

Annotea and Semantic Web Supported Collaboration

Marja-Riitta Koivunen, Ph.D.
Annotea project

Abstract

Like any other technology, the Semantic Web cannot succeed if the applications using it do not serve the needs of the users. Annotea is a Semantic Web based project for which the inspiration came from users' collaboration problems in the Web. It examined what users did naturally and selected familiar metaphors for supporting better collaboration.

The selected metaphors were a good match also for demonstrating the use of the Semantic Web technologies. Metadata was generated in the form of Annotea objects. It enhanced collaboration by adding flexibility to the applications and easy creation of different views. Furthermore, Annotea objects also let users to make the metadata available beyond its original purpose for many other Semantic Web applications.

The first phase of Annotea introduced Web annotations and replies, that formed reply threads, and the second phase, bookmarks and topics. All of these concepts are commonly used familiar metaphors that are general enough to suit for various purposes. As a result, normal users can easily create RDF metadata that can be merged, queried and mixed with other metadata.

1 Introduction

The original Web supports information sharing and collaboration between wide varieties of users, without requiring them to be computer scientists. The Semantic Web (SW) [5] focuses in providing more semantically exact data for machines and agents so that, as a consequence, the agents can better support users in finding the right information. But first the metadata containing the semantics needs to be generated.

Often the metadata is generated by the users of the SW. While it can be done in a decentralized manner together with other users it can still be tedious. Humans are not at their best in providing or understanding complex, machine readable information. Furthermore, they are seldom motivated to provide the information just to help the Semantic Web. For the Semantic Web applications to succeed they need to bridge the gap between the needs of the human users and the requirements of the SW machines and agents.

With Annotea [12] we wanted to experiment how we could enhance the collaboration in the Web with the help of the Semantic Web technologies [20, 6] that offer flexible tools for sharing the user data and semantics, easy extensibility, and

effortless merging and querying of the data. The idea of this Semantic Web Supported Collaboration (SWSC) was to support and enhance users' natural collaboration tasks and habits while examining and demonstrating the possibilities of the Semantic Web [13]. In that way, the metadata generation would not feel like an extra effort.

As Annotea tools were targeted for normal users we wanted to use familiar metaphors to support the collaboration. We also wanted users to be able to create metadata as an integral part of the tasks that they were already motivated in performing instead of explicitly creating metadata for the Web. The created metadata is gathered into Annotea objects: annotations, replies, bookmarks, and topics. These concepts were created in two phases.

During the first development phase we focused on examining how to help users, especially users in W3C working groups, to review and discuss the Web documents in the document context in addition to discussion lists. We developed Web annotations and replies that could be used for sharing comments, questions, discussion threads etc. on the context of the Web documents or other Web resources including the annotations and the replies themselves. The basic Annotea architecture was developed during this phase [12]. An important part of the architecture is the ability to store and retrieve the metadata from several annotation servers.

During the second phase we concentrated in enhancing organizing and grouping. The Web users typically grouped resources by organizing links to them under HTML text headings. They could not easily share and reuse categories and other semantics attached to the resources. Annotea shared bookmarks and topics [15, 16] were developed to support the sharing of links to interesting Web documents or other resources and the sharing of link categories. Annotea design allows the users to use simple topics that they find natural even though the topics could be informal and very subjective, especially at the beginning of new work. When the users' understanding about the domain gradually evolves, the organization of the topics can evolve also. When the user learns about similar concepts in more standard ontologies and feels comfortable in using them she can create links that tie her own topics to these concepts.

The SW technologies help Annotea to fulfill the users' needs in many ways. The use of standard SW metadata makes it easy to share annotations, bookmarks and topics with other users, share bookmarks and topics between different browsers, and query and present the annotation, bookmark and topic data in various views. Furthermore, with the SW technologies the users get additional benefits: the metadata can be easily combined with others users' metadata and it is also ready to be used by many other applications.

In the following sections, we will describe in more detail the basic Annotea components, the Annotea metaphors and discuss how the users can benefit from the metaphors and the generated metadata.

2 Basic Annotea Components

Fig.1 presents the basic Annotea architecture. We have various RDF metadata stores storing Annotea objects, a user interface providing different views to the objects in

the context of the Web documents or other Web resources, and users collaborating via these objects.

