

Online Annotation System and Student Clustering Platform

Miao-Han Chang

School of Computing and Information Systems
Athabasca University
Edmonton, Canada
chang.miaohan@gmail.com

Rita Kuo

Academic and Industrial Research Centre
Knowledge Square, Inc.
Buffalo, United States
rita.mcsl@gmail.com

Maiga Chang

School of Computing and
Information Systems
Athabasca University
Edmonton, Canada
maiga@ms2.hinet.net

Kinshuk

School of Computing and
Information Systems
Athabasca University
Edmonton, Canada
kinshuk@athabascau.ca

Hsu-Yang Kung

Department of Management
Information System
National Pingtung University of
Science and Technology
Pingtung, Taiwan
kung@mail.npust.edu.tw

Abstract—When students read a book, they usually annotate important words on the text. Students may miss some important information while reading the book. If someone can remind them the potential important annotations, students might get better academic performance later in terms of having better marks for written exams, quizzes, and other learning activities. This research implements an online platform which can help students annotate reading activities with different ways. The online platform can also give students annotation suggestions according their annotation behaviours. On the other hand, teachers can use this platform to review students' annotations in order to discover students' potential learning problems. The platform can also help teachers cluster students into different groups according to their annotation behaviours. Besides explaining and demonstrating the platform with examples, this research also designs an experiment plan for evaluating the system in the future.

Keywords—Annotation; Chromosome; Patterns; Clustering

I. INTRODUCTION

In traditional learning, teachers always give students reading assignments that ask students to read pieces of articles on papers or in a book. Students usually take notes and highlight important words/sentences while reading the articles. Nowadays, digitalized materials are common to be used and students may use digital devices like computers and tablets for their studying.

Students have different annotation preferences while reading. They may annotate words in different ways (e.g., underlining, highlighting, or double-underlining). For example, when Jack, John, and Berry are reading a text – "Every year in the U.S. factories release over 3 million tons of toxic chemicals into the land, air and water" – in the "Pollution" article [7], their annotations might be different. Jack only circles the word

"air"; John underlines the whole sentence; and, Berry highlights the three words – "air", "water", and "land".

Students might be intended to not annotate some important keywords or simply by accident while reading an article. When they review the annotated article for preparing exams and doing homework, they may also skip those un-annotated words due to they believe that all important words have been well annotated earlier. For example, Jack may forget to mention the toxic chemicals released to land and water while answering a question of "Environment Pollution" in the mid-term exam because he only annotated "air" and didn't annotate "water" and "land" earlier. The incomplete answer for the mid-term question may make him lose marks.

To avoid of missing any important thing, students always borrow friends' textbooks and notes and take their friends' annotations as reference while preparing forthcoming written exam and quiz. In previous example, Jack may borrow Berry's annotated article before the exam. If he does so for exam preparation, he would probably notice the two missing annotated words, "water" and "land", on his copy of the article and would probably have better answer for the question later when write the exam.

If an annotation system can provide students annotation recommendation, they might catch the missing annotations up easier and fully understand the concepts and knowledge that they are supposed to be learnt. This research aims to design an online platform which is capable of suggesting students the annotations made from other students who have similar annotation behaviours; teachers can used the platform to review student annotations and modify clustering results. Section 2 introduces the existing computer based annotation clustering approaches. The problems of implementing the system are stated in Section 3. Section 4 describes the system workflow and explains its functionality with examples. Section

0	0	0	1	1	1	0	0	0	a	U+0038	8
0	0	1	0	0	0	0	0	0	a	U+0040	@
0	0	1	0	0	0	1	0	0	a	U+0044	D

With UTF-8 character's help in terms of representing different annotation ways applied to a character, an UTF-8 string, "000@@@DDD00088888", can be used to present the annotations of a sentence in a text; for instance, a student annotated three words in the pollution article as shows in Fig 1. The "water" has five characters and has been underlined, bold, and italic by a student; the proposed approach can use five "8" to represent the three different ways that the student annotates the "water" word.

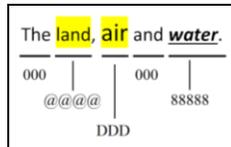


Fig 1. An example of an annotated text represented by UTF-8 characters

B. How to give students feedback according to their annotation behaviours?

After the system gathers students' annotations with UTF-8 string chromosomes, how to give students potential annotations that they might need is the next question. When students have similar annotations, they may have similar thoughts or understandings on an article. If a student didn't notice some important information in his/her reading, showing the annotations from other students who have similar annotations may remind him/her that he/she has missed the important information. As the example mentioned in the beginning of Section 1, Jack and Berry have focused on where the toxic chemicals are released into, but John has focused on how many tons of toxic chemicals are released into the environment. When Jack needs annotation suggestions, Berry's annotation is the better one because Jack and Berry have similar annotations and they may have similar knowledge base or thought on the "Pollution" article.

