An Abstract Framework for Modeling Argumentation in Virtual Communities

Tudor Groza, National University of Ireland, Ireland

Siegfried Handschuh, National University of Ireland, Ireland

John G. Breslin, National University of Ireland, Ireland

Stefan Decker, National University of Ireland, Ireland

ABSTRACT

Classic argumentative discussions can be found in a variety of domains from traditional scientific publishing to today's modern social software. An interactive argumentative discussion usually consists of an initial proposition stated by a single creator, followed by supporting propositions or counter-propositions from other contributors, usually part of the same virtual community. Thus, the actual argumentation semantics is hidden in the content created by the contributors. Although there are approaches that try to deal with this challenge, most of them focus on a particular domain, limiting the scope of the argumentation to that domain only. In this article, the authors describe an abstract model for argumentation which captures the semantics independently of the domain. Following a modularized approach, the authors also take into account additional important aspects of the argumentation, like the provenance information or its evolution (the temporal side). Consequently, they present a possible usage of the framework in the context of virtual communities. [Article copies are available for purchase from InfoSci-on-Demand.com]

Keywords: Argumentation Framework; Ontologies

INTRODUCTION

Argumentation can be found and captured in a variety of fields ranging from scientific publications, to ontology engineering, agent interaction or modern social software. An interactive argumentative discussion usually starts with an initial proposition stated by a single creator. This is then followed by supporting propositions or counter-propo-

sitions from other contributors. The actual semantics of the argumentation is hidden in the content created by the participants and therefore it is difficult to leverage this for use by machines.

Externalization represents the process of transforming implicit knowledge (such as the knowledge hidden in the argumentative discussions) into explicit knowledge, thus making it machine-processable (Nonaka et. al, 1995). One way of achieving externalization is by using formal models (ontologies) to capture the argumentation. There exist an important number of argumentation models, most of them following the direction given by the IBIS methodology (Kunz et. al, 1970). One of the main issues with each of these models is the focus on a particular knowledge domain, therefore limiting the view of the argumentation to the scope of that domain only.

When trying to model argumentation in a new domain, one faces the challenge of choosing the 'best' option from the current ones in existence, with the remark that only partial re-use is possible due to domain restrictions. As a result, in most cases researchers will tend to create a new model for their specific domain. This clearly shows the lack of an abstract enough model which allows a straightforward specialization for different specific needs. In addition, such a model should be able to fulfill a series of requirements dealing some pragmatic issues, like modularization, provenance or evolution.

In this article, we propose an abstract argumentation framework, which covers all of the above-mentioned issues. The framework is comprised of two layers: (i) a document model, capturing the environment in which the argumentation is present, and (ii) the argumentation model itself.

By having two layers, we follow a modularized approach, making a clear distinction between the document providing the provenance information and the identification of the argumentation elements and the argumentation per se. Based on the specific domain, a third layer can be added, thus introducing domain knowledge into the model. In addition, since such knowledge has its own particular terminology and language, this layer could also be comprised of linguistic features, providing the means to build semi-automatic knowledge acquisition tools.

In the following, we introduce the use-cases and requirements driving our framework. Then, we describe the framework itself, and before concluding, we re-visit the use-cases and present the relevant related work in the field.

USE CASES

As already mentioned, argumentation can be found and modeled in a variety of domains. Such domains usually have in common, besides the presence of argumentative discussions, also the organizational environment, while differing through the type of domain knowledge involved in the argumentation, or the physical environment (e.g. publications, forums, blogs, etc) in which the argumentation takes place. A common example of organizational environment is represented by virtual communities. These can be seen, from a simplistic perspective, as a group of people sharing a loose common interest, via virtual communication paths. Among such virtual communities, we could mention: online communities, scientists focused on a particular domain, lawyers, etc. An important remark is that all the previous

examples share, in one way or another, the presence of argumentative discussions. For example, dissemination represents a communication channel between scientists, spanned across multiple publications. They make claims, state positions and argument these positions, and thus, creating a virtual argumentative discourse network. In the following, we will focus on one particular use case, covering to a big extent all the others, i.e. online communities.

