Content Analysis of Collaboratively Constructed Knowledge Artifacts: Issues and Opportunities for Research

Abstract: New social media for collaborative knowledge construction, often associated with "Web 2.0," represent an emerging context for the research of learning and instruction. Wikis, for example, allow methods of collaborative knowledge construction that were very difficult to achieve with previous technologies. Applications of wikis and related technologies (e.g., Drupal and Django) are blooming in every corner of society, influencing the ways in which people learn and exchange with one another. They are also making their way into our research, resulting in new methodological challenges concerning the analysis of collaboratively constructed materials. This symposium will discuss new methods of content analysis in wiki-environments. The symposium includes an international set of presenters from Canada, Belgium and Norway, representing a small but growing research community that are engaged in such investigations. Each paper will present the research context and method of content analysis that was developed to evaluate collaboration in a wiki environment.

Introduction to Symposium

Across the learning sciences, wikis and other forms of collaborative knowledge construction are increasingly making their way into our classroom and informal learning environments (e.g. Harrer, Moskaliuk, Kimmerle, & Cress, 2008; Forte & Bruckman, 2007; Greenhow, Robelia, & Hughes, 2009). Wikis appeal to educators and researchers alike for several reasons: they enable students to contribute, edit, and maintain a shared document, they are relatively easy to learn, and the revision history is preserved for future reference and analyses. However, despite the growing popularity of wikis, little is known about the processes that students engage in when contributing to a wiki. What kinds of practices do they adopt when writing a wiki page? How do they interact with the content –their own and that of their peers – and how do those interactions change over time? Can we measure student learning in such task? How can we evaluate the effectiveness, or the specific dynamics of collaboration?

As observed by Jeremy Roschelle in his closing comments at the 2009 CSCL conference, the learning sciences still lack standard methods for analyzing the interactions that occur between students in a collaborative learning task. Challenges to analyzing collaborative processes include choosing an appropriate unit of analysis, segmentation, and developing a coding scheme that captures the learning goals of interest to the research (Strijbos, Martens, Prins, & Jochems, 2006). In-depth analyses of collaborative materials can also be constrained by the design of the technology environment and the kinds of analytic tools that are available to researchers. The idiosyncrasies of such research – the particular technology environments, collaborative designs, and research foci – have made it difficult to design analytic methods that can be shared or normed across the research community. A survey of the existing literature reveals little, if any, published approaches for how to analyze the conceptual content or collaborative progress within a wiki-based activity.

In the absence of established analytic methods within the academic community, researchers must design their own metrics for analyzing wiki-based collaborative knowledge construction. This symposium was conceived as a way of widening this discussion within the forum of the ICLS meeting. We have gathered four distinct research groups – two from Canada and two from Europe – who have all confronted the same challenge. We invite each of these groups to present their research paradigm as well as a specific methodological innovation or approach. Our discussant, Christine Greenhow, who is not affiliated with any of these research programs, will synthesize the ideas and advances in this work and lead a discussion amongst participants.

Paper 1: Development of a Content Analysis Approach for Collaboration in a Wiki Environment

Bram De Wever, Hilde Van Keer, Department of Educational Studies, Ghent University, Belgium
Email: bram.dewever@ugent.be, hilde.vankeer@ugent.be

This paper focuses on developing a content analysis of collaboration in a wiki environment. This contribution focuses on developing a content analysis approach studying collaboration in a wiki environment. It first discusses a number of units of analysis with different level of granularity. Next, a framework for content analysis is presented, focusing on one level and distinguishing between content and presentation matter.

In order to find out what is actually happening during the creation of a wiki, we can focus on different units of analysis. One unit of analysis is the complete wiki, i.e. the complete collection of wiki pages created by students. Observations at this level can focus on how the development takes places over time: What is the first page created? What are the following pages? What is the order of page creation? Information on the priorities and the need to change pages can be revealed and linked to explicit or implicit planning issues.

Another unit of analysis is the individual wiki page. By taking a detailed look at the editing history of each page, relevant information can be retrieved: How many times is the page edited? What were the major changes in the different edits? How many people worked together on this specific page? Who edited the pages at what time? What was the pacing of the edits? Were the edits labeled as "minor edits"? However, not all the information on the page level is to be found in the page history logs. Much relevant information can be gathered by analyzing the content of the wiki page's final version. Analysis might consider how many links to other pages are included, how many external links are included, and how well the page addresses the given assignment. One might also analyze the formatting or layout of the page, as well as what types of media are included. Moreover, detailed information can be found by comparing the consecutive versions of a certain wiki page. Was information added, deleted, or moved from one version to the next? Which information was subject to change? Were specific types of media (pictures, audio, video) added (or deleted)?

