Knowledge Sharing Platform Technology for Using Metadata on the Semantic Web

Hiroyuki Sato†*, Yutaka Abe, and Yukihisa Katayama*

Abstract

On the Semantic Web, the semantic attributes of information resources are expressed by using general-purpose metadata expressions. Here, we explain technology for a peer-to-peer platform for sharing not only existing metadata, but also metadata that has been extracted from each user's actions in the local environment of that user's personal computer. That metadata expresses the semantic relationships among resources and is shared according to the interests of individual users. Gathered metadata can also be employed by other users as knowledge.

1. Introduction

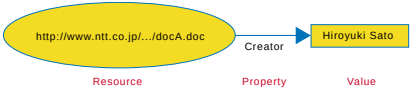
The Resource Description Framework (RDF) [1] specifies a grammar for describing the semantic attributes of information resources that can be identified by a uniform resource identifier (URI), such as documents on the Internet. RDF is in the process of being standardized by the World Wide Web Consortium (W3C) as a metadata description language that employs XML (Extensible Mark-up Language) as the representation format. XML can represent any data structure but does not specify its semantic significance. The RDF syntax, on the other hand, consists of a set of three elements—resource, property, and value—as shown in Fig. 1. Taking a document as an information resource, for example, if the property "creator" is assigned the value "Hiroyuki Sato," then the fact that the creator of the document is Hiroyuki Sato can be represented unambiguously. This characteristic makes it possible to use RDF as a general-purpose metadata description language that can be processed by machine.

In this article, we introduce knowledge sharing platform technology for implementing knowledge management and other such services by sharing metadata according to this general data model.

† NTT Information Sharing Platform Laboratories
Musashino-shi, 180-8585 Japan
E-mail: sato.hiroyuki@lab.ntt.co.jp

2. The Semantic Web

The Semantic Web [2] is an attempt to provide a




```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
  <xmlns:dc="http://purl.org/dc/elements/1.1/">
  <rdf:Description rdf:about="http://www.ntt.co.jp/…/docA.doc">
   <dc:creator>Hiroyuki Sato</dc:creator>
   </rdf:Description>
</rdf:RDF>
```

Fig. 1. Example of RDF syntax and representation by XML.

system for solving complex problems by treating Web resources from a semantic perspective based on RDF. It involves the use of RDF to create metadata for various information resources. Doing so makes it possible, to take a simple example, to perform a precise search for only those documents that have been created by "Hiroyuki Sato", a kind of search that is not possible with conventional keyword search methods. If an ontology concerning a vocabulary of the properties is provided together with the metadata, then even more advanced kinds of search may be expected. If it were possible to express in an ontology that "creator" and "author" have similar meanings, then a document that has "Hiroyuki Sato" as the value of the property "author" could also be returned by the same search.

In addition to specifying the RDF syntax, W3C is proceeding with standardization of the RDF Vocabulary Description Language (RDF Schema), which will allow the definition of property vocabularies, and Web Ontology Language (OWL), which is a specification for using XML to describe ontologies. W3C is thus advancing the construction of a foundation for the next-generation Web, which will make possible intelligent services implemented by means of automatic machine processing.

3. Knowledge sharing platform

Metadata can represent the relationships among resources as well as resource attributes such as "creator". With RDF, both can be represented within the same data model. Making use of that characteristic, we propose a knowledge sharing platform that makes it possible to automatically collect a large volume of RDF metadata and provide to the user various kinds of background information concerning the content items that are linked by the metadata. This platform does not simply handle the metadata associated with content on the Internet that has a URL (Uniform Resource Locator); the content documents that are on the user's PC are also given unique URIs and metadata about those documents is accumulated in the local environment. A mechanism for sharing that metadata among multiple users as necessary by using a peer-to-peer (P2P) protocol is also provided.

4. Aggregation of metadata

The question of who shall create the metadata, which has a cost associated with it, is one of the major problems for the Semantic Web. If services that provide advanced search functions that employ metadata become common, then each Web site might use RDF descriptions to make its pages available for searching. At present, however, there is no beneficial effect to offset the cost of producing the metadata, so most people can be expected to have a negative attitude towards the creation of metadata. Therefore, for the knowledge sharing platform, we took the approach of extracting metadata that represents the relationships among the content documents from each individual user's actions in handling the documents in the local environment of that user's personal computer. For example, when the user accesses a certain document on the network and uses an editing program to edit it in the local environment and then saves the document under a different name (using the editing program's "Save as ..." function), an RDF description that links the new and old documents with the property name "update" to describe the relationship between the two documents is created.

