Semantic Authoring By Tagging with Annotea Social Bookmarks and Topics

Marja-Riitta Koivunen

Annotea

marja@annotea.org

ABSTRACT

Annotea is a framework that supports Semantic Web based collaboration and authoring via Annotea objects, such as social annotations, bookmarks, and topics. Social bookmarks and topics can be used for semantic authoring in many levels. They let ordinary users tag interesting Web documents with their own personal concepts (folksonomies). These concepts can be simple or form hierarchies, and they can be linked to more standard concepts when users become aware of them. Collections of the concepts and the bookmarks under them can be stored to bookmark documents or services as metadata and easily shared with other users. Annotea users can also literally semantically author pages with link collections and categories. They can use the title and description information of the topics and bookmarks to create the page content. This content from the bookmark and topics hierarchies is then easily converted to XHTML Web pages by using XSLT scripts. Annotea presentation objects further support authoring and customized presentations by separating the presentation and the application data. While Annotea precedes tagging and social bookmarking services and tools popular today it still provides better flexibility, extendibility, and reuse of the bookmark and topic data with help of the Semantic Web technologies.

Categories and Subject Descriptors

H.5.3 [**Group and Organization Interfaces**]: Computer-supported Cooperative Work, Web-based Interaction; I.2.4 [**Knowledge Representation Formalisms and Methods**]: Semantic Networks – *Semantic Web*

General Terms

Design, Human Factors

Keywords

Social bookmarks, social topics, Semantic Web, semantic authoring, ontologies, tagging, folksonomies.

1. INTRODUCTION

Annotea is a framework for semantic annotations [26], bookmarks, and topics focusing on sharing of the objects and collaboration in document context[18]. Annotea [4, 17] gets its inspiration from observing users' needs and problems and developing user scenarios based on those observations [18]. Annotea sees humans as meaning makers. They create concepts, make associations, comment, clarify, annotate, review, organize, bookmark, tag, and file information into categories or organize it into folders. When humans learn something new they augment and modify their existing cognitive models with new personal concepts. Typically, these concepts are vague at first but as they become more precise they can be shared with other users and developed further collaboratively.

Ordinary users who are not familiar with semantic Web can do semantic authoring with intuitively familiar Annotea objects, such as annotations, discussions threads, bookmarks, and topics (tags). For instance, Annotea social bookmarks allow users to recall earlier visited Web resources, such as Web documents, services, music, images and blogs. As a side effect, the user also authors metadata about the bookmarked resources. When published and shared, this information can be useful not only for the user but also for other users and applications.

Users can also choose to use Annotea to author and maintain Web pages consisting of link collections and categories. In this case, users semantically author bookmark and topic information, such as title and description information, which can be used to form Web pages that were earlier manually written and maintained.

Annotea topics [18, 19, 21, 16] support user defined simple concepts (tags). Instead of forcing the users to invest in learning standard ontologies before they can start developing their ideas, Annotea supports the innovation process by letting users start from their own personal ontologies (folksonomies). Gradually, when the user learns about connections her concepts and standard, commonly used concepts, it is possible to link the personal topics to them.

Annotea uses the Semantic Web technologies for implementing the solutions but hides the Semantic Web from the ordinary users under the familiar objects and easy user interfaces. From the user point of view, the Annotea topics can be used for Web tagging just like in the popular Web tagging services [11, 13] that followed Annotea. However, the Semantic Web metadata basis [9] makes Annotea more flexible, easily extendible, and interoperable supporting user needs and applications also beyond tagging. The user can select bookmark stores relevant to

her tasks and seamlessly merge Annotea objects from these stores to support task related queries.

Annotea objects can be presented to the users in many innovative ways and by different tools. Typically, the objects are explored in the context of the Web resources they refer to, examined in selected topic hierarchies, and searched by using different properties. Amaya browser/editor [1] and lately Annotea Ubimarks [7] have been used to demonstrate some capabilities of Annotea objects. In addition, many other clients, such as Annozilla [8], have been developed outside the Annotea team. The bookmark files can also be presented in normal browsers as XHTML pages by using selected XSLT style sheets.

