

Argumentation 3.0: how Semantic Web technologies can improve argumentation modeling in Web 2.0 environments ¹

Jodi SCHNEIDER ^a Alexandre PASSANT ^a Tudor GROZA ^a and
John G. BRESLIN ^{a,b}

^a *Digital Enterprise Research Institute, National University of Ireland, Galway*

^b *School of Engineering and Informatics, National University of Ireland, Galway*

Abstract. Argumentative discussions are common in Web 2.0 applications, but the social Web still offers limited or no explicit support for argumentation. As Web 2.0 applications become more popular, modeling argumentation happening in these systems becomes important, to enable reuse and further understanding of on-line discussions. After reviewing four genres of online conversations—Web bulletin boards, Wiki talk pages, blog comments, and microblogs—and four current Web 2.0 argumentation systems, the paper suggests how Semantic Web technologies can be used to provide an interoperability layer for argumentation modeling across applications.

Keywords. informal argumentation, social networking, argumentation tools, Semantic Web, distributed conversations

1. Motivation

Social networking systems have increased in popularity, and substantive conversations occur in ‘Web 2.0’ media (such as forums, wikis, blogs, microblogs, etc). People argue implicitly (i.e. in comments on blogs), but these arguments must be inferred; the argumentative structure is rarely explicit. Meanwhile, structured argumentation tools, including web-based tools (such as Debategraph² and Compendium³ [4]), have slower adoption outside specialized domains such as enterprise and egovernment applications.

Many social media discussions could benefit from improved visualization and better presentation, for instance the ability to distinguish questions, disagreements, and elaborations or the ability to navigate by argument, rather than chronological order. We are inspired by earlier systems such as WIT, Hypernews, and Zest, which integrated social and argumentation features, and by recent surveys such as [9], which reviewed models for expressing the argumentation and rhetoric of scientific publications. The WIT discussion system⁴ aimed to make the current state of a discussion clear, by having the user indicate

¹This work was supported by Science Foundation Ireland under Grant No. SFI/09/CE/I1380 (Líon2).

²<http://debategraph.org/> and as used, e.g. in [7]

³<http://compendium.open.ac.uk/>

⁴<http://www.w3.org/WIT/User/Overview.html>

“whether he was agreeing, disagreeing or asking for clarification of a point” [1]. Hypernews⁵ asks users to indicate what kind of message they are posting⁶, then displays the message type as an icon in the forum’s thread view. Zest [24], a prototype email browser, supported lightweight integration of IBIS-based [12] argument maps, using “criticons” such as ([?], [#], [+], [-]) to mark paragraphs as questions, statement, supporting argument, or opposing argument; a fifth criticon, [!], indicated resolution of a discussion.

Lightweight annotation schemes based on similar techniques might find adoption on the Social Web, and in addition to improved visualizations on each individual site, we would also like to enable cross-website navigation driven by arguments. That is, we would like to identify, across various wikis, weblogs and other applications, who is arguing (positively or negatively) about a particular product, topic, or position. Public policy and shared events may be discussed across various different platforms but need to be viewed globally: Consider conversations provoked by the U.S. Health Care policy debates or the infamous ‘hand ball’ in the World Cup qualification of 2009. While many people talk about the same topic, there is little support for gathering conversations, in part because we lack shared identifiers for these topics (e.g. URIs). Cross-website navigation would make it possible to display all the arguments related to a URI (that URI could represent a blog post or a topic); we call this “*object-centred argumentation*”, since social media is centered on objects of interest, around which conversations develop [11].

We believe that Semantic Web technologies, that focus on interoperability between applications by relying on common data formats (RDF) and models or *ontologies* (RDFS/OWL), could play an important role in this. On the one hand, various characteristics of social media systems have been modeled using Semantic Web technologies [6], for instance via models such as FOAF⁷ and SIOC [2]. On the other hand, separate models based on Semantic Web technologies have been proposed for argumentation, such as IBIS-OWL⁸, SALT [10], and Scholarly Ontologies [19], and further the Argument Interchange Format (AIF) has been expressed in RDF [3]. Despite a similar metadata modeling layer via RDF(S)/OWL, there are still many gaps between these models.