In the description above, every version of the wiki page could be seen as a unit of analysis. However, since every version of the page could comprise multiple elements that are changed, some authors (e.g., Peters and Slotta, this symposium) opt to use every changed element as a unit of analysis, which they call transactions.

All of the previously discussed units of analysis can be hierarchically nested: All transactions are nested within one version. Every version is then again nested within a page. And every page is in turn nested within the overall wiki (see Figure 1 – upper portion). However, another unit of analysis, namely the user (student) does not completely fit with this hierarchically nested structure. Although transactions and versions can be hierarchically nested within users (every transaction is nested within one and only one version, and every version is edited by one and only one user), users themselves are cross-indexed with respect to wiki pages. That is, users can edit multiple pages (see Figure 1 – lower portion).

The user is an important unit of analysis when it comes to analyzing collaboration within a wiki. At this level, the following questions can be addressed: How many edits were made by a specific user? How many pages were edited by a specific user – or how many pages were edited on average by the users? How often did a user edit the wiki? Most of this information cannot be found in a log file, but must be constructed from one of the hierarchical units identified above. For instance, how many minor/major edits were done by a certain user? How many additions, figures, etc. has a user added? By combining information from the different levels, the full picture of a user's participation can be developed. Indeed, a number of interesting questions can only be answered by combining information. Moreover, users can be clustered into different groups or types, depending on their condition within the research study or their activities within the wiki-environment.

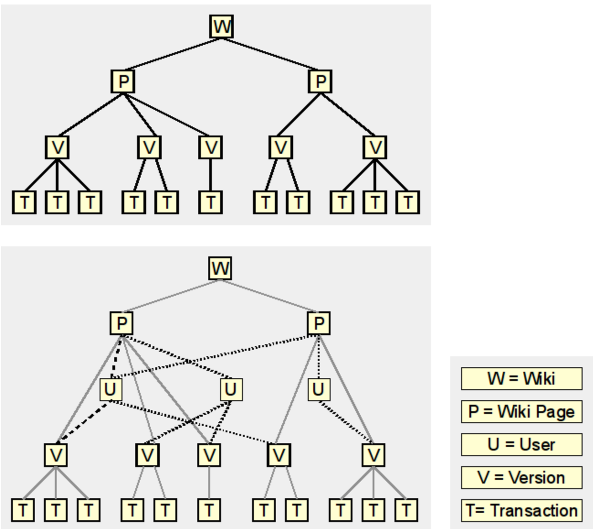


Figure 1. Four hierarchical levels within a wiki (upper figure) become more complex when a layer is added for users (lower figure).

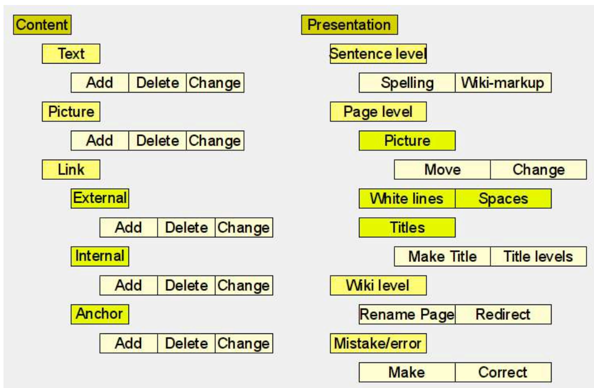


Figure 2. Coding schemes for wiki edits, including specific codes for content (left) and presentation or formatting of pages (right).

We developed a general framework for performing content analysis on collaboratively developed wiki materials. To begin, we focused on the single wiki page as our unit of analysis, which provides a basis for discussing the deeper levels framework (i.e., at the transaction level). In the proposed symposium, we will compare our approach with the methods of analyzing wiki data presented by the other participants.

Figure 2 above presents an overview of the coding scheme for each wiki page, for every version of the page. Since a version can contain different changes ("transactions"), multiple codes can be assigned to each unit of analysis. Content codes include text, pictures, and links that can be added, deleted, or changed. For links, we differentiate between external links (linking to other references on the WWW), internal links (linking to other pages within the wiki), and links to anchors (linking within the same wiki page). Other forms of media could be added to this scheme, e.g. videos, if necessary. Whereas the Add and Delete categories are straightforward, the Change category is not. Technically speaking, changing any text involves deleting some parts or adding others. Therefore, one could argue that a Change category is not needed. However, by including a specific "change" code, we focus on the fact that an existing idea is edited, which is not exactly the same as deleting an idea and adding one, where there is not necessarily a link between those two actions. Thus, when a part of a page is deleted (say in the introduction), or another part added (say in the conclusion of the page), such edits should be coded as Delete or Add, respectively. When one paragraph is rephrased, this should be coded as Change.