In Annotea the bookmarks and topics can be stored locally or shared both with individual users on different devices and with different user groups. Unlike simple tagging or bookmark formats, such as XBEL [27], Annotea bookmarks and topics help collaboration and finding relevant information even when users use different terms for similar concepts. This is is often the case in cross disciplinary fields, such as life science, user experience or even in marketing and research where users are awarded from finding original concepts instead of using the already existing ones. While services, such as Linkagogo [22] help in interpreting between different formats, the simple formats often don't have enough identification information to easily separate or merge concepts.

In addition to supporting tools for creating and retrieving bookmarks, the Annotea bookmarks can be used as user profiles for different services. This helps users find the information relevant to them at a certain context. Annotea bookmarks and topics can also contain information not just about the user but his or her social network. Again, this can be used to support the finding of more relevant information. Bookmarks and topics can also be used in applications organizing blog type information.

Finally, as shared bookmarks are easy to create we hope to eventually see enough commonly available bookmark and topic data to be able to see network effects in applications using the published bookmark repositories collectively. These applications can provide more relevant results. Furthermore, when data mining and search engines type applications mine the topics and different connections between the topics it is possible to make many new innovative discoveries connecting fields that were not aware of each other because they used different terminologies.

2. ANNOTEA ARCHITECTURE

Different annotation and tagging applications can be built on top of Annotea architecture that includes Annotea stores, Annotea objects, such as the bookmarks and topics, and protocols.

Annotea stores are RDF metadata stores that store Annotea objects. Objects from various stores can be easily merged and viewed together. Many currently popular tagging services violate the principle of user control and do not give users a choice of where to store the bookmarks and topics. The Annotea objects can be stored in local files, global Web servers and/or Web documents containing the metadata.

Web documents offer users an easy alternative to get started without a server. The same authentication and protection mechanisms can be used as for any other Web document. A document can also be used to archive snapshots of the selected Annotea objects stored on the server. For instance, a version of a document can be stored together with annotations for that version.

Annotea objects have URIs and properties. They all store some basic tracking information, such as who created them and when. The Annotea schemas [2, 3, 6] define only a few specifically Annotea related properties for Annotea objects. To improve reuse and decrease unnecessary inferencing the rest of the properties are adopted from other well-established schemas, such as Dublin Core. Annotea usage documents [20] and server protocols [25] help define the basic annotation/bookmarking framework [14] for Annotea objects. Finally, model user interfaces [1,7] demonstrate the use of the objects.

New properties can be easily added to Annotea objects. For instance, Annotea Ubimarks lets the author add a URI to his personal and social network information by using *foaf:maker* (see **Figure 1**). When provided, this social network information can be used as additional help when trying to find related documents.

Annotea [14, 17] has demonstrated storing annotations and replies to different annotation servers. Annotea protocol for the metadata stores is based on HTTP. Annotea bookmarks and topics can use a similar protocol or standard Web queries. We have done some experiments with this. Currently, the model implementations use HTTP, FTP, and possibly also other Web protocols to store the bookmarks and topics as Web documents. This approach can be applied for annotations and replies as well.

Annotea applications and user interfaces provide different views of the Annotea objects. When the interfaces are attached to browsers users can seamlessly collaborate via Annotea objects in the context of the Web documents or other Web resources. In addition, Annotea objects can be used in many other kinds of applications and user interfaces. For instance, bookmark collections on a selected topic can be used as user defined profiles for services or bookmark repositories for selected domains can be used to order the results from search agents, such as Yahoo Search or Google, in a more meaningful way for the user [24].

3. ANNOTEA BOOKMARK AND TOPIC OBJECTS

While the first phase of Annotea development concentrated on annotations, early on many user scenarios were developed, presented, and discussed about social bookmarks and topics (tags). These were implemented during the second development phase, when the need for categories was high on the users' wish list. Especially status categories were needed to mark the processed annotations.

