

A Prototype Facilitators Dashboard: Assessing and visualizing dialogue quality in online deliberation for education and work

Tom Murray, Leah Wing¹, Beverly Park Woolf,
Alexander Wise, Shijun Wu, Lori Clark, Lee Osterweil, Xiaoxi Xu
School of Computer Science, ¹Legal Studies Dept.
University of Massachusetts, Amherst, MA
Contact: tmurray@cs.umass.edu

Abstract: The emerging next generation ("Web 3.0") of socio-technological tool development is adding additional support for reflecting on and improving the *quality* of online information, communication, and action coordination. An important opportunity is that online systems can include tools that directly support participants in having higher quality and more skillful engagements. We are evaluating dialogue software features that support participants directly and "dashboard" tools that support third parties (mediators, teachers, facilitators, moderators, etc.) in supporting higher quality deliberation. In this paper we will focus on our work in educational settings (college classes) and on our development of a Facilitators Dashboard that visualizes dialogue quality indicators for use as facilitation tools or participant social awareness tools. We are particularly interested in supporting the "social deliberative skills" that interlocutors need to build mutual understanding and mutual regard in complex or contentious situations.

Keywords: Educational and Knowledge Building dialogue; deliberative skills; scaffolding; multiple representations; dashboards.

Introduction

As communication, collaboration, and knowledge building expand on the Internet the benefits and limitations of "Web 2.0" technologies become increasingly apparent [1]. The affordances of social networking, information sharing, and expansive search capabilities have lead to a dramatic increases in the *quantity* of information and connectivity without always supporting—and sometimes sacrificing—their *quality*. The emerging next generation ("Web 3.0") of socio-technological tool development adds additional support for reflecting on and improving the quality of online information, communication, and action coordination [2]. An important opportunity is that online systems can include tools that directly support participants in having higher quality and more skillful engagements. We are building and evaluating dialogue software features that support *participants* directly and "dashboard" tools that support *third parties* (mediators, teachers, facilitators, moderators, etc.) in supporting higher quality deliberation among participants. In this paper we will focus on our work in educational settings (college classes) and on our development of a Facilitators Dashboard that visualizes dialogue quality indicators for use by either third parties or participants. We are particularly interested in supporting the "social deliberative skills" that interlocutors need to build mutual understanding and mutual regard in complex or contentious situations [3][4].

Iandoli et al. [5] give examples of tools that provide visualizations and feedback about: the who, to-whom, when, and how-much (activity level) of conversation moves; show the structure and time-based evolution of topics and communication relationships; and provide meaning-making tools for relevance, importance, and summarization. They suggest that such tools "help people communicate in better and easier ways by reducing misunderstandings, facilitating the grounding process, and reducing its costs" (p. 73). Such tools not only compensate for some of what is lost in moving from face-to-face communication but can also provide novel means of visualizing, reflecting on, and improving collaborative processes that are not available in face-to-face communication. Bunder et al. [6] frame the issue in terms of "social and cognitive awareness tools" that "facilitate and institutionalize the natural processes of becoming aware about social and cognitive variables, thereby leading to adaptive behavior in collaboration" (p. 606). Third parties such as facilitators sometimes provide this type of structured support.

Communication, collaboration, and knowledge building have many facets, and we focus our research on a specific area: supporting the social deliberative skills and behaviors that allow interlocutors to build mutual understanding (or "negotiate meaning") in complex or contentious contexts. That is, when one is challenged to deeply consider and work with not just *an* idea or need, but *your* or *their* idea or need

in relationship to *mine* or *ours*. These skills include social perspective taking, question asking, self-reflection, and meta-dialogue (see [7] [3]). Iandoli et al. describe a "debate dashboard" to enhance online knowledge sharing. Our focus on mutual understanding and meaning negotiation is different from, but overlapping with, those who focus on supporting debate or argumentation quality, problem solving, or knowledge building activities (including Iandoli et al., and also see [9]). Similar to Iandoli et al., we wish to provide visual feedback or mirroring tools about (1) individual users and aggregations of users; (2) the interaction process, and (3) the content of a dialogue.

Below we will describe the Dashboard tool, the discussion forum software that it interacts with, our initial formative evaluations of the Dashboard, and future plans. This research is part of a larger project that includes (1) testing whether special discussion forum features ("reflective tools") support social deliberative skills, and (2) using state of the art text analysis and machine learning algorithms to analyze social deliberative behavior in online communication. In Murray et al. [3] we report on experimental trials that show a significant impact and large effect size for the reflective tools in college classroom online dialogues. For every classroom or large group that uses our tools, there are only one or two facilitators, so our evaluations of the Dashboard will be more qualitative and case-based.