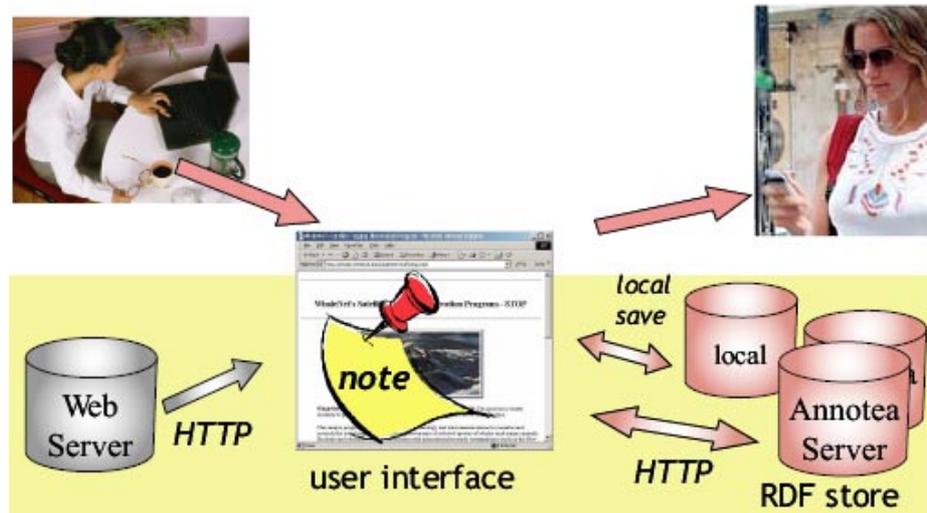


Fig. 1. The basic Annotea architecture.

2.1 Annotea objects

Annotea metaphors encourage users to create Annotea objects, such as annotations, replies, bookmarks, and topics. These are Web resources that have a URI, contain some RDF metadata, and normally include a property referring to some other Web resource. For instance, annotations have an “**annotates**” property for annotating a resource, such as a Web page or even another annotation. Bookmarks [14] have a “**bookmarks**” property for bookmarking Web pages or other resources such as annotations. In addition, the Annotea objects typically include a small set of core properties, such as a Dublin Core “**description**” for the description of the bookmark or a “**creator**” for the creator of the object. Other properties can be added if so wished.

2.2 Web browser user interface

The content of the Annotea objects can be presented in any Web browser user interface as XML text. However, to be usable for any user the normal Web browser needs to support Annotea metaphors. Currently, Annotations are supported in several browsers but they don't always have the same user interface or functionality. We developed the annotations and replies originally into Amaya [1] while Annozilla [3] provides a good implementation of annotations in Mozilla. We started also the bookmarks development with Amaya but currently the main Annotea shared bookmarks development is done in Firefox/Mozilla as Annotea Ubimarks [2]. In the

future we wish to collaborate more with Annotzilla development so that bookmarks and annotations offer a seamless user experience.

Some tools use Annotea objects but have extended them for their purposes. For instance, FilmEd [9] added time codes to be able to annotate films. Other browsers or tools that are not knowledgeable of the extensions may not be able to present the objects or parts of the objects. While the SW technologies help to make Annotea easily extendible the Web standards and browsers need to catch up in providing easier means for presenting these extensions to users.

2.3 Annotea metadata stores

Annotea objects metadata can be stored either locally, in Annotea servers or as published collections of Annotea objects in Web documents. The user can select from which of these metadata stores she wants to retrieve the Annotea objects. Similarly she can select a store for writing the created Annotea objects.

For historical reasons the current implementations use the Annotea servers for storing annotations and publish the bookmark or topic collections as Web documents. We started by providing servers for annotations but noticed that it is better to allow users to start without first figuring out how to install a server. In addition, a Web document containing annotations can be useful for archiving purposes because a version of the document can be saved with the related annotations for that document in the current review cycle.

Our future goal is to make this difference disappear and use Annotea servers also for storing bookmarks and similarly use the Web documents for storing collections of annotations.

3 Web Browser User Interface and Annotea Metaphors

Annotea uses several familiar metaphors to help users to attach the Semantic Web information to the Web resources as Annotea objects. All the metaphors were developed primarily to help solve user problems and secondarily to create metadata for the Web. We believe that this approach lead us to use more simple objects that can still be extremely beneficial, especially if the simplicity helps larger groups of users to provide the metadata.

This chapter first explains annotations and replies, then bookmarks and topics and finally discusses ways to mix and extend the metaphors.

