Bringing Discussion to Documents for the Creation of Richer Metadata

Michihiro Yamamoto and Ng S. T. Chong
{Yamamoto, Chong}@unu.edu
Media and Technology Laboratory
United Nations University

Abstract

Web documents are often the shared context of discussion in group collaborations. A typical example is collaborative peer review on the Web. Reviewers discuss and share their comments online. In other cases, the ongoing discussion itself can be the shared context for collaboration, like when people refer to and discuss around an existing issue in a Web forum. The annotations created in the form of discussion describe the shared context of collaboration, and hence are an importance source of metadata for the Web. FlexNetDiscuss (FND) was previously developed to facilitate this natural process of metadata creation and sharing. FND brings discussion to documents by enabling dynamic creation and sharing of advanced hyperlinks, within and between the discussion and HTML documents. Unfortunately, such metadata based on messages from discussions is represented in a format that is neither standard nor facilitates machine readability. We describe our efforts to extend FND to preserve the created links in a form that is manageable by RDF-aware applications and ready for RSS feeds to the interested users, who in turn can provide feedback through the system for further enrichment of the metadata. The system runs on top of any browser that handles DOM and XML.

Background

The Semantic Web has been defined as 'an extension of the current Web in which information is given well-defined meaning, better enabling computers and people to work in cooperation.' [Hendler et al, 2002] This extension seeks to build a global distributed database through the integration of data from independent communities without requiring each community to deliver data in a common structure. While each community is expected to contribute to the creation of metadata for Web resources, the relations among the different communities and their ideas are often ignored. It happens very often that knowledge originated in one community can have a relevant impact on ideas expressed by another community. Mining these relevant relationships and making them apparent to the users can help forge links within a community and across communities. This prompts for a framework that would allow us to capture Web discussions and their associated shared discussion contexts into a common representation, which could be used to produce rich metadata for such purposes as quantifying the strength of relationships within and between HTML documents and discussion messages, ranking contents, and identifying knowledge hotspots.

Objective

Our aim is to develop a framework that allows the creation of rich metadata for Web documents, which serves as valuable resources for the Semantic Web. The fact that metadata creation is a complex task requiring careful analysis of the resources makes the automated generation of metadata hard to achieve. Therefore, more human intervention is required for the creation and refinement of the metadata, which drives the need for the development of the effective collaboration environment. Our efforts in bringing discussion to documents serve to encourage users to actively participate in the metadata creation process, where the metadata collected by individual users is utilized during collaboration. Our challenge is to bring the individually collected metadata into the collaborative review for further enrichment of the metadata, while maximizing the potential of the capabilities and extensibility in FND [Chong, 2003a] [Chong, 2003b].

Semantic information collection

FND provides annotation support to enhance discussion around HTML documents, by enabling linking capabilities at the message level, where any message can be bi-directionally linked to a shared context that resides either in a collection of documents or the discourse itself.

Annotations in our world consist of the following parts: the link to the specified Web resource, link type and direction, location information in a discussion forum, and additional metadata related to the annotation. Links are implicitly bi-directional in the sense that the user only needs to specify the relationship in one direction. Bidirectional links serve to creating richer metadata by expressing multi-level contexts of discussion and multi-level relationships with Web resources.

Although annotations create meaningful relationships among Web resources, users can easily get lost in the highly linked collaboration space as the links grow in size and levels. To solve this problem, we propose an index structure called ObjectIndex to assist users with navigation along the links from the discourse space to the document space, and conversely, from the document space back to the discourse. The ObjectIndex groups the discourse-documents relationships hierarchically based on the discussion elements or page elements.

Figure 2 shows the organization by discussion elements, where the message in 'forum1' has a forward-link to the message in' forum0', and the message in 'forum0' has a backward-link to two messages. Forward links of type message are represented with an 'M' character and their backward link counterparts are represented with the same character displayed in inverse.

In the implementation, the ObjectIndex is displayed at the right top portion of the discussion space (as shown in Figure 3) and provides an intuitive interface to navigate the bidirectional links created between the Web resources, while enabling each link to be represented as RDF metadata [RDF, 2004].

