Chen, W. et al. (Eds.) (2017). Proceedings of the 25th International Conference on Computers in Education.
New Zealand: Asia-Pacific Society for Computers in Education

Synergizing Online Group Knowledge

Yen-An SHIH & Ben CHANG*
Graduate Institute of Learning and Instruction, National Central University, Taiwan
*bchang.tw@gmail.com

Abstract: A huge amount of users' data are created on the Internet day by day. How to synergize those users' data to provide useful information will be an important issue. Based on the phenomenon mentioned above, this study aims to refine the users' data on an experimental social network platform named CoCoing.info to a reintegrated concept map. The CoCoing.info platform has been online for a half year. During the period, 646 users enrolled, and 2,096 concept maps and 4,569 users' responses were created. To synergize those users' data into a connected concept map for further value-added applications, two computer algorithms, term-generation and term-association programs, were designed to filter the users' data and to reintegrate the filtered data. By the term-generation algorithm, 46,490 terms were filtered. Moreover, the term-association algorithm synergized as an 18,011-node concept map based on the filtered terms. The synergized concept map can potentially provide users to implement adaptive value-added applications on it.

Keywords: Learning analytics, big data, knowledge building, meta-knowledge

1. Introduction

With technology evolution, traditional web-based learning environment, such as MOODLE usage, has been getting challenges since general students use more social network applications to interact with their peers rather than the teacher-led activities. Moreover, with the Web 2.0 (O'Reilly, 2005) design method, those students, they can create a lot of content on their owned social network and give responses instantly. The student-created content has great potential values if we can apply useful technology to analyze and to synergize those users' data. Such kind of design thinking has been widely discussed in some research (Chatti et al., 2012; Ferguson, 2012), like big data research (Mayer-Schönberger & Cukier, 2013), learning analytics (LA) and educational data mining (EDM).

Nowadays, many popular and prevailing knowledge community (Scardamalia & Bereiter, 2006) were created. Wikipedia is one of the well-known platforms, which provides online services for users to create, edit, organize, and distribute knowledge. The purpose of knowledge community is to provide the open-sourced knowledge access and fair discussion to the Internet users. They may come from different backgrounds and abilities, but have the same study issue and come together to maintain the knowledge. They build the structural knowledge, propose innovative concepts, update the last idea, and post questions. In such kind of knowledge community, the users' work is primarily valued for what it contributes to the knowledge group (Scardamalia & Bereiter, 2014). Comparing to individual learning, knowledge community focus on learners' contribution, interaction and collaboration work.

A huge amount of knowledge is created every day on online knowledge communities. Hence, an online open-source knowledge community must handle the enormous quantity of data with the huge storage capacity. The circumstance brings and opens the era of big data in education. That means the learning process, path and pattern also can be recorded when learners are learning. By the recording data, we can conduct learning analytics method and figure out each learner's personal learning profile, for example, (1) what subjects are student's expertise, (2) what is student's interests, (3) what things student likes or dislikes, (4) and how does a student interact with knowledge community. However, before this, how to collect the huge data from learners and analyze the data become an important issue.

To sum above, in the web 2.0 environment, the volume of data is increasing faster than before. On the online knowledge community, the knowledge and learning data can automatically be saved in the database. Through learning analytics, the online knowledge data can be reintegrated and processed, and the result can further be provided to the learners. However, only a few prior studies discussed (1) how an online knowledge community can be designed to collect learner's knowledge data, (2) how to utilize the data, and (3) what is the usage of the analysis output on supporting learner's learning. To this end, this study aims to examine:

(1) How to design a social network platform in which the students can organize their personal social network, and have their learning activities on the platform as well?
(2) How to retrieve the concepts created by the students, and convert the retrieved concepts into a reintegrated concept map?