Online Communities

Online communities are formed by people through combinations of one-to-one (e.g. e-mail and instant messaging), one-to-many (web pages and blogs) and many-to-many (forums, wikis) forms of communication. Social software refers to the applications underlying these online communities, enabling people to connect or collaborate through computer-mediated communication (CMC). Prior to what is called Web 2.0, i.e. the wave of collaborative social sites now prevalent on the Web, online communities were mainly formed via BBS services, mailing lists, USENET, and web-based bulletin boards. Now, multi-forum sites, online social networks, weblogs and wikis are the hubs for many communities online.

However, despite the longevity of these communities, it is not possible to view or leverage any benefits from the argumentative structures that are implicit in the conversations that are taking place in the many millions of discussions contained in various social websites (an example of such a discussion is depicted in Figure 1). While some forum sites allow the use of icons to identify the type of replies that occur in a threaded discussion, very few make use of these identifiers to help users when they are searching for a particular type of response.

In CERN's "Discussion" system from the 1990s, a person could not „just reply".

Figure 1. Example of a threaded discussion in a forum
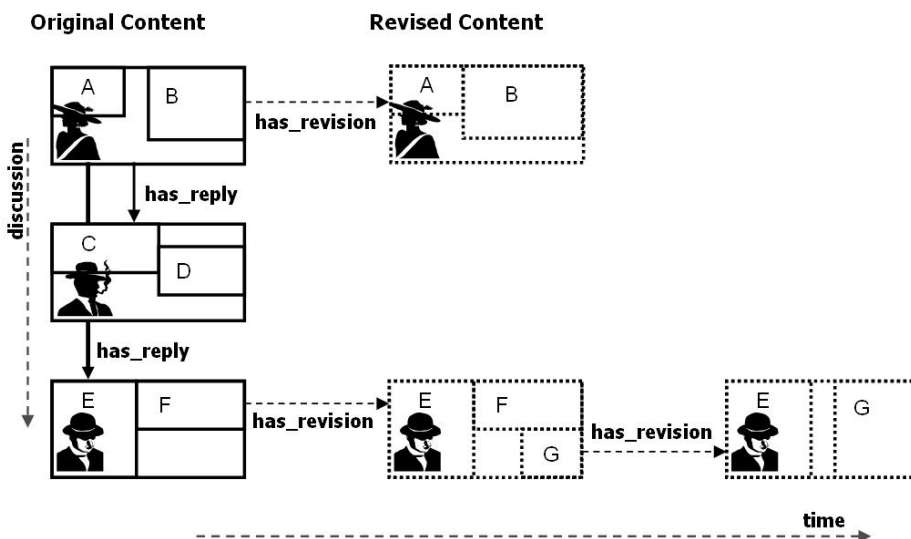
Rather, they had to state whether they were agreeing, disagreeing or asking for clarification of a point, and by enforcing this, the current state of a discussion or the role of a particular message in an argument was visible to all involved.

Tim Berners-Lee described in his book "Weaving the Web" how such argumentative context could be expressed in an open form to support ongoing discussions:

Imagine having servers for comments in different forums, perhaps family, school and company. Each point and rebuttal is linked, so everyone can see at a glance the direct agreements and contradictions and the supporting evidence for each view, such that anything could be contested by the people involved. If there was some sort of judicial, democratic process for resolving issues, the discussion could be done in a very clear and open fashion, with a computer keeping track of the arguments.

One example of a tool that can aid with such discussions, and makes a first step toward the externalization of the argumentation in online communities is the argumentation visualization site Debategraph[1]. The goal of Debategraph is "to make the best arguments on all sides of any debate freely available to all and continuously open to challenge and improvement by all". It is an impressive tool that has evolved from the work of just two people over the past few years, and was tested by Downing Street on their website following a speech by Tony Blair in 2008.

REQUIREMENTS

In this section we detail a set of requirements that emerge from the use cases presented in the previous section and from the need of maximizing the comprehensibility of the framework.

