Adding Provenance and Evolution Information to Modularized Argumentation Models

Tudor Groza, Siegfried Handschuh, John G. Breslin
DERI, National University of Ireland, Galway, Ireland
firstname.lastname@deri.org

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. 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 paper, we describe an abstract model for argumentation which captures the semantics independently of the domain. Following a modularized approach, we also take into account additional important aspects of the argumentation, like the provenance information or its evolution (the temporal side).

1. 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-propositions 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 [14]. One way of achieving externalization is by using formal models (ontologies) to capture the argumentation. There exists an important number of argumentation models, most of them following the direction given by the IBIS methodology [10]. 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 paper, 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.

The rest of the paper is structured as follows: in Sect. 2 we introduce two possible use cases for our framework. Then, in Sect. 3 we detail an important set of requirements which should be covered by an abstract framework. Sect. 4 details our approach, while Sect. 5 revisits the two previously introduced use-cases by showing the framework's applicability in different settings. Before concluding in Sect. 7 we provide a comprehensive overview of the related work (Sect. 6).

2. Use Cases

As mentioned earlier, argumentation can be found and modeled in a variety of domains. We chose two such representative domains, having as a common point the presence of argumentative discussions and differing in other points, such as the type of domain knowledge or environment in which the argumentation takes place: (i) online communities, and (ii) scientific publishing.

2.1. Online Communities

Figure 1. Example of argumentative discussion in an online community forum

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. While some forum sites allow the use of icons to identify the type of replies that occur in a threaded discussion (see Fig. 1), very few make use of these identifiers to help users when they are finding for a particular type of response. In CERN's "Discussion" system from the 1990s, a person could not "just reply". 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.
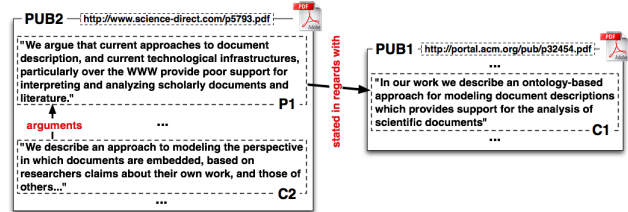
2.2. Scientific publishing

Figure 2. Example of argumentative discussion in different publications

Dissemination represents a communication channel between scientists, spanned across multiple publications. They make claims, state positions and argument these positions, creating a virtual argumentative discourse network. Fig. 2 depicts a typical example of such an argumentative discussion in this setting: an author makes a *claim in PUB1, while a second author, in PUB2, states a* position in regards to this claim and *arguments the position with an own* claim.

In order to realize a comprehensive model of the argumentation, one needs to know the authors involved in the discussion, the publications between which the discussion takes place and to be able to identify the actual elements in the publications which build the argumentative discourse network. By lowering the level of the analysis perspective, the publications are physically represented by documents, having multiple formats (usually PDF or Word) and usually scattered on the Web. Thus, transposing the same issues at this level, one needs to localize the document on the Web, to be able to represent the shallow metadata and chunks of texts present in the document.

3. Requirements

In this section we detail a set of requirements that emerge from both 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 [16]. 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 inter-linked 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. 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.

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 [3]).

Complementing argumentation with orthogonal models. Depending on the application domain, an argumentation framework should be easily complemented with orthogonal models for that domain. Considering the two use-cases previously introduced, examples of such models could be: SIOC [2] for social online communities, or RST (Rhetorical Structure of the Text) [13] for scientific publishing.
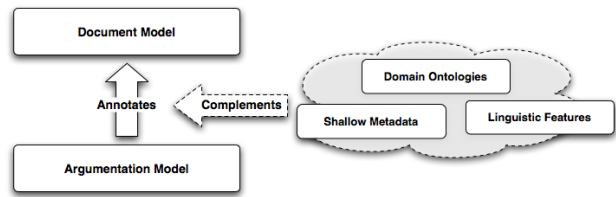

Figure 3. High level overview of the framework

4. Argumentation framework overview

Fig. 3 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.