Recent advances in computational psycholinguistics allow for a more systematic and deeper analysis of dialogues, which is necessary to uncover subtle cues that might be diagnostic of critical deliberation characteristics. In [10] [11], we report on our work in developing computational methods to measure deliberative skills from online discussions, which have shown promising results. In this paper we will describe our progress and plans for displaying the results of such text analysis in the Dashboard.

Mediem Deep Deliberation Software

Before we describe the Dashboard software we will describe the enhanced discussion forum software that it interacts with. Our research indicates that simple scaffolding features can increase skillful deliberation online. In [3], we compared students using the reflective tools in Mediem with a control group not using these tools and found a large effect in social deliberative skill support due to the reflective tools. Mediem is an off-the-shelf application (developed by Idealogue Inc.¹) that has been used (by others) in a number of dialogue contexts including interfaith discussions among college students.

The Mediem software was chosen for our study because it has a number of features designed to support deeper reflection and engagement. Mediem includes the three reflective tools. First is the Story feature, which gives participants a special place to say how the issue at hand relates to them personally, including relevant background information about themselves and "what is at stake" for them in the issue. Second is the Conversation Thermometer, a meta-dialogue tool that allows participants to rate (vote on) the quality of the conversation at any time. The choices can be customized by the administrator. Third is the Contribution Tag feature, which allows participants to give brief comments on other's contributions. It provides a fixed vocabulary similar to the sentence starters (or locution openers) used in other dialogue software, but the tags remain attached to the target post rather than starting a new post (see [12]).

Facilitators Dashboard Diagram Pane: Visualizing key indicators

Design considerations. We have prototyped a Facilitators Dashboard that provides third parties a "birds-eye view" of the state and flow of online engagements. We have piloted it with professional facilitators and also begun to pilot it as a feedback and "awareness tool" for participants. We built an API that allows the Dashboard to receive real-time updates on the dialogue state and posted text from the Mediem deep dialogue forum system. See Figure 1, which shows tools in the "Diagram" tab of the Dashboard. Similar to Iandoli et al., we aimed to visualize user, interaction, and content information, including participation levels, reply networks, and content or theme overviews—in both static and trend (timeline) visualizations. At a more ambitious level, we also aimed to use text and network analysis to identify skillful (or non-skillful) deliberation, emotional tone or sentiment. Further, we have made early forays into automatically identify dialogue phases (e.g. introductions, deliberation, impasses, persuasion) and turning/infection points or opportunities for intervention (e.g. silences or non-responsiveness, changes of phase or tone, sudden emotional tensions in multiple participants).

Unlike projects that help participants think logically or creatively, provide valid justifications, or design reasonable solutions (all of which are certainly important), and given our focus on social

¹ For more information about Mediem, which is Open Source software, see www.idealogueinc.com.

deliberative skills and meaning negotiation, we assumed that our facilitators would be most interested in supporting all voices being heard, participants acting respectfully, and encouraging reciprocal role taking, empathy, and self-reflection. Based on an informal analysis of the literature and our conversations with experts, Shrikant & Murray [13] identified a set of common problems encountered in online facilitation that facilitation tools might help monitor:

- Low or no participation of individuals or groups, or **silences** or lulls on the part of individuals, the entire group, or sub-groups (which can be due to disinterest/boredom in the discussion or discomfort/fear of speaking up);
- conversation **domination** by an individual or group;
- inappropriate or **disrespectful** behavior;
- **off-topic** conversation;
- tension-filled **disagreements**, or high emotional content;
- too much agreement or **politeness**;