3.1 Annotations and replies

After looking the W3C standard review process for a couple of months at the beginning of 1998 the author generated a couple of user scenarios where working group members and editors could see review comments as annotations in the context of the reviewed Web documents. The development of Annotea annotations started from those scenarios and targeted specifically to help collaboration between groups of

reviewers or other similar users while allowing a user to belong to several of these groups.

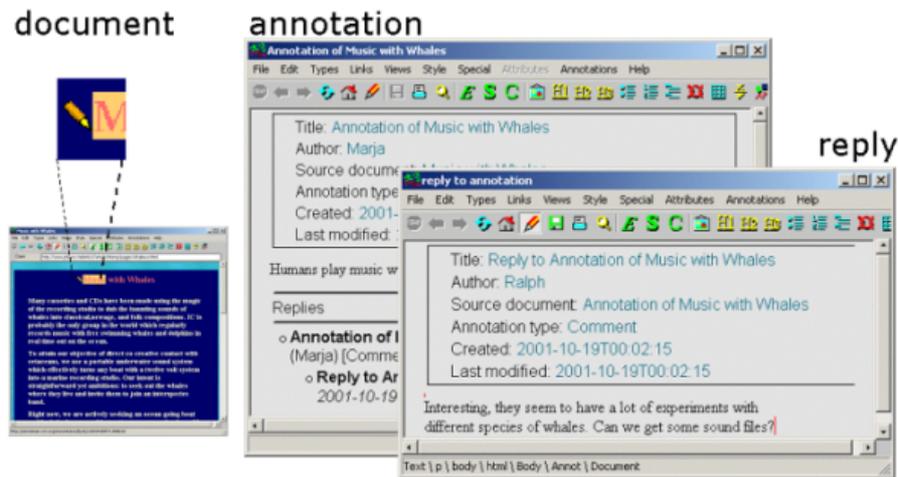


Fig. 2. Amaya interface to annotations and replies.

Unlike centralized annotation services, such as ThirdVoice [25], or even early versions of Mosaic [21] which by default offered one annotation server, Annotea was focusing in providing a mechanism where every Web user could customize the view to the annotations according to the collaborative groups they belonged to. The Annotea users could control whether they wanted to see annotations and select which servers they wanted to subscribe to retrieve the annotation metadata from. In addition, Annotea provided means to do more detailed filtering of the annotations, for instance, to selectively show the annotations created by a certain user.

Semantic Web technologies were used for implementing Annotea annotations for two reasons. They suited well for attaching information to resources and they helped develop SW technologies as part of W3C Live And Early Adoption (LEAD) philosophy [4].

Semantic Web made it easy to offer some control and flexibility for the users. For instance, we provided a default set of annotation types but also demonstrated how the groups themselves could define their own types if they so wished [17]. When users wanted to reply to annotations we easily extended Annotea to add reply objects with thread views as well.

The hardest part in this approach was to get the user interfaces implemented. Annotation content itself was relatively easy to present by using Web technologies, but making it as an integrated part of a browser and user experience was much harder. At the first phase we concentrated in co-operating with Amaya [1] browser developers to show a sample browser implementation. Several implementations were done to other browsers by other developers based on Amaya but with somewhat different functionalities and user interfaces.

Annotea annotations are implemented in browsers but they can also be implemented as part of other tools. For instance, the developers of the SWOOP

ontology editor [24] found that the users were often confused of how to properly use the concepts provided by ontology. They added Annotea annotations as an integral part of the tool to make it easy to annotate the ontologies with clarifying explanations and to point out potential problems.

3.2 Bookmarks and topics

During the first phase many informal discussions were performed with users and additional user scenarios were developed. The need for categories was high on the wish list, especially status categories were needed to mark the processed annotations. During the second phase, we selected to focus our scarce resources to broaden the scope and add new Annotea objects, bookmarks and topics. This would help us to make sure that our approach was extensible and the metaphors could work together. In addition, this approach gave us a chance to experiment with some other ideas before going back and improving the annotation implementations.

