Tools for Tracing the Development of Concepts through Discussions Mediated by a CSCL Environment: A Case Study

Gülgün Afacan Adanır, Informatics Institute, Middle East Technical University; Distance Education Center, Ankara University, gafacan@ankara.edu.tr
Murat Perit Çakır, Informatics Institute, Middle East Technical University, perit@metu.edu.tr

Abstract: This case study explores the use of learning analytics techniques to monitor students' conceptual development evidenced in time-stamped logs of a CSCL environment that provides chat, shared whiteboard, and wiki features. The study was conducted in a graduate level research methods course, which included online assignments that required students to collaboratively discuss questions related to statistical methods in chat sessions and submit their answers through co-authored wiki documents. This paper demonstrates the use of some of the existing learning analytics techniques to develop practical strategies and interfaces for helping instructors to effectively monitor the collaborative knowledge building processes taking place at micro and macro levels. In particular, we demonstrate the use of topic segmentation, tag clouds and concordance analysis for the identification of excerpts where key concepts are discussed by the students.

Keywords: learning analytics, knowledge building, chat, wiki

Introduction

CSCL is a field primarily concerned with the use of information and communication technologies to support learning through collaborative activities. Most CSCL systems offer a variety of communication channels with rich representational affordances, including shared workspaces, text-chat, wiki, discussion board and video-conferencing applications. Such tools enable instructors to support and manage rich learning experiences for their students. These tools also offer unique opportunities for the analysis and assessment of learning interactions as they provide time-stamped logs of all collaborative activities.

Learning occurs in an interactive and dynamic way in CSCL environments, so tracking the collaboration process is an important concern for monitoring and supporting learning activities in CSCL. Assessment of learning in CSCL generally involves two levels; namely product and process assessment. Product assessment involves the evaluation of the final outputs/outcomes to check whether a key skill has been appropriately employed or a specific concept has been mastered, whereas performance assessment is concerned with the quality of the learning process (Retalis, Petropoulou & Lazakidou, 2010). Moreover, assessment in CSCL is also considered as a means to enhance the collaboration process through feedback (Collazos, et al., 2007). For example, providing information on students' own activities can contribute to their awareness and meta-cognitive status, and as a result may facilitate self-regulation of the learning activity (Daradoumis, Martínez-Monés & Xhafa, 2006; Nurmela, Lehtinen & Palonen, 1999). In addition to this, records of learner successes/failures and recommendations for future learning activities based on such records may result in a more structured and efficient learning process (Wang, 2009).

Most CSCL applications automatically record information related to interactions of participants such as messages and documents exchanged (sending and reading timestamps, name of the sender, name of the readers, etc.) in log files (Pozzi, Manca, Persico & Sarti, 2007). However, the sheer volume of data generated by heterogeneous online communication channels, even in the context of a semester-long course, brings practical challenges for the monitoring and facilitation of collaborative learning activities by the educators. The challenges involved with processing this rich body of data forces educators to resort to superficial assessment of learning based on tests, without being able to take into account the micro-level processes of knowledge building which are key to the success of CSCL applications. Therefore, there is a need for tools and strategies for helping educators and researchers to make the best use of this rich data.

The big data phenomenon in business analytics and the increasing amount of data in online educational repositories have led to the emergence of the field of Learning Analytics. According to the definition contributed by the recently established Society for Learning Analytics Research, Learning Analytics is concerned with the measurement, collection, analysis and reporting of data about learners and their contexts of learning, for the purpose of understanding and optimizing learning and the environments in which it occurs (Siemens & Gasevic, 2012). The collection of data and devising analytics to make sense of the trails left by learners is a fundamental concern in this emerging field. Such trails involve information on key aspects of learning such as information access and use practices learners follow, the social networks they form, the content of interactions they engage with, and the knowledge artifacts they construct in the course of their learning process. Educational data mining and analysis of learning interactions within socio-technical systems are dominant themes in the emerging field of learning analytics (Siemens, 2012; Scherer et al., 2012).