To fill these gaps, our goal is to identify the needs of the Social Web community in terms of argumentation, and determine how social media argumentation patterns can be represented. Bridges—formal models and/or mappings—will be needed between the two sorts of ontologies: those for representing social media and those for representing argumentation, in order to bring the Social Web (‘Web 2.0’) and the Semantic Web together into a Social Semantic Web (‘Web 3.0’) for argumentation—‘Argumentation 3.0’.

This paper is organized as follows: First, we discuss use cases in social media, based on well-known applications. Second, we review existing Web-based tools for argumentation. Third, we work towards requirements for social media argumentation, based on a survey of users. Fourth, we review Semantic Web models for argumentation, considering their relation to social media. Finally we conclude, highlighting the need for an ecosystem approach.

⁵<http://www.hypernews.org/HyperNews/get/hypernews/reading.html>

⁶None, Question, Note, Warning, Feedback, Idea, More, News, Ok, Sad, Angry, Agree, Disagree

⁷<http://www.foaf-project.org/>

⁸<http://purl.org/ibis>

2. Argumentation in the Social Web

The Social Web includes many Web sites, each with its own affordances and interaction patterns, which affect which types of conversations are well-supported, and thus what kind and how much argumentation occurs⁹. In this section, we consider typical discussion environments from four types of social media: forums, wikis, blogs, and microblogs.

The message board has been a popular feature of Internet-based communication since mailing lists and Usenet newsgroups. Most forums employ some threaded display methods, where users post and reply to threads on a particular topic. Forums share some rhetorical characteristics of mailing lists, such as the tendency to quote previous comments, as well as some social conventions, such as the use of +1 to indicate agreement. On some boards, thumbs-up and thumbs-down signs are used, providing visual cues.

Wiki software saves a complete history of each page, and allows pages to be edited directly in a Web browser, facilitating collaboration. Wikipedia Talk pages often host argumentative discussions about editing articles, such as whether and how a topic should be covered, or whether particular sources are reliable. However, they lack some typical affordances: signatures, posting dates, and indentation are added manually, by social convention, whereas other systems store and display this information along with the message content. Only social convention prevents editing others' comments, and in some cases (e.g. responses to peer reviews), long comments are split by responses. While Talk pages are intended to support editing, discussions often remain after they have had their intended effect on page editing. When message volume is manageable, the topical, rather than chronological, order of wiki discussions has some advantages for coherence.

Blogs often include a comments section where readers can leave a response, and this format has been adopted by major newspapers, juxtaposing readers' reactions with the newspaper article cum blog post. This fragments the conversation about news items, since reactions to a newspaper story reside only on the platform for that paper, even though many news articles cover the same event, each from its own perspective. Blog comments can include long threads with substantive comments, or substantial back-and-forth replies between an author and one or more commenters. Replies may be threaded, and comments usually list the date and author (perhaps also with a visual cue such as an avatar). Even though a comment responds to a blog post at a given time, posts can be updated, and usually only the most recent version is publicly viewable.

Microblogging is a newer trend; Twitter¹⁰, is characteristic, and its brief posts (limited to 140 characters) are each globally available at a URI, and typically publicly viewable. In microblogging, each message stands by its own and forms part of a stream; some messages may also have secondary status as a reply to another user or as a retweet/repost broadcasting a prior message (similar to forwarding an email). The popularity of retweets [5] points to an inherent need to quote, even in very brief messages. While stream-based services make it difficult to maintain the coherence of a dialogue, short personal opinions, reactions, and interpretations are easy to post, and can be gathered through collation methods such as hashtags¹¹; preceding a word with a # symbol creates a link to other messages using the tag, but reduces the space available for message content.

⁹Argumentation may be explicit and commonplace (in task-based collaboration on wikis) or harder to spot (in microblogging where dialogue is comparatively rare).

¹⁰<http://twitter.com>

¹¹For instance, [17] used hashtags to gather tweets about the U.S. Presidential debates.

3. Web-based Argumentation Tools and Social Web Systems for Argumentation

This section provides an overview of four Social Web systems for argumentation: Cohere, Debategraph, Debatepedia, and LivingVote. These systems were chosen based on exploration of argumentation literature, Web searches, and Web browsing. At a minimum, the systems needed to be currently available and publicly viewable. They were chosen due to their influence, wide use, or novelty; after the selection period, another relevant system, the Climate Collaboratorium¹² came to light.