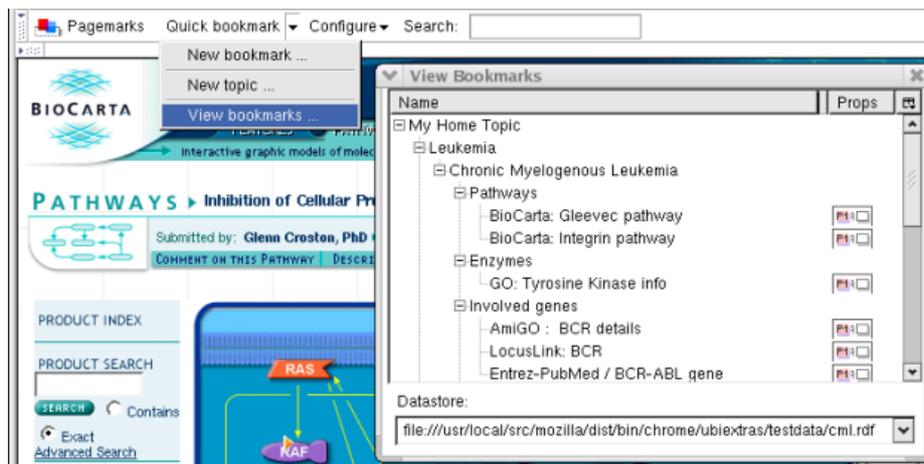


Fig. 3. Bookmark and topic hierarchy.

The shared bookmark metaphor with topics was perfect for Annotea as shared bookmarks could easily be seen as a variation of annotations. Furthermore, most users were not only familiar with bookmarks but had actually used traditional bookmark implementations. In addition, traditional browser bookmark user interfaces have a lot of enhancement possibilities that can benefit from the SW approach. For instance, the user interface can utilize the document context to remind the user that she has previously bookmarked the page or let the user define, share and link to other users' topics or categories. Furthermore, many other applications can utilize the bookmark metadata if it becomes widely available.

Annotea topics allow users to create and maintain shared classifications or informal categories [18]. A bookmark can be cataloged under one or more topics and presented to the user in a topic hierarchy (see Fig. 3.). When a user browses pages she

can see that someone has bookmarked a page from the pagemark icon that opens up to show a list of bookmarks.

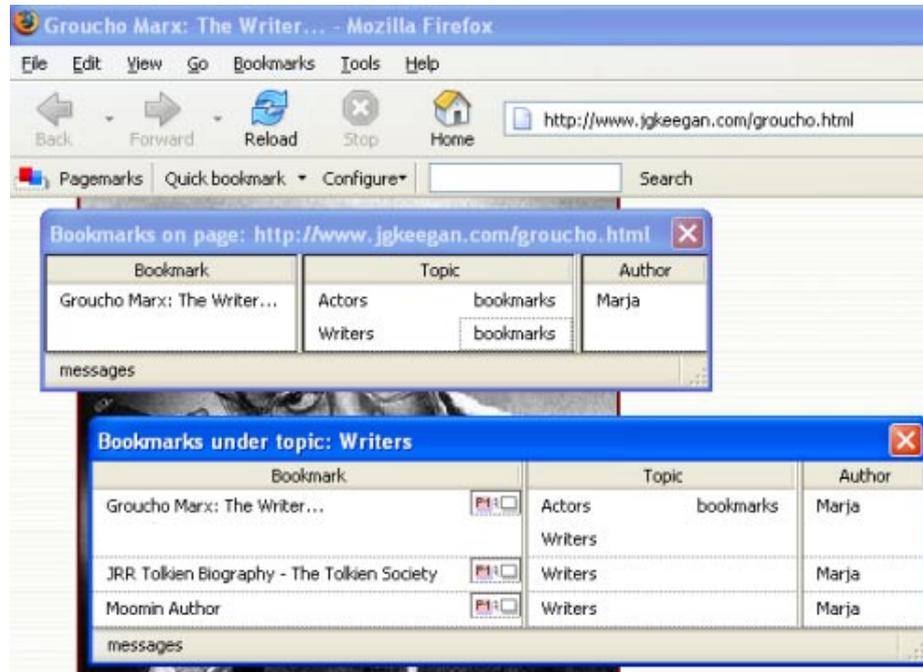


Fig. 4. Following the link chain from bookmarks on this page to related topics and bookmarks.

The Annotea topics can support early phases of innovations and research by letting the topics to be as vague as needed at the beginning and let the user to refine them as more learning happens or link them to concepts in well established ontologies when those are discovered and understood. Users collaborating in similar or related research areas can benefit from this information. They can see bookmarks on a current page and find related topics and other bookmarks to possibly interesting documents under these topics. For instance, in Fig. 4. the user sees that the page about Groucho Marx has been bookmarked by looking the “Pagemarks” icon on the left side of the toolbar, opens the Bookmarks on page window sees two topics “Actors” and “Writers”, and the bookmarks related to “Writers” to find other “Writers”.

