Envisioning a Discussion Dashboard for Collective Intelligence of Web Conversations

Jodi Schneider
Digital Enterprise Research Institute
National University of Ireland, Galway
jodi.schneider@deri.org

Alexandre Passant
Digital Enterprise Research Institute
National University of Ireland, Galway
alexandre.passant@deri.org

ABSTRACT
Collective Intelligence for wicked problems is urgently needed. To integrate relevant information from discussions across the Web, we need summarization and visualization tools. To show an at-a-glance view of a debate, we envision a discussion dashboard, using automatic information, supplemented where possible, by manual analysis. To make the dashboard easy to conceptualize, we organize it around a simple conceptual model: the Five W's–who, what, when, where, why. Our proposed discussion dashboard would present a full, yet digestible, picture of an argumentative Web discussion.

Author Keywords
online argumentation, discourse, collective intelligence, decision-making, summarization

ACM Classification Keywords
H.5.3 Information Interfaces and Presentation: Group and Organizational Interfaces—*Collaborative computing*

General Terms
Human Factors

INTRODUCTION
The problems currently facing humanity – "wicked problems" [7] like climate change, energy and resource shortages, and globalization – require brainstorming and decision-making on a vast scale. In particular, it is important to reconcile the views, ideas, and preferences of individuals (both experts and the general public) along with the known facts and the prioritized issues. Dealing with disagreement is even more difficult due to an overwhelming number of scattered conversations. Summarization interfaces and visualizations are needed to help guide people to the important topics and contentious areas of a conversation.

A considerable amount of discussion takes place in Web2.0 systems such as bulletin boards, blogs, microblogs and wikis, which enable large-scale, distributed communication [20]. Thus far, these systems have been very successful for detecting and amplifying agreement – pulling together information from a variety of individuals who mainly agree with one another [24]. Yet when there is disagreement, harvesting information from discussions is still challenging. A number of discussion systems focus explicitly on structuring agreement and disagreement, e.g. [5, 8, 18], however these systems require participants to explicitly participate, on dedicated platforms.

We see three major issues with current approaches for dealing with decision-making and disagreement on the Web. First, by requiring contributions to be made on a particular platform, we ignore large amounts of conversation made in the distributed environment of the Web. Second, generating collective intelligence from these discussions has thus far focused on statistical and social signals (e.g. [6]), which can sometimes predict outcomes (e.g. [28]), yet are less reliable when there is disagreement (e.g. [14]). Third, unless we want to blindly accede to the majority view, coarse collective intelligence is generally insufficient for decision-making: understanding the reasoning or explanations people give for their views clarifies the points on which disagreement occurs, and is important for untangling wicked problems (and even some mundane everyday ones).

To address these issues, first, a collective intelligence system should harvest information from distributed discussions across the Web. Second, statistical and social signals should be combined into an overview; in addition to orienting decision-makers and providing them with context, these automatic measures might, in combination, reveal when opinions are highly varied, and thus where more analysis of subtleties would be informative. Finally, when understanding decisions is particularly important, these statistical measures can be supplemented with more fine-grained analysis and representation of the content.

In this position paper, we address these latter two issues, by proposing a framework for a discussion dashboard. Following this introduction, we first discuss our vision for this discussion dashboard, drawing from existing work in summarization and collective intelligence for product reviews and learning analytics. Second, we present a model for the relevant aspects of conversations, drawing from five information-bearing questions: who, what, when, where, and why–also known as the Five W's [13]. To move from our theoretical framework towards planning an implementation, we then suggest analysis techniques that could address each of these questions, focusing on automatic approaches. Finally, we describe our ongoing work and future directions, and then conclude.

In ACM Computer Supported Cooperative Work (CSCW12) - Workshop: Collective Intelligence as Community Discourse and Action, February 11, 2012, Seattle, Washington, USA. Available at: `http://events.kmi.open.ac.uk/cscw-ci2012/programme-papers-demos/`

ENVISIONING A DISCUSSION DASHBOARD

Discussions generate overwhelming numbers of messages; we believe that finding the best solutions requires engagement with minority viewpoints, which indicate the points on which persuasion is needed and which may provide valueable perspectives and unique expertise. Yet finding the signal in the noise of the crowd is challenging; this is exactly why collective intelligence is needed for argumentative discussions.