Educational Data Mining (EDM) focuses on devising predictive relationships among features extracted from learner logs to better inform instruction (Baker &Yacef, 2009; Romero & Ventura, 2007; Romero et al., 2010). Automated discovery of learning needs and adapting learning resources to better cater to those needs are key components of the EDM approach. Typical EDM applications involve student modeling where successful as well as risky cases (e.g. a student who is likely to be dropping out) can be automatically detected, and recommender systems that allow students to interact with personalized content based on predictions about their learning needs/styles inferred from their past history (Stamper et al., 2010; Manouselis et al., 2012). Such applications extend the assessment of learning outside individual courses and allow educators to monitor the progress of students as members of a larger learning community (Hung, Hsu & Rice, 2012).

The socio-technical approach focuses on the content and the nature of the learning interactions mediated by learning environments as a systemic whole (Shum & Crick, 2012; Siemens, 2012). Building visualizations of social networks and studying the information flow within those networks with discourse analytic methods are of particular interest in this approach (Ferguson & Shum, 2011, 2012). Such tools are generally intended not only for research use, but also to support teachers' self-reflection on their teaching practice and to inform educational decision makers by providing a broader view of learning activities (Dyckhoff et al., 2012; Govaerts et al., 2012). Design of representations and analytic constructs that facilitate the coordinated analysis of learning traces distributed across individuals, collectivities and media in networked learning environments is another important thread in the socio-technical approach (Suthers & Rosen, 2012). Such tools aim to bring the learning traces distributed across multiple media and sites together to enable the investigation of emergent learning phenomena within a learning community.

In this study we employ a socio-technical approach to analyze the collaborative learning process taking place in a CSCL environment called Virtual Math Teams (VMT) that offers chat and wiki features. More specifically, we explored the use of learning analytic methods to investigate a learning group's conceptual development in a CSCL environment in the context of a semester long statistics course. Conceptual development was investigated according to the knowledge building theory (Scardamalia & Bereiter, 2006) which argues that knowledge is produced through the formation of common goals and negotiation of different perspectives. We attempt to examine how a particular group of students developed their understanding of some the key concepts in statistics during their collaborative activities distributed across multiple interaction spaces and spanning the entire semester. In particular, we aimed to illustrate the use of tag clouds and concordance analysis to locate segments where key statistical concepts were discussed, as part of a process analysis of conceptual development that spans micro and macro levels.

Methods and data

In this study the Virtual Math Teams (VMT) system was used to support and record the collaborative learning activities that took place in the context of a semester long course on research methods and statistics. The VMT system was developed as part of a research project that aims to support collaborative math problem solving activities at a distance (Stahl, 2009). Although the VMT system primarily attempts to serve the mathematics education domain, learning groups can use this platform to engage in collaborative learning activities in other domains as well.

The VMT online environment provides both quasi-synchronous and asynchronous collaboration tools to support collaborative learning activities. The chat component provides support for quasi-synchronous communication for the members of a learning team through the exchange of text-messages. At the same time, chat rooms offer shared whiteboards for drawing and organizing ideas. The chat platform also presents a shared web browser facility, which allows group members to collaboratively browse the web to support their group work. Finally, each chat room has a corresponding wiki page, through which learners can publish their collective findings in the form of co-authored wiki documents. The wiki component is based on MediaWiki.

The study has been conducted in the context of a graduate level Research Methods & Statistics course during 2013-2014 fall term at the Middle East Technical University (METU). There were 21 registered students in the course. Each registered student was assigned to a learning group and seven teams were constructed in total. All teams were required to complete course assignments by collaboratively working online in the VMT environment. That is, learning groups are initially required to perform online chat meetings, then publish their findings as co-authored wiki documents. The online activities were graded as group projects which constituted half of the total grade students obtained from the course. The remaining half of the grade was based on individual test scores students obtained from two conventional exams.