The Annotea topics can easily define concepts outside the conventional categories, such as status (see Fig. 5.). If the topics are separated from the bookmark stores the user can define which aspects she is interested in at each moment by subscribing those topic hierarchies. Similarly, the presented bookmark stores can be selected in a certain domain area or by the research group depending on how they are organized. There are many possibilities for enhancing this user interface.

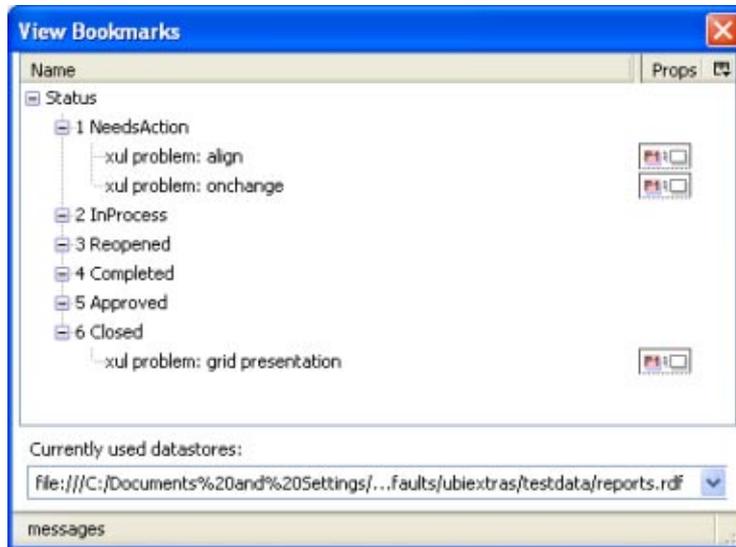


Fig. 5. Using topics for attaching status values for bookmarked XUL problems.

3.3 Mixing metaphors

The Annotea objects, especially annotations and bookmarks, do not differ very much from each other and can be easily thought as variations of the same class. However, we have also had long and intense discussions with users seeing annotations and bookmarks as a totally different concept and explaining that it would be confusing if bookmarks were annotations. On the other hand we have users who want to immediately extend the bookmarks so that they can refer not only to a Web resource with a URI but also to a part of a document in a similar way as annotations.

The SW metadata of the objects is easy to extend in different ways but designing the user interface to be both simple and expressive enough is critical. Our current view is to keep things simple and see what happens when users start using these interfaces more. Our hope is to be able to experiment with Annozilla and Ubimarks. Maybe the users can create annotations to point to parts of the document and then bookmark that annotation to give it a category and make it easy to find. This would match nicely the way the annotations and bookmarks are combined in Fig. 6 examining the real world usage of annotations and bookmarks.

3.4. Extending Annotea objects

New properties can be easily added to Annotea objects by using the SW technologies, however, SW does not yet offer good standard presentation and interaction definitions. Furthermore, if we want to support normal users to be able to do the extensions much more research is needed. Even if the users would select from a list of

predefined properties, or define a similar property to an already existing one, they would still need to do some learning of the properties.