Cohere¹³ is a knowledge-mapping website, which allows user to view and create maps, or import them from Compendium. Maps consist of ideas, which can be taken from the site's public, global pool of ideas, or added to a user's private collection. Cohere offers sorting options and several views, including map, timeline, argument, and argument listing views. Although all the data resides on Cohere's server, plugins provide some integration with external sites: users can clip ideas and save websites from Firefox (similar to social bookmarking), or tweet from Cohere (using a Jetpack extension). Ideas can be private or shared, allowing the possibility of finding arguments and ideas which interact with your own, and suggesting that truly distributed systems could be useful.

Debategraph¹⁴ is a wiki debate visualization tool which has been adopted for use at the Kyoto climate change summit and is being tested by EU projects such as WAVE¹⁵. Visualizations can be embedded in other websites, and Debategraph encourages users to add hyperlinks within graphs. Debategraph's user interface is elaborate, and its navigation methods may take some time to get used to: As the focus changes, so does the graph, and for a novice user it can be confusing to figure out how to get back to a previous view. The learning curve to effective use is its main disadvantage.

Debatepedia¹⁶ bills itself as the "the Wikipedia of pros and cons". Sponsored by the International Debate Education Association, Debatepedia is a collaborative community effort to summarize arguments. Each argument page provides an overview, then a list of issues, with pros and cons supported by news articles and similar sources. It provides an intuitive editing environment, where users can edit just the relevant section, such as the pro or con for a topic. Debatepedia's biggest weakness is the lack of alternate visualizations; this could be overcome by using the existing separation of pros and cons.

At Living Vote¹⁷, users discuss pro and con arguments of issues, creating argument maps. A tree view provides a coherent view of the argument, which can be drilled down, where arguments and their counterarguments are presented side-by-side. To vote, users must answer questions designed to test whether they've read the arguments. Living Vote also prunes unhelpful arguments and aims to provide a "complete, persistent, constantly changing and up-to-date record of everyone's opinion on an issue as well as the arguments that led them to that opinion, weighted by each voter's understanding and participation". Complex arguments, where a position supports one issue but argues against another, are not supported by Living Vote's current interface. Living Vote succeeds at summarizing large-scale policy debates and deserves further examination by those interested in vote-based approaches.

¹²<http://www.climatecollaboratorium.org/>

¹³<http://cohere.open.ac.uk/>

¹⁴<http://debategraph.org/>

¹⁵<http://www.wave-project.eu/>

¹⁶<http://debatepedia.idebate.org/>

¹⁷<http://www.LivingVote.org/>

4. What Users Say They Want in Argumentation

To begin to understand users' perceptions of which features they value in online discussions, we conducted two prototype surveys of users¹⁸. Survey responses were solicited first on Twitter (8 replies) and then for a second, slightly modified survey¹⁹, by email in our Semantic Web lab (23 replies). Respondents rated the following ten items as "Important", "Not Important", or "Maybe Important", based on the question "When you're commenting in an online discussion or argument, which of these are most important to you?" 1. It's easy to use [usability] 2. I can see the whole conversation (even if it's on multiple social networks) [integration] 3. I can quote or reference earlier parts of the discussion [referential context] 4. I can indicate the topic I'm discussing [topical context] 5. There's enough space to write my own message [appropriate length] 6. I can add or view comments on the document, at the part or section where they apply [view context] 7. I can tell who wrote earlier messages [author context] 8. I know whether a comment is up-to-date (not superseded) [currency/temporal context] 9. There are extra features useful in my domain/area of interest [modularity/topical integration] 10. It's easy to see the messages I care about [visualization/sorting].

Usability and appropriate length were the most important features to users; next were author, temporal, or referential context and visualization/sorting; then topical context or integration; and the least important features were view context and modularity/topical integration. The lack of emphasis on integration is interesting; from one perspective the Social Web is fundamentally distributed, yet, from the researcher perspective, integration seems fundamental to facilitating dialogue and argumentation in these distributed conversations. The question about appropriate length was proposed with microblogs in mind, but even those reaching the survey from an email link emphasized this aspect.

