ATHABASCA UNIVERSITY

BIO-INSPIRED CLUSTERING APPROACH BASED ON STUDENTS' ANNOTATION ON ONLINE READING MATERIALS

BY

MIAO-HAN CHANG

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE in INFORMATION SYSTEMS

FACULTY OF SCIENCE AND TECHNOLOGY

ATHABASCA, ALBERTA

MARCH, 2019

Approval of Thesis

The undersigned certify that they have read the thesis entitled

BIO-INSPIRED CLUSTERING APPROACH BASED ON STUDENTS' ANNOTATION ON ONLINE READING MATERIALS

Submitted by

Miao-Han Chang

In partial fulfillment of the requirements for the degree of

Master of Science in Information Systems

The thesis examination committee certifies that the thesis
and the oral examination is approved

Supervisor:
Dr. Maiga Chang
Athabasca University

Committee Members:
Dr. Vive Kumar
Athabasca University

Dr. Rita Kuo
New Mexico Tech

Dr. Hsu-Yang Kung
National Pingtung University of Science and Technology

External Examiner:
Dr. Stian Håklev
Ecole Polytechnique Fédérale de Lausanne

February 11, 20

Abstract

When students read a book, they usually annotate important words in the text. Students may perhaps miss some important information while reading the book. Students can perform better academically and score well in written exams, quizzes, and other learning activities if they know about the potentially important annotations in advance. With an annotation system's help, teachers can create online reading activities for students and review students' annotations on the e-text. This research aims to design and implement a bio-inspired clustering method. The proposed method can cluster students into groups according to their annotations on the online reading materials. The annotation system uses the clustering results to find content that might be important for student. This research implements an Online Annotation System that can let students annotate on the reading activities in different ways and give students annotation reminders base on the clustering results.

Keywords — Annotation; Chromosome; Patterns; Clustering

Table of Contents

List of Tables

List of Figures

List of Code Segments

Chapter 1. Introduction

1.1 Motivation

About ten years ago, students still prefer to print out the materials than the digital version, especially in academic reading because they prefer to make highlights, underlines, and write some notes on the learning materials (Liu, 2005). However, with the development of technology, students are habituated to using digitalized materials in digital devices, such as computers and tablets, while reading instead of using printed papers (Lopatovska, Slater, Bronner, El Mimouni, Lange, & Ludas Orlofsky, 2014). In 2014, Chen and Chen's study shows that no matter students prefer using printed version or digital version reading materials, their reading attitudes have no difference (Chen & Chen, 2014).

With more students using digital devices to read learning materials, their behaviour on the digital device might be similar to printed papers. When students are reading learning materials assigned by teachers, they usually take notes and highlight important words/sentences (Hoff, Wehling, & Rothkugel, 2009; Chen & Chen, 2014). 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, 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 may intend 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 might skip the, un-annotated words as 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 did not annotate "water" and "land" earlier. The incomplete answer for the mid-term question may make him lose marks.

To avoid missing any important thing, students may borrow friends' textbooks and notes in order to take their friends' annotations as a reference while preparing the forthcoming written exam and quiz. In the 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 a better answer for the question later when writing the exam.

1.2 Goal and contributions

This research aims to propose an annotations recommendation system that can provide annotation suggestions to students to help them catch the missing content up easier. With the annotation recommendation system, students might be easier to review the learning materials and help them to achieve higher academic achievements. Therefore, the goal in this research is designing an algorithm which can classify students' annotations into groups based on their annotated behaviour. Students' annotation behaviour can mix up with the highlight words in different colors, underline words in different colors, and other different kind annotation types.

Therefore, the objectives of this research are: (1) construct a web-based annotation

tool for researchers, teachers, and students; (2) propose an encoding method to store and represent a student's annotations when they use different annotating ways on a text; (3) design an approach to compare students' annotations; (4) Outline an approach which classifies students into groups based on their annotation behaviour; this approach would spend less time and have better accuracy when compared to other approaches from the previous studies; (5) find common features of annotations made by students in the same group.

There will be four contributions once the research is complete. The first contribution will have a plug-in where the students can annotate article when reading online materials and receive annotation recommendation based on their annotations behaviours. This plug-in can let other users easily plug this tool into their browsers and use it

Provide students several different annotated ways will raise the represented text's complexity (as Research Objective 1 shows in Section 2.5). The second contribution is designing a method to decrease the complexity of students' text with annotations in order to raise the accuracy of the classifying students into proper groups and the acceptance of the annotation suggestions to the students.

Another contribution of this research is designing a grouping method to classify students' annotations based on their annotation behaviours. After decreasing the complexity of these represented texts, the represented texts still have annotated feature on it. The proposed method is able to group students' annotations with complex annotation types, including annotations highlight on the article in different colors, underline, bold ,face, and etc.

The last contribution is defining students' annotation behaviour features and their

potential learning problems in the classified groups. Defining the students' annotation features and potential learning problems could help teachers adjust their teaching direction and understand which concepts in the materials are those most students do not understand and teachers can emphasize those concepts when reviewing the materials in the class.

1.3 Thesis organization

Chapter 2 discusses the relevant works that many researchers have done in the annotation systems, clustering methods, and bio-inspired. I will summarize the research problems, objectives, and issues in this chapter. Chapter 3 describes the system workflow and explains its functionality with examples. Chapter 4 focuses on bio-inspired clustering algorithm design. Chapter 5 talks the research questions I want to verify in the experiment, lists corresponding hypotheses I have, designs the experiment and related questionnaires I can use to collect qualitative and quantitative data, analyzes the collected data in both quantitative and qualitative way, evaluates the effectiveness of the game and discuss the important findings. Chapter 6 summarizes of this research and the potential benefits that students, teachers, and researchers may gain from this research.

Chapter 2. Annotation systems and behaviour clustering approaches

This chapter review studies relevant to this research; the studies include the reading activities in digital materials, digital material annotations, behaviour grouping, and bio-inspired adopting. Research problems and possible solutions for solving these problems are summarized at the end of this chapter.

2.1 Reading activities in digital materials

Some studies show that people are getting used to reading on the screen. In 2005, Liu has asked participants to think about their reading habits in the past ten years. Eighty-Three percent of 11 reported that their electronically reading has increased. In Chrzastowski and Wiley's research (2015), participants can choose what they want to use digital books or get printed books when they are doing reading activities. When participants want the printed book, the research team will send the printed book to students; on the other hand, when participants prefer using digital books, they can read the materials online. The result shows that participants prefer using digital books more than asking research team sending printed books.

Bounie, Eang, Sirbu, and Waelbroeck (2013) have discovered that Amazon sells digital books more than printed books. Especially in higher education, digital materials are popular in undergraduate and graduate students (Tashman, & Edwards, 2011). Lopatovska and colleagues' (2014) research shows the reasons that people use digital books because (1) digital books are convenient (58%); (2) of study for school's need (55%); (3) there is no printed version available (49%); (4) it is easy to use (48%); (5) digital books costs less (46%); (6) of the ability to search text (43%); (7) of the

interactive features (36%); and (8) they just want to use digital books (80%).

Because people's reading preference has changed to digital devices, Tashman and Edwards' (2011) research focuses on analyzing people's reading behaviour when they reading digital materials. The participants in their study record reading diaries when they do Active Reading activities. There are approximately 25% of the diaries shows that participants use both paper and computer for finishing their Active Reading tasks. They also find that 63% of diaries in doing their reading activities are performed on computer only.

Above studies show that people's reading habits are changed by digital reading materials and feel comfortable to read on the screen. If people get used to annotating on the paper-based reading materials, they might also have the same habit when they reading on the digital devices. The next section discusses the research related to annotation behaviour on digital devices.

2.2 Annotation on digital material

In traditional learning, teachers always give students reading assignments that ask students to read pieces of articles on papers or in a book. Reading and annotating articles are students' routine job of study. Chen and Chen's (2014) research discovers that when students use paper-based way to study, students are frequently highlighting or underlining words, phrases, or passages, writing short comments in blank space, between lines, or near figures. Before an exam, students can find some annotations that may important to themselves by reviewing other students' annotations.

Nowadays, digitalized materials are commonlyused and students may use digital

devices like computers and tablets for their studying. Hoff, Wehling, and Rothkugel (2009) classify functions based on existing annotation systems into four categories: media formats support (e.g., support web document, office documents, PDF, or multimedia.), annotation functions (e.g., user can annotate on articles), interactions management (e.g., is this annotation private, group, or public share and do I want to get notifications when other authors make new annotations), and repository implementation (e.g., the repository is local, global, or client-server). The four categories help researchers understand the gaps between the annotation features that students need and is currently a widely accepted annotation systems we have.

Some other researches provide annotation service to help users read and annotate articles on their computers. Yang, Chen and Shao (2004) have developed a web-based annotation platform – Personal Annotation Management System (PAMS) – where users can highlight, underline, attach notes and voice recordings to the text in an article. Su and colleagues (2010) and Yang and colleagues (2011) improve the PAMS system as PAMS 2.0 to know students' perceptions toward the collaborative annotation system and how the collaborative annotation system helps students improve their reading competence.

Another research provides a system that combines annotation service and collaborative learning together. Pearson, Buchanan, and Thimbleby (2012) aim to provide students an annotation system to help students learn better. They provide a collaborative system – BuddyBooks – to students and ask students to read the article in groups. While students read articles, members in the same group have to stay at the same place and discuss. When students read and annotate on the article, the actions are sent to other group members' pad. Every member can see others' annotations in different highlighted

colors. If one student wants other members to look at a particular paragraph in the article, he or she can just point out the location on his or her pad and other members will receive a notification at the side bar and can easily follow. The result shows that students believe sharing members' annotations are useful and enjoy the feature while doing a reading activity.

In Pearson, Buchanan and Thimbleby's research, they allow students to review other students' annotations only in small groups. I want to provide students an annotation recommended function so they can receive annotation suggestions to find useful ones instead of reviewing a small group of classmates' annotations on their own. To find useful annotation suggestions for students, I need to analyze the relations between students' annotations and their annotation behaviours.

Above-mentioned research allows students to use different ways to annotate their reading materials. The common annotation ways are underline, highlight, and note-taking. Other functions provided by annotation systems include, for example, students can attach multimedia resources (such as audio and video) to a word or sentence and students can collaborate with others in a group (such as point out an annotation).

Melenhorst (2005) records students' annotation behaviours in the annotation tool to identify the relationship between reading phases and annotation ways (e.g. highlighting words and sentences, taking notes, copying passages to notepad, etc.). With showing the relations between reading phases and annotation ways on a two-dimensional plane, Melenhorst has found that students use different annotation ways in different reading phases.

Seeing some researchers develop systems for students to study, I still want to know

whether these systems can help student learning better. Yang and collage's (2011) research as well as Chen and Chen's (2014) research have proved that using annotation system can improve students' reading comprehension to the article. In Chen and Chen's (2014) research, researchers give students article and give students comprehension test after reading. Students in using annotation system group have better explicit comprehension than students use paper based to read the article. Students in both groups (using the annotation system and not using the annotation system) have no significant difference in inferential comprehension.

Students have different purposes to read articles. Some students may prefer reading in the free time; some students may be pushed to learn by their teachers or parents; some students read the articles because they don't understand the topic they learned in the class today, etc. No matter what the purpose is to push students to study, they want to get themselves mastery in the materials. I want to cluster students' annotations into groups based on their annotation behaviours and provide students annotation reminders based on the clustering results. If a student gets a reminder, it means an annotation is missed or ignored by the student while other students of the same group annotated the particular words, phrases or sentences. The research can help students see some points that they might miss before and help them learn better.

2.3 Behaviour grouping

In grouping methods, the basic method is manually grouping. In Tashman and Edwards' research (2011), the analysis the non-text annotations and organize it as a two dimension. The first dimension is inter-page and intra-page; another dimension is

generative semiotics and simple semiotics. Tashman and Edwards defined the inter-page annotation such as visible marks across pages, a dog-ear, or bookmarks; the intra-page such as underline; generative semiotics such as sketches; simple semiotics such as underline. The clustering method that Tashman and Edwards's use is giving a general idea about what kind of annotation marks that participant's use when they read.

If I can use a computer to help us do the grouping job, I can consider two grouping methods – classifying and clustering. In my research, I choose the clustering method for my research instead of classifying because after training the clustering system with the training set, the new data can be decided which groups it belongs to automatically. This means before analysis the collecting data, I need to define how many groups students will have and what students annotations features are. However, students may have different annotation behaviours in different reading activities; a fixed group number and the features can't fit all the reading activities. For example, student's annotation behaviour in one reading activity might be appropriate to separate into three groups, but might be suitable to cluster into four groups in another reading activity. Using a clustering method based on their annotation behaviour, different reading actives will have different group numbers and features.

Regarding the clustering methods, the K-means algorithm (Tan, Steinbach, & Kumar, 2006) and the Hierarchical clustering methods (Tan, Steinbach, & Kumar, 2006) are in the top list of the clustering methods of the Machine Intelligence. The hardest decision about using k-means clustering method is to decide the k value because it will affect the clustering quality. For example, if make the k value too big, the similar items could divide into different groups. If make the k value too small, maybe the differences

between groups won't be clear. Another problem will affect the clustering result is the standard processor of the randomly chosen k in the very beginning. Using the Hierarchical clustering method can generate a tree map. Based on the tree map, I can find which leaves have the minimum distance to another. However, no matter choosing k-means or the Hierarchical clustering methods, both methods need to decide what the proper number of clusters is.

Another confusing method is the K-nearest neighbors' algorithm (also called k-NN) (Chang, Kao, Chu, & Chiu, 2009). This method needs to give sample data and these data are labeled. For example, there are 10 samples in the sample data and these 10 samples are marked as group A, group B, and group C. When a new data comes, this algorithm will find the nearest samples and base on these samples to make the decision for this new data belongs to which group. For example, if the k value is 3 when the new data comes, the algorithm will find the nearest 3 neighbors and base on the which group gets the biggest proportion and the new data will be classed to that group. This algorithm needs the pre-defining dataset, in this research; the reading activities for students to read have a different focus and different length, it is impossible to define students' annotations' style before they make their annotations.

In this research, I want to cluster students' annotations automatically. For deciding how many groups for a different reading activity, I will ask the teacher to cluster students' annotations into many groups. I want to use teacher-edited results to find out how many groups is fitting for which kind of reading activity.

This research adopts three grouping methods mentioned in the Su and colleagues' (2010) research; the three techniques are Rocchio, Linear Least Squares Fit (LLSF), and

pseudo-LLSF (pLLSF). Rocchio classification method (Liu, Yu, & Meng, 2002, Manning, 2008), is given a set of documents; each document has their features, and some features are shared in different documents. When the trained Rocchio needs to classify the new document, the Rocchio method will calculate which feature of the exciting documents is closest to the new document with the centroid vector for each feature. The Rocchio classification won't suit for this research because Rocchio needs a dataset for training; the teacher might set up different article in the same reading activities in the same course in different semesters. Therefore, it is impossible to have trained dataset for the reading activities.

The LLSF (Linear Least Squares Fit) in Matrix method uses a mapping function to compute two matrixes into one matrix. For example, the first matrix is an M × N matrix and the second one is an N × P matrix. After computing this two matrix into one matrix, this matrix will become an M × P matrix. It is possible to learn empirical associations between two matrixes; the M × N matrix and N × P matrix can become M × P matrix also means the M and P have co-occurrences relations. If the dimension M is the annotation features and the dimension P is the learning problems, this associate matrix can map students' annotations features to what kind of learning problems students might have. This method needs the researcher to give training and categorizing data. (Yang and Chute, 1992; Yang and Chute, 1994). Sometimes the training data of the matrix is too big and has a lot of noise; it makes researchers want to remove these noise. Liu, Yu, and Meng's (2002) research are reducing the size of source items matrix from LLSF and developed pseudo-LLSF. However, it is hard to decide what sets of data are the noise in the matrix (Deerwester, Dumais, Furnas, Landauer, & Harshman, 1990).

Comparing with the three methods, the Rocchio method needs to compute the vectors of the documents it has; when Rocchio got a new document, Rocchio needs to re-compute the entire document set. In my research, students could make annotations and update their annotation at any time. Giving a training set to Rocchio is necessary and how the accuracy of the training set setting will affect the Rocchio's classifying results are another two reasons that I cannot use Rocchio in my research. If both my system and Linear Least Squares Fit are imple=55mented together in Matrix method, the system can map two matrixes values into one matrix – one is annotations verse features and the other is features verse learning problems. I may have a chance to use Linear Least Squares Fit in Matrix and pseudo-LLSF.

Another method that can be applicable in my system is Correspondence Analysis to associate two non-value parameters. Melenhorst (2005) has divided participants' reading time into ten phases and defined features of using the annotation tools first. In the next step, the researcher uses Correspondence Analysis to associate the phases and features to build two-dimension coordinates in order to discover the relation between the phases and students' annotations' behaviours. Melenhorst's research provides a method can give non-value items coordinates. Because annotation way, annotation feature, and learning problems are also non-value items, I may be able to adopt this method in my research.

Another way to compute the coordinate is using bit-string chromosomes proposed by Ying and colleagues (Ying, Chang, Chiarella, Kinshuk & Heh, 2012; Chang, Kuo, Ying, Chiarella, Heh, & Kinshuk, 2013). The research uses four different approaches to compute bit-string chromosomes to a coordinate. All approaches divide a chromosome

into two parts for mapping the chromosome to a two-dimensional plane and use hierarchical clustering to cluster users, but the methods of deciding the coordinate on the plane in each approach are different.

Standard, Quantitative, and Cosine approaches compare chromosome in pairs. The standard approach adds the exponential value assigned in the position of difference in the chromosome (e.g., the ith position's value is 2i). The quantitative approach also sums up the exponential value assigned in the difference in the chromosome, but the assigned exponential value is corresponding to the number of difference in the chromosome (e.g., the ith difference is assigned to 2i). Cosine approach takes the chromosome as a vector and measures the cosine value of two vectors. The cosine value represents how similar the two vectors are. Unlike the first three approaches, Diffusion approach compares all chromosomes at the same time by finding the positions that not all the chromosomes have the same value.

Now, I have three different ways can compute my parameters: (1) use Linear Least Squares Fit in Matrix to compute two matrixes into one matrix and then use pseudo-LLSF to reduce the data size; (2) use Correspondence Analysis to transform non-value parameters into coordinates; (3) design a way to compute the chromosomes to coordinates. I list a clustering methods table (as shown in Table 1). I list the authors' name and the method they use in their paper.

Table 1.

Clustering methods in this research

Authors	Clustering methods
(Chang, Kao, Chu, & Chiu, 2009).	Using k-NN classification combined with GA
(Su, & Yang, 2010).	Using LLSF to associate the user' annotation and article categories.

(Melenhorst, 2005).	Using correspondence analysis to analyze the relations between study time and annotations behaviour.
(Huang, Chen, & Guo, 2012).	Developing a radar plot for students can tracking their study status by themselves.
(Chang, Kuo, Ying, Chiarella, Heh, & Kinshuk, 2013).	Using Cosine and Diffusion clustering approach to cluster students' annotations.
(Ying, Chang, Chiarella, & Heh, 2012).	Using Standard and Quantitative clustering approach to cluster students' annotations.

2.4 Bio-inspired adopting

Each biological individual in the earth has its unique deoxyribonucleic acid (DNA). DNA passes from adults to their offspring. DNA is a chain style polymer, which contains a big amount of genes. Gene stores biological information, for example, hair color, skin color, face shape…etc. There are protections between genetic information. Human's DNA has 23 chromosomes; each chromosome has an average of 30 to 200 million pairs (Monge & Crespo, 2015). These chromosomes compose by cytosine (C), guanine (G), adenine (A), or thymine (T). In this long chain DNA chromosome, Monge and Crespo (2014) compute the DNA complexity by using Shannon Entropy, Kolmogorov Complexity, and statistical complexity. Knowing the complexity of the DNA, researchers can determine whether a specific position contains gene information. In another way, computing the DNA complexity could filter out which parts are low complexity in order to know is this part contains gene information or not. In my research, computing the complexity will be used in weighting students' annotation chromosome.

Gene is very easy to adapt to representing data in various research areas. Maul, Buyer, and Yarwood (2015) have collected the soil samples from Brookings and the Rio Grande and divided the soil into 18 samples (9 samples from Brookings and 9 from the Rio Grande); the samples can be classified into 3 groups based on the moisture treatment

and carbon treatment for one year. In the next step, the researchers have analyzed all the samples and make a heat map about the abundances of phyla and Proteobacteria classes in order to make each sample becomes a long strip chromosome. Therefore, the chromosome composed by different abundances of phyla and Proteobacteria classes; each phylum and classes indicate by a color, which means the concentration of that kind of phyla/classes in this sample. After analyzing all the samples, they use hierarchical clustering method cluster the soil samples.

Another research uses chromosomes to store students' annotations. Ying and colleagues use bit-string chromosomes to represent and to store users' annotations (Ying, Chang, Chiarella, Kinshuk & Heh, 2012; Chang, Kuo, Ying, Chiarella, Heh, & Kinshuk, 2013). Every word in the text is represented by a bit-0 (no highlight) and one (has been highlighted). In addition, four different approaches of clustering users' annotations are by the research, which are Standard, Quantitative Cosine, and Diffusion.

2.4.1 bit-string chromosome clustering methods in Ying's research. To further extend Ying and colleagues' research, the four clustering methods are explained in the following section. The four existing clustering approaches use a few common functions that the four existing clustering approaches are using. For example, when a user reads and takes annotations, the system will collect the user's annotations in a bit-string form as Figure 1 shows. When the user annotates a word, the word will be represented by 1; those words without annotations will be represented by 0.

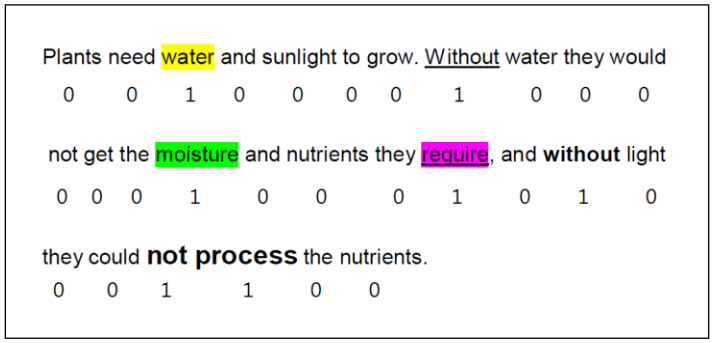


Figure 1. When users annotate an article, the annotations will be turned into a bit-string chromosome for storing purpose

When the system starts running the approaches, all of the four approaches cut the bit-string chromosome into two pieces as Figure 2 shows.

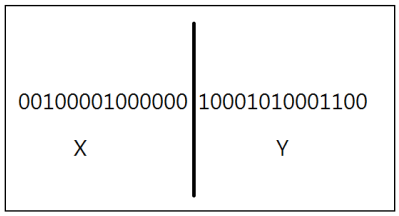


Figure 2. A bit-string chromosome is cutting into two pieces so the chromosome can be projecting on a two-dimension space

Code Segment 1 shows the pseudo code of cutting a bit-string chromosome into two pieces. When a bit-string chromosome is sent in, the class will assess whether or not the length of the chromosome is even (as Line #6 shows). If the chromosome's length is not even, then an additional zero is appended to the end of the chromosome (as Line #7 shows) so the chromosome can be cut into two pieces evenly (as Lines #10 to #11 show).

```
1   Function cutChromosomeInHalf() {
2       String dnaX="";
3       String dnaY="";
4       String ChromosomeString = given a symbolic string
```

```
5
6       if (ChromosomeString.length is not even) {
7           ChromosomeString=ChromosomeString + "0";
8       }
9
10      dnaX = ChromosomeString first half
11      dnaY = ChromosomeString second half
12  }
```

Code Segment 1. The pseudo code of the chromosome cutting function

Another common function is tree-clustering function (as Code Segment 2 shows). In the beginning of this function, all students' ID, coordinates, and `GroupQuantity` will be sent in. The `GroupQuantity` tells the function what the maximum number of clusters should be. If the maximum number of clusters is larger than the number of students, the `GroupQuantity` will be the number of students (as Line #5 shows).

The tree cluster method (as Lines #11 to #29 show) will measure the two closest coordinates and the correspondent student IDs, and record it as `removeUsersID` (as Line #16 shows). The coordinates in the middle of the two closest coordinates have to be calculated and taken as a new coordinate/branch (as Line #18 show). For example, `allCoordinates` may have {<ID_1, (x_1, y_1)>, <ID_2, (x_2, y_2)>, <ID_5, (x_5, y_5)>, <ID_3, (x_3, y_3)>, <ID_7, (x_7, y_7)>} If the ID_3 and ID_1 have minimum distance, the `removeUserID` will record ID_3 and ID_1. If the coordinates of ID_3 is (6, 10), ID_1 is (4, 8), the `newCoordinate` will be (5, 9) as it is the middle position between ID_1 and ID_3. Once the function has the `removeUsersID` and `newCoordinate`, it adds the two student IDs into the `treelist` (as Line #22 shows), removes the two IDs from `allCoordinates` (as Line #24 shows), and adds the `newCoordinate` into the `allCoordinates` for next run

(as Line #27 shows).

```
1   Function TreeCluster() {
2       itemBank allCoordinates = given all students' IDs and coordinates
3       int GroupQuantity = given a number for how many clusters you want.
4       treelistBank treelist;
5       int controlGroupQuantity = smaller number of student amount and
6       GroupQuantity.
7
8       int [] removeUsersID;
9       double [] newCoordinate;
10
11      do {
12          if(allCoordinates' size equals to the controlGroupQuantity){
13              jump out of the do loop.
14          }
15
16          removeUsersID ← find a pair of student IDs whose coordinates
17          have minimum distance in the allCoordinates.
18          newCoordinate ← get coordinates in the middle of the pair
19      of the coordinates stored in the removeUsersID
20
21
22          treelist ← add the two student IDs and correspondent
23          coordinates;
24          remove the two student IDs and correspondent coordinates from
25          allCoordinates;
26
27          allCoordinates ← add the coordinates stored newCoordinate;
28
29      } until (allCoordinates.size equals or is smaller than 1);
30
31      treelist ← If allCoordinates.size is 1, put this coordinates in
32  to treelist. This coordinate is the root of this tree.
```

```
33      allCoordinates = {empty}.
34  }
```

Code Segment 2. The pseudo code of the tree cluster function

When the system has the bit-string chromosome of student's annotations, the system sends the chromosome to the four approaches, gets the two-dimension coordinates of the chromosome from the four approaches, and sends the coordinates to the hierarchical clustering method.

The Code Segment 3 shows the pseudo code of Standard approach. After this function receives a `chromosomeBank` which includes all students' chromosomes (as Line #2 shows), this function gets the first chromosome from the `chromosomeBank` and set the chromosome as standard chromosome (as Line #7 shows). In Line #14, a standard coordinates, (0, 0), is stored into the `CoordinateBank`. Once the standard chromosome is set-up, the compared chromosomes are retrieved from the `chromosomeBank` one by one (as Lines #17 to #44 show). The function measures how different the compared chromosome is from the standard chromosome and gets the difference value by comparing the two chromosomes (Line #28 to #40 show).

```
1   Function StandardApproach() {
2       chromosomeBank allChromosomes = given all students' chromosomes
3       and IDs.
4       TreeInformationsBank StandardTree;
5       CoordinateBank standardCoordinate; //a coordinates vector
6
7       String getChromosome ← get a chromosome at first position in the
8   chromosome bank – allChromosomes
```

```
9
10      String standardChromosomesX, standardChromosomesY ← use
11  cutChromosomeInHalf function to cut the getChromosome.
12
13
14      standardCoordinate ← add standard coordinates (0, 0);
15      String compareChromosomesX, compareChromosomesY;
16
17      for (int i = 1 to allChromosomes' size){
18          String ChromosomeString = get the chromosome at position i in
19  the     chromosome bank – allChromosomes.
20
21          compareChromosomesX, compareChromosomesY = using
22      cutChromosomeInHalf function to cut the ChromosomeString and get
23      the X and Y chromosomes;
24
25          int DifferentValueX = 0;
26          int DifferentValueY = 0;
27
28          for (int j = 0 to ChromosomeString cut in half length){
29              if (standardChromosomesX [j] NOT equals to
30          compareChromosomesX [j]) {
31                  DifferentValueX = DifferentValueX + (int)Math.pow(2,
32                  j);
33              } end if
34
35              If (standardChromosomesY [j] NOT equals to
36              compareChromosomesY [j]) {
37                  DifferentValueY = DifferentValueY + (int)Math.pow(2,
38                  j);
39              } end if
40          } end for (j)
41
42          standardCoordinate ← add coordinate value (DifferentValueX ,
43      DifferentValueY);
```

```
44      } end for (i)
45
46      StandardTree = given standardCoordinate and GroupQuantity to
47      treeCluster function;
48  } end function
```

Code Segment 3. The pseudo code of the Standard approach function

The Code Segment 4 shows the pseudo code of Quantitative approach. After this function receives a chromosomeBank which includes all students' chromosomes (as Line #2 shows), this function gets the chromosome at first position in the chromosomeBank and set the chromosome as standard chromosome (as Line #7 shows). In Line #11, a Quantitative coordinate, (0, 0), is stored into the CoordinateBank. Once the standard chromosome setting is done, the compared chromosomes are retrieved from the chromosomeBank one by one (as Lines #15 to #43 shows). Then the function measures how different the compared chromosome is from the standard chromosome and gets the difference value by comparing the two chromosomes (as Lines #25 to #39 shows).

```
1   Function QuantitativeApproach() {
2       chromosomeBank allChromosomes = given all students' chromosomes
3   and IDs.
4       TreeInformationsBank QuantitativeTree;
5       CoordinateBank quantitativeCoordinate;//a coordinates vector
6
7       String standardChromosomesX, standardChromosomesY = get the
8       chromosome at first position in the chromosome bank –
9   allChromosomes    and use cutChromosomeInHalf function to cut
10  chromosome into two parts.
11
12      quantitativeCoordinate ← add Quantitative coordinates (0, 0);
```

```
13
14      String compareChromosomesX, compareChromosomesY;
15
16      For (int i = 1 to allChromosomes' size) {
17          String ChromosomeString = get the chromosome at position i in
18  the     chromosome bank – allChromosomes.
19          compareChromosomesX, compareChromosomesY = using
20      cutChromosomeInHalf function to cut the ChromosomeString and get
21      the X and Y chromosome;
22
23          int DifferentValueX, DifferentValueY;
24          int countPositionX, countPositionY;
25
26          for (int j = 0 to ChromosomeString cut in half length) {
27              if (standardChromosomesX[j] NOT equals to
28          compareChromosomesX[j]) {
29                  DifferentValueX = DifferentValueX + (int)Math.pow(2,
30              countPositionX);
31                  countPositionX++;
32              } end if
33
34              If (standardChromosomesY [j] NOT equals to
35              compareChromosomesY [j]) {
36                  DifferentValueY = DifferentValueY + (int)Math.pow(2,
37              countPositionY);
38                  countPositionY++;
39              } end if
40          } end for (j)
41
42          quantitativeCoordinate ← add coordinate value
43      (DifferentValueX ,    DifferentValueY);
44      } end for(i)
45
46      QuantitativeTree = given quantitativeCoordinate and GroupQuantity
47  to  treeCluster function;
```

```
} end function
```

Code Segment 4. The pseudo code of the Quantitative approach function

The function, `twoVectorCosineValue()`, calculates the difference value in-between two chromosomes. Code Segment 5 shows the pseudo code of calculating the difference value between two chromosomes.

```
1   Function twoVectorCosineValue(standardDNA, compareDNA) {
2       String standardDNA is a standard DNA passed in.
3       String compareDNA is a DNA for comparing passed in.
4       double dotAB, distanceA, distanceB;
5       for (int i=0 to standarDNA.length) {
6           dotAB = dotAB +
7           (Double.parseDouble(standarDNA.substring(i,i+1)) *
8           Double.parseDouble(compareDNA.substring(i,i+1)));
9
10          distanceA = distanceA +
11          Math.pow(Double.parseDouble(standarDNA.substring(i,i+1)), 2);
12
13          distanceB = distanceB +
14          Math.pow(Double.parseDouble(compareDNA.substring(i,i+1)), 2);
15      }
16      return dotAB / (Math.pow(distanceA, 0.5) * Math.pow(distanceB,
17  0.5));
    }
```

Code Segment 5. The pseudo code of the `twoVectorConsineValue` function

The Code Segment 6 shows the pseudo code of Cosine approach. After this function receives a `chromosomeBank` which includes all students' chromosomes (as Line #2 shows), this function gets the chromosome at first position in the `chromosomeBank` and

set the chromosome as standard chromosome (as Line #7 shows). In Line #11, a standard coordinate, (1, 1), is stored into the `CoordinateBank`. After the standard chromosome setting is done, the compared chromosomes are retrieved from the `chromosomeBank` one by one (as Lines #15 to #30 show). Then, the function measures how different the compared chromosome is from the standard chromosome and gets the value by using `twoVectorConsineValue()` function to compare the two chromosomes (as Lines #23 to #27 show).

```
1   Function CosineApproach() {
2       chromosomeBank allChromosomes = given all students' chromosome and
3       IDs.
4       TreeInformationsBank CosineTree;
5       CoordinateBank cosineCoordinate; //a coordinates vector
6
7       String standardChromosomesX, standardChromosomesY = get the
8       chromosome at first position chromosome in the chromosome bank –
9       allChromosomes and use cutChromosomeInHalf function to cut
10  chromosome   into two parts;
11      cosineCoordinate ← add cosine coordinates (1, 1);
12
13      String compareChromosomesX, compareChromosomesY;
14
15      For (int i = 1 to allChromosomes' size) {
16          String ChromosomeString = get the chromosome at position i
17          chromosome in the chromosome bank – allChromosomes;
18
19          compareChromosomesX, compareChromosomesY = using
20      cutChromosomeInHalf function to cut the ChromosomeString and get
21      the X and Y chromosome;
22
```

```
23          int ValueX, ValueY = 0;
24          ValueX = twoVectorCosineValue(standardChromosomesX,
25          compareChromosomesX);
26          ValueY = twoVectorCosineValue(standardChromosomesY,
27          compareChromosomesY);
28
29          cosineCoordinate ← add coordinate value (ValueX, ValueY);
30      } end for (i)
31
32      CosineTree = given cosineCoordinate and GroupQuantity to
33  treeCluster  function;
34  } end function
```

Code Segment 6. The pseudo code of the Quantitative approach function

The Code Segment 7 shows the pseudo code of Diffusion approach. After this function receives a `chromosomeBank` which includes all students' chromosomes (as shows in Line #2), this function compares position j in chromosomes at the same time (as Lines #9 to #18 show). If all chromosomes have same value at position j, then position j will be represented by an "H"; otherwise the position j will be represented by a character "D".

After the for-loop (i.e., Line #17), the `diffusionString` will be a D-H string (i.e., something like DHHDDHHHDDD). In the `diffusionString`, the areas filled with D indicate the differences among chromosomes. This function uses `positionBank` to store the information like how many such areas these chromosomes have and how many D each area contains. Taking abovementioned D-H string as example, there are three areas filled with D: the first area contains 1 Ds; the second area contains 3 Ds; and, the third area contains 2 Ds (as Lines #20 and #21 show). Therefore, the `PositionBank` will

store a vector like {(0, 1), (3, 2), (8, 3)}. Where (0, 1) means there is 1 D starting at position 0; (3, 2) means there are 2 continuous Ds starting at position 2; and (8, 3) means there are 3 continuous Ds starting at position 8. In such case,

`PositionBank.at(0).startPposition` is 0 and

`PositionBank.at(0).Dnumber` is 1 for (0, 1);

`PositionBank.at(1).startPosition` is 3 and `PositionBank.at(1).Dnumber` is 2 for (3, 2); and, `PositionBank.at(2). startPosition` is 8 and `PositionBank.at(2).Dnumber` is 3 for (8, 3).

With the `positionBank`, this function compares each chromosome to the position information and measure each chromosome's coordinate (as shows in Line #23 to #39). For example, a `ChromosomeString` is "00110110100" and the `diffusionString` is "DHHDDHHHDDD" as Figure 3 shows. There are two hot zone areas and three diffusion areas starting at position 0, 3, and 8. The three diffusion areas are represented by "0", "10", and "100" in the `ChromosomeString`. Both of areas 2 and 3 have one "1" and area 1 has no "1" existed. This function (as Lines #29 to #35 show) count only the number of "1" existed in a diffusion area.

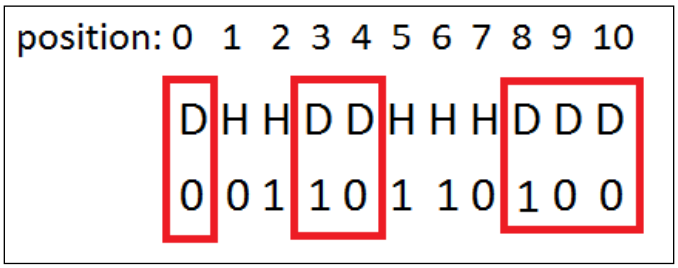


Figure 3. Explain the relationship between parameter ChromosomeString and diffusionString

```
1   Function DiffusionApproach() {
2       chromosomeBank allChromosomes = given all students' chromosomes
3   and IDs.
4       TreeInformationsBank DiffusionTree;
5       DiffusionCoordinateBank diffusionCoordinate;
6
7       String diffusionString = "";
8       String temSimble;
9       for (int j = 0 to the length of a chromosome) {
10          for (int i = 0 to allChromosomes' size) {
11              if (all chromosomes have same value at position j) {
12              temSimble = "H";
13              } else {
14              temSimble = "D";
15              } end if
16          } end for (i)
17          diffusionString = diffusionString + temSimble;
18      } end for (j)
19
20      PositionBank position = get how many groups of "D" in the
21      diffusionString and how many "D" in each groups.
22
23      For (int k = 0 to allChromosomes's size) {
24          String ChromosomeString = get the chromosome at position k in
25  the     chromosome bank – allChromosomes.
26          int x = 0, y = 0;
27
28          for (int i = 0 to position's size) {
29              int counter = how many "1" existed in the diffusion area
30  i of ChromosomeString starting from position.at(i).startPosition to
31  position.at(i).startPosition+position.at(i).Dnumber
32              if (counter not equals to 0) {
33                  y = y + counter;
34                  x = x + i;
35              } end if
```

```
36          } end for (i)
37
38          diffusionCoordinate ← add coordinate(x, y).
39      } end for (k)
40
41      DiffusionTree = given diffusionCoordinate and GroupQuantity to
42      treeCluster function;
43  } end function
```

Code Segment 7. The pseudo code of the Diffusion approach function

In Ying and colleagues' experiment, 40 student's annotations on 2894-word text had been used for evaluating the performances of the four approaches. They found that Diffusion approach only took 20.53 milliseconds and was the fastest one compared to Standard approach's 29.95 milliseconds, Quantitative approach's 27.43 milliseconds, and Cosine approach's 178.7 milliseconds. Although Cosine approach is slowest, it can still cluster 40 students' annotations on thousands of words article within a second and has highest accuracy rate with 0.7488 precision compared to Standard's 0.7146, Quantitative's 0.7027, and Diffusion's 0.7047. All of the four approaches have high speed and accuracy in clustering students' annotations; however, they can only deal with one single annotation type – highlight words in single color.

In my research, I would like to extend Ying and colleagues' research; I want to provide an article to student, that means the article could be more than 10 thousand of words. I also want to provide more annotation ways to students instead of only highlighting can be used. Therefore, student's annotation chromosome will be long and complex. For this reason, how to compute the long and complex chromosome is my next research issue to solve.

2.4.2 the bio-inspired methods applications. Monge and Crespo (2015) have developed a predictor for analyzing chromosome automatically. Monge and Crespo set the different window sizes to measure the precision and recall value on estimating of gene zones in human DNA. They got good precision and recall values when window size set on 1000 and 2000 but got poor results when set on 250 and 500. This result may cause by basic gene components which are bigger than 500. In my research, I may need to compute students' annotations by sentences, paragraphs or the whole article to adjust the window size.

Monge and Crespo also concerns the window shift problem; if the set window's shift size is too small, the shifting is more effective but will take larger computing time. Their suggestion is that the window shift could be set as 1/4 of the window size. Monge and Crespo's research in 2014 gives me an idea about computing students' annotations' chromosomes, I can find out which fragments are without genes and give these fragments lower weight.

Regarding the bio-inspired methods, the well-known methods will be regarded – genetic algorithms (GA), particle swarm optimization (PSO), and ant colony optimization (ACO). The genetic algorithm is chosen to solve the problem, this algorithm continues to create new solutions until a best solution is found. In the beginning , the genetic algorithms is randomly creates lots of chromosomes. Second step is using each chromosome to solve the problem to sees how good it is and assign them a fitness score. In third step, based on the chromosomes' fitness score they are divided into several

groups. According to each group's proportion, put it on a pie chart and spin this pie chart to choose two chromosomes as the parent chromosomes from the chosen group. After the parent chromosomes are chosen, randomly choose a position in the chromosome and swap all the bits from the chosen position to the end. The last step, after these two new chromosome come out, it will have a very small chance flip one bit (0 become 1, 1 become 0). Genetic algorithms will repeat from step three to the last step until the N new members come out and these N new members are the next generation. The genetic algorithms can't guarantee after how many generations will get a good result. (Brownlee, 2011).

The basic idea of particle swarm optimization is from bee's daily working. When a bee find a place can collect the honey, the bee will come back to report the information it got – which direction, how far, how much, and how rich. Each bee searches the best result it can find and remembers the position and what is the best result so far. Bees will exchange the information about what they got in the place they discovered. Bee will remember which place has the better result than the one it has for itself. Bees will keep searching and updating until some condition is met. (Blondin, 2009 and Darzi and collage, 2013).

Regarding the ant colony optimization, it is inspired from ants looking for food and ants always can find the shortest way reaching the food. When ants look for the food, they left the pheromone on their way to the food. The strength of pheromone will decay over time. Pheromone trail will build up fast on shorter path. Ants will follow the stronger pheromone trail. (Dorigo, & Stützle, 2004).

These three famous bio-inspired methods are looking for the best result, but my

research is not looking for the best result. Even more, my research does not have the best result. If an algorithm chooses a best annotation from all students, this best annotation can't give useful reminder to all students. Each student have their thoughts to make the annotations – maybe they annotate what they understand, maybe they annotate what they don't understand, and maybe they annotate the important thing in that article etc. Different thinking has different annotation behaviour. It is impossible to use one best annotation and give all the students useful suggestions. I saw some papers are talking about multiple cluster analysis, maybe I can consider using ant colony optimization.

2.5 Objectives and research issues

An article may have thousands of words or longer. When a teacher collects students' annotations of an article, he/she has difficulty in quickly identifying which two students have similar annotation behaviour. To help teachers cluster students according to students' annotations and remind students the missing keywords in their annotations, this research needs to solve four research objectives when we design a bio-inspired clustering approach. Each objective may have one or more issues.

Objective 1: Storing and representing student annotations.

When students read a text, they will annotate the text with their preferred annotation ways. Their annotations are stored in kind of structures. The structure should be able to represent a student's annotations for many thousands of words without having corresponding size growth.

Issue 1-1: How to represent different ways that a student may use for annotating a

text?

In the researches of Chang, Chiarella, Kinshuk and Heh (2012) and Chang, Kuo, Ying, Chiarella, Heh and Kinshuk (2013), researchers have used bit-string chromosomes to represent students' annotations. In previous researches, they use each bit to indicate whether students way take an annotation. This research provides students several ways to annotate and it is impossible to adopt the bit-string chromosome to represent students' annotations with several annotated ways. (This issue will be solved in chapter 3.1)

Issue 1-2: What data structure are used to store student annotations?

This research represents every alphabet as a new character. Students may annotate a word with multiple annotation ways. As bit can store and represent only one annotation way such as highlight, an ASCII (8 bits) character can be used for representing eight different annotation ways applied on a single character of a text. Though an ASCII character can represent multiple annotation ways that students may use, it can only represent eight annotation ways. UTF-8 character is a 32-bit character, which means an UTF-8 character represents 32 different annotation ways. I consider the use of UTF-8 encoding in order to make the proposed method more flexible and can cover more annotation ways that students may use for reading an e-text. (This issue is solved in chapter 3.1)

Issue 1-3: How to ensure that the data structure will not grow too much when a text has thousands of words?

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 Figure 4. 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. (This issue is solved in chapter 3.1)

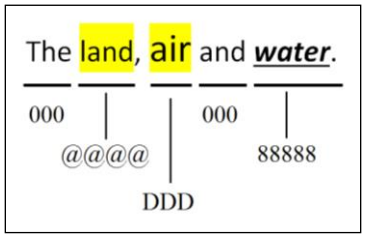


Figure 4. An example of an annotated text represented by UTF-8 characters

Objective 2: Clustering student annotations.