We wanted to make sure that Annotea could easily support not only annotation categories but also our other scenarios related to categories. We also wanted an easy user metaphor that would work well with the already existing metaphors. As a result, we decided to implement the social bookmark and topic objects as a variation of the semantic annotations [26]. Our goal was to experiment with these objects and use the experiences later to also improve the annotation implementations.

```
<?xml version="1.0"?>
<RDF:RDF xmlns:foaf="http://xmlns.com/foaf/0.1/"
    xmlns:dc="http://purl.org/dc/elements/1.1/"
    xmlns:bm="http://www.w3.org/2002/01/bookmark#"
    xmlns:a="http://www.w3.org/2000/10/annotation-ns#"
    xmlns:RDFS="http://www.w3.org/2000/01/rdf-schema#"
    xmlns:RDF="http://www.w3.org/1999/02/22-rdf-syntax-ns#">

  <bm:Bookmark RDF:about = "http://mydomain.org/AnnoteaURI/exBookmark"
      dc:title = "The Whole Brain Atlas"
      dc:date = "Mon, 15 May 2006 14:28:32 GMT"
      a:created = "Mon, 15 May 2006 14:28:32 GMT"
      dc:description = "The Whole Brain Atlas is intended as
        an introduction to basic neuroanatomy, with emphasis
        on the pathoanatomy of several leading central nervous
        system diseases."
      dc:creator = "Marja">
    <bm:recalls RDF:resource =
      "http://www.med.harvard.edu/AANLIB/home.html"/>
    <foaf:maker RDF:resource =
      "http://www.annotea.org/marja#marja"/>
    <bm:hasTopic RDF:resource = "http://mydomain.org/AnnoteaURI/exTopic"/>
    <RDFS:seeAlso RDF:resource =
      "http://big.sfn.org/NDG/site/eavData.asp?o=28952"/>
  </bm:Bookmark>

  <bm:Topic RDF:about = "http://mydomain.org/AnnoteaURI/exTopic"
      dc:title = "Brain collections"
      dc:date = "Mon, 15 May 2006 14:23:32 GMT"
      a:created = "Mon, 15 May 2006 14:23:32 GMT"
      dc:creator = "Marja">
    <foaf:maker RDF:resource =
      "http://www.annotea.org/marja#marja"/>
    <bm:subTopicOf RDF:resource = "http://mydomain.org/AnnoteaURI/exParentTopic"/>
  </bm:Topic>
</RDF:RDF>
```

Figure 1. A sample Annotea bookmark and topic as RDF.

The bookmark metaphor was great for our purposes. Ordinary users are not only familiar with it but have often used traditional bookmark interfaces. Furthermore, the bookmark user interfaces have a lot of enhancement possibilities that benefit from the Semantic Web approach. For instance, the user interface can utilize the document context to find similar bookmarks and topics..

3.1 Annotea Object Identification

Like other Annotea objects, the bookmark and the topic objects have their own identification URI. This URI ensures that Annotea objects retain their identity globally even when objects are merged together from different data stores, or moved to another location along with a bookmark file, or sent to another user by e-mail.

Not all users have their own Web domains. Therefore, the current Annotea Ubimarks implementation uses URIs that are unique UUIDs, such as "urn:uid:025f3d84-57bd-11da-924d-b619a79a047d". The UUIDs can be easily created without owning a Web domain. Resolving the objects content based on the UUID would need more work, but that is not usually a problem. .

Alternatively, the user can choose HTTP addresses in her own domain with date info by defining an URI path, for instance, "http://mydomain.org/AnnoteaURI/". Similarly, a bookmark server can create URIs. If the URI changes in the server or for some other reason, the correspondence between two Annotea objects can be defined by using the owl:sameAs property.

3.2 Bookmark and Topic Properties

Annotea does not define all bookmark and topic properties by itself. It uses properties from well known vocabularies, such as Dublin Core. Applications already familiar with these properties know how to deal with them also in Annotea object context without any extra processing. RDF for a sample bookmark and a topic is presented in **Figure 1**.

