Annotations in Semantic Web
- A position Paper

Authors:
Venkatasubramani S
Raman RKVS

***Email Addresses*:**
venkat@ncb.ernet.in
raman@ncb.ernet.in

Mailing Address:
National Centre For Software Technology,
68, Electronics city,
Bangalore, 561229,
India.
Ph: +91-80-852 33 00

Table of Contents

Topics

List of Figures

1. Abstract

After the recent revolutions of HTML and XML, the Internet community is preparing for another major break through. The new development is termed as 'Semantic Web'. In this position paper, we shall see the trends of semantic web, the building blocks and a feature review of one such building block. We shall begin with a general introduction about metadata, its role in semantic web, and metadata management with RDF. Then, we move on to see what are annotations (a form of metadata) in the context of semantic web and a practical implementation for annotation handling. The component is a toolbar plug in for MS Internet Explorer. This toolbar helps in metadata management.

2. Metadata – An introduction

It's a strange thing to note that despite the billions of web pages spread all over the Internet, we are devoid of 'information'. What we have is only data. There doesn't exist today, a tool that automatically reads all these data and presents us with what we exactly want. All those web search engines and search-bots roaming around the Internet rely mostly on keyword searching, while we want is context/semantic-based data. To understand what that means, let's consider a simple example. Assume that we are searching for ''. It won't be a surprise if our popular search engine threw up thousands of links pointing to , , , etc., - most of them irrelevant to the current context. The problem in this and many other cases is that the search engine or any other tool does not understand what exactly we want. All web pages are alike in the eyes of a robot or search engine. They do not have a way to find out what is relevant – neither what the user considers as relevant, nor the views of the author of the web page. It's here that metadata come in to play a key role in Internet. Metadata is 'data about data, usually understandable and process able by machines, automatically'. This data need not be only about web pages – it could be any data. For example, database management systems (DBMS) keep a metadata repository to find, query and manipulate the user data. The core idea is that: 'Metadata is any piece of data that helps in finding our original data efficiently and quickly – a sort of index'.

3. Metadata and the Semantic Web

Before we see how metadata can be used in the concept of Semantic Web, let us first briefly cover what is the 'Semantic Web'. Semantic Web, for some time around, is being called as the 'Web of Future'. There seems to be so many definitions around for 'Semantic Web'. To quote the more relevant ones:

Semantic Web is:

- an entity that includes documents, describing explicit relationships between things and containing semantic information intended for automated processing.
- an extension of the current web in which information is given well defined meaning, enhancing the capability of computers and people to work in co-operation.

- an idea of having data on the web defined and linked in a way, that it can be processed by machines - not just for display purposes, but for using it in various knowledge based applications.
- a query-able web-model that enables machines to make more sense of the web, with the result of making the web more useful for humans.

The key points to note in the above definitions are 'machine readability' and 'automated processing'. This directly leads us to the fact that some sort of additional data, apart from the actual contents of the web pages, is required so that any tool can access this data, process it according to its requirements and present the results to the user with more relevance and 'intelligence' rather than a simple text based keyword search. This additional data is the metadata we are talking about. Metadata about one document can occur within the document, or within a separate document, or it may be transferred accompanying the document. Some primitive form of metadata that is already in use is the META tag embedded in the HEAD tag of HTML documents. However, the usage is very limited and is highly unstructured. No standard exists today that defines the content of the META tags and how they should inferred. For example, an attribute of 'location' of a META tag is different from the attribute 'URL' even though they logically mean the same. To overcome all these drawbacks, W3C (World Wide Web Consortium) has come up with a specification for uniform resource description and metadata interchange. This specification is termed as Resource Description Framework (RDF). The annotation tool about which we shall be seeing soon has been developed to take advantage of RDF.