- **Modularization.** Modularization represents a key requirement for ontologies in order to achieve re-use and evolution (Rector, 2003). In the same way, an argumentation framework should clearly decouple domain knowledge from the provenance information representation and from the argumentation itself. This will lead to a model which will easily support evolution and integration with ontologies that capture specific needs.
- **Provenance information.** Argumentative discussions assume the presence of at least two actors. Thus, in order to create a comprehensive view, it is important to know who are the actors, and what are the means of the discussion.
- **Identification and Revision.** Argumentation is formed by a series of interlinked elements, usually represented by documents or chunks of text, which capture the argumentation semantics. The lower is the granularity of the elements, the more exact is the resulting model. Independently of the case, we need to be able to identify (uniquely if possible) these elements as a means to find the provenance information. In addition, by keeping track of the revisions brought to the elements we can visualize the evolution of the argumentative discussion over time (the temporal aspect of argumentation).

- **Support for domain knowledge and specific complementary models.** Although not directly coupled with the argumentation model, the framework has to provide support for embedding domain knowledge. One should be able to plug dynamically into the framework a specific ontology and to directly profit from the model as a whole. At the same time, depending on the application domain, an argumentation framework should be easily complemented with specific models for that domain. Examples of such models could be: SIOC (Breslin et. al, 2006) for social online communities, or RST (Rhetorical Structure of the Text) (Mann et. al, 1987) for scientific publishing.
- **Support for linguistic features.** As the current tendency is toward semi-automatic knowledge acquisition, there is a need for linguistic features to enable it. The framework should allow easy integration with a linguistic approach covering both domain knowledge and / or actual argumentation via verb tense and cue phrases for segment identification (e.g. LingInfo (Buitelaar et. al, 2006)).

ARGUMENTATION FRAMEWORK OVERVIEW

Figure 2 depicts a high level overview of our argumentation framework. As previously mentioned, in order to insure a high modularization, the framework consists of two layers: a document model and the argumentation model. In addition, the framework can be complemented with domain knowledge, other specific models or linguistic features, by adding a third layer, and thus providing the possibility of using the model as a whole for enhanced semi-automatic knowledge acquisition or reasoning purposes. In the following we will detail the two main layers of the framework.

The Document Model

The document model, as depicted in Figure 3, provides an abstract way to capture information about documents and their structure, independently of their format. It is composed by three main concepts:

- *Document*, the root concept, which acts as an information aggregator providing the entry point for describing the

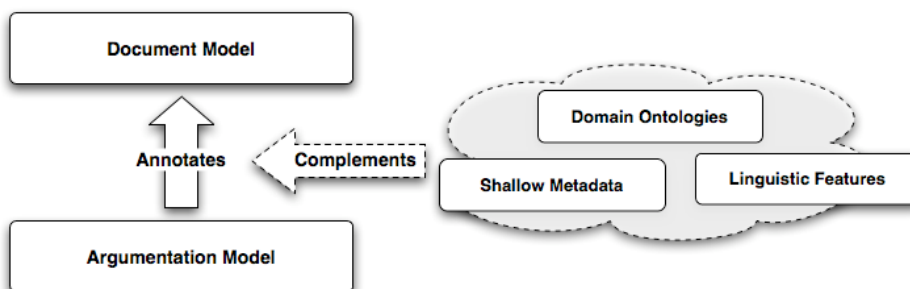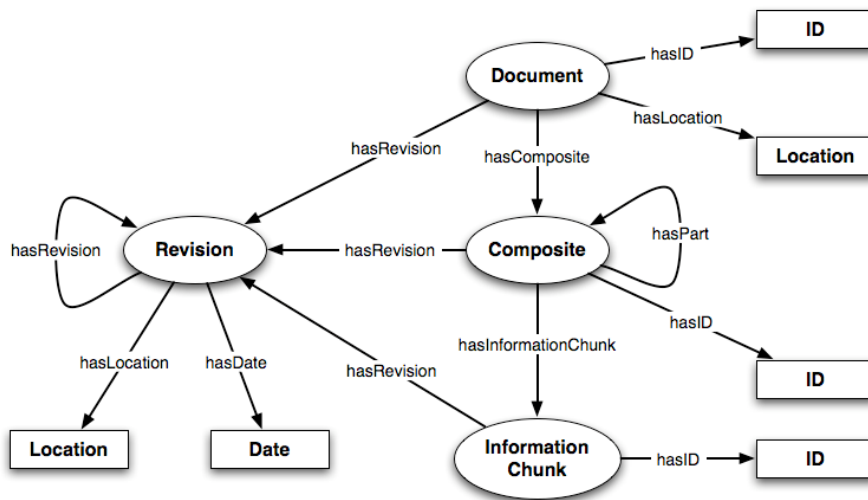*Figure 2. High level overview of the framework*

Figure 3. Detailed view of the document model

document in terms of its structure, provenance, or other aspects one wants to attach,

- *Composite*, the document's subdivision, providing the means for modeling the document's tree structure, and
- *Information Chunk*, the lowest granularity item, part of a document, able to capture variable sized information chunks as subdivisions of the *Composite*.