With regard to presentation or formatting of the wiki page, we differentiate between changes at the sentence or paragraph level, at the page level, and at more global wiki level. In addition, we include a code for revisions that involve simply correcting mistakes (e.g., typos). It is important to recognize the differences between content and presentation, and to note that some actions may be coded in both taxonomies. For instance, if somebody adds a title, this should be coded both as Text/Add in the Content scheme and Page level/Titles/Make title in the Presentation scheme, whereas simply changing some existing text to make it into a header or title would only result in the Presentation code, since the text was there beforehand.

This method of coding can permit a wide range of analyses, according to the specific needs of the research design. While we are in the beginning stages of utilizing such techniques in our own research, we are confident that it can be applied reliably, and that it can be further developed for other units of the wiki organizational scheme (Figure 1). Additionally, qualitative measures concerning the semantic content of the wiki pages themselves can be added. Thus, our present coding scheme could serve as at a first stage of analysis, to identify the versions and collaborative patterns in which content is edited (added, deleted, change). In a second stage, these versions can be studied in a more detailed way. This symposium will allow the authors to discuss such procedures with co-participants as well as with the audience.

Paper 2: Analyzing Student Collaborations in a Wiki-based Science Curriculum

Vanessa L. Peters, James D. Slotta, Ontario Institute for Studies in Education, University of Toronto, Canada
Email: vlpeters@gmail.com, jslotta@gmail.com

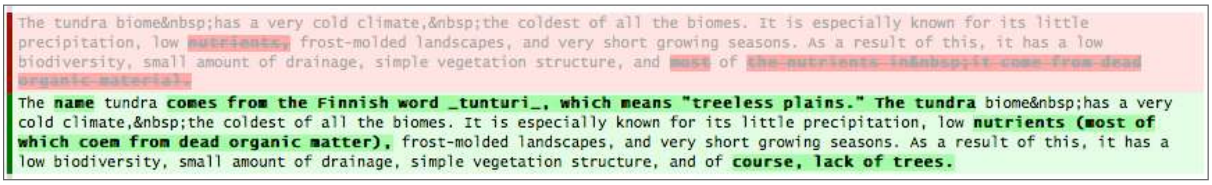
As part of a larger research study (see Peters & Slotta, 2010), we designed a collaborative wiki-based activity that engaged secondary students in co-creating a community knowledge base to serve as a resource for a subsequent inquiry activity. Over eight weeks, 112 students from four sections of a high school biology course contributed to a common wiki repository about Canadian biodiversity. A wiki provided the ideal functionality for this activity as it enabled students to easily access and edit each other's contributions, reorganize the repository, and link pages to establish connections between related themes or ideas.

While it was important to preserve the open-ended feeling of collaborative editing that typifies wikis, it was equally important to have a simple and structured way for students to create wiki pages that addressed specific science content. We therefore created a hybrid wiki environment that included a customized "Create a New Page" web form (developed using Ruby on Rails) that collects basic information about their resource page, then generates a new wiki page with pre-specified headers to help scaffold the content in students' wiki entries.

Analysis and Findings

The activity was evaluated in terms of collaborative editing, usefulness of wiki materials in later inquiry activities, and student learning outcomes. Students were actively engaged in the construction of knowledge resource within the wiki, creating 31ecozone pages with a mean word count of 2809.13 (SD = 1408.73). Each wiki page was run through Copyscape©, a web-based utility that compares web-based material to check for instances of plagiarism, with negligible results.

To analyze students' knowledge construction practices, we coded their contributions to the wiki as individual "transactions," defined as the distinct changes to a wiki page that occur during authoring. A transaction could be as simple correcting a spelling error, or it could be the revision of a paragraph that involves adding, deleting and moving text. It must be noted that, while wiki systems generally track all edits according to "versions" (i.e., where each version corresponds to one editing session), any editing session might contain multiple transactions. Thus, coding for transactions meant looking within each version edit, to see whether the author conducted multiple edits before saving the page. Transaction codes included: *move*, *add, delete* and *format*. For each of the first three codes, we then coded the object of the action: *text*, *image*, *external link*, or *internal link*. In the case of text actions, we coded whether the resulting text was *embedded* or *detached* (see Figure 1). Finally, we coded whether the text was made to one's own text, or that of a peer. This distinction was important because collaborative knowledge construction entails building on and extending the ideas of others within the learning community.