Bookmarks and topics have many common properties with other Annotea objects. For instance, the objects can have a title (dc:title), a longer description (dc:description), information about their author (dc:creator and foaf:maker), the creation date (a:created), the modification date (dc:date), and links to other related information (RDFS:seeAlso)
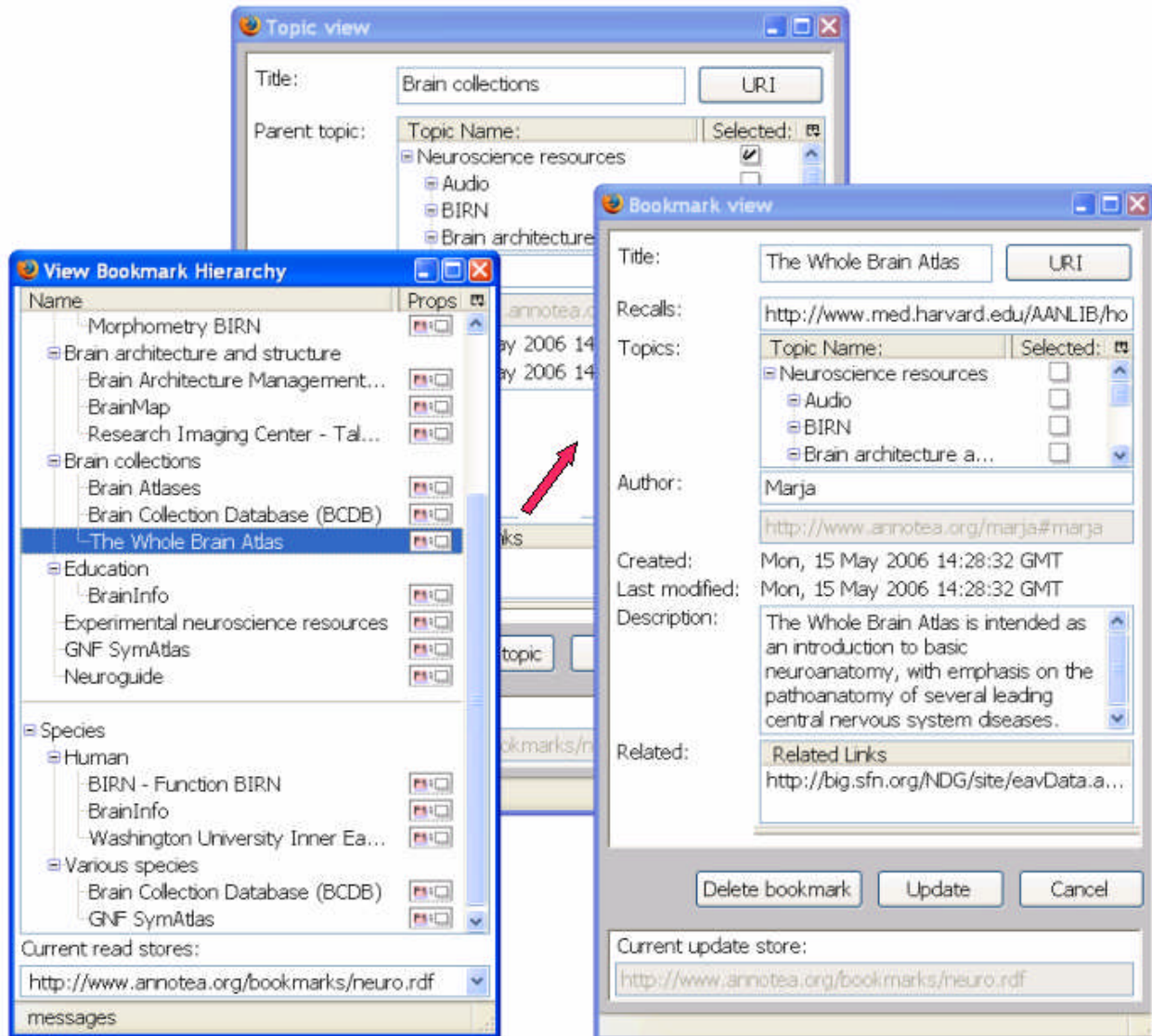
Figure 2. Clicking properties icon to open a bookmark view with common bookmark properties in Annotea Ubimarks. A topic view is presented in the background.

In addition, a bookmark has information about the page it recalls (`bm:recalls`), and zero or more topics (`bm:hasTopic`). A topic has information about its parent topic (`bm:subTopicOf`). In addition, new properties can be easily added.