This research needs (1) to design a method to weigh the structure which represents a student's annotations so two structures can be compared its similarity; (2) to design a fast enough method to compare the difference and the distance of two structures; (3) to design a clustering method to cluster all structures into several groups.

Issue 2-1: How to weight a data structure?

This research need to find a way to compare the difference of two data structures so researcher can tell if a student's annotation behaviour is close to another student's. (This issue is solved in chapter 3.2)

Issue 2-2: How to compare data structures and measure the differences?

Researcher prefers to compare all chromosomes (i.e., all data structures that represent all students' annotations on a text) at the same time but use the way of feature extraction and comparison that the pattern recognition research area has. (This issue is solved in chapter 3.2)

Issue 2-3: How to cluster data structures?

For clustering data structures, it has to compare data structures and to see the similarity among structures. (This issue is solved in chapter 3.3)

Objective 3: Having high performance clustering approach.

High performance in this research means the clustering results are accurate and the clustering process is fast. This research needs to prove the proposed methods are fast and accurate enough with appropriate experiment design and quantitative and qualitative data analysis.

Issue 3-1: What is the benchmark?

This research shall use the clustering results altered by teachers as the benchmark because the student clustering results are intended for teachers. (This issue will be solved in chapter 4.4)

Issue 3-2: How to measure the accuracy of the clustering results?

I will use precision and F-measure (F0.5, in particular) to measure the accuracy of the proposed approach and previous approach(es) by comparing with the benchmark. (This issue will explain in chapter 5.4)

Issue 3-3: How to measure the speed of the proposed approach?

System records the time before the proposed approach stars running and when the proposed approach finishes the clustering process. Using these time stamps measure the speed of proposed approach. (This issue will explain in chapter 5.4)

Issue 3-4: How to calculate the performance of the results?

When I get the results of the accuracy and time spent, I have to evaluate which one

is better. If there are three results (1) accuracy is 85%, time spend is 27 sec. (2) accuracy is 80%, time spend is 28 sec. (3) accuracy is 70%, time spend is 20 sec. I will measure which result has higher acceptance by students. (This issue will explain in chapter 5.4)

Objective 4: Providing teachers feedback for potential learning problem that students have.

This research wants to provide teachers clear idea of which cluster's students have potential learning problems and teachers can adjust their teaching direction or material for students.

Issue 4-1: How to find behaviour features?

Asking teachers write down students' behaviour features when they adjust the clustering results. (This issue will explain in chapter 5.4)

Issue 4-2: How to find potential learning problems?

Asking teachers explain that does each behaviour feature has potential learning problems. (This issue will explain in chapter 5.4)

Issue 4-3: How to match the behaviours' features and learning problems?

Finding connections between the behaviours' features and the potential learning problems. Between clusters, they may share the same behaviours' features they may have same learning problems. A learning problem may be shown in different behaviours' features' group. This issue will ask teacher to match the annotation features and potential learning problems. (This issue will explain in chapter 5.4)

Chapter 3. Bio-inspired clustering approach

3.1 Storing and representing student annotations

In previous research Ying and colleagues (Ying, Chang, Chiarella, Kinshuk & Heh, 2012; Chang, Kuo, Ying, Chiarella, Heh, & Kinshuk, 2013), researchers have used bit-string chromosomes to represent students' annotations. In this research, I use each bit to indicate whether an annotation way is taken by students. For instance, highlighting is an annotation way and this research uses the second bit of an ASCII character to represent whether or not a student highlights something in an e-text. If each bit represents one annotation way, an ASCII character can only represent maximum 8 ways because ASCII character is a byte with 8 bits. I would like to allow students to annotate e-text in more ways; therefore, I choose to use UTF-8 character instead of ASCII character to represent students' annotations. The use of UTF-8 encoding to make the chromosome more flexible and can cover more annotation ways that students may use in real case. UTF-8 character is a 32-bit character, which means an UTF-8 character can maximum represent 32 different annotation ways. Base on UTF-8 encoding rule, this research can provide 21 annotation ways for students.

Table 2 lists examples of the use of UTF-8 characters to represent a student's annotations on a character. When the student highlights a character, the character's annotation can be recorded as U+0040 (i.e., "@"). On the other hand, if students annotate a character by highlighting it and enlarging its size, the character's annotation should be recorded as U+0044 (i.e., "D").

Table 2.

Example of UTF-8 encoding in multiple annotation styles

Chang word's color to red	Chang word's color to blue	Highlight in blue color	Highlight in yellow color	Underline	Bold face	Italics	Increase word size	Double-line	Strikethrough	Example	UTF-8 code	UTF-8 character		
0	0	0	0	0	0	0	0	0	0	a	U+0000			
0	0	0	0	1	1	1	0	0	0	***a***	U+0038	8		
0	0	0	1	0	0	0	0	0	0	a	U+0040	@		
0	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 Figure 4 (in page 34) . 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.

In this UTF-8 string, "0", "D", "@", and "8" are repeated. To make the string shorter for analysis and comparison purpose, one possible solution is the use of regular expression. In the regular expression, the use of "+" indicates the symbol beforehand can be one or more. By applying regular expression, the original UTF-8 string can be converted to "$0^{+}@^{+}D^{+}0^{+}8^{+}$". Another possible solution is to count how many times a

symbol is appeared consecutively in the string. For instance, "D" appears consecutively three times and "DDD" can be replaced by "D^3". In such case, the original UTF-8 string can be converted to "$0^3@^4D^30^38^5$".

3.2 Calculating a value for student annotations

The complexity of chromosome string is raised when adopting the UTF-8 chromosome for store students' annotation. Because students would choose different ways to annotate on the article and make different UTF-8 annotation chromosomes, this section designs the method to reduce the complexity of the UTF-8 chromosomes to weight UTF-8 chromosome which solve the Objective 2 in this research.

According to Brent's research in 1987, Huffman is the fastest method to reduce complexity comparing to the other three methods (Move-to-front (MTF), Lempel-Ziv-Welch (LZW), and linearity of the maximal matching (SLH)) introduced in his research. On the other hand, Manna compares the Huffman coding and Arithmetic coding in compression ratio and compression speed. Arithmetic coding gets better performance in compression ratio, but Huffman coding uses less time in decompression speed (Manna, 2013).

I also study Ramya and Pushpa's research in 2016 and Mathpal, Mittal, and Mehta's research in 2017. There are five methods introduced in these two studies. Indata compression, Ramya and Pushpa (2016) introduce three different lossless data compression – LZW, Huffman, and Shannon-Fano. Mathpal and colleagues' research introduces the Arithmetic code and Run Length encoding. The Huffman and Shannon-Fano's compression ratio is powerful than LZW. In these famous data compression

methods, considering the LZW and the Arithmetic need to see the next character and students' movement and thinking cannot be predicted. Therefore, these two compression methods are not considered

The Run Length encoding is an easy way to adopt into students' chromosomes, but the length of the compression string won't be the same and the length will be influenced by how much different annotation ways that student use. The more annotation types that students use, the output will be longer, and the complexity is still not reduced. Therefore, the Run Length encoding is not in the consideration. The Shannon-Fano and Huffman is very suitable for reducing the complex of the annotation chromosome. Because of students' annotations' number could change all the time, the symbol tree should not change all the time too. Therefore, this research rules out the Shannon-Fano and adopts the Huffman methods to reduce student's annotations complexity.

This research doesn't use the traditional Huffman tree to represent students' annotations. When Huffman method start the compression, it will calculate the number of appearance for each alphabet. According to the appearance time, an optimum code will be produced. The alphabet appears more frequently, and the optimum code will be shorter.

For example, when system receive an UTF-8 string "000@@@@DDD00088888", the first step of the Huffman method is calculating the appearance times for each alphabet as Figure 5 shows. System will put these characters in the order and put these characters as the step 1 as Figure 6 shows. When the tree grows, the first parent will choose the appearance frequency in the last two character for the first node. The character D appears 3 times and the character @ appear 4 times, these two characters is the last two

appearance frequency. Therefore, the first node is linking the D and @ and give this node a new appearance as 7 as the step 2 in Figure 6 shows. The next node is following the same rule – the last two appearance leaf is the 0 and 8, and the algorithm will make these two leaf as a new node. After the D and @ became a new node, the left node is the 0 (6 times), 8 (5 times), and the first node (7 times). The last two appearances are the 0 (6 times) and 8 (5 times), make these two become a second node and appearance is 11.

Character	Frequency
0	6
@	4
D	3
8	5

Calculate the appearance

Character	Frequency
0	6
8	5
@	4
D	3

Make it in the order by appearance.

Figure 5. The frequency of a string for Huffman method

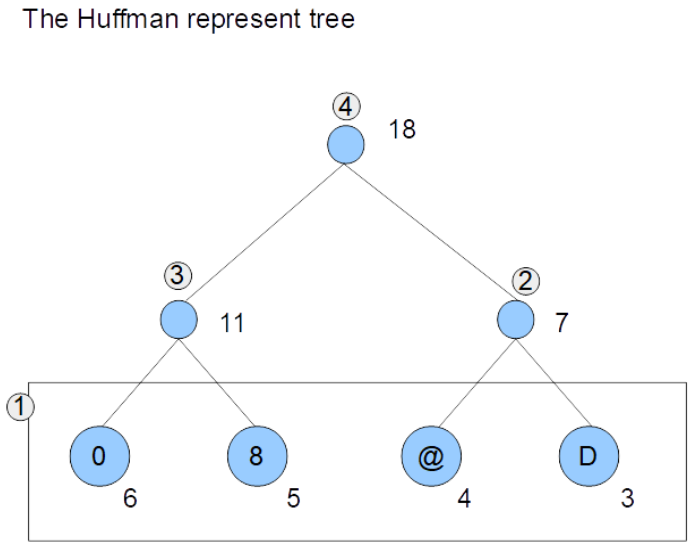
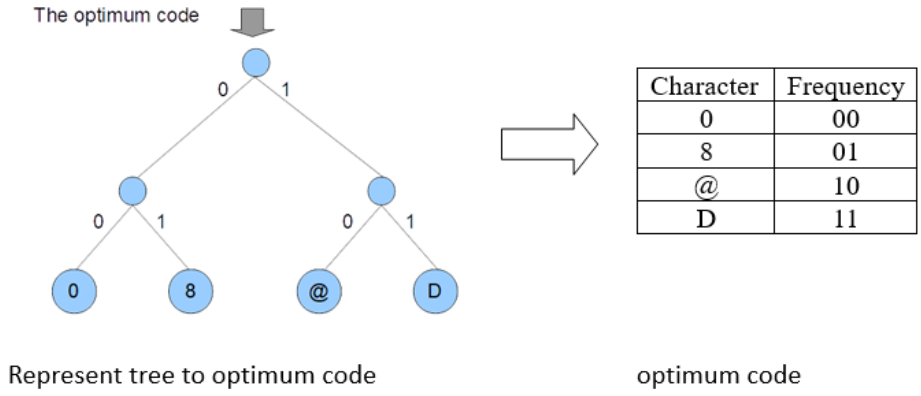


Figure 6. The Huffman represent tree

After the Huffman tree has made, the represent tree needs to be transferred to an optimum code. This step is labeling each parent to its left child with digit 0 and right child with digit 1. The optimum code for each character is the path label from the root to the leaf. For example, the optimum code for character @ is 10. From the tree root, digit 1 is right leaf, follow the right path will see another node; second digit is 0, so go to left will see the @ in the end leaf. The optimum code is shown in Figure 7.



Character	Frequency
0	00
8	01
@	10
D	11

Figure 7. Transfer the Huffman tree to an optimum code

The complexity of the UTF-8 chromosome depends on how much different annotation ways that students use on the article. If the student prefers to use two different annotation ways, there are three possible annotation ways that could happen. For example, if this student likes to use the underline and circle, when student use the underline in one character, the character will replace to a new represented UTF-8 character. When student use the circle on another character, the spot will replace to the second represented character. When the underline and circle annotate on the same spot, the represented character will have a different represented character other than the

underline only and the circle only represented character. If the student likes to use three annotation ways on the article, there are 7 possibility combinations for these three annotation ways; four ways will have 15 possibility combinations; and so on.

Take the pollution article in in Figure 8 for example. Two student read the pollution article and makes the annotations on it. Suppose these two students have similar reading style, they both think the word “land”, “air”, and “water” are important. They all use one annotation way on “land”, two annotation ways on “air”, and three annotation ways on “water”. Suppose using more annotation ways on one spot indicates the concept in that spot is more important the spots with fewer annotation ways. Therefore, both students believe keyword “water” is more important than “land”. However, one student prefers to use the highlight, and the other student believe underline is easier for reviewing. Although the two students give same weights (the number of annotation ways) on the three keywords in the pollution article, the UTF-8 chromosome from their annotations is complete different because of the UTF-8’s code of their annotation ways are different.

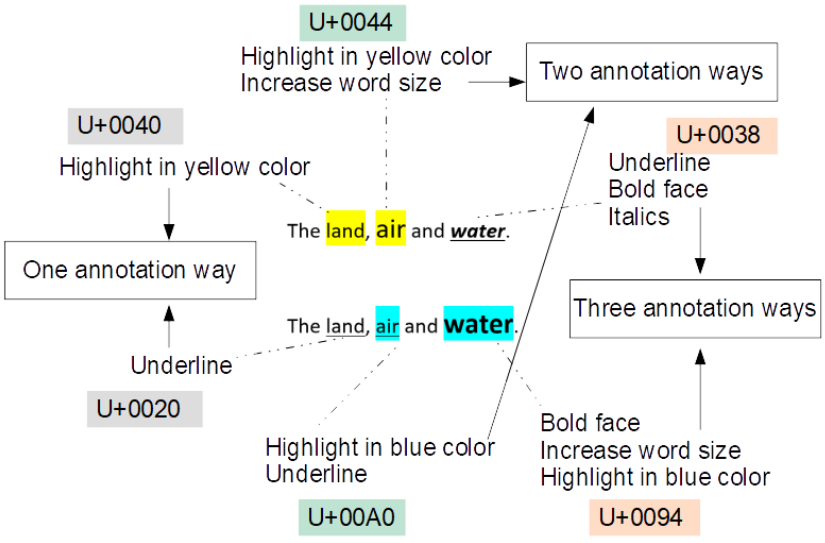


Figure 8. An extend example from the pollution article

For reducing the complexity of the UTF-8 chromosome structure, this research is adopting the Huffman code to compress the UTF-8 chromosome. The system calculates how many different UTF-8 codes are on this article and the counting of each UTF-8 codes appears. Take Figure 9 for example, the word “land” has four alphabets, so the counting of @/U+0040 is 4; “air” has three alphabets, so the counting of D/U+0044 is 3; “watch” has five alphabets, so the counting of 8/U+0038 is 5.

After counting how many alphabets were used for the specific annotation way, the proposed method makes these different ways in the order by the times. Based on the ordered list, the system gives a symbol to each annotation way. When all the symbols are prepared, the method replaces student annotations to the corresponded symbol. As Figure 9 shows, the original UTF-8 chromosome is including the “space”, because the “space” place could get annotate too. For example, if the space annotation by underline, this space will replace to a correspond symbol. If not, put a digital 0. When system calculates the times, the 0 in the UTF-8 doesn’t in count. Therefore, the calculated timetable only has character @, D, and 8.

After sorting the calculate table by the appeared time, the next step is making a Huffman tree and getting the optimum code if I follow the Huffman method. However, the main goal for adopting Huffman method is reducing the complexity, not for data compress. Therefore, this research skips the Huffman tree and give another defined code instead of the optimum code from Huffman tree; the UTF-8 chromosome is revised to Huffman chromosome – “000022220333000000111110”. Figure 9 is an example of transforming UTF-8 chromosomes in Figure 8 to Huffman chromosomes. The □ and

the △ in the right hand side represent two different characters that are unable to be displayed on screen. After the Huffman procedure, second student's UTF-8 chromosome will get the same Huffman chromosome as "000022220333000001111110" as first student's Huffman chromosome.

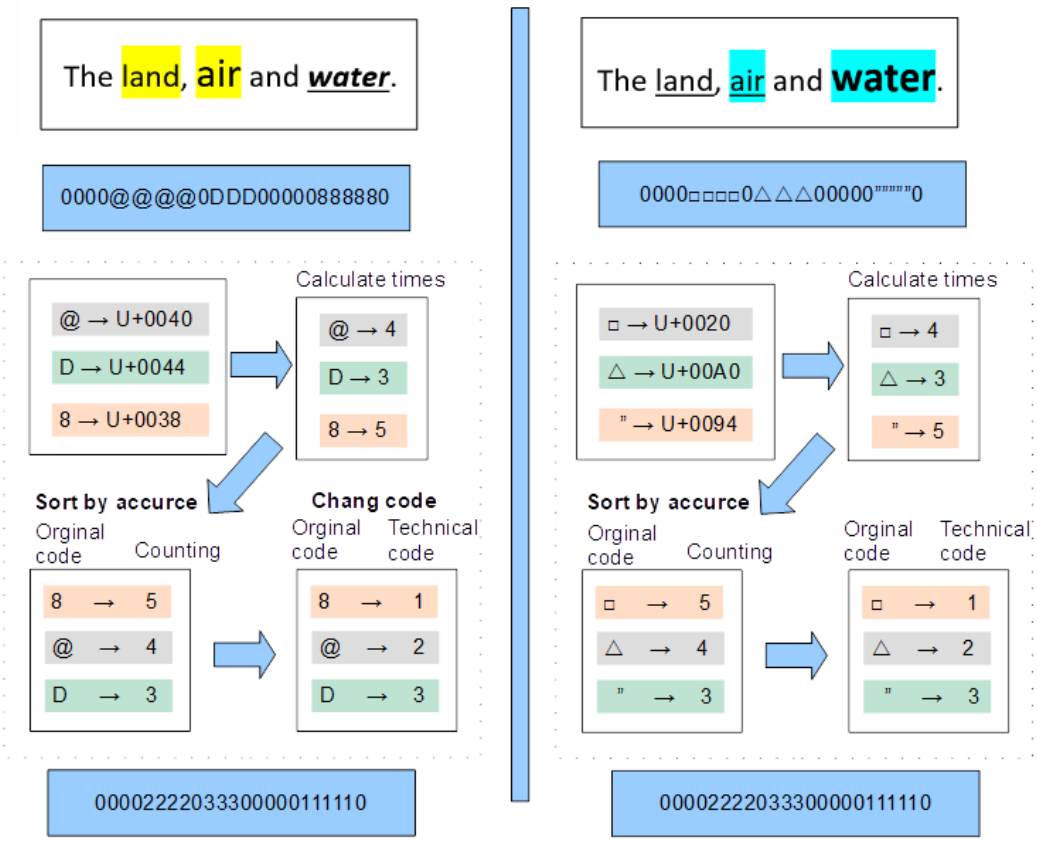


Figure 9. An example of transform UTF-8 chromosome to Huffman code chromosome

After the system changes the UTF-8 chromosome to the Huffman one, the different annotate ways can represent different important level to the student and two students' annotations has the consistency evaluation. Therefore, the issue #2-2 – How to compare data structures and measure the differences – can be solved. The comparing method of

measuring the differences between two Huffman chromosomes is displayed as pseudo code in *Code Segment 8* showing how to calculate the difference value between Huffman chromosomes for GRACE algorithm.

This function receives a `chromosomeBank` which includes all students' Huffman-string chromosomes (as Line #2 shows) and `groupNum` which sets for how many groups will be clustered. If the `groupNum` is not equal to 1, prepare a `valueBank` for the all combinations and the calculated value for each combination. For example, if there are four chromosomes IDs like A, B, C, and D, the combination will have AB, AC, AD, BC, BD and CD six combinations. Therefore, the `valueBank` will need to put these six combinations and the calculated values for each combination. In Line #10 and #12 have two for loop, these two For loop get the Huffman chromosomes and these two For loop will run the all combinations. The Line #11 gets an entire chromosome and Line #13 gets another one. The Line #14 to #25 calculate the values. Checking each position (the For loop in Line #16 will run the entire chromosome) between two chromosomes are the same or not, if the position is the same value, which number that this position is plus 1 and square it (as Line #19 to #21 show). If the position is not the same value, gets the absolute value for one minus another one (as Line #23 shows). After reviewing the chromosomes, will get the total same value and different value. The value of these two chromosomes is `differentValue/sameValue`. After puts these combination IDs and the value into the `calculatedInfo`., reset the `differentValue` and the `sameValue` and start to calculate the next combination's value. After calculating all the combination's value, sorting the `calculatedInfo` from small to high by the calculated value.

```
1   Function calculateValues () {
2       chromosomeBank allChromosomes ← given all students' Huffman
    chromosomes  and IDs.
3       int groupNum ← set for how many groups
4       if (groupNum is 1){
5           groupMember ← add all students IDs.
6       } else {
7           ValueBank calculatedInfo; // prepare for user combination and
                                      //the value calculate for each
          combination.
8                                    //(combination: ABCDE => AB, AC, AD, AE, BC…)
9           double sameValue, differentValue=0;
10          int compare1, compare2;
11          for (int i=0 to allChromosomes' size){
                String ChromosomeString1 = get the chromosome at position i
12              in the chromosome bank – allChromosomes.
13              for (int k = i+1 to allChromosomes' size){
                    String ChromosomeString2 = get the chromosome at
14  position                    k in the chromosome bank – allChromosomes.
                    for (int j=0 to the length of a chromosome –
15                  ChromosomeString1){
                        compareChar1= ChromosomeString1 get the number at
16                      position j in the ChromosomeString1
                        compareChar2= ChromosomeString2 get the number at
17                      position j in the ChromosomeString2
18
19                      int temCheckValue = compareChar1- compareChar2;
                        if (temCheckValue is 0 & this position is not
20  equal                       to the 0){
21                          int temValue= compareChar1-0;
                            sameValue =
22                          sameValue+Math.pow((temValue+1),2);
23                      } else {
                            differentValue =
24                          differentValue+Math.abs(temCheckValue);
```

```
25                          } End if
26                      } End for (j)
27                      If (sameValue is 0){
28                          sameValue=1;
29                      } End if
30                      Double calculateValue = (differentValue/sameValue);
                        calculatedInfo ← add the user ID of ChromosomeString1
31  and                 user ID of ChromosomeString2 and calculateValue;
32
33                      differentValue=0;
34                      sameValue=0;
35
36                  }End for (k)
37              }End for (i)
38
                calculatedInfo ← make data in order by calculateValue (from
39  small           to big value)
40
41              prepareGroups();
42              grouping();
43
44          }end if
    }end function
```

Code Segment 8. The pseudo code of calculating function for the GRACE algorithm

The example of the Code Segment 8 is shown in Figure 10, which compares two Huffman chromosomes in the same time. The difference of two chromosomes is using the different annotation behaviour divided by the common annotation behaviour. If the value in the same position of the two chromosomes are different as the position 1, 2, 4, 7, 8, 9, and 12 in Figure 10, these positions are used to get the value of the different annotation behaviour. The method finds the difference of the value in the same position of the two

chromosomes. For example, the difference of the two values in position 1 is |0-2| = 2.0; the difference of position 9 is |2-3| = 1.0. System will accumulate all the different number and make the accumulated number as `differentValue` in the algorithm.

The common annotation behaviour in Figure 10 are in position 3, 5, 6, 10, 11, 13, and 14; these positions are used to calculate the `sameValue` in the algorithm. If the position has the same value and the value is 0, it indicates that both of students have no annotation on this character and the algorithm will ignore this position. If the position has same value and the value is not equal to "0", the system will get the square value of the number and plus 1. Take position 6 for example, the method will add 1 on value 1 and square the result; the position 10 will get $(2+1)^2$ in the end. The sum of the common annotation behaviour is $2^2+2^2+3^2+4^2 = 33.0$ and it is the value of `sameValue` in the algorithm. The difference of the two chromosome M and N is using `differentValue` / `sameValue` and the result is 11/33 = 0.333.

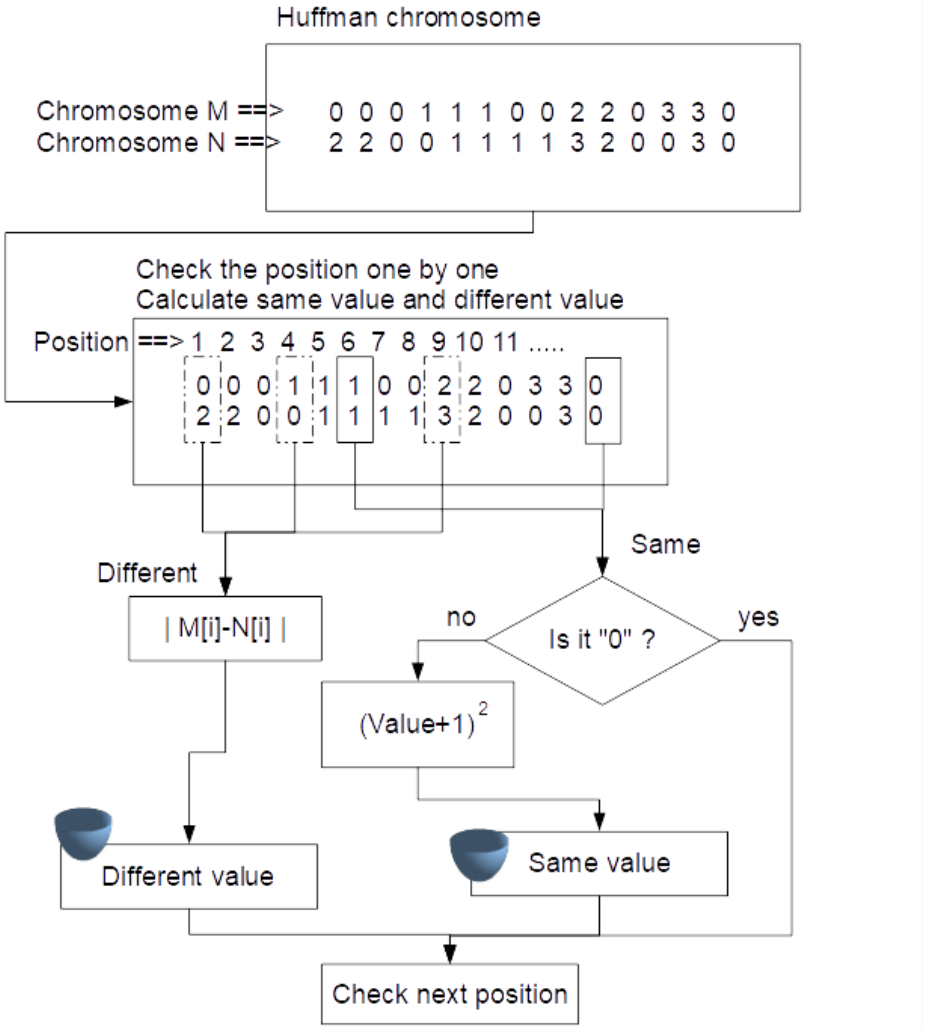


Figure 10. A method to weight Huffman chromosomes

3.3 Clustering student annotations

This research uses the shortest distance method to solve the issues #2-3: How to cluster data structures. The first step is listing every two chromosomes' distance and finding the shortest distance to be a group's foundation groups. Before the clustering methods start, the foundation of the groups' combination needs to be selected and the

Code Segment 9 is the function for how to select the foundation combinations. In Line #2 will receive a `ValueBank` which contains the combinations and their value from the `calculateValues` function. The `groupNum` in Line #3 is the same setting as the Line #3 in Code Segment 8. The for-loop in Line #7 in Code Segment 9 is setting the groups foundation one by one. The for-loop in Line #8 is for checking the all combinations in `ValueBank`. The if in Line #9 is finding one combination which both are not chosen as a foundation group for other groups. If find the combination both are not chosen before, put this combination into the `groupMember` and remove this combination from the `calculatedInfo`.

```
1  Function prepareGroups() {
2      ValueBank calculatedInfo;
3      int groupNum ← set for how many groups
4      groupMember groupMember; //prepare for groups.
5      boolean check0=false;
6      boolean check1=false;
7      for (int i = 0 to groupNum){
8          for (int k = 0 to calculatedInfo's size){
9              if (one of the user ID in calculatedInfo[k] is picked
10 before){
11                 break;
12             } else {
13                 groupMember[i] ← pick up first record from
14                 calculatedInfo and remove this record from the
15                 calculatedInfo.
16             }end if
17         }end for (k)
18     }end for (i)
   }end function
```

Code Segment 9. The pseudo code of preparing groups function

Figure 11 and Figure 12 shows an example of how to cluster five chromosomes into two groups. Figure 11 use the algorithm listed in Code Segment 8 to calculate the distance of the chromosome distance in pairs and sort the pairs from closest distance to the greatest. After getting the ordered list, Figure 12 shows how to find the base groups from the list. The step 1 is getting the first combination from the list. The step 2 is checking the next one from the list, follow the line 4 in the Figure 12 will see the M-N. Because N is already a member of group 1, M-N can't be chosen for a new group. Therefore, system will check the next combination in the list and the O-Q will be the next one. O-Q has the same problem to be a new group, because O is the member of the group 1. The next one is the P-Q combination, no matter P or Q are not chosen for any group. Therefore, the P-Q is chosen for the new group.

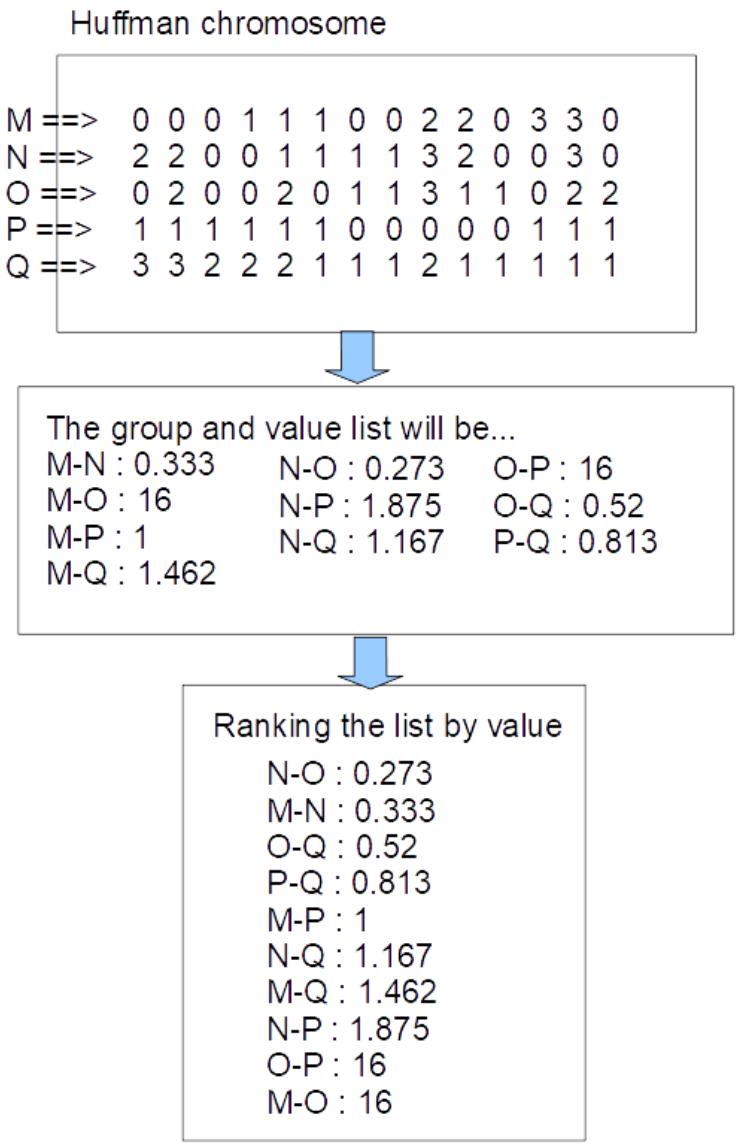


Figure 11. The calculated value and the list in order

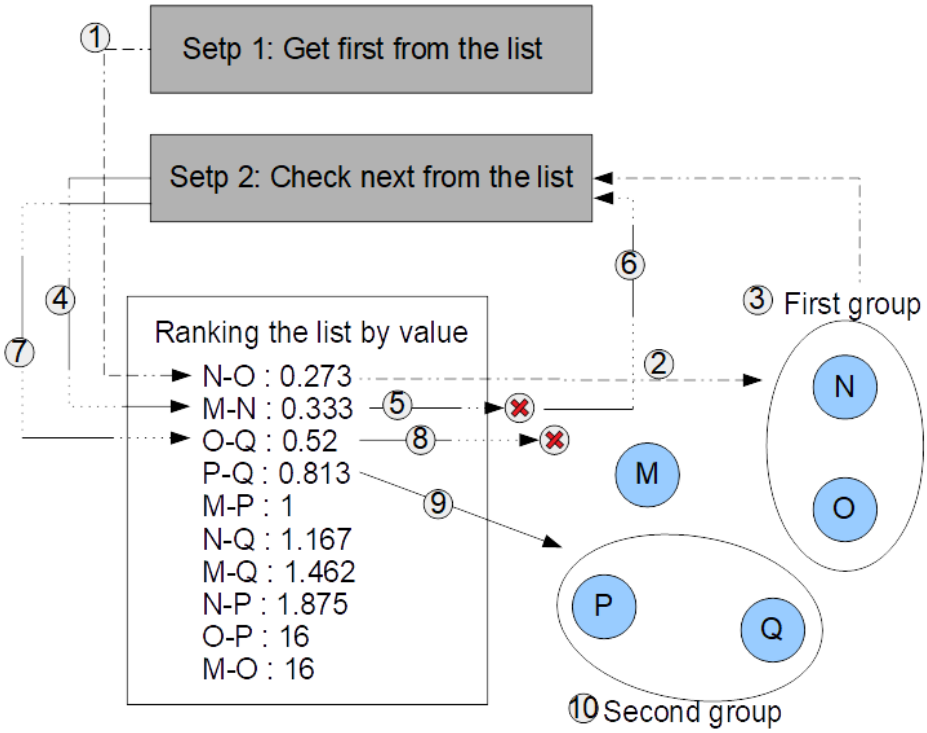


Figure 12. An example of preparing the foundation of the group for clustering

When the foundation of the group members has prepared, the clustering function is the next step. Before starting the grouping, the Line #2 and #3 in the Code Segment 10 need to receive the `ValueBank` and the `groupMember` from the prepareGroups() function. The for-loop in Line #5 run the all list in the `ValueBank`. When getting the combination (in Line #6), checking both of chromosomes if one of them belongs to one of the `groupMember`. If one of them belongs to `groupMember`, put another one of the combinations to the `groupMember` and remove the combination from the `ValueBank`. Checking all the combination from the `ValueBank` until the `ValueBank` empty. When the `ValueBank` empty, the grouping is done.

```
1  Function grouping() {
2      ValueBank calculatedInfo;
3      groupMember groupMember;
4      Do {
5          For (int k = 0 to calculatedInfo's size) {
6              String userID = get the user IDs at position k in the
               calculatedInfo. (it will contain two userIDs)
7              if (one of the two user ID is in one of the groupMember) {
8                  groupMember ← put another user ID into the group that
         the            "IF" found.
9                  calculatedInfo ← remove the record from position k in
         the            calculatedInfo.
10             } end if
11         } end for (k)
12     } while (calculatedInfo is not empty)
13 } end function
```

Code Segment 10. The pseudo code of grouping function

A followed example from the Figure 12, when the base groups are already prepared for the clustering, the Ranking list will remove the chosen group as Figure 13 shows. Therefore, the list won't have the N-O and P-Q. The first combination that will be checked is the M-N, N is in one group, therefore, M will be grouped into the group where N is. The next one on the stack is O-Q, but this combination will be skipped because of the O and Q both are already in other groups. After O-Q have checked, the M-P, N-Q, M-Q…and so on will be skipped until N-R. N-R will be grouped into where N is too. Another exception is – what if both chromosomes are not chosen? In this algorithm, this combination will be skip too. After all the combination are checked on the list, check this list again until no chromosome left.

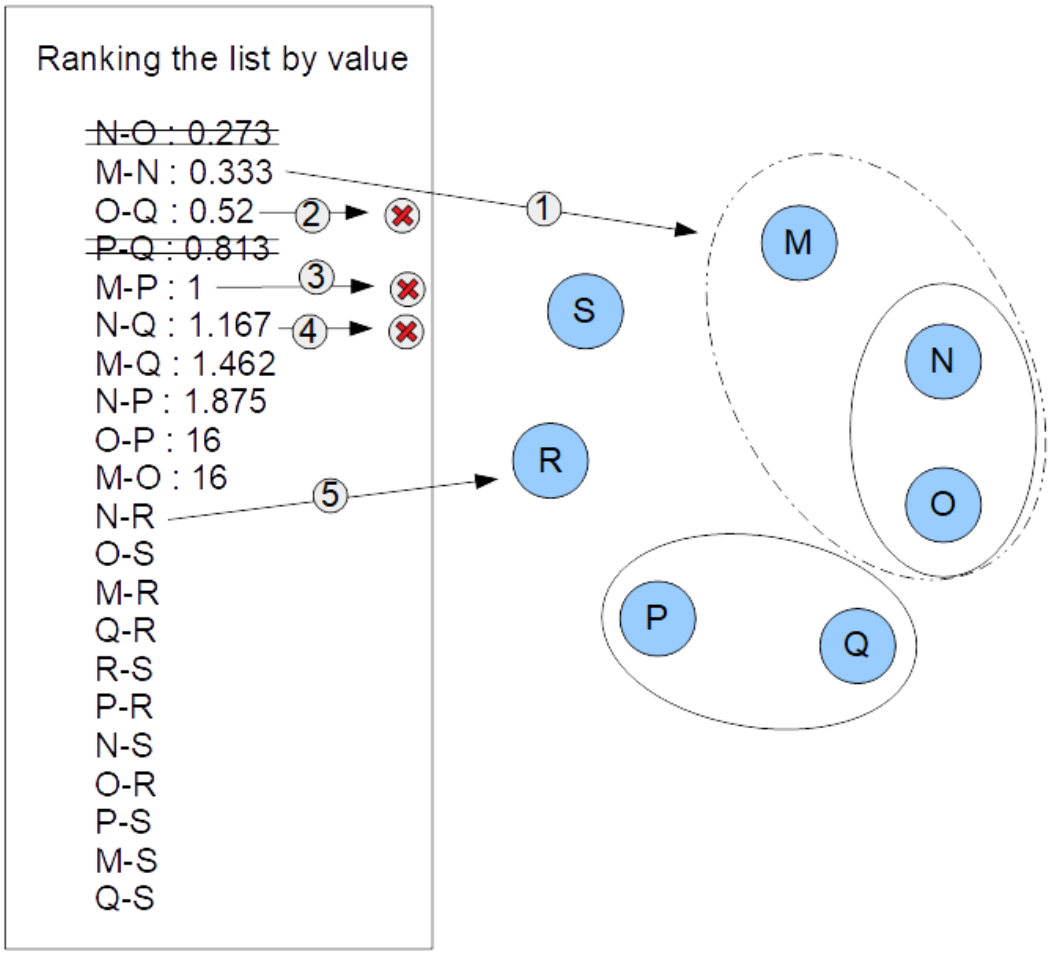


Figure 13. Clustering chromosomes into the group they belong to

3.4 Annotations suggestion provider bio-inspired algorithm

The proposal to invent the GRACE algorithm is providing students some useful suggestions as Code Segment 11 shows. In the line #3, algorithm locate where student annotated. The line #4 and #5 find the last sentence of where the student takes an annotation. This research only checks the question mark (?), exclamation mark (!), and the end mark (.) for the end of a sentence. In the line #6, the algorithm gets the list who

are in the same cluster of this student. This list can limit the searching results from the database. The line #7 is looking for where is the suggestion spot candidates. The candidates is the all places that this student didn't take any annotation. The information that line #6 and line #7 get will became the conditions when searching the log record from the database.

```
1  Function choseSuggestion() {
2      String readingArticle ← the reading activity that student is
3  reading now.
4      String annotatePlace ← where student annotate right now.
       String endOfLastSentence ← base on the annotatePlace, find out
   where is
5              the end of the last sentence from the readingArticle. (only
   check
               the mark "?", "!", and ".")
6      String startOfLastSentence ← base on the endOfLastSentence, find
7  out where
               is the start of the last sentence from the readingArticle.
   (only check
               the mark "?", "!", and ".")
8      String clusterResult ← get the same cluster list that this student
   is.
       String suggestionSpot ← put where is not annotated by this student
9  in
               his/her last sentence (get the position between the
               startOfLastSentence and the endOfLastSentence will get the
   last
               sentence).
       String getSuggestion ← search from the log in the database, find
   out who
               annotated on the no-annotation spot of this student's last
```

```
sentence.
            Randomly get one record from the log.
}end function
```

Code Segment 11. Suggestion chosen function

An suggestion example is shown in Figure 14 and Figure 15, when a student reads, he/she may annotate on an article as the step 1 in Figure 14 shows. System find the last sentence as shown in the step 2 and find the no annotation spot as step 3.

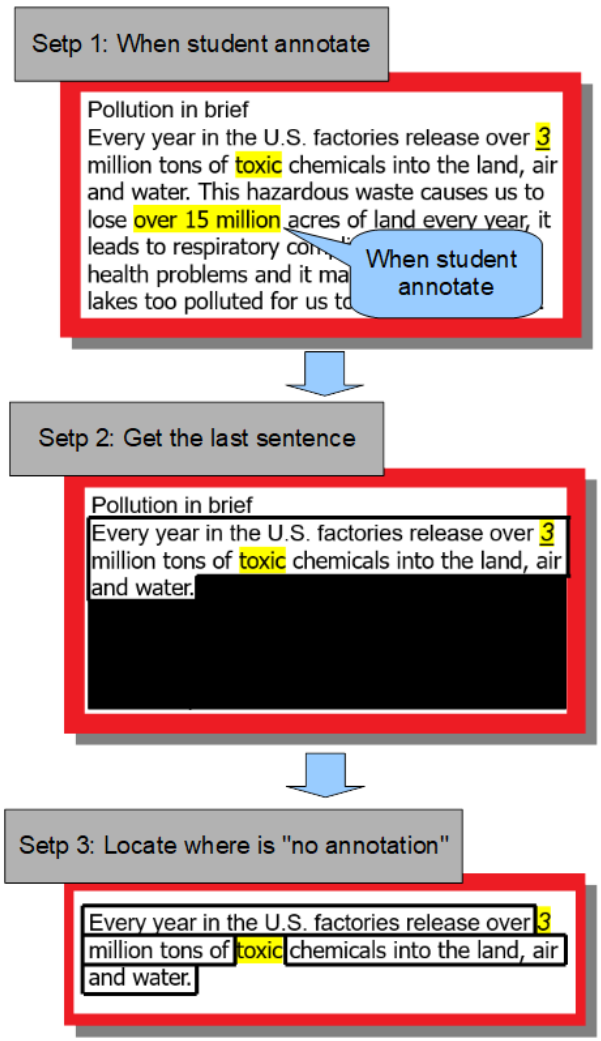


Figure 14. When student annotate on an article

After the system locates where is the no annotation place in the sentence, the suggestion algorithm will find the cluster that the student belongs to and find all the logs about what other students annotate on these spots in this group. The Figure 15 shows a followed example, when a student annotate, in the line #6 of the Code Segment 11 will find who is in the same cluster. If there are three other students in the same cluster, five annotations from the three students in the sentence are selected – the Student M uses the green highlight on the “3 million”; the student P uses blue highlight on the “over 3 million tons” and uses italic and red highlight on the “land, air and water”; the student Q uses the underline on the “U.S. factories”. Because the student already annotated on the “3”, therefore, any log has the position “3” will not in the search results. In the Figure 15 shows, there only 3 recommendation candidates left. The suggestion will randomly pick one log from the 3 valid search results.