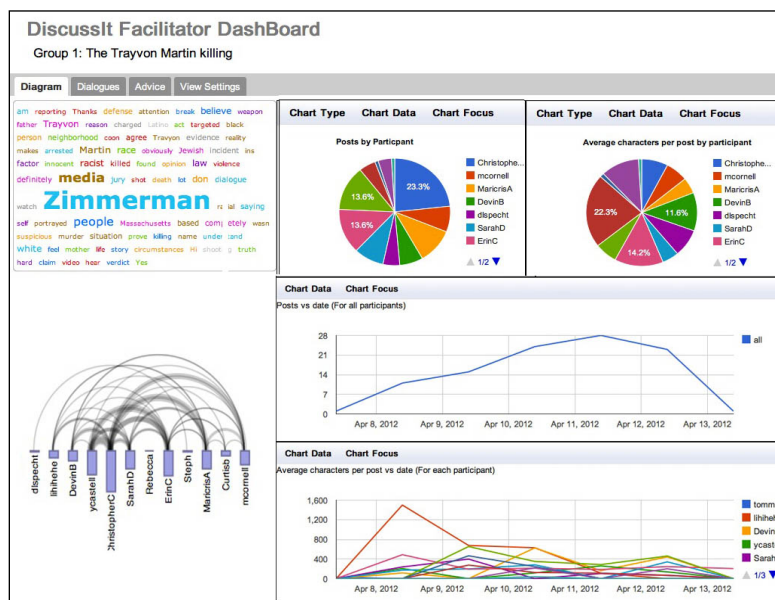


Figure 1: Facilitator Dashboard: Diagram Pane

We add to this list indicators of deliberative skillfulness that we would be interested in identifying and monitoring automatically (if possible):² social question asking and perspective-seeking behaviors; appreciation; social perspective taking; meta-dialogue (reflecting on the quality of the dialogue); proposing and weighing alternatives; systems or big picture thinking; peer mediation and facilitation; question answering and responsiveness; and topic initiation. The Diagram panel of the Dashboard shown in Figure 1 shows direct data displays that involve no analysis. Later we will discuss the Dialogue and Advice panels, which involve more in-depth processing. As the reader will note in the descriptions below, we have only begun to address all of these issues. But we feel that presenting the full list is a helpful indicator for where the field needs to go, and a map for our continued research.

Dashboard description: Diagram pane. Figure 1 shows data from a classroom discussion about the fatal shooting of Trayvon Martin by George Zimmerman which was a hot topic in the news during the time of this activity. When the facilitator begins using the Dashboard they select from a list of the deliberation projects, classes, or discussion groups registered with the Mediem software and the Dashboard (not shown in the Figure). Pie and bar charts show participation levels (number of participant posts and

² Note that in our research to measure the impact of reflective tools, we have a manual coding scheme for these items. In our text analysis and machine learning research we are also trying to create models to automatically recognize them.

average size of posts). Timelines show trends in these same metrics. A social network diagram shows who is replying to whom, with the thickness of the lines proportional to the number of replies. . A Word Cloud (see en.wikipedia.org/wiki/Word_cloud) graphically shows word frequencies through font sizes (the color and location of the words has no meaning in this representation).

When participants register on the Mediem site they have the option of entering demographic information, the structure of which is determined by the forum administrator. For the classroom trials we include gender and school grade level, but religious or political affiliation or other demographic information could also be captured and visualized. Figure 1 shows pie charts by individual. Figure 1 shows a view by gender. In this figure females are posting more often but males are posting longer posts. For the Trayvon Martin discussion shown in Figure 1, the class instructor was able to see that upper class students were more engaged and lowerclass students were more silent, which she found to be useful information.

At first glance the word cloud for our classroom discussion in Figure 3 may not seem particularly informative. It shows that in the discussion about the Trayvon Martin killing the students were, unsurprisingly, talking about the shooter, Zimmerman, quite a bit. In fact this pattern was common in all of the discussions we looked at—the words that stood out most were directly related to the topic at hand. One feature that we plan to add is the ability to add additional "stop words" that the Word Cloud will ignore (it already has standard stop words in the software). In this example, if Zimmerman were removed (temporarily) from the list then the differences between other terms would become more salient (as the font sizes would scale based on the range of most to least frequent words). However, one of our facilitators, trained in social justice issues, used the tool in a way that was unexpected to us. She was quick to point out that the conversation in this predominantly white classroom focused on Zimmerman, who is white, and there was surprisingly little mention of Martin, who is black. Each facilitator will draw their own conclusions from such observations, but we were encouraged to see that the tool supported insightful observations even in what seemed to be an unrevealing visualization to the untrained eye.

Dialogue and Advice Panes: Text Analysis

As mentioned above, one component of our project is researching automatic text analysis and machine learning algorithms (and soon also relationship networks) to identify deliberative skill, other indicators related to dialogue quality, and trends or opportunity points. Text analyses methods have advanced significantly in recent years. According to [14] the "increased use of automated text analysis tools can be attributed to landmark advances in such fields as computational linguistics, discourse processes... , cognitive science..., and corpus linguistics..." (p. 34). We are using three types of technologies. The first two, LIWC and Cohmetrix, are pre-existing text analysis tools that take text segments as inputs and output dozens of measurement or classification metrics. The third technology is a set of machine learning methods we are using that take text, reply and demographic information, and some of the LIWC and Cohmetrix outputs as input or training features, and output classification analysis (e.g. whether a segment of text demonstrates good "deliberative skill" or "self reflection").