5. Making use of existing resources and technologies

We also believe that it is important to make effective use of existing metadata.

5.1 Utilizing file properties

The creator's name and other such properties can be extracted from files that have been saved from Microsoft PowerPoint or other such software. Files created by Adobe software have RDF metadata embedded in them. In the knowledge sharing platform, metadata can also be extracted from the information contained in HTML META tags and treated as an RDF metadata file that is separate from the original file.

5.2 Utilizing a directory-type search service

A directory-type search service is being provided by NTT-X's goo[*1]. We believe the hierarchical relationships among directory names (for example, the word "J1" is under "J-League", etc.) can be represented as an RDF schema. We provide a way to extract directory data as an RDF schema and allow its use as a simple ontology for information on the Internet.

5.3 Utilizing the concept base

Furthermore, to make use of existing technology,

*1 goo is a portal similar to Yahoo!

the concept base, which is a topic of research and development at NTT Communication Science Laboratories, is capable of presenting terms that are determined to be similar in concept on the basis of linguistic corpora and Japanese dictionaries and vocabulary systems. Utilizing that technology in the platform may make it possible to absorb semantic variations in the specification of metadata properties such as “creator” and “author” when performing searches for resources.

6. Semantic metadata handling method

The knowledge sharing platform comprises context bureaus and knowledge sharing clients, as shown in Fig. 2.

The context bureau, an application program that resides and operates on a personal computer, manages the metadata that is generated on the computer as each user references and edits documents. In addition to accumulating metadata, it also uses a P2P protocol to search for metadata that has been accumulated by multiple remote context bureaus upon request from a knowledge sharing client.

An application program that is capable of connecting to a context bureau is called a knowledge sharing client. Various services can be offered to the user by connecting multiple clients to a single context bureau. The client acquires the information on relationships among content documents from user operations in application programs, thus generating metadata that is then sent to the context bureau. There are clients that have a function for searching context bureaus for content-related metadata and displaying the results.

7. Knowledge management

The collection of metadata for content documents inside and outside of a company and the sharing of that metadata as knowledge are illustrated in Fig. 2. Making use of information that is buried within systems that are closed to individuals is an important issue in knowledge management.

This figure shows that a user who obtains a certain application form template file from the Web site of a related section in the company (Fig. 2, lower left) knows that another user has applied in the past (Fig. 2, upper left) and that the file of the filled-out form has been saved and its content can be referred to

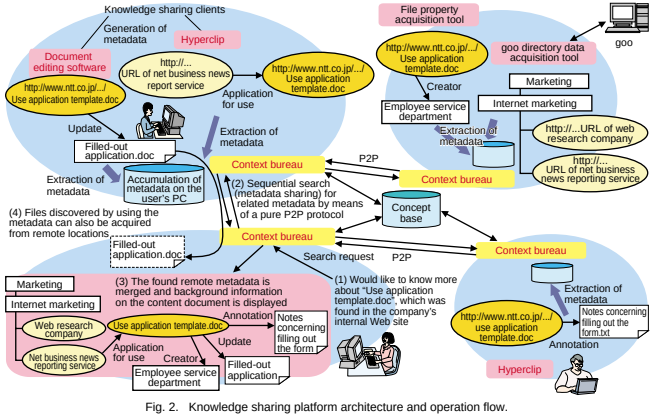


Fig. 2. Knowledge sharing platform architecture and operation flow.

immediately. It also shows that a third user (Fig. 2, lower right) has written some notes concerning the filling out of that form. Also, information extracted from a goo directory (Fig. 2, upper right) enables the user to acquire the location of the service that has been applied for in the directory.

8. Hyperclip: an application program for the platform

Hyperclip [3] is a tool that we have proposed as a client application for the knowledge sharing platform. It resides on the desktop of the user's personal computer and allows the user to 'clip' various kinds of content documents in a bookmark-like manner. This clipping is accomplished by a simple drag-and-drop operation as the user deals with various kinds of local files and e-mail messages, or browses pages on the Web. If the icon of a new content item is dragged onto an existing clip, a dialog window prompts the user for the relationship between the two content items, allowing the user to select content relationships that correspond to RDF properties ("is referenced by", "has new version", "is comparable with", "is opposite to", *etc.*) from a pull-down menu.