The assignments cover standard statistical methods including descriptive statistics, exploring data with graphs, correlation/regression methods and methods for testing hypotheses about group differences such as t-test, ANOVA and their non-parametric equivalents. The aim of the online activities was to help students develop their understanding of key statistics concepts through collaborative assignments where they attempted to conduct a specific type of analysis by using the SPSS software. Some concepts such as identification of independent/dependent variables, their scale of measurement, whether variables satisfy parametric assumptions (i.e. normality and homogeneity of variance), the notion of null hypothesis and statistical significance were common to all online activities due to their central role in statistical analysis. Developing a deep understanding of each of these concepts was targeted as learning goals of the course. Our case study focuses on learners' progress in one of these key dimensions during the entire term, namely identifying variables and checking parametric assumptions. The chat logs that were analyzed as part of the case study were obtained from the fourth assignment during the semester, which included the following instructions:

> A study of reading comprehension in children compared three methods of instruction. First, all participants' reading comprehension levels were assessed with a pre-test. Then, participants were split into 3 groups, where they were exposed different methods of instruction to develop their reading comprehension skills. Finally, all group members were given a post-test that is comparable to the pre-test in terms of content. The data for the study is stored in reading.sav file.
>
> 1. Identify the dependent and independent variables of this study. At what level of scale each variable is measured?
> 2. Are the dependent variables normally distributed? Perform the appropriate tests in SPSS and report their results (Note: use the appropriate group level for these tests.)
> 3. Focus on the pre-test results only. Draw a bar chart with 95% confidence intervals. Is there a difference among the groups? Which test would be appropriate to test whether there is a statistically significant difference among the groups and why? What is the null hypothesis? Do the test and report the test results (you should use the reporting guidelines in the book). If there is an overall difference, which pair of groups differ from each other? Again, explain what statistical test you are using to make that argument.
> 4. Next, focus on the post-test results. Draw a bar chart with 95% confidence intervals. Is there a difference among the groups? Which test would be appropriate to test whether there is a statistically significant difference among the groups and why? What is the null hypothesis? Do the test and report the test results (you should use the reporting guidelines in the book). If there is an overall difference, which pair of groups differ from each other? Again, explain what statistical test you are using to make that argument.
> 5. Finally, focus on each instruction group separately. Which test should you use to compare the difference between the pre and post test scores of each student in each instruction group? Do the appropriate test(s) and report the results in the formal reporting format.

Data collection

This study focuses on excerpts obtained from the online sessions of a single team in this corpus. The participants' actions in the chat environment were recorded as chat log files, which were automatically logged by the VMT system. Teams used a single chat room for each assignment, so one chat log file was generated for each group. The chat log contained the author, date, start time, post time, duration, and event type for each action entry. Remaining columns are allocated for indicating chat messages and other activities of students (e.g. awareness messages such as user is typing, drawing on the whiteboard etc.).

Chat discussions continue with learners' submission of solutions as wiki content, which are represented with screenshots in Figure 1. Wiki activities of learners are listed in the "View History" page and listed from initial to recent one. Each wiki activity is tagged with its author and time information. Successive activities can be compared to identify learners' editions and removals related to the wiki content. The VMT also provides Wiki activities in textual format, hence facilitates the analysis of the evolution of the wiki content.

Data analysis

After collecting data, we consider the following steps for the analysis of chat logs:

- Segmentation Analysis

- Removal of stop words
- Use of Tag Clouds to identify recurrent concepts
- Concordance Analysis to identify the context of concepts
- Interaction Analysis of episodes for tracking learners' development of concepts