The tundra biome has a very cold climate, the coldest of all the biomes. It is especially known for its little precipitation, low ~~nutrients,~~ frost-molded landscapes, and very short growing seasons. As a result of this, it has a low biodiversity, small amount of drainage, simple vegetation structure, and ~~most~~ of ~~the nutrients in it come from dead organic material.~~

The **name** tundra **comes from the Finnish word _tunturi_, which means "treeless plains." The tundra** biome has a very cold climate, the coldest of all the biomes. It is especially known for its little precipitation, low **nutrients (most of which coem from dead organic matter),** frost-molded landscapes, and very short growing seasons. As a result of this, it has a low biodiversity, small amount of drainage, simple vegetation structure, and of **course, lack of trees.**

Figure 1. Wiki transaction showing "embedded" text that has been added and deleted.

Using the transaction as our unit of analysis, we coded the complete revision history of the wiki (i.e., the community knowledge base) to identify the kinds of contributions that students made, when they made them, and the relative frequency of contributions by different students. This coding allowed us to measure which students made substantive contributions at what points in the curriculum, whether participation was more or less equitable and whether ideas grew iteratively or simply as a piecemeal collection.

As a result of our analysis we identified several trends in terms of students' editing practices when contributing to a wiki. We also found a relationship between students' individual editing practices and their scores on their final unit exam. This presentation will provide an overview of the research and describe the methodology that was used in the study. We will detail the coding, analysis of collaboration and growth patterns within the wiki, and conclude with recommendations for educators wishing to implement a wiki technology in their classrooms.

Paper 3: Does Scale Matter: Using Different Lenses to Understand Collaborative Knowledge Building

Elizabeth S. Charles, Dawson College, Canada, echarles@dawsoncollege.qc.ca
Nathaniel Lasry, John Abbott College, Canada, lasry@johnabbott.qc.ca
Chris Whittaker, Dawson College, Canada, cwhittaker@dawsoncollege.qc.ca

Web-based environments for communicating, networking and sharing information, often referred to collectively as "Web 2.0," have become ubiquitous – e.g., Wikipedia, Facebook, Flickr, or YouTube. Understanding how such technologies can promote participation, collaboration and co-construction of knowledge, and how such affordances could be used for educational purposes has become a focus of research in the Learning Science and CSCL communities (e.g., Dohn, 2009; Greenhow et al., 2009). One important mechanism is self-organization, which includes the regulation of feedback loops and the flows of information and resources within an activity system (Holland, 1996). But the study of such mechanisms calls for new ways of thinking about the unit of analysis, and the development of analytic tools that allow us to move back and forth through levels of activity systems that are designed to promote learning. Here, we propose that content analysis can focus on the flows of resources (i.e., content knowledge, scientific artifacts, epistemic beliefs) in terms of how they are established and the factors affecting whether they are taken up by members of the community.

We present our analytic approach, and some empirical data that tests out its constraints. Our approach is informed by a "systems thinking" perspective and relies on aspects of Activity Theory (Engeström, 1999) as well as distributed cognition (e.g., Salomon, 1993, Hutchins, 1995). In doing so, we attempt to look across several levels of granularity within an activity system. To start we look at the smallest unit of sense making between individuals and investigate how their contributions lead to the construction of an artifact, which is a form of interactional analysis (e.g., Zemel & Cakir, 2009). On another level, we look at how that artifact becomes a shared object, whether it is used by others in the community, and how it becomes culturally imbued (typical of socio-cultural analysis). This means that we systematically interweave the use of two units of analysis: (1) micro-scale, and (2) macro-scale. Our interest here is to determine whether the insights made by one or both levels reveal substantially different things about the system and its actors. As well, we shed light on co-regulatory mechanisms that may be at play, such as collective cognitive responsibility (Scardamalia, 2002).

We bridge these levels of granularity by producing a visualization of the discourse contributions, expanding upon earlier work by the first author (Karin, Charles & Kolodner, 2006). We demonstrate that such visualizations can facilitate the investigation of how collective states of knowledge and beliefs emerge within an activity system such as a collaboratively edited wiki.

The data we analyze is a case study of an online activity involving a class of 27 first year science students enrolled in an introductory physics course, with topics such as Newton's laws, and the basic principles of forces and motion. These data are drawn from a larger study that investigated the impact of different instructional approaches to collaborative learning, across 4 classes, at a post-secondary institution in the province of Quebec The province of Quebec, Canada, has a unique 2-year pre-university college system, equivalent to grades 12 and first-year university in other parts of the world. However, the science curriculum in our program is taught roughly at the level of an introductory university courses.