2. Related Work

2.1 Collaborative Knowledge Creation

The new communication technology provides the potential web-based environment for the students to learn on the Internet, and the online resources can be widely accessed via many portal websites, and the content can be supported with multimedia such as image, audio, video and flash animation to give learners more information than traditional text content. On the websites, the online learning content is not only created by teachers, Web 2.0 allows students from individual learning to collaborative create knowledge, share ideas and discuss with peers through online tools on a learning community. The handy usage of knowledge distribution and access triggers the trend that student has interest and engages in creating and sharing personal material. A student in the online learning community transfers the role from traditional knowledge receiver to knowledge creator and information provider. Meanwhile, the learning content is also widely accessed and evaluated by others, each learner can provide feedback and enhance the quality of knowledge contribution by discussion (Yücel & Usluel, 2016). The online environment improves the traditional teaching method and provides a space for the distributed team work to collaboratively contribute and maintain knowledge via technology device.

Past studies indicated that online collaborative knowledge work could provide the chance which teacher and students can create knowledge without the limitation of time and space (Shukor, Tasir, Van der Meijden, & Harun, 2014; Lampe, Wohn, Vitak, Ellison, & Wash, 2011). Such collaborative learning environment motivates students to create more knowledge and personal opinion because peers play an important role as guide of knowledge creation (Yücel & Usluel, 2016). In the Liaw, Chen, and Huang's study (2008), researchers developed a web-based collaborative learning system for students to learn with the online learning material and share their knowledge which relates the learning content and to examine their attitudes. The result indicated that collaborative learning environment may promote the efficiency of knowledge creation and students actively share ideas. To sum up, the collaborative knowledge creation environment has many advantages on improving student's learning performance. However, few studies discussed how to construct a collaborative learning environment and manage the knowledge data. To fill this gap, this study developed an online platform called "CoCoing.info" to implement the collaborative learning environment and to examine the effectiveness.

2.2 Educational Data Mining

Educational data mining (EDM) has been widely discussed since the mature development of computer science. EDM is a multidisciplinary research field in terms of using the technics of statistics, machine learning and data mining (Dutt, Ismail, & Herawan, 2017). EDM relies on computing power to deal with the processing of data collection, data extraction, analysis, and even system training with the large volume of data. EDM focuses on the analysis of profiles and students' learning data in which learner acts on the digital learning environment in order to discover students' learning process and profile (Levy & Wilensky, 2011). The analysis result can be used in monitoring students' learning status as well as providing feedback to the student for improving the learning problems, and to understand how and what student learns in that environment. EDM not only enriches the value of educational data by the contribution of learners who have different backgrounds but also provides the adaptive assistance to each learner according to s/he learning status.

There are many examples of the applications of EDM. For instance, students' plagiaristic behavior on assignments is a challenging problem for the teacher because it is cumbersome to examine by a human being. To solve this problem, Akçapınar (2015) applied the EDM technic in text mining on the volume of 4268 reflection texts with 59 participants. The plagiaristic ratio was provided to students as a reminder while they were posting. There is a significant improvement on the student's plagiaristic behavior according to the result of the experiment. Romero, Ventura, & García (2008) introduced how to apply the EDM on mining Moodle data. Some studies used EDM on adaptive learning with a proposed data mining system (Lin, Yeh, Hung, & Chang, 2013; Romero, Ventura, Zafra, & De Bra, 2009). However, only few research discussed how the implement the EDM on the integrating the group knowledge and discuss the mining process. To fill this gap, this research aims to process of how to create the knowledge relation by text mining and generating the integrating concept map.

3. Data Source: CoCoing.info Social Network Platform

To exam the goals of how to reintegrate group knowledge, an online learning platform called CoCoing.info had been developed to collect the knowledge data from learners and conducted in the experiment. As shown in Figure 1, students contribute knowledge and learn in the ubiquitous computing environment. The knowledge data was converted into terms by the term-generation algorithm and saved in the database. Those terms were further used to connect concepts and knowledge by the term-association algorithm and create a reintegrated concept map to support further value-added applications. The scenario of reintegrating group knowledge on CoCoing.info is shown in Figure 1.