In our work we are using the Cohmetrix multiple-level text analysis system [15], the LIWC "Linguistics Inquiry Word Count" application, and a variety of machine learning methods (see [4]). LIWC (Linguistic Inquiry Word Count [16]) is a well researched but "shallow" dictionary-matching text categorization system yielding about 80 linguistic categories (e.g. positive emotion words, pronouns, and causation words—some of the categories are defined by hundreds of words in the dictionary entry). CohMetrix [15] performs a series of deep-processing analysis (including semantic cohesion, latent semantic analysis, and reading complexity level) yielding about 100 metrics.

We have implemented the most straight-forward of these methods, LIWC-based dictionary-matching processing, into the Dashboard. An important thing about LIWC is that, though its dictionary-matching method is simple (like keyword-matching), hundreds of studies have been done using it (and contributed to its development) so many of the categories it uses are well researched in terms of how use of these linguistic categories correlate with important psychological or social phenomena. Other analysis methods will be incorporated into the Dashboard in the future (one of the constraints and unknowns is that some of the text processing methods will probably not run in real time, and thus, while useful in our deliberation analysis research, would not be appropriate for the Dashboard). The reader can look at the "common problems encountered in online facilitation" in the Design Considerations sub-section for an indication of the types of things we hope to eventually assess using automatic methods. We will sometimes refer to each metric or found category that our software evaluates in terms of an "agent" that that assesses

and reports its findings, even though, depending on the feature, "agent" may be more a metaphor than an implementation description.

Kushal et al. [17] describe ForumReader, a dashboard tool for "large scale online discussion" that also uses some text analysis. Like other systems, ForumReader's text analysis is content-based, and is used to identify main topics of conversation, related topics, and topic clusters. It uses text-summarizing and clustering algorithms. Our interest is more in supporting reflection on the quality of a conversation than its content, which should be facilitated by LIWC and CohMetrix analysis. The Dashboard contains software agents that watch for patterns or categories in the dialogue, and flags occurrences that reach a certain threshold. The current version of the system matches a set of about 20 word categories found in the LIWC dictionaries. When the condition is met, the word is highlighted in the text and a line detailing the analysis is appended in yellow below the post. Of the 80 text categories LIWC classifies, the Dashboard is tagging these: first person singular & plural, second person singular & plural, swear, posemo, negemo, anx (anxiety/anxious), anger, sad, certain, sexual, assent, tentative, negate, and inclusive. There were arrived at through conversations with our facilitation experts, and the need to keep the number of categories manageable. LIWC was not conceived of as a system of analyzing deliberative dialogue, and our experts suggested three additional categories of interest that were implemented: always-never (always, every, never, none, all, never, everywhere, nowhere); should (should); and question-words (how, what, when, where, who, whom, whose, why).

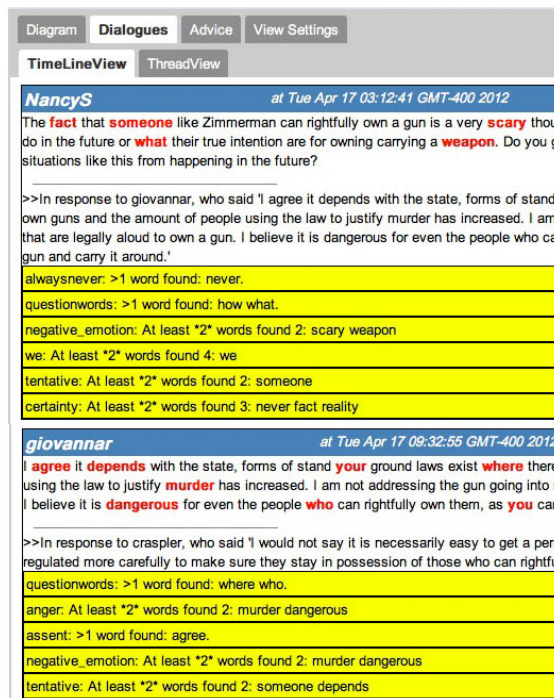


Figure 2: Dashboard: Dialogue Pane

As mentioned, this preliminary interface has proven, as expected, to be too busy and not graphically intuitive enough for facilitators to use easily. In the Next Steps section we describe our plans for improving it.