Visualization and summarization environments aimed at human processing are needed, and in most cases, little work has been done on presenting the algorithmic summaries to decision-makers. Product reviews and learning analytics provide inspiration for the sort of collective intelligence that is possible. These domains are useful because Web-based conversations in these areas have received particular sensemaking attention: product reviews have been subjected to intense individual and commercial scrutiny (though many of the deeper analyses remain unshared and unavailable for reuse), while learning analytics builds on a traditionally manual analysis by adding statistical and summarization-based approaches.

Collective Intelligence in Product Reviews

Summarization and visualization for collective opinions is most commonly seen in e-commerce; for example, Amazon.com has proven useful in part because of the massive amounts of data it collects and shares about consumer opinions. This collective intelligence information is presented alongside product information, in effect orienting prospective customers about a product. We get a summary of the collective opinion, as well as the variance in that opinion: Amazon, for instance, provides not only an average star rating for its products, but also visualizes the spread in the ratings. Previous customers' habits affect the page in other ways, displaying items frequently bought at the same time or by the same customer.

Each product page also invites a social space of pro/sumers, highlighting reviews, forums, and related discussions. Content summarization, however, is limited. While readers can indicate whether a review was helpful, or respond to it with their own comments, mainly, reviews function as an undifferentiated space for commentary, advice sharing, and occasional discussion. Some of this data in turn drives the display: for instance, 'helpfulness' is used to determine the order in which reviews are shown, and the 'most helpful' positive and critical reviews are prominently displayed.

In our view, it is the combination of these diverse pieces of information that is helpful; presenting average ratings without their spread would be far less helpful. There is also further potential for analyzing and visualizing more in-depth information, for instance about product features, by segmenting the important factors and summarizing the polarity of opinions [19]. By combining information from multiple sources and presenting it together, a prospective customer is able to quickly determine whether a product is suitable.

Collective Intelligence in E-Learning

Combining information from multiple sources has also proven useful in e-learning. The field of learning analytics has emerged as e-learning has provided a rich source of data (e.g. online discussions, login patterns, time on task) for studying how people learn. For instance Gerosa et al. [15] used the network structure of e-learning forum discussions to identify how much interaction a topic has generated, on the theory that mediators should intervene when the volume of conversation does not match the depth of a discussion. The authors later experimented with different grading policies, to encourage more interaction by promoting early replies. Students also categorized their own posts; analysis showed that the message types depended on the number of previous replies, and that message types and sizes corresponded. This work indicates that combining automatic analysis can provide collective intelligence signals about a conversation, pointing out the most likely places with interesting discussion.

More recently, De Liddo et al. [9] presented a discourse-centric approach to learning analytics. Students' manual annotation of their own messages enabled summarization of the rhetorical roles and moves, based on post type and links between posts. Further, they bring together various network measurements, such as ego-network measurements around both individuals and posts in a professional conversation.

In e-learning environments, unlike on the wider Web, manual annotation and posting strategies can be incentivized. Next we turn our attention to Web discussions, envisioning what sort of analysis could be beneficial for understanding discussions in this less controlled environment.

Towards Collective Intelligence in Web Discussions

To bring these ideas to Web discussions, we envision packaging a number of shallow, algorithmic approaches into a discussion dashboard, with the goal of enabling a better understanding of a conversation, without reading it. Combining existing visualizations to provide orientation to debates has previously been suggested [22], with the goal of improving interaction feedback, particularly in the areas of participants, the interaction process, and the content. Our approach differs in its model and its ideas for implementation; in particular, the purpose of the discussion is important to our model and we propose to use different types of automatic summarization. Our outlook is similar to that of [6], which identified structural differences in message board forums depending on their topic: overall statistics such as number of posters and reply rates can provide context about the type of conversation. With similar coarse statistics, we expect that it should be possible to highlight the turning points in a discussion, or to identify which viewpoints have diverse support.

This best effort automatic approach of our discussion dashboard can later be amplified with content, discourse, and argumentation analysis; combining automatic and manual approaches has already been found to be fruitful [10]. Yet currently, such deep analysis is only feasible in limited cases, since manual analysis is required. Next we present our model.

MODEL

Our model focuses on five key features of the basic structure of conversations: the Five W's–who, what, when, where, why–which help provide context about a situation. Similar features have been used in provenance and data curation ontologies such as the W7 model [23]. In decision-making, these information-bearing questions all need to be taken into consideration.

Who

It can be very important to attach statements to the people and organizations putting them forward; this is recognized in intelligence analysis and political discourse, for example, where misleading information can be intentionally propagated and where biases may significantly shape the discourse.