The identification aspects are present in the framework through the identification property. Each of the concepts has attached an *ID* property, the difference being given by the scope of the identification. At *Document* level this can be used with a global scope, while for the other concepts it could have a rather local (inside the document) scope. The root level has also a *location* property that can augment the identification with the goal of modeling the provenance information of the document.

The last aspect of the document model is temporal aspect. By introducing the *Revision* concept we want to be able to capture the evolution of the document at every level. Implicitly, together with the document structure and content, the evolution will be reflected also in the argumentation model. Thus, one will be able to trace the way in which the argumentation evolved based on the different revisions that the document had suffered.

An example of instantiation of the document model could be represented by a *Publication*. A publication is a *Document*, having multiple *Sections* (i.e. *Composites*), which in turns comprise several *Paragraphs* or *Sentences* (i.e. *Information Chunks*). The section numbers could represent the local identifiers for the sections. Another example could be a discussion *Thread*. A discussion thread is a *Document*, composed by multiple *Posts*, which contain several *Paragraphs* of text.

The Argumentation Model

Our argumentation model follows closely the initial approach introduced by IBIS (Kunz et. al, 1970) and then developed further by DILIGENT (Tempich et. al, 2005). As shown in Figure 4, the model consists of four main concepts: *Idea*, *Issue*, *Position* and *Argument*. A typical argumentation flow would start with one expressing an *Idea*, which could raise an *Issue*, or about which others could state their *Positions*. A *Position* could take the form of a *Request*, or could *Agree*, *Disagree* or be *Neutral* in regards to an *Issue* or an *Idea*. One could *Argument* the *Issues* that were raised or the existing *Positions*, either by providing a *Justification* or by starting a *Challenge* which would leave the argumentative thread open for debate.

It is important to note that our main goal is to provide an abstract form of modeling argumentation (similar to an upper level ontology) which can then be customized for personal needs. This implies that one can specialize, for example, the *Argument* class with a different subclass, and then use it in her own model. The same can be stated regarding possible other relations that can be introduced.

The Model as a Whole

Having described the document and argumentation models, in the following we detail how are the two models connected and how can one plug additional information, to create a unitary model. Figure 5 depicts an example of the way in which the different layers of the framework are linked. An instance of an *Issue* is present in a particular *Document* (or a *Revision* of that document) and has as representation in the document an *Information Chunk*. In a similar way, one can model a *Position*, an *Argument* or an *Idea*.

Adding additional knowledge to the model is straightforward. One can attach shallow metadata by introducing the author of the document (as in the example, by using *dc:creator*), or directly attach different domain concepts, for example *Protein123*.