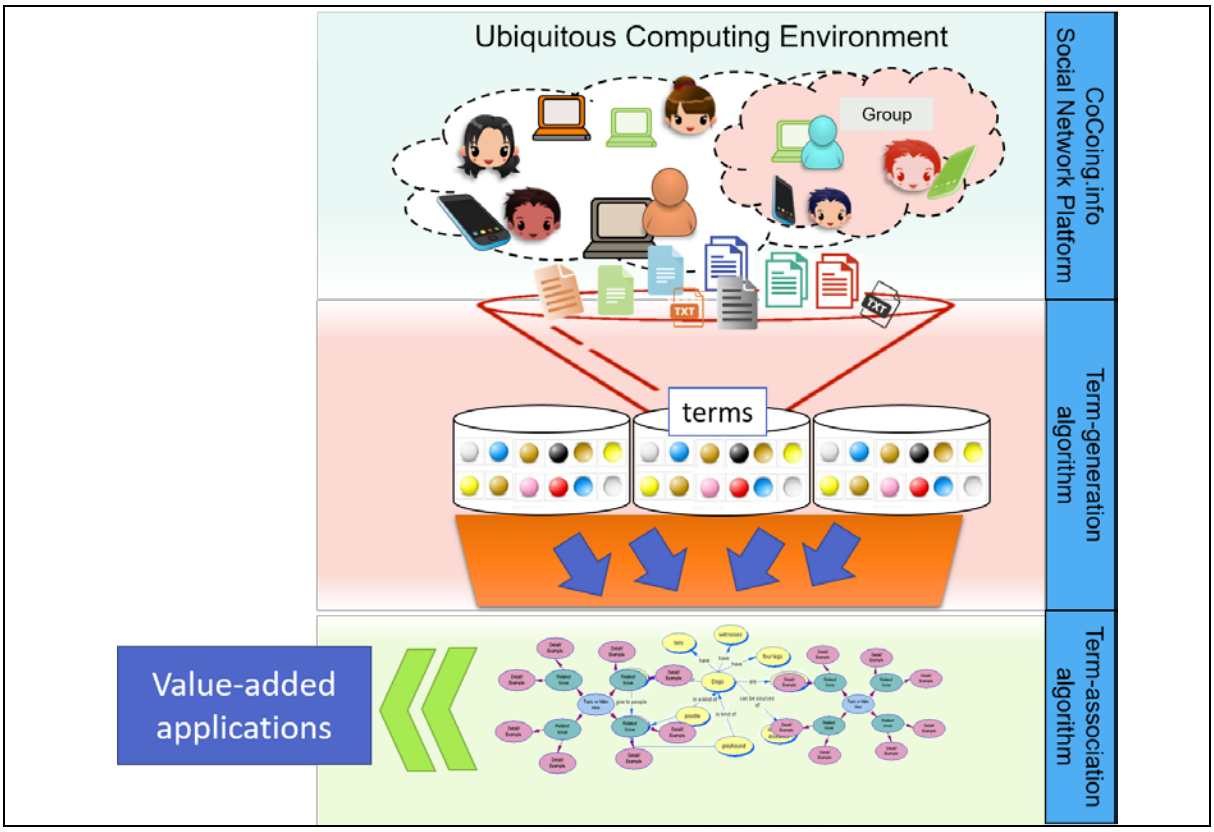

Figure 1. The scenario of reintegrating group knowledge on CoCoing.info.

3.1 CoCoing.info Background Introduction

There are two major designs on CoCoing.info for facilitating learners producing knowledge and cooperative constructing ideas. The first major design is the concept construction, and the other is the social learning network design. The structure of CoCoing.info is shown in Figure 2.