An example of Hyperclip's graphical user interface (GUI) is shown in Fig. 3. The upper part of Fig. 3 shows the situation where one user has registered information for his or her own use. A clipping was made of the file named "Use application template" and then another document in a different format that was found at the same time was clipped and the two documents were linked with the relationship "is similar to". The relationship between content items is represented by semantic icons in the GUI. Double-clicking on a semantic icon allows the relationship to be edited. Double-clicking on the title of a content item brings up the relevant application program to open the file for viewing.

Next, by right-clicking on the document title, the user can search for remote metadata that other users have generated. For example, to search for content related to "Use application template," a search can be made with the URI of the content item as a search key. If there are many relevant items, the scope of the search can be narrowed by using property values. The lower part of Fig. 3 shows a display of content items that are related to the content item that serves as the search key. We can see that a file that contains an application form that has been filled out on the basis of the template is registered by remote user "hiroyuki" and a file of notes concerning the filling out of the form has been registered by "yutaka". The degree to which a user's documents and profile are accessible by others can be set by the user.

9. Application to communities

If the relationships among content items on a network can be described by metadata that is independent of the content documents, then other users that access the metadata can add different annotations or new relationships about the content without affecting the original content or metadata, thus expanding the metadata space. If the RDF metadata maintained by individual users is collected and shared, it is possible to form a dynamic community that differs from BBS (Bulletin Board System) like communication spaces,

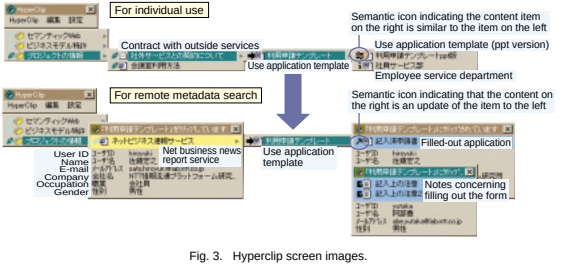


Fig. 3. Hyperclip screen images.

which are provided only to users that access a certain place on a network.

10. Conclusion

Centralized management of the huge quantity of contradiction-containing RDF metadata that is generated by individual users with the knowledge sharing platform is considered to be an impossible task. Therefore, we employed a P2P approach in which the search scope and groups can be specified according to need. We believe that the development of a platform that dynamically collects any RDF metadata without restriction to a particular domain, repository management, or ontology translation among multiple services will contribute to the evolution of the future Semantic Web. We further believe that, by embracing the RDF metadata specified by the grammar as an intermediate form for metadata conversion, this technology can be applied in the near future to the cooperation among multiple metadata services as well as to services that employ RDF. If the utility value of general-purpose metadata becomes higher than the cost of creating it, we believe that RDF will rapidly come into widespread use. We will continue with the standardization of a platform for immediately providing the optimum solution when that time comes.

References

[1] http://www.w3.org/RDF/
[2] http://www.w3.org/2001/sw/
[3] H. Sato, Y. Abe, and A. Kanai, "Hyperclip: a Tool for Gathering and Sharing Meta-Data on Users' Activities by using Peer-to-Peer Technology" WWW2002 Workshop on Real-world RDF and Semantic Web Applications. http://www.cs.rutgers.edu/~shklar/www11/final_submissions/paper12.pdf

Hiroyuki Sato
Member, Software Architecture Project, NTT Information Sharing Platform Laboratories.
He received the B.S. degree from the University of Library and Information Science and the M.S. degree from Keio University in 1994 and 1996, respectively. In 1996, he joined NTT. In 1998, he received the outstanding paper award at ICCE98. He is a member of the Information Processing Society of Japan and the Human Interface Society.

Yutaka Abe
Member, Software Architecture Project, NTT Information Sharing Platform Laboratories.
He received the B.S. and M.S. degrees in computer science from Keio University, in 1999 and 2001, respectively. In 2001, he joined NTT.

Yukihisa Katayama
Senior Research Engineer, Network Application Platform Project, NTT Information Sharing Platform Laboratories.
He received the B.S. degree in electronic-mechanical engineering and M.S. degree in computer science from Nagoya University, in 1989 and 1991, respectively. In 1991, he joined NTT. He is a member of Japan Society for Software Science and Technology, The Society of Instrument and Control Engineers, and The Robotics Society of Japan.