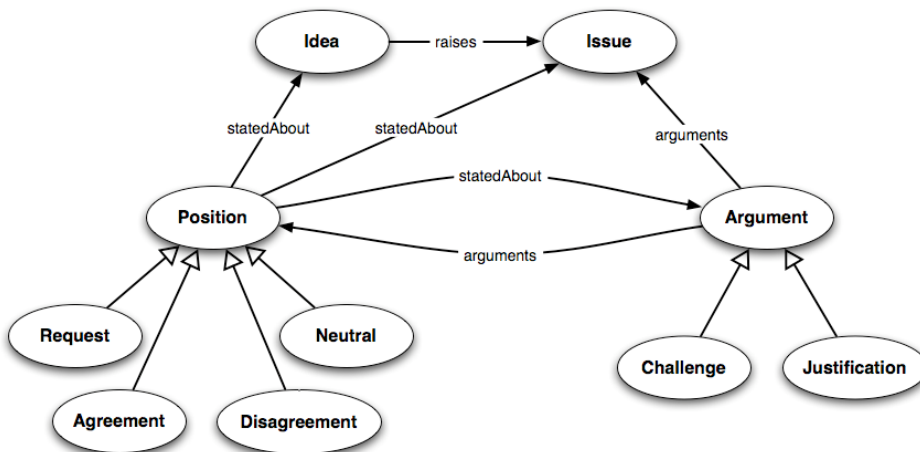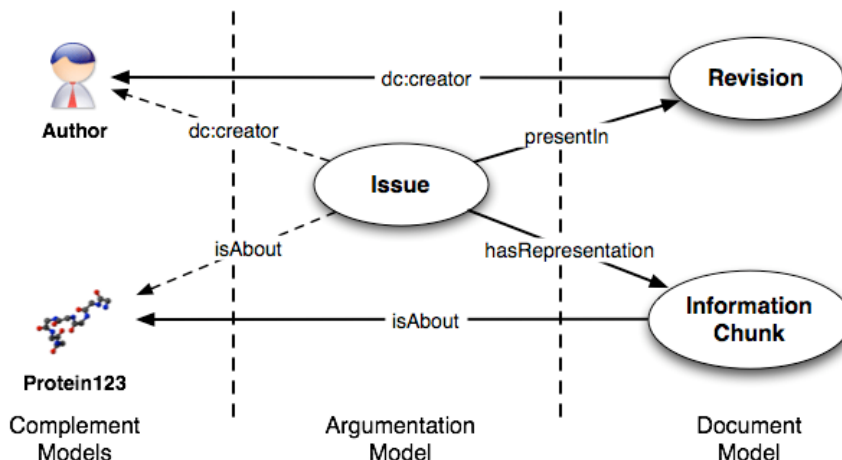*Figure 4. Detailed view of the argumentation model*

Figure 5. Linking the document to the argumentation

The place where the new knowledge is plugged into the framework is established by the user's needs. One can attach a domain concept at the document level, thus stating something about the document as a whole, or at the *Information Chunk* level, and therefore lowering the granularity of the stated information. The fact that the framework is layered has a consequence over the entire model: stating facts about the document implies stating facts about the argumentation, and vice-versa.

USE CASE – REVISITED

In Section 2 we listed a series use cases that act both as motivation and as starting point for collecting prerequisites for our framework. In the following we revisit the use case on which we focused, i.e. online communities, and show the way in which one can model it by means of our argumentation framework.

Semantically-Interlinked Online Communities

Semantically-Interlinked Online Communities (SIOC) is a Semantic Web framework that enables interconnection and interoperation within and between social websites, based on an ontology that describes the main terms occurring in these sites. Some of the relevant concepts in SIOC include *Sites* that host *Forums* which contain *Posts*. Posts are linked to each other in a threaded discussion structure via *has_reply* and *reply_of* links in either direction.

Threaded discussion structures are normally local to a particular community, but with services such as trackbacks on blogs allowing distributed reply structures, threaded discussions (and argumentative derivatives thereof) can also be distributed across sites. The SIOC idea for distributed threads is closely related to the cite-rel draft by Ryan King and Eran Globen[2] from the microformats community. Their *rel-reply* roughly corresponds to *sioc:has_reply*, and *rev-reply* corresponds to *sioc:reply_of*;

rev-update and *rel-update* may correspond to *sioc:previous_version* and *sioc:next_version* respectively; and *via* may be compared with *sioc:related_to*. Using the *rev* or *reply* links would allow distributed threads to form, e.g. if used from the post content. *cite-rel* brings the reply idea a step further by introducing *rel-forward* and *rev-forward* (basically, *sioc:has_reply* and *sioc:reply_of* with quoted content). Ultimately, we need ways to say that a post is in agreement or disagreement with a previous post or even with specific parts of a previous post by creating other reply types.

An argumentation extension to SIOC will include the possibility to formulate agreement and disagreement between SIOC content items. This can be represented by creating sub-properties of existing properties in SIOC (e.g. *has_reply*, *related_to*), and then by relating these to the argumentation model terms described in Section 4. Some of these are illustrated in Figure 6. At the abstract level, a *Thread* can be rep-

resented as a *Document*, while each *Post* can be an *InformationChunk*.

