Hyperclip: a Tool for Gathering and Sharing Meta-Data on Users' Activities by using Peer-to-Peer Technology

Hiroyuki Sato, Yutaka Abe and Atsushi Kanai
NTT Information Sharing Platform Laboratories, NTT Corporation
3-9-11, Midori-cho Musashino-shi
Tokyo 180-8585 Japan
Tel: +81 422 59 2614
sato.hiroyuki@lab.ntt.co.jp

ABSTRACT
We have developed a tool called "Hyperclip" for gathering and sharing meta-data which represents background information about electronic content to enhance intellectual creative work without imposing much burden on users. In this research, we call this meta-data "context"; it is composed of links among various contents, and has the flexibility to allow users to add descriptions freely. We propose the idea of handling context as knowledge that is extracted from users' activities such as referring to, creating, and reviewing electronic documents. This paper explains a method of describing the context using XML (RDF or XLink). We also propose an architecture for sharing the context among users via peer-to-peer technology. This architecture is called knowledge sharing platform. Hyperclip is implemented on this platform. These help users better understand how the existing contents can be reused by showing context, for example "which content was referred when it was created" and "who used it and with what".

Keywords
XML, RDF, XLink, Knowledge sharing, Knowledge management, Peer-to-peer, Content, Context, Meta-data, Semantic Web

1. INTRODUCTION
With the advance of electronic media, people who work in ordinary offices are now often performing their jobs through "document-based activities" such as referring to, creating, and reviewing electronic documents. The amount of content, such as HTML, MS-Word, MS-PowerPoint, PDF documents, on networks has been steadily increasing. As a result, it has become difficult to efficiently reuse the huge amount of existing content for intellectual creative work.

To solve the problem, in this research, we propose the idea of handling semantic links among content as meta-data about content extracted from users' document-based activities. We have developed a tool called "Hyperclip" for gathering and sharing meta-data which represents background information about electronic content without imposing much burden on users. We represent the link using the features of Resource Description Framework (RDF)[1] or XML Linking Language (XLink)[2]. We have also developed an architecture for sharing links among users by using peer-to-peer (P2P) technology that is used as platform for Hyperclip.

2. HOW TO REUSE EXISTING CONTENTS FOR CREATIVE WORK?
2.1 Showing the background would help users
To enhance intellectual creative work, it is important to help users to be aware of useful content buried in computers and on networks. To create something new and to do a new job, one needs

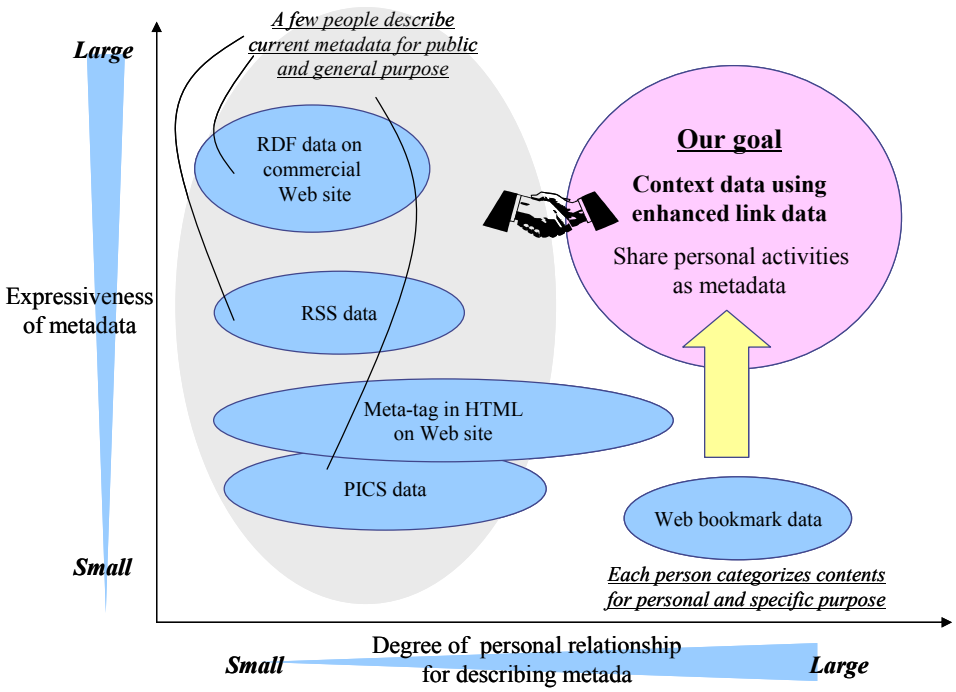

Figure 1. Present state of metadata and our goal area.

reference information to get existing knowledge and know-how. Today, many search engine services on networks succeed in enabling users to find a huge amount of content that corresponds to keywords.

In addition to the content found by search engines, background information such as "when the content was created and in what kind of situation" or "who used it and how" would help people understand how the content can be used. This would enable users to understand if the content is useful and how to reflect it in their own creative work. There are some existing methods of adding metadata, which represents various properties, to content. RDF and Dublin Core[3] provide users with rules and elements for describing meta-data that show the characteristics of content.

2.2 Problems with metadata

Meta-tags are suitable for representing static information such as the content creator's name and the creation date, but they are not suitable for describing dynamically changing background information. Some background information about content is based on user activity, which differs from one individual to another. It is difficult to describe the background as one static characteristic by using a fixed meta-tag.