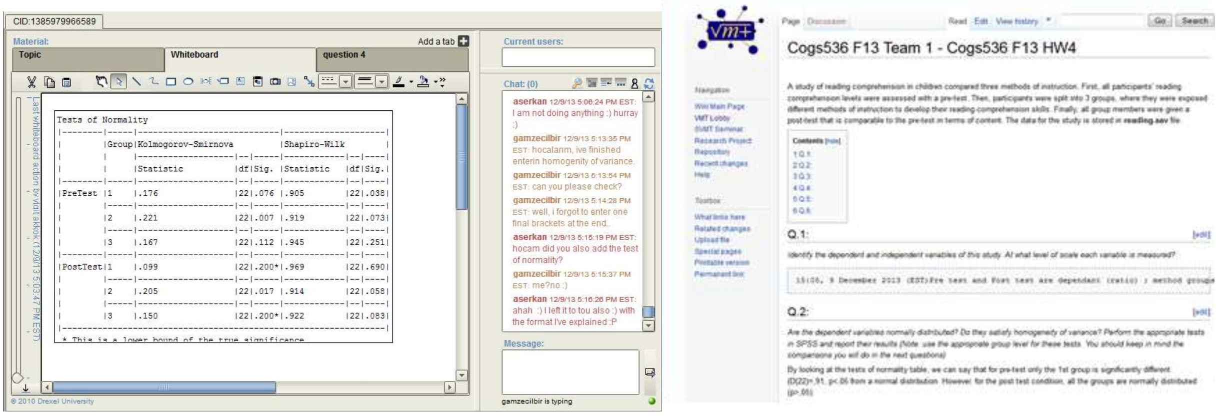


Figure 1. VMT chat and wiki components

Segmentation analysis aims to capture how participants organize their chat interaction into long sequences (i.e. chunks of activity). For this purpose, chat logs are investigated to identify activity boundaries where new activities are initiated and current activities are terminated or suspended. That is, transitions where learners either (1) close one activity to initiate a new one, or (2) temporarily suspend an ongoing activity and start a temporary one as an insertion sequence, are identified by investigating topic/activity change markers (Zemel, Xhafa & Cakir, 2007). As a result, chat logs are organized into segments.

Next, stop words (e.g. words that provide no content such as prepositions, conjunctions, determiners) are eliminated from chat logs in order to prepare the data for further analysis. Then, chat logs are subjected to content analysis to identify recurrent keywords in that session. In particular, tag clouds are computed for pre-processed log files where the size of the word indicates its frequency. In our study, tag clouds are employed to identify the statistical concepts learners discuss during their chat sessions.

Thirdly, concordance analysis is employed to identify the context in which key terms of interest occurred in the chat logs. In this case study, we focus on tracking the evolution of a group of learners' understanding of variable types and parametric assumptions. Once we identify the contexts in which key statistics concepts were mentioned by the team, we focus on the sequential organization of chat messages and whiteboard actions in that episode to observe how learners referred to and made use of these concepts.

Finally, the wiki content is analyzed as a reflection of the discussions that took place during the team's chat session. Wiki content constituted the final deliverable submitted by the team, so its content is organized to be read as a summary of the team's findings. The way team members use key concepts in their wiki pages are further analyzed to trace their conceptual development.

Results

The team consists of three students, whose demographic characteristics are provided in the Table 1.

Table 1: Demographic characteristics of students

Subject Handle	**A_S**	**G_C**	**Y_A**
Age	Over 29	22-29	Over 29
Gender	Male	Female	Male
Grade	PhD	Masters	Masters
Undergraduate major	Physics	Foreign Language Education	Electric and Electronics Eng.
Graduate major	Biomedical Engineering	Cognitive Science	Cognitive Science
Current GPA	3.00-3.50	3.00-3.50	3.00-3.50