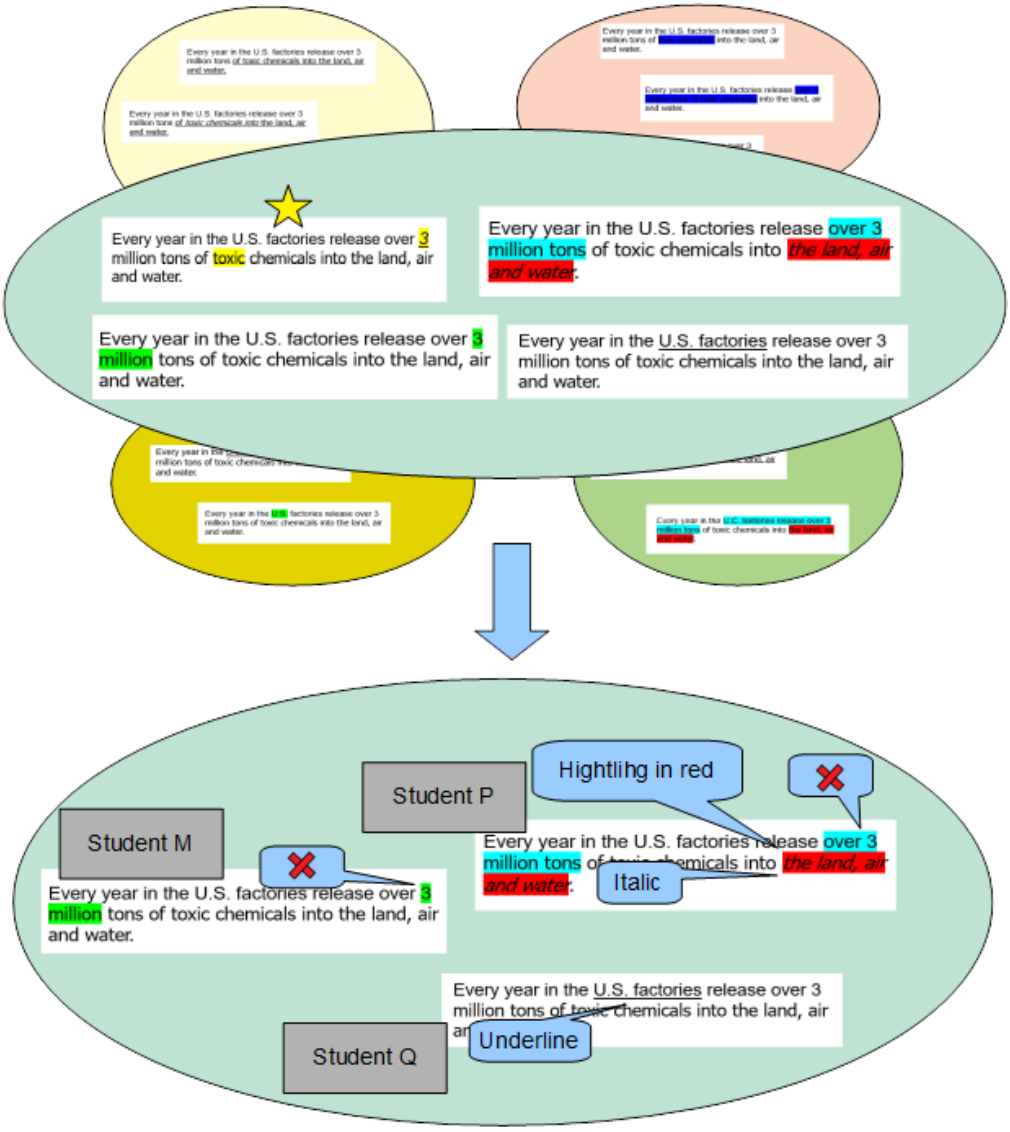


Figure 15. Find the valid log from database

Chapter 4. Online annotation system

4.1 Online annotation system architecture and working flow

Liu (2005) did a research and found that eighty-three percent of participants indicated that they increased read electronically. Similarly, Chrzastowski and Wiley (2015) found that students prefer the digital ones when they were offered options of having hard copies and digital ones. Bounie and colleagues (2013) also found that Amazon sells more digital books than printed books, especially in higher education according to the research had done by Lopatovska and colleagues (2014).

Almost in every courses, teachers will give reading assignments for students to read pieces of articles on papers or in a text. When students read, they usually make annotations on the reading materials. There are different annotations like sidebar notes, words or sentences highlighting, underlining and so on (Tashman & Edwards, 2011). Every student have preferred ways of making annotations; some of them may want to double underlining the words they thought important but others may choose to simply highlighting the words.

As we all know that no one's annotations are perfect and good for exam preparation, we also know that annotations may represent a person's perception of the importance and familiar degree toward the content. Under such circumstance, if three students' annotations are similar to each other and they are not overlooked or overrate their understandings for the content, then we might be able to say that they have similar degree of understanding and perceptions toward the content. However, we all know that no one is perfect and overlooking and self-overrating do happen. In such case, they might benefit from each other's annotations – which help them re-examine and double check why they

didn't make annotation on certain words but others did. They may, confirm that the missing annotation is on purpose because they are already too familiar with the content to annotate. Or perhaps in some cases, the missing annotation may ring the bell for them to make their annotation more complete.

For this reason, I develop GRACE (General Rapid Annotation Clustering Enhancement) platform that is composed of a frontend online annotation system and a backend bio-inspired clustering service (Chang, Kuo, Chang, Kinshuk, Kung, 2015). The frontend system allows teachers to create online reading activities and students to make annotations with a variety of ways – highlight, underline, bold, italic, and the use of sidebar notes and different colors. The backend service automatically group students according to their annotations from time to time.

With the backend service's help, the frontend system can prompt annotation recommendations for a student to double check in real time and provide grouping results for teachers to review so they may be able to identify potential learning problems their students may have. For instance, if a group of students' annotations show that they always annotate those irrelevant or not so important words and sentence, then the teachers can do a mini lecture in the class (in traditional settings) or post important notice on discussion board (in e-learning environments) to remind those students and make them be aware of their problems.

The chapters are organized in the following way. Sections 4.2 uses cases and screenshots to explain how teachers use the GRACE platform for creating online reading activities as the teacher side's Creating reading activity in the left side of the Figure 16 shows. Section 4.3 uses cases and screenshots to explain how students use GRACE

platform for making annotations on the material in a course as the students' side in the right side as shown in Figure 16. Section 4.4 shows how teachers can use GRACE platform to see their students' annotations, compare a group of students' annotations, and identify the potential learning problems a group of students may have as the teacher side's Reviewing annotations and Evaluating clustering results.

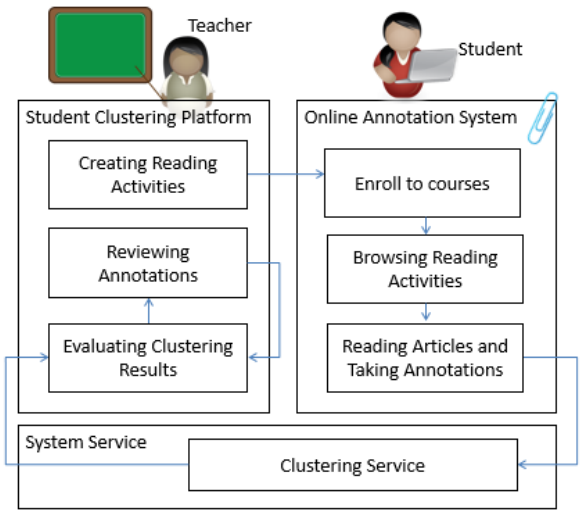


Figure 16. The system workflow of the Online Annotation System and the Student Clustering Platform

4.2 Teachers' use cases

GRACE platform can be free accessed online 1. Both of teachers and students can see similar page and self-register an account to use the platform as shown in Figure 17 As a teacher, teachers need to click the link after "I am a teacher" to teacher's page. After teachers register and sign in, they can see and manage their courses freely as shown in Figure 18. If they want to create a course to include reading activities for their students, they will need to enter correspondent course information, including course year, season/semester/term, number, and name.

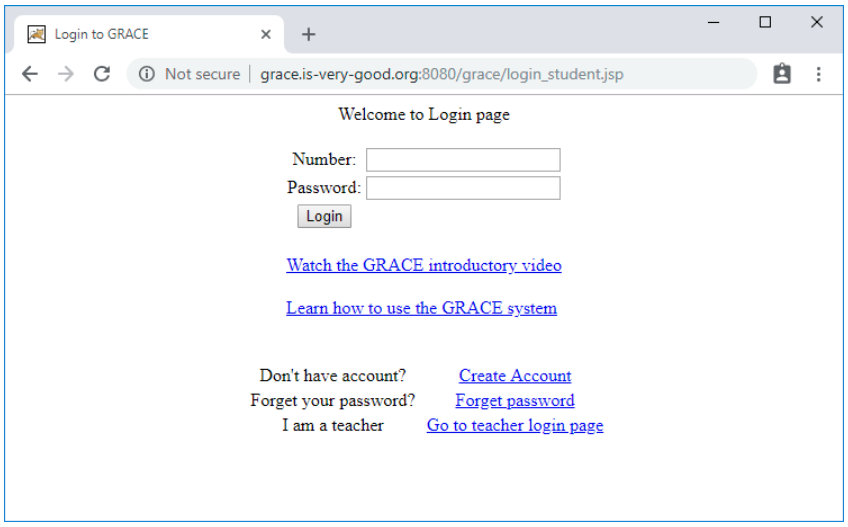


Figure 17. Login page that users can sign-in or self-register an account

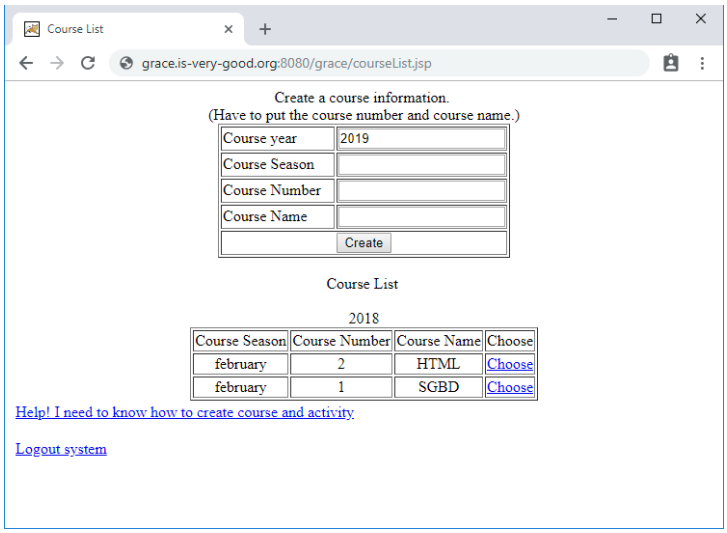


Figure 18. Course creation and management

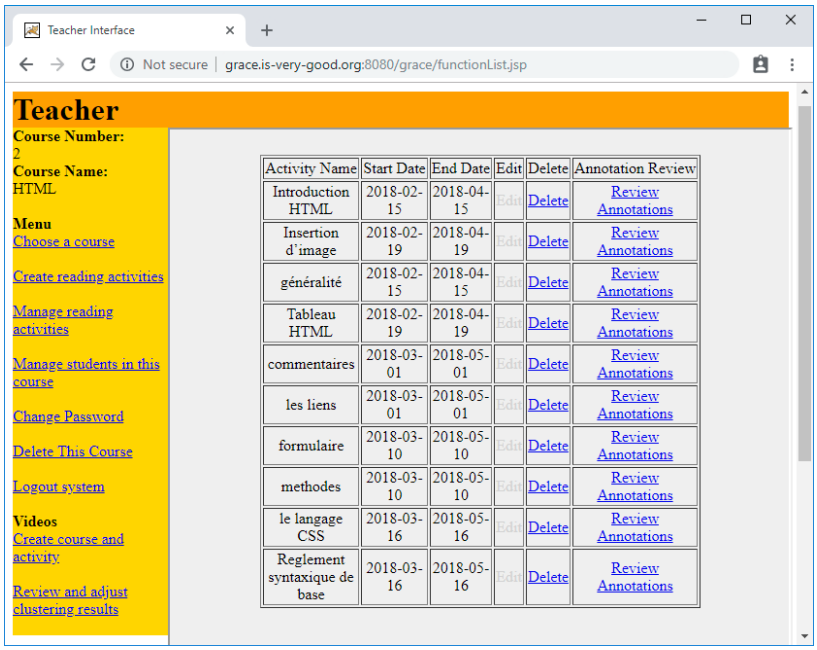


Figure 19. Reading activities that a course has

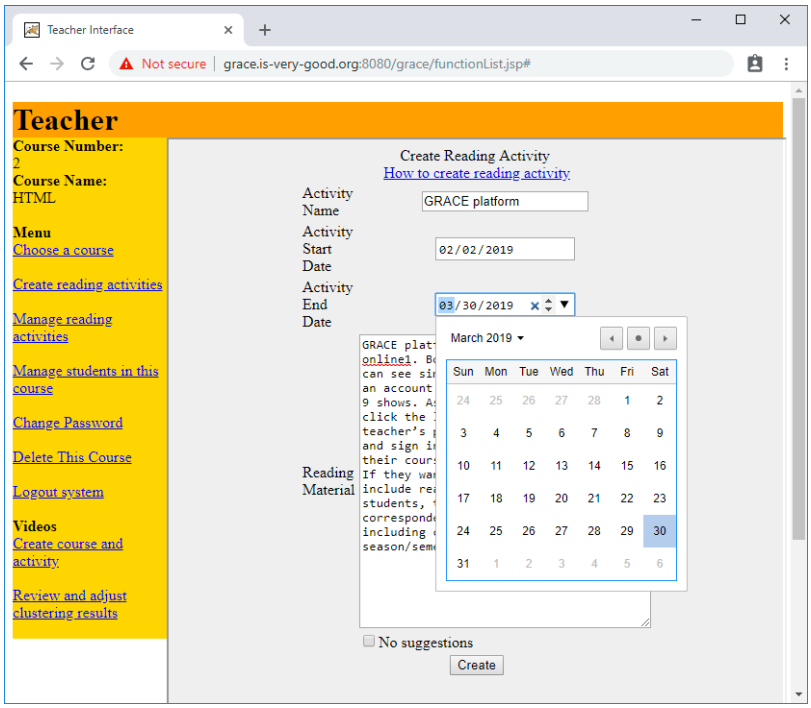


Figure 20. Teachers can create an online reading activity

They may directly go to a course that they created earlier, by click “Choose” link, to manage reading activities for their students as well as to review their students’ annotations. They can always exit from a course and enter to another one by clicking “Choose a course” link at left-hand side menu shown in Figure 19. A course can have many reading activities and teachers may create one themselves easily by clicking “Create reading activities” link at left-hand side menu. They will be able to see the user interface like Figure 20 shows. They need to, first of all, decide the start and end dates for the activity. Students will not be allowed to read before the start date and can still read the material but cannot make any annotation further after the end date. Teachers then need to fill in the form with the reading activity’s name and the reading material. At bottom of the page, there is a check box named “No suggestion”. If teachers just want to use GRACE platform for their students to read and make annotation but do not want their students to receive any annotation recommendations while reading, then they can check this box. If the box is unchecked, students will be prompted annotation recommendation every time when they make an annotation on the material. Of course, students can also disable the feature themselves at any time while reading and making annotations.

When anytime teachers want to check out all the reading activities they created for the course or edit particular reading activity, they can click “Manage reading activities” link at left-hand side menu to get back to Figure 19. From there, they can click “Edit” or “Delete” link for specific reading activity to update its information includes reading material or remove it permanently from the course. In addition, the “Edit” will be locked after the reading date started (as Figure 21 shows).

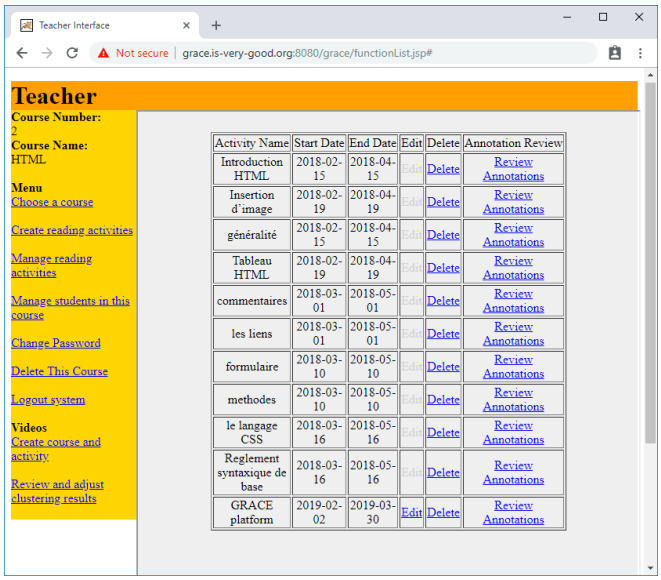


Figure 21. After reading start, the Edit button will lock

4.3 Students' use cases

After students register an account and sign in, they can see all courses and enroll any courses they want as Figure 22 shows. Having this flexibility is to reduce the workload that teachers have. With automatic enrollment feature, teachers can simply tell their students which course they should enroll after sign-in the platform. Of course, in some cases, if teachers want to self-add a particular student to their course, they can simply enter the student's account name by clicking "Manage students in this course" link at left-hand side menu as Figure 23 shows.

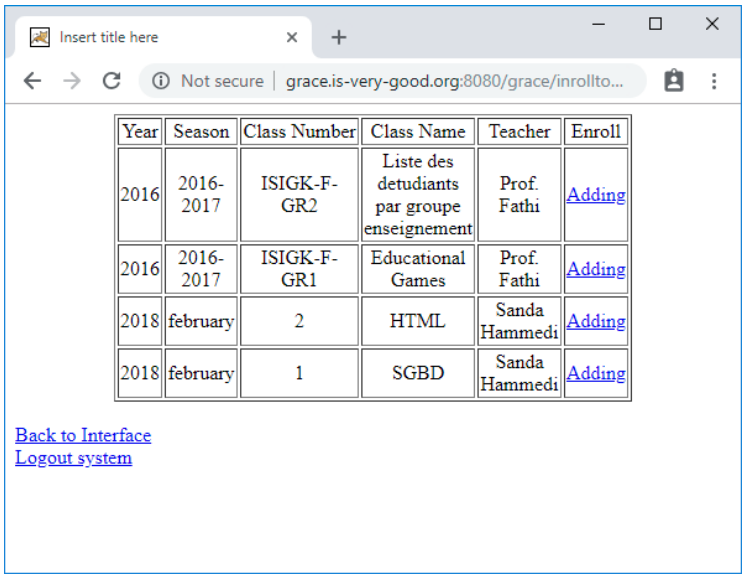


Figure 22. Students can self-enroll any course they want

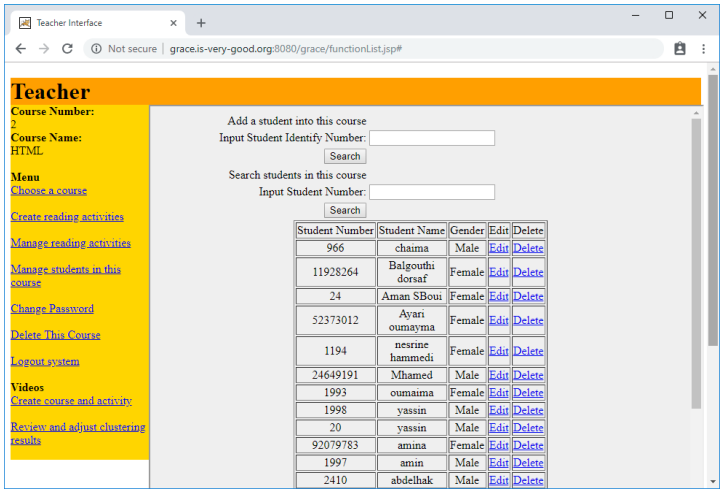


Figure 23. Teachers can manage students who enroll their course

When students sign in the platform, they can click “Reading Activities” link on the menu as Figure 24 shows to check what reading activities they have. As Figure 25 shows, they can find all reading activities from all courses they have enrolled. They may start working on any reading activity by clicking “Reading” link as long as it is in the time

frame between the start and end dates.

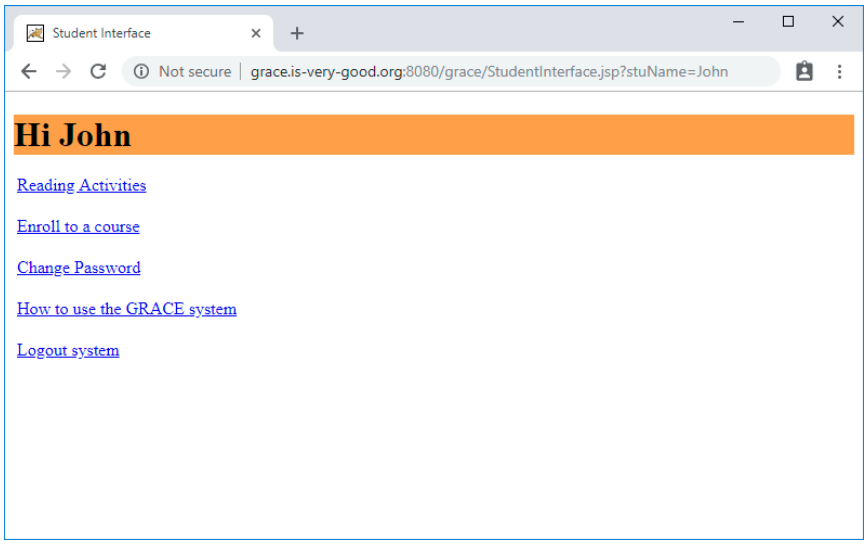
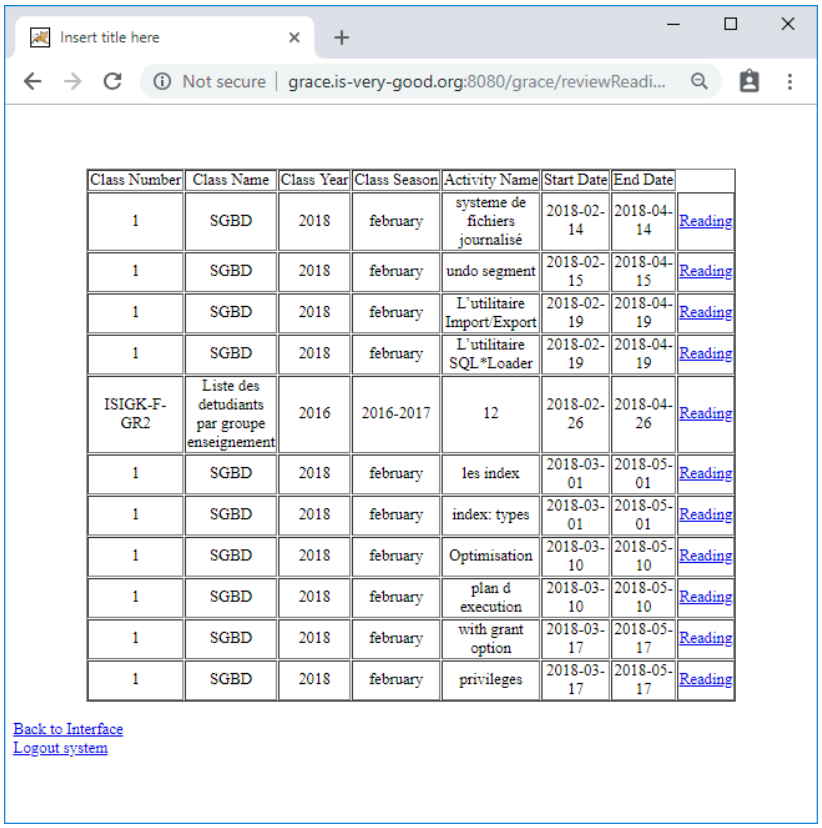


*Figure 24.*The menu students can use to switch among courses



Class Number	Class Name	Class Year	Class Season	Activity Name	Start Date	End Date	
1	SGBD	2018	february	systeme de fichiers journalisé	2018-02-14	2018-04-14	Reading
1	SGBD	2018	february	undo segment	2018-02-15	2018-04-15	Reading
1	SGBD	2018	february	L'utilitaire Import/Export	2018-02-19	2018-04-19	Reading
1	SGBD	2018	february	L'utilitaire SQL*Loader	2018-02-19	2018-04-19	Reading
ISIGK-F-GR2	Liste des detudiants par groupe enseignement	2016	2016-2017	12	2018-02-26	2018-04-26	Reading
1	SGBD	2018	february	les index	2018-03-01	2018-05-01	Reading
1	SGBD	2018	february	index: types	2018-03-01	2018-05-01	Reading
1	SGBD	2018	february	Optimisation	2018-03-10	2018-05-10	Reading
1	SGBD	2018	february	plan d execution	2018-03-10	2018-05-10	Reading
1	SGBD	2018	february	with grant option	2018-03-17	2018-05-17	Reading
1	SGBD	2018	february	privileges	2018-03-17	2018-05-17	Reading

Figure 25. Students can see all reading activities that different courses have

When students click an eligible reading activity, they will see the reading material as well as the annotation options on the screen as

Figure 26 shows. At the top panel on the screen, they can find that they can use four different color to highlight words in the reading material. They are also allowed to make selected words be underline, bold or italic. If they want, they can increase the selected words' font size or even attach a written note to the words. In the panel, there are two options for the students to set for their annotations: single choice and multiple choice. It is because students only annotate content with one kind of annotations, e.g., highlight or underline, in most of time. However, in any case the students want to highlight the selected words and underline them, they can choose "Multiple choice" instead. On the other hand, if they don't want to make any annotation but read the material, they can check "Reading Mode" checkbox to disable annotation feature.

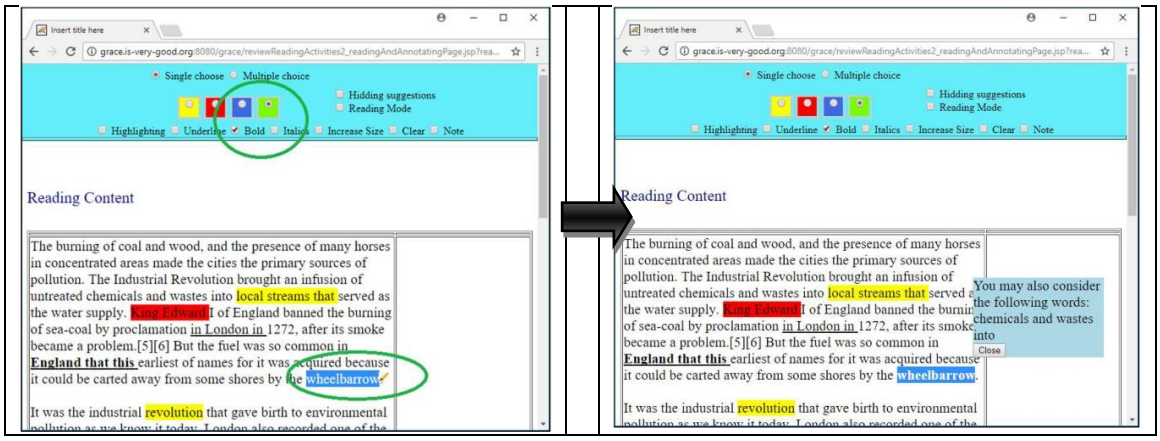


Figure 26. Annotations that a student can make on the material and then students can receive annotation suggestion

Unless the teachers chose to make a reding activity "Reading Mode Only" by checking "No suggestions" checkbox (see Figure 20), the platform will prompt an

annotation suggestion for students every time when they make an annotation on the reading material. As soon as the students annotate the word “wheelbarrow”,

Figure 26 shows that the platform tries to remind the students to review the importance of the words “chemicals and wastes into” that other students whose annotation behaviours similar to the students have. If the students think those words they were not annotating are not important, then they can click “Close” button to dismiss the reminder. On the other hand, if they suddenly find that “oh I missed that”, then they can make any necessary annotations they want and the platform may prompt another reminder for them if there is any. Last but not the least the students can always free to disable/enable the annotation suggestion feature by checking/unchecking “Hiding suggestions” so they will not be disturbed while reading.

4.4 Clustering results and benchmark

Teachers can click “Review Annotations” link any time(see Figure 19) to check their students’ annotations for a specific reading activity. Since the platform uses the bio-inspired clustering method “GRACE” behind the scene and continuously grouping students according to the similarity of their annotations, the teachers can see the grouping results when they enter the annotation review page as Figure 27 shows. When GRACE groups students, it tries to find different ways to do that. In the case of Figure 27 teachers can tell that GRACE is capable of grouping students into a single large group as well as two to six smaller groups. Teachers can choose 3-group result and click “Show ALL” button to see what student annotations look like and how students are grouped; for instance, Student #106 is in Group 1 when GRACE divides students into three groups.

Similarly, teachers can click to check out students and their annotations in particular group; for example, Figure 28 shows all students in Group 4 when GRACE divides everyone into three groups and Student #110 is one of them.

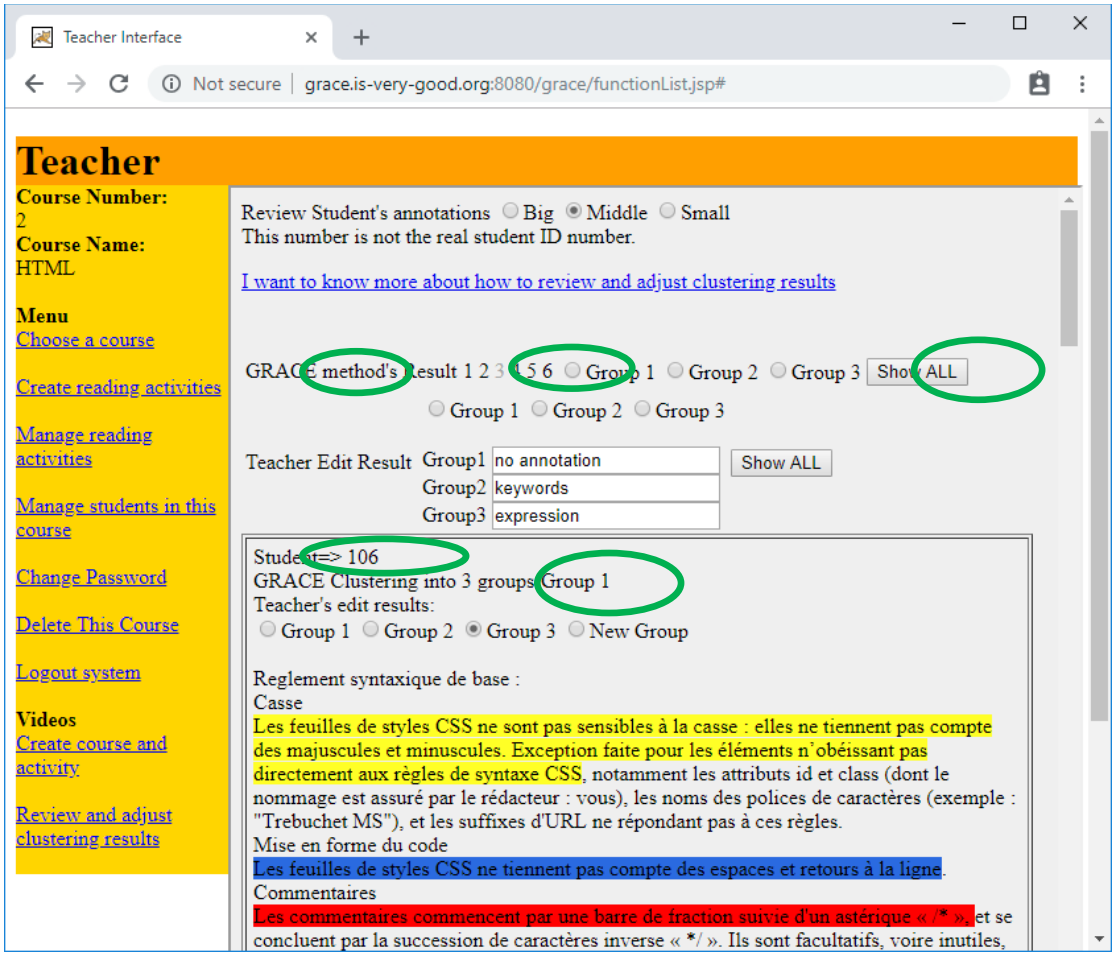


Figure 27. Student annotations and the grouping results

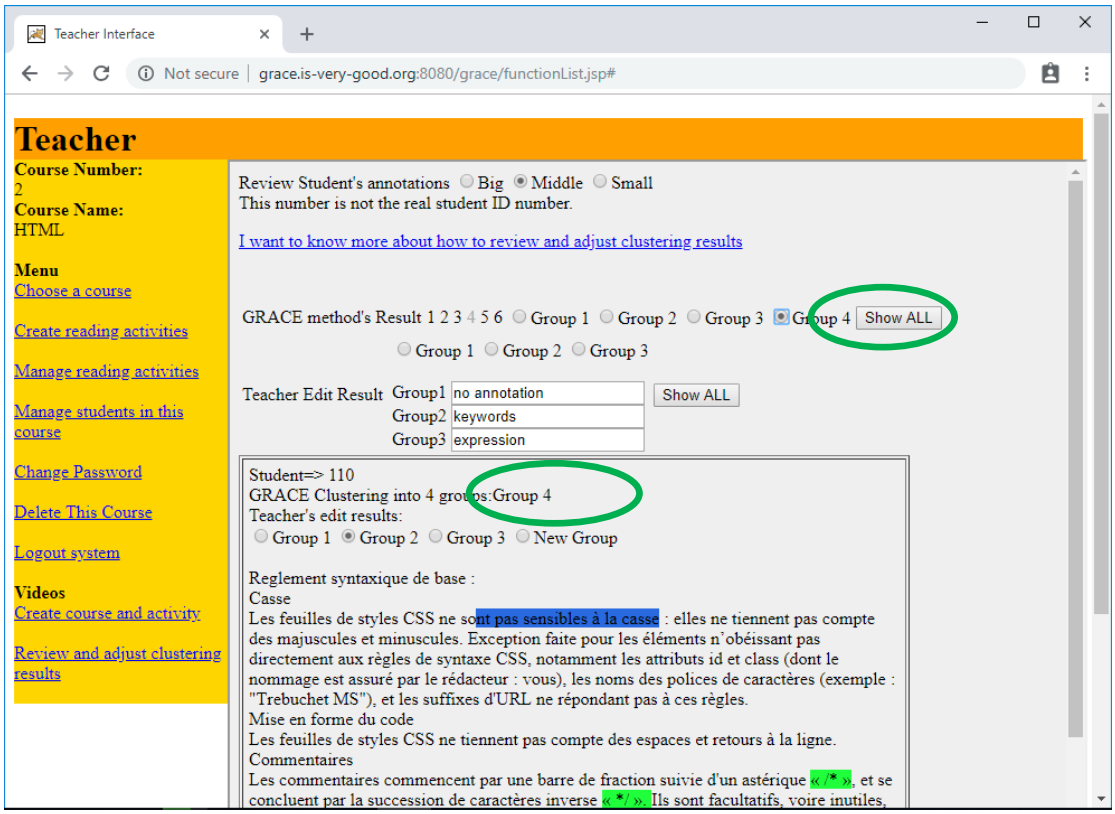
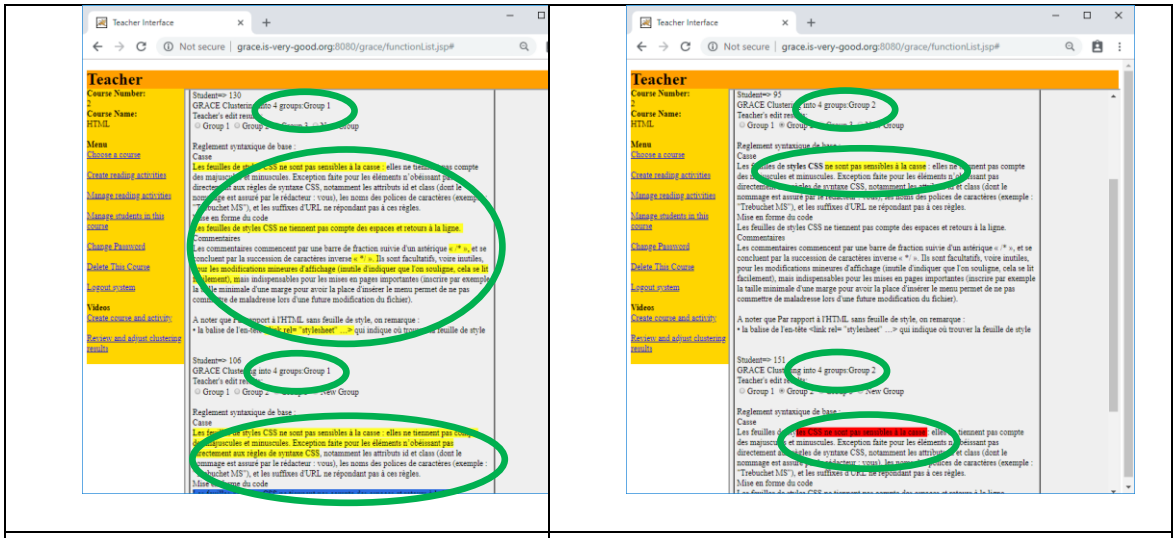


Figure 28. Teacher can choose to see students of particular group

With the platform, teachers can easily check out everyone's annotations and they might be able to identify the similarity and difference between the annotations made by students clustered into different groups, by GRACE. As

Figure 29 shows, the annotations obviously are different from Group 1 students to Group 2.



(a) Annotations that students in Group 1 have.

(b) Annotations that students in Group 2 have.

Figure 29. Teachers may easily to find the annotation difference that different group's students have

When teachers are not satisfying with the grouping results made by GRACE, they can always override it by clicking any existing group or even choosing to put a student into a new group. Figure 30 shows the teacher believes Student #101 doesn't belong to Group 1 in the 3-group clustering results made by GRACE and think the student's annotation is more similar to other students in Group 4. No matter whether or not teachers override the grouping results, they can label groups according to their perceptions toward the annotations that same group students have. Sometimes teachers may find that all students in a group have similar learning problems like overlooking fundamental concepts the reading material describes or ignoring the connections between key person and his or her invention, from student annotations. On the other hand, teachers may also be capable of identify common characteristics that a group of students have; for instance, students in a particular group can always catch the most important

events and relationships mentioned in the material. In both of the cases, teachers can label a group with their findings as Figure 31 shows. With the labels, teachers can effectively and put more efforts and energy on those groups of students to help them recognize their problems and make them learn better.

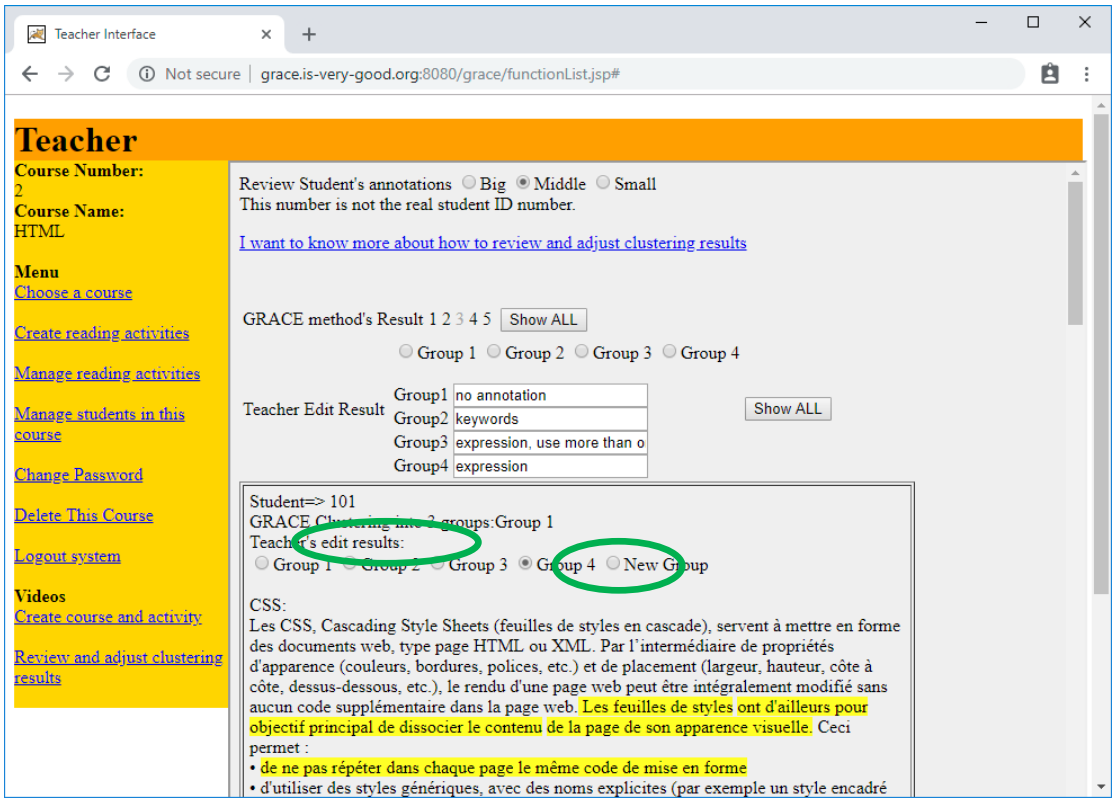


Figure 30. Teacher can override grouping results

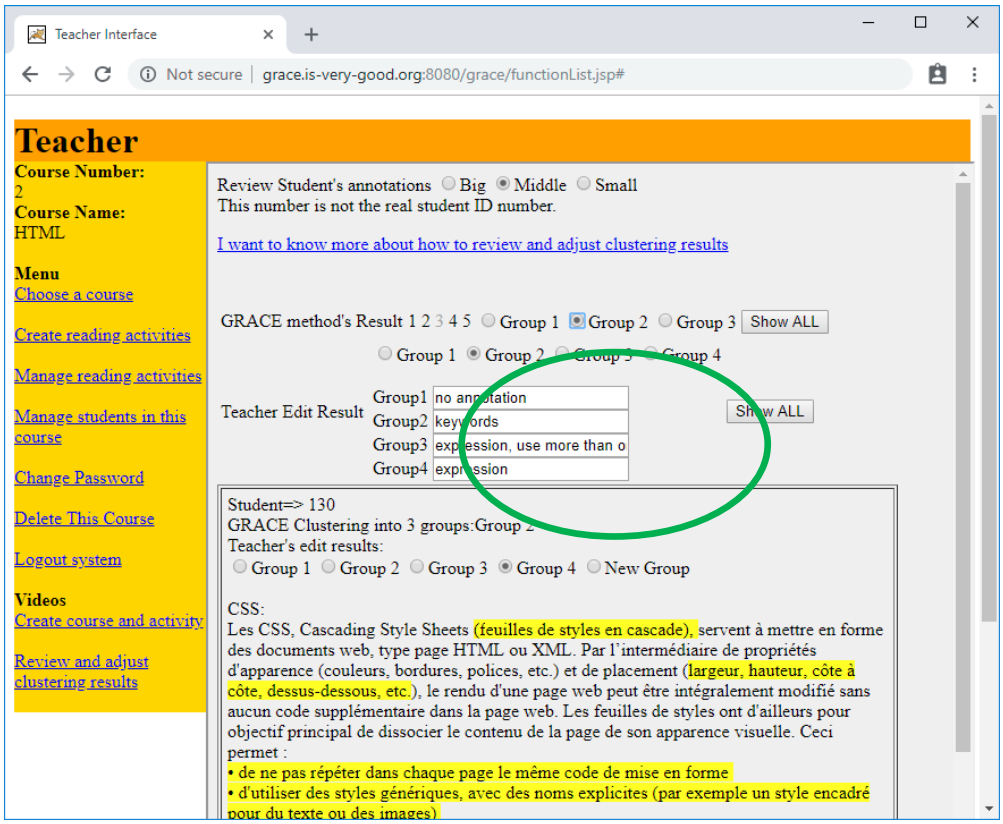


Figure 31. Teacher can identify the potential learning problems or the characteristics that students in a group may have

Chapter 5. Experiment and discussion

5.1 Research model and hypotheses

5.1.1 research model and questions. For the research model, I suppose some factors may affect student' attitudes toward to use the Online Annotation System. I think there are three directions may affect students' attitudes toward the system – students' general information, students' think/feel about the proposed system, and their experience on e-reader in the past. Students' general information is collecting students' basic information, their history of using e-readers/ browsers, and their using hobbit on using browsers. For the students' think/feel about the proposed system, I use the system usability questionnaire to ask students' think and feel about the proposed system. For students' experience on e-reader in the past, I use the Diffusion of Innovation (DoI) questionnaire to ask how they think about using e-reader.

The following are the research questions based on the research model:

Q1: Will students' Diffusion of Innovation affect their given score in System usability score?

Q2: Will the System usability score affect their perceived attitude toward to use Online Annotation System?

Q3: Will the Diffusion of Innovation affect their perceived attitude toward to use Online Annotation System?

Q4: Will the system usability internal factor affect another factor?

Within the research questions, I draw a macro view (as show in Figure 32) and this macro view contains the three directions detail that will be students' general information,

system's usability, and Diffusion of Innovation.