4. VIEWING ANNOTEA BOOKMARKS AND TOPICS

Annotea objects can be presented to users without a need to know anything about the Semantic Web. The bookmark metaphor nicely supports semantic authoring and presentation of Annotea bookmark and topic objects, their collections, and context sensitive views to these objects.

The Annotea topics can be anything the user wants them to be. They can start with vague meaning and gradually become more robust and better understood. When user discovers other related concepts in a better established ontologies, her own concepts can be linked to those.

4.1 Bookmark and Topic Object View

Users can explore the bookmark or the topic information in simple forms. For instance, in **Figure 2** the "Bookmark view" presents the bookmark's RDF properties and values.

Properties can be added to Annotea objects and different presentations may show different sets of properties. With guidance information about the properties, such as the presentation order and the preferred widgets, the presentations could be created automatically. This is the next step for the new Annotea presentation objects.
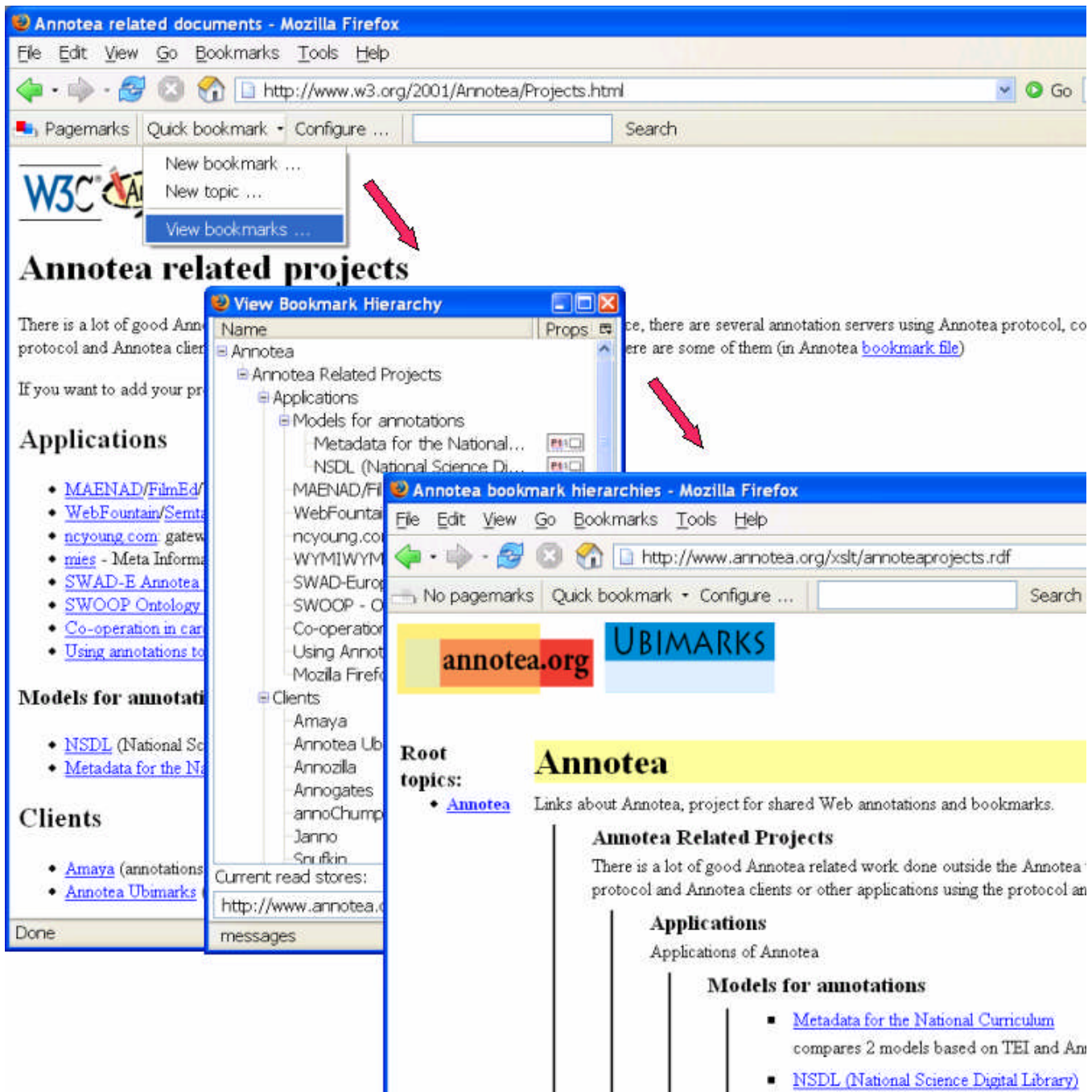
Figure 3. Annotea related projects page started as a manually edited Web page (bottom window). Next it evolved into a Semantic Web based Annotea bookmark hierarchy maintained with Annotea Ubimarks (middle window). Finally, the bookmark hierarchy semantics was presented as a Web page by using XSLT transformation (top window) making the original page obsolete.