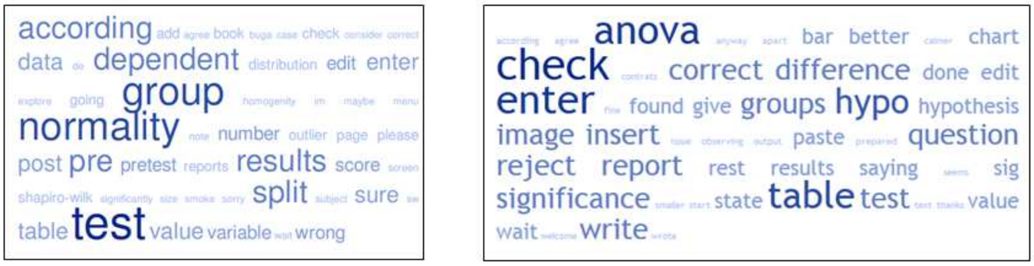
The team has performed a series of online meetings while working on their group assignments. In this study, we provide results from their fourth assignment to illustrate our analysis. The chat log for the fourth assignment consists of 376 chat lines. The task description was provided in the methods section.

Chat segmentation results

Chat logs of the assignment have been quickly investigated to identify specific discussion topics. The team has performed three online chat meetings and considered various topics during these sessions. The topics mainly consist of coordination issues and learners' discussions related to the assignment. We consider segments related to the assignment in the subsequent steps of our analysis.

Key concepts the team considered

The next chat processing step aims to explore key concepts that the team has employed during their online collaborative studies. For this purpose, we mainly focus on segments that capture the discussion of the members on the assignment. These task-related logs, which are identified during the segmentation analysis step, belong to two different online meeting of the team. In parallel to our purpose, we have employed tag clouds on these task-related logs to identify concepts that the team considered while they were collaboratively solving the questions involved in the assignment. We used the TagCrowd to obtain the two tag clouds displayed in Figure 2 that represent the most frequently observed key terms in the task-related segments of both logs.



<u>Figure 2.</u> Key concepts discussed in the group meetings

In this case study, we focus on the learners' discussion of two specific statistical concepts – variables and normality test. These two concepts frequently occurred in the first tag cloud. For instance, the "normality" concept is reflected with 'normality' and 'test' keywords together with their large sizes. Similarly, the "variables" concept is identified by 'dependent' and 'variable' key words.

Concordance analysis

Concordance analysis is applied to indicate locations of the keywords within the chat logs. Compared to segmented excerpts, the lines obtained from concordance analysis are minimal and more specific. For instance, the team's task related activities are identified as segmented excerpts, whereas chat lines related to learners' understanding of the variable concept are obtained through the concordance analysis.

The results of the concordance analysis demonstrated that the "variable" concept was discussed between chat lines 33 and 36, and the "normality test" was discussed between chat lines 98 and 137. Although each sentence within these chat lines doesn't consist of the keywords, we consider all messages since they are components of an ongoing interaction and have pragmatic and semantic relationships with the lines containing the keywords. Simply ignoring the lines that do not include the keywords brings problems of intelligibility, since chat unfolds sequentially and the meaning of each utterance need to be analyzed within this sequential context.

Learners' development of concepts

Once the relevant excerpts are obtained through segmentation, tag cloud and concordance analysis steps, we focus on the interactional content where the "variables" and "normality test" concepts were discussed by the team. Our purpose is to understand how learners made progress throughout chat activities while working on these concepts.

<u>Variables</u>

The excerpt obtained from the concordance analysis related to the variables concept is provided below.

1. A_S: we can start to discuss dependent and independent variables and the level of scale
2. G_C: i think Pre test and Post test are dependent (ratio); method groups are independent (nominal-categorical) variables.?

3. Y_A: That is exactly my opinion too.
4. A_S: I agree with your opinions

A_S initiated the discussion about the "variables" concept with his remark in line 1, possibly in response to the first question in the assignment. This is taken up by G_C in line 2, where she proposed the test scores as the dependent and the group categories as the independent variables. She also proposed that test scores are measured at the ratio level and group is a nominal-categorical variable. In the next two lines, Y_A and A_S agreed with G_C. The team quickly came to an agreement around G_C's proposal. Note that this was the team's fourth assignment where they answered similar questions for the previous three assignments. Coming to an agreed answer for the same question took more time and turns in those previous cases, so the team seemed to have progressed in detecting and categorizing variables involved in a given research design description. However, one could criticize the argument that test scores are measured at the ratio scale, since a score of 0 does not necessarily imply absence of reading comprehension skills.