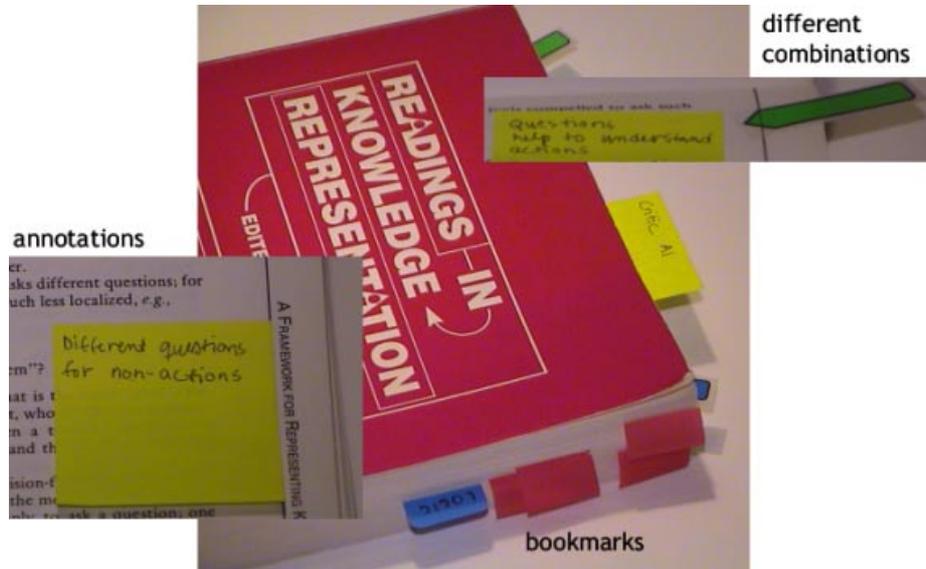


Fig. 6. Bookmarks and annotation concepts mix in the real world.

Ontology editors and browsers, such as Protege [22], make the definition of new classes, properties, instances and presentations for them simple. However, the users of these kinds of editors need to have an understanding of classes and properties. While new Annotea objects or subclasses of them could be defined with such an editor, a person with some familiarity with the basic ontology concepts and as well as requirements for the Annotea objects is needed. With ontologies, it is often easy to define and standardize the user interface appearance as the set of possible properties are predefined.

A definition language is needed for presenting the added properties in different views. Currently, the properties not known to the developer beforehand are often presented in an unordered list with property name, string value or a link. Another approach is to show only the known properties and there are use cases for both.

Another aspect we want to be able to extend is the addressing mechanisms and some of the projects using Annotea have already implemented extensions. For instance, we want to be able to use SVG for defining piece of an image that is annotated or time codes to define part of a video. We also want to experiment with different context information. While the metadata extensions need some design, they are easy to accomplish with the underlying SW metadata. However, the user interface is problematic.

We need an extension mechanism for adding user interface definitions related to the metadata extensions for already existing Annotea implementations. So far, we have experimented a little bit with IsaViz [11] by adding simple presentation rules for presenting the properties and their order. In the future, we hope to be able to add piece of presentation code as part of Annotea object extension to make the user interface

extensions simpler. The user interface extension could be a combination of presentation rules and scripting, for instance using definitions similar to XUL [26] or XForms [8].

In most cases, we don't expect normal users to write the Annotea object extensions but when new extensions are provided by experts they should be able to easily use them and understand them. Here, the direct benefit from SW technologies is for the developers. When it is easy to add new functionality, test them and change them according to user feedback, it is more probable that users will get what they need.

4 Sample scenarios benefiting from Annotea metadata

Annotation and bookmark metaphors make it easy for users to do what they are already familiar with. Annotations and bookmarks also solve user problems related to collaboration which motivates their creation. As a result from using Annotea objects for their own needs the users create SW metadata that can be easily reused by various other applications. Here are couple of examples of such applications.

Spam annotations support collaborative spam filtering:

Spam messages in discussion lists can be annotated by trusted users and the messages filtered away while not losing any information from the archives. This is used at W3C discussion lists [19].

Bookmark and topic collections can be used as user controlled profiles:

Users can use parts of their bookmark collections as user profiles when they visit services, for instance, a user visiting Amazon.com can ask similar books to the ones he has bookmarked on the Web in addition or in place of the information the service already has gathered of the user.

Using bookmarks and topics for finding and categorizing related information:

Data mining techniques can be used to find related resources by using connection paths provided by Annotea topic objects. This works even when the user is not subscribing the data stores and following the links like in Fig. 4. The topic objects can also help in naming the automatically found clusters of information in user understandable ways.

Annotea bookmarks and topics objects could be used to enhance tools providing automatic collaborative browsing, such as Magpie [7]. For instance, the tool could provide links to related projects both automatically and by utilizing user generated bookmarks. With the help of topic object data these projects can be presented in category hierarchies initiated from users' own understanding.

Organizing search engine results:

Users can use bookmark or annotation collections with search engines to organize the search results. For instance, when using Google the bookmarks of an expert user group in a searched domain can be used to first show the resources in that domain

[23]. Alternatively given topics and their related topics found from the Web can be used to organize the results.

Easy feedback channel for normal users:

Annotations and bookmarks can be used as a feedback channel for many services and they can also be integrated with other SW applications. For instance users could bookmark the resources in applications, such as Museum of Finland [10], by using Annotea topic objects. If the application provides a bookmark server, the data from that can be used to analyze and further develop the used ontologies.