It may be necessary to extend these terms for use cases where more detailed discourse representation is required. Also, one may want to define exactly what it is that parts of a discussion will be in agreement or disagreement with: for example, a statement (an opinion or a well-known fact), a question, a topic, etc. Similarly, there may be a need for more fine-grained argumentation: rather than agreeing or disagreeing with an entire post, someone may refer to a knowledge chunk within that post. In such a case, the *Post* could be seen as a *Composite*, further decomposed into multiple *InformationChunk*s.

Even in just a single thread there can be other challenges in terms of identifying agreement and disagreement. When user revisions of content are brought into play, there can be significant complexity relating to implementing such reply types. For example, there are at least two types of

Figure 6. Example of externalization of the argumentation present in the threaded discussion in forum
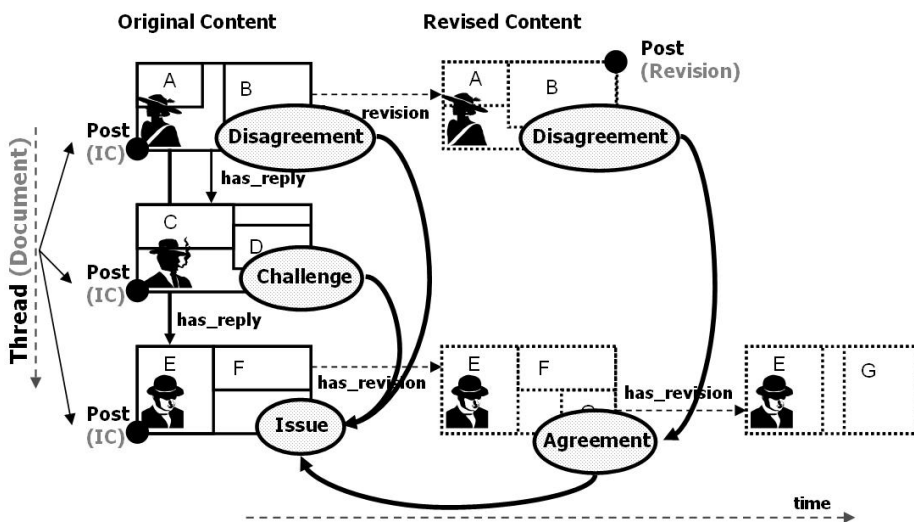
agreement or disagreement: those between responses to the discussion topic (thread starter), and others between revised versions of a content item (starter post or reply).

RELATED WORK

The background and previous research performed in argumentation covers, in general, an important number of related directions. In this section we will focus on three main aspects:

- background theories and models, which were used as inspiration by the majority of the currently existing argumentation models,
- existing argumentation models, having similar goals with our approach, and
- specific models that were used to complement these argumentation models.

In terms of background theories and models, we found three of them as being relevant for covering the directions from which most of the argumentation models were inspired. The first of them is IBIS (Issue Based Information Systems) (Kunz et. al, 1970). IBIS introduced a methodology for argumentation-based decision making in information systems, adopted by most of the current ontology-based argumentation frameworks, like DILIGENT (Tempich et. al, 2005), the Compendium methodology (Mancini et. al, 2006) or SALT – Semantically Annotated LaTeX – (Groza et. al, 2007).

A second important background theory is the Speech Acts Theory (Cohen et. al, 1979) that models the language aspects of speech acts and their planning in human communication. Although not directly (re)

used in argumentation models (one of the main application areas being email workflow modeling), this theory represented groundbreaking research that later lead, for example, to the third important theory, i.e. the Dialogue Games Theory. The Dialogue Games theory (Carlson, 1983) proposed a novel direction for the general Game Theory by considering discourse analysis and the logics and rhetorics of the human communication. This approach can be found as inspiration in most of the agent-based argumentation models.

One of the early argumentation models was the one of Conklin et. al, i.e. gIBIS (Conklin et. al 1987). This was following closely the original IBIS model and applied its methodology in team-based deliberation. gIBIS served as inspiration for later models like:

- DILIGENT (Tempich et. al, 2005), which applies argumentation in ontology engineering,
- Compendium (Mancini et. al, 2006), that follows a semiotic (Ogden et. al, 1923) approach for dealing with knowledge visualization and design rationale, while complementing argumentation with Cognitive Coherence Relations (Mancini et. al, 2006), or
- The Zeno argumentation framework (Gordon, 1997) applied in mediation systems.