4.2 Viewing Collections of Bookmarks and Topics

Annotea topics allow authors create and maintain shared classifications or topics [15]. A bookmark can be cataloged under one or more topics and presented to the user in a hierarchy view resembling a folder hierarchy. In **Figure 3** the "View Bookmark Hierarchy" window shows a hierarchy of

"Annotea related projects". Clicking the bookmark title in the hierarchy view shows the recalled page while clicking the properties icon presents the bookmark view (see **Figure 2**). Clicking a topic always shows its properties in a topic view.

The Semantic information in **Figure 3** bookmark hierarchy was originally copied from the Web page shown in the bottom browser window. The bookmark document was so easy to update with Annotea Ubimarks that in many cases the updates to

the original Web page lacked behind. Fortunately, it was quite easy to create a corresponding XHTML page directly from the bookmark document semantics with an XSLT transformation. The top browser window in **Figure 3** shows one example of such a transformed page. The style of the generated page can be easily changed to bear even closer resemblance to the original page. This is semantic authoring at its best as the user gets benefits from semantics that was created without any extra work or learning.

4.3 Viewing Bookmarks and Topics in Context

Context dependent views of Annotea objects are characteristic to Annotea applications. For instance, Annotea Ubimarks user interface provides information about related Web documents in the current document context. When a user browses through the Web pages and finds something interesting she can look for the "Pagemarks" icon. This icon means that one or more users has

bookmarked the current page. The "Pagemarks" icon can be clicked to open a view listing all the bookmarks on the page. These bookmarks and their topics can be easily followed to find related documents in the currently subscribed bookmark stores, some of which may provide information from Web crawlers.

Users collaborating in related research areas can benefit from the bookmark and topic information. They can find related topics and bookmarks. For instance, in **Figure 4** the user has bookmarked a gene page in Amigo. When she browses the page later and clicks the "Pagemarks" and the topic icons, she finds more information about the genes involved in CML leukemia made by other researchers in other departments of the institution or in other institutions. In the future, the user generated information can be combined with automatic categorizations, such as created in MagPie [12] or in Trailblazer [10] to help the user to find also non-bookmarked relevant documents and make new innovative associations.

Figure 4. Following the chain from a pagemark and its topic to related bookmarks.

```
<pres:Presentation RDF:about = "http://.../exPresentation">
  <pres:guidesResource RDF:resource = "http://.../exTopic" />
  <pres:ordersBookmarks>
    <RDF:Seq>
      <RDF:li RDF:resource = "http://.../exBookmark1" />
      <RDF:li RDF:resource = "http://.../exBookmark2" />
    </RDF:Seq>
  </pres:ordersBookmarks>
</pres:Presentation>
```

Figure 5. A simple Annotea presentation object.