Normality test

Initial lines of the excerpt obtained from the concordance analysis are provided below.

1. Y_A: so about the normality tests
2. Y_A: I've got some results from the explore menu item
3. Y_A: some look normal, some not.
4. G_C: sorry before we move on, did you split the file?
5. A_S: actually I splitted the file but I got weird results... I'm doing somethings wrong
6. Y_A: no i didn't. should I?
7. G_C: no, i just wanted to be sure
8. A_S: I think its better to split hocam
9. A_S: how did you do it without split Y_A?

Before the first chat message, Y_A shared the results of his tests of normality in the whiteboard area. He then indicates in the first three lines that he applied the normality tests by using the Explore feature of SPSS, and found that some variables were normally distributed, whereas others were not. Through these chat messages Y_A reported his initial finding about the distribution of data, without specifically identifying the normal and non-normal cases. In line 4, G_C asks whether Y_A had considered splitting the data before testing for normality. Next, A_S comments that he obtained weird results when he tried splitting the data, and states that he had probably done something wrong. In the next line, Y_A responds to G_C that he didn't split the file, and asked if he should had done so. G_C's response in the next line indicates that she did not see this as a necessity, but she was mainly reminding her teammates about a possible issue. In line 8, A_S argues that it is better to split (hocam is a Turkish term used as a colloquial way to address a fellow student or colleague) and asks Y_A how he did the analysis without splitting. In the following conversation (not provided in the transcript) Y_A provided a summary of his steps where he explained how he conducted the normality test by using the explore menu in SPSS by defining pre-test and post-test as dependent variables. This short exchange among the team members indicate that they took issue with an important concern, namely identifying the correct level to check for the normality assumption. The problem statement states that there are three independent groups in the experiment, whose scores should be tested for normality separately. Splitting the data set is one way to achieve this in SPSS depending on how the data is organized. This discussion provides evidence that the team members are aware of finding the appropriate level to apply the test, but they have neither justified nor demonstrated this explicitly.

In the segment identified by the concordance analysis, Y_A shared the output he obtained from SPSS for the Kolmogorov-Smirnov (K-S) and the Shapiro-Wilks (S-W) tests. Y_A stated that group-2 has non-normal distribution in both pre and post test cases according to the K-S test, whereas only group-1's pre-test scores are not normally distributed according to the S-W test. Then, Y_A asked others whether they agree with these results. G_C stated that she applied the test and found the same results as Y_A. Previously G_C couldn't produce results in SPSS. Y_A's statements seemed to help G_C to replicate the analysis on her computer.

Next, the team discussed what they should do with the variables that violate normality. Y_A argued that all the scores could be considered fairly normal, since the sample size of 22 was not so small and the q-q plots looked fairly on the diagonal. G_C agreed with Y_A. Then, A_S reminded the team that when the sample size is less than 30, S-W is a more conservative test of normality, and argued that S-W could be the more reliable test in this case. During this discussion it turned out that the reason why A_S found weird results was due to an incorrect splitting he applied on the data. Y_A's comments helped A_S correct his analysis.

In the following part of the discussion, Y_A proposed that the deviation from normality in the variable of concern was due to an outlier, which he noticed on the q-q plot, and wondered if that could be a typo in the data. G_C agreed on the presence of an outlier but argued it could be a genuine data point, as no information about minimum and maximum possible scores were given in the problem statement. The team then agreed that the outlier was not to be treated as a typo. A_S asked if they will ignore the outlier and consider pre-test scores of group-1 as normal. Y_A proposed to explain the significance of the S-W test due to the presence of this outlier score, and continue with a parametric test for subsequent analysis. G_C and A_S's agreement concluded the discussion on checking the assumption of normality in this log.