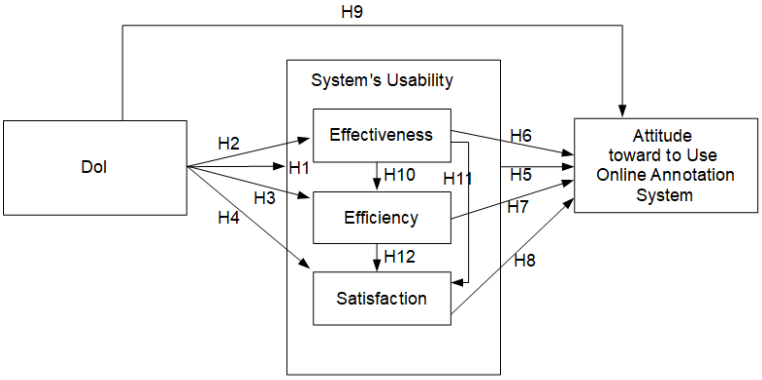


Figure 32. Macro view of research model

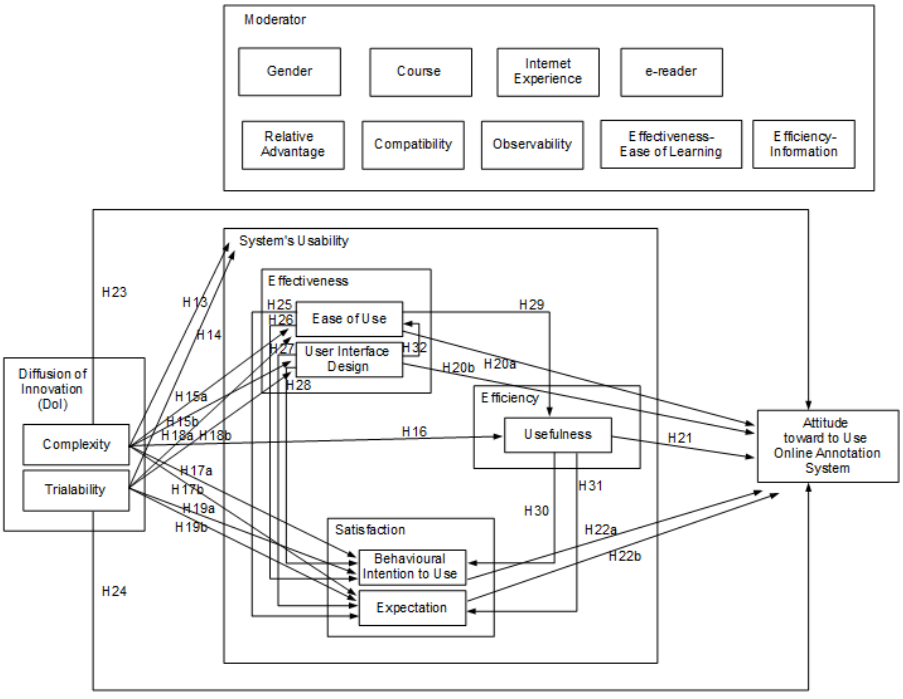


Figure 33. Micro view of research model

5.1.2 hypotheses. In this section, according to the research model I propose several hypotheses. I used the components that are described in Section 5.1.1 to develop the hypotheses from macro view and micro view of the model. The following is the research hypotheses for the questions above and would like to test them via the analysis of collected data:

H1. The Diffusion of Innovation will positively affect students' System usability score.

H2. The Diffusion of Innovation will positively affect students' perceived effectiveness when using Online Annotation System.

H3. The Diffusion of Innovation will positively affect students' perceived efficiency when using Online Annotation System.

H4. The Diffusion of Innovation will positively affect students' perceived satisfaction when using Online Annotation System.

H5. The system usability score will positively affect students' perceived attitude toward to use Online Annotation System.

H6. The perceived effectiveness will positively affect students' perceived attitude toward to use Online Annotation System.

H7. The perceived efficiency will positively affect students' perceived attitude toward to use Online Annotation System.

H8. The perceived satisfaction will positively affect students' perceived attitude toward to use Online Annotation System.

H9. The Diffusion of Innovation will positively affect students' perceived attitude

toward to use Online Annotation System.

H10. The perceived effectiveness will affect students' perceived efficiency when using Online Annotation System.

H11. The perceived effectiveness will affect students' perceived satisfaction when using Online Annotation System.

H12. The perceived efficiency will affect students' perceived satisfaction when using Online Annotation System.

H13. The perceived complexity will positively affect students' System usability score.

H14. The perceived trial ability will positively affect students' System usability score.

H15a. The perceived complexity will positively affect students' perceived ease of use.

H15b. The perceived complexity will positively affect students' perceived user interface design.

H16. The perceived complexity will positively affect students' perceived usefulness.

H17a. The perceived complexity will positively affect students' perceived behavioural intention to use.

H17b. The perceived complexity will positively affect students' perceived expectation.

H18a. The perceived trialability will positively affect students' perceived ease of use.

H18b. The perceived trialability will positively affect students' perceived user interface design.

H19a. The perceived trial ability will positively affect students' perceived behavioural intention to use.

H19b. The perceived trial ability will positively affect students' perceived expectation.

H20a. The perceived ease of use will positively affect students' perceived attitude toward to use Online Annotation System.

H20b. The perceived user interface design will positively affect students' perceived attitude toward to use Online Annotation System.

H21. The perceived usefulness will positively affect students' perceived attitude toward to use Online Annotation System.

H22a. The perceived behavioural intention to use will positively affect students' perceived attitude toward to use Online Annotation System.

H22b. The perceived expectation will positively affect students' perceived attitude toward to use Online Annotation System.

H23. The perceived complexity will positively affect students' perceived attitude toward to use Online Annotation System.

H24. The perceived trial ability will positively affect students' perceived attitude toward to use Online Annotation System.

H25. The perceived ease of use will affect students' perceived expectation.

H26. The perceived ease of use will affect students perceived behavioural intention to use.

H27. The perceived user interface design will affect students' perceived expectation.

H28. The perceived user interface design will affect students' perceived behavioural intention to use.

H29. The perceived ease of use will affect students' perceived usefulness.

H30. The perceived usefulness will affect students perceived behavioural intention to use.

H31. The perceived usefulness will affect students' perceived expectation.

H32. The perceived user interface design will affect students' perceived ease of use.

5.1.3 questionnaire design: The Diffusion of Innovation questionnaire.

Diffusion of Innovation (DoI) is a questionnaire to understand what the level of new technology acceptance for a person is. It is a questionnaire to know is this person an early adopter or major adopter type. For example, when an innovation comes out, this person will want to use/try it right away, seen many people use it and then he or she want to use it too, or never want to use it. Rogers, E. M., (2003) used five categories to describe the difference acceptance levels – innovators, early adopters, early majority, late majority, laggards. In Park and Chen (2007) research, this research uses Diffusion of Innovation to understand participants' acceptance of the smartphone.

The questionnaire design in Park and Chen (2007) research adopted into the Diffusion of Innovation design for this research. This research adopted the Diffusion of Innovation questionnaire design from Park and Chen's research (2007), because of all the

items in Diffusion of Innovation questionnaire have the reliability and validity analysis. This research chooses five factors – the Relative Advantage, Compatibility, Trialability, Observability, and Complexity. The Relative Advantage is asking if he/she have seen or use similar kind of innovation before, will he/she want to try the now one now. The Compatibility is asking if he/she thinks this innovation can help with his/her work. The Trialability is asking if he/she has a lot of opportunity can try it out or see other people use it; the Observability is asking he/she does this innovation is common to see in his/her life. The Complexity is asking is this innovation is easy for he/she to learn and use.

This evaluation plan includes two five-point Likert-scale questionnaires. The first questionnaire is Diffusion of Innovation questionnaire (DoI) and this questionnaire has 24 items. The purpose of Diffusion of Innovation questionnaire is to get idea of students' experiences and thoughts of using any kind of e-reader applications so the connection between their perceptions toward the system and their attitudes toward e-reader applications are found. For instance, will students who think the use of e-reader application can help them learn better make them perceived more positive toward the Online Annotation System? When I design the Diffusion of Innovation questionnaire, I decide to use other research's' questionnaire. I chose Quadir et. al. (2017) research for the Relative Advantage, Compatibilty, Complexity, and two items of the Triability. I take 3 items of the Triability and the 2 times for the Observability from Park and Chen's (2007) research. The factors, correspondent items, and the source of the study that Diffusion of Innovation questionnaire has to be found in

Table 3. For the Compatibility factor and the Observability, this research creates additional items for both factors (item 9, 21, 22, 23, and 24). Because I don't know how

much items will be removed after I analyze the collected data. Therefore, I create more items to make each factor have at least four items for analysis.

Table 3.

Diffusion of Innovation questionnaire

Factor	Items	Source Studies
Relative Advantage	1. Using an e-reader application with annotation functions enables me to understand the key concepts of the reading activities more quickly.	(Quadir et. al., 2017)
	2. Using an e-reader application with annotation functions improves the quality of annotations I make.	(Quadir et. al., 2017)
	3. Using an e-reader application with annotation functions makes easier to do reading activities.	(Quadir et. al., 2017)
	4. Using an e-reader application with annotation functions improves my learning performance.	(Quadir et. al., 2017)
	5. Using an e-reader application with annotation functions gives me greater control over my study schedule.	(Quadir et. al., 2017)
Compatibility	6. Using an e-reader application with annotation functions is compatible with all aspects of my study in school.	(Quadir et. al., 2017)
	7. Using an e-reader application with annotation functions is completely compatible with my current study in the class.	(Quadir et. al., 2017)
	8. I think that using an e-reader application with annotation functions fits well with the way I like to study.	(Quadir et. al., 2017)
	9. Using an e-reader application with annotation functions fits well with the device I prefer to use.	Self-developed
Complexity	10. My interaction with e-reader application with annotation functions is clear.	(Quadir et. al., 2017)
	11. My interaction with e-reader application with annotation functions is understandable.	(Quadir et. al., 2017)
	12. Learning to use an e-reader application with annotation functions is easy for me.	(Quadir et. al., 2017)
	13. Overall, I believe that an e-reader application with annotation functions is easy to adopt into my study.	(Quadir et. al., 2017)
Trialability	14. I've had a great deal of opportunities to try an e-reader	(Quadir et. al.,

	application with annotation functions for studying.	2017)
	15. I know where I can go to satisfactorily try out various uses of an e-reader application with annotation functions for studying.	(Quadir et. al., 2017)
	16. Before deciding whether or not to adopt an e-reader application with annotation functions, I would need to use it on a trail basis.	(Park & Chen, 2007)
	17. Before deciding whether or not to adopt an e-reader application with annotation functions, I would need to properly try it out	(Park & Chen, 2007)
	18. I would like to be permitted to use an e-reader application with annotation functions on a trial basis long enough to see what it can do.	(Park & Chen, 2007)
	19. It is easy for me to see people using e-reader application with annotation functions in the school.	(Park & Chen, 2007)
	20. I have had a lot of opportunities to see people using e-reader application with annotation functions to study.	(Park & Chen, 2007)
	21. It is easy for me to see others' annotations when we all use e-reader application with annotation functions.	Self-developed
Observability	22. I can see how others annotate the content of an article or book when we use e-reader application with annotation functions.	Self-developed
	23. I see people searching and finding the desired content quickly in an e-reader application.	Self-developed
	24. I can tell how different that I annotate an article or book from others when we use e-reader application with annotation functions.	Self-developed

The System Usability questionnaire

After students use the Online Annotation System doing their reading activities and before the semester is end, I ask students to fill out the usability questionnaire. This questionnaire asks students whether they think the system is useful for them as well as

their willingness of using the system later. The questionnaire has forty-one five-point Likert point items for four higher-level factors: Like, Effectiveness, Efficiency, and Satisfaction. Each higher-level factor has sub-factors. There are 5 items for Like factor; 16 items for the three sub-factors of Effectiveness factor; 10 items for the two sub-factors of Efficiency factor; and 10 items for the two sub-factors of Satisfaction factor. These items are coming from different questionnaires, I use the USE questionnaire from Lund (2001); I use the SUS questionnaire from Brooke (1996); I use the TAM questionnaire form AlQudah (2014); and I use CSUQ questionnaire from Tullis and Stetson (2004). I put all the items together and decide which item belongs to which factor. After I finish the classifying, for the Attitude toward to Use Online Annotation System factor only have one question, I develop four questions to make this factor have at least four items. For the User Interface Design also needs more items, I develop 3 more items for this factor.

The Like factor aims to ask students does they like the Online Annotation and willing to keep use it. The Ease of Learning factor aims to ask students does they think the Online Annotation System is easy to use and the Ease of Use factor aims to ask students does they think the system is easy to learn. The User Interface Design is asking students do they think the user interface is friendly for them. The Information factor is asking students the information that system provide is clear for them to find their needs. The Usefulness aims to ask students do they think this system help their study. The Behavioural Intention to Use factor is asking students do they like to keep using the system in the future. The Expectation factor is asking students do this system meets their needs.

Table 4.

Usability questionnaire

Factor (HL)	Sub-factor	Items	Source Studies
Att	N/A	1. I believe it is a good idea to use an Online Annotation System.	Self-developed
		2. Once I started using the Online Annotation System I found it is hard to stop.	Self-developed
		3. I like to use the Online Annotation System.	Self-developed
		4. As a student I like to use Online Annotation System to study.	Self-developed
		5. The Online Annotation System is pleasant to use.	(Lund, 2001)
Effectiveness	EoL	6. I could imagine that most people could learn how to use the Online Annotation System very quickly.	(Brooke, 1996)
		7. I needed to learn a lot of things before I could get going with the Online Annotation System.	(Brooke, 1996)
		8. Learning to use the Online Annotation System is easy for me.	(AlQudah, 2014)
		9. It is easy for me to remember how to do the reading activities in the Online Annotation System.	(AlQudah, 2014)
		10. I find it takes a lot of efforts to become skillful at using the Online Annotation System.	(AlQudah, 2014)
		11. I quickly became skillful with the Online Annotation System.	(Lund, 2001)
	EoU	12. I think the Online Annotation System is easy to use.	(Brooke, 1996)
		13. I think that I would need the support of a technical person to be able to use the Online Annotation System.	(Brooke, 1996)
		14. I find the various functions in the Online	(Brooke, 1996)

		Annotation System were well integrated.	
		15. I think there is too much inconsistency in the Online Annotation System.	(Brooke, 1996)
		16. I find the Online Annotation System very cumbersome to use.	(Brooke, 1996)
	UI	17. The interface of the Online Annotation System is pleasant.	(Tullis & Stetson, 2004)
		18. The user interface of the Online Annotation System is confusing.	Self-developed
		19. The Online Annotation System requires minimal steps for doing my reading activity.	Self-developed
		20. The logical design of this Online Annotation System is good, I have no difficulty in using it.	Self-developed
		21. The Online Annotation System is user friendly.	(Lund, 2001)
Efficiency	Information	22. Whenever I make a mistake while using the Online Annotation System I recover easily and quickly.	(Tullis & Stetson, 2004)
		23. The information (such as courese list, reading activity list, activity starting date, and activity ending date) provided by the Online Annotation System is clear.	(Tullis & Stetson, 2004)
		24. It is easy to find the information I needed.	(Tullis & Stetson, 2004)
		25. The information provided by the Online Annotation System is easy to understand.	(Tullis & Stetson, 2004)
		26. I find the Online Annotation System unnecessarily complex.	(Brooke, 1996)
		27. I can use the Online Annotation System without written instructions.	(Lund, 2001)
	Use	28. I believe I understand the reading materials more in-depth by using the Online Annotation System.	Self-developed
		29. Using the Online Annotation System gives	(AlQudah,

		me greater control over my time to finish my reading activities.	2014)
		30. The Online Annotation System enables me to accomplish the reading activity more quickly.	(AlQudah, 2014)
		31. Using the Online Annotation System improves my learning performance.	(AlQudah, 2014)
Satisfaction	BI	32. I think that I would like to use the Online Annotation System frequently.	(Brooke, 1996)
		33. I feel very confident using the Online Annotation System.	(Brooke, 1996)
		34. I plan to use an Online Annotation System in the future.	(AlQudah, 2014)
		35. Assuming that I have access to an Online Annotation System, I intend to use it.	Self-developed
		36. I intend to continue to use the Online Annotation System in the future.	(AlQudah, 2014)
		37. I will recommend others to use the Online Annotation System.	Self-developed
	Exp	38. This Online Annotation System has all the functions and capabilities I expect it to have.	Self-developed
		39. I expect that I would use the Online Annotation System in the future	(AlQudah, 2014)
		40. The Online Annotation System meets my needs.	(Lund, 2001)
		41. The Online Annotation System works the way I want it to work.	(Lund, 2001)

Att: Attitude toward to Use Online Annotation System; EoL: Ease of Learning; EoU: Ease of Use; UI: User Interface Design; Use: Usefulness; BI: Behvioural Intention to Use; Exp: Expectation

The Teacher's Interview questions

At the end of experiment end, students were asked to fill out another questionnaire, which asks for their experience of using the Online Annotation System. The details and explanation of the questionnaire can be found at

Table 4. At the same time, this research also asks the teacher to review the clustering results. If the teacher doesn't agree with the results, he or she can edit the results. After teacher the edit results (as shown in Figure 30 in page 75), I interview the teacher and asks her the questions to get idea of whether the system can help her to teach better, whether the clustering results meet her expectation, and how she thinks about the system. The detailed interview questions are listed in Table 5.

In teacher's interview questions, the question 1 to question 3 are asking teacher their thinking of the Online Annotation System. The question 5 to question 10 is asking do she think the Online Annotation System can help her for her teaching. In question 11 to question 14 is asking do she want to keep using this system in the future and what do she expect for this system can provide.

Table 5.

Interview questions for teacher

Questions
1. Do you like the Online Annotation System? Why? Can you give us some examples or reasons?
2. How do you feel about the system? Could you please elaborate it further?
3. Is the Online Annotation System easy to use?
4. What do you think about the management functions of reading activities?
5. What do you think about the management functions of clustering results?
6. What do you think about the clustering results of students?
7. Can you find student's learning problems when you review students' annotations? How? Can you share couple of examples with us?
8. Could you identify any behaviour feature that connects to potential learning problems? Do you see any features that can be used to distinguish students' learning problems?
9. Does finding students' learning problems is more quickly by reviewing the clustering results? Why? Could you please explain the reasons?
10. Do you think the use of the Online Annotation System improve your teaching performance? Please elaborate it further or give examples.

11. Would you want to use the Online Annotation System in the future? Why? Can you share with us your reasons?
12. Would you recommend others to use the Online Annotation System? Why? Can you share with us your reasons?
13. Does the Online Annotation System meet your needs? How? Could you please elaborate it further, perhaps with real cases?
14. Any feature that you think the Online Annotation System needs to have? Or is anything currently missing in the Online Annotation System?

5.1.4 moderators. In macro view, I mainly focus on the students' general experiment components and the factors of system usability. In user general experiment component of Figure 32, I suppose students' gender and age may affect their attitude toward to use the system. Different majors have different features, so I suppose that students who use computer more or learn more computer technology will have more positive attitude toward to use the system. Another idea is about when is the first time that students use the e-reader and browser may affect their attitude toward to use the system. Students purpose to use browsers and how much time they spent on it may affect their attitude. For more details about my research model, I draw a micro view (as show in Figure 33) of my research model. For the micro of the research model, I put the second layer factors of the system usability and the factors of the Diffusion of Innovation in the Figure 33.

In the questionnaire, I ask a few questions for the moderators. I suppose different gender could have different acceptance for the system because most people believe males play games and accept new technology easier than females; therefore, male students might have more positive attitudes toward the Online Annotation System than female. Therefore, gender is one moderator in the research model I would like to analyze.

The experiment plan finds two courses participants, the participants in HTML course are freshman. I suppose the course for freshmen will be more general concept. The Database Management Systems course are sophomore and the course should have the practice skill than the general concepts. Students might have different attitude toward to the Online Annotation System when using it when reading generic concept and practical skill articles.

Internet experience is another moderator because I believe students' experience in Internet usage will affect their attitudes toward the Online Annotation System. Therefore, the questions in the moderator includes "when was the first time you used the internet?", "when was the first time you use browsers?", "how many hours do you use internet in week days?", and "what are you doing when you use internet?".

I am also interested in students' past experience in e-reader because students who start using e-reader in their earlier age might have higher Diffusion of Innovation score as well as giving more positive attitude toward the Online Annotation System. Therefore, I ask students when did they first heard of the e-reader and when was the first time they use the e-reader.

5.2 Experiment design

5.2.1 experiment plan. This research has two classes of students to participate the experiment and both classes will use the same stage to verify the hypotheses. The research plans to have a quasi-experiment. The experiment uses a mixed methodology to verify 32 hypotheses. I will analyze the data with quantitative research methodology. The experiment has four stages. After students finish their reading activities, there is a quiz for

students.

Stage 1. Collecting Students' e-reader experience

In the beginning of the experiment, the teacher will give all students a Diffusion of Innovation questionnaire to understand their e-reader using experience and student's background information (e.g., gender).

Stage 2. Students can Use Any Way They Like to Read

At this stage, the teacher will assign two reading activities to each class. Students can use any way they like to read the reading activities. For example, they can print the articles out on a paper or use any devices such as smart phone/tablet/laptop to read the articles. This stage will be a base line for before using the Online Annotation System.

Stage 3. Reading Article with Online Annotation System

At this stage, the teacher will assign two reading activities to all students every week for four weeks. Students use the Online Annotation System to read and to annotate, if they want.

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.

Stage 4. Evaluating the Proposed Methods

After students finishing all reading activities, the researchers will ask students to

complete a Usability 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 the teacher can find out students' potential learning problems via the reviews of students' annotations.

5.2.2 data collection. The perceived usability towards the proposed system from both students and teachers are necessary to be known. This section talks about the experience and process that I had in a Tunisian university to execute an evaluation study. I start from introducing how to recruit participants include teacher(s) and students to explaining the stages that both teacher(s) and students would be done.

First, the teachers and students who are teaching and studying in Higher Institute of Computer Science and Management of Kairouan (ISIGK), Kairouan, Tunisia, are the potential participants of the evaluation study. The courses that teachers teach include 85 students in Database management system (in French is SGBD) and 20 students HTML course in second year and first year of the college student. I first approach to the teachers and ask for their willingness of adopting the proposed system in their courses so students need to use the Online Annotation System for reading activities and the student clustering results would be provided for them. I chose two different courses in the same major. The reason for choosing the Database Management System is that I suppose the course is for sophomore students and support more advanced concepts in computer science discipline. The HTML course is for freshmen students and I suppose the course for freshmen only focuses on the basic course in the computer science with more practical skills.

Table 6.

Students' study year and total students in each course

Course name	Course Level	Total students in the course
Database Management Systems	Sophomore	85
HTML course	Freshmen	20

5.3 Students annotation data analysis

5.3.1 precision and recall analysis. The experiment recruited one teacher in one university in Tunisia. The teacher taught two classes in February to March, 2018; the courses are Database Management Systems and HTML. In the section 5.2.1, the Stage 1 is fill out the questionnaire. The reading activities in Stage 2 is allowed students use any way they like to read the reading activity, including print out the reading activity that teacher provided, read the reading activity on the computer, smart phone, or their pad. The reading activities in the Stage 3 need to use Online Annotation System to read, but system did not collect any behaviour. The reading activities in Stage 3 is for students getting familiar with the Online Annotation System. System start collecting students' behaviour in Stage 4, Stage 4 is the stage that I ask teacher to group students' annotations into different groups.

The experiment in Stage 4 was asking teacher classify students into groups based on students' annotations on the articles; teachers could decide how many groups he would like to classify into based on his experience. The system also collects students' annotation behaviour on the Online Annotation System in this stage. However, not all students took annotations on the reading activities. As Table 7 shows, out of 85 enrolled students in the Database Management Systems class, there are 32 students have used the Online

Annotation System to read article “plan d execution” (activity #131, execution plan) but only 10 of them have made annotation in the system. Table 2 also list the number of students use the Online Annotation System in reading and annotation in other reading activities in the stage 4.

Table 7.

Descriptive statistics of students’ annotation on a reading activity in the stage 4 of the experiment

Course	Activity name (in French)	Activity name (in English)	Activity ID	# of recorded students	# of chromosome with/without annotation
Database Management Systems	plan d execution	execution plan	131	32	10 / 22
	Optimisation	Optimization	132	27	14 / 13
	privileges	privileges	137	22	13 / 9
	with grant option	with grant option	138	23	11 / 12
HTML	formulaire	form	133	19	10 / 9
	methodes	methods	134	19	11 / 8
	le langage CSS	CSS language	135	15	10 / 5
	Reglement syntaxique de base	Basic syntax rules	136	14	12 / 2

When cluster algorithm running, students’ annotations change to different formats chromosomes. In previous research, researchers use the bit-string chromosome on word-based format to represent students’ annotations. This research uses the UTF-8 to restore student’s annotations, but the GRACE algorithm uses the Huffman code for the clustering method. The different format not just the represent methods is different, but also the length has the difference too. Therefore, Table 8 lists the different formats length. When designing the GRACE algorithm, the idea comes from the bit-string chromosomes from the previous research. The research extends the word-based bit-string represent way

to alphabet-based for representing more detail information than the word-based bit-string. This research wants to represent more annotation ways, therefore, extend the alphabet-based bit-string again and the Huffman is the transformed alphabet-based string. In the Table 8 has a question, if the Huffman-string extends from the alphabet-based bit-string, why these two formats has different length? The alphabet-based extends from previous research, the method that previous research didn't count the space as an annotation bit. The Huffman-string wants to represent students' annotations for entire article, the space could annotate by some marks too.

Table 8.

Students' annotation chromosome length on different format

	Activity ID	Alphabet	Word	Huffman
Database Management Systems	131	1725	303	2141
	132	1436	263	1699
	137	1229	231	1460
	138	914	174	1090
HTML	133	823	156	979
	134	1334	265	1599
	135	1023	196	1219
	136	1054	210	1264

To understand the accuracy and efficiency of the clustering strategy in GRACE, I use the annotation groups classified by the teacher as the benchmark to compare the precision; recall, F2, F0.5, and average calculate time of the GRACE as well as the four methods proposed by Ying. This research organized the word-based chromosomes analysis and alphabet-based chromosomes analysis, but this research only focus on discussing the alphabet-based chromosomes analysis. Because of the word-based

chromosomes is the original method, I organize the word-based chromosome analysis. However, the word-based chromosome analysis is too different with the GRACE method, therefore, the organized the word-based bit-string chromosomes for the Database Management Systems and HTML course are attached in the Appendix C.

The GRACE algorithm is using the Huffman code to do the clustering. The Standard, Quantitative, Cosine, and Diffusion are using the alphabet-based bit-string format. The previous research is using word-based and this research extend the word-based to the alphabet-based method. Because the alphabet-based can't represent different annotation ways, this research use the alphabet-based as a foundation and use the Huffman code to represent students' annotations. Therefore, it is comparable for the previous research methods in alphabet-based and the transformed Huffman code base on the alphabet-based.

The alphabet-based bit-string chromosome average time taking are listed in

Table 9. In the beginning, I would like to compare the efficiency of the five methods. The Standard method gets the quickest average calculating time in both course – 0.76 seconds in Database Management Systems course and 0.368 seconds in HTML course. The Quantitative method is the second best, which uses 0.777 seconds in Database Management Systems course and 0.378 in HTML course. The Cosine clustering method takes the longest time in calculating in both courses – 22.107 seconds in Database Management Systems course and 20.051 seconds in HTML course.

Table 9.

Have all the students: alphabet-based bit-string chromosomes average calculating times (seconds)

Course	Reading ID	131	132	137	138	Mean	SD
Database Management Systems	GRACE	1.096	1.708	1.440	0.872	1.279	0.369
	Standard	0.792	1.152	0.672	0.424	0.760	0.303
	Quantitative	0.732	1.212	0.656	0.508	0.777	0.305
	Cosine	16.564	28.752	26.072	17.040	22.107	6.226
	Diffusion	7.516	12.768	9.504	5.652	8.860	3.043
HTML	GRACE	0.604	1.064	0.972	1.140	0.945	0.237
	Standard	0.252	0.288	0.488	0.444	0.368	0.116
	Quantitative	0.268	0.336	0.512	0.396	0.378	0.104
	Cosine	14.748	26.928	15.496	23.032	20.051	5.917
	Diffusion	2.420	6.028	6.792	5.692	5.233	1.931

For the precision, recall, F2, and F0.5 values, After I see the Cosine method has the lowest value in the four previous methods, the GRACE algorithm has the lower value in precision (22.42%), recall (22.84%), F2 (22.6%), and F0.5 (22.37%) values than what Cosine have (precision: 40.40%; recall: 40.19%; F2: 40.15%; and F0.5: 40.28%). If pick the all lowest mean values HTML course, the precision in Standard has 38.6 7%, the recall in Quantitative has 35%, the F2 in Standard has 36.51%, and the F0.5 in Standard has 37.85, and the GRACE algorithm also have lower values than the lowest value in the previous methods (Precision: 29.84%, Recall: 31.23%, F2: 30.46%, and F0.5: 29.71%). The results in the Table 10 and

Table 11 show that the performance of GRACE algorithm is not very well when comparing with the algorithm in the previous study. Looking back to the algorithm design, I find out that the GRACE algorithm does not consider how to classify students when they have no annotation on the reading activity. If students did not make any annotation on the article, the GRACE algorithm calculates their GRACE value as 0. For example, if there are four students did not make any annotations and system want to choose three combinations for the group's foundation, two groups of the three will have one combination from two of the four no-annotations and another group will get another no-annotations combination.

Table 10.

Have all the students: alphabet-based chromosomes on Database Management Systems class

Method	Evaluation	131	132	137	138	Mean	SD
GRACE	Precision	15.38%	8.33%	34.38%	31.58%	22.42%	12.58%
	Recall	18.18%	13.46%	30.56%	29.17%	22.84%	8.35%
	F2	17.54%	11.99%	31.25%	29.62%	22.60%	9.35%
	F0.5	15.87%	9.02%	33.54%	31.07%	22.37%	11.84%
Standard	Precision	66.16%	82.50%	41.25%	44.85%	58.69%	19.31%
	Recall	55.11%	56.41%	45.83%	34.17%	47.88%	10.28%
	F2	57.02%	60.22%	44.84%	35.88%	49.49%	11.24%
	F0.5	63.61%	75.51%	42.09%	42.21%	55.86%	16.55%
Quantitative	Precision	94.64%	19.12%	39.58%	61.67%	53.75%	32.33%
	Recall	54.17%	25.00%	43.75%	53.33%	44.06%	13.56%
	F2	59.23%	23.55%	42.85%	54.81%	45.11%	15.95%
	F0.5	82.34%	20.06%	40.35%	59.80%	50.64%	26.64%
Cosine	Precision	34.38%	41.52%	50.00%	35.71%	40.40%	7.11%
	Recall	39.20%	45.03%	44.44%	32.08%	40.19%	6.01%
	F2	38.13%	44.28%	45.45%	32.75%	40.15%	5.89%
	F0.5	35.24%	42.18%	48.78%	34.92%	40.28%	6.58%
Diffusion	Precision	63.67%	52.27%	42.31%	66.67%	56.23%	11.16%

	Recall	60.42%	35.42%	43.75%	48.33%	46.98%	10.43%
	F2	61.04%	37.86%	43.45%	51.15%	48.37%	10.05%
	F0.5	62.99%	47.73%	42.59%	61.97%	53.82%	10.23%

Table 11.

Have all the students: alphabet-based bit-string chromosomes on HTML class

Method	Evaluation	133	134	135	136	Mean	SD
GRACE	Precision	8.33%	40.38%	53.13%	17.50%	29.84%	20.56%
	Recall	13.89%	31.25%	50.63%	29.17%	31.23%	15.07%
	F2	12.25%	32.73%	51.11%	25.74%	30.46%	16.18%
	F0.5	9.06%	38.15%	52.61%	19.02%	29.71%	19.46%
Standard	Precision	51.56%	35.42%	34.38%	33.33%	38.67%	8.64%
	Recall	35.00%	34.38%	34.38%	41.67%	36.35%	3.55%
	F2	37.40%	34.58%	34.38%	39.68%	36.51%	2.53%
	F0.5	47.10%	35.20%	34.38%	34.72%	37.85%	6.18%
Quantitative	Precision	64.06%	62.50%	38.89%	27.78%	48.31%	17.89%
	Recall	40.00%	40.63%	34.38%	25.00%	35.00%	7.23%
	F2	43.25%	43.68%	35.19%	25.51%	36.91%	8.54%
	F0.5	57.18%	56.42%	37.89%	27.17%	44.67%	14.68%
Cosine	Precision	41.07%	49.11%	50.00%	39.58%	44.94%	5.37%
	Recall	38.89%	62.50%	52.50%	45.83%	49.93%	10.05%
	F2	39.31%	59.27%	51.98%	44.43%	48.75%	8.73%
	F0.5	40.62%	51.31%	50.48%	40.69%	45.77%	5.92%
Diffusion	Precision	38.24%	89.02%	32.64%	38.64%	49.63%	26.40%
	Recall	30.00%	78.13%	34.38%	33.33%	43.96%	22.85%
	F2	31.35%	80.08%	34.01%	34.27%	44.93%	23.47%
	F0.5	36.25%	86.60%	32.97%	37.44%	48.32%	25.59%

Figure 34 shows four chromosomes, which do not have any annotations on the article. In the Ranking list, these no-annotation chromosomes will fill up the top of the ranking list. The first group is the A-B combination and second group will get the C-D combination. No matter how many groups need to be clustered, the no-annotation combinations will be chosen for the group foundation until the no-annotation combinations runout.

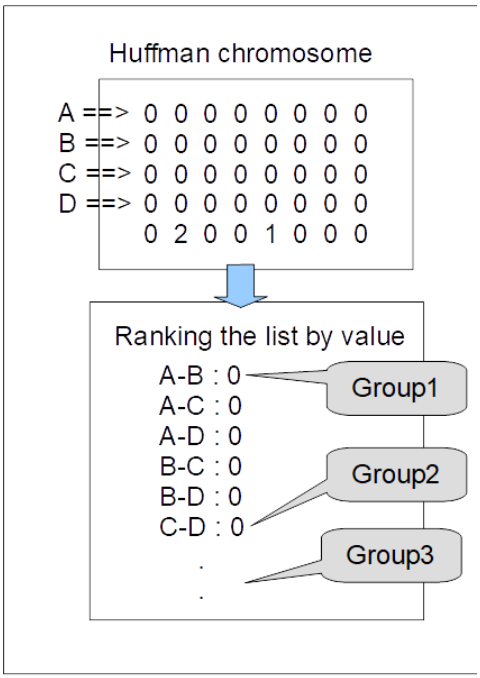


Figure 34. An example for choosing foundation from no-annotation chromosomes

The teacher also set up a “no annotations” group when he analyzed students’ annotation in each reading activity. The teacher put the articles with no annotation and the article only has annotations which might be those marked by students accidently in the “no annotations” group. Therefore, I remove the chromosomes in the “no annotations” group marked by the teacher and the group itself.

After remove the chromosomes in the “no annotations” group, the average time taken are listed in the Table 12. Regarding the efficiency of the five methods, the Quantitative method gets the fastest average calculating time in both course – 0.744 seconds in Database Management Systems course and 0.348 seconds in HTML course. The second fastest method is the Standard method, the Database Management Systems course gets 0.764 seconds and the HTML course gets 0.354 seconds. The least efficient

method is Cosine. The Cosine method gets 24.58 seconds in the Database Management Systems course and 20.031 seconds in HTML course. The GRACE takes 1.431 seconds in the Database Management Systems course and 1.467 seconds in HTML course. If I compare GRACE with the Standard and Quantitative, GRACE is slower but considering the chromosome complexity, a little bit longer is acceptable.

Table 12.

Alphabet-based bit-string chromosomes average calculating times (seconds) after removing non-annotations' chromosome

Course	Reading ID	131	132	137	138	Mean	SD
Database Management Systems	GRACE	1.080	1.560	1.336	1.748	1.431	0.288
	Standard	0.792	1.168	0.680	0.416	0.764	0.312
	Quantitative	0.744	1.164	0.676	0.392	0.744	0.319
	Cosine	26.504	28.548	26.212	17.056	24.580	5.123
	Diffusion	7.844	13.056	9.844	5.760	9.126	3.106
HTML	GRACE	1.480	1.012	0.948	2.428	1.467	0.683
	Standard	0.292	0.264	0.484	0.376	0.354	0.099
	Quantitative	0.296	0.316	0.456	0.324	0.348	0.073
	Cosine	14.764	26.816	15.372	23.172	20.031	5.926
	Diffusion	2.468	5.944	6.808	5.756	5.244	1.907

The analysis of the precision, recall, F0.5, and F2 in the alphabet-based results are shown in (Database Management Systems course) and Table 14 (HTML course). The precision values for Database Management Systems course the Cosine gets the highest value (42.18%), Diffusion (41.86%), GRACE (39.65%), Quantitative (38.44%), and Standard (29.51%). In HTML course, the highest value is the Diffusion (44%), Quantitative (35.06%), GRACE (33.07%), Cosine (30.47%), and Standard (26.49%). In this research Cosine gets 42.18% for the first place in Database Management Systems course, Cosine

could also get 30.47% for the fourth place in HTML course. For the Standard and Quantitative methods, these two methods stay in the similar value.

Table 13.

Remove no-annotations' chromosome: alphabet-based bit-string chromosomes on Database Management Systems class

		131	132	137	138	Mean	SD
GRACE	Precision	45.83%	47.50%	38.89%	26.39%	39.65%	9.60%
	Recall	37.50%	39.58%	33.33%	25.83%	34.06%	6.07%
	F2	38.92%	40.95%	34.31%	25.94%	35.03%	6.66%
	F0.5	43.88%	45.67%	37.63%	26.28%	38.37%	8.77%
Standard	Precision	45.83%	34.72%	11.11%	26.39%	29.51%	14.63%
	Recall	45.83%	33.33%	16.67%	25.83%	30.42%	12.33%
	F2	45.83%	33.60%	15.15%	25.94%	30.13%	12.92%
	F0.5	45.83%	34.44%	11.90%	26.28%	29.61%	14.27%
Quantitative	Precision	37.50%	51.39%	35.71%	29.17%	38.44%	9.34%
	Recall	37.50%	39.58%	33.33%	28.33%	34.69%	4.97%
	F2	37.50%	41.49%	33.78%	28.50%	35.32%	5.53%
	F0.5	37.50%	48.50%	35.21%	29.00%	37.55%	8.13%
Cosine	Precision	43.75%	40.28%	51.67%	33.04%	42.18%	7.74%
	Recall	43.75%	35.42%	52.08%	32.50%	40.94%	8.83%
	F2	43.75%	36.29%	52.00%	32.61%	41.16%	8.58%
	F0.5	43.75%	39.20%	51.75%	32.93%	41.91%	7.92%
Diffusion	Precision	29.17%	46.59%	47.92%	43.75%	41.86%	8.64%
	Recall	35.42%	35.42%	52.08%	40.00%	40.73%	7.87%
	F2	33.96%	37.20%	51.19%	40.70%	40.76%	7.48%
	F0.5	30.23%	43.83%	48.70%	42.94%	41.42%	7.88%

Table 14.

Remove no-annotations' chromosome: alphabet -based bit-string chromosomes on HTML class

		133	134	135	136	Mean	SD
GRACE	Precision	25.00%	33.93%	33.33%	40.00%	33.07%	6.16%
	Recall	25.00%	43.75%	40.63%	33.33%	35.68%	8.35%
	F2	25.00%	41.36%	38.92%	34.48%	34.94%	7.21%
	F0.5	25.00%	35.52%	34.57%	38.46%	33.39%	5.83%
Standard	Precision	30.95%	22.92%	27.08%	25.00%	26.49%	3.43%
	Recall	30.00%	25.00%	34.38%	25.00%	28.59%	4.52%
	F2	30.19%	24.55%	32.62%	25.00%	28.09%	3.96%
	F0.5	30.76%	23.31%	28.28%	25.00%	26.84%	3.33%
Quantitative	Precision	38.89%	45.83%	26.25%	29.29%	35.06%	8.98%
	Recall	30.00%	56.25%	37.50%	29.17%	38.23%	12.58%
	F2	31.44%	53.80%	34.54%	29.19%	37.24%	11.26%
	F0.5	36.71%	47.60%	27.93%	29.26%	35.37%	9.02%
Cosine	Precision	25.00%	37.50%	25.00%	34.38%	30.47%	6.44%
	Recall	25.00%	37.50%	34.38%	33.33%	32.55%	5.34%
	F2	25.00%	37.50%	31.98%	33.54%	32.00%	5.22%
	F0.5	25.00%	37.50%	26.44%	34.16%	30.78%	6.02%
Diffusion	Precision	38.89%	72.22%	26.25%	38.64%	44.00%	19.72%
	Recall	30.00%	62.50%	37.50%	29.17%	39.79%	15.60%
	F2	31.44%	64.23%	34.54%	30.67%	40.22%	16.09%
	F0.5	36.71%	70.04%	27.93%	36.28%	42.74%	18.65%

In the recall value for Database Management Systems course, the Cosine gets the highest value (40.94%), Diffusion (40.73%), Quantitative (34.69%), GRACE (34.06%), and Standard (30.42%). In HTML course is the Diffusion gets the highest value (39.79%), Quantitative (38.23%), GRACE (35.68%), Cosine (32.55%), and Standard (28.59%). Cosine value also suddenly drop as precision value. For the suddenly value drop, the performance between the reading activities have big gap or have similar values

need to be found. The standard deviation value for the Cosine are under 10% in both courses. on the contract, the Standard, Quantitative, and Diffusion all have standard deviation above 10%. Combining the mean value and standard deviation values, the Standard, Quantitative, and Diffusion all seems stable (the mean value for two courses doesn't have much gap) for the performance, but the standard deviation values show the gap could have big difference for individual reading activity performance. At first, Cosine's big gap mean value make me wonder is this method not stable, but the standard deviation value is small – the range is only 5% to 8%. The performance for Cosine maybe is not "not stable", it's performance could affect by different types courses. For the GRACE algorithm, the precision and recall values are in the middle of all these methods, the mean value for two courses is similar like Standard and Quantitative, the standard deviation value's range is 5% to 9% like Cosine's, and the average calculate time is quicker than Cosine and Diffusion. Maybe not as quick as Standard and Quantitative, but the time is close enough to these two methods. In addition, the GRACE algorithm can cluster more complex chromosome than Standard and Quantitative.

5.3.2 kappa analysis. To compare the agreement between teachers's grouping results and the clustering results in GRACE as well as the four methods proposed by Ying, this research uses the Cohen's Kappa (Landis & Koch, 1977) to evaluate the data I collected in the experiment. Kappa value can examine the agreement between two raters. For examining these five methods, seem teacher edits clustering results as a standard result. Labeling the groups for teacher's edit result and find the corresponding student is in which group in other method. After the kappa analysis, a list for how a Kappa

represent is given from Landis & Koch, 1997 as shown in Table 15.

Table 15.

Kappa value interpret from Landis & Koch (1997)

Kappa Statistic	Strength of Agreement
< 0.00	Poor
0.00-0.20	Slight
0.21-0.40	Fair
0.41-0.60	Moderate
0.61-0.80	Substantial
0.81-1.00	Almost Perfect

After the teacher groups students based on students' annotations, I manually map the clusters in the five methods to the groups assigned by the teacher. When I start do the mapping, I randomly start with any group. The mapping rule is based on which teacher edited group has higher proportion of the chromosomes in the clustering group; students' annotations will not affect the manually mapping. For example, when teacher cluster students into three groups, called Group T1, Group T2, and Group T3 as shown in the left-and side of Figure 35. The right hand side is the clustering results from one of the five methods implemented in this study, and the clusters are called the Group X, Group Y, and Group Z. The chromosome B, C and D in the Group X can correspond to the Group T1 of the teacher edit groups and A in the Group X can correspond to the Group T3 of the teacher edit groups. Because Group X has larger proportion items of what Group T1 (3/4 = 75%) then Group T3 (1/4 = 25%), Group X is mapping to Group T1. The E and F in the Group Z can correspond to the Group T2 and the I and J can be funded in the Group T3. In the Group Z, there are two sets of the chromosomes can correspond to two different

groups – E and F and correspond to Group T2 (2/5 = 40%) and I and J can correspond to Group T3 (2/5 = 40%). Because these two gets the same proportion, I will check the left side's groups. The E and F get the 2/4 (50%) of the Group T2 and the I and J get the 2/5 (20%) of the Group T3. Therefore, I will map the Group Z to the Group T2. However, in this case, I could get the same proportion again in the teacher edit groups. If I gets the same proportion in the system clustering results and the teacher edit groups, I will randomly correspond any two groups, because of every groups gets the same proportion.