Formative Evaluation

We will not detail the numerous suggestions from our collaborators and advisors, many of which were implemented (and some of which are reflected in the design descriptions and rationale here). Recall that we worked closely with three professional mediator/facilitators over 18 months, and contracted feedback of 10 high-profile professional facilitator/mediators and leaders in that field, for a short-term consulting assessment. Overall the facilitators were very enthusiastic about the tools and their potential, finding the Dashboard "extremely helpful" and "extremely useful." In real-time classroom dialogue facilitation we instructed facilitators to take a very low-key role, motivating participation and supporting a safe space for

all to contribute, but trying not to influence the conversation very much (thus we asked them to leave some of their facilitation skills unutilized). Therefore, most of the feedback on the Dashboard has been in noting what was interesting or potentially very useful to know about a group, but little of it lead to interventions in this phase of the formative evaluation. For example, a facilitator might note that males, or seniors, were slightly dominating a conversation. In these trials they did not try to remedy this, but noted how useful the tools would be if they did want to ameliorate such phenomena.

Peer facilitation and group auto-reflection. We mentioned above that we have a goal to test the Dashboard as a group-awareness (or group auto-reflection) tool with participants and peer-mediators or facilitators. Our early forays into this area have consisted of showing participants (college classes) the dashboard in follow-up focus group sessions after they have completed an online dialogue assignment. Verbal responses have been enthusiastic, but it is also clear that we will need a simpler version of the Dashboard for participants as compared with what we offer trained facilitators.

Next Steps

Our future plans include evaluating our tools in civic deliberation and online dispute resolution contexts and we have engaged potential collaborators in both of these areas. Above we mentioned our plans for further evaluation of the Dashboard, and noted that we have gathered many comments from facilitators on ways to improve and extend it. In this section we describe some of these design plans. We will add an Alert feature that will highlight high priority dialogue events and properties. As mentioned we plan to add other types of text analysis beyond the dictionary matching LIWC-style agents (as is possible for real-time analysis). These include: deeper structural linguistic properties of the sort measured by Cohmetrix, and agents that search for trends, patterns, or inflection points (facilitation opportunities) over time, and post Alerts. As noted, the "common problems encountered in online facilitation" in the Design Considerations sub-section shows the types of things we hope to eventually assess. Of the dialogue speech acts that we manually code for, which we are attempting to build machine models to recognize, we have noted several that might signal important turning points in a conversation: changes of mind, peer-initiated mediation, apologies, high emotional tone, and acknowledgments/appreciations.

We have mentioned that the information in the text Dialogue pane needs more graphical portrayal, which will include pie charts, bar charts, and trend lines for the deliberative properties identified by text analysis. In addition, the current analysis only tags *words* and annotates *posts* according to word categories found in the post. We plan to include visualization tools that show occurrences and frequencies of triggered agents at four levels: post, participant, sub-group, and whole-group. For example, a chart could illustrate how instances of Appreciation or Meta-dialogue compare between participants or groups, or how they trend over the course of a dialogue. A facilitator will be able to focus in on a particular individual (or group) and display all of the analysis for that person. We will also include Settings to toggle on and off particular analysis agents, to reduce clutter and allow facilitators to focus on dialogue properties relevant to the context. The parameters for agents (e.g. triggering threshold values) will be moved from an initialization text file to a user-friendly Settings interface.

We have begun work to include social network analysis in the Dashboard [18]). This will analyze reply or reference structure between participants to measure and visualize: mutuality/reciprocity, centrality/influence, density, cliques and selective responsiveness, and initiators vs. responders.

Conclusions

We have described a novel Facilitators Dashboard tool that visualizes dialogue quality indicators for use as facilitation tools or participant social awareness tools that includes textual analysis, and described our initial attempts to use it in educational settings. We are particularly interested in supporting the "social deliberative skills" that interlocutors need to build mutual understanding and mutual regard in complex or contentious situations. Developing methods to scaffold SD-skills in online deliberation, for participants and third parties, could have an impact in many online contexts, e.g. knowledge-building, situated learning, civic engagement, and dispute resolution.

One of the goals of education is to produce competent national and global citizens capable of participating in democratic self-governance and capable of wrestling with the difficult questions and dizzying array of information and opinion they face in our technologically advanced society. Students engaged in extended collaborative knowledge building, discussion, or problem solving eventually

encounter moments of tension in which they are challenged to understand each other's perspectives and opinions. Engaging with others on complex topics requires not only learning the relevant facts and concepts and making logical inferences, but also engaging with the perspectives and opinions of others who may not share one's views or goals. Doing so requires skills that can be systematically supported. Our work points to how such skills can be supported in online deliberation, collaboration, and dispute resolution, in educational settings and beyond.