Other relevant argumentation models include the one proposed by Torroni et. al in (Torroni et. al, 2007) for dealing with agent-based argumentation in the Semantic Web, or in the case of communities of Web Services the one introduced by Bentahar et. al in (Bentahar et. al, 2007).

The bio-zen[3] initiative is an attempt to represent data, information and knowledge

from research in all facets of life sciences on the Semantic Web. The goal of this project is the unification of information that is now scattered through a multitude of different data structures, exchange formats and databases. As part of this, the Semantically-Interlinked Scientific Communities (SISC) effort aims to improve how scientific data and knowledge is currently being represented and communicated.

Bio-zen and SISC uses SIOC, FOAF, DC, Creative and Science Commons, OBO and HCLS ontologies and technologies as its foundation. The bio-zen ontology framework also has a proposed bio-zen DOLCE use case for the W3C Semantic Web Health Care and Life Sciences Interest Group (HCLSIG). SIOC has been adopted by authors of this initiative for the representation of basic scientific discourse in scientific publications or on the web.

Two interesting properties use in bio-zen (that are along the lines of argumentative discussion and IBIS) are *supported-by* and *in-conflict-with*, allowing bio-zen "to represent the basics of scientific discourse (e.g. one can make the statement that a certain posting / document / dataset is supported or in conflict with some other posting / document / dataset)".

Although similar to the existing models, our framework brings novelty through its layered approach and by coupling the argumentation itself and the environment in which this takes place. As compared to the above-mentioned solutions, it provides an innovative way for attaching domain or linguistic knowledge, which could improve the knowledge acquisition and enrichment process. In addition, our framework captures also the temporal aspect of argumentation, by keeping track of the evolution of the argumentative discussion by modeling revisions of the 'hosting' environment.

CONCLUSION

The main goal of the abstract argumentation framework presented in this article is to enrich the current existing solutions by introducing a layered approach which makes a clear distinction between domain knowledge, the actual argumentation and the environment in which the argumentation is captured. We focused on aspects that were not considered in most of the existing models, like modularization, provenance or identification. We proved the applicability of our framework by using it in modeling argumentative discourse networks captured in scientific publications and argumentative discussions present in semantically interlinked online communities.

Future work on our framework will include an analysis on how we can transpose the current model into one or several ontology design patterns (Gangemi, 2005), and more specifically into content patterns. In addition, we will seek other possible application domains for the framework, like for example software engineering or software development.

REFERENCES

Bentahar, J., Maamar, Z., Benslimane, D., & Thiran, P. (2007). An Argumentation Framework for Communities of Web Services. *IEEE Intelligent Systems, 22*(6).

Breslin, J., Decker, S., Harth, A., & Bojars, U. (2006), SIOC: An Approach to Connect Web-Based Communities. *The International Journal of Web-Based Communities, 2*(2).

Buitelaar, P., Declerck, T., Frank, A., Racioppa, S., Kiesel, M., Sintek, M., Engel, R., Romanelli, M., Sonntag, D., Loos, B., Micelli, V., Porzel, R., & Cimiano, P. (2006). LingInfo: Design and Applications of a Model for the Integra-

tion of Linguistic Information in Ontologies. In *Proceedings of OntoLex06, a Workshop at LREC, Genoa, Italy*.

Carlson, L.M. (1983). *Dialogue Games: An Approach to Discourse Analysis*. D. Reidel Publishing Company.

Cohen, P.R., & Perrault, C.R. (1979). Elements of a Plan-Based Theory of Speech Acts. *Cognitive Science, 3*, 177–212.

Conklin, J., & Begeman, M.L. (1987). gIBIS: A Hypertext Tool for Team Design Deliberation. In *HYPERTEXT '87: Proceeding of the ACM conference on Hypertext*, (pp. 247–251). ACM Press.

Gangemi, A. (2005). Ontology Design Patterns for Semantic Web Content. In *Proceedings of the 4th International Semantic Web Conference (ISWC2005). Springer Verlag Berlin-Heidelberg*.