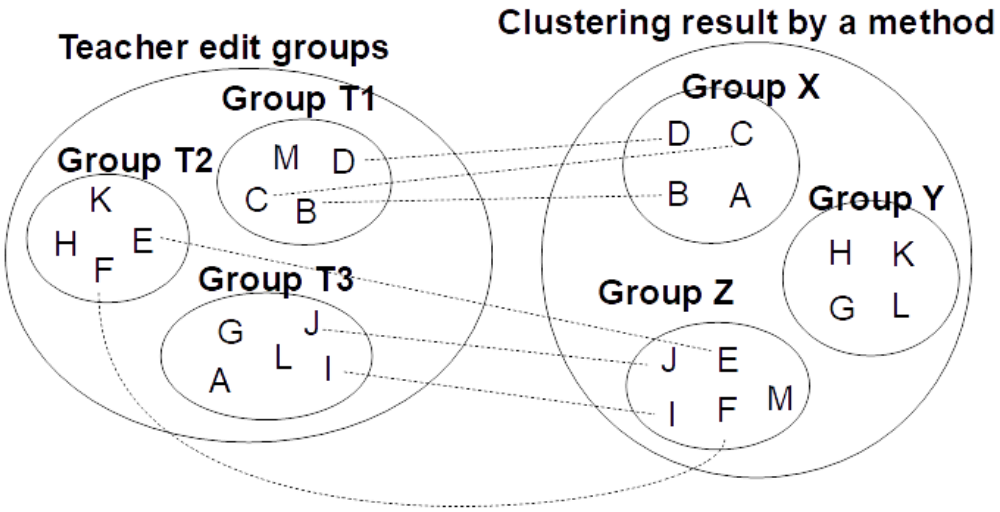


Figure 35. Manually corresponded example

According to the kappa interpret on Table 15, Table 16 lists the analysis results between teacher edits clustering results and other five clustering methods and the interpreted meaning for each value. Table 16 uses the simple letter to represent methods' name, T for teacher edit, S for Standard method, Q for Quantitative method, C for Cosine method, D for Diffusion methods, and G for GRACE method. The N of valid cases in the

Table 16 is the chromosomes, which has students' annotations (as shown in Table 7 in the page 96)

Table 16.

Alphabet-based clustering kappa analysis result between teacher edited and other methods

	database				HTML			
	Reading ID	N of Valid Cases	κ	Strength of Agreement	Reading ID	N of Valid Cases	κ	Strength of Agreement
T vs. S	131	10	0.403	moderate	133	10	0.2	fair
T vs. Q		10	0.242	fair		10	0.2	fair
T vs. C		10	0.394	fair		10	0	slight
T vs. D		10	0.219	fair		10	0.2	fair
T vs. G		10	0.265	fair		10	0	slight
131 average value			0.3046	fair	133 average value		0.12	Slight
T vs. S	132	14	0.197	slight	134	11	-0.031	poor
T vs. Q		14	0.311	fair		11	0.5	moderate
T vs. C		14	0.197	slight		11	0.295	fair
T vs. D		14	0.188	slight		11	0.761	substantial
T vs. G		14	0.279	fair		11	0.254	fair
132 average value			0.2344	fair	134 average value		0.3558	fair
T vs. S	137	13	-0.17	poor	135	10	0.032	slight
T vs. Q		13	0.133	slight		10	0.091	slight
T vs. C		13	0.527	moderate		10	0.032	slight
T vs. D		13	0.435	moderate		10	0.091	slight
T vs. G		13	0.228	fair		10	0.167	slight
137 average value			0.2306	fair	135 average value		0.0826	Slight
T vs. S	138	11	0.035	slight	136	12	0	slight
T vs. Q		11	0.127	slight		12	0.167	slight
T vs. C		11	0.29	fair		12	0.333	fair
T vs. D		11	0.621	substantial		12	0.167	slight
T vs. G		11	0.035	slight		12	0.333	fair

138 average value	0.2216	fair	136 average value	0.2	Slight
database average value	0.2478	fair	HTML average value	0.1896	slight
total average value (database and HTML)				0.2187	fair

After mapping all the clusters in the five methods with the groups decided by the teacher, I analyze the kappa coefficient (κ) as shown in the Table 16. The best value of kappa is 0.761, this number belongs to the Substantial agreement, and another Substantial agreement is 0.621. These are the only two Substantial agreements in the all values. The worst value of kappa is -0.031 and -0.17, these two values are the only two values below 0. The Standard method has two poor agreements in reading 134 and 137, which has the lowest performance; On the other hand, the Diffusion methods gets two substantial agreement in reading 134 and 138, which are the best performance in the evaluation. -

The Cohen's kappa analysis results are summarized in

Table 17. There is no case reached the “Almost Perfect” agreement and 17 cases (42.5%) in the “Slight” agreement. In the

Table 17, the highest kappa agreement in each clustering method is: the GRACE value is Fair (62.5%), Cosine value is Fair (50%), Quantitative value Slight (50%), and Standard value is Slight (50%) agree to teacher's clustering result. In previous four clustering methods, sometimes they will get a good agreement like Substantial also, they will get low agreement like Poor. Each agreement is contributed by different clustering methods and the number is shown in the

Table 18.

Table 17.

kappa interpret statistic

	Poor	Slight	Fair	Moderate	Substantial	Almost Perfect	Total
Standard	2 (25%)	4 (50%)	1 (12.5%)	1 (12.5%)	0 (0%)	0 (0%)	8
Quantitative	0 (0%)	4 (50%)	3 (37.5%)	1 (12.5%)	0 (0%)	0 (0%)	8
Cosine	0 (0%)	3 (37.5%)	4 (50%)	1 (12.5%)	0 (0%)	0 (0%)	8
Diffusion	0 (0%)	3 (37.5%)	2 (25%)	1 (12.5%)	2 (25%)	0 (0%)	8
GRACE	0 (0%)	3 (37.5%)	5 (62.5%)	0 (0%)	0 (0%)	0 (0%)	8
Total	2 / 40 (5.00%)	17 / 40 (42.50%)	15 / 40 (37.50%)	4 / 40 (10.00%)	2 / 40 (5.00%)	0 / 40 (0.00%)	40

Table 18 compares Cohen's kappa in GRACE and Ying's four methods. Ying's clustering methods majority drop at Slight (43.75%), other agreements also have few numbers, and this numbers are contributed from all the four clustering methods. The agreement for GRACE algorithm only has Fair (62.5%) and Slight (37.5%). Standard gets 2 poor (25%), 0 Substantial, and the highest agreement drops on the Slight, the Standard method may have low performance when clustering the chromosomes. Diffusion's highest agreement drops on the Slight, but it also gets 25% chance drop on Substantial, this method seems very not stable. Quantitative and Cosine's values centralize between the Slight and Moderate, these two methods seem stable, the difference between them is Quantitative could mostly drop on the Slight and Cosine drop on the Fair. The GRACE only have Slight and Fair agreement and the highest agreement drop on the Fair, the GRACE method seems have very stable agreement and clustering results.

Table 18.

kappa interpret statistic in comparing the four bit-string clustering methods and the GRACE algorithm

	Poor	Slight	Fair	Moderate	Substantial	Almost Perfect	Total
Standard	2 (25%)	4 (50%)	1 (12.5%)	1 (12.5%)	0 (0%)	0 (0%)	8
Quantitative	0 (0%)	4 (50%)	3 (37.5%)	1 (12.5%)	0 (0%)	0 (0%)	8
Cosine	0 (0%)	3 (37.5%)	4 (50%)	1 (12.5%)	0 (0%)	0 (0%)	8
Diffusion	0 (0%)	3 (37.5%)	2 (25%)	1 (12.5%)	2 (25%)	0 (0%)	8
Total	2 / 32 (6.25%)	14 / 32 (43.75%)	10 / 32 (31.25%)	4 / 32 (12.50%)	2 / 32 (6.25%)	0 / 40 (0.00%)	32
GRACE	0 / 8 (0%)	3 / 8 (37.5%)	5 / 8 (62.5%)	0 / 8 (0%)	0 / 8 (0%)	0 / 8 (0%)	8

5.4 Data analysis

5.4.1 reliability and validity analysis for questionnaires. For the experiment, this research design two questionnaire – the Diffusion of Innovation questionnaire and the system usability questionnaire. The Diffusion of Innovation questionnaire is for knowing students' experience of using e-reader. The system usability is for knowing students perceived about the Online Annotation System. Students fill the Diffusion of Innovation questionnaire before the experiment start. The study code they fill on the Diffusion of Innovation questionnaire should be used in the system usability questionnaire again after the experiment end. The Diffusion of Innovation questionnaire has 164 responses. As

Table **6** shows, the total students in both course 105. I suppose some response in the 164 responses are fill the questionnaire more than once, if they didn't fill the same study code, I can't consider any two are fill by one student. These 164 responses feedback have some students did not fill the study code, did not finish all the questions, and didn't fill the questionnaire by their thinking or experience. After deleting these invalid questionnaires, have 100 Diffusion of Innovation questionnaires are valid.

For the system usability questionnaire, the total number of the response is 37. After deleting the invalid questionnaires, there is 19 questionnaires left. Next step is matching the Diffusion of Innovation and system usability questionnaires' study code and I got 9 left. These 9 questionnaires are using in reliability and validity analysis. In Table 19 shows the all moderators' statstic. Regarding the question for what are you doing when you use Internet, I give students few choices for them can easily check. I also provide "other" option for them, but no one answer other. What they do when they use internet as shown in the Table 20.

Table 19.

The moderators' statistic

Moderator	Category	Students
Gender	Female	5 (6.173%)
	Male	4 (4.938%)
Course	HTML course	5 (6.173%)
	Database management system course	4 (4.938%)
The total internet use items	0-4 kinds	3 (3.704%)
	5-7 kinds	6 (7.407%)
When is the first time use browser	Under 10	4 (4.938%)
	Above 11	5 (6.173%)
When is the first time use internet	Under 14	6 (7.407%)

	Above 15	3 (3.704%)
When is the first time students know the e-reade	Under 12	2 (22.222%)
	From 13 to 17	4 (44.444%)
	Above 18	3 (33.333%)
When is the first time students use the e-reader	Under 12	2 (22.222%)
	From 13 to 17	4 (44.444%)
	Above 18	3 (33.333%)
How many hours that students use browser in week days. (the total hours / 5)	Less than 1	2 (22.222%)
	Between 1 and 2	5 (55.556%)
	More than 2	2 (22.222%)

Table 20.

Statistic of what students will do when they use browser

	Yes	No
Playing game	7 (77.78%)	2 (22.22%)
Posting article	1 (11.11%)	8 (88.89%)
Reading article	5 (55.56%)	4 (44.44%)
Reading news	3 (33.33%)	6 (66.67%)
Study	7 (77.78%)	2 (22.22%)
Using social	8 (88.89%)	1 (11.11%)
Watching movie	8 (88.89%)	1 (11.11%)

In the next step, the research use SPSS 25 to verify both the validity and reliability for the Diffusion of Innovation and system usability questionnaires. The Diffusion of Innovation questionnaire has 24 items in five factors – Relative Advantage has 5 items, Compatibility has 4 items, Complexity has 4 items, Trialability has 5 items, and Observability has 6 items. The system usability has 4 higher factors – the Like, Effectiveness, Efficiency, and Satisfaction. In Effectiveness factors has three lower factors – Ease of Learning has 6 items, Ease of Use has 5 items, and User Interface

Design has 5 items. In the Efficiency has two lower factors – the Information has 6 items and Usefulness has 4 items. The Satisfaction has two lower factors – Behavioural Intention to Use has 6 items and the Expectation has 4 items.

Based on the Cronbach's, alpha value will raise after deleting the item and the validity smaller than 0.6 in round 1 and round 2, I deleted the items 3, 7, 13, 14, 15, 23, and 24 in Diffusion of Innovation questionnaire and the item 2, 7, 10,14, 22, 28, and 32 in system usability questionnaire and I got stable results in round 3.

For the revised Diffusion of Innovation questionnaire, after the Cronbach's alpha value for each item gets above 0.7, the first factor needs to be deleted is the Compatibility factor because of this factor is too dispersion as Table 21 shows. After deleting the Compatibility factor, the Observability factor and the Relative Advantage has been deleted for the same reason. Regarding the reliability of the revised Diffusion of Innovation, Cronbach's alpha value is 0.88, which sits on an "Good" range and shows the questionnaire is reliable (George and Mallery 2010). The revised Diffusion of Innovation is shown in Table 22. The left item 10, 11, and 12 in complexity and the item 16, 17, and 18 in trialability will be analyzed for the deeper analysis.

Table 21.

The validity analysis of the Diffusion of Innovation questionnaire in round 3

	Component 1	Component 2	Component 3	Component 4	Component 5
Compatibility-6	.924				
Trialability-18	.803				
Relative Advantage-1	.789				
Compatibility-8	.746				
Relative Advantage-2	.732				
Observability-20		.947			

Observability-19		.878			
Trialability-17		-.640			
Trialability-16		-.582			
Relative Advantage-5			.881		
Observability-22			.860		
Relative Advantage-4			.614		
Complexity-12				.909	
Complexity-10				.789	
Complexity-11				.707	
Observability-21				-.699	
Compatibility-9					.935

Table 22.

Final validity analysis of the Diffusion of Innovation questionnaire

		factor 1	factor 2
Trialability-17	Before deciding whether or not to adopt an e-reader application with annotation functions, I would need to properly try it out	0.951	
Trialability-18	I would like to be permitted to use an e-reader application with annotation functions on a trial basis long enough to see what it can do.	0.918	
Trialability-16	Before deciding whether or not to adopt an e-reader application with annotation functions, I would need to use it on a trail basis.	0.897	
Complexity-10	My interaction with e-reader application with annotation functions is clear.		0.916
Complexity-12	Learning to use an e-reader application with annotation functions is easy for me.		0.872
Complexity-11	My interaction with e-reader application with annotation functions is understandable.		0.777
Eigenvalue		3.921	1.259
% of variance		65.347	20.979
Overall $\alpha = 0.880$, total variance explained is 86.326%			

For the system usability questionnaire, after the Cronbach's alpha value for each

item gets above 0.7, the first factor that I remove is the Ease of Learning factor because of the items in Ease of Learning are the splitter in this factor (as shown in Table 23). Regarding the efficiency factor, there are only two sub factors left, and all the items are dispersion into two factors. I remove the information sub factors because of the usefulness is more important than information to the Online Annotation System the validity analysis as shown in Table 24. There are only have item 29, 30, and 31 in usefulness of the Efficiency left for the deeper analysis. The item 39 is belong to the expectation of the satisfaction factor, after checking the question in item 39, this item change to behavioural intention to use of the satisfaction in round 5. Regarding the reliability of the revised system usability, Cronbach's alpha value is 0.969, which sits on an "Excellent" range and shows the questionnaire is reliable (George and Mallery 2010).

Table 23.

The effectiveness factor of validity analysis in round 3

	Component 1	Component 2	Component 3
Effectiveness-Ease of Use12	.942		
Effectiveness-Ease of Use13	.919		
Effectiveness-Ease of Learning8	.863		
Effectiveness-Ease of Learning11	.836		
Effectiveness-Ease of Use16	.705		
Effectiveness-Ease of Use15	.610		
Effectiveness-User Interface Design19		.932	
Effectiveness-User Interface Design21		.819	
Effectiveness-Ease of Learning6		.795	
Effectiveness-Ease of Learning9		.751	
Effectiveness-User Interface Design20		.745	.601
Effectiveness-User Interface Design18			.872
Effectiveness-User Interface Design17			.749

Table 24.

The efficiency factor of validity analysis in round 3

	Component 1	Component 2
Efficiency-Information26	.853	
Efficiency-Usefulness29	.798	
Efficiency-Information24	.775	
Efficiency-Information25	.773	
Efficiency-Usefulness31	.692	
Efficiency-Usefulness30	.544	
Efficiency-Information23		.854
Efficiency-Information27		.808

The Attitude Toward to Use in revised system usability only has Like factor (as shown in Table 25). The Attitude Toward to Use Cronbach's alpha value is 0.951, which sits on an "Excellent" range and shows the questionnaire is reliable (George and Mallery 2010).

Table 25.

Final validity analysis of the attitude toward use factor in system usability questionnaire

		Attitude toward to use
Attitude toward Use-Like4	As a student I like to use Online Annotation System to study.	0.949
Attitude toward Use-Like3	I like to use the Online Annotation System.	0.945
Attitude toward Use-Like1	I believe it is a good idea to use an Online Annotation System.	0.943

Attitude toward Use-Like5	The Online Annotation System is pleasant to use.	0.912
Eigenvalue		3.515
% of variance		87.868
Overall α = 0.951, total variance explained is 87.868%		

The efficiency in revised system usability only Usefulness left (as shown in Table 26). The efficiency's Cronbach's alpha value is 0.852, which sits on an "Good" range and shows the questionnaire is reliable (George and Mallery 2010).

Table 26.

Final validity analysis of the usefulness factor in system usability questionnaire

		Usefulness
Efficiency-Usefulness31	Using the Online Annotation System improves my learning performance.	0.955
Efficiency-Usefulness30	The Online Annotation System enables me to accomplish the reading activity more quickly.	0.872
Efficiency-Usefulness29	Using the Online Annotation System gives me greater control over my time to finish my reading activities.	0.843
Eigenvalue		2.382
% of variance		79.410
Overall α = 0.852, total variance explained is 79.410%		

The effectiveness in revised system usability has two lower factors left – the User Interface Design and Ease of Use (as shown in

). The effective' Cronbach's alpha value is 0.89, which sits on a "Good" range and shows the questionnaire is reliable (George and Mallery 2010).

Table 27.

Final validity analysis of the Effectiveness factor in system usability questionnaire

		User Interface Design	Ease of Use
Effectiveness-User Interface Design20	The logical design of this Online Annotation System is good, I have no difficulty in using it.	0.942	
Effectiveness-User Interface Design21	The Online Annotation System is user friendly.	0.939	
Effectiveness-User Interface Design19	The Online Annotation System requires minimal steps for doing my reading activity.	0.811	
Effectiveness-User Interface Design18	The user interface of the Online Annotation System is confusing.	0.698	
Effectiveness-User Interface Design17	The interface of the Online Annotation System is pleasant.	0.61	
Effectiveness-Ease of Use13	I think that I would need the support of a technical person to be able to use the Online Annotation System.		0.957
Effectiveness-Ease of Use12	I think the Online Annotation System is easy to use.		0.872
Effectiveness-Ease of Use16	I find the Online Annotation System very cumbersome to use.		0.78
Effectiveness-Ease of Use15	I think there is too much inconsistency in the Online Annotation System.		0.744
Eigenvalue		4.987	1.983
% of variance		55.409	22.039
Overall α = 0.890, total variance explained is 77.448%			

The satisfaction in revised system usability has two lower factor left – the Behavioural Intention to Use and Expectation (as shown in Table 28). The satisfaction's Cronbach's alpha value is 0.938, which sits on an "Excellent" range and shows the

questionnaire is reliable (George and Mallery 2010).

Table 28.

Final validity analysis of the Satisfaction factor in system usability questionnaire

		Behavioural Intention to Use	Expectation
Satisfaction-Behavioural Intention to Use 35	Assuming that I have access to an Online Annotation System, I intend to use it.	0.963	
Satisfaction-Behavioural Intention to Use 36	I intend to continue to use the Online Annotation System in the future.	0.892	
Satisfaction-Behavioural Intention to Use 37	I will recommend others to use the Online Annotation System.	0.871	
Satisfaction-Behavioural Intention to Use 34	I plan to use an Online Annotation System in the future.	0.836	
Satisfaction-Behavioural Intention to Use 39	I expect that I would use the Online Annotation System in the future.	0.722	
Satisfaction-Behavioural Intention to Use 33	I feel very confident using the Online Annotation System.	0.697	
Satisfaction-Expectation40	The Online Annotation System meets my needs.		0.927
Satisfaction-Expectation38	This Online Annotation System has all the functions and capabilities I expect it to have.		0.923
Satisfaction-Expectation41	The Online Annotation System works the way I want it to work.		0.837
Eigenvalue		6.248	1.504
% of variance		69.419	16.713
Overall $\alpha = 0.938$, total variance explained is 86.131%			

5.4.2 quantitative analysis. When analyzing the usability, this research uses two different usability scores. The first one is the SUS value based on Brooke study (marked

as SUS in the following analysis) in 1996. The average value of SUS is 55.556. The meaning of the value means the system is acceptable according to Bangor, Kortum, and Miller's (2009) research. The second usability score is the average value of the items in the Usability factor after removing uncategorized item in the reliability and validity analysis (marked as Usability in the following sections). The values of other factors are the average value of students' responses of the items in each factor. The average value of the Usability is 3.476. The meaning of the value is students' perceived for the system is OK trend to the agree. The SUS and usability score are in the middle and trend to the better side.

To answer Q1 – Will students' Diffusion of Innovation affect their given score in System usability score, I use Pearson correlation to analysis the relationship between Diffusion of Innovation and usability. The Table 29 shows the relation between Diffusion of Innovation and Usability is significant correlated ($r = 0.678$, $p = 0.045$); however, there is no significant relation between Diffusion of Innovation and SUS ($r = 0.555$, $p = 0.121$). The results show that H1 – The Diffusion of Innovation will positively affect students' System usability score. – is confirmed.

Table 29.

Correlation analysis between Diffusion of Innovation and usability score

		SUS	Usability
Diffusion of Innovation	Pearson correlation	0.555	0.678*
	Sig.	0.121	0.045
	N	9	9

*: $p < 0.05$; **: $p < 0.01$

Because of the Diffusion of Innovation affects the system usability score, I would

like to know whether or not the Diffusion of Innovation affect the factors in the usability as well as which factors in Diffusion of Innovation will affect the usability scores. The correlation analysis results in factors in Diffusion of Innovation and usability are listed on the Table 30.

Table 30.

Correlation analysis between Diffusion of Innovation and usability factors

		Efficiency	Effectiveness	Satisfaction
Diffusion of Innovation	Pearson correlation	0.599	0.804**	0.712*
	Sig.	0.088	0.009	0.031
	N	9	9	9

*: $p < 0.05$; **: $p < 0.01$

As Table 30 shows, students' experience of using e-reader does not have significate correlation with student's perceived effectiveness for the Online Annotation System. It shows that H2 is not sustained - students' experience of using e-reader will not affect their perceived effectiveness for the Online Annotation System. However, Diffusion of Innovation has significant positive relation to students' perceived efficiency ($r = 0.678$, $p = 0.045$) and perceived satisfaction ($r = 0.804$, $p = 0.009$) when using Online Annotation System – H3 and H4 are confirmed.

Because of the Diffusion of Innovation affects the factors in the system usability, I would like to know whether the factors in the Diffusion of Innovation affect the factors in the usability. The correlation analysis results in factors in Diffusion of Innovation and usability are listed on the Table 31.

Table 31.

Correlation analysis between Diffusion of Innovation factors and System usability factors

		Usability Scores		Effectiveness		Efficiency	Satisfaction	
		SUS	Usability	EoU	UI	Use	BI	Exp
Complexity	Pearson correlation	0.149	0.327	0.176	0.393	0.341	0.418	0.197
	Sig.	0.703	0.390	0.650	0.296	0.369	0.263	0.611
	N	9	9	9	9	9	9	9
Trialability	Pearson correlation	0.729*	0.839**	0.595	0.547		0.926**	0.758*
	Sig.	0.026	0.005	0.091	0.127		0.000	0.018
	N	9	9	9	9		9	9

EoU: Ease of Use; UI: User Interface Design; Use: Usefulness; BI: Behvioural Intention to Use; Exp: Expectation
*: $p < 0.05$; **: $p < 0.01$

The first things to know is the relation between the trial ability and system usability. The experience of trying to use e-reader has significate correlate with students perceived for the Online Annotation System and the correlation r is 0.839 ($p = 0.005$) (H14). The experience of trying to use e-reader affects students perceived for the Online Annotation System. The results show that H14 – The perceived trial ability will positively affect students' System usability score. – is confirmed.

The experience of trying affects students perceived for the Online Annotation System. How does the trialability affects the lower factors of the system usability? The experience of trying to use e-reader doesn't have significate correlate with students' perceived ease of use (H18a). The experience of trying to use e-reader doesn't affect students' perceived ease of use. The results show that H18a – The perceived trialability will positively affect students' perceived ease of use. – is fail.

Another lower factor in the Effectiveness is User Interface Design. The experience of trying to use e-reader doesn't have significant correlation with students' perceived user interface design (H18b). The experience of trying to use e-reader doesn't affect students' perceived user interface design. The results show that H18b – The perceived trialability will positively affect students' perceived user interface design. – is fail.

The H19a is the perceived trialability will positively affect students' perceived behavioural intention to use. The experience of trying to use e-reader has significate correlate with students' perceived behavioural intention to use and the correlation r is 0.926 (p = 0.000). The experience of trying to use e-reader affects students' perceived behavioural intention to use. The results show that H19a is confirmed.

Second lower factor in the satisfaction is expectation. The experience of trying to use e-reader has significate correlate with students' perceived expectation and the correlation r is 0.758 (p = 0.018) (H19b). The experience of trying to use e-reader affects students' perceived expectation. The results show that H19b – The perceived trialability will positively affect students' perceived expectation. – is confirmed.

First, the Complexity is not related to the sub-factors in the System's Usability. The hypotheses H13, H15, H16, and H17 are rejected. However, when I test the correlation between Trialability and the sub-factors in the System Usability, Trialability only has no significant relation to Ease of Use and Use Interface Design – the sub-factors of Effectiveness. Trialability has significant positive relation to the other sub-factors. It shows that the hypotheses H14, H19a, and H19b are proved.

The question 2 in my research is – will the system usability score affect student's perceived attitude toward to the Online Annotation System? The correlation analysis

between system usability factors and attitude toward to use Online Annotation System is listed on the Table 32.

Table 32.

Correlation analysis between system usability factors and Attitude Toward to use Online Annotation System

		Usability Scores		Effectiveness	Efficiency	Satisfaction
		SUS	Usability			
Att	Pearson correlation	0.838**	0.937**	0.895**	0.877**	0.806**
	Sig.	0.005	0.000	0.001	0.002	0.009
	N	9	9	9	9	9

*: $p < 0.05$; **: $p < 0.01$

First thing to know is the correlation between the system usability and the attitude toward to use Online Annotation System. The SUS value ($r = 0.838$ and $p = 0.005$) and the system usability score ($r = 0.937$ and $p = 0.000$) both have significate correlate with students' perceived attitude toward to use Online Annotation System (H5). The system usability affects students' perceived attitude toward to use Online Annotation System. The results show that H5 – The system usability score will positively affect students' perceived attitude toward to use Online Annotation System. – is confirmed.

Because of the Online Annotation System affects students' perceived attitude toward to use Online Annotation System, I would like to know whether the factors of the Online Annotation System affect students' perceived attitude toward to use Online Annotation System? The correlation analysis results in factors in Online Annotation System and students' perceived attitude toward to use Online Annotation System are listed on the Table 32.

The first higher factor is the Effectiveness. The perceived effectiveness has

significate correlate with students' perceived attitude toward to use Online Annotation System and the correlation *r* is 0.895 ($p = 0.001$) (H6). The perceived effectiveness affects students' perceived attitude toward to use Online Annotation System. The results show that H6 – The perceived effectiveness will positively affect students' perceived attitude toward to use Online Annotation System. – is confirmed.

Second hither factor is the Efficiency. The perceived usefulness has significate correlate with students' perceived attitude toward to use Online Annotation System and the correlation *r* is 0.877 ($p = 0.002$) (H7). The perceived usefulness affects students' perceived attitude toward to use Online Annotation System. The results show that H7 – The perceived efficiency will positively affect students' perceived attitude toward to use Online Annotation System. – is confirmed.

The last higher factor is Satisfaction. The perceived satisfaction has significate correlate with students' perceived attitude toward to use Online Annotation System and the correlation *r* is 0.806 ($p = 0.009$) (H8). The perceived satisfaction affects students' perceived attitude toward to use Online Annotation System. The results show that H8 – The perceived satisfaction will positively affect students' perceived attitude toward to use Online Annotation System. – is confirmed.

The higher-factors in the System's Usability have significant relation to the attitude toward to use Online Annotation System. The hypotheses H5, H6, H7, and H8 are confirmed.

As Table 32 shown, the three higher factors of the system usability have the significate correlation with students perceived attitude toward to use Online Annotation System. I would also like to know whether the sub-factors in Effectiveness, Efficiency,

and Satisfaction has signification relation to the students' attitudes toward the Online Annotation System. The bivariate correlation analysis results are listed in Table 33.

Table 33.

Correlation analysis between system usability factors and Attitude Toward to use Online Annotation System

		Effectiveness		Efficiency	Satisfaction	
		EoU	UI	Use	BI	Exp
Att	Pearson correlation	0.822**	0.725*	0.877**	0.752*	0.724*
	Sig.	0.007	0.027	0.002	0.019	0.028
	N	9	9	9	9	9

Att: Attitude toward to Use Online Annotation System; EoU: Ease of Use; UI: User Interface Design; Use: Usefulness; BI: Behvioural Intention to Use; Exp: Expectation
*: $p < 0.05$; **: $p < 0.01$

In the Effectiveness has two lower factors. One of the lower factors is the ease of use. The perceived ease of use has significate correlate with students' perceived attitude toward to use Online Annotation System and the correlation *r* is 0.822 (p=0.007) (H20a). The perceived ease of use affects students' perceived attitude toward to use Online Annotation System. The results show that H20a – The perceived ease of use will positively affect students' perceived attitude toward to use Online Annotation System. – is confirmed.

Another lower factor is user interface design. The perceived user interface design has significate correlate with students' perceived attitude toward to use Online Annotation System and the correlation *r* is 0.725 ($p = 0.027$) (H20b). The perceived user interface design affects students' perceived attitude toward to use Online Annotation System. The results show that H20b – The perceived user interface design will positively affect students' perceived attitude toward to use Online Annotation System. – is confirmed.

The lower factor in the Efficiency has usefulness. The perceived usefulness has significate correlate with students' perceived attitude toward to use Online Annotation System and the correlation *r* is 0.877 ($p = 0.002$) (H21). The perceived usefulness affects students' perceived attitude toward to use Online Annotation System. The results show that H21 – The perceived usefulness will positively affect students' perceived attitude toward to use Online Annotation System. – is confirmed.

Regarding the Satisfaction's lower factor, behavioural intention to use is one of the lower factors. The perceived behavioural intention to use has significate correlate with students' perceived attitude toward to use Online Annotation System and the correlation *r* is 0.752 ($p = 0.019$) (H22a). The perceived behavioural intention to use affects students' perceived attitude toward to use Online Annotation System. The results show that H22a – The perceived behavioural intention to use will positively affect students' perceived attitude toward to use Online Annotation System. – is confirmed.

The last lower factor is the expectation. The perceived expectation has significate correlate with students' perceived attitude toward to use Online Annotation System and the correlation *r* is 0.724 ($p = 0.028$) (H22b). The perceived expectation affects students' perceived attitude toward to use Online Annotation System. The results show that H22b – The perceived expectation will positively affect students' perceived attitude toward to use Online Annotation System. – is confirmed.

As Table 33 shown, every lower factor have significate correlate with students' perceived attitude toward to use Online Annotation System. Everyone can affect students' perceived attitude toward to use Online Annotation System.

The sub-factors in each higher-factors in the System's Usability have significant

relation to the attitude toward to use Online Annotation System. The hypotheses H20a, H20b, H21, H22a, and H22b are confirmed.

After I analysis the correlation between the Diffusion of Innovation and the system usability and correlation between the system suability and the attitude toward to use Online Annotation System, I want to know does the Diffusion of Innovation will affect the Attitude toward to use the Online Annotation System (Q3)? To answer the question 3, the correlation analysis between Diffusion of Innovation and the attitude toward to use the system is listed in Table 34.

Table 34.

Correlation analysis between Diffusion of Innovation and Attitude Toward to use Online Annotation System

		Diffusion of Innovation		
		ALL	Complexity	Trialability
Attitude	Pearson correlation	0.478	0.080	0.664
	Sig.	0.193	0.839	0.051
	N	9	9	9

*: $p < 0.05$; **: $p < 0.01$

The H9 is analyzing the correlation between the Diffusion of Innovation and the Attitude toward to use Online Annotation System. The experience of using e-reader doesn't have significate correlate with students' perceived attitude toward to use Online Annotation System. The experience of using e-reader doesn't affect students' perceived attitude toward to use Online Annotation System. The results show that H9 – The Diffusion of Innovation will positively affect students' perceived attitude toward to use Online Annotation System. – is fail.

Even the Diffusion of Innovation doesn't affect students' perceived attitude toward

to use Online Annotation System. I would like to know whether the factors of Diffusion of Innovation affect students' perceived attitude toward to use Online Annotation System? The correlation analysis results in factors in Diffusion of Innovation and students' perceived attitude toward to use Online Annotation System are listed on the Table 34.

In the Diffusion of Innovation has two factors, one of the factors is the Complexity. The perceived complexity of experience for using e-reader doesn't have significant correlate with students' perceived attitude toward to use Online Annotation System (H23). The perceived complexity of experience for using e-reader doesn't affect students' perceived attitude toward to use Online Annotation System. The results show that H23 – The perceived complexity will positively affect students' perceived attitude toward to use Online Annotation System. – is fail.

Another related analysis is the Trialability factor. The perceived trialability doesn't have significate correlate with students' perceived attitude toward to use Online Annotation System (H24). The perceived trialability doesn't affects students' perceived attitude toward to use Online Annotation System. The results show that H24 – The perceived trialability will positively affect students' perceived attitude toward to use Online Annotation System. – is fail.

The sub-factors of the Diffusion of Innovation don't have any significate correlate with students' perceived attitude toward to use Online Annotation System. The hypotheses H9, H23, and H24 are rejected.

I would like to know whether the factors in the system usability affects other factors in the system usability (Q4)? The correlation analysis results between factors in system

usability are listed on the Table 35 and Table 38.

Table 35.

Correlation analysis between effectiveness and other same level's factors

		Efficiency	Satisfaction
Effectiveness	Pearson correlation	0.898**	0.739*
	Sig.	0.001	0.023
	N	9	9

*: $p < 0.05$; **: $p < 0.01$

The H10 is the perceived effectiveness will affect students' perceived efficiency when using Online Annotation System. The correlation is shown in the Table 35. The perceived effectiveness has significate correlate with students' perceived efficiency and the correlation r is 0.898 ($p = 0.001$). The perceived effectiveness affects students' perceived efficiency. The results show that H10 is confirmed.

After confirmed the effectiveness affects students' perceived efficiency, I would like to know whether the lower factors in each hither factor affect another lower factor. The correlation analysis results in Ease of Use and other system usability internal factors are listed on the Table 36.

Table 36.

Correlation analysis between Ease of Use and other system usability internal factors

		Efficiency	Satisfaction	
		Use	BI	Exp
EoU	Pearson correlation	0.687*	0.537	0.538
	Sig.	0.041	0.136	0.135
	N	9	9	9

EoU: Ease of Use; Use: Usefulness; BI: Behvioural Intention to Use; Exp: Expectation
*: $p < 0.05$; **: $p < 0.01$

The perceived ease of use has significate correlate with students' perceived usefulness and the correlation *r* is 0.687 ($p = 0.041$) (H29). The perceived ease of use affects students' perceived usefulness. The results show that H29 – The perceived ease of use will affect students' perceived usefulness. – is confirmed.

The relation between effectiveness and efficiency has significant relation. The hypotheses H10 and H29 are confirmed.

The effectiveness affects the efficiency and the lower factors – the ease of use also affects the usefulness. This result means the perceived ease of use affects students perceived usefulness for the Online Annotation System.

The H11 is the perceived effectiveness will affect students' perceived satisfaction when using Online Annotation System. The correlation between the effectiveness and satisfaction as shown in the Table 35. The perceived effectiveness has significate correlate with students' perceived satisfaction and the correlation *r* is 0.739 ($p = 0.023$). The perceived effectiveness affects students' perceived satisfaction. The results show that H11 is confirmed.

Because of the effectiveness affects students' perceived satisfaction, I would like to know whether the lower factors in the effectiveness affect the lower factors in the satisfaction? The correlation analysis results in the lower factors in the effectiveness and lower factors in the satisfaction are listed on the Table 36 and Table 37.

A related analysis is between the ease of use and expectation. The perceived ease of use doesn't have significant correlation with students' perceived expectation (H25) the analysis is shown in Table 36. The perceived ease of use doesn't affect students' perceived expectation. The results show that H25 – The perceived ease of use will affect

students' perceived expectation. – is fail.

Another related analysis is ease of use and behavioural intention to use. The perceived ease of use doesn't have significate correlate with students' perceived intention to use the Online Annotation System (H26) the analysis is shown in Table 36. The perceived ease of use doesn't affect students' perceived intention to use the Online Annotation System. The results show that H26 – The perceived ease of use will affect students' perceived behavioural intention to use. – is fail.

The higher factor Effectiveness has another lower factor –user interface design. The perceived user interface design does not have significate correlate with students' perceived expectation (H27) the analysis is shown in Table 37. The perceived user interface design does not affect students' perceived expectation. The results show that H27 – The perceived user interface design will affect students' perceived expectation. – is fail.

Table 37.

Correlation analysis between User Interface Design and other system usability internal factors

		Effectiveness	Satisfaction	
		EoU	BI	Exp
UI	Pearson correlation	0.483	0.734*	0.424
	Sig.	0.187	0.024	0.255
	N	9	9	9

EoU: Ease of Use; UI: User Interface Design; Use: Usefulness; BI: Behvioural Intention to Use; Exp: Expectation
*: $p < 0.05$; **: $p < 0.01$

The user interface design could affect the behavioural intention to use as H28 says. The perceived user interface design has significate correlate with students' behavioural intention to use and the correlation r is 0.734 ($p = 0.024$) the analysis is shown in Table

37. The perceived user interface design affects students' perceived behavioural intention to use. The results show that H28 is confirmed.

Another hypothesis related to the user interface design is the H32. The perceived user interface design doesn't have significate correlate with students' perceived ease of use the analysis is shown in Table 37. The perceived user interface design doesn't affect students' perceived ease of use. The results show that H32 – The perceived user interface design will affect students' perceived ease of use. – is fail.

First, the effectiveness has significant relation to the satisfaction. The hypotheses H11 is confirmed. Nevertheless, when I test the correlation between the sub-factors in the effectiveness and the sub-factors in the satisfaction, there are only the user interface design and the behavioural intention to use has significant relation – only H28 is confirmed. There is no significant relation in the H25, H26, H27, and H32, which mean these hypotheses are rejected.

The last internal factors affect is: the perceived efficiency will affect students' perceived satisfaction when using Online Annotation System (H12). The perceived efficiency has significate correlate with students' perceived satisfaction and the correlation *r* is 0.893 ($p = 0.001$) the analysis is shown in Table 38. The perceived efficiency affects students' perceived satisfaction. The results show that H12 is confirmed.

Table 38.

Correlation analysis between efficiency and satisfaction

		Efficiency
Satisfaction	Pearson correlation	0.893**
	Sig.	0.001

	N	9

*: p < 0.05; **: p < 0.01

Because of the perceived efficiency affects students' perceived satisfaction, I would like to know whether the lower factor in the efficiency affect the lower factors in the satisfaction. The correlation analysis results in the lower factor in efficiency and the lower factors in the satisfaction are listed on the Table 39.

The analysis between the usefulness and behavioural intention to use. The perceived usefulness has significate correlate with students' perceived behavioural intention to use the Online Annotation System and the correlation *r* is 0.83 (*p* = 0.006) (H30) the analysis is shown in Table 39. The perceived usefulness affects students perceived behavioural intention to use the Online Annotation System. The results show that H30 – The perceived usefulness will affect students perceived behavioural intention to use. – is confirmed.

Table 39.

Correlation analysis between Usefulness and other system usability internal factors

		Satisfaction	
		BI	Exp
Use	Pearson correlation	0.830**	0.809**
	Sig.	0.006	0.008
	N	9	9

Use: Usefulness; BI: Behvioural Intention to Use; Exp: Expectation
*: p < 0.05; **: p < 0.01

Another analysis is the usefulness and expectation. The perceived usefulness has significate correlate with students' perceived expectation and the correlation *r* is 0.809 (*p* = 0.008) (H31) the analysis is shown in Table 39. The perceived usefulness affects

students' perceived expectation. The results show that H31 – The perceived usefulness will affect students' perceived expectation. – is confirmed.

The efficiency affects satisfaction and look deep into the lower factors' analysis. The usefulness affects student's perceived behavioural intention to use and expectation. This means students' perceived usefulness affects if students want to keep using the Online Annotation System or not.

The relation between efficiency and satisfaction has significant relation. The hypotheses H12, H30, and H31 are confirmed.

5.4.3 moderator analysis. There are some moderators are the questions from the original questionnaires. When analyzing the reliability and validity, some factors are removed because of the weak reliability or. I choose one of the questions from that factor as the moderator. The five extra moderators are list in the Table 40.

Table 40.

Five moderators from deleted factors

Short name for moderators	Questions in the questionnaires
Relative Advantage	Using an e-reader application with annotation functions enables me to understand the key concepts of the reading activities more quickly.
Compatibility	I think that using an e-reader application with annotation functions fits well with the way I like to study.
Observability	I have had a lot of opportunities to see people using e-reader application with annotation functions to study.
Effectiveness – Ease of Learning	Learning to use the Online Annotation System is easy for me.
Efficiency –	The information provided by the Online Annotation System is easy to

Information	understand.

Regarding the gender moderator, Table 41 shows there is no significant difference between male and female in their given scores toward each factor. Students attitude are similar; their thinking or attitude doesn't have difference by their gender.

Table 41.

Independent t-test result for extract moderator and each factor (gender difference analysis)

Factor	Subfactor	Gender	Descriptive statistic			*t*-test		
			N	Mean	SD	*t*	*df*	*p*
Moderators	Relative Advantage	Female	5	3.800	1.304	-0.220	7	0.832
		Male	4	4.000	1.414			
	Compatibility	Female	5	4.200	0.447	0.471	7	0.652
		Male	4	4.000	0.816			
	Observability	Female	5	2.800	1.643	0.518	7	0.621
		Male	4	2.250	1.500			
	Effectiveness – EoL	Female	5	3.800	1.304	-0.574	7	0.584
		Male	4	4.250	0.957			
	Efficiency - Information	Female	5	3.600	1.517	-0.471	7	0.652
		Male	4	4.000	0.816			
Diffusion of Innovation	Complexity	Female	5	3.333	0.816	0.114	7	0.912
		Male	4	3.250	1.371			
	Trialability	Female	5	3.200	1.406	-0.637	7	0.544
		Male	4	3.833	1.575			
	ALL	Female	5	3.267	0.997	-0.360	7	0.729
		Male	4	3.542	1.301			
Effectiveness	EoU	Female	5	3.250	1.159	-1.347	7	0.220
		Male	4	4.125	0.629			
	UI	Female	5	3.560	0.590	0.170	7	0.870
		Male	4	3.450	1.310			
	ALL	Female	5	3.422	0.751	-0.568	7	0.588
		Male	4	3.750	0.987			

Efficiency	Use	Female	5	3.333	0.972	-0.447	7	0.669
		Male	4	3.667	1.277			
Satisfaction	BI	Female	5	3.200	0.811	-0.011	7	0.992
		Male	4	3.208	1.524			
	Exp	Female	5	3.133	1.406	-0.581	7	0.579
		Male	4	3.583	0.687			
	ALL	Female	5	3.178	0.958	-0.215	7	0.836
		Male	4	3.333	1.221			
SUS		Female	5	51.000	23.822	-0.039	7	0.970
		Male	4	61.250	19.843			
Usability		Female	5	3.368	0.823	-0.382	7	0.714
		Male	4	3.610	1.086			
Att		Female	5	3.700	1.204	-0.235	7	0.821
		Male	4	3.875	0.968			

Att: Attitude toward to Use Online Annotation System; EoL: Ease of Learning; EoU: Ease of Use; UI: User Interface Design; Use: Usefulness; BI: Behvioural Intention to Use; Exp: Expectation
*: $p < 0.05$; **: $p < 0.01$

Asking students which course they take because I suppose that difference year of the college could have different perceived for the Online Annotation System. The course different analysis with each factors and extract moderator are listed in the Table 42. In this table doesn't show any significant difference between two different courses in each factor.

Table 42.

