the precision of estimation using this method is very high. The high precision may be the result of relatively few updates in the material and updates that did not have an effect on other parts of the page.

4.2 Following Updates with Randomly Generated Annotations

We conducted the second experiment to increase the samples of updates. To increase the number of annotations, we planned to randomly generate annotations according to a trend in the annotations of learners, the morpheme number distribution. The distribution of 1008 annotations collected in the first experiment is shown in Fig. 5.

Exploratory Experiment. We conducted an exploratory experiment to confirm the adequacy of using this trend to generate annotations. Using the material from the first experiment, 1008 annotations were generated, starting from randomized morphemes, and the lengths of these annotations were decided according to the distribution.

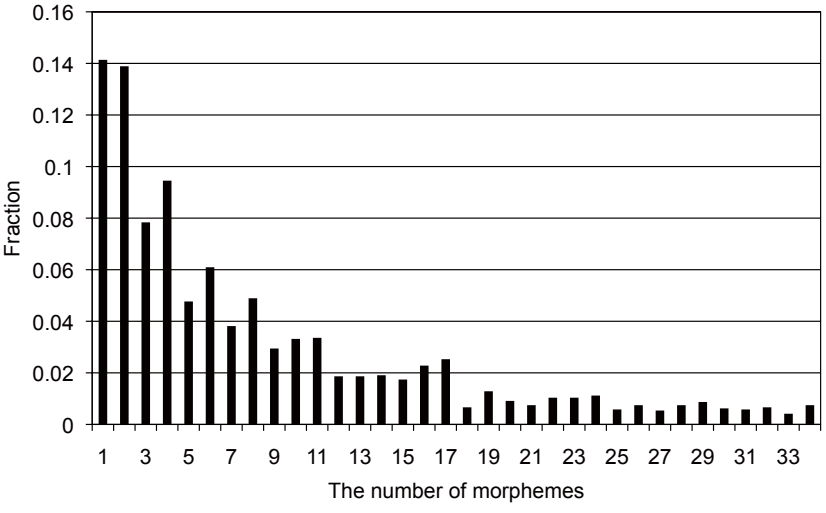

Fig. 5. Morpheme number distribution of learners' annotations

The results of the exploratory experiment are shown in Table 2. There were 409 annotations, 40.6% of all annotations, which were needed to follow updates.

There was a small difference in the result of discontinuing the estimation. The reason for the difference was that the number of annotations straddling multiple paragraphs was larger than in annotations made by learners, with some paragraphs being added between paragraphs, and the string in the annotation position changing. There were 198 annotations which straddled multiple paragraphs in the generated annotations, and there were 51 in annotations made by learners. However, the difference between the result of the first experiment and the exploratory experiment was small, so we used the generated annotations in place of the annotations from learners.

Experiment. We collected 30 materials before and after the update, and 3000 annotations were generated on pages before the update (100 annotations for each page). The results of this experiment are shown in Table 3. There was 2106 annotations, 70.2% of all annotations, which were needed to follow updates.

These results indicate that the number of annotations on which estimation was stopped was larger than in both the first and the exploratory experiments; however, the number of annotations that were misplaced was very low.

This experiment was conducted with a large number of annotations, which were randomly generated according to a trend from annotations made by actual learners; although, the trend used in this experiment was simple, and there was a difference in properties between annotations from learners and the generated annotations, as mentioned in the section 4.2. Confirming the precision of the

estimation method is also very important when providing this method to actual learners.

Moreover, *Writable Web* offers functions to support manually moving the annotations to the proper positions, so we believe that there are no practical issues using this method.

During this experiment, we also checked the time to estimate the proper position of 100 annotations on each page. The mean and standard deviation of processing time were 119.2 seconds and 102.0, and the shortest and the longest processing times were 18 seconds and 557 seconds, respectively. On the page that took the proposed method 557 seconds (the longest processing time) to estimate, there were long annotations with more than 10 morphemes.

We believe the proposed method is very simple, and that we can improve the proposed method to estimate the proper position more quickly.

5 Conclusions

The present study proposed a following method of annotations on updated contents and described the result of experiments to evaluate the method. We found that the method can estimate the proper positions of annotations on contents having small updates, and confirmed that the method avoids moving annotations to incorrect positions by discontinuing the estimation.

We plan on improving the method by increasing the precision, decreasing the processing time of the method, and by tuning the thresholds. Moreover, we would like to evaluate the precision of the improved method by providing it to actual learners.

Acknowledgement

This research was partially supported by the Ministry of Education, Culture, Sports, Science and Technology, Grant-in-Aid for Young Scientists (B), 18700642, 2006–2008 and 21700808, 2009–2011.

References

1. Cadis, J., Gupta, A., Gruding, J.: Using Web Annotations for Asynchronous Collaboration Around Documents. In: Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work, pp. 309–318 (2000)
2. Koivunen, M.: Annotea and Semantic Web Supported Collaboration. In: Proceeding of UserSWeb, 1st Workshop on End User Aspects of the Semantic Web, vol. 137(1), pp. 5–17 (2005)
3. Kunimune, H., Yokoyama, K., Niimura, M., Fuwa, Y.: An Annotation Sharing System for e-Learning Materials. In: Proceedings of ED-MEDIA 2007, pp. 3366–3371. Association for the Advancement of Computing in Education, Norfolk (2007)

4. Kunimune, H., Yokoyama, K., Takizawa, T., Hiramatsu, T., Fuwa, Y.: A Web-based Asynchronous Discussion System and Its Evaluation. In: Lovrek, I., Howlett, R.J., Jain, L.C. (eds.) KES 2008, Part III. LNCS, vol. 5179, pp. 507–514. Springer, Heidelberg (2008)
5. Odaka, T., Murata, T., Gao, J., Suwa, I., Shirai, H., Takahashi, I., Kuroiwa, J., Ogura, H.: A Proposal on Student Report Scoring System Using N-gram Text Analysis Method. The IEICE Transaction on Information and Systems J86-D-I(9), pt. 1, 702–705 (2003)
6. O'Hara, K., Sellen, A.: A comparison of Reading Paper and On-line Documents. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 335–342 (1997)