4.4 Defining Custom Bookmark Orders

The order of the bookmarks can be defined by using Annotea presentation objects [5]. They can be attached to another resource, such as a topic, to guide its presentation. For instance, it can define the order of the selected bookmarks under a topic. The rest of the topics are presented alphabetically. The presentation objects let users define customized orders for a shared hierarchy. A sample presentation object is presented in **Figure 5**.
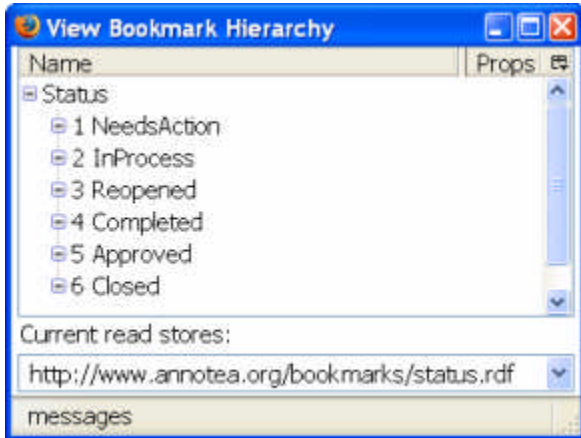

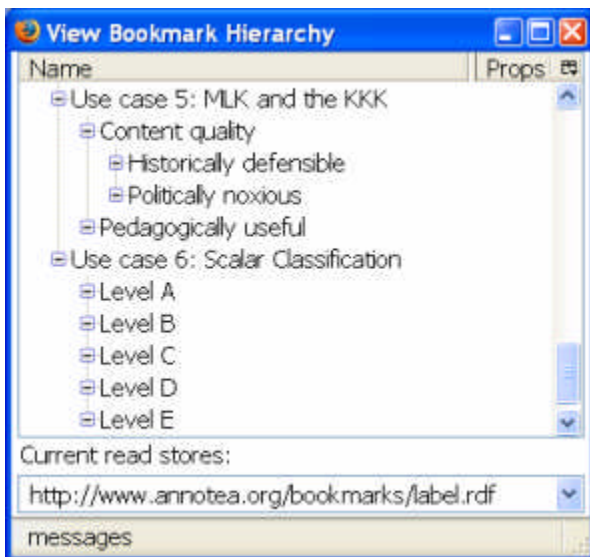**Figure 6. Using topics for attaching status values for bookmarked problem descriptions.**

Figure 7. Using topics for content labels.

4.5 Defining Multiple Root Topics

The topic hierarchies do not have to have a common root topic. It is possible to have several roots. This lets the user also view hierarchies even when some topics are temporarily unavailable. Hierarchies containing several root topics can be also viewed in multifaceted displays, such as the ones used in mSpace [23].

Because of multiple roots it is easy to define concept hierarchies that are intended to be used in parallel with other concepts.

Workflow status (see **Figure 6**) and content label hierarchies (**Figure 7**) are such examples.

5. CONCLUSIONS

Annotea has been pioneering with social bookmarking, tagging, and folksonomy applications on the Web. Annotea social bookmarks and topics provide ordinary users familiar metaphors for authoring and sharing reusable and easily extendible Semantic Web metadata without a need to learn or understand the Semantic Web.

Many Web pages containing links and categories can be chosen to be semantically authored as Annotea bookmark files and transformed with XSLT into XHTML pages. This makes maintaining the links easy.

Annotea users are not limited to use one service for their bookmarks and topics. They can subscribe several local or global data stores containing the various Annotea objects they are interested in at each moment. Web documents offer users an easy starting point. Semantic Web gives supports for easy merging and querying of Annotea metadata from several different sources.

Annotea object metadata is uniquely identifiable and includes information of its creator and creation dates. It can be easily extended, customized, combined, and reused in many applications and services. It can provide additional organization for data mining and search engine applications, chatrooms, and blogs.

Different views to the data, including document context dependent views, can be easily created. Overall, Annotea bookmarks and topics can provide huge benefits and flexibility for Semantic authoring without adding complexity for the user interface.

