

Metadata Based Annotation Infrastructure offers Flexibility and Extensibility for Collaborative Applications and Beyond

Marja-Riitta Koivunen

World Wide Web Consortium
MIT Laboratory for Computer Science
marja@w3.org

Ralph Swick

World Wide Web Consortium
MIT Laboratory for Computer Science
swick@w3.org

ABSTRACT

In this position paper, we describe three user scenarios that benefit from metadata based annotation infrastructure. We explain how a basic annotation schema can be extended to support new scenarios. We also describe and evaluate some other features and modifications that are useful when implementing these scenarios. The most laborious part in the scenarios is the design and implementation of new user interfaces; the metadata infrastructure itself easily supports the needs of the different applications and new schemas.

Keywords

Annotation infrastructure, metadata, collaboration scenarios

1 INTRODUCTION

The World Wide Web is a collaborative space that lets users share their thoughts, their work, their images, and other aspects of their life by publishing Web pages. But publishing is not enough; feedback and interaction is needed for collaboration. E-mails and netnews distributed and archived in discussion lists are two of the earliest and most important collaborative applications of the Internet. Other applications such as irc¹, Netmeeting², and "buddy list" applications provide real time sense of presence, communication and sharing of resources.

Sharing content through Web pages is important but also is limited as readers can seldom share comments or questions by writing back to the pages, even when they are members of a closed collaborative group. Instead, with the Web today we still observe much effort spent by users on forming and trying to understand different e-mail conventions for commenting on documents that are on-line in the Web.

Shared annotations that do not require write access to the annotated page can support very rich communications about the Web pages. When these annotations are seen as metadata about the pages or parts of them, and when the metadata vocabulary is grounded in semantically rich ontologies that are themselves published in the Web, a lot

of possibilities that extend beyond basic annotation capabilities are opened.

This paper describes a simple collaborative annotation scenario and then broadens the scope of the annotations in a couple of additional scenarios. We briefly explain the basic metadata infrastructure for annotations that is provided by our system, known as Annotea [1], and the features that are needed to support the additional scenarios.

2 SCENARIOS

We present three scenarios describing the use of annotations in different illustrative contexts. The first scenario explains the use of annotations for basic collaboration, the second one shows an interpretation of shared bookmarks as annotations, and the last scenario examines the use of annotations for communicating evaluation results.

2.1 Scenario: Using Annotations for Collaboration

University of Oslo organizes a seminar focusing on writing research reports and collaboration. The goal of the seminar is not only to produce a report but also to learn from other students' use of research methods and collaborative techniques and their approaches to problem solving.

One student group elects to write a report on the communication of whales. They collaborate by using the Web to publish new material, to search and share hypertext links to references and to annotate the material they uncover. The group's discussions of their research material is facilitated through a threading mechanism that links together some of their annotations in chronological order. They use an annotation (metadata) server dedicated to this seminar in conjunction with other annotation servers to which they normally subscribe.

The group gathers lists of references on a shared Web page. The lists include an estimation of the papers' relevance and a preliminary categorization of the reference. As the students read each paper they mark the paper as interesting or uninteresting and refine the categorization. They use annotations to mark or question unclear text, point out interesting perspectives, add keywords and share other general comments with each other.

¹ <http://www.ietf.org/rfc/rfc2812.txt>

² <http://www.microsoft.com/windows/netmeeting/>

Later they dedicate one person to write more detailed replies to selected research questions pointed out in the annotations and write a short summary. This starts fruitful discussions in the context of the reference document and the new summaries. By using annotations to conduct their commentary on their reading, the group avoids contention for write access to a single shared document and potential loss of data from conflicting updates.

2.2 Scenario: Using Annotations for Shared Bookmarking

In the first stage of gathering references for their report on whale communication, the group uses traditional Web search tools to locate references on the Web. They create 'bookmark' annotations in their dedicated seminar annotation server to those references that appear relevant. When they create these bookmarks they also select a category from a list of categories defined by a shared ontology or, if no existing category is a good match, they define new categories, adding each such category to a special seminar ontology that is stored in their shared Web space. The classification category is more metadata about the bookmark annotation, one of a variety of such extensions that the group can store with their metadata.

When a user goes to a bookmarked page she sees the existing bookmarks as annotations. The user can also ask for a list of bookmarks, in which case, a page is dynamically created showing bookmarks under different categories. The user may query all the bookmark annotations on the annotation servers or filter the list to show only certain bookmarks. The user may also ask for just the bookmarks that belong to the concepts in a given ontology.