This research applies the clustering approaches proposed by Ying and colleagues [1][6] to separate students into groups. Students in the same group have similar annotations; therefore, when a student needs annotation suggestions, the system will identify which group the student belongs to and compare his/her annotations to the annotations that other students in the same group have to find out which annotations him/her doesn't take while others have. In previous example, Jack and Berry are classified into the same group because they have similar annotations. The system may suggest Jack to annotate "water" and "land" based on Berry's annotations. After students get the annotation suggestions from the system, they can decide whether or not the suggestion is useful for them.

C. How to help teachers review students' annotations and find the students' annotation behaviour features?

Teachers may also want to review students' annotations after students finished their reading activities in order to discover students' learning problems. Hence, the system

designed in this research needs to allow teachers to review students' annotations. For example, a teacher, Emma, wants to know why her students, Amy and Mark, couldn't catch up the main point of the reading activity, Pollution. After reviewing their annotations on the article, Emma finds out that Amy and Mark didn't annotate keywords about where the toxic chemicals are released. Emma can ask Amy and Mark to review the article again and this time try to focus on the three ways of how the toxic chemicals may release.

If the teacher has many students in a class and wants to discover students learning problems, he/she may spend a lot of time comparing students' annotations one by one. Therefore, the system needs to provide the teacher the function of clustering students by their annotations. The teacher then can review students' annotations in the same group to find out students' learning problems. Moreover, teachers can take notes for each cluster to describe the annotation behaviour features that students belong to the cluster have so they can be easily reminded about students' learning conditions. For example, Emma has 50 students in the Environmental Science course. She decides to cluster students into four groups in order to determine the annotation behaviour features that the students in the same group have. She may discover students' learning problems by checking the features of the groups one by one and make her work easier.

If teachers do not agree with the clustering results generated by the system, teachers can also modify the clustering results according to their observations. For example, Emma has reviewed the clusters generated by the system and found the students in one of the clusters are preferred to annotate articles in keywords basis, so she decides to take a note - "keywords" - on this cluster. However, she finds that one of the students in the cluster, John, different from others is preferred to annotate articles in sentences basis. In this case, Emma can drag-and-drop John from this cluster to another cluster in which most of students annotated text in sentences basis.

D. How to improve the accuracy of the clustering results?

Researchers in the research area of data clustering can use the proposed platform to analyze the accuracy of different clustering approaches. Researchers can have both of the system clustering results and the teacher-modified clustering results easily and compare the data to measure the precision and the recall of their proposed clustering approaches.

After getting the approach's precision and recall values, researchers can use teachers' modified clustering results as benchmark to find out which students are classified into incorrect cluster. To determine why the samples are incorrectly classified, the researchers can review students' annotations and find the differences. Researchers can revise the approach accordingly to enhance its accuracy.

Teachers' notes on the clusters can also help researchers refine the clustering approach. The researchers can understand why teachers thought these students should be classified into the same group based on the notes. The researchers can then realize teachers' classification criteria accordingly.

IV. THE SYSTEM

This research designs the Online Annotation System which provides students a platform to read articles and take annotations. Students can see annotation suggestions from the system based on the results of the four clustering approaches mentioned in Section II. Teachers can use the Student Clustering Platform to review students' annotations and modify the clustering results. Fig. 2 shows the workflow of teachers and students use the system.

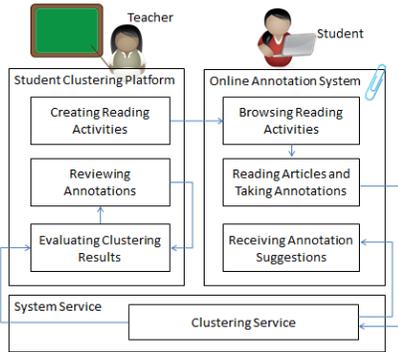


Fig. 2. The system workflow of the Online Annotation System and the Student Clustering Platform.

In the beginning, teachers can choose a course which they would like to manage the reading activities. For example, Emma, a teacher who teaches Java, plans to add a new reading activity about BlueJ topic in her Java course. She can use the "Creating Reading Activities" function to set up the activity name, start and end dates of the activity, and text-based reading material. After Emma creates the reading activity, students can see it in the Online Annotation System. Berry, one of the students in Emma's Java class, finds out that she has a new reading activity. She can click the "Reading" link next to the reading activity to start her online reading.