What

A decision-making discussion on a single issue may involve discussions of subissues and related topics. Assuming that the relevant discussions have already been assembled, we can then focus on topic drift and topic evolution as well as on deeper analysis of the content.

Full analysis would indicate what positions are taken, and the relationships between positions, such as attack and support. Various intermediate levels of analysis could be helpful.

When

The order of messages can be important, for instance, when a person changes their mind or when an organization updates or retracts a statement. Conversation outcomes can also be affected by the message order (i.e. 'first mover advantage' and 'last word').

Further, some information is inherently time-sensitive, and its validity may be called into question due to its age, or due to the arrival of more recent information from other sources.

Where

Geographic information can be relevant to assumed viewpoints and biases. It can also determine who the imputed stakeholders are, for some types of discussions.

The Web location of messages can also be relevant; for instance, microblog posts are briefer and often less formal than blog posts. Forum and listserv posts often rely on the surrounding messages for coherence and context. Temporal aspects of a discussion are also related to the genre; [11] observed that listservs have short, intense exchanges, organized as a tree, while blogs promote slower diffusion and may have multiple ancestor posts. Thus the source genre is relevant when combining messages from different genres, and in some cases messages may not be understandable without the context of the other messages to which they reply.

Why

The purpose of a discussion is a key aspect to understanding it. Philosopher and argumentation scholar Douglas Walton has identified seven types of dialogue: Persuasion, Inquiry, Discovery, Negotiation, Information-Seeking, Deliberation, and Eristic (Figure 1), determined from the participant's goal and the goal of the dialogue. We have analyzed how important knowledge, emotion, and personal values and opinions are in each dialogue.

Understanding the purpose can provide readers with insight into a single discussion. For instance, persuasion dialogues may call for more skepticism in reading, since participants may overstate their positions. Meanwhile, outcomes of deliberations depend on the participants' circumstances and values, and may not be suitable for direct application to another situation. When discussions are read without such context, perhaps reached via a search engine, they can be more difficult to interpret and reuse.

The knowledge/emotion/values classification can also guide the reuse and aggregation of multiple discussions. Knowledge-oriented discussion types are straightforward to reuse; opinion-oriented discussion types may require caveating or balancing; emotion makes a discussion more interesting [16] but also can indicate the potential for bias. Grouping similar types of discussions may make them easier or less jolting to read.

SUGGESTED INITIAL DASHBOARD COMPONENTS

In this section we use the Five W's model to suggest components that could be assembled for an initial dashboard implementation.

Who: Reply Networks

As mentioned above, we want to indicate the people and organizations involved in the conversation, as well as their relationships to each other. In addition to providing statistics about the number of participants, we can show the reply networks between pairs of people, perhaps indicating authors with icons or avatars. Disagreement in verbal dialogue can be signaled by interruption; graphing these interruptive replies generates useful overviews of political disagreements [17].

What: Topic detection

We want to determine and represent the topic, subissues, and topic drift. This can draw from metadata about the discussion (such as subject lines and hashtags), from topic identification, and from word frequency. To get a coarse idea of the number of positions, we can look for polarized, high sentiment words in responses, which may signal disagreement.

When:Timelines

Conversation timelines can be useful for understanding the temporal flow of a conversation. Identifying old, possibly outdated, information, is also aided by timeline views.

Extracting the structure of a conversation gives insight since the depth of the reply network tends to indicate what has been most heavily discussed [15]. Timeline information alone can be useful for reconstructing single discussions; for combining multiple discussions, more careful reconstruction is needed.

Where: Source websites and geographic information

Indicating source websites, as we suggested above, is straightforward when aggregating discussions, just using their provenance. Geographic information is somewhat less available,

Dialogue Type	Initial Situation	Participant's Goal	Goal of Dialogue	Knowledge	Emotion	Personal Opinions & Values
Inquiry	Need to Have Proof	Find and Verify Evidence	Prove (Disprove) Hypothesis	High	Low	Low
Discovery	Need to Find an Explanation of Facts	Find and Defend a Suitable Hypothesis	Choose Best Hypothesis for Testing	High	Low	Low
Information-Seeking	Need Information	Acquire or Give Information	Exchange Information	High	Low	Low
Deliberation	Dilemma or Practical Choice	Coordinate Goals and Actions	Decide Best Available Course of Action	High	Varies	High
Persuasion	Conflict of Opinions	Persuade Other Party	Resolve or Clarify Issue	Varies	High	High
Negotiation	Conflict of Interests	Get What You Most Want	Reasonable Settlement Both Can Live With	Low	Varies	High
Eristic	Personal Conflict	Verbally Hit Out at Opponent	Reveal Deeper Basis of Conflict	Low	High	High