6. ACKNOWLEDGMENTS

7. REFERENCES

[1] Amaya homepage, http://www.w3.org/Amaya

[2] Annotea annotation schema, http://www.w3.org/2000/10/annotation-ns

[3] Annotea bookmark schema, http://www.w3.org/2002/01/bookmark#

[4] Annotea homepage at W3C, http://www.w3.org/2001/Annotea

[5] Annotea presentation schema (draft), http://www.annotea.org/schemas/2006/04/presentation

[6] Annotea thread schema, http://www.w3.org/2001/03/thread

[7] Annotea Ubimarks homepage, http://www.annotea.org/mozilla/ubi.html

[8] Annozilla home page, http://annozilla.mozdev.org/

[9] Brickley, D., and Guha R.V. (eds.). RDF Vocabulary Description Language 1.0: RDF Schema, W3C Recommendation 10 February 2004. http://www.w3.org/TR/2004/REC-rdfschema-20040210/

[10] Croke, P., Johnston, A., Tighe, K., Using Named Entities as a basis to share associative trails between Semantic Desktops, http://www.semanticdesktop.org/xwiki/bin/download/Wiki/UsingNamedEntitiesAsABasisToShareAssociativeTrailsBetweenSemanticDesktops/19_croke_jack_final.pdf

[11] Del.icio.us homepage, http://del.icio.us/

[12] Domingue, J.B., Dzbor, M., Motta, E. (2004) Collaborative Semantic Web Browsing with Magpie, In Proc. of the 1st European Semantic Web Symposium (ESWS), May 2004, Greece. http://kmi.open.ac.uk/people/dzbor/public/2004/iswc2004-dzbor-etal.pdf

[13] Flickr homepage, http://www.flickr.com

[14] Kahan, J., Koivunen, M., Prud'Hommeaux, E., and Swick, R. 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.

[15] Koivunen, M. Scenario: Organizing CML cancer research knowledge by using Annotea shared bookmarks, 2003, http://www.w3.org/2003/12/cmlcase/cml.html

[16] Koivunen, M., Annotea shared bookmarks: Semantic Web at your fingertips, In Proc. Of the ISWC 2004 Conference Demonstrations Session. http://www.annotea.org/ISWC2004/annoteademo.html

[17] Koivunen, M. Annotea and Semantic Web Supported Collaboration, in Proc. Of the ESWC2005 Conference, Crete, May 2005. http://kmi.open.ac.uk/events/usersweb/papers/01_koivunen_final.pdf

[18] Koivunen, M. and Swick, R. Metadata Based Annotation Infrastructure offers Flexibility and Extensibility for Collaborative Applications and Beyond, In Proc. of the KCAP 2001 Conference, http://www.w3.org/2001/Annotea/Papers/KCAP01/annotea.html

[19] Koivunen, M., Swick, R., Kahan, J., Prud'Hommeaux, E., Topics in Annotea: Extending the Basic Architecture,.WWW2002 Developers Day, Honolulu, May 2002.

[20] Koivunen, M., Swick, R., Kahan, J., Prud'hommeaux, E., An Annotea Bookmark Schema, 2003, http://www.w3.org/2003/07/Annotea/BookmarkSchema-20030707

[21] Koivunen, M., Swick, R., and Prud'hommeaux, E. Annotea Shared Bookmarks, 2003, In Proc. of KCAP 2003, http://www.w3.org/2001/Annotea/Papers/KCAP03/annoteabm.html

[22] Linkagogo homepage, http://www.linkagogo.com/

[23] Schraefel, M.C., Carr, L., De Roure, D., et al., You've Got Hypertext. Journal of Digital Information, 2004. 5(1): id 253.

[24] Shiraishi, N. (2004) The RDF Trust Model Using RDF Bookmark and it's Application. In Proc. of WWW2004 Workshop on Content Labeling -Technical and Socio-Cultural Challenges and Solutions. http://web.sfc.keio.ac.jp/~kaz/www2004/papers/ns.pdf

[25] Swick, R., Prud'Hommeaux, E., Koivunen, M., Kahan, J., Annotea protocols, http://www.w3.org/2001/Annotea/User/Protocol.html.

[26] Uren, V.S., Cimiano, P., Iria, J., et al., Semantic Annotation for Knowledge Management: Requirements and a Survey of the State of the Art. Journal of Web Semantics, 2006. 4(1): p.14-28. http://www.websemanticsjournal.org/ps/pub/showDoc.Fulltext/document.pdf?lang=en&doc=2005-34&format=pdf&compression=

[27] XML Bookmark Exchange Language, http://pyxml.sourceforge.net/topics/xbel/