4. Role of RDF in metadata management

RDF, as we have seen already, is a framework for metadata. It provides interoperability between applications that exchange machine-understandable information on the Web. RDF emphasizes facilities to enable automated processing of Web resources and as such provides the basic building blocks for supporting the Semantic Web. RDF metadata can be used in a variety of application areas. Some typical applications of RDF includes:

- Resource discovery to provide better search engine capabilities.
- Cataloging for describing the content and content relationships available at a particular Web site, page, or digital library by intelligent software agents to facilitate knowledge sharing and exchange.
- Content rating.
- Describing collections of pages that represent a single logical "document".
- Describing intellectual property rights of Web pages, and many more.

5. Annotations – An overview

Annotation is one of the most common forms of metadata in the context of Semantic Web. Annotations can be comments, notes, explanations, questions, references, examples, advice, correction or any other type of external remark that can be attached to a Web document or a selected part of the document. As they are external, it is possible to annotate any Web document independently, without needing to edit that document. From the technical point of view, annotations are usually seen as metadata, as they give

additional information about an existing piece of data. They can be stored either at the user workstation for his private use or at one/more common servers. Annotations stored in public servers can be made available to anybody with proper authentication. By this way everybody can share their views/comments about a particular web site regarding its content, usefulness, etc., Since anybody can make annotations in any page, the visitors of the site can get a wide range of views – with out being restricted to the views of the author of the page alone, as in the case of META tags. Also the META tags applies, in general to the whole page – whereas annotations can be made for a particular section/text in the page.

6. The Semantic Web Project at NCST, Bangalore

The project on Semantic Web was started at NCST Bangalore, India in October 2001. The aim of the project is to develop an integrated solution for the realization of Semantic Web, individual components being:

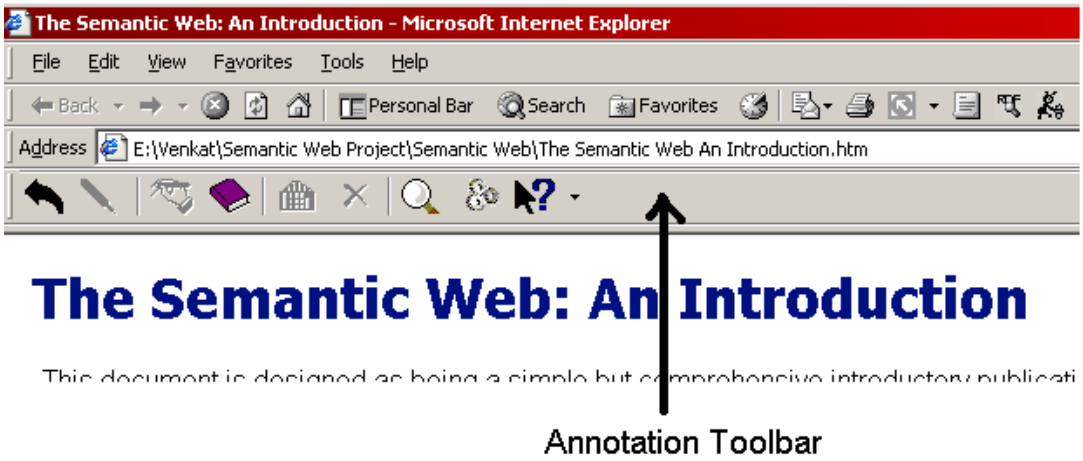
- Developing a web page annotation tool for making annotations
- Development of Ontologies – 'Ontology creation tool'
- Development of Inference engines to query metadata repository
- Designing software agents that communicates with all the above

This paper briefly explains the annotation tool.

7. The Annotation toolbar component

The annotation toolbar component is a plug in for MS Internet Explorer developed in Visual C++. Figure 7.1 shows a screen shot of the annotation toolbar in the Internet Explorer. The various functionalities of the toolbar are explained in the following sections.

Fig: 7.1 Annotation Toolbar Plug in for Internet Explorer



8. Annotation toolbar in action

The following functionalities are provided by the toolbar component.