2.3 Scenario: Using Annotations to Present Evaluation Results

Kim is a teaching assistant in a collaborative seminar. He wants to make sure that the students remember that the readers of their documents may have different physical or cognitive abilities in receiving and interacting with the information. Kim uses the Web Accessibility Initiative³ guidelines and some automatic tools for assessing the markup used within Web pages. These accessibility assessment tools rely on EARL, a metadata language expressing what is or may be wrong in a page, citing by URI the specific guideline that describes the accessibility issue.

Kim stores the EARL analysis of each document in the same annotation server that holds the seminar's other annotations. Kim also adds to the server some inferencing rules that represent a transformation from the EARL

vocabulary to the annotation vocabulary. The EARL vocabulary is a superset of the annotation vocabulary, so Kim includes some style rules that instruct presentation clients in the rendering of the extra properties of the EARL metadata.

When students view their pages they see the EARL report items as annotations on the pages as a result of processing the inferencing rules. Now they can address the accessibility issues in the pages and add additional metadata to the annotations to note them as fixed or to request help from Kim. When Kim helps the group, he sends a mail to the mailing list explaining the problem and adds a link to the EARL annotation so that others in the group can benefit from the example.

When the work is done the group can run the accessibility evaluation tools again. The document author can choose to delete the earlier report annotations at this time or she may just mark them as obsolete. The group may also freeze a copy of the evaluated page with the original annotations.

3 ANNOTEA METADATA INFRASTRUCTURE

The metadata infrastructure of the Annotea project makes it easy to support the annotation scenarios presented above. The Annotea infrastructure provides flexibility and an easy framework to extend the annotation capabilities to other applications. The basic infrastructure and the extensions needed for the previous scenarios are discussed in the following sections.

3.1 Basic Annotea Annotations

In the first scenario, the students annotate Web pages and use reply threads as supported by the Annotea infrastructure.

Annotea sees annotations as metadata about a whole document or a part of a document. This metadata is written in RDF/XML [2], and can be stored in annotation servers using the HTTP protocol. An annotation client queries annotations related to a document from one or several annotation servers and presents them in document context.

The Annotea annotation model uses multiple RDF schemas e.g. Dublin Core⁴ (*dc:*) with the Annotation schema to define the basic annotation properties (see Figure 1). The *annotates* property refers to the annotated document, the *context* property refers to the actual place of the annotation within the document, the *body* property contains the content of the annotation, the *dc:title* property is a descriptive annotation title. The other properties further describe the annotation.

⁴ <http://dublincore.org/>

³ <http://www.w3.org/WAI/>

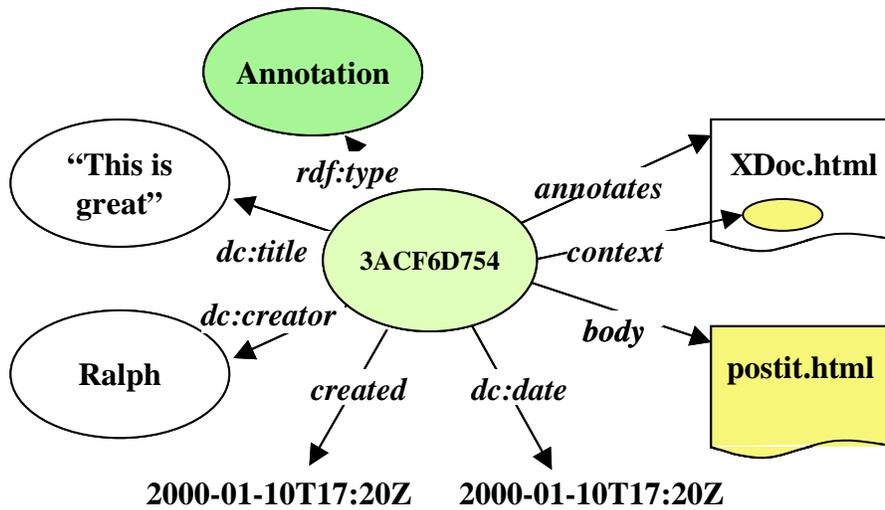


Figure 1: The basic annotation schema

With RDF it is also easy to add new properties to the annotations. The DAML+OIL⁵ ontology construction vocabulary [3] provides a framework for describing new properties with precise semantics and placing those semantics in the Web.

⁵ <http://www.daml.org/2001/03/daml+oil-index>

3.2 Extending the Annotation Schema for Reply Threads

Annotea has a concept of a reply that relates to an annotation or another reply. Replies can form discussion threads that start from an annotation.

The reply schema looks similar to an annotation schema. It has two new properties, the *reply-to* property, which defines which annotation or reply was the previous one in the thread, and the *root-of-thread*, which is the first annotation in the thread. The generic metadata-based design of our annotation server made it easy to incorporate these additional properties.