We selected this instructional case because it exemplifies collaborative knowledge building and refinement from a productive online discussion forum, which led in turn to a class wiki. This development was intriguing because it emerged from an activity system with few initial resources (i.e., weakly scaffolded) and open-ended participation goals. The teacher's instructions were simply to explore two questions following a physics lab on the topic of acceleration and its effect on weight. Participation was encouraged and monitored, though no grades were assigned. In other words, we anticipated that co-regulation would be a large factor in the success of this activity system.

We started on the micro-scale, analyzing the interactions at the level of the threaded postings. These are the tightly coupled (or chained) postings that respond to an initial idea (see Figure 4, below). Along the way to being submitted to the wiki, the idea is transformed. Sometimes it was a simple clarification (B, E, F, H). Sometimes it took on new meaning, which we consider to be a cultural artifact (C, G, I, J).

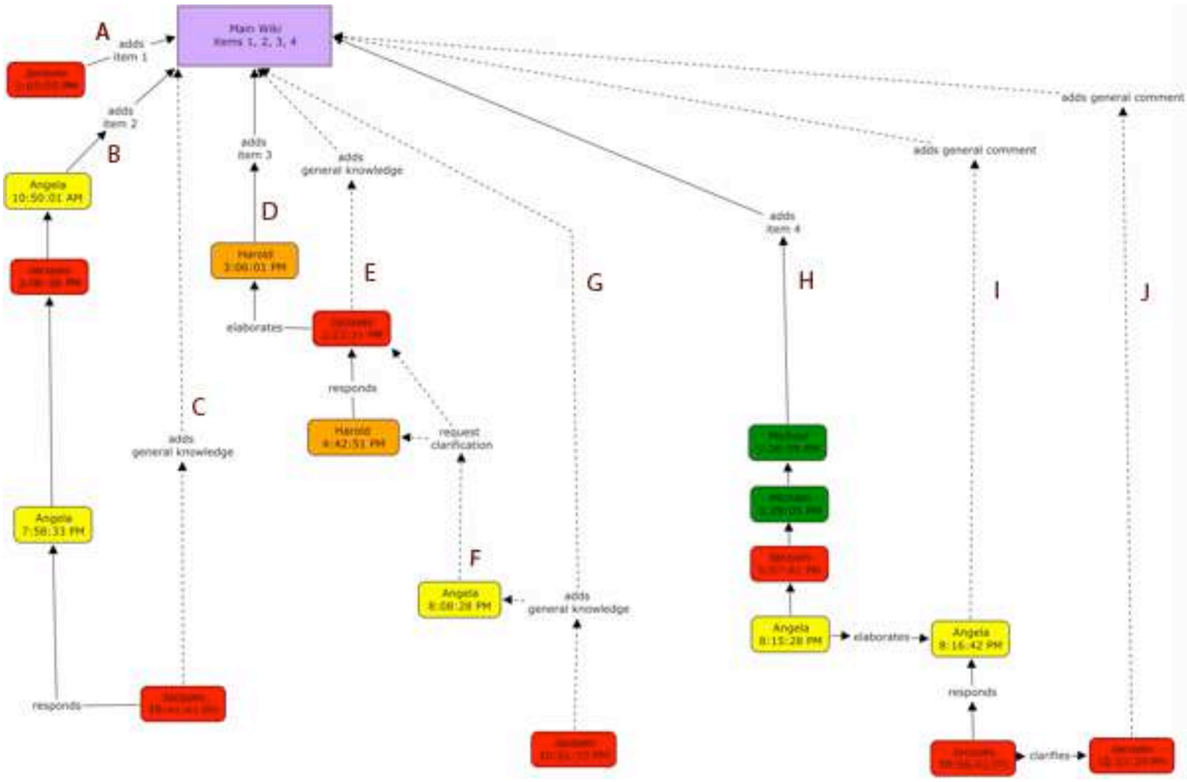


Figure 4. Micro scale representation of contributions to class wiki, deriving from previous discussions.

When there were no new artifacts being produced, we switched to the macro-scale unit of analysis (see Figure 5, below). Though not threaded, new postings instead referred to identified artifacts: (1) content knowledge (the lilac colored boxes at right) and (2) epistemic knowledge (the blue boxes on the left). Both ideas are taken up by the community and constitute their collective knowledge. This visualization thus represents these two kinds of artifacts and their flow through the activity system.

Finally, we also recorded how several of the actors continued to regulate the flow of these resources, in particular two students, Jacques and Angela (see both Figures). When viewed from the macro-scale, such repeated actions can be seen as the emergence of practice. When such practice is tightly coupled to the improvement of a shared artifact, we argue that it demonstrates the construction of a shared goal and propose this may exemplify what Scardamalia (2002) refers to as collective cognitive responsibility.