Gordon, T. (1997). The Zeno Argumentation Framework. In *Proceedings of International Conference on Artificial Intelligence and Law*.

Groza, T., Handschuh, S., Moeller, K., & Decker, S. (2007). *SALT - Semantically Annotated LATEX for Scientific Publications*. In ESWC 2007, Innsbruck, Austria.

Kunz, W., & Rittel, H. (1970). *Issues as elements of information system*. Working paper 131, Institute of Urban and Regional Development, University of California.

Mancini, C., Scott, D., & Shum, S.B. (2006). Visualising Discourse Coherence in Non-Linear Documents. *Traitement Automatique des Langues, 47*(1), 101–108.

Mancini C., & Shum, S.B. (2006). Modelling discourse in contested domains: a semiotic and cognitive framework. *Int. J. Hum.-Comput. Stud., 64*(11),1154–1171.

Mann, W.C., & Thompson, S.A. (1987). Rhetorical structure theory: A theory of text organization. *Technical Report RS-87-190, Information Science Institute*.

Nonaka, I., & Takeuchi, H. (1995). *The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation*. Oxford University Press.

Ogden, C.K., & Richards, I.A. (1923). *The Meaning of Meaning: A Study of the Influence of Language upon Thought and of the Science of Symbolism*. Magdalene College, University of Cambridge.

Rector, A.L. (2003). Modularisation of Domain Ontologies Implemented in Description Logics and related formalisms including OWL. In *Proceedings of the 2nd International Conference on Knowledge Capture (K-CAP 2003)*, (pp. 121–129). ACM Press.

Tempich, C., Pinto, H.S., Sure, Y., & Staab, S. (2005). *An Argumentation Ontology for Distributed, Loosely-controlled and evolv-InG Engineering processes of oNTologies (DILIGENT)*. In ESWC 2005, (pp. 241–256), Heraklion, Crete, Greece, May.

Torroni, P., Gavanelli, M., Chesani, F. (2007) Argumentation in the Semantic Web. *IEEE Intelligent Systems, 22*(6).

ENDNOTES

1   http://debategraph.org/

2   http://microformats.org/wiki/cite-rel

3   http://neuroscientific.net/index. php?id=43

Tudor Groza received his BSc and MSc degrees in computer science from the Technical University of Cluj Napoca, Romania, in partnership with DaimlerChrysler AG Berlin. Since 2006, he has been a PhD student at Digital Enterprise Research Institute, National University of Ireland, Galway. His main research interests include methods for the representation and extraction of shallow and deep metadata from scientific publications, argumentation modeling and the social semantic desktop. He has written (and co-authored) a number of peer-reviewed international publications (including a book chapter) on the above-mentioned topics.

Siegfried Handschuh is a SFI Stokes Lecturer at the National University of Ireland, Galway (NUIG) and research leader of the Semantic Collaboration team at the Digital Enterprise Research Institute (DERI). Siegfried holds honours degrees in both computer science and information science and a PhD from the University of Karlsruhe. He published over 70 papers as books and journal, book chapters, conference, and workshop contributions, mainly in the areas of annotation for the Semantic Web, knowledge acquisition and social semantic collaboration.

John Breslin is currently a lecturer at the Department of Electronic Engineering in the National University of Ireland, Galway. He is also an associate researcher and leader of the Social Software Unit at the Digital Enterprise Research Institute at NUI Galway, researching semantically-enabled social networks and community portals. He is the founder of the SIOC project, which aims to semantically-interlink online communities. He has received a number of awards for website design, including two Net Visionary awards from the Irish Internet Association for the Irish community website boards.ie, which he co-founded in 2000. Breslin is a member of the IEI, IET and IEEE.

Stefan Decker is a professor at the National University of Ireland, Galway, and director of the Digital Enterprise Research Institute. Previously he worked at ISI, University of Southern California (2 years, research assistant professor and computer scientist), Stanford University, Computer Science Department (Database Group) (3 Years, postdoc and research associate), and Institute AIFB, University of Karlsruhe (4 years, PhD Student and junior researcher). He is one of the most widely cited Semantic Web scientists, and his current research interests include semantics in collaborative systems, Web 2.0, and distributed systems.