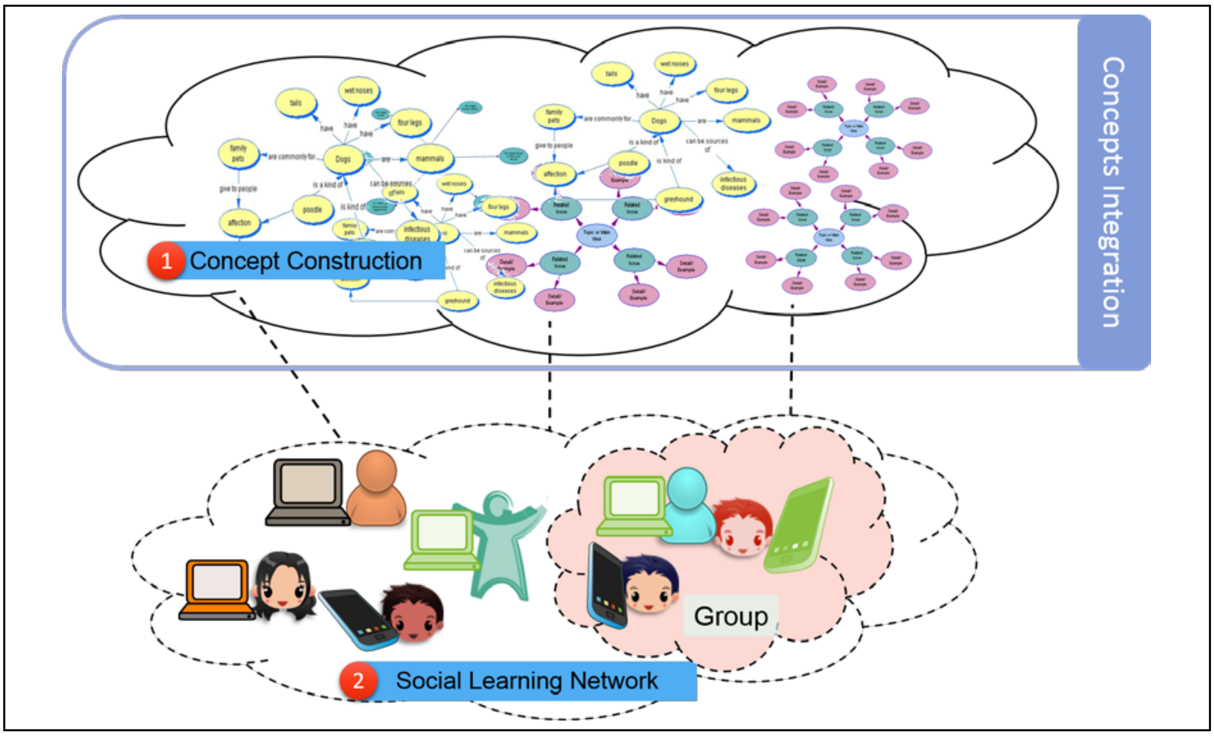

Figure 2. The structure of CoCoing.info.

3.1.1 Concept Construction

Knowledge is a thinking product which exists in the personal mind. However, how to transform the invisible thinking into the visible and structured knowledge is a challenge. In this study, we utilized the concept map tool and implemented into CoCoing.info platform for facilitating learners building knowledge. The node and line are two basic elements in a concept map. With the main idea, it can be expanded to many nodes and connected by lines. CoCoing.info provides functions for learner easily building a structured concept map. By the clicking function buttons, the learner can create the many colorful nodes and links on a map, and attach relative images and files as learning resources on each node to illustrate the learning content. Furthermore, a node can connect to online learning resource with URL so that learner can easily integrate online learning material into a concept map. The concept map construction page and the description of each function are shown in Figure 3.