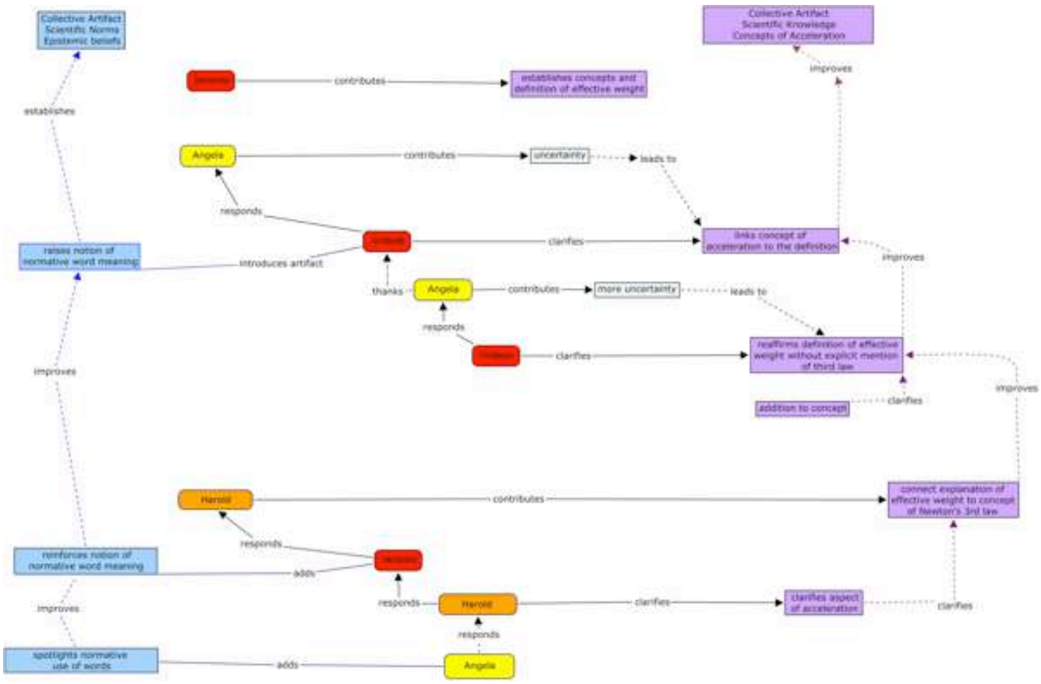


Figure 5. Macro level representation of the flow of content and epistemic knowledge in the activity system.

Paper 4: Learning Through Collaborative Creation of Shared Knowledge Objects: Technological Support and Analytic Challenges

Crina Damşa, Intermedia, University of Oslo, Norway, crina.damsa@intermedia.uio.no,
Patrick Sins, Utrecht University, the Netherlands, p.h.m.sins@uu.nl
Bert Reijnen, Stoas University of Applied Sciences, The Netherlands, b.reijnen@stoashogeschool.nl

Recent studies of collaboration demonstrate that simply bringing people together in groups with some task, or pooling the group's knowledge, are insufficient conditions to lead to productive real collaboration (Barron, 2003; Salas, Shawn, & Burke, 2005). One recent approach to learning called Knowledge Creation (Paavola & Hakkarainen, 2005) depicts learning as a collaborative activity aimed at creating new knowledge mediated through the creation and development of shared knowledge objects (e.g., research reports, instructional material or scientific models). According to this approach, collaborative learning does not serve only individual learning or only social interaction. Rather, collaboration is seen as organized around common knowledge objects whose creation and development defines their purpose.

To design educational contexts that apply this idea involves designing new pedagogical methods and technological tools that are grounded in the knowledge creation approach to learning. One of the prominent approaches to such innovative designs is technological support of object-mediated collaboration, in the form of virtual collaboration spaces. However, such designs are often undermined by the fact that little is known about how to analyze processes of object-mediated collaboration and learning. Consequently, our study seeks to understand how learners create knowledge in a virtual environment that supports object-mediated collaborative learning. We describe an analytical framework that supports an integrative approach to analyzing how interactions between learners in the environment contribute to the progression of their knowledge objects.

The study was conducted in a collegiate institution in the Netherlands that prepares pre-service teachers for lower secondary education in the field of agricultural and animal studies. The current study involves 15 students in one of the Professional Situations a 5-month curriculum unit. Students worked in groups of 4-5 on long-term projects constructing shared knowledge objects and employing the Knowledge Practices Environment (KPE) to support their work. KPE is a virtual collaboration space offering facilities for collaboration, with peers and working with knowledge objects in shared work spaces. Each shared space encompasses a workplace which presents users with three views: a Process view (to support planning and organizing the process), a Content item view (to allow creating, sharing and collaboratively editing of documents, wiki pages, or notes) and a Community view (to enable management of the community). Each group of students worked in their own shared space, with the emphasis on using wiki or document versioning for collaborative text production and object-bound chat for supporting focused discussions around the produced material (see Figure 6).