Independent t-test result for extract moderator and each factor (course difference analysis)

Factor	Subfactors	Course	Descriptive statistic			*t*-test		
			N	Mean	SD	*t*	*df*	*p*
Moderators	Relative Advantage	HTML	5	4.200	1.304	0.804	7	0.448
		Database	4	3.500	1.291			
	Compatibility	HTML	5	4.200	0.447	0.471	7	0.652
		Database	4	4.000	0.816			
	Observability	HTML	5	2.200	1.304	-0.770	7	0.467

		Database	4	3.000	1.826			
	Effectiveness - EoL	HTML	5	3.800	1.304	-0.574	7	0.584
		Database	4	4.250	0.957			
	Efficiency - Information	HTML	5	3.800	1.643	0.058	7	0.955
		Database	4	3.750	0.500			
Diffusion of Innovation	Complexity	HTML	5	3.733	1.090	1.561	7	0.162
		Database	4	2.750	0.687			
	Trialability	HTML	5	4.000	1.374	1.265	7	0.246
		Database	4	2.833	1.374			
	ALL	HTML	5	3.867	1.010	1.644	7	0.144
		Database	4	2.792	0.927			
Effectiveness	EoU	HTML	5	3.400	1.294	-0.768	7	0.468
		Database	4	3.938	0.554			
	UI	HTML	5	3.880	0.807	1.458	7	0.188
		Database	4	3.050	0.900			
	ALL	HTML	5	3.667	0.984	0.380	7	0.715
		Database	4	3.444	0.691			
Efficiency	Use	HTML	5	3.800	1.095	1.014	7	0.344
		Database	4	3.083	0.995			
Satisfaction	BI	HTML	5	3.633	0.916	1.390	7	0.207
		Database	4	2.667	1.179			
	Exp	HTML	5	3.600	0.925	0.790	7	0.456
		Database	4	3.000	1.361			
	ALL	HTML	5	3.622	0.870	1.296	7	0.236
		Database	4	2.778	1.092			
SUS		HTML	5	57.000	24.711	1.616	7	0.150
		Database	4	53.750	20.156			
Usability		HTML	5	3.688	0.957	0.778	7	0.462
		Database	4	3.210	0.858			
Att		HTML	5	3.800	1.230	0.067	7	0.948
		Database	4	3.750	0.935			

Att: Attitude toward to Use Online Annotation System; EoL: Ease of Learning; EoU: Ease of Use; UI: User Interface Design; Use: Usefulness; BI: Behvioural Intention to Use; Exp: Expectation

*: $p < 0.05$; **: $p < 0.01$

When asking students when did they use internet for first time, I suppose the earlier they use the internet, their acceptance for the e-reader is higher. However, there are 6 students use the internet for their first time are between 7 to 12 years old. This range is about study in the elementary school and if divided into three group there are not enough students in the smallest group. This is the mean reason that I divide into two groups. The *t*-test for the two groups with each factors and extract moderators are listed in the Table 43. When is the first time that start use internet have significant difference in the extract moderator Compatibility-8, question is "I think that using an e-reader application with annotation functions fits well with the way I like to study."

The significant difference also on the SUS value, Trialability of Diffusion of Innovation, Attitude toward to Online Annotation System, Intention, Expectation, and Satisfaction of the system usability, and all these mean values are above 15 higher than under 14. I think this is because even they use the internet when they were in elementary school, but their parents could still control their total time for use internet for every day or every week. When they get the time can use internet, child will choose having fun than study. Even they already in their 20s, they still used to use internet for fun.

Table 43.

Independent t-test result for extract moderator and each factor (when is the first time use internet)

Factor	Subfactor	First time use Internet	Descriptive statistic			*t*-test		
			N	Mean	SD	*t*	*df*	*p*
Moderators	Relative Advantage	Under 14	6	3.667	1.366	-0.720	7	0.495
		Above 15	3	4.333	1.155			
	Compatibility	Under 14	6	3.833	0.408	-2.546	7	0.038*
		Above 15	3	4.667	0.577			

	Observability	Under 14	6	3.167	1.472	2.023	7	0.083
		Above 15	3	1.333	0.577			
	Effectiveness – EoL	Under 14	6	3.667	1.211	-1.323	7	0.227
		Above 15	3	4.667	0.577			
	Efficiency - Information	Under 14	6	3.500	1.378	-0.978	7	0.361
		Above 15	3	4.333	0.577			
Diffusion of Innovation	Complexity	Under 14	6	3.333	1.011	0.144	7	0.889
		Above 15	3	3.222	1.262			
	Trialability	Under 14	6	2.833	1.295	-2.469	7	0.043*
		Above 15	3	4.778	0.385			
	ALL	Under 14	6	3.083	1.139	-1.249	7	0.252
		Above 15	3	4.000	0.726			
Effectiveness	EoU	Under 14	6	3.375	1.159	-1.119	7	0.300
		Above 15	3	4.167	0.382			
	UI	Under 14	6	3.333	1.056	-0.815	7	0.442
		Above 15	3	3.867	0.462			
	ALL	Under 14	6	3.352	0.933	-1.133	7	0.294
		Above 15	3	4.000	0.333			
Efficiency	Use	Under 14	6	3.111	1.089	-1.638	7	0.145
		Above 15	3	4.222	0.509			
Satisfaction	BI	Under 14	6	2.667	0.913	-2.871	7	0.024*
		Above 15	3	4.278	0.347			
	Exp	Under 14	6	2.833	1.006	-2.443	7	0.045*
		Above 15	3	4.333	0.333			
	ALL	Under 14	6	2.722	0.791	-3.274	7	0.014*
		Above 15	3	4.296	0.231			
SUS value		Under 14	6	45.833	19.343	-3.031	7	0.019*
		Above 15	3	75.000	7.500			
Usability		Under 14	6	3.087	0.830	-2.284	7	0.056
		Above 15	3	4.253	0.321			
Att		Under 14	6	3.292	0.900	-2.593	7	0.036*
		Above 15	3	4.750	0.433			

Att: Attitude toward to Use Online Annotation System; EoL: Ease of Learning; EoU: Ease of Use; UI: User Interface Design; Use: Usefulness; BI: Behvioural Intention to Use; Exp: Expectation

*: $p < 0.05$; **: $p < 0.01$

When asking student when did they use the browser for the first time, I suppose students first time use internet will earlier than the first time use browser or in the same age, but all students first time using browser are under 12. Some in use internet earlier some is use browser earlier, therefore, I decided to divide students into two group from the middle – under 10 and above 11. There are two questions in the questionnaire is relate to the browser – what they do with browser and how many hours they use in week day. For the "what they do with browser" questions, I sum each students' behaviour into a total item they use on the browser to the "total internet use items". The independent *t*-test result shows the significant difference on the two extra analysis, extract moderator Observability-20, and SUS value.

In the two extra question analysis, students who use the browser in their first time under 10 on the total use hours in week day is a lot lower than above 11. The total internet use items also have significant difference between these two groups. Maybe because they use browser later also means they know what browser can do more. Therefore, when they start to use the browser, they want to try everything they know. When they try, they think it's interesting, so they keep the activity as a habit and keep using it. On the other hand, the earlier they use the browser, some fun thing could be controlled by their presents (for example, watch movies and playing games) and later when they know new thing and try it out, maybe they just try it out and didn't keep using it.

In the question in Observability-20 is "I have had a lot of opportunities to see people using e-reader application with annotation functions to study." This result could fit the

reason that later they start use browser they see more what browser can do and willing give it a try.

Table 44.

Independent t-test result for extract moderator and each factor (when is the first time use browser)

Factor	Subfactor	First time use browsers	Descriptive statistic			*t*-test		
			N	Mean	SD	*t*	*df*	*p*
Extra Moderators	Usually in a week day, how many hours you use Internet with browsers?	Under 10	4	3.750	2.062	-2.690	4.842	0.045*
		Above 11	5	12.600	6.986			
	Total internet use items	Under 10	4	3.000	1.826	-2.605	7	0.035*
		Above 11	5	5.400	0.894			
Moderators	Relative Advantage	Under 10	4	3.750	1.500	-0.276	7	0.790
		Above 11	5	4.000	1.225			
	Compatibility	Under 10	4	4.250	0.957	0.522	3	0.638
		Above 11	5	4.000	0.000			
	Observability	Under 10	4	1.250	0.500	-3.800	7	0.007*
		Above 11	5	3.600	1.140			
	Effectiveness - EoL	Under 10	4	4.250	0.957	0.574	7	0.584
		Above 11	5	3.800	1.304			
	Efficiency - Information	Under 10	4	4.000	0.816	0.471	7	0.652
		Above 11	5	3.600	1.517			
Diffusion of Innovation	Complexity	Under 10	4	2.917	1.198	-0.999	7	0.351
		Above 11	5	3.600	0.863			
	Trialability	Under 10	4	4.000	1.587	0.973	7	0.363
		Above 11	5	3.067	1.300			
	ALL	Under 10	4	3.458	1.235	0.163	7	0.875
		Above 11	5	3.333	1.074			
Effectiveness	EoU	Under 10	4	3.938	0.554	0.768	7	0.468
		Above 11	5	3.400	1.294			
	UI	Under 10	4	3.350	1.100	-0.453	7	0.664

		Above 11	5	3.640	0.829			
	ALL	Under 10	4	3.611	0.824	0.132	7	0.899
		Above 11	5	3.533	0.918			
Efficiency	Use	Under 10	4	3.667	1.186	0.447	7	0.669
		Above 11	5	3.333	1.054			
Satisfaction	BI	Under 10	4	3.500	1.581	0.702	7	0.505
		Above 11	5	2.967	0.606			
	Exp	Under 10	4	4.000	0.720	1.845	7	0.107
		Above 11	5	2.800	1.121			
	ALL	Under 10	4	3.667	1.273	1.133	7	0.295
		Above 11	5	2.911	0.718			
SUS		Under 10	4	69.375	12.809	3.018	7	0.019*
		Above 11	5	44.500	21.316			
Usability		Under 10	4	3.740	1.060	0.775	7	0.464
		Above 11	5	3.264	0.791			
Att		Under 10	4	4.250	1.061	1.260	7	0.248
		Above 11	5	3.400	0.962			

Att: Attitude toward to Use Online Annotation System; EoL: Ease of Learning; EoU: Ease of Use; UI: User Interface Design; Use: Usefulness; BI: Behvioural Intention to Use; Exp: Expectation
*: $p < 0.05$; **: $p < 0.01$

When asking students what they do when they use browser, I gave them seven items to choose – playing games, posting article, reading book, reading news, study, using social network application, and watching movie. I think divide 7 items into three groups doesn't show much different, therefore I divide 7 items into two group 0-4 kinds and 5-7 kinds. Before analysis, I decide to put two more question for analysis, one is asking students when did they know e-reader for the first and when did they use e-reader for the first time. I think e-reader is one of the applications and these two questions is relate to the e-reader.

The two groups have significant difference on the two extra analysis questions, Observability-20, and the SUS value. The mean value for when is the first time they use

e-reader is the age, the 0-4 kinds their first time use e-reader is 11.333 years old and 5-7 is 16.833 years old. I think this reason have the same logic, the students who use browser or internet later they like try more and will keep use more items. The same reason also on the Observability-20 question. The SUS value has the different story, 0-4 kind give higher score (65) than the 5-7 kinds (50.833), the reason probably because they use more different applications, they have more related experience. When they use the Online Annotation System, they compare the Online Annotation System with their e-reader using experiences.

Table 45.

Independent t-test result for extract moderator and each factor (the total internet use items)

Factor	Subfactor	Total internet use items	Descriptive statistic			*t-test*		
			N	Mean	SD	*t*	*df*	*p*
Extra Moderators	When did your know e-reader for the first time?	0-4 kinds	3	11.333	4.041	-2.714	7	0.030*
		5-7 kinds	6	16.833	2.229			
	When did you first use any kind of e-readers (i.e., software or device that you use it to read articles or books)?	0-4 kinds	3	12.667	2.082	-2.521	7	0.040*
		5-7 kinds	6	18.000	3.286			
Moderators	Relative Advantage	0-4 kinds	3	4.000	1.732	0.174	7	0.867
		5-7 kinds	6	3.833	1.169			
	Compatibility	0-4 kinds	3	4.000	1.000	-0.370	7	0.722
		5-7 kinds	6	4.167	0.408			
	Observability	0-4 kinds	3	1.000	0.000	-4.719	5	0.005*
		5-7 kinds	6	3.333	1.211			

	Effectiveness - EoL	0-4 kinds	3	4.000	1.000	0.000	7	1.000
		5-7 kinds	6	4.000	1.265			
	Efficiency - Information	0-4 kinds	3	4.000	1.000	0.370	7	0.722
		5-7 kinds	6	3.667	1.366			
Diffusion of Innovation	Complexity	0-4 kinds	3	3.000	1.453	-0.591	7	0.573
		5-7 kinds	6	3.444	0.861			
	Trialability	0-4 kinds	3	3.889	1.925	0.581	7	0.580
		5-7 kinds	6	3.278	1.272			
	ALL	0-4 kinds	3	3.444	1.512	0.103	7	0.921
		5-7 kinds	6	3.361	0.963			
Effectiveness	EoU	0-4 kinds	3	3.750	0.500	0.218	7	0.834
		5-7 kinds	6	3.583	1.242			
	UI	0-4 kinds	3	3.267	1.332	-0.547	7	0.602
		5-7 kinds	6	3.633	0.742			
	ALL	0-4 kinds	3	3.481	0.958	-0.209	7	0.840
		5-7 kinds	6	3.611	0.843			
Efficiency	Use	0-4 kinds	3	3.444	1.347	-0.070	7	0.946
		5-7 kinds	6	3.500	1.027			
Satisfaction	BI	0-4 kinds	3	3.333	1.893	0.172	2.268	0.877
		5-7 kinds	6	3.139	0.687			
	Exp	0-4 kinds	3	3.778	0.694	0.837	7	0.430
		5-7 kinds	6	3.111	1.259			
	ALL	0-4 kinds	3	3.481	1.492	0.467	7	0.655
		5-7 kinds	6	3.130	0.836			
SUS		0-4 kinds	3	65.000	11.456	3.031	7	0.019*
		5-7 kinds	6	50.833	24.580			
Usability		0-4 kinds	3	3.560	1.220	0.188	7	0.856
		5-7 kinds	6	3.433	0.820			
Att		0-4 kinds	3	4.000	1.146	0.429	7	0.681
		5-7 kinds	6	3.667	1.080			

Att: Attitude toward to Use Online Annotation System; EoL: Ease of Learning; EoU: Ease of Use; UI: User Interface Design; Use: Usefulness; BI: Behvioural Intention to Use; Exp: Expectation

*: $p < 0.05$; **: $p < 0.01$

The Table 46 is one of the choices what students will do when they use browser and

students say when they use browser, they use it for reading book. All the choices are list in the Table 20 (in page 114). The numbers in the other choices have many differences that is the reason does not analysis the other choices. Does students use browser to reading book doesn't have any significant difference between who answer yes and no. The Table 20 has a study choice for students, if put the reading book and study in the choices in the same time, students could consider the reading book is more like the relaxing reading, not the serious reading like study. If considering student see the reading book as relaxing reading, then the reading book is one of the having fun activity like watch movie. I think this is the reason the reading book doesn't have any significant different between two answers.

Table 46.

Independent t-test result for what student do with browser? [reading book]

Factor	Subfactor	Student use browser to read book	Descriptive statistic			*t-test*		
			N	Mean	SD	*t*	*df*	*p*
Moderators	Relative Advantage	Yes	5	3.600	1.140	-0.742	7	0.482
		No	4	4.250	1.500			
	Compatibility	Yes	5	4.200	0.447	0.471	7	0.652
		No	4	4.000	0.816			
	Observability	Yes	5	3.000	1.000	0.986	7	0.357
		No	4	2.000	2.000			
	Effectiveness - EoL	Yes	5	3.800	1.304	-0.574	7	0.584
		No	4	4.250	0.957			
	Efficiency - Information	Yes	5	3.600	1.517	-0.471	7	0.652
		No	4	4.000	0.816			
Diffusion of Innovation	Complexity	Yes	5	3.600	0.863	0.999	7	0.351
		No	4	2.917	1.198			

	Trialability	Yes	5	3.600	1.116	0.246	4.581	0.816
		No	4	3.333	1.925			
	ALL	Yes	5	3.600	0.855	0.634	7	0.546
		No	4	3.125	1.390			
Effectiveness	EoU	Yes	5	3.550	1.385	-0.307	4.840	0.772
		No	4	3.750	0.408			
	UI	Yes	5	3.600	0.825	0.310	7	0.766
		No	4	3.400	1.120			
	ALL	Yes	5	3.578	0.938	0.038	7	0.971
		No	4	3.556	0.796			
Efficiency	Use	Yes	5	3.667	1.054	0.563	7	0.591
		No	4	3.250	1.167			
Satisfaction	BI	Yes	5	3.267	0.683	0.165	3.876	0.877
		No	4	3.125	1.601			
	Exp	Yes	5	3.467	1.017	0.382	7	0.714
		No	4	3.167	1.347			
	ALL	Yes	5	3.333	0.749	0.251	4.366	0.813
		No	4	3.139	1.398			
SUS		Yes	5	52.500	27.099	-1.463	7	0.187
		No	4	59.375	14.631			
Usability		Yes	5	3.512	0.891	0.128	7	0.902
		No	4	3.430	1.030			
Att		Yes	5	3.650	1.207	-0.389	7	0.709
		No	4	3.938	0.944			

Att: Attitude toward to Use Online Annotation System; EoL: Ease of Learning; EoU: Ease of Use; UI: User Interface Design; Use: Usefulness; BI: Behvioural Intention to Use; Exp: Expectation

*: $p < 0.05$; **: $p < 0.01$

When asking students when is their first time know the e-reader. When analyzing what is the difference in different age category, I decide to divide into three group, when they know e-reader is under age 12, between 13 and 17, and older than 18. The ANOVA result is shown in Table 47, there are no any significant difference between age differences in each factor. No matter when did they know the e-reader doesn't mean they

really know what e-reader can do, is the e-reader they know they will like, or does the e-reader really meet their need. Therefore, when did they know the e-reader doesn't affect their behavioural for the thinking of the e-reader.

Table 47.

ANOVA result for extract moderator and each factor (when is the first time students know the e-reader)

Factor	Subfactor	Students' first time know the e-reader	N	Mean	SD	*F*	*df*	*p*
Moderators	Relative Advantage	Under 12	2	3.500	2.121	0.245	(2, 6)	0.79
		From 13 to 17	4	4.250	0.957			
		Above 18	3	3.667	1.528			
	Compatibility	Under 12	2	4.000	1.414	0.152	(2, 6)	0.863
		From 13 to 17	4	4.250	0.500			
		Above 18	3	4.000	0.000			
	Observability	Under 12	2	1.000	0.000	1.556	(2, 6)	0.286
		From 13 to 17	4	3.000	1.826			
		Above 18	3	3.000	1.000			
	Effectiveness - EoL	Under 12	2	4.000	1.414	0.913	(2, 6)	0.451
		From 13 to 17	4	4.500	0.577			
		Above 18	3	3.333	1.528			
	Efficiency - Information	Under 12	2	3.500	0.707	0.496	(2, 6)	0.632
		From 13 to 17	4	4.250	0.500			
		Above 18	3	3.333	2.082			
Diffusion of Innovation	Complexity	Under 12	2	2.167	0.236	2.360	(2, 6)	0.175
		From 13 to 17	4	3.417	0.957			
		Above 18	3	3.889	0.962			
	Trialability	Under 12	2	3.333	2.357	0.046	(2, 6)	0.955
		From 13 to 17	4	3.667	1.440			
		Above 18	3	3.333	1.453			
	ALL	Under 12	2	2.750	1.296	0.389	(2, 6)	0.694
		From 13 to 17	4	3.542	1.092			
		Above 18	3	3.611	1.206			
Effectiveness	EoU	Under 12	2	3.750	0.707	0.937	(2, 6)	0.442

		From 13 to 17	4	4.063	0.375			
		Above 18	3	3.000	1.639			
	UI	Under 12	2	3.100	1.838			
		From 13 to 17	4	3.500	0.346	0.29	(2, 6)	0.754
		Above 18	3	3.800	1.058			
	ALL	Under 12	2	3.389	1.336			
		From 13 to 17	4	3.750	0.190	0.141	(2, 6)	0.872
		Above 18	3	3.444	1.281			
Efficiency	Use	Under 12	2	3.333	1.886			
		From 13 to 17	4	3.500	0.694	0.021	(2, 6)	0.979
		Above 18	3	3.556	1.388			
Satisfaction	BI	Under 12	2	2.917	2.475			
		From 13 to 17	4	3.417	0.799	0.121	(2, 6)	0.888
		Above 18	3	3.111	0.770			
	Exp	Under 12	2	3.667	0.943			
		From 13 to 17	4	3.250	1.450	0.091	(2, 6)	0.914
		Above 18	3	3.222	1.072			
	ALL	Under 12	2	3.167	1.964			
		From 13 to 17	4	3.361	0.999	0.035	(2, 6)	0.966
		Above 18	3	3.148	0.788			
SUS		Under 12	2	63.750	15.910			
		From 13 to 17	4	57.500	21.213	0.309	(2, 6)	0.745
		Above 18	3	47.500	29.475			
Usability		Under 12	2	3.380	1.669			
		From 13 to 17	4	3.640	0.563	0.097	(2, 6)	0.909
		Above 18	3	3.320	1.105			
Att		Under 12	2	3.875	1.591			
		From 13 to 17	4	4.125	0.661	0.546	(2, 6)	0.605
		Above 18	3	3.250	1.323			

Att: Attitude toward to Use Online Annotation System; EoL: Ease of Learning; EoU: Ease of Use; UI: User Interface Design; Use: Usefulness; BI: Behvioural Intention to Use; Exp: Expectation

*: $p < 0.05$; **: $p < 0.01$

Another moderator could affect student's thinking of the Online Annotation System

is when is students first time to use e-reader. In the when is the first time use e-reader analysis, I divide student's age into three groups as when is their first time to know e-reader. The ANOVA result is shown in Table 48, there are no any significant difference between age differences in each factor. No significant difference means no matter how early they use e-reader, students try the e-reader early or late don't affect they like to use e-reader or the Online Annotation System. If their first time use e-reader is late, doesn't mean when they use because they have confidence in e-reader or like to use the Online Annotation System. The first time they use e-reader maybe just try out what is e-reader what it can do, doesn't mean anything.

Table 48.

ANOVA result for extract moderator and each factor (when is the first time students use the e-reader)

Factor	Subfactor	Students' first time use the e-reader	N	Mean	SD	*F*	*df*	*p*
Moderators	Relative Advantage	Under 12	2	3.500	2.121	0.245	(2, 6)	0.790
		From 13 to 17	4	3.750	1.258			
		Above 18	3	4.333	1.155			
	Compatibility	Under 12	2	4.000	1.414	0.250	(2, 6)	0.787
		From 13 to 17	4	4.000	0.000			
		Above 18	3	4.333	0.577			
	Observability	Under 12	2	1.000	0.000	1.788	(2, 6)	0.246
		From 13 to 17	4	2.750	1.500			
		Above 18	3	3.333	1.528			
	Effectiveness - EoL	Under 12	2	4.000	1.414	3.316	(2, 6)	0.107
		From 13 to 17	4	3.250	0.957			
		Above 18	3	5.000	0.000			
	Efficiency - Information	Under 12	2	3.500	0.707	0.410	(2, 6)	0.681
		From 13 to 17	4	3.500	1.732			

		Above 18	3	4.333	0.577			
Diffusion of Innovation	Complexity	Under 12	2	2.167	0.236	2.116	(2, 6)	0.202
		From 13 to 17	4	3.750	0.631			
		Above 18	3	3.444	1.347			
	Trialability	Under 12	2	3.333	2.357	0.031	(2, 6)	0.970
		From 13 to 17	4	3.417	1.198			
		Above 18	3	3.667	1.764			
	ALL	Under 12	2	2.750	1.296	0.386	(2, 6)	0.696
		From 13 to 17	4	3.583	0.908			
		Above 18	3	3.556	1.417			
Effectiveness	EoU	Under 12	2	3.750	0.707	1.538	(2, 6)	0.289
		From 13 to 17	4	3.063	1.214			
		Above 18	3	4.333	0.520			
	UI	Under 12	2	3.100	1.838	0.643	(2, 6)	0.559
		From 13 to 17	4	3.350	0.700			
		Above 18	3	4.000	0.529			
	ALL	Under 12	2	3.389	1.336	1.204	(2, 6)	0.363
		From 13 to 17	4	3.222	0.743			
		Above 18	3	4.148	0.463			
Efficiency	Use	Under 12	2	3.333	1.886	0.278	(2, 6)	0.766
		From 13 to 17	4	3.250	0.877			
		Above 18	3	3.889	1.072			
Satisfaction	BI	Under 12	2	2.917	2.475	0.149	(2, 6)	0.865
		From 13 to 17	4	3.125	0.712			
		Above 18	3	3.500	0.866			
	Exp	Under 12	2	3.667	0.943	0.106	(2, 6)	0.901
		From 13 to 17	4	3.167	0.882			
		Above 18	3	3.333	1.764			
	ALL	Under 12	2	3.167	1.964	0.066	(2, 6)	0.937
		From 13 to 17	4	3.139	0.699			
		Above 18	3	3.444	1.160			
SUS		Under 12	2	63.750	15.910	1.781	(2, 6)	0.247
		From 13 to 17	4	41.875	18.750			

	Above 18	3	68.333	22.407			
Usability	Under 12	2	3.380	1.669			
	From 13 to 17	4	3.210	0.746	0.445	(2, 6)	0.660
	Above 18	3	3.893	0.740			
Att	Under 12	2	3.875	1.591			
	From 13 to 17	4	3.313	1.087	0.791	(2, 6)	0.496
	Above 18	3	4.333	0.629			

Att: Attitude toward to Use Online Annotation System; EoL: Ease of Learning; EoU: Ease of Use; UI: User Interface Design; Use: Usefulness; BI: Behvioural Intention to Use; Exp: Expectation
*: $p < 0.05$; **: $p < 0.01$

I ask students usually how many hours that they use in week days, students gave me a total number and I divide the total hours by 5 (days) – the first group is less than one hour, the second group is between 1 and 2 hours, and the third group is more than 2 hours in week days. The significant difference value is on the Compatibility, Effectiveness-Ease of Learning, and Ease of Use factor. In the question Compatibility, students who use browser less than one hour and more than three hours have 5 and 4 for the mean values. Maybe the first group (less than one hours use) every day they use browser just for study and the third group (more than 2 hours use) every day they use browser not just study, they may do some other relaxing activity and having fun with that activity.

Table 49.

Use ANOVA for – how many hours that students use browser in week day (the total hours / 5)

Factor	Subfactor	Students' use internet hr/day	N	Mean	SD	*F*	*df*	*Sig.*
Moderators	Relative Advantage	Less than 1	2	4.000	1.414			
		Between 1 and 2	5	4.200	1.304	0.580	(2, 6)	0.588
		More than 2	2	3.000	1.414			
	Compatibility	Less than 1	2	5.000	0.000	7.833	(2, 6)	0.021*

		Between 1 and 2	5	3.800	0.447			
		More than 2	2	4.000	0.000			
	Observability	Less than 1	2	1.500	0.707			
		Between 1 and 2	5	2.800	1.789	0.573	(2, 6)	0.592
		More than 2	2	3.000	1.414			
	Effectiveness - EoL	Less than 1	2	5.000	0.000			
		Between 1 and 2	5	4.200	0.837	6.091	(2, 6)	0.036*
		More than 2	2	2.500	0.707			
	Efficiency - Information	Less than 1	2	4.000	0.000			
		Between 1 and 2	5	4.200	0.837	1.749	(2, 6)	0.252
		More than 2	2	2.500	2.121			
Diffusion of Innovation	Complexity	Less than 1	2	2.500	0.236			
		Between 1 and 2	5	3.600	1.278	0.789	(2, 6)	0.496
		More than 2	2	3.333	0.000			
	Trialability	Less than 1	2	4.667	0.471			
		Between 1 and 2	5	3.400	1.673	1.247	(2, 6)	0.353
		More than 2	2	2.500	0.236			
	ALL	Less than 1	2	3.583	0.118			
		Between 1 and 2	5	3.500	1.467	0.203	(2, 6)	0.822
		More than 2	2	2.917	0.118			
Effectiveness	EoU	Less than 1	2	4.375	0.177			
		Between 1 and 2	5	3.950	0.570	8.736	(2, 6)	0.017*
		More than 2	2	2.125	0.884			
	UI	Less than 1	2	4.000	0.566			
		Between 1 and 2	5	3.360	1.043	0.311	(2, 6)	0.744
		More than 2	2	3.400	1.131			
	ALL	Less than 1	2	4.375	0.236			
		Between 1 and 2	5	3.950	0.792	1.506	(2, 6)	0.295
		More than 2	2	2.125	1.021			
Efficiency	Use	Less than 1	2	2.500	0.236			
		Between 1 and 2	5	3.600	1.011	1.352	(2, 6)	0.328
		More than 2	2	3.333	1.414			
Satisfaction	BI	Less than 1	2	4.667	0.471	1.653	(2, 6)	0.268

		Between 1 and 2	5	3.400	1.221			
		More than 2	2	2.500	0.000			
	Exp	Less than 1	2	5.000	0.236			
		Between 1 and 2	5	3.700	1.090	1.731	(2, 6)	0.255
		More than 2	2	2.750	1.179			
	ALL	Less than 1	2	4.000	0.236			
		Between 1 and 2	5	3.360	1.063	2.198	(2, 6)	0.192
		More than 2	2	3.400	0.393			
SUS		Less than 1	2	78.750	5.303			
		Between 1 and 2	5	56.000	17.642	4.751	(2, 6)	0.058
		More than 2	2	31.250	12.374			
Usability		Less than 1	2	4.420	0.198			
		Between 1 and 2	5	3.368	0.831	2.282	(2, 6)	0.183
		More than 2	2	2.800	0.905			
Att		Less than 1	2	5.000	0.000			
		Between 1 and 2	5	3.700	0.622	4.336	(2, 6)	0.068
		More than 2	2	2.750	1.414			

Att: Attitude toward to Use Online Annotation System; EoL: Ease of Learning; EoU: Ease of Use; UI: User Interface Design; Use: Usefulness; BI: Behvioural Intention to Use; Exp: Expectation

*: $p < 0.05$; **: $p < 0.01$

Chapter 6. Findings and discussion

6.1 Findings and discussion

When I compare the difference between before and after remove the “no annotation” chromosomes, I notice that the precision, recall, F2, and F0.5 values in previous methods are dropping. The shows the mean values of the precision, recall, F2, and F0.5 and standard deviation values of the precision, recall, F2, and F0.5 between the all chromosomes and remove “no annotation” chromosomes. When students’ number is less, I suppose the precision and recall values in all methods are better than more students. However, the previous methods get the worse value than “all students”. Maybe the more students that system have, the results will be more accurate. But, when I look at the standard deviation value, the more students the Standard, Quantitative, and Diffusion have the standard deviation value are bigger too. This means the more students the Standard, Quantitative, and Diffusion have, their clustering methods are more not stable.

Table 50.

The mean value and standard deviation value for total 8 reading activities

		All students		Delete students	
		Mean	SD	Mean	SD
GRACE	Precision	26.13%	16.27%	36.36%	8.26%
	Recall	27.04%	12.14%	34.87%	6.81%
	F2	26.53%	12.93%	34.99%	6.43%
	F0.5	26.04%	15.42%	35.88%	7.39%
Standard	Precision	48.68%	17.50%	28.00%	9.97%
	Recall	42.12%	9.42%	29.51%	8.65%
	F2	43.00%	10.25%	29.11%	8.91%
	F0.5	46.85%	15.05%	28.22%	9.71%
Quantitative	Precision	51.03%	24.36%	36.75%	8.67%

	Recall	39.53%	11.17%	36.46%	9.06%
	F2	41.01%	12.63%	36.28%	8.27%
	F0.5	47.65%	20.17%	36.46%	8.04%
Cosine	Precision	42.67%	6.32%	36.33%	9.09%
	Recall	45.06%	9.27%	36.74%	8.11%
	F2	44.45%	8.29%	36.58%	8.20%
	F0.5	43.03%	6.50%	36.34%	8.82%
Diffusion	Precision	52.93%	19.09%	42.93%	14.14%
	Recall	45.47%	16.53%	40.26%	11.45%
	F2	46.65%	16.82%	40.49%	11.62%
	F0.5	51.07%	18.28%	42.08%	13.27%

The Diffusion of Innovation only has significant positive relation with Effectiveness and Satisfaction; there is no significant relation between Diffusion of Innovation and Efficiency. Diffusion of Innovation evaluates students' acceptance in new technology acceptance. The results show that people who are easier to accept a new technology believe the Online Annotation System is effectiveness and has higher satisfaction toward the system. However, they may not believe the system is efficient. To understand why the relationship between Diffusion of Innovation and Efficiency, I need to review the tests in the micro-view.

The factors in Diffusion of Innovation are complexity and trialability. Based on the analysis results in Table 30, the Pearson correlation analysis shows that there is no significate correlation between the complexity and the system usability factors. On the other hand, the trialability has significate correlation with the system usability, behavioural intention to use, and expectation. This result shows that when an innovation comes out, if students have tried an e-reader application before, they willing give other e-reader innovation a chance. And also, if they have tried an e-reader application before,

they will have more interested to use the Online Annotation System and will have the expectation for the Online Annotation System.

Although the trialability also plays an important role in students' satisfaction toward the Online Annotation System, it has no significant relation to students' attitude toward using Online Annotation System. I also exam the relation of students' attitude toward using annotation system to their perceived behavioural intention of using the Online Annotation System and their expectation of what Online Annotation System should work. The results show that the two relationships are significantly related. However, the Pearson Correlation Coefficient between Trialability and Attitude is 0.664 with $p = 0.051$, which is very close to significant bound. Maybe the sample size is not good enough, so the test could not reflect the correlation between the two factors.

I am also interested in whether or not the lower-factors will affect other lower-factors in the Usability factor. The results show that students' perceived effectiveness doesn't affect their satisfaction toward the system, but students' perceived friendly designed user interface affects their perceived behavioural intention to use. This means the interface designing a friendly user interface would give students' a better user experience and have higher intention of using the system in the future.

There are some unexpected findings. When I analysis whether or not student's gender affects their behaviour when they use the Online Annotation system, I suppose their behaviour have significant difference. The result shows there are no difference between the gender. The possible reason is students will have same attitudes toward the new technology – such as the Online Annotation System – when they use it for learning. If they system is useful to them, both male and female students would like to use the

system for improving the academic achievement.

Another unexpected finding is there is no significant difference in two courses, no matter it is a freshmen or sophomore class. Maybe because of all students are studying in computer major, even they are in different year of the collage they still have the similar acceptance for the Online Annotation System.

There is one question regarding the possibility of students losing the focus on the reading because of the variety annotation ways on the e-readers. Liu's (2005) research found that eighty-three percent of participants indicated that they increased read electronically in past ten years. In Qayyum and Smith's (2015) research, most participants shows comfortable on doing everything online. Also, participants in Qayyum and Smith's research show the slow reading and re-reading can point to a better achievement. Liu's (2005) research also shows that when participants read on the screen, over 50 percentage of participants indicated that they increase the browsing and scanning, keyword spotting, one-time reading, reading selectively, and non-linear reading actions. The research results also show that that annotate on the reading material is important because the annotation can make readers the information they want in a short time. However, there is no evidence that various annotation options will distract students' focus.

Chapter 7. Conclusions

7.1 Summary

This research designs a bio-inspired clustering approach to cluster students' annotations based on students' annotation behaviour and implement this bio-inspired clustering approach and give students' annotation suggestions. Give students annotation suggestion could help students get better academic performance. This research has implemented two systems for students doing reading activities and teachers reviewing students' annotations and writing the feature for clustering groups and potential learning problems. This research can help teachers to identify student's problems in learning via their annotations and get clear idea of which parts of the text (or learning units) that most of students don't understand and have misconceptions. For researcher, in the research area of data clustering, this research proposes a bio-inspired method, which can cluster data entry in huge size fast and accurate. The huge size data means the chromosome is very long. For example, if an article has three thousand words and each word has at least three alphabets, the annotation behaviour might be transformed to a Huffman chromosome with at least nine thousand nucleobases and make the computation more complex. The proposed method can also be used for clustering sequential big data.

There is one question regarding the possibility of students losing the focus on the reading because of the variety annotation ways on the e-readers. Liu's (2005) research found that eighty-three percent of participants indicated that they increased read electronically in past ten years. In Qayyum and Smith's (2015) research, most participants shows comfortable on doing everything online. Also, participants in Qayyum and Smith's research show the slow reading and re-reading can point to a better

achievement. Liu's (2005) research also shows that when participants read on the screen, over 50 percentage of participants indicated that they increase the browsing and scanning, keyword spotting, one-time reading, reading selectively, and non-linear reading actions. The research results also show that that annotate on the reading material is important because the annotation can make readers the information they want in a short time. However, there is no evidence that various annotation options will distract students' focus.

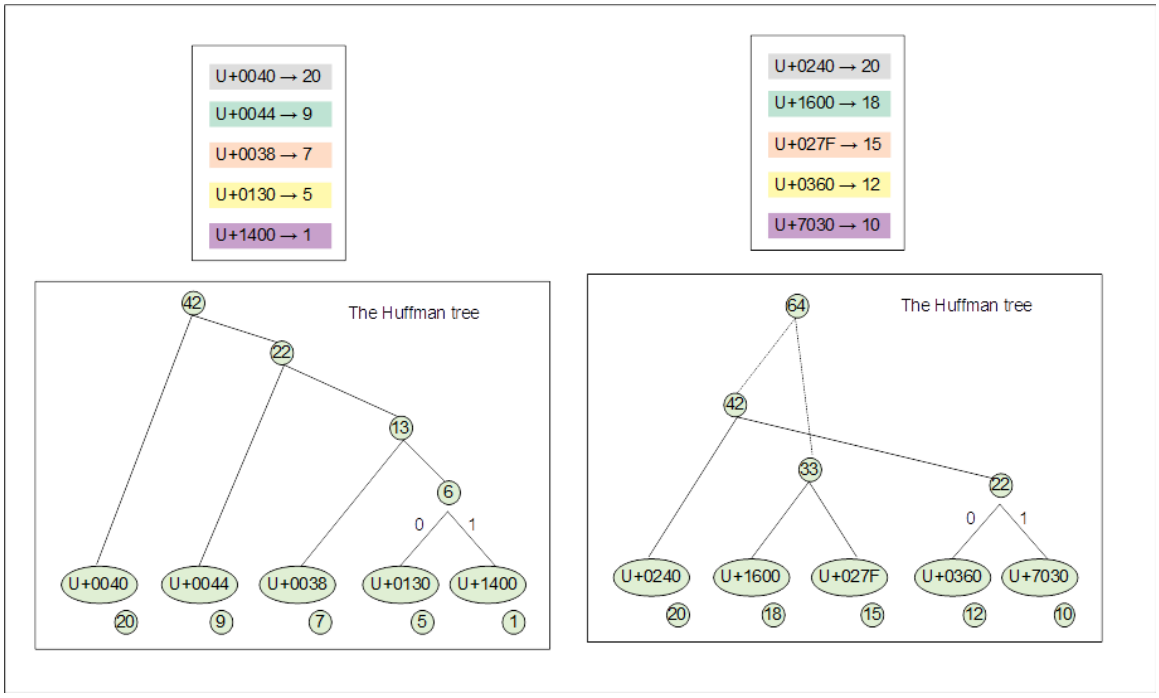


Figure 36. An example for Huffam tree

The platform and its bio-inspired student clustering methodology is innovative, powerful and flexible for any teachers to adopt into their courses, the platform is brand new. This research evaluates the effectiveness and usability of the platform. This platform would like to invite anyone, participating in the study, who is willing to adopt the platform for their courses and classes and having their students to do reading activities

online.

The GRACE algorithm didn't have the best performance comparing with the four previous clustering methods. If I want to improve the performance of the GRACE algorithm, the method of measuring the difference between students' annotation chromosomes could be redesigned.

The clustering method designed in this research also could cause a problem. For example, there are 10 students' annotation chromosomes, if there are four students have similar annotation behaviour and their values are small in their combinations, according to the clustering algorithm, these four students will be separated into two groups and each group will be chosen for one group's foundation combination. After the system finish the clustering process, teacher will see two groups students have similar annotation behaviour.

When I design the GRACE algorithm, students' annotation content is not considered as the factor for clustering students; however, integrating semantic analysis in the research is also a possible feature in the future. Foltz's (1996) research uses matrix to calculate the relation between words and documents. If I transfer the annotation features as one dimension in the matrix and the learning problems as the second dimension, this associate matrix can predict students' annotations features and be able to determine what kind of learning problems students might have.

The experiment results show that the performance of GRACE algorithm has no difference in the two selected courses. Maybe it is because the topics in the courses are close to procedural/strategy knowledge and students need more practical practice than reading in these courses. The GRACE algorithm might be more suitable for courses

which are tide to conceptual knowledge, such as history or biology which requires more reading.

7.2 Future works

For the mapping the clustering results to teacher edited results, I mapped it by my own judgment and because of I do the mapping manually, even I have the rules for the mapping, I still may adjust the mapping results in no awareness. For example, if I got another example as Figure 37 shows. If I still pick the Group X as the first group I want to find the mapping, the Group X (2/4 = 50%) will map to the Group T1 (2/3 = 66.67%), Group Y (2/4 = 50%) will map to the Group T2 (2/4 = 50%), and Group Z (2/4 = 50%) will map to the Group T3 (2/5 = 40%). But, because I randomly pick a group to start, if I start from Group Z (2/4 = 50%), Z will be mapped to the Group T2 (2/4 = 50%). The results will have different possibility which group is the next. If I chose the Group X for the next mapping, Group X will be mapped to the Group T1 and Group Y (1/4 = 25%) will be mapped to the Group T3 (1/5 = 20%).

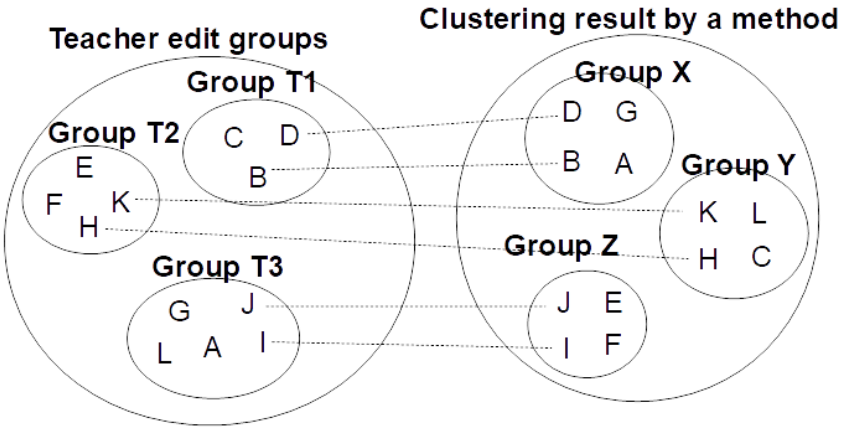


Figure 37. Another manually corresponded example

If I do the mapping one by one, the results will have different possibility. Because of the group is not much, sometimes I can see the good way in my first sight. For example, if I do pick the Group Y for the first to start mapping, I could see this will have bad results. Therefore, I will choose another group to start. I feel I randomly pick a group, but I was not so random. Sometimes I avoid the bad results in purpose, sometimes I avoid the bad results without awareness. If I implement the algorithm and let computer run the mapping for me, the kappa, precision, and recall analysis could be more precisely. Maybe the results will not as good as mapping manually, but the results will be more reliable.

This research designs a bio-inspired clustering algorithm for providing students useful annotation suggestions, but I didn't evaluate the recommendation mechanism; students does not get the recommendation from the system.

In addition, when the right time to provide the suggestions is another important issue. System should provide students just right after they take annotations or after they finish the reading and then provide all suggestions in the same time. If a researcher wants to provide the suggestions in the same time, how much suggestion should give maybe will cause a new problem?

Another related issue is how many sentences is the proper length for providing. In the section 3.4 Annotations Suggestion Provider Bio-inspired algorithm use the last sentence, when I making the decision I don't know the one sentence is good length or three sentences is better length?

Because of the suggestion providing issues that I can't solve, when I start the experiment, I turn off the suggestion providing function. I want to make the clustering

results useful, I decide to evaluate how good is this algorithm can do. The new experiment design can add the suggestion function for helping students' study.

I original plan in the experiment was to analyze students' learning performance in each reading activity by designing quizzes before and after using the Online Annotation System. However, after I finished the experiment, I couldn't find the same student ID in every quizzes I collected and was unable to analyze students' performance. The research design in the future should make sure the researchers can easily find student ID in every quiz.

When I compare the GRACE with the other clustering algorithms in previous research, the GRACE is quick enough comparing with the Diffusion and Cosine methods. I believe the GRACE algorithm can measure the huge size Huffman chromosomes very quickly, but the experiment in this research didn't have the long reading activities for students. The future research can ask teacher to provide long reading activities when design a new experiment.

There is a possibility to adopt the semantic into the GRACE. If system can make the student-annotation semantic tree right after students making the annotation instead of making the article-word tree, I think the semantic can be adopted into the GRACE system in the future.