8.1 Making Annotations

The user can select a particular text in the web page currently displayed by the browser and make an annotation for it. Alternatively, annotations can be made for the whole document also. When the user is making the annotation, he is asked to give details about the annotation, which constitutes the attributes of the annotation. They are explained below:

Table 8.1: Annotation Attributes

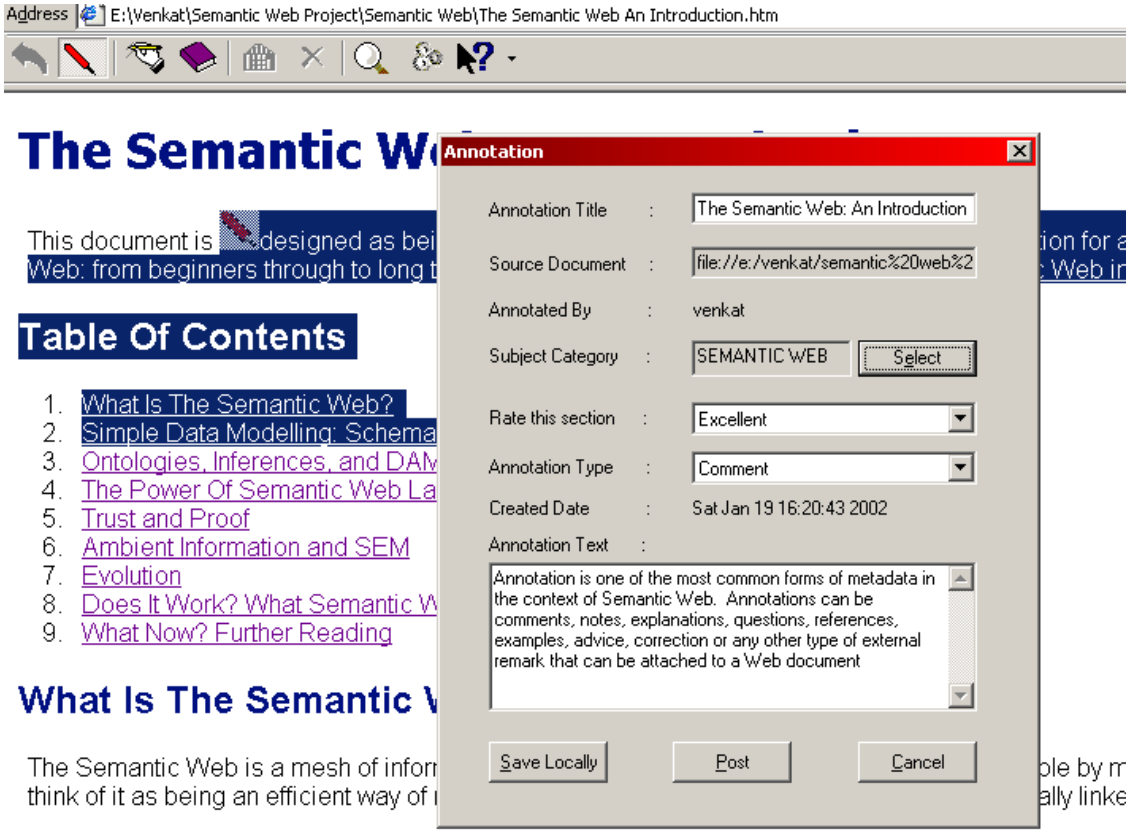
Attribute	Explanation
1) Annotation title	Title for the current annotation
2) Source Document*	Web page in which the annotation is being made
3) Annotation author	User who is making this annotation
4) Subject Category	Please see section 8.1.1
5) Annotation type	Can be any one of the following: *Advice, Change, Comment, Correction, Example, Explanation, Question, See Also*
6) Annotation Rating	User can rate the section in the web page under any one of: *Excellent, Very Good, Good, Average, Poor*
7) Created Date*	Current date and time
8) Annotation text	The actual views/comments of the author

* - Filled in automatically.