Figure 1. Walton's seven types of dialogue, based on [30], expanded with our analysis of knowledge, emotion, and personal opinion/values.

though it is used in the aggregate on some occasions, for instance to produce dynamic maps of the visitors to a blog. Geographic information may serve a purpose in serious discussions, for instance, in contentious discussions about placenames; since naming conventions tend to be heavily correlated with geographic ties [27], indications about such ties can provide context for moderators about who is taking part in the discussion.

Why: Knowledge/Emotion/Values

Understanding the purpose of a discussion helps in interpreting and assigning meaning to its contents. We plan to detect the prevalence of knowledge, emotion, and values as a first approximation to the purpose.

Knowledge-based discussions often cite statistics, experts, and studies, which can be text-mined; they may also commonly use argumentation schemes such as expert opinion. High sentiment and low sentiment messages can be found through sentiment analysis [21], which we also use as a first indication of whether people agree and how strongly their views are expressed. Values are abstract qualities such as utility, beauty, respect, and patriotism; these can be found with gazetteers.

FUTURE DIRECTIONS

Our immediate line of work is to prepare a prototype of the automatic summarization, to be applied initially to a corpus of Wikipedia deletion discussions [25]. In our view, bridging human and machine analysis is key, and we also see potential for supplementing the approaches proposed in this paper with manual analyses and emerging algorithmic ones.

Several manual approaches have been particularly fruitful for argumentation analysis. Benn et al. identified viewpoint clusters in the abortion debate, using the KDA ontology and cognitive coherence relations to construct a knowledge base from which oppositions between clusters are calculated [4]. Factors analysis has been used to identify the pivotal aspects for a dispute or decision, for instance product features in [19]. Amplifying revert graphs for Wikipedia pages with user demographics– such as users' self-presentation of their heritage and geographic affiliation– has also shed light on disagreements and conflicts between users [27].

Meanwhile, steps towards automation are already being taken through research in detecting disagreement with natural language processing, [1, 29, 31]; highlighting contentious issues [12]; and detecting stance [3, 26] and persuasion [2, 32]. We hope that detecting argumentation could ultimately allow rich knowledge representations to be constructed from natural language; in the meantime, sensemaking analysts will continue to structure arguments, and knowledge representations should be shared among interoperating tool chains. Such representations, however constructed, can then be used in many ways, as suggested in [25].

CONCLUSIONS

We propose a discussion dashboard, that will integrate information from across the Web, to provide a glanceable summary of the state of a debate. The Five W's–who, what, when, where, why–are used to organize a model of what is important in conversations. We envision using automatic information, supplemented where possible, by manual analysis and newly emerging algorithmic approaches. Such a discussion dashboard would present a full, yet digestible, picture of an argumentative Web discussion.

Acknowledgements

This work was supported by Science Foundation Ireland under Grant No. SFI/09/CE/I1380 (Líon2). Thanks to Katie Atkinson, Marieke van Erp, Paul Groth, Conor Hayes, and Adam Wyner for helpful conversations, and to the anonymous reviewers.

REFERENCES

1. Abbott, R., Walker, M., Anand, P., Tree, J. E. F., Bowmani, R., and King, J. How can you say such things?!?: Recognizing disagreement in informal political argument. In *ACL HLT 2011*.

2. Anand, P., King, J., Boyd-Graber, J., Wagner, E., Martell, C., Oard, D., and Resnik, P. Believe me–we can do this! annotating persuasive acts in blog text. In *CMNA at AAAI 2011*.

3. Anand, P., Walker, M., Abbott, R., Tree, J. E. F., Bowmani, R., and Minor, M. Cats rule and dogs drool!: Classifying stance in online debate. In *ACL HLT 2011*.

4. Benn, N., Buckingham Shum, S., Domingue, J., and Mancini, C. Ontological foundations for scholarly debate mapping technology. In COMMA 2008.

5. Cartwright, D., and Atkinson, K. Using computational argumentation to support e-participation. *IEEE Intelligent Systems 24*, 5 (2009), 42–52.