Wiki reflection

According to the wiki logs, G_C wrote the results about variables as follows: "Pre test and Post test are dependent (ratio); method groups are independent (nominal-categorical) variables. "Y_A shared the results of the normality tests as a table, whereas G_C contributed the interpretation "By looking at the tests of normality table, we can say that for pre-test only the 1st group is significantly different (D(22)=,91, p<,05) from a normal distribution. However, for the post test condition, all the groups are normally distributed (p>,05)".

The wiki summary does not capture all the details of the team's chat discussion. The team stated their answer for the variable types in the same way as it was articulated in one of the chat messages. They presented the normality analysis with the correct groupings, but provided the standard interpretation of the K-S test results. The wiki posting for this particular question seem to suggest that the group members changed their mind about treating the pre-test score of group-1 as normally distributed. In particular, they didn't mention their noticing about the outlier and its effect on the normality test. However, in the remaining parts of the question, the team employed a parametric test to complete their analysis, which seemed to be a consequence of this discussion.

Discussion and conclusion

A significant advantage of CSCL environments is that they provide system logs that record details of interactions experienced among students. Since these logs capture instances where learners ask questions, look for information and make reasoning together, learning becomes visible to the instructors. The growing use of computer-mediated communication channels such as social networking, chat, instant messengers and wikis as components of CSCL applications has resulted in large repositories of such learning interactions. Although CSCL tools offer advantages to eliminate the student isolation issue, such environments also result in some methodological and pedagogical challenges. For example, analyzing hundreds of lines of collaborative interactions of student teams is a time consuming task for instructors. Therefore, instructors generally focus on learning outputs while evaluating learner performance in CSCL environments. In this kind of evaluation, each team member is often assumed to equally contribute to the final deliverable, and each obtains the same grade as a result of evaluation. Yet, dividing students into groups and requiring them to collaborate do not simply result in equal participation and effective discussion. Thus, a detailed monitoring of the collaboration process is necessary to support teachers to perform a fair assessment of group work and provide support when needed (Wang, 2009).

In this study, we aimed to bring together basic ideas from text-mining and interaction analysis methods to prototype an interface that will help instructors follow the conceptual development of their students with respect to the specific learning goals of their course. For that purpose, we used tokens and phrases that signal a change in the course of the discussion as a basis for the initial segmentation of the chat data. This step provided the much-needed pre-processing to improve the representative power of the tag-cloud analysis performed in the next stage. The keywords deemed important by the instructor based on the course goals can be then used at this point to navigate through chat and wiki logs. Concordance analysis aims to help teachers identify those interactional episodes where the teams discussed the key concepts. The case study summarized above allowed us to observe how a team of students discussed key issues involved with identifying variable types and their scale of measurement as well as their distributions. Capturing such instances across multiple log files would give the teacher a much better view of the progression of ideas across multiple sessions and teams, as well as the difficulties students might be having with specific concepts and methods. In future work we are planning to automate more of the key steps in the presented process in an effort to develop a dashboard interface for instructors where they can visualize the data at different granularities and zoom in/out of collaboration logs depending on the level of analysis they deem relevant for their educational goals. For instance, linguistic markers used for initiating changes in the course of a discussion in chat can be used to mark potential segment boundaries in an automated manner. Since chat data is noisy with missing phrases and incorrect spellings, it is in general difficult to employ natural language processing techniques on chat data. However, segmentation can be improved further by considering word repetition patterns and coherence indicators for candidate segments. However, such techniques will inevitably fall short in dealing with the complexity of the meaning making processes taking place in the logs. Therefore, our main goal will be to support the teachers with practical analytics and navigational tools so that they can effectively trace the fragments of students' knowledge building discourse distributed in many components of modern CSCL systems.

References

Baker, R. S. J. d., &Yacef, K. (2009). The State of Educational Data Mining in 2009: A Review and Future Visions. *Journal of Educational Data Mining, 1*(1), 3-17.