In the Reading interface (as Fig. 3 shows), students can annotate the text in many ways, such as highlight, underline, bold face, italic, and enlarge the font size. They can also use multiple ways on the same passage. For example, the words – "of personal lifestyle and travel too" – in area (B) of Fig. 3 have been enlarged and made bold by Berry. On the other hand, when students don't want to see their annotations, they can check the "Hiding your annotation" checkbox in area (A) of Fig. 3, and they will find that their annotations are no longer seeable.

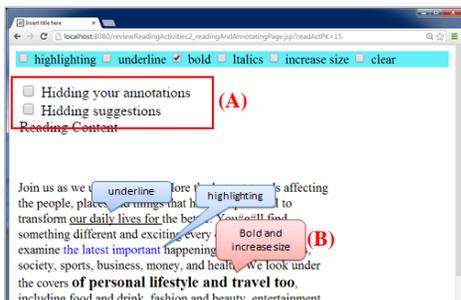


Fig. 3. Students can decide which ways they would like to annotate the text by checking the boxes on the top of the screen..

When students read an article, they may want to know if they have missed anything that are important. The system gives students annotation suggestions for a specific paragraph when they finish to annotate the last sentence of the paragraph so they have fresh memory on what they have done so far. After students see the annotation suggestions, they can decide whether the suggestions are useful for them or not. If students don't want to see suggestions, they can check the "Hiding suggestions" checkbox in area (A) of Fig. 3, and they will see no annotation suggestion on the screen.

After students finished their reading activity, teachers may want to review students' annotations before teachers start to modify the clusters based on their opinions. The system provides teachers the Reviewing Annotations function which help them review all students' annotations in a particular reading activity at the same time. If teachers are interested in particular students' annotations, they can select the students and click the "Chosen Review" button. For example, after Emma's students finish the "BlueJ" reading activity, Student #10, #28, #37, and #43 are classified into the same cluster. Emma wants to know why these students are in the same group and uses the Reviewing Annotations function to review students' annotations. She finds out that those students taking notes mostly on BlueJ's history; therefore has clear idea of the students in this cluster are more interested in BlueJ's history.

After teachers review students' annotations, teachers can use the Evaluating Clustering Results shown in Fig. 2 to move students from one cluster to another and to identify the annotation features of the clusters. Teachers have two functions to use in the Evaluating Clustering Results stage. The two functions are Modifying Clustering Results and Denoting Cluster Features. Teachers can use Modifying Clustering Results function to review and modify the clustering results generated by the four approaches mentioned in Section II. For example, after five students have finished the reading activity, Emma decides to separate students into three groups. Emma can choose which clustering approach she would like to use by clicking one of the clustering approach buttons on the top of Fig. 4 as area (A) shows. The clustering results are shown in area (B) and the five students are divided into three clusters: Student #1 and Student #5 are in the first cluster, Student #2 is in the second cluster, and the last two students are in the third cluster.

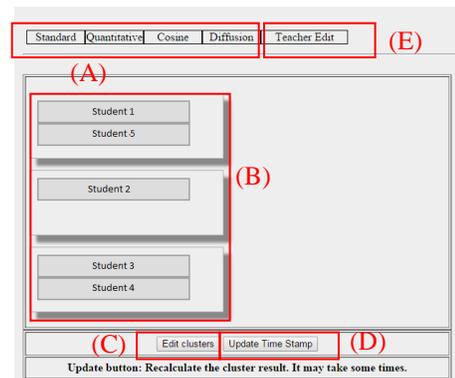


Fig. 4. Teachers can see and revise the clustering results generated by the system.

If teachers do not agree with the clustering results, they can modify the results by clicking “Edit Clusters” button shown in area (C) of Fig. 4. They can either move a student from one cluster to another or create a new cluster. For example, Emma finds out that Student #1’s annotation feature is different from all other students, so she clicks the “Add a New Cluster” in area (A) of Fig. 5 to create a new cluster in area (C) and drag Student #1 from the first cluster to the new cluster.