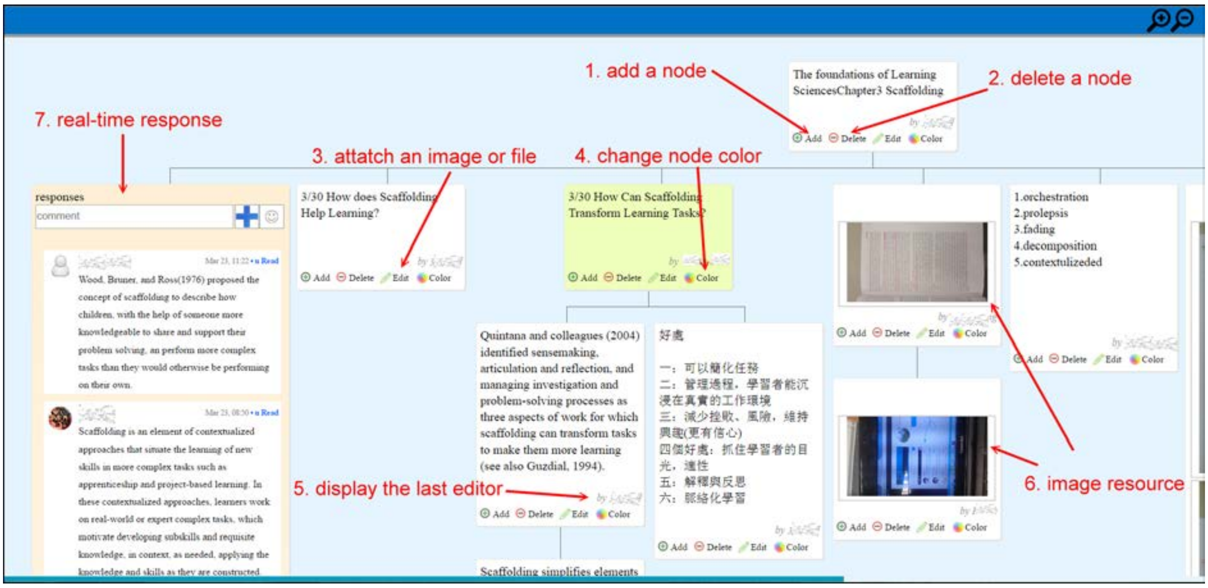

Figure 3. The concept map construction page on CoCoing.info.

3.1.2 Social Learning Network

The social network is an important element among students when they are learning. The relationship between peers is a pattern of how students interact and influence others. On the CoCoing.info platform, there are several functions for learners to create and represent their social network from the classroom. For example, they can add their peers as a friend by name or email on CoCoing.info, and two or more friends can gather and form a learning group. They can become the friend relationship with each other. The social network among students is a complicated net connection working on CoCoing.info. With the integration of social network and concept map, CoCoing.info can support the collaborative learning environment. The learner can share a concept map with friends or groups. Learners can simultaneously and collaboratively construct the same concept map with peers, and teacher can join students and provide guidance when they are constructing. Each node displays the last editor's name. In the process of concept map constructing, learners and teachers can discuss, provide feedbacks, and exchange ideas simultaneously with each other by the response function.

3.2 Participants

There were 646 users enrolled in the system. They were ten to twenty-four years old covering in elementary school, junior high school, and college students. All of the teachers had been educated how to use the functions on CoCoing.info by researchers with face-to-face teaching. Therefore, students registered their personal account and went through the operation guidance about CoCoing.info over three-hour class periods for two weeks by teachers to ensure they were skilled. In this study, the teachers were as the role of supporter and provided guidance while learners constructing their concept maps.

4. Data Synergy: Term-generation and Term-association Algorithms

4.1 A Process Model for Learning Analytics

On CoCoing.info learning platform, learners can easily create personal and cooperative concept maps, and the concept map data is automatically stored in CoCoing.info database. To achieve the data reusing, we proposed the model of learning analytics and try to mine out the useful information from the concept map data. The analytics output could provide the feedback to students in order to enhance their learning performance and extend their knowledge and perspective. The process mode of learning analytics is shown in Figure 4.