References

- [1] O'Reilly, T. and Musser, J. "Web 2.0 principles and best practices", O'Reilly Radar, November, 2006
- [2] Keen, A. *The cult of the amateur: How the democratization of the digital world is assaulting our economy, our culture, and our values.* NY: Doubleday Currency, 2007
- [3] Murray, T., Xu, X. & Woolf, P.B.. An Exploration of Text Analysis Methods to Identify Social Deliberative Skills. In *Proceedings of 16th International Conference on Artificial Intelligence in Education (AIED-2013)*. Memphis, TN, July 2013. K. Yacef et al. (Eds.): LNAI 7926, pp. 811–814, 2013.
- [4] Murray, T., Stephens, A.L., Woolf, B.P., Wing, L., Xu, X. & Shrikant, N. Supporting Social Deliberative Skills Online: the Effects of Reflective Scaffolding Tools. *Proceedings of 15th International Conference on Human-Computer Interaction (HCI-2013)*. Las Vegas, July 2013.
- [5] Iandoli, L., Quinto, I., De Liddo, A. & Shum, S.B. A debate dashboard to enhance online knowledge sharing. *VINE: The journal of information and knowledge management systems*, Vol. 42 No. 1, 2012, pp. 67-93, 2012.
- [6] Buder, J., Bodemer, D., Dehler, J., & Engelmann, T. SCAN tools for collaborative learning. In *Proceedings of the 9th international conference on Computer supported collaborative learning-Volume 1* (pp. 606-615). International Society of the Learning Sciences, 2009.
- [7] Murray, T., Woolf, B., Xu, X., Shipe, S., Howard, S. & Wing, L.. "Supporting social deliberative skills in online classroom dialogues: Preliminary results using automated text analysis." *Proceedings of 11th International Conference on Intelligent Tutoring Systems (ITS-2012)*. S.A. Cerri and B. Clancey (Eds.). LNCS 7315, pp. 669–671, 2012. June 2012, Chania, Greece, 2012.
- [9] Constantino-Gonzales, M. A., Suthers, D., & Escamilla, J. Coaching Web-based Collaborative Learning based on Problem Solution Differences and Participation. *International Journal of Artificial Intelligence in Education*, 13(2-4), 263-299, 2003.
- [10] Xu, X., Murray, T. Woolf, B P. & Smith, D. Mining Social Deliberation in Online Communication -- If You Were Me and I Were You , In *Proceedings of the 6th International Conference on Educational Data Mining (EDM)*, 2013.
- [11] Xu, X., Murray, T. Woolf, B P, " Text Analysis of Deliberative Skills in Undergraduate Online Dialogue: Using L1 Regularized Logistic Regression with Psycholinguistic Features", In *Proceedings of the 12th International Conference on e-Learning, e-Business, Enterprise Information Systems, and e-Government (EEE)*, 2013.
- [12] Soller, A., Martinez, A., Jermann, P., & Muehlenbrock, M. From mirroring to guiding: A review of state of the art technology for supporting collaborative learning. *International Journal of Artificial Intelligence in Education*, 15, 261-290, 2005.
- [13] Shrikant, N. & Murray, T. On Line Facilitation Guidelines for Supporting Social Deliberative Skills in Online Deliberation. Unpublished manuscript available from the authors. 2012.
- [14] Graesser, A. C., McNamara, D. S., & Louwerse, M. M. 2 Methods of Automated Text Analysis. *Handbook of Reading Research*, 4, 34, 2010.
- [15] Graesser, A. C., & McNamara, D. S. Computational analyses of multilevel discourse comprehension. *Topics in Cognitive Science*, 3, 371–398, 2011.
- [16] Tausczik, Y. R. & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*.
- [17] Kushal Dave, Martin Wattenberg, Michael Muller. Flash forums and forumReader: navigating a new kind of large-scale online discussion. In *CSCW '04: Proceedings of the 2004 ACM conference on Computer supported cooperative work (2004)*, pp. 232-241, 2004
- [18] D'Andrea, A., Ferri, F. & Grifoni, P. "An Overview of Methods for Virtual Social Network Analysis". In Abraham, A. Hassanien, A. E., & Snášel, V. (Eds.) *Computational Social Network Analysis: Trends, Tools and Research Advances*. Springer. p. 8, 2009.