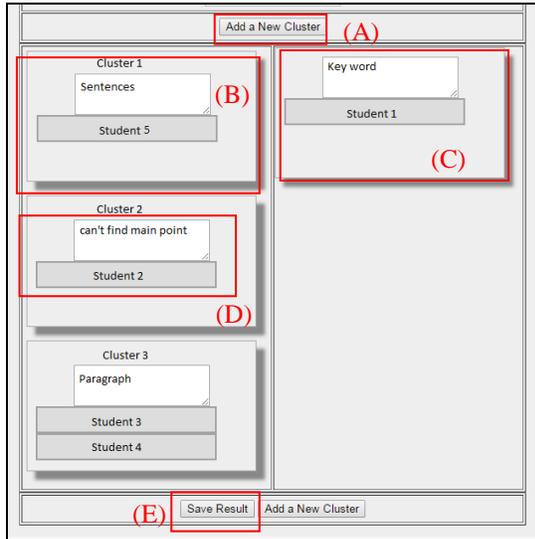


Fig. 5. Teachers can edit the clustering results by drag and drop students from one cluster to another.

Teachers can also take notes for students’ annotation features. For example when Emma finds that students in Cluster 1 prefers to annotate articles in sentences, she puts “Sentences” in the textbox of Cluster 1 to denote the annotation feature of this cluster. After she satisfies with the revised clustering results, she clicks the “Save Result” button shown in area (E) of Fig. 5. Teachers can review the revised results at anytime by clicking the “Teacher Edit” button shown in area (E) of Fig. 4.

V. EXPERIMENT PLAN

This research will have three classes of students to participate the experiment. One of the classes will be the Control Group, and the other two are Experiment Group 1 and Experiment Group 2. The research plans to have a quasi-experiment. The experiment uses a mixed methodology to verify three hypothesis. Researchers will analyze the data with quantitative research methodology and use qualitative research methodology to analyze the teacher’s words in the interview. The experiment has four stages.

Stage 1. Collecting Students’ Knowledge Baseline

In the beginning of the experiment, the teacher will give all students a pre-test to understand their prior knowledge of the learning activities in the course.

Stage 2. Reading Article without Annotation Suggestions

At this stage, the teacher will assign four reading activities to all students. Students in Control Group are allowed to read (and to annotate, if they want) in their preferred ways

(e.g., print the articles out on a paper or use any devices such as smart phone/tablet/laptop to read the articles). Students in the two experiment groups will use the Online Annotation System to read (and to annotate, if they want) the articles. At this stage, students in both experiment groups will not receive annotation suggestion. After the students finish a reading activity, they will take a quiz (i.e., post-test) for the evaluation of what they have learnt from the reading activity. At the end of this stage, the researchers will ask the teacher to review the clustering results generated by the system and modify the results according to his/her thoughts. Teacher’s modified clustering results can be used for evaluating the accuracy of the clustering approaches used in the system. After researchers compare the performance of the four approaches, the most efficient one that has higher accuracy will be chosen for clustering students at Stage 3.

Stage 3. Reading Articles with Annotation Suggestions

The teacher will ask students to finish five reading activities at this stage. Unlike Stage 2, students in Control Group will use the Online Annotation System without annotation suggestion. On the other hand, the two experiment groups will receive annotation suggestions while using the system at different timings. The students in Experiment Group 1 will receive annotation suggestions *while* they are reading; the students in Experiment Group 2 will receive the annotations suggestions *after* they finish the reading activity. All students still need to take a quiz after they finish a reading activity at this stage.

Stage 4. Evaluating the Proposed Methods

After students finish all reading activities, the researchers will ask students to complete two questionnaires: a Usability Scale questionnaire and a TAM questionnaire. The collected data will be used to assess the usability of the proposed system and students’ acceptance degree towards the use of the system. The researchers will also have an interview with the teacher. The purpose of the interview is to understand whether or not the teacher can find out students’ potential learning problems via the reviews of students’ annotations.

With the Online Annotation System and Student Clustering Platform, this research has three hypotheses toward the system and the platform.

H1: The Online Annotation System can help students learn better.

Researchers can verify this hypothesis by comparing students’ performances of the quizzes which are answered after students finished a reading article. At Stage 2 of the experiment, the control group students will do their reading activities in their preferred ways and the experiment groups’ students will use the proposed system to do their reading activities. At the end of this stage, Researchers will use t-test to evaluate if there is any significant difference in terms of students’ performances of the answering quizzes between the control group and the experiment groups. Researchers expect to see the average performance of

experiment group students is significantly better than control group students.

Furthermore, the researchers will use t-test to evaluate the quiz performances of students between Stage 2 and Stage 3 to understand if the annotation suggestions can help students get better performance. Researchers expect to see the average performance of students who receive annotation suggestions is significantly better than those who don't. Researchers also expect to see students in Experiment Group 1 perform better than those in Experiment Group 2 and students in Experiment Group 1 have more positive perceptions toward the usability of the annotation suggestions.