Figure 6. Object development in wiki pages and object versions

The integrative approach described above involves combining the analysis of the interactions involved in the creation of objects with the analysis of the object up-take and progress of the objects (see also contribution by Charles, Lasry, & Whittaker, above). Based on an analysis framework for verbal and written interactions previously constructed by two of the authors (Damşa, Kirschner, Andriessen, Erkens, & Sins, in press) we distinguish between epistemic and regulative actions. Epistemic actions are those that bring about the construction and progress of the shared knowledge-object. These are actions involved in or resulting from productive interactions. They can be described qualitatively in such terms as "creating awareness," "explicating vision," "alleviating lack of knowledge," "elaborations," a" negotiations to create explanations and shared understanding" or "generative collaborative actions and problem solving." Regulative actions are concerned primarily with the processes necessary to direct, organize and support the collaborative process, and include categories of projective, coordinating and relational actions.

With respect to analyzing students' knowledge objects, two analytic perspectives are distinguished. First, the normative perspective, which involves determining the quality of the final object according to domain-specific criteria, established together with the teachers using an evaluation form. Second, the content-evolution perspective, which analyzes the progression of the contents of the constructed objects. Inspired by Suthers (2006) this analysis focuses on the epistemic content that is brought in by group members (i.e., ideas) and that evolves in terms of conceptual complexity during the period in which the object is developed. Another category of this analysis concerns the way this content was shaped, elaborated, synthesized, selected, and revised over time. These analysis categories were drawn mainly from the literature on text writing and elaboration.