This research didn't ask teacher provides long reading materials and the number of the clusters. The future research can ask teacher to provide long reading materials. I believe the long reading materials will have more features determine the different purposes (i.e. don't understand or important for the exam) in different annotation ways. When students have more purpose to annotate, teacher could cluster students more easily

based on the annotation features. Moreover, the recommendation annotation function could be more useful when the long reading. After the teacher helps the researcher to identify the features in different clusters, maybe the system could help the teacher to determine the problems students might have when they are reading the materials.

During the experiment, I have asked the teacher to provide students a small quiz every week and the quiz is relevant to what they have read in the previous week in order to understand student's academic performance in each week when they use the Online Annotation System. If the data is collected, the system could identify what students' annotation behaviour is corresponding to what type of students' academic performance. The system could provide the academic performance prediction based on students' annotation bahaviour to teachers in advanced so the teachers could support proper guidance to students who might fail in the learning activity.

References

AlQudah, A. A. (2014). Accepting Moodle by academic staff at the University of Jordan: Applying and extending tam in technical support factors. *European Scientific Journal*, 10(18), 183-200.

Bangor, A., Kortum, P., and Miller, J. (2009). Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of usability studies*, 4(3), 114-123.

Bounie, D., Eang, B., Sirbu, M., & Waelbroeck, P. (2013). Superstars and outsiders in online markets: An empirical analysis of electronic books. *Electronic Commerce Research and Applications*, 12(1), 52-59.

Blondin, J. (2009). Particle swarm optimization: A tutorial. Retrieved January 05, 2017 from site: http://www.cs.armstrong.edu/saad/csci8100/pso_tutorial.pdf

Brownlee, J. (2011). Clever algorithms: nature-inspired programming recipes. Jason Brownlee.

Brooke, J. (1996). SUS: A Quick and Dirty Usability Scale. In Jordan, P.W., Thomas, B., Weerdmeester B.A. & McClelland, I.L. (eds). *Usability Evaluation in Industry*, pp. 189-194. London: Taylor & Francis.

Chang M., Kuo, R., Chang, M., Kinshuk, and Kung, H. (2015, August). Online Annotation System and Student Clustering Platform. In the *Proceedings of the 8th International Conference on Ubi-Media Computing* (pp. 202-207). UMedia.

Chang, M., Kuo, R., Ying, K., Chiarella, A. F., Heh, J., & Kinshuk (2013). Clustering Annotations of a Digital Text with Bio-inspired Approaches. *Hybrid Technology*, 1(1), 1-10.

Chrzastowski, T. E., & Wiley, L. N. (2015). E-book Use and Value in the Humanities: Scholars' Practices and Expectations. *Library Resources & Technical Services*, *59*(4), 172.

Chen, C. M., & Chen, F. Y. (2014). Enhancing digital reading performance with a collaborative reading annotation system. *Computers & Education*, *77*, 67-81.

Chang, Y. C., Kao, W. Y., Chu, C. P., & Chiu, C. H. (2009). A learning style classification mechanism for e-learning. *Computers & Education*, *53*(2), 273-285.

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by Latent Semantic *Analysis. Journal of the American society for information science*, 41(6), 391-468. Retrieved January 05, 2017 from site: http://ajbasweb.com/old/ajbas/2013/February/490-499.pdf

Dlott, G., Maul, J. E., Buyer, J., & Yarwood, S. (2015). Microbial rRNA: rDNA gene ratios may be unexpectedly low due to extracellular DNA preservation in soils. *Journal of microbiological methods*, 115, 112-120.

Darzi, S., Kiong, T. S., & Salem, B. (2013). Overview of Particle Swarm Optimization (PSO) on its Applications and Methods. *Australian Journal of Basic and Applied Sciences,* 7(2), 490-499.

Dorigo, M & Stützle, T. (2004). Ant Colony Optimization. London, LON: A Bradford Book.

Foasberg, N. M. (2014). Student reading practices in print and electronic media. *College & Research Libraries*, 75(5), 705-723.

George, D., & Mallery, P. (2010). SPSS for windows step by step: a simple guide and

reference 18.0 update (11th ed.). Boston: Allyn & Bacon.
Huang, S. H., Chen, C. M., & Guo, J. C. (2012, April). Using collaborative reading annotation system with self-regulated learning mechanisms to promote reading performance in English. In *Consumer Electronics, Communications and Networks (CECNet), 2012 2nd International Conference on* (pp. 3687-3690). IEEE.
Hoff, C., Wehling, U., & Rothkugel, S. (2009). From paper-and-pen annotations to artefact-based mobile learning. *Journal of Computer Assisted Learning*, *25*(3), 219-237.
Kodituwakku, S.R., & Amarasinghe, U.S. (2010). Comparison of lossless data compression algorithms for text data. *Indian Journal of Computer Science and Engineering (IJCSE)*, 1(4), 416-425.
Liu, F., Yu, C., & Meng, W. (2002, November). Personalized web search by mapping user queries to categories. In *Proceedings of the eleventh international conference on Information and knowledge management* (pp. 558-565). ACM.
Liu, Z. (2005). Reading behavior in the digital environment: Changes in reading behavior over the past ten years. *Journal of documentation*, *61*(6), 700-712.
Lopatovska, I., Slater, A., Bronner, C., El Mimouni, H., Lange, L., & Ludas Orlofsky, V. (2014). In transition: academic e-book reading in an institution without e-books. *Library Review*, *63*(4/5), 261-275.
Lund, A. M. (2001). Measuring Usability with the USE Questionnaire. Usability and User Experience Newsletter of the STC Usability SIG. 8.
Manning, C. D., Raghavan, P. & Schütze H. (2008). Rocchio classification, Introduction to Information Retrieval, Retrieved September 26, 2016 from http://nlp.stanford.edu/IR-book/html/htmledition/rocchio-classification-1.html
Mathpal, D., Mittal, D., & Mehta, S. (2017). A Research Paper on Lossless Data Compression Techniques. *International Journal for Innovative Research in Science & Technology (IJIRST)*, 4(1), 190-194.
Melenhorst, M. (2005, July). Observing professionals taking notes on screen. In *IPCC 2005. Proceedings. International Professional Communication Conference, 2005.* (pp. 540-545). IEEE.
Monge, R. E., & Crespo, J. L. (2014, July). Comparison of complexity measures for DNA sequence analysis. In *Bio-inspired Intelligence (IWOBI), 2014 International Work Conference on* (pp. 71-75). IEEE.
Monge, R. E., & Crespo, J. L. (2015). Analysis of Data Complexity in Human DNA for Gene-Containing Zone Prediction. *Entropy*, *17*(4), 1673-1689.
Park, Y., & Chen, J. V. (2007). Acceptance and adoption of the innovative use of smartphone. *Industrial Management & Data Systems*, 107(9), 1349 - 1365.
Pearson, J., Buchanan, G., & Thimbleby, H. (2012, September). Investigating collaborative annotation on slate pcs. In *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services* (pp. 413-416). ACM.
Qayyum, M. A. & Smith, D. (2015). Learning from student experiences for online assessment tasks. *Information Research*, 20(2), 674. Retrieved April 15, 2019 from site: http://www.informationr.net/ir/20-2/paper674.html#.XLS3MpgzaUk
Quadir, B., Yang, J. C., Chen, N. S., & Shih, M. J. A. (2017). The Effects of Perceived

Innovation Game Attributes by Learners on Learning Performance in a Game-Based Achievement Learning System. *In Innovations in Smart Learning* (pp. 151-160). Springer Singapore.

Ramya, K.A., & Pushpa, M. (2016). Comparative Study on Different Lossless Data Compression Methods. *International Journal of Scientific Engineering and Applied Science (IJSEAS)*, 2(1), 273-278.

Rogers, E. M., (2003). Diffusion of innovations (5th ed.). New York, Free Press

Su, A. Y., Yang, S. J., Hwang, W. Y., & Zhang, J. (2010). A Web 2.0-based collaborative annotation system for enhancing knowledge sharing in collaborative learning environments. *Computers & Education*, *55*(2), 752-766.

Su, A. Y. S., & Yang, S. J. H. (2010). Improving annotation categorization performance through integrated social annotation computation. *Expert Systems with Applications,* 37(12), 8736-8744.

Tashman, C. S., & Edwards, W. K. (2011, May). Active reading and its discontents: the situations, problems and ideas of readers. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2927-2936). ACM

Tan, P-N, Steinbach, M., & Kumar, V. (2006). Introduction to data mining. In Library of congress (pp. 496-526). Retrieved from: https://www-users.cs.umn.edu/~kumar/dmbook/ch8.pdf

Tullis, T. S., & Stetson, J. N. (2004, June). A Comparison of Questionnaires for Assessing Website Usability. In *Proceedings of the Usability Professionals Association Conference* (pp. 7-11). UPA 2004.

Wolfe, J. L. (2000, June). Effects of annotations on student readers and writers. In *Proceedings of the fifth ACM conference on Digital libraries* (pp. 19-26). ACM

Yang, S. J., Chen, I. Y. L., & Shao, N. W. (2004). Ontology Enabled Annotation and Knowledge Management for Collaborative Learning in Virtual Learning Community. *Educational Technology & Society*, *7*(4), 70-81.

Yang, S. J., Zhang, J., Su, A. Y., & Tsai, J. J. (2011). A collaborative multimedia annotation tool for enhancing knowledge sharing in CSCL.*Interactive Learning Environments*, *19*(1), 45-62.

Yang, Y., & Chute, C. G. (1992, August). A linear least squares fit mapping method for information retrieval from natural language texts. In *Proceedings of the 14th conference on Computational linguistics-Volume 2* (pp. 447-453). Association for Computational Linguistics.

Yang, Y., & Chute, C. G. (1994). An example-based mapping method for text categorization and retrieval. *ACM Transactions on Information Systems (TOIS)*, *12*(3), 252-277.

Ying, K., Chang, M., Chiarella, A. F., & Heh, J. S. (2012, July). Clustering Students Based on their Annotations of a Digital Text. In *Technology for Education (*T4E), 2012 IEEE Fourth International Conference on (pp. 20-25). IEEE.

"Pollution" (n.d.), Pollusion. Retrieved February 21, 2019 from http://webpage.pace.edu/jb44525n/page5.html

Appendix A: Reading activities in French (original version) and English version (translated by Google translate)

All reading activities for HTML course

Reading activity title: Introduction HTML (in English: Introduction HTML)

Material in French:	Material in English: (translated by Google translate)
Introduction L'HyperText Markup Language, généralement abrégé HTML, est le langage de balisage conçu pour représenter les pages web. C'est un langage permettant d'écrire de l'hypertexte, d'où son nom. HTML permet également de structurer sémantiquement et logiquement et de mettre en forme le contenu des pages, d'inclure des ressources multimédias dont des images, des formulaires de saisie, et des programmes informatiques. Il permet de créer des documents interopérables avec des équipements très variés de manière conforme aux exigences de l'accessibilité du web. Il est souvent utilisé conjointement avec le langage de programmation JavaScript et des feuilles de style en cascade (CSS). Cet exemple contient du texte, cinq balises et une référence d'entité : • <TITLE> est la balise ouvrante de l'élément TITLE. • </TITLE> est la balise fermante de l'élément TITLE. • Exemple de HTML est le contenu de l'élément TITLE. • <A HREF=cible.html> est la balise ouvrante de l'élément A, avec :	Introduction HyperText Markup Language, usually abbreviated HTML, is the markup language designed to represent web pages. It is a language for writing hypertext, hence its name. HTML also allows semantically and logically structuring and formatting the content of pages, including multimedia resources including images, input forms, and computer programs. It makes it possible to create interoperable documents with a wide variety of equipment in accordance with the requirements of web accessibility. It is often used in conjunction with the JavaScript programming language and Cascading Style Sheets (CSS). This example contains text, five tags, and an entity reference: • <TITLE> is the opening tag of the TITLE element. • </ TITLE> is the closing tag of the TITLE element. • HTML example is the content of the TITLE element. • <A HREF=cible.html> is the opening tag of element A, with: • HREF = target.html, the HREF attribute whose value is target.html.

• HREF=cible.html, l'attribut HREF dont la valeur est cible.html. • <P> est la balise ouvrante de l'élément P. Toutefois, elle est utilisée ici comme s'il s'agissait d'un séparateur de paragraphe, et c'est même ainsi qu'elle est souvent présentée dans les plus anciennes documentations de HTML. Il s'agit de la balise ouvrante du paragraphe. La balise fermante de l'élément P, qui est optionnelle, est ici omise. L'élément P est implicitement terminé lorsqu'un nouveau paragraphe commence ou que l'élément parent est fermé (cas présent). • Les balises peuvent être indifféremment écrites en minuscules ou majuscules.	• <P> is the opening tag of the P element. However, it is used here as if it were a paragraph separator, and that is how it is often presented in the oldest HTML documentation. This is the opening tag of the paragraph. The closing tag of the P element, which is optional, is omitted here. The P element is implicitly terminated when a new paragraph begins or the parent element is closed (present case). • Tags can be indifferently written in lowercase or uppercase letters.

Reading activity title: généralité (in English: generality)

Material in French:	Material in English: (translated by Google translate)
Généralité Le HTML définit le contenu d'une page. Le CSS permet, lui, d'arranger le contenu et de définir la présentation : couleurs, image de fond, marges, taille du texte… Le CSS a besoin d'une page HTML pour fonctionner. C'est pour cela que nous allons d'abord apprendre les bases du HTML avant de nous occuper de la décoration en CSS. Au fil du temps, les langages HTML et CSS ont beaucoup évolué. Dans la toute première version de HTML (HTML 1.0) il n'était même pas possible d'afficher des images ! Les versions de HTML • HTML 1 • HTML 2 • HTML 3 • HTML 4 • HTML 5 : c'est la dernière version. De plus en plus répandue, elle fait beaucoup parler d'elle car elle apporte de nombreuses améliorations comme la possibilité d'inclure facilement des vidéos, un meilleur agencement du contenu, de nouvelles fonctionnalités pour les formulaires, etc. C'est cette version que nous allons découvrir ensemble. Les versions de CSS • CSS 1 • CSS 2 • CSS 3 : c'est la dernière version, qui apporte des fonctionnalités particulièrement attendues comme les bordures arrondies, les dégradés, les ombres, etc.	Generality HTML defines the content of a page. CSS allows you to arrange the content and define the presentation: colors, background image, margins, text size ... CSS needs an HTML page to work. That's why we will first learn the basics of HTML before we take care of CSS decoration. Over time, HTML and CSS languages have evolved a lot. In the very first version of HTML (HTML 1.0) it was not even possible to display images! HTML versions • HTML 1 • HTML 2 • HTML 3 • HTML 4 • HTML 5: This is the latest version. Increasingly popular, it is widely talked about as it brings many improvements such as the ability to easily include videos, better layout of content, new features for forms, etc. It is this version that we will discover together. CSS versions • CSS 1 • CSS 2 • CSS 3: this is the latest version, which brings particularly expected features like rounded borders, gradients, shadows, etc. There are actually many software dedicated to the creation of websites. Here are some software:

Il existe effectivement de nombreux logiciels dédiés à la création de sites web. Voici quelques logiciels : • Sublime Text ; • Notepad ; • Brackets ; • jEdit ; • PSpad ; • ConTEXT … Le navigateur est le programme qui nous permet de voir les sites web.Le travail du navigateur est de lire le code HTML et CSS pour afficher un résultat visuel à l'écran. Si votre code CSS dit « Les titres sont en rouge », alors le navigateur affichera les titres en rouge. Le rôle du navigateur est donc essentiel ! ces navigateurs se ressemblent beaucoup. Mais les navigateurs n'affichent pas toujours un même site web exactement de la même façon. Pourquoi ? Cela est dû au fait que les navigateurs ne connaissent pas toujours les dernières fonctionnalités de HTML et CSS. Par exemple, Internet Explorer a longtemps été en retard sur certaines fonctionnalités CSS.	• Sublime Text; • Notepad; • Brackets; • I'm saying ; • PSpad; • ConTEXT ... The browser is the program that allows us to see the websites. The browser's job is to read the HTML and CSS code to display a visual result on the screen. If your CSS code says "Titles are in red," then the browser will display the titles in red. The role of the browser is essential! these browsers are very similar. But browsers do not always display the same website exactly the same way. Why ? This is because browsers do not always know the latest features of HTML and CSS. For example, Internet Explorer has long been behind some CSS features.

Reading activity title: Insertion d'image (in English: Image insertion)

Material in French:	Material in English: (translated by Google translate)
Insertion d'image : L'élément <IMG> : On utilise l'élément <IMG> pour placer les images. Deux attributs sont obligatoires : • L'attribut src pour spécifier le nom du fichier image à charger. Les fichiers image doivent impérativement être aux formats GIF ou JPG. • L'attribut alt pour indiquer un contenu alternatif, c'est à dire un texte à afficher à la place de l'image lorsque, pour différentes raisons, elle n'apparaît pas. L es attributs optionnels de l'élément <IMG> L'attribut align : Il indique le mode d'alignement de l'image. Il y a 5 valeurs possibles : top, middle, bottom, left, right. Les trois premiers concernent l'alignement vertical par rapport à la ligne. Les attributs width et height : Ils permettent de déterminer la hauteur et la largeur de l'image, en pixels. Il faut toujour éviter de les utiliser pour redimensionner l'image. Mieux vaut utiliser un logiciel de dessin pour modifier la taille de l'image source : • Essayer d'agrandir une image par ce moyen donne de mauvais résultats (voir à droite) • Réduire une image par ce biais n'altère pas la qualité de l'affichage. En revanche, dans ce cas, l'image téléchargée sur le réseau est trop grosse, ce qui fait perdre beaucoup de temps à l'utilisateur. Les images cliquables : Utiliser une image dans un lien :	image ion: The <IMG> element: The <IMG> element is used to place the images. Two attributes are required: • The src attribute to specify the name of the image file to load. Image files must be in GIF or JPG format. • The alt attribute to indicate an alternative content, ie a text to display in place of the image when, for various reasons, it does not appear. The optional attributes of the <IMG> element The attribute align: It indicates how to align the image. There are 5 possible values: top, middle, bottom, left, right. The first three relate to vertical alignment with respect to the line. The width and height attributes: They make it possible to determine the height and the width of the image, in pixels. Always avoid using them to resize the image. Better to use a drawing software to change the size of the source image: • Trying to enlarge an image by this means gives poor results (see right) • Reducing an image in this way does not affect the quality of the display. On the other hand, in this case, the image downloaded on the network is too big, which is wasting the user a lot of time. Clickable images: Use an image in a link: The use of an image as a label in a link is of course no problem. Just place an IMG element in an A

L'utilisation d'une image comme étiquette dans un lien ne pose bien entendu aucun problème. Il suffit de placer un élément IMG dans un élément A. Attention toutefois : par défaut les images cliquables ont une bordure de la couleur des liens. Pour la supprimer, il faut utiliser l'attribut border (en lui attribuant la valeur 0). Exemple d'insertion d'image cliquable : <A href="exemples/lefrance.html"><IMG src=lefranc_.gif alt="Le France"></A>	element. Be careful though: by default the clickable images have a border of the color of the links. To delete it, use the border attribute (giving it the value 0). Example of clickable image insertion: <A href="examples/lefrance.html"> <IMG src = lefranc_.gif alt = "France"> </A>

Reading activity title: Tableau HTML (in English: HTML table)

Material in French:	Material in English: (translated by Google translate)
Tableau HTML HTML permet de réaliser des tableaux avec réglage de l'encadrement, de la taille et de l'espacement des cellules. Chaque cellule peut contenir du texte, des listes, des images, des liens hypertextes, des éléments de formulaire... Structure d'un tableau : Un tableau est décrit par différents éléments : • L'élément TABLE correspond au tableau lui-même • L'élément TR est utilisé pour définir chacune des lignes du tableau • L'élément TD est utilisé pour chaque cellule L'élément TABLE Un tableau est donc débuté par une balise <TABLE> et se termine sur une balise </TABLE>. Entre les deux, on définira les lignes et les cellules. Les principaux attributs applicables à l'élément TABLE sont : Border : pour spécifier l'épaisseur de la bordure. A noter que si on ne précise pas la taille des bordures, il n'y en a pas. Cellpading : pour spécifier l'espace entre bordures et contenu des cellules. Cellspacing : pour spécifier l'épaisseur des bordures entre cellules. Width : permet de déterminer quelle proportion de la largeur de la fenêtre doit être occupée. Entêtes : L'élément TH permet de définir des cellules d'entête. Les navigateurs visuels par exemple	HTML table HTML allows you to create tables with framing, cell size and spacing. Each cell can contain text, lists, images, hypertext links, form elements ... Structure of a table: A table is described by different elements: • The TABLE element corresponds to the table itself • The TR element is used to define each row of the table • The TD element is used for each cell The TABLE element A table is started with a <TABLE> tag and ends with a </ TABLE> tag. Between the two, we will define the lines and the cells. The main attributes applicable to the TABLE element are: Border: to specify the thickness of the border. Note that if we do not specify the size of the borders, there is none. Cellpading: to specify the space between borders and cell contents. Cellspacing: to specify the thickness of the borders between cells. Width: Determines how much of the width of the window should be occupied. Heads: The TH element is used to define header cells. Visual browsers for example use this information to put these cells in bold. Legends: The CAPTION element allows you to place a

utilisent cette information pour mettre ces cellules en gras. Légendes : L'élément CAPTION permet de placer une légende au-dessus ou au-dessous d'un tableau (selon que l'attribut align a la valeur top ou bottom).	legend above or below a table (depending on whether the align attribute is top or bottom).

Reading activity title: commentaires (in English: comments)

Material in French:	Material in English: (translated by Google translate)
Les commentaires Les lignes de commentaires vont être des lignes de texte que l'on va écrire au milieu de notre code, afin de donner des indications sur ce que fait le code en question. Les commentaires seront invisibles pour vos visiteurs, ils ne servent qu'aux développeurs créant ou lisant le code. Les commentaires vont être très utile dans deux situations : 1. Dans le cas d'un gros / long projet, afin de bien se rappeler soi même pourquoi nous avons écrit tel ou tel code, ou encore pour se repérer dans le code ; 2. Si l'on souhaite distribuer son code, ou si l'on travaille à plusieurs, cela fait beaucoup plus professionnel et permet aux autres développeurs de comprendre beaucoup plus rapidement et facilement le code distribué. Les commentaires peuvent être mono-ligne ou multi-lignes. Pour écrire un commentaire en HTML, nous allons nous y prendre de la façon suivante : < !- - je suis un commentaire, je ne serai pas affiché - -> Faîtes bien attention à la syntaxe des commentaires : il y a un point d'exclamation au début mais il n'y en a pas à la fin.	Comments The lines of comments will be lines of text that we will write in the middle of our code, to give indications on what the code in question does. The comments will be invisible to your visitors, they are only for developers creating or reading the code. The comments will be very useful in two situations: 1. In the case of a big / long project, to remember yourself why we wrote a particular code, or to find a way in the code; 2. If you want to distribute your code, or if you work with others, it's a lot more professional and allows other developers to understand the distributed code much more quickly and easily. Comments can be single-line or multi-line. To write a comment in HTML, we will do it as follows: <! - - I have a comment, I will not be displayed - - > Pay attention to the syntax of the comments: there is an exclamation point at the beginning but there is none at the end.

Reading activity title: les liens (in English: the links)

Material in French:	Material in English: (translated by Google translate)
Un lien, en HTML, va généralement servir à "transporter" un utilisateur d'un endroit vers un autre après que celui-ci ait cliqué dessus. Pour créer des liens en HTML, nous allons utiliser l'élément a accompagné de son attribut href (hypertext reference) qui va prendre comme valeur la cible du lien. La cible d'un lien est tout simplement l'adresse de la page de destination du lien. Nous allons pouvoir distinguer deux grands types de liens : les liens permettant de se déplacer d'une page vers une autre à travers un même site et les liens permettant de se rendre sur d'autres sites. On appellera ce premier type de liens des liens internes tandis que le second type va correspondre aux liens externes. Dans les deux cas, seule la façon dont on va construire la valeur de l'attribut href va changer. On peut faire un lien d'une pagea.htmlvers une pageb.html, mais on peut aussi faire un lien vers un autre site (par exemple,http://www.siteduzero.com). Dans les deux cas, le fonctionnement est le même. Si vous voulez faire un lien vers un autre site, il suffit donc de copier son adresse (on parle d'URL) enhttp://. Notez que certains liens commencent parfois parhttps://(sites sécurisés, comme OpenClassrooms) ou d'autres préfixes (ftp://,…).	in HTML, it will usually be used to "transport" a user from one place to another after clicking on it. To create links in HTML, we will use the element accompanied by its href attribute (hypertext reference) which will take as value the target of the link. The target of a link is simply the address of the landing page of the link. We will be able to distinguish two major types of links: links to move from one page to another through the same site and links to visit other sites. We will call this first type of links internal links while the second type will correspond to external links. In both cases, only the way we will build the value of the href attribute will change. We can make a link from a page.html to a pageb.html, but we can also link to another site (for example, http: //www.siteduzero.com). In both cases, the operation is the same. If you want to link to another site, just copy its address (we speak of URL) in http: //. Note that some links sometimes begin at https: // (secure sites, like OpenClassrooms) or other prefixes (ftp: //, ...).

Reading activity title: formulaire (in English: form)

Material in French:	Material in English: (translated by Google translate)
Toute page HTML peut être enrichie de formulaires interactifs, qui invitent vos visiteurs à renseigner des informations : saisir du texte, sélectionner des options, valider avec un bouton… tout est possible ! Nous arrivons cependant aux limites du langage HTML car il faut ensuite pouvoir analyser les informations que le visiteur a saisies… et cela ne peut pas se faire en langage HTML. Comme nous allons le voir, le traitement des résultats doit s'effectuer dans un autre langage, par exemple le PHP. En attendant, nous avons un grand nombre de nouvelles balises HTML à découvrir. Bienvenue dans le monde merveilleux des formulaires, un monde où les boutons, les cases à cocher et les listes déroulantes vivent en harmonie. Créer un formulaire Lorsqu'il vous prend subitement l'envie d'insérer un formulaire dans votre page HTML, vous devez pour commencer écrire une balise<form> </form>. C'est la balise principale du formulaire, elle permet d'en indiquer le début et la fin.	Any HTML page can be enriched with interactive forms, which invite your visitors to enter information: enter text, select options, validate with a button ... anything is possible! However, we come to the limits of the HTML language because we must then be able to analyze the information that the visitor has entered ... and this can not be done in HTML. As we will see, the results must be processed in another language, for example PHP. In the meantime, we have a large number of new HTML tags to discover. Welcome to the wonderful world of forms, a world where buttons, checkboxes and drop-down lists live in harmony. Create a form When you suddenly feel like inserting a form into your HTML page, you must first write a <form> </ form> tag. This is the main tag of the form, it allows to indicate the beginning and the end.

Reading activity title: methodes (in English: methods)

Material in French:	Material in English: (translated by Google translate)
• Problème n°1 : comment envoyer le texte saisi par le visiteur ? Par quel moyen ? • Problème n°2 : une fois que les données ont été envoyées, comment les traiter ? Souhaitez-vous recevoir le message automatiquement par mail ou préférez-vous qu'un programme se charge de l'enregistrer quelque part, puis de l'afficher sur une page visible par tout le monde ? Pour fournir les réponses à ces deux problèmes, vous devez ajouter deux attributs à la balise<form>: • method: cet attribut indique par quel moyen les données vont être envoyées (réponse au problème n°1). Il existe deux solutions pour envoyer des données sur le Web : o method="get": c'est une méthode en général assez peu adaptée car elle est limitée à 255 caractères. La particularité vient du fait que les informations seront envoyées dans l'adresse de la page (http://…), mais ce détail ne nous intéresse pas vraiment pour le moment. La plupart du temps, je vous recommande d'utiliser l'autre méthode :post. o method="post": c'est la méthode la plus utilisée pour les formulaires car elle permet d'envoyer un grand nombre d'informations. Les données saisies dans le formulaire ne transitent pas par la barre d'adresse. • action: c'est l'adresse de la page ou du programme qui va traiter les informations (réponse au problème n°2). Cette page se chargera de vous envoyer un e-mail avec le message si c'est ce que vous voulez, ou bien d'enregistrer le message avec	• Problem # 1: How to send the text entered by the visitor? In what way ? • Issue # 2: Once the data has been sent, how to process it? Would you like to receive the message automatically by email or would you prefer a program to save it somewhere and then post it on a page visible to everyone? To provide the answers to both of these problems, you must add two attributes to the <form> tag: • method: This attribute indicates how the data will be sent (answer to problem # 1). There are two ways to send data to the web: o method = "get": this is a method that is generally not very suitable because it is limited to 255 characters. The peculiarity comes from the fact that the information will be sent in the address of the page (http: // ...), but this detail does not really interest us for the moment. Most of the time, I recommend using the other method: post. o method = "post": this is the most used method for forms because it allows to send a lot of information. The data entered in the form does not pass through the address bar. • action: it is the address of the page or program that will process the information (answer to problem # 2). This page will send you an e-mail with the message if that's what you want, or save the message with all the others in a database. This can not be done in HTML and CSS, we will usually use another language that you may have heard about: PHP.

tous les autres dans une base de données. Cela ne peut pas se faire en HTML et CSS, on utilisera en général un autre langage dont vous avez peut-être entendu parler : PHP.	

Reading activity title: le langage CSS (in English: CSS language)

Material in French:	Material in English: (translated by Google translate)
CSS: Les CSS, Cascading Style Sheets (feuilles de styles en cascade), servent à mettre en forme des documents web, type page HTML ou XML. Par l'intermédiaire de propriétés d'apparence (couleurs, bordures, polices, etc.) et de placement (largeur, hauteur, côte à côte, dessus-dessous, etc.), le rendu d'une page web peut être intégralement modifié sans aucun code supplémentaire dans la page web. Les feuilles de styles ont d'ailleurs pour objectif principal de dissocier le contenu de la page de son apparence visuelle. Ceci permet : • de ne pas répéter dans chaque page le même code de mise en forme • d'utiliser des styles génériques, avec des noms explicites (par exemple un style encadré pour du texte ou des images) • de pouvoir changer l'apparence d'un site web complet en ne modifiant qu'un seul fichier • de faciliter la lecture du code de la page La puissance et de l'intérêt des CSS peut être démontrée en modifiant radicalement l'apparence d'une page, sans changer son code HTML... Bref les CSS permettent de gagner en productivité et en maintenabilité des sites web, tout en offrant des possibilités graphiques incontestables. Lorsqu'on se lance dans la conception de pages web, il faut apprendre les CSS !	CSS: Cascading Style Sheets (CSS) are used to format web documents, such as HTML page or XML. Through appearance properties (colors, borders, fonts, etc.) and placement (width, height, side-by-side, top-bottom, etc.), the rendering of a web page can be completely modified without any additional code in the web page. The main purpose of style sheets is to separate the content of the page from its visual appearance. This allows : • Do not repeat in each page the same formatting code • use generic styles, with explicit names (for example a framed style for text or images) • to be able to change the appearance of a complete website by modifying only one file • to facilitate the reading of the code of the page The power and the interest of CSS can be demonstrated by radically modifying the appearance of a page, without changing its HTML code ... In short, the CSS make it possible to gain in productivity and maintainability of the websites, while offering undeniable graphic possibilities. When you start designing web pages, you have to learn CSS!

Reading activity title: Reglement syntaxique de base (in English: Basic syntax rules)

Material in French:	Material in English: (translated by Google translate)
Reglement syntaxique de base : Casse Les feuilles de styles CSS ne sont pas sensibles à la casse : elles ne tiennent pas compte des majuscules et minuscules. Exception faite pour les éléments n'obéissant pas directement aux règles de syntaxe CSS, notamment les attributs id et class (dont le nommage est assuré par le rédacteur : vous), les noms des polices de caractères (exemple : "Trebuchet MS"), et les suffixes d'URL ne répondant pas à ces règles. Mise en forme du code Les feuilles de styles CSS ne tiennent pas compte des espaces et retours à la ligne. Commentaires Les commentaires commencent par une barre de fraction suivie d'un astérique « /* », et se concluent par la succession de caractères inverse « */ ». Ils sont facultatifs, voire inutiles, pour les modifications mineures d'affichage (inutile d'indiquer que l'on souligne, cela se lit facilement), mais indispensables pour les mises en pages importantes (inscrire par exemple la taille minimale d'une marge pour avoir la place d'insérer le menu permet de ne pas commettre de maladresse lors d'une future modification du fichier). A noter que Par rapport à l'HTML sans feuille de style, on remarque : • la balise de l'en-tête <link rel= "stylesheet" …> qui indique où trouver la feuille de	Basic syntax rules: Broken CSS style sheets are not case-sensitive: they are not case-sensitive. Except for items that do not directly obey CSS syntax rules, including id and class attributes (which are nicked by the editor: you), font names (eg "Trebuchet MS"), and URL suffixes that do not meet these rules. Code formatting CSS stylesheets do not include spaces and line breaks. comments Comments begin with a slash followed by an asterisk "/ *", and end with the reverse sequence of characters "* /". They are optional, if not useless, for minor changes to the display (needless to say that it is emphasized, this is easy to read), but essential for large layouts (for example, write the minimum size of a margin to have the place to insert the menu makes it possible not to commit clumsiness during a future modification of the file). Note that compared to HTML without style sheet, we notice: • the <link rel = "stylesheet" ...> header tag that indicates where to find the style sheet

style	

All reading activities for Database Management Systems course

Reading activity title: systeme de fichiers journalisé (in English: journaled file system)

Material in French:	Material in English: (translated by Google translate)
Système de fichiers journalisé Le système de fichiers journalisé est un système de fichiers tolérant/résistant aux pannes qui permet d'assurer l'intégrité des données en cas de problème matériel, de panne de courant (ou débranchement à chaud) ou d'arrêt brutal du système. Cette fonctionnalité est assurée par la tenue d'un journal référençant les opérations d'écriture sur le support physique avant que ce dernier ne soit réellement mis à jour. Le système de fichiers doit permettre une reprise d'activité à la suite d'une coupure brutale, telle un arrêt électrique. Les métadonnées doivent alors rester cohérentes et à jour. La journalisation permet d'optimiser le contrôle d'intégrité du système de fichiers, réduisant ainsi le temps de redémarrage du système, critère important dans les environnements qui ont besoin d'une haute disponibilité. Objectif : La journalisation du système de fichier assure la cohérence des données en utilisant un journal. Ce journal est un fichier spécial qui enregistre les changements destinés au système de fichier, dans une mémoire circulaire. À intervalles réguliers, le journal est appliqué sur le système de fichier. Si une interruption électrique intervient, le journal peut être utilisé comme point de départ afin de récupérer les informations non sauvegardées, et ainsi assurer l'intégrité des données du système de fichier.	Journaled file system The journaled file system is a fault-tolerant / fault-tolerant file system that ensures data integrity in the event of a hardware problem, power failure (or hot plug-in), or sudden shutdown of the system. This functionality is ensured by keeping a log referencing the write operations on the physical medium before it is actually updated. The file system must allow a recovery of activity following a sudden cut, such as an electrical shutdown. The metadata must then remain consistent and up to date. Logging helps optimize file system integrity control, reducing system reboot time, which is important in environments that require high availability. Purpose: Logging the file system ensures data consistency by using a log. This log is a special file that records changes to the file system in a circular memory. At regular intervals, the log is applied to the file system. If an electrical interruption occurs, the log can be used as a starting point to retrieve unsaved information, thus ensuring the integrity of the file system data. Methods : Physical Logs: The physical log records data changes on the media before they are made. This method penalizes performance because each write requires a double write on the physical medium, one for the journal, and one for the actual data. It is

Méthodes : Journaux physiques : Le journal physique enregistre les modifications de données sur le support avant que celles-ci soient opérées. Cette méthode pénalise les performances, car chaque écriture nécessite une double écriture sur le support physique, une pour le journal, et une pour les données effectives. Elle est néanmoins acceptée en raison de la garantie de la cohérence des données qu'elle permet. Journaux logiques : Le journal logique ne stocke que les métadonnées, sacrifiant la tolérance aux pannes pour de meilleures performances. Il permet lui aussi le rejeu des opérations, mais peut lier des métadonnées journalisées à des données non journalisées, causant ainsi une corruption des données. Ecritures à risque : Le cache d'écriture de la plupart des systèmes d'exploitation trie les opérations d'écriture en fonction de leur taille afin de maximiser les performances. Pour éviter un déséquilibre entre les métadonnées et les données, les écritures de données doivent être opérées avant celles des métadonnées.	nevertheless accepted because of the guarantee of data consistency that it allows. Logical Logs: The Logbook only stores metadata, sacrificing fault tolerance for better performance. It can also replay operations, but can link logged metadata to non-journaled data, causing data corruption. Risk Writing: The write cache of most operating systems sorts write operations according to their size to maximize performance. To avoid imbalance between metadata and data, data writes must be done before metadata.

Reading activity title: undo segment (in English: undo segment)

Material in French:	Material in English: (translated by Google translate)
Les segments d'annulation Chaque fois qu'une instruction « INSERT », « UPDATE » ou « DELETE » met à jour un ou plusieurs enregistrements dans une table les blocs qui contient les enregistrements sont stockés dans les segments « UNDO ». Les segments d'annulation sont des zones de stockage gérées automatiquement par Oracle. Ils sont stockés dans un tablespace de type « UNDO ». Chaque fois qu'une instruction « INSERT », « UPDATE » ou « DELETE » met à jour une ou plusieurs lignes dans la table, un verrou LMD ROW EXCLUSIVE est placé. Il permet à des transactions multiples de mettre à jour la table aussi longtemps qu'elles ne mettent pas à jour les mêmes lignes. Oracle gère le stockage, la rétention et l'emploi de l'espace pour les données d'annulation par l'intermédiaire des segments de type SMU (System-Managed Undo). Aucun objet permanent n'est placé dans le tablespace de type « UNDO ». Les undo segments sont utilisés aux fins de gérer : - La lecture cohérente des données de la base. - L'annulation d'une transaction. - La récupération des transactions après un arrêt brutal du serveur ou la perte intempestive d'une connexion. La restauration des transactions est possible car toutes les modifications apportées aux segments « UNDO » sont également protégées par des fichiers journaux. - La conservation des blocs « UNDO » après la fin des transactions pour pouvoir mettre en œuvre	The cancellation segments Whenever an "INSERT", "UPDATE", or "DELETE" statement updates one or more records in a table, the blocks that contain the records are stored in the "UNDO" segments. Undo segments are storage areas that are automatically managed by Oracle. They are stored in a tablespace of type "UNDO". Whenever an "INSERT", "UPDATE", or "DELETE" statement updates one or more rows in the table, an ROW EXCLUSIVE LMD lock is placed. It allows multiple transactions to update the table as long as they do not update the same rows. Oracle manages the storage, retention, and use of space for undo data through System-Managed Undo (SMU) segments. No permanent object is placed in the "UNDO" type tablespace. The undo segments are used to manage: - Consistent reading of data from the database. - The cancellation of a transaction. - The recovery of the transactions after a sudden stop of the server or the untimely loss of a connection. Transaction restore is possible because all changes to the "UNDO" segments are also protected by log files. - The conservation of "UNDO" blocks after the end of transactions to be able to implement "FLASHBACK" technologies. The transcendental recovery of data in the state where it was several hours ago.

les technologies « FLASHBACK ». La récupération transcendantale des données dans l'état où elles étaient plusieurs heures auparavant.	

Reading activity title: L'utilitaire SQL*Loader (in English: The SQL * Loader utility)

Material in French:	Material in English: (translated by Google translate)
L'utilitaire SQL*Loader L'utilitaire SQL*Loader (sqlload) permet de charger dans une base de données Oracle des données provenant de fichiers externes (générés par des logiciels autres qu'Oracle). SQL*Loader utilise un fichier de contrôle et des fichiers de données, charge les données, et génère un fichier pour les enregistrements erronés, un fichier pour les enregistrements rejetés et un fichier journal. Le fichier de contrôle contient la description des données à charger (noms et structures) et leur destination dans la base de données (liste des tables et colonnes où effectuer le chargement, correspondance entre les champs des fichiers de données et les colonnes). Il existe des utilitaires pour importer un fichier DBase, un fichier Lotus ou encore tout fichier ASCII. Les fichiers de données sont au format binaire ou caractères, de longueur fixe ou variable. Le fichier pour les enregistrements erronés correspond à des enregistrements n'ayant pas pu être interprétés. Le fichier pour les enregistrements rejetés correspond à des enregistrements ne satisfaisant pas à une condition spécifiée dans le fichier de contrôle (clause when). Le fichier journal est un compte-rendu de l'exécution de l'opération de chargement.	The SQL * Loader utility The SQL * Loader utility (sqlload) allows you to load data from external files (generated by software other than Oracle) into an Oracle database. SQL * Loader uses a control file and data files, loads the data, and generates a file for erroneous records, a file for rejected records, and a log file. The control file contains the description of the data to be loaded (names and structures) and their destination in the database (list of tables and columns to perform the loading, correspondence between the fields of the data files and the columns). There are utilities to import a DBase file, a Lotus file or any ASCII file. Data files are in binary format or characters, of fixed or variable length. The file for the erroneous records corresponds to records that could not be interpreted. The file for rejected records corresponds to records that do not meet a condition specified in the control file (when clause). The log file is a report of the execution of the load operation.

Reading activity title: L'utilitaire Import/Export (in English: The Import / Export utility)

Material in French:	Material in English: (translated by Google translate)
L'utilitaire Import/Export L'utilitaire Import/Export permet l'échange de données entre différentes bases Oracle. C'est un outil de migration, de défragmentation, et de transfert de données d'un environnement vers un autre. L'importation : L'utilitaire Import (imp) permet l'importation de données à partir d'une base de données Oracle ; on parle de « restauration logique » à partir d'un fichier de commandes SQL de mises à jour de données. L'exportation : L'utilitaire Export (exp) permet l'exportation de données à partir d'une base de données Oracle ; on parle de « copie logique » générant dans un fichier des commandes SQL de mises à jour de données. Les objets concernés sont : les structures des tables, les données des tables, les privilèges, les vues, les clusters, les synonymes, les séquences, les contraintes d'intégrité. Trois modes d'exportation sont possibles : mode table, mode utilisateur ou mode base de données. Le mode d'exportation base de données peut concerner tous les objets y compris ceux appartenant à l'utilisateur SYS (entire database), tous les objets exceptés ceux de SYS (full database), incrémental c.-à-d. concerner tous les objets y compris ceux de SYS et cela depuis la dernière exportation en mode incrémental ou cumulatif ou complet (incremental), cumulatif c.-à-d. concerner tous les objets y compris ceux de SYS et cela depuis la dernière exportation en mode	The Import / Export utility The Import / Export utility allows the exchange of data between different Oracle databases. It is a migration, defragmentation, and data transfer tool from one environment to another. Import: The Import utility (imp) allows the import of data from an Oracle database; we are talking about "logical restore" from a SQL file of updates of data. Export: The Export utility (exp) allows the export of data from an Oracle database; we are talking about "logical copy" generating in a file SQL commands for data updates. The objects concerned are: table structures, table data, privileges, views, clusters, synonyms, sequences, integrity constraints. Three export modes are possible: table mode, user mode or database mode. The database export mode can affect all objects including those belonging to the user SYS (entire database), all objects except those of SYS (full database), incremental ie. concern all the objects including those of SYS and that since the last incremental or cumulative or complete (incremental) export, cumulative i. concern all objects including those of SYS since the last cumulative or complete (cumulative) export, complete ie concern all objects including those of SYS and resetting the counters for incremental and cumulative modes (complete). A good strategy is to make a first full-mode export, regular incremental exports, and cumulative exports from time to time.

cumulatif ou complet (cumulative), complet c.-à-d. concerner tous les objets y compris ceux de SYS et en remettant à zéro les compteurs pour les modes incrémental et cumulatif (complete). Une bonne stratégie consiste à faire une première exportation en mode complet, régulièrement des exportations en mode incrémental, de temps en temps des exportations en mode cumulatif.	