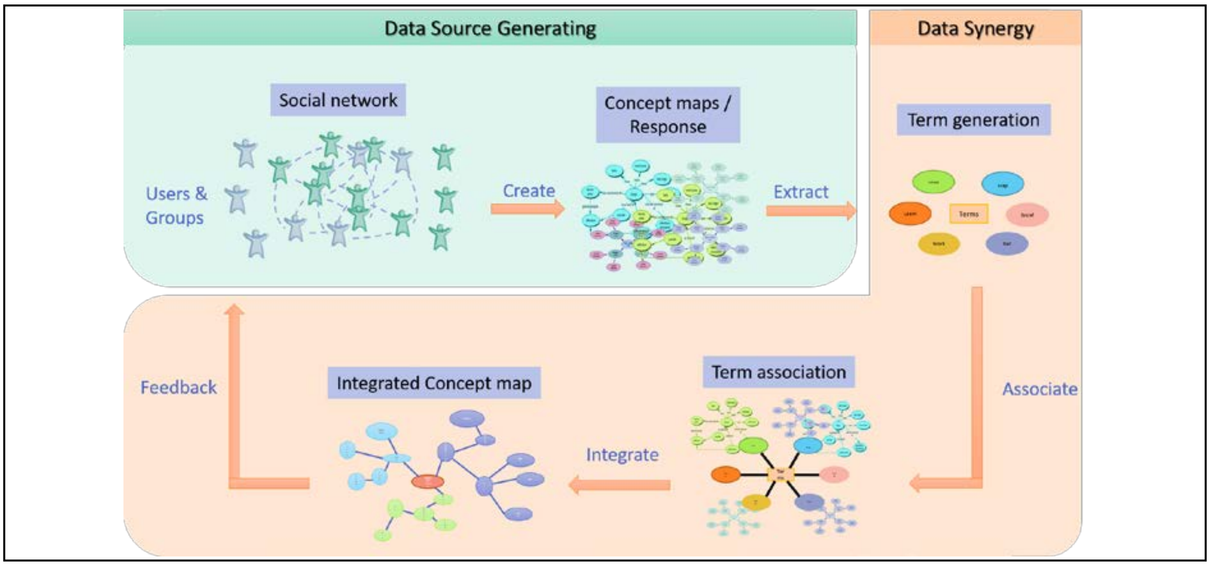

Figure 4. The process model of learning analytics.

4.2 Term-generation Algorithm

In CoCoing.info database, the concept maps and responses data was selected as the resource of term generation and training. A total of 2,096 concept maps and 18,011 nodes which generated by learners without null or empty value from the period of November 1, 2016 to April 30, 2017 were selected as experiment data, and the 4,569 responses data was selected as well. The content of nodes and responses included English and Chinese text. To generate terms, the content was split into a single word as the basic element. Each word went through others and combined as a new phrase. The new phrase may contain two or more word in the phrase. With the loop process of phrase composition, we counted the sum of repetition and recorded it as the weight for each phrase. According to the weight number, a new term may be created if the number is higher than others. Therefore, the new term can be used in analytic and continue changing its weight number by the new data and the loop process. The term extraction algorithm is shown in Table 1.

Table 1: The term-generation algorithm pseudo code.

```
Do Loop
        Extract text from nodes and responses
        If source is Chinese characters
                Set position at x word
                Split text from position x, y words as a term
        IF source is English words
                Filter alphabet from text
                Combine alphabets as a term
        Push term into array cell
End of loop
Save text, array, array length as a record
```

According to the term-generation algorithm mentioned above, a total of 46,490 terms were generated based on the 2,096 concept maps and 4,569 users' responses data stored in the CoCoing.info database. Figure 5 displays part of the terms list. In Figure 5, a user's sentence is parsed to a list of terms for further analysis.

Figure 5. The screen capture of term-generation page.

4.3 Term-association Algorithm

As mentioned above, the relation between nodes is represented as a line in a concept map. Each node may have its parent node and child node. However, the lines in a concept map cannot interact with different concept maps. For this purpose, the terms work as an index which represents the advance relation between nodes if the terms are similar. We compared the term list among two nodes and counted its similarity with the algorithm. There were 46,490 terms generated and the highest occurrence number was counted by 3,176 times. With the highest number, we can calculate the weight for each term by the occurrence number. However, not all of terms can be used in the analysis, the lower weight terms were removed in the pre-processing and only high weight terms were kept to the comparison. The similarity number ranges from 0 “no similarity” to 100 “high similarity”. The association algorithm is shown in Table 2.

Table 2: The term-association algorithm pseudo code.

```
Collect terms of text A and text B as array A and array B from database
    For i= 1 to the length of array A
        For j= 1 to the length of array B
            Compare term i and term j
            If term i and term j are matched
                Match count +1
        End of For loop
    End of For loop
Calculate the minimum of array A and array B
Similarity = match count / minimum
Visualize terms relation based on similarity result
```

4.4 Terms Relation Visualization: Integrative Concept Maps Generating

By the term-association algorithm, the similarity of the relation among nodes can be calculated. Therefore, the integrative concept map can be generated based on the similarity calculated results. Currently, the reintegrated concept map is an 18,011 nodes concept map which is a huge map. The reintegrated concept map is expanding day by day automatically by the term-generation algorithm and term-association algorithm once users continually use the system. Figure 6 only displays part of the reintegrated concept map. However, the whole reintegrated concept map can be retrieved via the Internet for further value-added applications.