4.1. The document model

The document model [1] provides an abstract way to capture information about documents and their structure, independently of their format. It is composed by three main concepts: (i) *Document*, the root concept, which acts as an information aggregator providing the entry point for describing the document in terms of its structure, provenance, or other aspects one wants to attach, (ii) *Composite*, the document's subdivision, providing the means for modeling the document's tree structure, and (iii) *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 have 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 modelling 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.
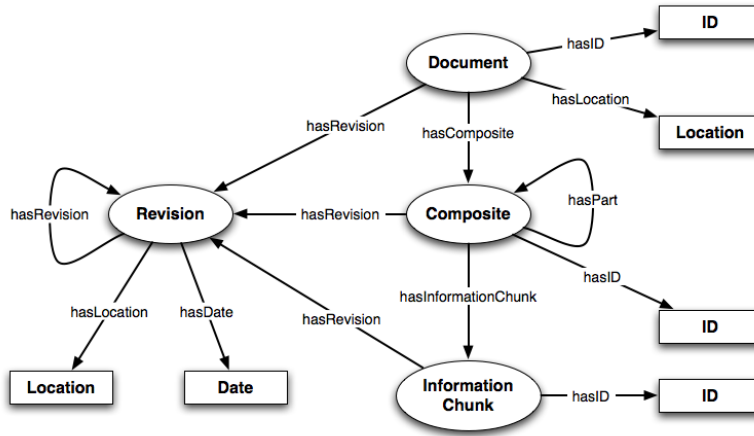

Figure 4. Detailed view of the document model

A document can have multiple revisions (created by different people), while a particular revision, can as well suffer changes, and therefore having revisions of its own. As an example, one could consider the case of a scientific publication edited by multiple authors. On the argumentation side, 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, that contain several paragraphs of text.

4.2. The argumentation model

Our argumentation model [2] follows closely the initial approach introduced by IBIS [10] and then developed further by DILIGENT [17]. As shown in Fig. 5, the model consists of four main concepts: *Statement*, *Issue*, *Position* and *Argument*. A typical argumentation flow would start with one expressing a *Statement*, which could raise an *Issue*, or about which others could state their *Position*s. A *Position* could take the form of a request, or could agree, disagree or be neutral in regards to an issue or a statement. The raised issues or the existing positions could then be argumented (*Argument*), either by providing a *Justification*, or

---

2http://ontologies.smile.deri.ie/argonto.rdfs

by creating 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.
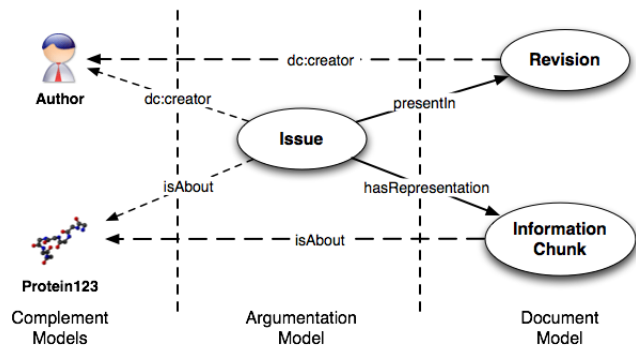
4.3. The model as a whole

Figure 6. Linking the document to the argumentation

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. Fig. 6 depicts an example of the way in which the different layers of the framework are linked. An issue is present in a particular document (or a revision of
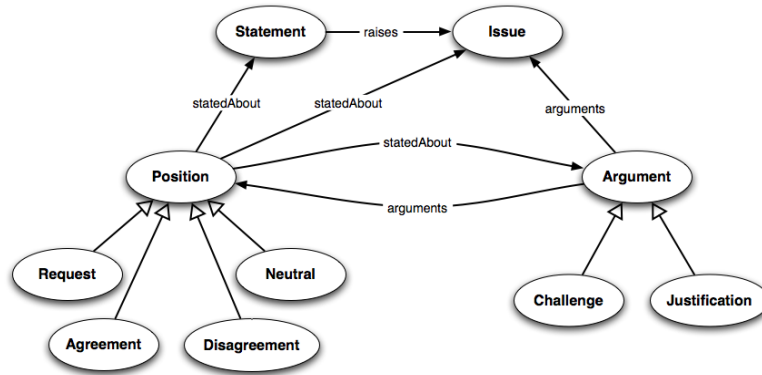
Figure 5. Detailed view of the argumentation model

that document) and has as representation in the document an Information Chunk. In a similar way, one can model a position, an argument or a statement.