The user can select the subject category (from the pop-up tree view) and store the annotation under that category. This is an important feature, as this categorization will enable the annotation author or any other user to query for annotations in a particular category. After completing all the required details about the annotation, the user has a choice of either storing it in his/her local machine or posting the same annotation to a central server so that the annotation is available for all users to view. The annotations in the current page are indicated by the icon 

Fig 8.1 shows how the annotation is made.

Fig: 8.1 Making an Annotation



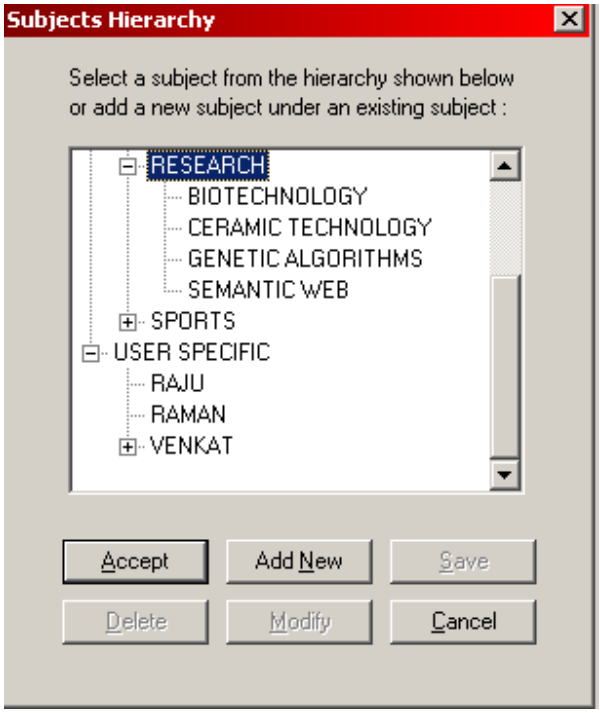
8.1.1 The Subjects Hierarchy view

The subject hierarchy is shown to the user in the form of a tree. The user can browse through the hierarchy and select the appropriate subject under which his annotation can then be categorized. If the user feels that the there is no subject in the given hierarchy that closely matches the one he has thought about for his annotation, a new subject can be added under any of the existing subjects. One additional feature is the provision for subject categorization according to users, apart from the 'General' category. This feature is useful especially in intranets (e.g. in the case of a Research Organization). Annotations can be stored in the corresponding subject categories under different users and others can easily lookup the annotations made for particular subjects, once they come to know that a person X is doing research on topic Y.

Any authenticated user can add a new subject category or make annotations under the 'General' category. Whereas in the 'User Specific' category, only the corresponding user can add new subject categories or annotations under existing categories, under any existing subject categories.

Fig 8.2 shows the subjects hierarchy view

Fig 8.2 Subjects Hierarchy View



8.1.3 Annotation Servers

When the user is making a new annotation, we saw that the new annotation can be saved locally on the machine or posted to a central server. This server is named as annotation server. This server receives requests from the clients (from the annotation toolbar component in the browser). Some typical client requests include annotation addition/modification/deletion, query for annotations in particular subjects, request all annotations in a particular page, etc. The server to which the user posts his/her annotations is named as 'Post Server' and the one from which the user retrieves annotations is called as 'Source Server'. Typically, there will be one Post Server and one or more Source Servers.