Bienkowski, M., Feng, M., & Means, B. (2012). Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics: An Issue Brief. Retrieved from http://www.ed.gov/edblogs/technology/files/2012/03/edm-la-brief.pdf

Dyckhoff, A. L., Zielke, D., Bültmann, M., Chatti, M. A., & Schroeder, U. (2012). Design and Implementation of a Learning Analytics Toolkit for Teachers. *J of Educational Technology & Society, 15*(3), 58-76.

Ferguson, R., & Shum, S. B. (2011). *Learning analytics to identify exploratory dialogue within synchronous text chat*. In Proc. of the 1st Int. Conf. on Learning Analytics & Knowledge.

Ferguson, R., & Shum, S. B. (2012). *Social learning analytics: five approaches*. In Proc. of 2nd Int. Conf. on Learning Analytics & Knowledge, Vancouver, Canada.

Govaerts, S., Verbert, K., Duval, E., & Pardo, A. (2012). *The student activity meter for awareness and self-reflection*. Paper presented at the CHI '12 Extended Abstracts on Human Factors in Computing Systems.

Hung, J.-L., Hsu, Y.-C., & Rice, K. (2012). Integrating Data Mining in Program Evaluation of K-12 Online Education. *Journal of Educational Technology & Society, 15*(3), 27-41.

Liddo, A. D., Shum, S. B., Quinto, I., Bachler, M., &Cannavacciuolo, L. (2011). *Discourse-centric learning analytics*. Paper presented at the 1st Inter. Conf. on Learning Analytics and Knowledge.

Manouselis, N., Drachsler, H., Verbert, K., & Duval, E. (2012). *Recommender Systems for Learning*: Springer.

Manyika, J., M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, et al. (2011). Big Data: The Next Frontier for Innovation, Competition, and Productivity. McKinsey. Retrieved from http://www.mckinsey.com/Insights/MGI/Research/Technology_and_Innovation/Big_data_The_next_frontier_for_innovation

Romero, C., & Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. *Exp. Syst. Appl., 33*(1), 135-146.

Romero, C., Ventura, S., Pechenizkiy, M., & Baker, R. S. J. d. (Eds.). (2010). *Handbook of Educational Data Mining*. Boca Raton, FL: CRC Press.

Shum, S. B., & Crick, R. D. (2012). *Learning dispositions and transferable competencies: pedagogy, modelling and learning analytics.* Paper presented at the 2nd Int. Conf. on Learning Analytics & Knowledge, Vancouver, Canada.

Shum, S. B., & Ferguson, R. (2012). Social Learning Analytics. *Journal of Educational Technology & Society, 15*(3), 3-26.

Siemens, G. (2012). *Learning analytics: envisioning a research discipline and a domain of practice*. In Proc. of the 2nd Int. Conf. on Learning Analytics & Knowledge.

Siemens, G., & Baker, R. S. J. d. (2012). *Learning analytics and educational data mining: towards communication and collaboration*. Paper presented at the 2nd Inter. Conf. on Learning Analytics and Knowledge, Vancouver, Canada.

Siemens, G., &Gasevic, D. (2012). Guest Editorial - Learning and Knowledge Analytics. *Journal of Educational Technology & Society, 15*(3), 1-2.

Stamper, J., Koedinger, K., Baker, R. J. d., Skogsholm, A., Leber, B., Rankin, J., et al. (2010). PSLC DataShop: A Data Analysis Service for the Learning Science Community. In V. Aleven, J. Kay & J. Mostow (Eds.), *Intelligent Tutoring Systems* (Vol. 6095, pp. 455-455): Springer Berlin Heidelberg.

Suthers, D. D., Hoppe, H. U., Laat, M. d., & Shum, S. B. (2012). *Connecting levels and methods of analysis in networked learning communities*. In Proc. of 2nd Int. Conf. on Learning Analytics & Knowledge, Vancouver, Canada

Wang, Q. (2009). Design and evaluation of a collaborative learning environment. *Computers & Education, 53*(4), 1138-1146.