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. 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.

5. Use cases revisited

In Sect. 2 we introduced two use cases to motivate the need for our framework. In the following we revisit these use cases and show the way in which one can model them by means of our argumentation framework.

5.1. Semantically-Interlinked Online Communities

Semantically-Interlinked Online Communities (SIOC) [2] 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.
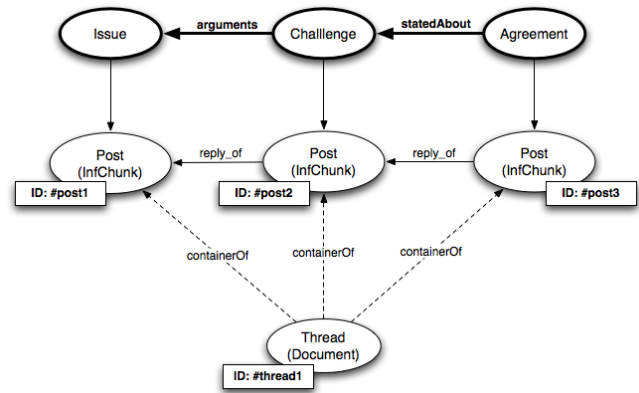

Figure 7. Example of argumentative discussion in online communities

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") for the existing classes (e.g. "Post", "Thread"), and then by relating these to the argumentation model terms described in Sect. 4. Some of these are illustrated in Fig. 7. At the abstract level, a *Thread* can be represented 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 (as in the scientific publishing model – see below). In such a case, the *Post* could be seen as a *Composite*, further decomposed into multiple *InformationChunk*s.

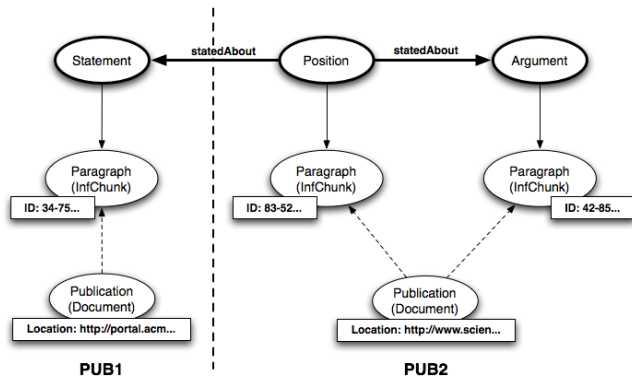## 5.2. Semantically-enhanced scientific publishing

Figure 8. Example model for argumentative discussion in scientific publications

SALT (Semantically Annotated LaTeX) [9] is a semantic authoring framework targeting the enrichment of scientific publications with semantic metadata. The goal of the framework is to define a clear formalization for externalizing the knowledge captured in the scientific discourses. SALT defines a special LaTeX markup syntax for annotating claims and arguments, and models the relationship between them by means of the SALT Rhetorical Ontology. The bridge between the rhetorical model and the publication itself is realized via the SALT Document Ontology. As most of the existing models in the area, SALT is focused on a particular domain, i.e. scientific publications. That is why, the Document Ontology describes publications in terms of *Section*s or *Paragraph*s. In addition, it follows a naïve argumentation model.

The solution proposed in Sect. 4 corrects the shortcomings of SALT. Firstly, it introduces a rich argumentation model that can be further customized in SALT and complemented with the other sides of the SALT Rhetorical Ontology. For example, one could envision an *Statement* as being represented by a *Claim*, which in turn could have attached an *Explanation* via a rhetorical relation. Secondly, the SALT Document Ontology represents a specialization of the currently proposed Document model, and thus enabling an easy mapping between itself and other models with similar goals.