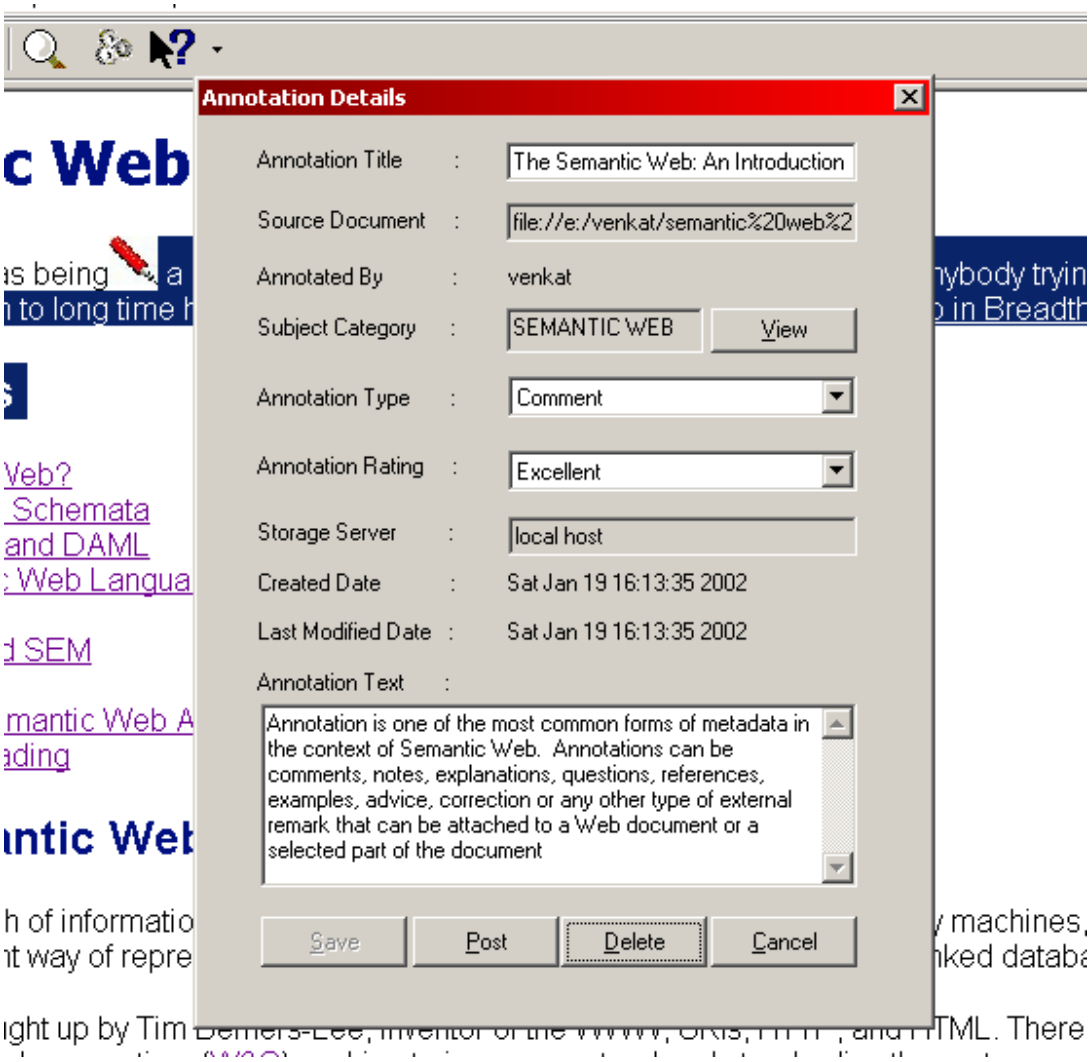
8.2 Viewing Annotations

The annotations stored already, either in the local machine or in the central server can be loaded in the current page. After loading the annotations, the 'Show/Hide Annotations' toggle button can be used to view/hide all the annotations in the current page. All the annotations in the page are indicated by the icon

at the beginning of each annotation (identified by the start of the selection text made by the annotation author). The user can then single click on the annotation icon to highlight the text selected by the annotation author. A double click on the annotation icon will pop

up a dialog box giving all the details about the annotation like the annotation title, annotation author, subject category under which the author has classified this annotation, created and last modified date and finally the actual comments/views made by the annotation author. Alternatively, a right-click on the annotation icon pops up a context menu from which commands like highlighting the text, viewing the annotation details etc., can be accessed.