As we can see in Figure 6, the nodes were displayed as the block with light blue color. The concept map relation was represented as the solid line with black color. The nodes were drawn closer if they were in the relation of concept map which showed as the high-density area, and there were few outliers which were deleted and had no relation among nodes as well. Different similarities were drawn with the dotted line with the red color (51~60), orange color (61~70), green color (71~80), blue color (81~90), and pink color (>90), respectively.

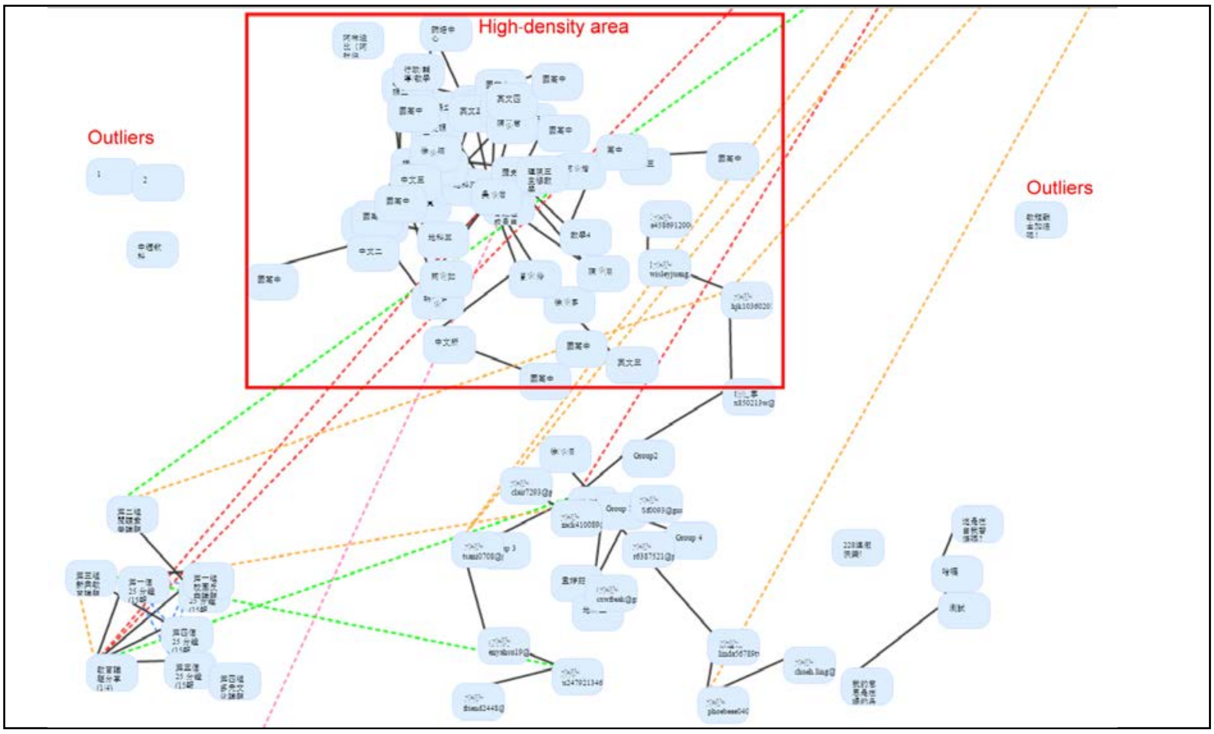

Figure 6. The screen capture of integrative concept map.

5. Discussions and Conclusions

No doubt that on the Internet, the users generate a big volume of data every day. Novel data analysis approaches, such as educational data mining and learning analytics, can retrieve those data to provide useful information. The research trend of applying EDM and LA in the educational system is getting important (Romero & Ventura, 2007).

In this study, an online platform named CoCoing.info was implemented which is applied to study on how to design an online knowledge community, and how to collect the structured users' data. Meanwhile, two computer algorithms, term-generation and term-association, were adopted to filter the users' data and to reintegrate the filtered data into a reintegrated concept map automatically.