Fig. 8 depicts a simplified modeling example of an argumentative discourse network spanned across multiple scientific publications. It shows how one can abstract from the SALT Document Ontology and also profit of the underlying document's format (i.e. PDF) and use the identification properties at the *InformationChunk* instances level as pointers inside the document. The advantage of such an approach is the lack of duplicate information. By having the identifier and the document, there is no need to attach the text of the Statement / Position / Argument to the *InformationChunk* instance. Similarly, one can take a different approach, in which the identifiers can be represented by XPath expressions (in the case of XML / HTML documents). Or, attach a Text property to one of the concepts and generate own identifiers.

6. 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: (i) background theories and models, which were used as inspiration by the majority of the currently existing argumentation models, (ii) existing argumentation models, having similar goals with our approach, and (iii) 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) [10]. IBIS introduced a methodology for argumentation-based decision making in information systems, adopted by most of the current ontology-based argumentation frameworks, like DILIGENT [17], Compendium methodology [12] or SALT [9]. A second important background theory is the Speech Acts Theory [5] 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 modelling), 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 [4] 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 [6]. This was following closely the original IBIS model and applied its methodology in team-based deliberation. gIBIS served as inspiration for later models like: (i) DILIGENT [17], which applies argumentation in ontology engineering, (ii) Compendium [11], that follows a semiotic [15] approach for

dealing with knowledge visualization and design rationale, while complementing argumentation with Cognitive Coherence Relations [12], or (iii) The Zeno argumentation framework [8] applied in mediation systems. Other relevant argumentation models include the one proposed by Torroni et. al in [18] 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 [1].

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 a innovative way for attaching domain or liguistic 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 modelling revisions of the 'hosting' environment.

7. Conclusion

The main goal of the abstract argumentation framework presented in this paper 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 modelling 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 [7], 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.

Acknowledgments

References

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

[2] J. Breslin, S. Decker, A. Harth, and U. Bojars. SIOC: An Approach to Connect Web-Based Communities. *The International Journal of Web-Based Communities*, 2(2), 6 2006.

[3] P. Buitelaar, T. Declerck, A. Frank, S. Racioppa, M. Kiesel, M. Sintek, R. Engel, M. Romanelli, D. Sonntag, B. Loos, V. Micelli, R. Porzel, and P. Cimiano. LingInfo: Design and Applications of a Model for the Integration of Linguistic Information in Ontologies. In *Proceedings of OntoLex06, a Workshop at LREC, Genoa, Italy*, 2006.

[4] L. M. Carlson. *Dialogue Games: An Approach to Discourse Analysis*. D. Reidel Publishing Company, 1983.

[5] P. R. Cohen and C. R. Perrault. Elements of a Plan-Based Theory of Speech Acts. *Cognitive Science*, 3:177–212, 1979.

[6] J. Conklin and M. L. Begeman. gIBIS: A Hypertext Tool for Team Design Deliberation. In *HYPERTEXT '87: Proceeding of the ACM conference on Hypertext*, pages 247–251. ACM Press, 1987.

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

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

[9] T. Groza, S. Handschuh, K. Möller, and S. Decker. SALT - Semantically Annotated LaTeX for Scientific Publications. In *ESWC 2007*, Innsbruck, Austria, 2007.

[10] W. Kunz and H. Rittel. Issues as elements of information system. Working paper 131, Institute of Urban and Regional Development, University of California, 1970.

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

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

[13] W. C. Mann and S. A. Thompson. Rhetorical structure theory: A theory of text organization. Technical Report RS-87-190, Information Science Institute, 1987.

[14] I. Nonaka and H. Takeuchi. *The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation*. Oxford University Press, 1995.

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

[16] A. L. Rector. 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)*, pages 121–129. ACM Press, 2003.

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

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