Fig: 8.3 Viewing, Modifying and Deleting Annotations



8.3 Modifying existing Annotations

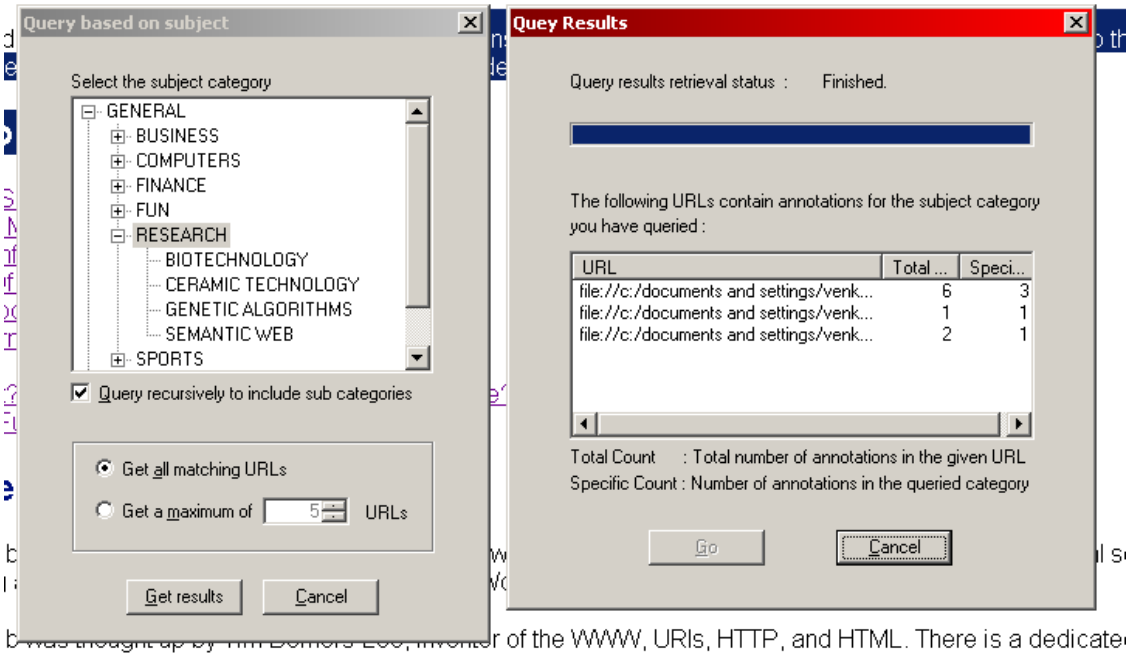
After the annotations are loaded in the current page, they can be modified individually if the annotation author happens to be the same as the current user. The

details that can be modified include: annotation title, subject category, annotation type, rating, and the annotation text. If the user feels that a particular annotation is not relevant anymore or if his views do not hold any longer, he can opt to even delete the annotation he has made.

8.4 Searching for Annotations

The user can also search for annotation s belonging to a particular subject or annotations made by a particular user. The search criteria can include recursive search – where the result includes annotations belonging to sub-categories also. The result of the search includes the list of web pages that has annotations for the queried category, total number of annotations in that page and the number of annotations belonging to the queried category. The user can then navigate to any of the pages displayed by double clicking on the URL in the list box.