One respondent suggested five additional aspects for reading and scanning online discussions, which should be considered in further studies: 1. a view of the reputation or role of the contributor (reputation) 2. an overview of a participant's interests/past contributions by topic (summarization) 3. indicator of sentiment in an argument discussion thread (sentiment) 4. indicator of possible repetitiveness/circularity (redundancy) 5. indication of a consensus emerging in an argument (consensus).

A disadvantage of studies of this kind is that it is hard to distinguish regular interaction with the medium from specific argumentation support. Similarly, case studies of how people are using social media for argumentation could be helpful, however those require a clear understanding of what argumentation can and should mean in the context of social media, a question which deserves further consideration.

One fundamental question is what amount of complexity users are willing to adopt in order to reap the benefits of argumentation; previous research has emphasized incremental formalization [18] because users do not generally understand the larger structure of an argument from the outset (see e.g. [19], page 29), and even experienced users can have difficulty holding a complex argumentation model in their heads (page 27, *ibid*). This leads us to believe that only a simple argumentation model will gain use in social media, unless the complexity can be mitigated by good interfaces and familiar metaphors.

¹⁸Further information at <http://jodischneider.com/pubs/suppinfo/2010COMMA/>

¹⁹Items were randomized for the first survey but not for the second survey, and one item, [view context] was at first described as "I can annotate documents or messages". The second survey also gave slightly more context, including the keywords in brackets and the indication that "We're trying to establish requirements for arguments (discussions where there's disagreement or differences in viewpoints)."

5. Semantic Web Models for Argumentation

The Semantic Web's strengths include modularity—easing integration of domain knowledge and topical models—and integration—allowing distributed systems to interoperate—along with the ability to reference anything (providing author, topic, or referential context), at any granularity addressable with a URI. Push-based Semantic Web protocols such as sparqlPuSH²⁰ can facilitate currency. Yet users' most important criteria—usability and appropriate length—depend on particular models and the implementations of those models, whose simplicity, for instance, varies considerably. To understand these factors, we next focus on five Semantic Web models for argumentation: AIF, AIF-RDF, DILIGENT, OWL IBIS, and SWAN/SIOC, and comment on their applicability to social media.

AIF [3], an RDF-based format for argument interchange, has been combined with the Dialogue Game Description Language (DGDL), to create Argument Blogging [22]. With a JavaScript plugin, bloggers add argumentative relations to their own posts, selecting whether they want to support or refute a highlighted statement, or attack an inference between statements. Even though blogs are still published at their usual location, this code allows the distributed conversation to be centralized and stored in a single database, facilitating collation and visualization. Argument Blogging's effectiveness is in its ability to hide the complication of the AIF ontology, simplifying to just three choices. Similar approaches could be used on forums and wikis, but for microblogging, this approach would need some modification, perhaps using a registry and a brief command language such as Zest's. Wide adoption would require cooperation of major blog hosts, since many bloggers use hosted systems rather than administering their own blog.

Rahwan's work on the AIF-RDF ontology and ArgDF system looks promising [16], [15]. AIF-RDF's strengths lie in its use of standards (AIF and RDF) and its ability to represent full argument schemes. To make a new argument, in the latest public demo of ArgDF²¹, a user must first choose an argumentation scheme by name, then add statements in the appropriate structure. However, most users do not have formal training in logic or argumentation, and while they may recognize (and even use) complex arguments, their ability to formalize these arguments lags behind. Incremental formalization suggests that users might first present their arguments, and then edit them to match an appropriate scheme; if this scheme were suggested (perhaps with NLP detection of the argument type [13]), AIF-RDF could back a powerful and user-friendly environment. However, developing a generic interface for AIF-RDF, which could accompany existing social media, seems challenging due to the inherent abstraction of argumentation schemes.

DILIGENT [20] is primarily a methodology for engineering an ontology; the acronym comes from certain letters in the phrase “DIstributed, Loosely-controlled and evolvInG”. Argumentation is used to track the process, in part because externalizing the information exchanged during ontology creation could help avoid rehashing discussions as new people join the process or later review the ontology. DILIGENT's strength is its previous use in collaborative ontology engineering, using IM and wikis [21], in which Talk pages are used to store “elaborations, arguments, positions and decisions ordered chronologically”. However DILIGENT has over 30 terms, including terms such

²⁰<http://code.google.com/p/sparqlpush/>

²¹<http://dundee.argdf.org/>

as Issue, position-on, Justification, and Decision; the intricacies which make it an ideal model for design support detract from its use by the layman.