H2: Students' perceptions toward the Online Annotations System is positive.

For verifying this hypothesis, the research team will ask students to fill up a revised Usability Scale developed by Lu and colleagues [3] after students finish all reading activities. The research team will also ask students to complete a revised Technology Acceptance Model (TAM) questionnaire developed by Davis [2] in the end of the experiment. The research team will use t-test to evaluate students' perceptions toward the system and use structural equation model to analyze which factor affects how students think about this system. Researchers expect to see students have high acceptance toward this system and think this system is easy to use and very helpful for them.

H3: Teachers can find learning problems that students may have with the Student Clustering Platform.

For verifying this hypothesis, at the end of Stage 2, the research team will ask teachers to review the clustering results generated by the system and modify the results according to their thoughts. The research team assumes that teachers could discover students' learning problems when they review students' annotations. For example, teachers may get clear idea of which parts of the text (or learning units) that most of students don't understand and have misconceptions via reviewing students' annotations. At the end of the experiment, the research team will interview the teachers to see if they can find their students' potential learning problems via the use of the system. Researchers will analyze teacher's response in the interviews with qualitative data analysis methodology.

VI. CONCLUSION

This research has implemented two systems for students doing reading activities and teachers reviewing students' annotations and finding students' potential learning problems online. The first system is the Online Annotation System which allows students to take annotations on an article in different ways and give them annotation suggestions based on their

previous annotations and clustering results. The other system is the Students Clustering Platform which allows teachers to review students' annotations and clustering students into groups according to their annotations automatically and manually if they are not satisfied with the results clustered by the system. This paper also explains the system's functions from both of student's and teacher's viewpoint. In the end, the paper describes an experiment plan that the research team is going to use for verifying the usefulness of the systems.

The proposed systems have several potential benefits for students, teachers, and researchers. Students can use the Online Annotation System to take annotations on the articles easily; they may also have better academic achievement by receiving useful annotation suggestions from the system. Teachers can identify students' learning problems via reviewing students' annotations and get clear idea of which parts of the text (or learning units) that most of students don't understand and have misconceptions. Last but not least, researchers can use teacher-modified clustering results to measure the accuracy of the clustering approaches. They can also use the cluster characteristics noted by teachers to determine the annotation features of clusters for improving the proposed clustering approaches.

The experiment is planned to be hosted at one of public schools in Canada. The research team will find three classes which have the same teacher in History course in a secondary school to do the experiment for three months. After the experiment, the research team will gather the feedback from students and the teacher to evaluate systems' usability and users' acceptance. The research team will also revise the clustering approaches in order to enhance the accuracy of student clustering according to teachers' revised clustering results.

REFERENCES

- [1] M. Chang, R. Kuo, K. Ying, A. F. Chiarella, J. Heh, and Kinshuk, "Clustering Annotations of a Digital Text with Bio-inspired Approaches," *Hybrid Technology*, vol. 1, no. 1, pp. 1-10, 2013.
- [2] F. D. Davis, "Perceived usefulness, perceived ease of use, and user acceptance of information technology," *MIS Quarterly*, vol. 13, no. 3, pp. 319-340, September 1989.
- [3] C. Lu, M. Chang, Kinshuk, E. Huang, and C. W. Chen, "Usability of Context-Aware Mobile Educational Game," *Knowledge Management & E-Learning*, vol. 3, no. 3, pp. 448-477, 2011.
- [4] A. Y. S. Su, S. J. H. Yang, W. Hwang, and J. Zhang, "A Web 2.0-based collaborative annotation system for enhancing knowledge sharing in collaborative learning environments," *Computers & Education*, vol. 55, no. 2, pp. 752-766, September 2010.
- [5] S. J. H. Yang, J. Zhang, A. Y. S. Su, and J. J.P. Tsai, "A collaborative multimedia annotation tool for enhancing knowledge sharing in CSCL," *Interactive Learning Environments*, vol. 19, no. 1, pp. 45-62, January 2011.
- [6] K. Ying, M. Chang, A. F. Chiarella, Kinshuk, and J. Heh, "Clustering Students based on Their Annotations of a Digital Text," In the *Proceedings of 4th IEEE International Conference on Technology for Education*, Andhra Pradesh, India, pp. 20-25, July 18-20, 2012.
- [7] "Pollution" (n.d.), *Pollusion*. Retrieved March 1, 2015 from <http://webpage.pace.edu/jb44525n/page5.html>