Fig: 8.4 Searching for Annotations



of the WWW, URIs, HTTP, and HTML. There is a dedicate

8.5 Customizing the settings

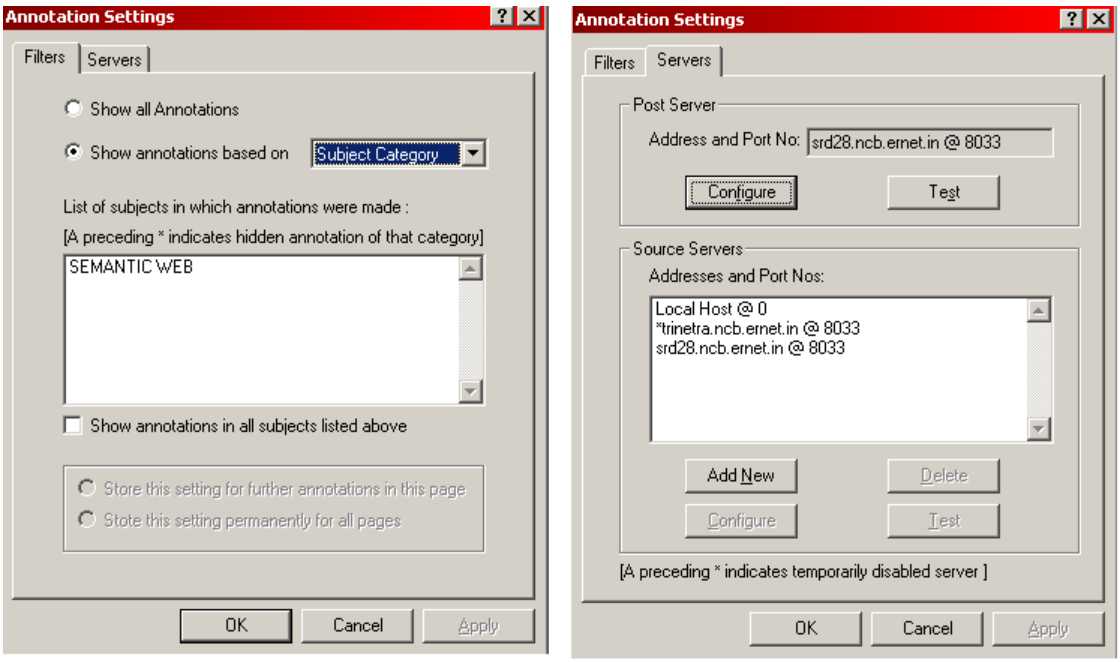
8.5.1 Annotation Filters

After loading the annotations for a page, the user can apply the filters to view/hide annotations according to the criteria chosen by him. The filter criteria includes the attributes of annotations namely, the subject category, rating, type of annotation, author and the server from which the annotation was loaded. The user can select one or more categories and opt for viewing/hiding annotations in those categories. For example, if there are annotations for 'Computers', 'Finance' & 'Business' subject categories in a particular page, the user can hide the annotations belonging to 'Business' and view only those belonging to 'Finance' & 'Computers'.

8.5.2 Annotation Servers

The user is provided with an option to configure any of the servers (Post and Source Servers). The server name, port number, user name and password are the configurable details for each server. Additionally, the user can opt to disable any server temporarily. This feature is useful either when the user is not interested in retrieving annotations from a particular server for some time or if the server is down.

Fig: 8.5 Annotation Configuration



9. Proposed further developments

Currently, the annotations are done manually. The user has to go through the contents of the web page and make the annotations. Instead, if this process could be automated, the whole web page can be annotated with the help of ontologies.

After making the annotations and collecting all the metadata, any user can then use some specialized agents to search 'semantically' for a particular context. Many such agents can interact among themselves sharing the details collected. The agents in turn can use inference engines to access the metadata collected in the repository.

10. Conclusion

The concept of semantic web is one of the hot topics among the researches. Some people predict that semantic web will be the 'Web of Future'. We can expect to hear about many interesting research topics and hot discussions in semantic web and its implementations in near future. This annotation toolbar component is a first step towards realization of semantic web.

11. References

1) RDF Specification Development, W3C, URL: http://www.w3.org/RDF/

2) Metadata Architecture, W3C, URL: http://www.w3.org/DesignIssues/Metadata.html

3) Semantic Web Activity Statement, W3C, http://www.w3.org/2001/sw/Activity

4) Semantic Web Development Proposal, W3C,
URL: http://www.w3.org/2000/01/sw/DevelopmentProposal

5) The Semantic Web: An Introduction, Sean B. Palmer,
URL: http://infomesh.net/2001/swintro/