The IBIS OWL Model²² is an RDF representation of IBIS, providing URIs and PSIs for these ten terms: Idea, Question, Argument, Decision, Reference, Note, Map, refersTo, pro, and con. The brevity, simplicity and clarity of the terms are promising for casual use across social media, but that use has not yet been tested.

SWAN/SIOC [14] harmonizes the argumentation aspects of models for neuromedicine and online communities. SWAN/SIOC's argumentation is based around specifying 11 types of relationships between items such as inconsistentWith, motivatedBy, and discusses. The system mediates and moderates this complexity, leading to the success of the model in scientific online communities such as a Parkinson's disease discussion site²³, where the intricacy of SWAN/SIOC is suited to representing the relationships between scientific arguments. These distinctions would not match some more general discussions, where they might be overly complex.

6. Conclusions & Future Work

With the rise of Web 2.0 systems, discussions happen everywhere on the Web, and argumentation is often present in those discussions. Semantic Web models have been developed and used to structure social media as well as in argumentation.

For forums and blogs, the approach taken by Argument Blogging needs mainly evangelism and integration with hosted systems to help the World Wide Argument Web [16] emerge, and a federated network in order to make it scalable and resistant to disruption.

For wikis, argumentation should use the inherent structure—the change over time. For instance, comments could be connected to particular text chunks, as in Commentpress²⁴ [8]. On some heavily-trafficked wiki Talk pages, FAQs are used to represent the current consensus, guarding against repetition and redundancy, but discouraging the involvement of new users; with a dialogue game system, perhaps based on MAGtALO [23], newcomers could interactively persuade or be persuaded, and the existing community would need to get involved only when new arguments warrant reexamining the consensus.

For microblogging, brevity is an overwhelming limitation; even indicating what is being argued about can be challenging, if a URI must fit in the message. Twitter Annotations²⁵ and client-based semantic annotations such as SMOB²⁶ provide a way forward, without depending on special fields for URIs or argumentative structure within a message.

Despite the proliferation of Semantic Web models for argumentation, a unified standard for argumentation on the Social Web is still lacking. We cannot yet collate arguments *across* social media, to find, for instance where microbloggers express approval of or provide brief counterpoints to blog posts, or where forums and listservs point to and extend wiki Talk page disputes. In short, we lack an ecosystem approach. In future research, we will work towards an argumentation ecosystem for social media, aligning Semantic Web models for argumentation with those for social media.