Figure 2. Concept of ObjectIndex

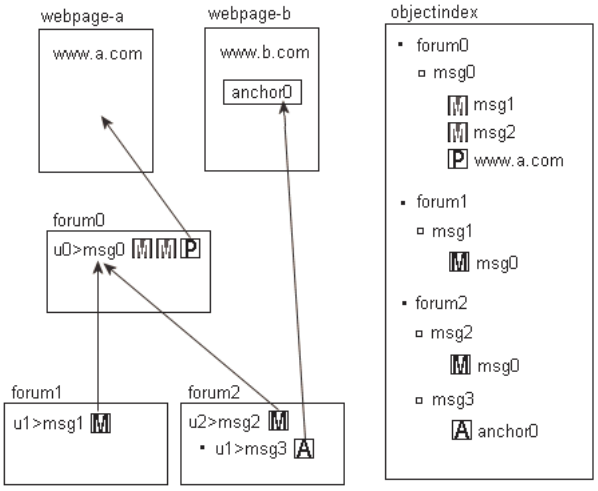
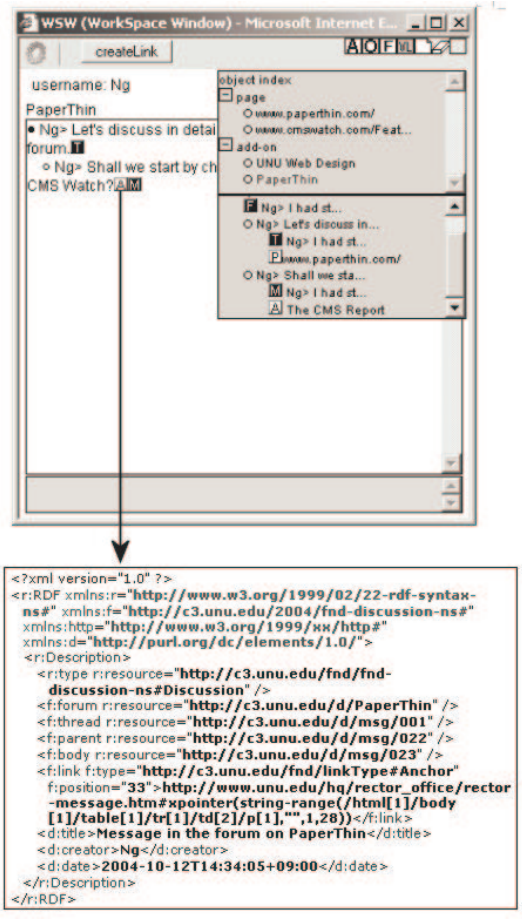


Figure 3. ObjectIndex GUI



Linking structure description

There have been many efforts to describe the relationships between resources using RDF and XLink [XLink, 2001]. In the RDF framework, metadata is described using expressions comprised of collections of resource-property-value triples, which can be represented with directed graphs and URI references. RDF shares the same functionality as XLink in that both of them are used for describing the relationship between Web resources while more advanced support for describing the complex linking structures is available only in XLink.

We have applied and implemented the concepts in XLink for describing the bidirectional linking structure that connects annotated Web resources. The bidirectional link between two Web resources (as shown in Figure 3) can be represented through XLink elements (as shown in Figure 4), where the annotation holds two arc-type elements and each element serves to connect the same two messages in the opposite direction. To produce RDF metadata from this model, we devised a way to separate each arc-type element in an individual RDF model using our designed vocabulary 'fnd' as shown in Figure 5.

Metadata syndication

While the metadata collected through document-centric discussion can effectively describe the contexts of Web documents, the description quality is not always high enough for users to judge the relevance or freshness of the Web documents, which motivates us to let more users review and comment on the metadata by extending FND to have a feedback prompting mechanism.

Instead of publishing all the available metadata of Web documents to the Web, we let each metadata reach the most suitable audiences by using RSS feeds syndication [RSS, 2001], where users can subscribe arbitrary Web documents and get notified each time the update occurs in the associated contexts. Users can contribute to enriching the metadata by responding to the discussion messages, which can be browsed and commented through a preferred RSS reader.