5 Conclusions

SW technologies can support users directly by helping them to generate reusable standard metadata or indirectly by helping the developers provide different views to the data. If we want a wide variety of users to contribute to the SW by providing metadata and benefit from the metadata they need both motivation in the form of helping them in their tasks and good metaphors that hide or make the technology understandable. With Annotea we used SWSC starting from analyzing users' collaboration needs for motivation and Annotea metaphors for making the necessary SW technologies understandable.

Annotea metaphors successfully hide the underlying SW technologies from the users so that they can use SW fluently without even knowing about it. Users do need to know how to subscribe the data stores containing the various Annotea objects. Stores can be local files, global servers and or Web documents containing the metadata. Web documents offer users an easy alternative to get started without investing in installation of a server. They can also be used to archive snapshots of the selected Annotea objects outside the server.

SW offers the developers an easy and flexible interface for merging metadata from several different sources and doing queries against it. Different views to the data can be created easily, and it is easy to let the users follow tracks of data from the information on the current page to possibly related information. Extending the SW data in Annotea objects is easy as well. Defining a simple user interface for extensions is relatively easy, but adding more complex user interface definitions to the extensions needs more research.

The biggest direct benefit from the use of SW technologies and metadata is that the user generated metadata can be easily combined and reused in many other applications, such as user profiles for services, data mining and search engine applications.

Acknowledgements

Many people have contributed to Annotea in one way or other during its development and helped to make it what it is today. I want to give special thanks to Jose Kahan, Eric Prud'hommeaux and Ralph Swick.

References

- [1] Amaya home page, <http://www.w3.org/Amaya>
- [2] Annotea Ubimarks homepage, <http://www.annotea.org/mozilla/ubi.html>
- [3] Annozilla home page, <http://annozilla.mozdev.org/>
- [4] Berners-Lee, T. Web Architecture from 50,000 feet, <http://www.w3.org/DesignIssues/Architecture.html#Collaboration>
- [5] Berners-Lee, T., Hendler, J. and Lassila, O. The Semantic Web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities, *Scientific American*, May 2001.
- [6] 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-rdf-schema-20040210/>
- [7] Domingue, J., Dzbor, M., and Motta, E. 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/ESWS-domingue-dzbor-motta-final.pdf>
- [8] Dubinko, M., et. al. XForms 1.0, W3C Recommendation 14 October 2003, <http://www.w3.org/TR/xforms/>
- [9] FilmEd homepage, <http://metadata.net/filmed/>
- [10] Hyvönen, E. et. al , Finnish Museums on the Semantic Web: The user's Perspective on MuseumFinland, In Proc. of the Museums and the Web 2004 Conference, <http://www.archimuse.com/mw2004/papers/hyvonen/hyvonen.html>
- [11] IsaViz homepage <http://www.w3.org/2001/11/IsaViz/>
- [12] 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>
- [13] 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>
- [14] Koivunen, M., Swick, R., Kahan, J., Prud'hommeaux, E., An Annotea Bookmark Schema, 2003, <http://www.w3.org/2003/07/Annotea/BookmarkSchema-20030707>
- [15] 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>
- [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. Defining New Annotation Types in Amaya, January 2004, <http://www.w3.org/2001/Annotea/User/Types.html>
- [18] Koivunen, M. Scenario: Organizing CML cancer research knowledge by using Annotea shared bookmarks, 2003, <http://www.w3.org/2003/12/cmlcase/cml.html>
- [19] Koivunen, M., AnnoSpam: Filtering Spam According to Annotations, demo slides, September 2003

- [20] Lassila, O., and Swick, R. (eds.), Resource Description Framework (RDF) Model and Syntax Specification, W3C Recommendation, 22 February 1999
<http://www.w3.org/TR/1999/REC-rdf-syntax-19990222>.
- [21] NCSA Mosaic: Annotations Overview, 1997,
<http://archive.ncsa.uiuc.edu/SDG/Software/XMosaic/Annotations/overview.html>
- [22] Protege home page, <http://protege.stanford.edu/>
- [23] 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>
- [24] SWOOP Ontology Editor home page, <http://www.mindswap.org/2004/SWOOP/>
- [25] ThirdVoice, 1998, <http://c2.com/cgi/wiki?ThirdVoice>
- [26] XML User Interface Language (XUL) home page, <http://www.mozilla.org/projects/xul/>