²²<http://purl.org/ibis>

²³<http://www.pdonlineresearch.org/>

²⁴<http://www.futureofthebook.org/commentpress/>

²⁵<http://apiwiki.twitter.com/Annotations-Overview>

²⁶<http://smob.me/>

References

- [1] Tim Berners-Lee. *Weaving the Web*. Harper Paperbacks, 1st edition, 2000.
- [2] John G. Breslin, Andreas Harth, Uldis Bojars, and Stefan Decker. Towards Semantically-Interlinked Online Communities. In *The Semantic Web: Research and Applications*, pages 500–514. Springer, 2005.
- [3] Carlos Chesnevar, Jarred McGinnis, Sanjay Modgil, Iyad Rahwan, Chris Reed, Guillermo Simari, Matthew South, Gerard Vreeswijk, and Steven Willmott. Towards an Argument Interchange Format. *The Knowledge Engineering Review*, 21(04):293–316, 2006.
- [4] Jeff Conklin, Albert Selvin, Simon Buckingham Shum, and Maarten Sierhuis. Facilitated hypertext for collective sensemaking: 15 years on from gIBIS. In *Proceedings of the 12th ACM conference on Hypertext and Hypermedia*, pages 123–124. ACM, 2001.
- [5] danah boyd, Scott Golder, and Gilad Lotan. Tweet, tweet, retweet: Conversational aspects of retweeting on Twitter. In *HICSS-43*. IEEE, 2010.
- [6] Ben Kei Daniel, Alexandre Passant, Sheila Kinsella, Uldis Bojars, John G. Breslin, and Stefan Decker. Understanding online communities by using Semantic Web technologies. In *Handbook of Research on Methods and Techniques for Studying Virtual Communities: Paradigms and Phenomena*. Information Science Publishing, 2010.
- [7] Aldo de Moor, Jack Park, and Madalina Croitoru. Argumentation map generation with conceptual graphs: the case for ESSENCE. In *CS-TIW 2009*, 2009.
- [8] Kathleen Fitzpatrick. CommentPress: new (Social) structures for new (Networked) texts. *Journal of Electronic Publishing*, 10(3), 2007.
- [9] Tudor Groza, Siegfried Handschuh, Tim Clark, Simon Buckingham Shum, and Anita de Waard. A short survey of discourse representation models. In *Proceedings of the Workshop on Semantic Web Applications in Scientific Discourse (SWASD 2009) at SWC-2009*, volume 523. CEUR-WS.org, 2009.
- [10] Tudor Groza, Siegfried Handschuh, Knud Möller, and Stefan Decker. SALT - semantically annotated LaTeX for scientific publications. In *The Semantic Web: Research and Applications*. Springer, 2007.
- [11] K.D. Knorr-Cetina. Sociality with objects: Social relations in postsocial knowledge societies. *Theory, Culture and Society*, 14(4):1–30, 1997.
- [12] Werner Kunz and Horst W.J. Rittel. Issues as elements of information systems. Working Paper 131, Center for Planning and Development Research, University of California at Berkeley, 1970.
- [13] Raquel Mochales Palau and Marie-Francine Moens. Automatic argumentation detection and its role in law and the Semantic Web. In *Law, Ontologies and the Semantic Web - Chamelling the Legal Information Flood*, pages 115–129. IOS Press, 2009.
- [14] Alexandre Passant, Paolo Ciccarese, John G. Breslin, and Tim Clark. SWAN/SIOC: aligning scientific discourse representation and social semantics. In *Proceedings of the Workshop on Semantic Web Applications in Scientific Discourse (SWASD 2009) at ISWC-2009*, volume 523. CEUR-WS.org, 2009.
- [15] Iyad Rahwan. Mass argumentation and the Semantic Web. *Web Semantics: Science, Services and Agents on the World Wide Web*, 6(1):29–37, 2008.
- [16] Iyad Rahwan, Fouad Zablith, and Chris Reed. Laying the foundations for a World Wide Argument Web. *Artificial Intelligence*, 171(10-15):897–921, 2007.
- [17] David A. Shamma, Lyndon Kennedy, and Elizabeth F. Churchill. Tweet the debates. In *Proceedings of the first SIGMM workshop on Social Media*, pages 3–10. ACM, 2009.
- [18] Frank M. Shipman and Catherine C. Marshall. Formality considered harmful: Experiences, emerging themes, and directions on the use of formal representations in interactive systems. *Computer Supported Cooperative Work*, 8(4):333–352, 1999.
- [19] Simon J. Buckingham Shum, Victoria Uren, Gangmin Li, Bertrand Sereno, and Clara Mancini. Modeling naturalistic argumentation in research literatures: Representation and interaction design issues. *International Journal of Intelligent Systems*, 22(1):17–47, 2007.
- [20] Christoph Tempich, H. Sofia Pinto, York Sure, and Steffen Staab. An argumentation ontology for Distributed, loosely-controlled and evolving engineering processes of ontologies (DILIGENT). In *The Semantic Web: Research and Applications*, pages 241–256. Springer, 2005.
- [21] Christoph Tempich, Elena Simperl, Markus Luczak, Rudi Studer, and H. Sofia Pinto. Argumentation-Based ontology engineering. *IEEE Intelligent Systems*, 22(6):52–59, 2007.
- [22] Simon Wells, Colin Gourlay, and Chris Reed. Argument blogging. In *CMNA 2009*, Pasadena, CA, 2009.
- [23] Simon Wells and Chris Reed. MAgtALO: an Agent-Based system for persuasive online interaction. In *AISB 2008*, 2008.
- [24] Ka-Ping Yee. Zest: Discussion mapping for mailing lists. In *CSCW 2002*. ACM, 2002.