Figure 3. Link between two messages

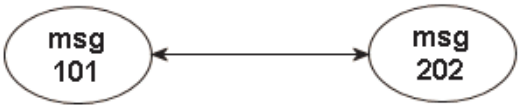


Figure 4. Link modeled through XLink element

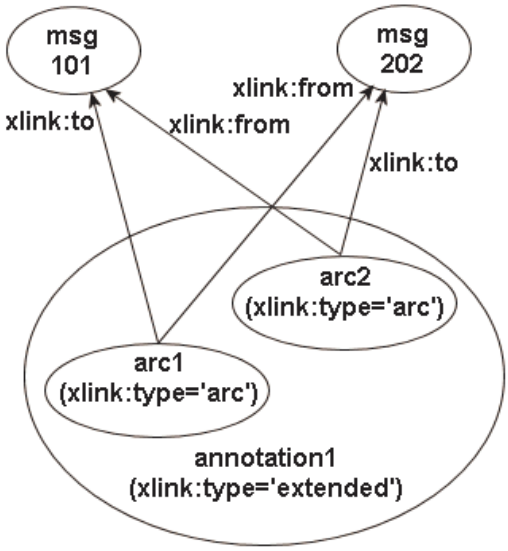
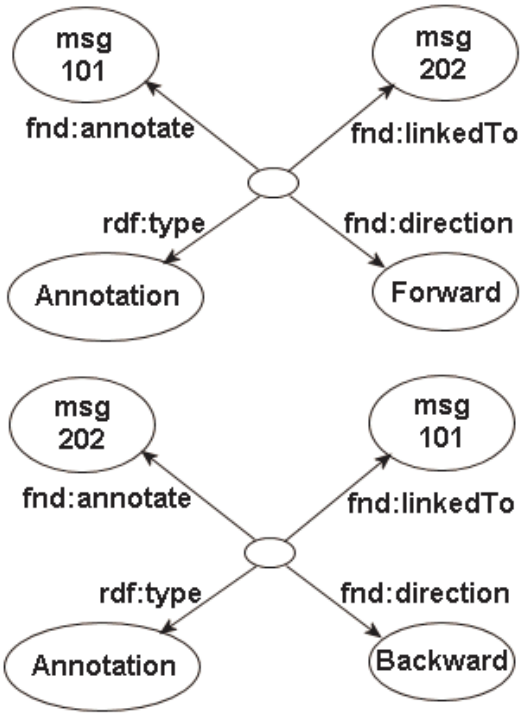


Figure 5. A bidirectional link modeled through two sets of the FND RDF model



Concluding remarks and future directions

Our framework enhances current online discussion systems with advanced annotation and linking features. It also facilitates the discovery of complex relationships among Web resources and the associated semantic information as defined by the annotations that result from group collaborations.

While making these relationships explicit and encoding the metadata in a generic representation is an important step to create a bridge between Web mediated communications and the Semantic Web, the rich metadata created does not necessarily serve the information needs of the participants of the online communities. To bridge this gap, as a future direction we would apply the metadata generated in our framework for ranking knowledge hotspots and facilitate its use as hints for generating recommendations matching the individual preferences and interests of users, using collaborative filtering techniques.

References

[Chong, 2003a] Chong, N, FlexNetDiscuss - A Model and Architecture for Collaboration using Two-way Associative Hyperlinks, 14th ACM. Conference on Hypertext and Hypermedia, 2003.
[Chong, 2003b]Chong, N. FlexNetDiscuss: a Generalized Contextual Group Communications System to Support Learning Inside and Outside the Classroom, 3rd IEEE International Conference on Advanced Learning Technologies, July 2003.
[RDF, 2004] RDF Primer, http://www.w3.org/TR/rdf-primer/
[RSS, 2001] RDF Site Summary (RSS) 1.0, http://web.resource.org/rss/1.0/spec, 2001.
[XLink, 2001] XML Linking language (XLink) Version 1.0, http://www.w3.org/TR/xlink.
[Hendler et al, 2002] Hender, J., Berners-Lee, T., Miller, E., Integrating Applications on the Semantic Web, Journal of the Institute of Electrical Engineers of Japan, Vol 122(10), October, 2002, p. 676-680.