A half year period of the announcement of the CoCoing.info platform, a total of 2,096 concept maps, 18,011 nodes, and 4,569 responses were collected in the database. Moreover, there were 46,490 terms generated by the terms-generation algorithm, and the highest occurrence number in terms is 3,176 times. The results indicated that the relative concepts can be organized and connected. Moreover, the concepts from different learners can be combined together for further applications. Those results provide further learning platform designers a guideline and approach on designing adaptive learning, appropriated responses, and users modeling system. More specifically, those adaptive systems mentioned above need a huge concept map as the fundamental knowledge set to provide appropriate reactions, and this study illustrated how such a huge concept map can be created and reintegrated automatically.

Comparing to the big data research, this study still bases on small samples. However, the design has provided a useful reference model and the results have revealed some interesting insights that further designers' references. Furthermore, the algorithms discussed in this study of meta-knowledge generation can be applied in different tasks in education. For example, the meta-knowledge can provide the learning suggestions for the learner to improve learning performance, and improve teacher's pedagogy as well. The meta-knowledge also can be used in the training of machine tutor as the scaffolding while student learning.

References

Akçapınar, G. (2015). How automated feedback through text mining changes plagiaristic behavior in online assignments. *Computers & Education, 87*, 123-130.

Chatti, M. A., Dyckhoff, A. L., Schroeder, U., & Thüs, H. (2012). A reference model for learning analytics. *International Journal of Technology Enhanced Learning, 4*(5-6), 318-331.

Dutt, A., Ismail, M. A., & Herawan, T. (2017). A Systematic Review on Educational Data Mining. *IEEE Access*.

Ferguson, R. (2012). Learning analytics: drivers, developments and challenges. *International Journal of Technology Enhanced Learning, 4*(5-6), 304-317.

Lampe, C., Wohn, D. Y., Vitak, J., Ellison, N. B., & Wash, R. (2011). Student use of Facebook for organizing collaborative classroom activities. *International Journal of Computer-Supported Collaborative Learning, 6*(3), 329-347.

Levy, S. T., & Wilensky, U. (2011). Mining students' inquiry actions for understanding of complex systems. *Computers & Education, 56*(3), 556-573.

Liaw, S. S., Chen, G. D., & Huang, H. M. (2008). Users' attitudes toward Web-based collaborative learning systems for knowledge management. *Computers & Education, 50*(3), 950-961.

Lin, C. F., Yeh, Y. C., Hung, Y. H., & Chang, R. I. (2013). Data mining for providing a personalized learning path in creativity: An application of decision trees. *Computers & Education, 68*, 199-210.

Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt.

O'Reilly, T. (2005). *What is Web 2.0? Design patterns and business models for the next generation of software.* Cambridge: O'Reilly.

Romero, C., & Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. *Expert systems with applications, 33*(1), 135-146.

Romero, C., Ventura, S., & García, E. (2008). Data mining in course management systems: Moodle case study and tutorial. *Computers & Education, 51*(1), 368-384.

Romero, C., Ventura, S., Zafra, A., & De Bra, P. (2009). Applying Web usage mining for personalizing hyperlinks in Web-based adaptive educational systems. *Computers & Education, 53*(3), 828-840.

Scardamalia, M., & Bereiter, C. (2006). Knowledge building: Theory, pedagogy, and technology. In K. Sawyer (Ed.), *Cambridge handbook of the learning sciences* (pp. 97–118). New York: Cambridge University Press.

Scardamalia, M., & Bereiter, C. (2014). Knowledge building and knowledge creation: Theory, pedagogy, and technology. In K. Sawyer (Ed.), *Cambridge handbook of the learning sciences (2nd ed.)* (pp. 397-417). New York: Cambridge University Press.

Shukor, N. A., Tasir, Z., Van der Meijden, H., & Harun, J. (2014). Exploring students' knowledge construction strategies in computer-supported collaborative learning discussions using sequential analysis. *Educational Technology and Society, 17*(4), 216-228.

Yücel, Ü. A., & Usluel, Y. K. (2016). Knowledge building and the quantity, content and quality of the interaction and participation of students in an online collaborative learning environment. *Computers & Education, 97,* 31-48.