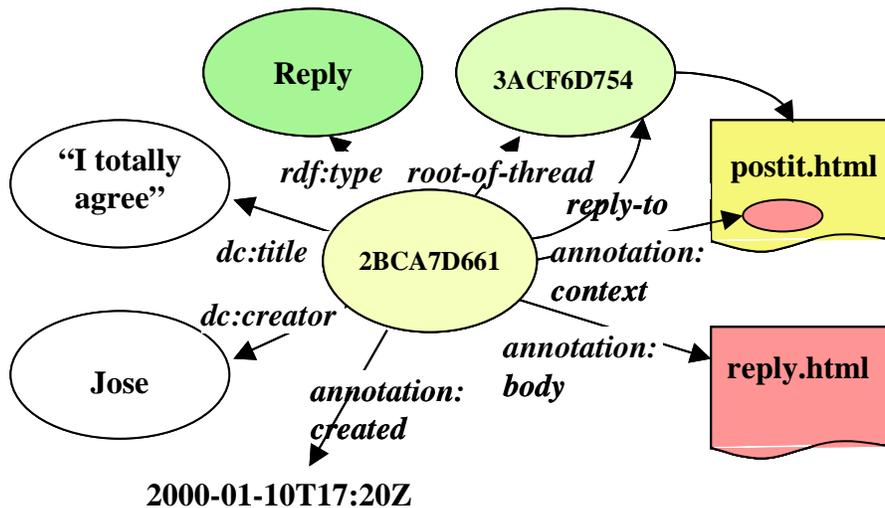


Figure 2: The reply schema

3.3 Using Annotea for Shared Bookmark Annotations

Shared bookmarks can be easily seen as annotations of type bookmark. In addition, they need a *category* property. Again, no changes are needed to our annotation server. The

bookmark annotations can be presented as annotations on the pages with a special icon to visually differentiate them. For that an icon property can be added to the metadata.

Addition of new properties for annotation schemas necessitates a user interface change so that the client can present them. The presentation style for a property can be described in the same metadata framework as properties of properties. We expect to work on a schema for describing presentation characteristics as part of future development. Existing ontology construction applications provide user interfaces for ontology definitions and these are well suited to the definition of categories for classifying bookmarks.

The generic metadata approach to describing bookmarks naturally lends itself to supporting a variety of views on the bookmark database. User-customizable queries can select bookmarks by any criteria desired.

3.4 Accessibility Evaluation Report Items as Annotea Annotations

Annotations can also be used to present automatically generated report items, such as accessibility evaluation items or markup validation items. If the report items are described in the metadata format it is straight-forward to map them to an annotation schema. For instance, the EARL report item reporting an accessibility problem has semantics that map easily into an annotation of a part or the whole of the evaluated Web page. This mapping can be expressed as a collection of inference rules over the properties produced by the EARL tools.

The generic metadata framework provides the necessary flexibility to decide on a case by case basis whether to archive, delete, or revise annotations when a document is reprocessed through the evaluation tool. The tool can maintain state information for successive runs in the same metadata store.

4 CONCLUSIONS

A metadata based annotation infrastructure such as Annotea can easily support a broad range of different annotation needs. The generic property mechanism of RDF allows us to construct ontology-neutral data stores. Applications can

use several ontologies simultaneously to describe different aspects of their annotations.

Most work in the scenarios is needed in the customization of the user interfaces for the different annotation applications. More research is needed to ease the presentation of the metadata, especially new properties from ontologies the application (or user) may not have previously seen.

The RDF model provides a convenient mechanism on which to layer client-side or server-side inferencing for mapping between ontologies. Further work to build effective end-user tools to take advantage of this capability is in progress.

ACKNOWLEDGMENTS

We thank Jose Kahan, Eric Prud'hommeaux, Art Barstow, Eric Miller and other W3C staff for their many ideas that have contributed to this paper. We also thank Charles McCathieNeville for his help when we were experimenting with EARL scenarios.

REFERENCES

1. José Kahan, Marja-Riitta Koivunen, Eric Prud'Hommeaux, Ralph R. Swick, [Annotea: An Open RDF Infrastructure for Shared Web Annotations](#), in Proc. of the *WWW10 International Conference, Hong Kong, May 2001* (<http://www10.org/cdrom/papers/488/index.html>).
2. Ora Lassila and Ralph R. Swick, eds., [Resource Description Framework \(RDF\) Model and Syntax Specification](#), W3C Recommendation, February 1999 (<http://www.w3.org/TR/1999/REC-rdf-syntax-19990222>).
3. Frank van Harmelen, Peter F. Patel-Schneider and Ian Horrocks, eds, [Reference description of the DAML+OIL \(March 2001\) ontology markup language](#), Joint United States / European Union ad hoc Agent Markup Language Committee (<http://www.daml.org/2001/03/reference.html>).