6. Chan, J., Daly, E. M., and Hayes, C. Decomposing discussion forums and boards using user roles. In *ICWSM 2010*.

7. Conklin, J. *Dialogue Mapping: Building Shared Understanding of Wicked Problems*. John Wiley and Sons, Ltd., 2005.

8. De Liddo, A., and Buckingham Shum, S. Capturing and representing deliberation in participatory planning practices. In *Online Deliberation (ODET 2010)*.

9. De Liddo, A., Buckingham Shum, S., Quinto, I., Bachler, M., and Cannavacciuolo, L. Discourse-centric learning analytics. In *1st International Conference on Learning Analytics Knowledge (LAK 2011)*.

10. De Liddo, A., Sándor, Á., and Buckingham Shum, S. Contested collective intelligence: Rationale, technologies, and a human-machine annotation study. *Computer Supported Cooperative Work*, 1–32.

11. de Moor, A., and Efimova, L. An argumentation analysis of Weblog conversations. In *The 9th International Working Conference on the Language-Action Perspective on Communication Modelling (LAP 2004)*.

12. Ennals, R., Trushkowsky, B., and Agosta, J. M. Highlighting disputed claims on the Web. In *WICOW at WWW 2010*.

13. Flint, L. *Newspaper writing in high schools, containing an outline for the use of teachers*. Pub. from the Department of Journalism Press in the University of Kansas, 1917.

14. Gayo-avello, D. Limits of electoral predictions using Twitter. *ICWSM 2011*.

15. Gerosa, M. A., Pimentel, M. G., Fuks, H., and Lucena, C. J. P. No need to read messages right now: helping mediators to steer educational forums using statistical and visual information. In *CSCL 2005*.

16. Ahn, H., Geyer, W., Dugan, C., and Millen, D. R. "How incredibly awesome!" - click here to read more. In *ICWSM 2010*.

17. Kaptein, R., Marx, M., and Kamps, J. Who said what to whom?: capturing the structure of debates. In *SIGIR 2009*.

18. Kirschner, P. A., Buckingham-Shum, S. J., and Carr, C. S. *Visualizing Argumentation: Software Tools for Collaborative and Educational Sense-Making*. Springer, 2003.

19. Liu, B., Hu, M., and Cheng, J. Opinion observer: analyzing and comparing opinions on the Web. In *WWW 2005*.

20. O'Reilly, T. What is Web 2.0: Design patterns and business models for the next generation of software, 30 September 2005. `http://www.oreillynet.com/lpt/a/6228`.

21. Pang, B., and Lee, L. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval 2*, 1-2 (2008), 1–135.

22. Quinto, I., Buckingham Shum, S., De Liddo, A., and Iandoli, L. A debate dashboard to enhance on-line knowledge sharing. In *International Forum on Knowledge Asset Dynamics (IFKAD 2010)*.

23. Ram, S., and Liu, J. Understanding the semantics of data provenance to support active conceptual modeling. In *Active Conceptual Modeling of Learning 2006*.

24. Robu, V., Halpin, H., and Shepherd, H. Emergence of consensus and shared vocabularies in collaborative tagging systems. *ACM Trans. Web 3* (September 2009), 14:1–14:34.

25. Schneider, J. Building a Standpoints Web to support decision-making in Wikipedia. In *CSCW2012 Doctoral Colloquium*.

26. Somasundaran, S., and Wiebe, J. Recognizing stances in ideological on-line debates. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*.

27. Suh, B., Chi, E., Pendleton, B., and Kittur, A. Us vs. them: Understanding social dynamics in Wikipedia with revert graph visualizations. In *IEEE VAST 2007*.

28. Tumasjan, A., Sprenger, T., Sandner, P., and Welpe, I. Predicting elections with Twitter: What 140 characters reveal about political sentiment. In *ICWSM 2010*.

29. Walker, M. A., Anand, P., Abbott, R., Tree, J. E. F., Martell, C., and King, J. That's your evidence?: Classifying stance in online political debate. *Submitted to: Decision Support Sciences* (2011).

30. Walton, D. Types of dialogue and burdens of proof. In *COMMA 2010*.

31. Wyner, A., Mochales-Palau, R., Moens, M., and Milward, D. Approaches to text mining arguments from legal cases. In *Semantic Processing of Legal Texts*, no. 6036 in Lecture Notes in Computer Science. Springer, 2010, 60–79.

32. Young, J., Martell, C., Anand, P., Ortiz, P., and Gilbert IV, H. T. A microtext corpus for persuasion detection in dialog. In *Workshops at the Twenty-Fifth AAAI Conference on Artificial Intelligence* (2011).