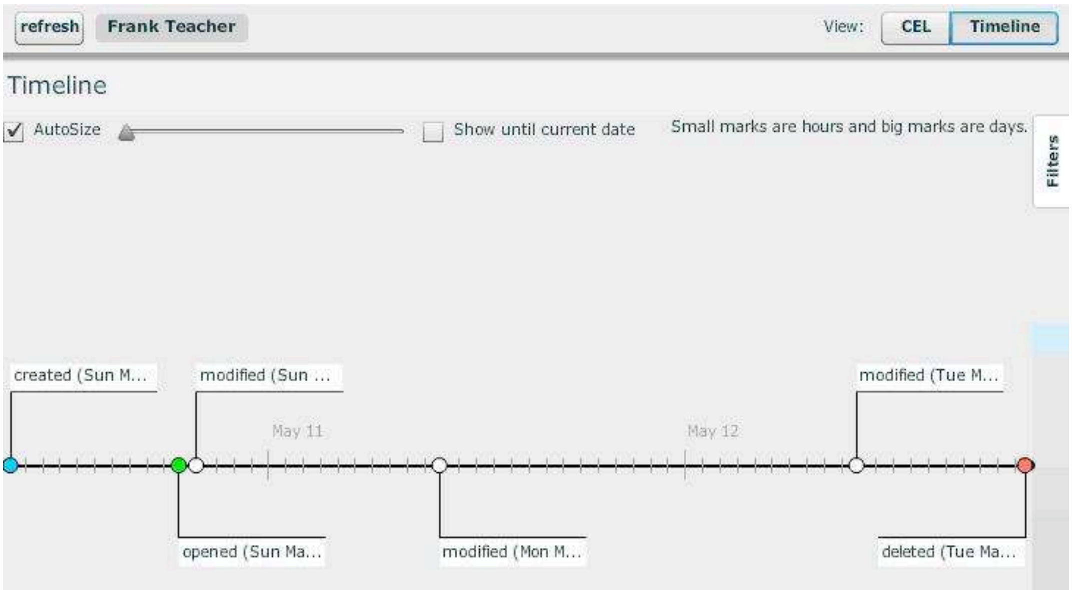


Figure 7. The object-based timeline analysis of an artefact evolution

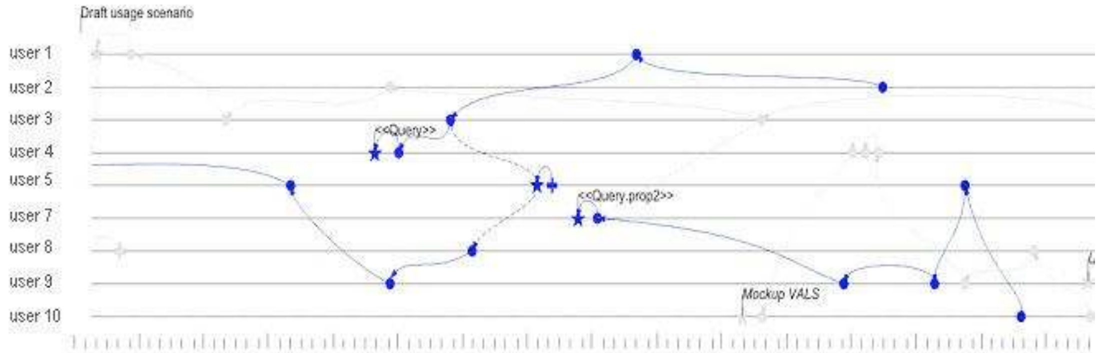


Figure 8. The user-based timeline analysis of an artefact evolution

Our analysis tool makes it possible to trace changes in the knowledge objects, by providing visualization of performed activities related to a selected object (or set of objects) or for a particular user (or group) on a timeline. (see Figure 7 and 8, above). To support the analysis of students' interactions feeding into their shared knowledge objects, the system connects the chat discussions that took place between students related to specific objects (e.g., document or versions of a wiki), serving to unveil the way in which verbal actions are taken-up and how ideas expressed are materialized in the knowledge objects.

References

Barron, B. (2003). When smart groups fail. The Journal of the Learning Sciences, 12, 307-359.

Bereiter, C., & Scardamalia, M. (1989). Intentional learning as a goal of instruction. In L. B. Resnick (Ed.) *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 361-392). Hillsdale, LEA.

Damşa, C. I., Kirschner, P. A, Andriessen, J. E. B., Erkens, G., Sins, P. H. M. (in press). Shared epistemic agency – An empirical study of an emergent concept. Journal of the Learning Sciences.

Dohn, N.B. (2009) Web 2.0: Inherent tensions and evident challenges for education. International Journal of Computer-Supported Collaborative Learning, 4(3), 1556-1607.

Engeström, Y. (1999). Activity theory and individual and social transformation. In Y.Engeström, R. Miettinen, & R. Punamaki, (Eds.), Perspectives on activity theory (pp. 19–38). Cambridge, MA: Cambridge University Press.

Greenhow, C., Robelia, B., & Hughes, J. E. (2009). Web 2.0 and classroom research: What path should we take now? Educational Researcher 38(4), 246-259.

Harrer, A., Moskaliuk, J., Kimmerle, J., & Cress, U. (2008). Visualizing wiki-supported knowledge building: Co-evolution of individual and collective knowledge. Proceedings of WikiSYm '08, Porto, Portugal.

Holland, J.H. (1995). Hidden order: How adaptation builds complexity. Reading, MA: Addisoon-Wesley Publishing Co.

Hutchins, E. (1995) Cognition in the wild. Cambridge, MA: MIT Press.

Karkin, S., Charles, E.S., & Kolodner. (2006). Visualizing Discussion by the Use of the Conversation Chain Model. Poster in proceedings of the 7th International Conference of the Learning Sciences: Bloomington, IN.

Paavola, S. & Hakkarainen, K.: The Knowledge Creation Metaphor – An Emergent Epistemological Approach to Learning. Science & Education, 14(6), 535--557 (2005)

Peters, V. L., & Slotta, J. D. (2010). Scaffolding knowledge communities in the classroom: New opportunities in the Web 2.0 era. In M. J. Jacobson & P. Reimann (Eds.), *Designs for learning environments of the future: International perspectives from the learning sciences* (p. 205-232). Secaucus, NJ: Springer.

Salas, E., Sims, S., & Burke, C. (2005). Is there a "Big Five" in teamwork? Small Group Research, 36, 555-599.

Salomon, G. (1993). Distributed cognitions. New York: Cambridge University Press.

Scardamalia, M. (2002). Collective cognitive responsibility for the advancement of knowledge. In B. Smith (Ed.), Liberal education in a knowledge society (pp. 67-98). Chicago: Open Court.

Strijbos, J., Martens, R., Prins, F., & Jochems, W. (2006). Content analysis: What are they talking about? Computers & Education, 46, 29-48.

Suthers, D. D. (2006). A qualitative analysis of collaborative knowledge construction through shared representations. Research and Practice in Technology Enhanced Learning, 1(2), 1-28.

Zemel, A. & Çakir, M. (2009). Student and Team Agency in VMT. In G. G. Stahl (Ed.), Studying Virtual Math Teams, (pp.). New York, NY: Springer Publishing.