Reading activity title: les index (in English: indexes)

Material in French:	Material in English: (translated by Google translate)
INDEX un index est une structure de données utilisée et entretenue par le système de gestion de base de données (SGBD) pour lui permettre de retrouver rapidement les données. L'utilisation d'un index simplifie et accélère les opérations de recherche, de tri, de jointure ou d'agrégation effectuées par le SGBD. L'index placé sur une table va permettre au SGBD d'accéder très rapidement aux enregistrements, selon la valeur d'un ou plusieurs champs. Un index permet simplement de récupérer plus rapidement des lignes d'une table en se basant sur des colonnes qui contiennent des valeurs très différentes les unes des autres. Prenons l'exemple du livre de recettes : tu cherches comment faire une tarte à la framboise, tu as deux solutions : soit tu traites les pages une par une pour tomber sur la bonne (long et fastidieux), soit tu te réfères directement au sommaire (index) qui t'indiquera à quelle page aller pour lire ta recette. Ça marche exactement pareil avec l'index d'une table : imagine que tu mettes un index sur la colonne « email » de ta table, eh bien lors de ta requête : SELECT nom_famille, prenom, age FROM auteurs WHERE email = 'gerard@publi.com' l'index créé sur cette colonne « email » a permis	INDEX an index is a data structure used and maintained by the database management system (DBMS) to enable it to quickly retrieve data. Using an index simplifies and speeds up the search, sort, join, or aggregate operations performed by the DBMS. The index placed on a table will allow the DBMS to access the records very quickly, depending on the value of one or more fields. An index simply allows faster retrieval of rows from a table based on columns that contain values very different from each other. Take the example of the recipe book: you are looking for how to make a raspberry pie, you have two solutions: either you treat the pages one by one to get the good one (long and tedious), or you refer directly to the summary (index) which will tell you which page to go to read your recipe. It works exactly the same with the index of a table: imagine that you put an index on the column "email" of your table, well at your request: SELECT family_name, firstname, age FROM authors WHERE email = 'gerard@publi.com' the index created on this "email" column allowed faster access to the line containing the information sought because it was "registered" in advance. It is also important to note that an index is

d'accéder plus rapidement à la ligne contenant les informations recherchées car elle était « enregistrée » à l'avance. Il est d'ailleurs important de noter qu'un index est automatiquement créé lorsque tu appliques une contrainte de clé primaire ou d'unicité à ta colonne (et c'est d'ailleurs assez logique, car toutes les valeurs seront différentes les unes des autres !).	automatically created when you apply a primary key or uniqueness constraint to your column (and this is quite logical, because all the values will be different from each other). others!).

Reading activity title: index: types (in English: index: types)

Material in French:	Material in English: (translated by Google translate)
Les index sont en particulier exploités par l´optimiseur de requêtes: L'optimiseur est le composant des SGBD qui recherche la manière la plus économique d'exécuter une requête. L'optimiseur examine les différents scénarios possibles et estime le nombre d'opérations nécessaires pour chaque scénario, puis opte pour le scénario qui en demande le moins. Le nombre d'opérations nécessaires dépend de la présence d'index, ainsi que du nombre de lignes de la table et de la répartition des valeurs. Type d'index : • La structure la plus courante pour les index est l'arbre B (B-tree). En stockant les différentes valeurs du champ dans un arbre équilibré, le SGBD pourra hiérarchiser les enregistrements d'après un champ dont la plage de valeurs est infinie (ou presque). • Un autre type d'index est l'index bitmap. Il consiste en une simple table indiquant, pour chaque valeur possible du champ, la liste des enregistrements ayant cette valeur pour ce champ. Cependant, pour être efficace, il nécessite que le SGBD puisse accéder directement à une valeur donnée. Il n'est donc applicable que sur les colonnes pour lesquelles le nombre de valeurs est limité et ordonné. • On trouve également des index par table de hachage. L'inconvénient majeur d'un tel index est de ne permettre que les sélections par égalité, puisqu'il ne conserve pas la notion d'ordre. Si n est le nombre d'enregistrements d'une table, l'utilisation d'une	In particular, indexes are exploited by the query optimizer: The optimizer is the component of DBMS that looks for the most economical way to execute a query. The optimizer examines the different possible scenarios and estimates the number of operations required for each scenario, then chooses the scenario that requires the least. The number of operations required depends on the presence of indexes, as well as the number of rows in the table and the distribution of values. Type of index: • The most common structure for indexes is tree B (B-tree). By storing the different values of the field in a balanced tree, the DBMS will be able to prioritize the records according to a field whose range of values is infinite (or almost). • Another type of index is the bitmap index. It consists of a simple table indicating, for each possible value of the field, the list of records having this value for this field. However, to be effective, it requires that the DBMS can directly access a given value. It is therefore only applicable on columns for which the number of values is limited and ordered. • There are also indexes by hash table. The major disadvantage of such an index is to allow only the selections by equality, since it does not preserve the notion of order. If n is the number of records in a table, using a balanced hash table can reduce the number of records to go to √n, the square root of n (where the table is composed of √n hash values

table de hachage équilibrée peut permettre de réduire le nombre d'enregistrements à parcourir à √n, la racine carrée de n (la table étant alors composée de √n valeurs de hachage accédant chacune à √n enregistrements). La même remarque sur l'efficacité existe pour l'index bitmap : le SGBD doit pouvoir accéder directement à une valeur de hachage donnée, sans avoir à parcourir la liste des valeurs de hachage possibles.	each accessing √n records). The same remark about efficiency exists for the bitmap index: the DBMS must be able to directly access a given hash value, without having to go through the list of possible hash values.

Reading activity title: plan d execution (in English: plan d execution)

Material in French:	Material in English: (translated by Google translate)
Il est important d'exécuter les requêtes SQL le plus efficacement possible. Pour cela, la base doit utiliser les informations et l'infrastructure (index par exemple) à sa disposition, réaliser des filtres, faire des correspondances de données correctement selon le nombre de lignes etc… C'est la tâche de l'optimiseur. Définition d'un plan d'exécution Le plan d'exécution est un moyen de présenter les différentes opérations qu'Oracle va réaliser pour répondre à votre requête. Lorsque vous demandez le contenu d'une table, Oracle va devoir lire les blocs de cette table. Mais dans certains cas, un index peut accélérer la lecture ; s'il l'utilise Oracle va donc lire l'index puis les blocs de la table référencés par l'index. ---------- ---------- | Id| Operation | Name | Rows | Bytes | Cost (%CPU)| Time | ---------- ---------- | 0 | SELECT STATEMENT | | 4 | 16208 | 3 (0)| 00:00:01 | | 1 | TABLE ACCESS BY INDEX ROWID| CONSULTANT | 4 | 16208 | 3 (0)| 00:00:01 | |*2 | INDEX RANGE SCAN | CONSULTANT_PK | 4 | | 2 (0)| 00:00:01 | ---------- ---------- Predicate Information (identified by operation id):	It is important to execute SQL queries as efficiently as possible. For this, the database must use the information and infrastructure (index for example) at its disposal, make filters, make data matches correctly according to the number of lines etc ... This is the task of the optimizer. Definition of an execution plan The execution plan is a way to present the different operations that Oracle will perform to answer your request. When you request the contents of a table, Oracle will have to read the blocks of this table. But in some cases, an index can speed up reading; if it uses it Oracle will read the index then the blocks of the table referenced by the index. ---------- ---------- | id | Operation | Name | Rows | Bytes | Cost (% CPU) | Time | ---------- ---------- | 0 | SELECT STATEMENT | | 4 | 16208 | 3 (0) | 00:00:01 | | 1 | TABLE ACCESS BY INDEX ROWID | CONSULTANT | 4 | 16208 | 3 (0) | 00:00:01 | | * 2 | INDEX RANGE SCAN | CONSULTANT_PK | 4 | | 2 (0) | 00:00:01 | ---------- ---------- Predicate Information (identified by operation id): ---------- - 2 - access ("ID_CONSULTANT" <5) What to read?

----------	Always read the plan from bottom to top. Oracle will browse the CONSULTANT_PK index in step 2; this step is prefixed with a * which is a reference to "Predicate Information" which indicates that in step 2 the lines are filtered (normal, the index is on the ID).
2 - access("ID_CONSULTANT"<5)	Then Oracle will make an access to the table for each ROWID found in the index: Step 1.
Qu'en lire ?	Note: The ROWID is the identifier of a row of a table on an entire database. It is nothing less than the pointer from the index to the line.
Toujours lire le plan de bas en haut. Oracle va parcourir l'index CONSULTANT_PK à l'étape 2 ; cette étape est préfixée d'un * qui est un renvoi à "Predicate Information" qui indique qu'à l'étape 2 on filtre les lignes (normal, l'index porte sur l'ID).	With the execution plan, other information is given as the cost of a step or the time spent on it.
Puis Oracle va faire un accès à la table pour chaque ROWID trouvé dans l'index : étape 1.	
Note : Le ROWID est l'identifiant d'une ligne d'une table sur toute une base de donnée. C'est rien de moins que le pointeur de l'index vers la ligne.	
Avec le plan d'exécution, d'autres informations sont données comme le coût d'une étape ou bien le temps passé sur celle-ci.	

Reading activity title: Optimisation (in English: Optimisation)

Material in French:	Material in English: (translated by Google translate)
L'optimisation des bases de données est souvent perçue comme étant une tâche du DBA, cependant sur de gros volumes de données et un grand nombre de bases, le DBA ne peut pas connaître la logique métier de toutes ses bases ni passer réécrire les requêtes de chacun des développeurs. La conception d'applications performantes passera nécéssairement par la connaissance des différents éléments permettant d'avoir des bases répondant aux besoins et évolutives. Il est fréquent qu'une requête mal écrite coûte 10 fois plus de temps et de ressources qu'elle ne le devrait. Lorsque l'application n'est pas seule sur le serveur de base de données, les ralentissements ou la surconsommation de ressource peuvent s'avérer gênants non seulement pourl'application mais aussi les autres projets/ applications partageant les ressources. Plan d'execution : Il est souvent nécéssaire sur des requêtes d'interroger une ou plusieurs tables qui feront donc l'objet d'un ou plusieurs chemins d'accès. L'ensemble des chemins d'accès utilisés pour une requête a pour nom "Plan d'exécution". Trouver le meilleur chemin ? On a un GPS ? (l'optimiseur) Trouver le chemin le plus intéressant pour accéder aux données, c'est le boulot de l'optimiseur Oracle. Evidemment le terme "chemin le plus intéressant" est assez subjectif (et règlable par le DBA dans le cas d'Oracle) mais on demande le plus souvent à Oracle d'utiliser le chemin qu'il estime avoir le	Database optimization is often perceived as a DBA task, however, on large volumes of data and a large number of databases, the DBA can not know the business logic of all its databases or rewrite queries. each of the developers. The design of high-performance applications will necessarily require knowledge of the different elements to have bases that meet the needs and evolve. It is common for a poorly written request to cost 10 times more time and resources than it should. When the application is not alone on the database server, slowdowns or over-consumption of resources can be annoying not only for the application but also for other projects / applications sharing the resources. Execution plan : It is often necessary on queries to query one or more tables which will thus be the object of one or more paths. The set of access paths used for a query is named "Execution Plan". Find the best way? We have a GPS? (The optimizer) Finding the most interesting way to access data is the job of the Oracle Optimizer. Obviously the term "most interesting way" is quite subjective (and adjustable by the DBA in the case of Oracle) but it is most often asked to Oracle to use the path he believes has the best "cost" (cost -based). The cost takes into account the use of resources (CPU, disk access, memory ...) for a given plan. Reading on a disc is slow. Even worse, having to

meilleur "coût" (cost-based). Le coût tient compte de l'utilisation des ressources (CPU, accès disques, mémoire ...) pour un plan donné. Lire sur un disque c'est lent. Pire encore, devoir effectuer un tri de données sur le disque s'avère souvent extrêmement coûteux...	sort through the disk is often extremely expensive ...

Reading activity title: privileges (in English: privileges)

Material in French:	Material in English: (translated by Google translate)
Rôles et privilèges sont définis pour sécuriser l'accès aux données de la base Ces concepts sont mis en oeuvre pour protéger les données en accordant (ou retirant) des privilèges a un utilisateur ou un groupe d'utilisateurs. Un rôle est un regroupement de privilèges. Une fois créé il peut être assigné à un utilisateur ou à un autre rôle. Les privilèges sont de deux types • Les privilèges de niveau système Qui permettent la création, modification, suppression, exécution de groupes d'objets les privilèges CREATE TABLE, CREATE VIEW, CREATE SEQUENCE par exemple permettent à l'utilisateur qui les a reçu de créer des tables, des vues et des séquences. • Les privilèges de niveau objet Qui permettent les manipulations sur des objets spécifiques les privilèges SELECT, INSERT, UPDATE, DELETE sur la table SCOTT.EMP par exemple permettent à l'utilisateur qui les a reçu de sélectionner, ajouter, modifier et supprimer des lignes dans la table EMP appartenant à l'utilisateur SCOTT. Assigner des privilèges système à un utilisateur Lorsqu'un utilisateur est créé avec l'instruction CREATE USER, il ne dispose encore d'aucun droit car aucun privilège ne lui a encore été assigné Il ne peut même pas se connecter à la base !	Roles and privileges are defined to secure access to database data These concepts are implemented to protect data by granting (or removing) privileges to a user or group of users. A role is a collection of privileges. Once created it can be assigned to a user or another role. Privileges are of two types • System level privileges For example, CREATE TABLE, CREATE VIEW, and CREATE SEQUENCE privileges allow the user who received them to create tables, views, and sequences for creating, modifying, deleting, and executing object groups. • Object level privileges Which allow the manipulation on specific objects the privileges SELECT, INSERT, UPDATE, DELETE on the table SCOTT.EMP for example allow the user who received them to select, add, modify and delete lines in the table EMP belonging to the SCOTT user. Assign system privileges to a user When a user is created with the CREATE USER statement, they still have no rights because no privileges have been assigned to them yet He can not even connect to the base! It must therefore be assigned the necessary privileges. It must be able to connect, create tables, views,

Il faut donc lui assigner les privilèges nécessaires. Il doit pouvoir se connecter, créer des tables, des vues, des séquences. Pour lui assigner ces privilèges de niveau système il faut utiliser l'instruction GRANT dont voici la syntaxe.	sequences. To assign it these privileges of system level it is necessary to use the instruction GRANT which here is the syntax.

Reading activity title: with grant option (in English: with grant option)

Material in French:	Material in English: (translated by Google translate)
With grant option : Nous pouvons accorder des privilèges uniquement avec l'option grant (select, update, insert …) A ---------> B ---------> C Si A veut supprimer des privilèges de C, ce n'est pas possible. Uniquement l'utilisateur qui a accordé le privilège peut le supprimer. nous devons révoquer (supprimer) le privilège de B, qui révoque automatiquement les privilèges de C. Accorder des privilèges à quelqu'un d'autre : Si un privilège d'objet est révoqué d'un utilisateur auquel il a été accordé WITH GRANT OPTION, ce privilège serait également supprimé de toute personne à qui ce privilège est accordé. Par exemple, si Damir accordait à John le privilège SELECT sur la table DAMIR.JOHN1 WITH GRANT OPTION, et que John accordait le privilège SELECT à Tim, alors si Damir émettait la commande REVOKE SELECT sur DAMIR.JOHN1 de JOHN, Tim ne possède plus ce privilège. Cela est dû au fait que lorsque les privilèges d'objet sont révoqués, la révocation est également répercutée sur toute personne à laquelle le privilège a été accordé par l'utilisateur auprès duquel elle est révoquée.	With grant option: We can grant privileges only with the option grant (select, update, insert ...) A ---------> B ---------> C If A wants to remove privileges from C, this is not possible. Only the user who granted the privilege can delete it. we must revoke (remove) the privilege of B, which automatically revokes the privileges of C. Grant privileges to someone else: If an object privilege is revoked from a user granted WITH GRANT OPTION, that privilege would also be removed from anyone to whom that privilege is granted. For example, if Damir gave John the SELECT privilege on the DAMIR.JOHN1 WITH GRANT OPTION table, and John granted the SELECT privilege to Tim, then if Damir issued the REVOKE SELECT command on DAMIR.JOHN1 of JOHN, Tim no longer has this privilege. This is because when object privileges are revoked, the revocation is also passed on to any person to whom the privilege has been granted by the user from whom it is revoked.

Appendix B: Questionnaires

Student information

1. Study code: (Student number)

2. Gender: □ Male □ Female

3. What is your age? Or put your birth year.

4. You are now studying for □ Bachelor Degree □ Master Degree □ Doctoral Degree

5. You study in

Technology	□ Agronomy □ Computer science □ Engineering □ Medicine
Sciences	□ Biology □ Chemistry □ Earth and space sciences □ Mathematics □ Physics
Arts	□ Performing arts □ Visual arts
Humanities	□ Geography □ History □ Languages and literature □ Philosophy
Social sciences	□ Economics □ Law □ Political science □ Psychology □ Sociology

6. When you use the Online Annotation System, what is the category of the course you take?

□ Technology □ Sciences □ Arts □ Humanities □ Social sciences

7. When did you know e-reader for the first time?

When I was years old. □ I have never heard it before

8. When did you first use any kind of e-readers (i.e., software or device that you use it to read articles or books)?

When I was years old. □ I have never heard it before

9. When is your first time to use Internet?

When I was years old. □ I have never heard it before

10. When was your first time to use browser? When I was years old.

11. Usually what do you do with browser? (choose more than one)

□ playing games □ posting article □ reading book □ reading news
□ study □ Using social network application (e.g., Facebook, twitter, etc.)
□ watching movie □ other (need to have a textbox for them to enter something)

12. Usually in a week day, how many hours you use Internet with browsers? About hours per day.

The experience of using e-reader

(5 for strong agree, 4 for agree, 3 for normal, 2 for disagree, and 1 for strong disagree)

Item#	Questions	Agree 5	4	3	2	Disagree 1
1	Using an e-reader application with annotation functions enables me to understand the key concepts of the reading activities more quickly.	□	□	□	□	□
2	Using an e-reader application with annotation functions improves the quality of annotations I make.	□	□	□	□	□
3	Using an e-reader application with annotation functions makes easier to do reading activities.	□	□	□	□	□
4	Using an e-reader application with annotation functions improves my learning performance.	□	□	□	□	□
5	Using an e-reader application with annotation functions gives me greater control over my study schedule.	□	□	□	□	□
6	Using an e-reader application with annotation functions is compatible with all aspects of my study in school.	□	□	□	□	□
7	Using an e-reader application with annotation functions is completely compatible with my current study in the class.	□	□	□	□	□
8	I think that using an e-reader application with annotation functions fits well with the way I like to	□	□	□	□	□

	study.	
9	Using an e-reader application with annotation functions fits well with the device I prefer to use.	□ □ □ □ □
10	My interaction with e-reader application with annotation functions is clear.	□ □ □ □ □
11	My interaction with e-reader application with annotation functions is understandable.	□ □ □ □ □
12	Learning to use an e-reader application with annotation functions is easy for me.	□ □ □ □ □
13	Overall, I believe that an e-reader application with annotation functions is easy to adopt into my study.	□ □ □ □ □
14	I've had a great deal of opportunities to try an e-reader application with annotation functions for studying.	□ □ □ □ □
15	I know where I can go to satisfactorily try out various uses of an e-reader application with annotation functions for studying.	□ □ □ □ □
16	Before deciding whether or not to adopt an e-reader application with annotation functions, I would need to use it on a trail basis.	□ □ □ □ □
17	Before deciding whether or not to adopt an e-reader application with annotation functions, I would need to properly try it out	□ □ □ □ □
18	I would like to be permitted to use an e-reader application with annotation functions on a trial basis long enough to see what it can do.	□ □ □ □ □
19	It is easy for me to see people using e-reader application with annotation functions in the school.	□ □ □ □ □
20	I have had a lot of opportunities to see people using e-reader pplication with annotation functions to study.	□ □ □ □ □
21	It is easy for me to see others' annotations when we all use e-reader application with annotation functions.	□ □ □ □ □
22	I can see how others annotate the content of an article or book when we use e-reader application with	□ □ □ □ □

	annotation functions.	
23	I see people searching and finding the desired content quickly in an e-reader application.	□ □ □ □ □
24	I can tell how different that I annotate an article or book from others when we use e-reader application with annotation functions.	□ □ □ □ □

The experience of using Online Annotation System

(5 for strong agree, 4 for agree, 3 for normal, 2 for disagree, and 1 for strong disagree)

Item#	Questions	Agree 5	4	3	2	Disagree 1
1	I believe it is a good idea to use an Online Annotation System.	□	□	□	□	□
2	Once I started using the Online Annotation System I found it is hard to stop.	□	□	□	□	□
3	I like to use the Online Annotation System.	□	□	□	□	□
4	As a student I like to use Online Annotation System to study.	□	□	□	□	□
5	The Online Annotation System is pleasant to use.	□	□	□	□	□
6	I could imagine that most people could learn how to use the Online Annotation System very quickly.	□	□	□	□	□
7	I needed to learn a lot of things before I could get going with the Online Annotation System.	□	□	□	□	□
8	Learning to use the Online Annotation System is easy for me.	□	□	□	□	□
9	It is easy for me to remember how to do the reading activities in the Online Annotation System.	□	□	□	□	□
10	I find it takes a lot of efforts to become skillful at using the Online Annotation System.	□	□	□	□	□
11	I quickly became skillful with the Online Annotation System.	□	□	□	□	□
12	I think the Online Annotation System is easy to use.	□	□	□	□	□
13	I think that I would need the support of a technical person to be able to use the Online Annotation	□	□	□	□	□

	System.	
14	I find the various functions in the Online Annotation System were well integrated.	□ □ □ □ □
15	I think there is too much inconsistency in the Online Annotation System.	□ □ □ □ □
16	I find the Online Annotation System very cumbersome to use.	□ □ □ □ □
17	The interface of the Online Annotation System is pleasant.	□ □ □ □ □

18	The user interface of the Online Annotation System is confusing.	□ □ □ □ □
19	The Online Annotation System requires minimal steps for doing my reading activity.	□ □ □ □ □
20	The logical design of this Online Annotation System is good, I have no difficulty in using it.	□ □ □ □ □
21	The Online Annotation System is user friendly.	□ □ □ □ □
22	Whenever I make a mistake while using the Online Annotation System I recover easily and quickly.	□ □ □ □ □
23	The information (such as courese list, reading activity list, activity starting date, and activity ending date) provided by the Online Annotation System is clear.	□ □ □ □ □
24	It is easy to find the information I needed.	□ □ □ □ □
25	The information provided by the Online Annotation System is easy to understand.	□ □ □ □ □
26	I find the Online Annotation System unnecessarily complex.	□ □ □ □ □
27	I can use the Online Annotation System without written instructions.	□ □ □ □ □
28	I believe I understand the reading materials more in-depth by using the Online Annotation System.	□ □ □ □ □
29	Using the Online Annotation System gives me greater control over my time to finish my reading activities.	□ □ □ □ □
30	The Online Annotation System enables me to accomplish the reading activity more quickly.	□ □ □ □ □
31	Using the Online Annotation System improves my learning performance.	□ □ □ □ □
32	I think that I would like to use the Online Annotation System frequently.	□ □ □ □ □
33	I feel very confident using the Online Annotation System.	□ □ □ □ □
34	I plan to use an Online Annotation System in the future.	□ □ □ □ □
35	Assuming that I have access to an Online Annotation	□ □ □ □ □

	System, I intend to use it.	
36	I intend to continue to use the Online Annotation System in the future.	□ □ □ □ □
37	I will recommend others to use the Online Annotation System.	□ □ □ □ □
38	This Online Annotation System has all the functions and capabilities I expect it to have.	□ □ □ □ □
39	I expect that I would use the Online Annotation System in the future	□ □ □ □ □
40	The Online Annotation System meets my needs.	□ □ □ □ □
41	The Online Annotation System works the way I want it to work.	□ □ □ □ □

Appendix C: Word-based bit-string clustering precision and recall analysis

Have all the students: word-based bit-string chromosomes on database class

		131	132	137	138	Average
Standard	Precision	59.38%	65.00%	76.56%	40.10%	60.26%
	Recall	53.03%	43.91%	50.00%	36.25%	45.80%
	F2	54.19%	46.96%	53.73%	36.96%	47.96%
	F0.5	57.99%	59.30%	69.21%	39.27%	56.44%
	Average Time	0.228	0.304	0.236	0.168	
Quantitative	Precision	57.14%	57.14%	50.74%	27.50%	48.13%
	Recall	39.58%	39.58%	35.42%	29.17%	35.94%
	F2	42.18%	42.18%	37.69%	28.82%	37.72%
	F0.5	52.49%	52.49%	46.70%	27.82%	44.87%
	Average Time	0.212	0.34	0.224	0.168	
Cosine	Precision	15.63%	41.52%	52.08%	38.39%	36.90%
	Recall	37.50%	45.03%	44.44%	37.08%	41.01%
	F2	29.30%	44.28%	45.79%	37.34%	39.18%
	F0.5	17.69%	42.18%	50.35%	38.12%	37.09%
	Average Time	15.18	5.704	5.172	3.576	
Diffusion	Precision	70.37%	52.27%	42.31%	40.79%	51.44%
	Recall	47.92%	35.42%	47.92%	34.17%	41.35%
	F2	51.18%	37.86%	46.68%	35.31%	42.76%
	F0.5	64.34%	47.73%	43.32%	39.27%	48.66%
	Average Time	0.52	0.94	0.8	0.52	

Have all the students: word-based bit-string chromosomes on HTML class

		133	134	135	136	Average
Standard	Precision	64.06%	50.00%	42.36%	38.64%	48.76%
	Recall	40.00%	46.88%	42.50%	33.33%	40.68%
	F2	43.25%	47.47%	42.47%	34.27%	41.87%
	F0.5	57.18%	49.34%	42.39%	37.44%	46.59%
	Average Time	0.176	0.152	0.164	0.124	
Quantitative	Precision	38.24%	62.50%	23.86%	38.64%	40.81%
	Recall	30.00%	62.50%	40.63%	33.33%	41.61%
	F2	31.35%	62.50%	35.62%	34.27%	40.94%
	F0.5	36.25%	62.50%	26.01%	37.44%	40.55%
	Average Time	0.156	0.12	0.12	0.248	
Cosine	Precision	41.96%	49.11%	41.67%	22.92%	38.91%
	Recall	41.67%	62.50%	44.38%	25.00%	43.39%
	F2	41.73%	59.27%	43.81%	24.55%	42.34%
	F0.5	41.90%	51.31%	42.18%	23.31%	39.67%
	Average Time	3.08	5.508	2.96	4.864	
Diffusion	Precision	40.00%	27.65%	43.75%	41.67%	38.27%
	Recall	40.00%	46.88%	48.13%	41.67%	44.17%
	F2	40.00%	41.15%	47.18%	41.67%	42.50%
	F0.5	40.00%	30.12%	44.56%	41.67%	39.09%
	Average Time	0.26	0.4	0.492	0.508	

Remove the no annotations' chromosome: word-based bit-string chromosomes on database class

		131	132	137	138	Average
Standard	Precision	43.33%	34.38%	50.00%	26.39%	38.53%
	Recall	37.50%	33.33%	50.00%	25.83%	36.67%
	F2	38.54%	33.53%	50.00%	25.94%	37.00%
	F0.5	42.02%	34.16%	50.00%	26.28%	38.12%
	Average Time	0.284	0.368	0.196	0.216	
Quantitative	Precision	30.83%	40.63%	36.36%	40.00%	36.96%
	Recall	29.17%	39.58%	27.08%	30.00%	31.46%
	F2	29.49%	39.79%	28.54%	31.58%	32.35%
	F0.5	30.48%	40.42%	34.03%	37.50%	35.61%
	Average Time	0.256	0.316	0.224	0.152	
Cosine	Precision	39.17%	40.28%	51.67%	33.04%	41.04%
	Recall	39.58%	35.42%	52.08%	32.50%	39.90%
	F2	39.50%	36.30%	52.00%	32.61%	40.10%
	F0.5	39.25%	39.20%	51.75%	32.93%	40.78%
	Average Time	4.596	5.64	5.168	3.464	
Diffusion	Precision	48.21%	46.59%	22.50%	36.61%	38.48%
	Recall	39.58%	35.42%	22.92%	35.83%	33.44%
	F2	41.05%	37.20%	22.83%	35.98%	34.27%
	F0.5	46.20%	43.83%	22.58%	36.45%	37.26%
	Average Time	0.456	0.896	0.86	0.528	

Remove the no annotations' chromosome: word-based bit-string chromosomes on HTML class

		133	134	135	136	Average
Standard	Precision	40.63%	43.75%	26.25%	41.67%	38.08%
	Recall	35.00%	31.25%	37.50%	37.50%	35.31%
	F2	36.00%	33.14%	34.54%	38.27%	35.49%
	F0.5	39.36%	40.51%	27.93%	40.76%	37.14%
	Average Time	0.128	0.196	0.204	0.164	
Quantitative	Precision	38.89%	72.22%	39.58%	38.64%	47.33%
	Recall	30.00%	62.50%	59.38%	29.17%	45.26%
	F2	31.44%	64.23%	53.98%	30.67%	45.08%
	F0.5	36.71%	70.04%	42.41%	36.28%	46.36%
	Average Time	0.148	0.152	0.168	0.188	
Cosine	Precision	25.00%	41.67%	37.50%	34.38%	34.64%
	Recall	25.00%	43.75%	56.25%	33.33%	39.58%
	F2	25.00%	43.32%	51.14%	33.53%	38.25%
	F0.5	25.00%	42.07%	40.18%	34.16%	35.35%
	Average Time	3.004	5.508	3.028	4.924	
Diffusion	Precision	40.63%	33.33%	26.25%	38.64%	34.71%
	Recall	35.00%	28.13%	37.50%	29.17%	32.45%
	F2	36.00%	29.04%	34.54%	30.67%	32.56%
	F0.5	39.36%	32.14%	27.93%	36.28%	33.93%
	Average Time	0.244	0.468	0.48	0.492	

Appendix D: The manual mapping data for kappa analysis

The word-based cluster dataset for kappa analysis
T for teacher edited clustering result
S for Standard method clustering result
Q for Quantitative method clustering result
C for Cosine method clustering result
D for Diffusion method clustering result
G for GRACE method clustering result

This appendix shows the manual mapping results. Teacher edited clustering result is the benchmark. The different numbers in the teacher edited clustering represent the different groups. Each row represents a student, this student in teacher edited clustering is in which group; in Standard method clustering is in another group. For example, see the reading activity 131, the student in the first row is in the group 1 when he/she was clustered by teacher; in the Standard methods, he/she is in the group 4; in the Quantitative method, he/she is in the group 4; in the Cosine method, he/she is in the group 2; in the Diffusion method, he/she is in the group 4; in the GRACE method, he/she is in the group 4.
The first student in the reading activity 132 doesn't mean he/she is the same student as the first student in 131. A row in a reading activity only represent a student belongs to which group in the correspondence clustering method.

word-based clustering

131						132						133						134					
T	S	Q	C	D	G	T	S	Q	C	D	G	T	S	Q	C	D	G	T	S	Q	C	D	G
1	4	4	2	4	4	1	1	4	3	4	4	1	3	3	3	3	3	1	3	1	1	1	4
1	2	4	2	1	4	1	4	4	3	4	4	1	1	3	3	3	3	2	2	3	2	1	4
1	1	4	1	4	4	1	4	4	3	4	4	1	3	3	2	3	3	2	3	2	1	2	4
2	4	4	2	2	4	1	4	4	1	4	4	1	3	3	1	3	3	3	3	4	3	4	4
2	2	2	2	1	4	2	3	4	2	3	4	1	1	3	1	3	3	3	3	4	3	4	3
2	4	4	3	3	4	2	2	4	3	4	4	2	2	2	1	1	3	3	3	4	3	4	1
3	2	4	3	4	4	2	4	3	4	2	4	2	3	1	2	2	3	3	3	4	2	4	2
3	3	2	3	3	4	2	4	2	2	4	4	2	3	3	2	2	3	3	3	4	3	4	1
3	1	1	1	4	4	3	4	4	3	4	4	2	3	3	2	2	3	3	3	4	1	4	3
3	3	3	4	4	4	3	4	1	3	4	4	2	3	3	1	3	3	3	3	4	3	4	4
4	4	4	2	4	4	3	3	3	3	3	4	3	3	3	2	3	3	3	3	4	4	4	2
4	4	4	4	4	4	3	3	3	3	1	4	3	3	3	3	3	2	4	3	4	2	3	4
4	4	4	3	4	2	3	2	4	4	4	4	3	3	3	3	3	3	4	3	4	2	3	4
4	4	4	2	4	1	3	4	4	4	4	4	3	3	3	2	3	3	4	4	4	4	4	4
4	4	4	3	4	3	4	4	4	2	4	4	3	3	3	3	3	1	4	2	4	4	4	4
4	4	4	3	4	2	4	3	4	2	4	4	3	3	3	3	3	1	4	1	4	3	3	4
4	4	4	2	4	4	4	4	4	2	4	1	3	3	3	2	3	3	4	4	4	4	2	4
4	4	4	1	4	4	4	4	4	2	4	4	3	3	3	3	3	3	4	4	4	1	2	4
4	4	4	1	4	4	4	4	4	4	4	3	3	3	3	3	3	2	4	2	4	3	4	4
4	4	4	2	4	1	4	4	4	2	4	1												
4	4	4	4	4	4	4	4	4	1	4	4												
4	4	4	3	4	3	4	4	4	4	4	2												
4	4	4	3	4	4	4	4	4	1	4	4												
4	4	4	1	4	4	4	4	4	4	4	2												
4	3	4	4	4	4	4	4	4	4	4	3												
4	4	4	1	4	4	4	4	4	4	4	4												
4	4	4	1	4	4	4	4	4	1	4	4												
4	4	4	1	4	4																		
4	4	4	4	4	4																		
4	4	4	4	4	4																		
4	4	4	4	4	4																		
4	4	4	4	4	4																		

135						136						137						138					
T	S	Q	C	D	G	T	S	Q	C	D	G	T	S	Q	C	D	G	T	S	Q	C	D	G
1	4	4	1	3	4	1	3	3	1	3	1	1	4	4	2	4	3	1	3	3	1	3	1
2	2	2	4	2	2	1	3	3	1	3	1	1	1	4	1	4	3	1	3	2	2	2	1
3	4	4	1	3	3	2	2	1	2	1	1	1	4	4	3	4	3	1	2	1	3	1	1
3	4	4	1	3	3	2	1	3	3	2	1	2	2	4	3	2	3	1	3	3	1	1	1
3	4	4	1	3	4	2	3	2	1	2	2	2	4	4	1	2	3	1	1	3	1	3	1
3	3	4	3	3	4	2	3	2	3	2	2	2	4	1	2	2	3	2	3	3	2	3	1
3	4	4	3	3	4	2	3	3	1	2	1	2	2	2	2	3	3	2	2	3	1	3	1
4	1	2	2	1	4	2	2	3	1	3	1	3	4	4	3	4	3	2	1	3	1	3	1
4	2	4	2	2	1	3	3	3	3	3	3	3	3	4	2	2	3	2	3	3	2	3	1
4	4	3	2	2	1	3	3	3	3	3	1	3	2	3	1	1	3	2	1	2	1	2	1
4	4	4	2	4	4	3	3	3	2	3	1	3	4	1	3	2	3	2	2	3	2	3	1
4	2	4	4	3	4	3	3	3	2	3	1	3	4	2	2	3	3	3	3	3	1	3	1
4	2	1	3	2	2	3	3	3	1	3	1	3	2	4	3	2	3	3	3	3	3	3	2
4	4	4	4	3	4	3	3	3	2	3	3	4	4	4	3	4	3	3	3	3	3	3	1
4	4	4	4	3	4							4	4	4	4	4	1	3	3	3	3	3	1
												4	4	4	4	4	3	3	3	3	2	3	1
												4	4	4	2	4	3	3	3	3	3	3	1
												4	4	4	4	4	2	3	3	3	2	3	3
												4	4	4	2	4	4	3	1	3	2	3	1
												4	4	4	2	4	4	3	3	3	2	3	3
												4	4	4	1	4	1	3	3	3	3	3	2
												4	4	4	4	4	2	3	3	3	3	3	1
																		3	3	3	3	3	1

alphabet-based clustering

131						132						133						134					
T	S	Q	C	D	G	T	S	Q	C	D	G	T	S	Q	C	D	G	T	S	Q	C	D	G
1	1	4	1	4	4	1	3	1	3	4	4	1	3	1	3	3	3	1	3	3	1	1	4
1	1	4	1	4	4	1	1	4	3	4	4	1	1	1	3	3	3	2	4	2	2	2	4
1	2	1	3	4	4	1	3	2	3	4	4	1	3	3	2	3	3	2	3	3	1	4	4
2	1	2	1	1	4	1	4	4	1	4	4	1	3	3	1	3	3	3	3	3	3	3	4
2	2	4	1	2	4	2	3	1	2	3	4	1	2	3	1	3	3	3	3	3	3	3	3
2	4	4	2	2	4	2	4	4	3	4	4	2	3	3	1	2	3	3	3	3	3	3	1
3	2	4	2	3	4	2	3	1	4	2	4	2	3	3	2	1	3	3	3	3	2	3	2
3	1	3	2	2	4	2	2	1	2	4	4	2	2	2	2	3	3	3	3	3	3	3	1
3	3	3	3	3	4	3	3	1	3	4	4	2	3	3	2	3	3	3	3	3	1	3	3
3	4	4	4	3	4	3	3	2	3	4	4	2	3	3	1	3	3	3	3	3	3	3	4
4	4	4	1	4	4	3	3	2	3	3	4	3	3	3	2	3	3	3	3	3	4	3	2
4	4	4	4	4	4	3	3	1	3	1	4	3	3	3	3	3	2	4	4	3	2	4	4
4	4	4	2	4	2	3	4	4	4	4	4	3	3	3	3	3	3	4	2	1	2	4	4
4	4	4	1	4	1	3	3	2	4	4	4	3	3	3	2	3	3	4	3	3	4	3	4
4	4	4	2	4	3	4	4	4	2	4	4	3	3	3	3	3	1	4	4	3	4	3	4
4	4	4	2	4	2	4	3	4	2	4	4	3	3	3	3	3	1	4	1	3	3	4	4
4	4	4	1	4	4	4	4	4	2	4	1	3	3	3	2	3	3	4	3	3	4	4	4
4	4	4	3	4	4	4	4	4	2	4	4	3	3	3	3	3	3	4	4	4	1	4	4
4	4	4	3	4	4	4	4	4	4	4	3	3	3	3	3	3	2	4	2	3	3	3	4
4	4	4	1	4	1	4	4	4	2	4	1												
4	4	4	4	4	4	4	4	4	1	4	4												
4	4	4	2	4	3	4	4	4	4	4	2												
4	4	4	2	4	4	4	4	4	1	4	4												
4	4	4	3	4	4	4	4	4	4	4	2												
4	1	4	4	4	4	4	4	4	4	4	3												
4	4	4	3	4	4	4	4	4	4	4	4												
4	4	4	3	4	4	4	4	4	1	4	4												
4	4	4	3	4	4																		
4	4	4	4	4	4																		
4	4	4	4	4	4																		
4	4	4	4	4	4																		
4	4	4	4	4	4																		

135						136						137						138					
T	S	Q	C	D	G	T	S	Q	C	D	G	T	S	Q	C	D	G	T	S	Q	C	D	G
1	4	3	3	3	4	1	1	3	1	3	1	1	2	3	2	4	3	1	1	1	2	3	1
2	3	3	2	4	2	1	1	3	1	3	1	1	2	4	1	4	3	1	2	1	1	1	1
3	3	3	3	3	3	2	2	2	3	1	1	1	4	4	3	4	3	1	2	1	3	1	1
3	3	3	3	3	3	2	3	3	2	3	1	2	2	3	3	3	3	1	2	1	2	1	1
3	3	3	3	3	4	2	3	1	1	2	2	2	2	3	1	3	3	1	2	3	2	3	1
3	3	3	1	3	4	2	3	2	2	2	2	2	3	1	2	3	3	2	1	2	1	3	1
3	3	3	1	3	4	2	1	3	1	3	1	2	3	2	2	2	3	2	2	1	2	3	1
4	3	3	4	2	4	2	1	3	1	3	1	3	3	3	3	4	3	2	2	2	2	2	1
4	4	4	4	4	1	3	2	1	2	3	3	3	1	2	2	3	3	2	2	3	1	3	1
4	2	1	4	4	1	3	3	2	2	3	1	3	3	3	1	1	3	2	2	3	2	2	1
4	4	4	4	1	4	3	1	3	3	3	1	3	4	2	3	3	3	2	2	1	1	3	1
4	4	4	2	3	4	3	3	3	3	3	1	3	4	4	2	2	3	3	2	3	2	3	1
4	2	2	1	4	2	3	1	3	1	3	1	3	2	3	3	3	3	3	2	3	3	3	2
4	1	2	2	3	4	3	3	3	3	3	3	4	4	4	3	4	3	3	2	3	3	3	1
4	3	3	2	3	4							4	4	4	4	4	1	3	2	3	3	3	1
												4	4	4	4	4	3	3	2	3	1	3	1
												4	4	4	2	4	3	3	2	3	3	3	1
												4	4	4	4	4	2	3	2	3	1	3	3
												4	4	4	2	4	4	3	2	3	1	3	1
												4	4	4	2	4	4	3	3	3	1	3	3
												4	4	4	1	4	1	3	3	3	3	3	2
												4	4	4	4	4	2	3	3	3	3	3	1
																		3	3	3	3	3	1

Kappa result in word-based

			N	Kappa Value	Meaning
databse	131	T * S	10	0.265	fair
		T * Q	10	0.077	slight
		T * C	10	0.265	fair
		T * D	10	0.265	fair
	132	T * S	14	0.164	slight
		T * Q	14	0.323	fair
		T * C	14	0.197	slight
		T * D	14	0.188	slight
	137	T * S	13	0.44	moderate
		T * Q	13	0.052	slight
		T * C	13	0.527	moderate
		T * D	13	-0.061	poor
	138	T * S	11	0.035	slight
		T * Q	11	0.214	fair
		T * C	11	0.29	fair
		T * D	11	0.441	moderate
html	133	T * S	10	0.4	moderate
		T * Q	10	0.2	fair
		T * C	10	0	slight
		T * D	10	0.4	moderate
	134	T * S	11	0.17	slight
		T * Q	11	0.761	substantial
		T * C	11	0.492	moderate
		T * D	11	0.167	slight
	135	T * S	10	0.091	slight
		T * Q	10	0.275	fair
		T * C	10	0.211	fair
		T * D	10	0.091	slight
	136	T * S	12	0.5	moderate
		T * Q	12	0.167	slight
		T * C	12	0.333	fair
		T * D	12	0.167	slight

Appendix E: REB Certificates

The REB Certificate

CERTIFICATION OF ETHICAL APPROVAL

The Athabasca University Research Ethics Board (AUREB) has reviewed and approved the research project noted below. The AUREB is constituted and operates in accordance with the current version of the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans (TCPS) and Athabasca University Policy and Procedures.

Ethics File No.: **22397**

Principal Investigator:

Miss. Miao-Han Chang, Graduate Student

Faculty of Science & Technology\Master of Science in Information Systems

Supervisor:

Dr. Maiga Chang (Supervisor)

Project Title:

Bio-inspired clustering approach based on students' annotation on online reading materials

Effective Date: December 06, 2016 **Expiry Date**: December 5, 2017

Restrictions:

Any modification or amendment to the approved research must be submitted to the AUREB for approval.

Ethical approval is valid *for a period of one year*. An annual request for renewal must be submitted and approved by the above expiry date if a project is ongoing beyond one year.

A Project Completion (Final) Report must be submitted when the research is complete *(i.e. all participant contact and data collection is concluded, no follow-up with participants is anticipated and findings have been made available/provided to participants (if applicable))* or the research is terminated.

Approved by: **Date:** December 6, 2016

Ali Akber-Dewan, Chair

School of Computing & Information Systems, Departmental Ethics Review Committee

---

Athabasca University Research Ethics Board

University Research Services, Research Centre

1 University Drive, Athabasca AB Canada T9S 3A3

E-mail rebsec@athabascau.ca

Telephone: 780.675.6718

The REB Renewal

Athabasca University
RESEARCH CENTRE

CERTIFICATION OF ETHICAL APPROVAL - RENEWAL

The Athabasca University Research Ethics Board (AUREB) has reviewed and approved the research project noted below. The AUREB is constituted and operates in accordance with the current version of the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans (TCPS) and Athabasca University Policy and Procedures.

Ethics File No.: **22397**

Principal Investigator:
Miss. Miao-Han Chang, Graduate Student
Faculty of Science & Technology\School of Computing & Information Systems

Supervisor:
Dr. Maiga Chang (Supervisor), Associate Professor, Faculty of Science & Technology

Project Title:
Bio-inspired clustering approach based on students' annotation on online reading materials

Effective Date: November 7, 2017 **Expiry Date**: November 06, 2018

Restrictions:

Any modification or amendment to the approved research must be submitted to the AUREB for approval.

Ethical approval is valid *for a period of one year*. An annual request for renewal must be submitted and

approved by the above expiry date if a project is ongoing beyond one year.

A Project Completion (Final) Report must be submitted when the research is complete *(i.e. all participant contact and data collection is concluded, no follow-up with participants is anticipated and findings have been made available/provided to participants (if applicable))* or the research is terminated.

Approved by: **Date:** November 7, 2017

Joy Fraser, Chair

Athabasca University Research Ethics Board

Athabasca University Research Ethics Board

University Research Services, Research Centre

1 University Drive, Athabasca AB Canada T9S 3A3

E-mail rebsec@athabascau.ca

Telephone: 780.675.6718