People will not cooperate with the description of metadata if it is troublesome to input, so it is difficult to gather metadata for complicated situations about the background of contents from most Web sites especially ones that are administered non-commercially.

2.3 Representing "context" by using a link

In this research, we focused on links among content. There are some discussions about link that has meanings[4][5][6]. To improve the reusability of the content for creative work, we propose a way of describing a link among content as background information and we propose sharing the link among users. The link can be independent of the content, can be dynamically updated by other users, and

```
<?xml version="1.0"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:context="http://www.ntt.co.jp/2000/context#">
  <rdf:Description rdf:about="http://www.ntt.co.jp/docA.txt">
    <context:modifiedVersion rdf:resource="http://www.ntt.co.jp/docB.txt" />
  </rdf:Description>
</rdf:RDF>
```

Figure 2. Description of "modifying a document" by using RDF.

can represent the work process by using two or more items of content. Linked contents form a group, which enables various kinds of background and characteristics to be shown by their contents and the links among them. Background information represented by a link that is based on the user's activities is called "context" in this paper.

Figure 1 shows the position and goal of our research. Current metadata is described by a small number of people, such as the content author and official organizations. It is used for public and general purposes, for example filtering or indexing [7][8]. Our goal is to enhance existing metadata by collaboratively gathering many kinds of personal activities related to content.

2.4 Simple examples of context

The following are examples of context, which guide a person who wants to access or reuse some particular content efficiently for creative work.

2.4.1 Grouping the same content into different categories

Different users whose objectives are also different often classify the same content on the network in different ways. The location of content in a directory structure, such as a document file system or Web bookmark, can be important background information.

2.4.2 Arranging contents according to the specific situation of each user

Some links among contents can show a kind of process for specific work. Each user's activity when they handle contents in document-based work, such as quotation and reference, can help to form the context.

3. METHOD OF DESCRIBING CONTEXT

The context must be independent from existing content on the networks, because, as mentioned above, it can be updated by different users who are not related to the content creator. The context must be flexible with respect to being changed dynamically at different times. To fulfill these conditions, we utilize the RDF data model or XLink features.

3.1 Describing context using RDF

Figure 2 shows a description of the relationship between two items of content (http://www.ntt.co.jp/docA.txt and http://www.ntt.co.jp/docB.txt) using RDF. Here, the expression "modifiedVersion" is a "predicate (property)" of RDF and is used to show that the "subject" of RDF (http://www.ntt.co.jp/docA) also has a modified version. And the "object" of RDF is another resource (http://www.ntt.co.jp/docB.txt). The statement in Figure 2 means that the modified version of http://www.ntt.co.jp/docA.txt is http://www.ntt.co.jp/docB.txt.

```
<?xml version="1.0"?>
<context xmlns:xlink="http://www.w3.org/1999/xlink"  xlink:type="extended">
  <locator xlink:type="locator"
    xlink:href="http://www.ntt.co.jp/docA.txt"  xlink:label="docA" />
  <locator xlink:type="locator"
    xlink:href="http://www.ntt.co.jp/docB.txt"  xlink:label="docB" />
  <arc xlink:type="arc"
    xlink:from="docA"  xlink:to="docB"
    xlink:arcrole="http://www.ntt.co.jp/2000/context/modifiedVersion" />
</context>
```

Figure 3. Description of "modifying a document" by using XLink.

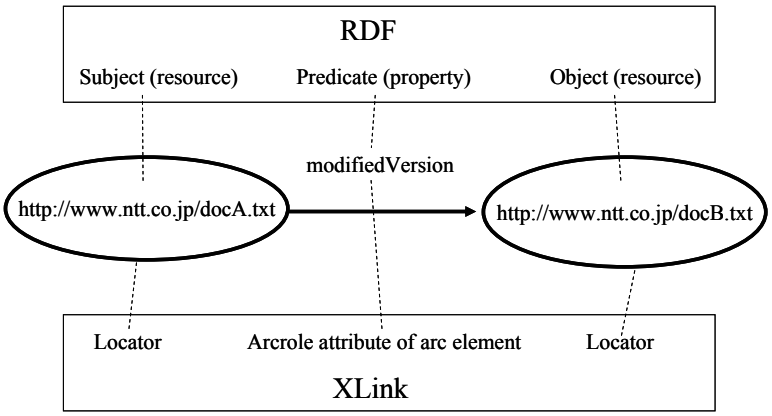

Figure 4. Representation of "modifying a document" by directed graph.

3.2 Describing context using XLink

XLink enables elements to be inserted into XML documents to create and describe links between different resources. It uses XML syntax and enhances the simple hyperlinks of today's HTML. One of its differences from HTML is that XLink can express links that associate one resource with an arbitrary number of other resources simultaneously. Figure 3 shows the same relationship as in Figure 2 using XLink. Two locator elements that have the "href" attribute for the reference of each URL of the document are linked with arc. The value of the "arcrole" attribute indicates that the document was modified.

3.3 Representing context by a directed graph

The above are very simple ways of representing context. Figure 4 shows that Figures 2 and 3 can be represented by the same directed graph. These two methods have two common features. They

- express links that reside in a location separate from the linked resources and
- describe the meanings of the links.

4. ARCHITECTURE FOR KNOWLEDGE SHARING PLATFORM

We have developed an architecture for sharing contexts by using peer-to-peer (P2P) technology, because it is important for creative work to share contexts among a lot of distributed users[9]. Figure 5 shows the concept of exchanging context data extracted from a user's activities with other

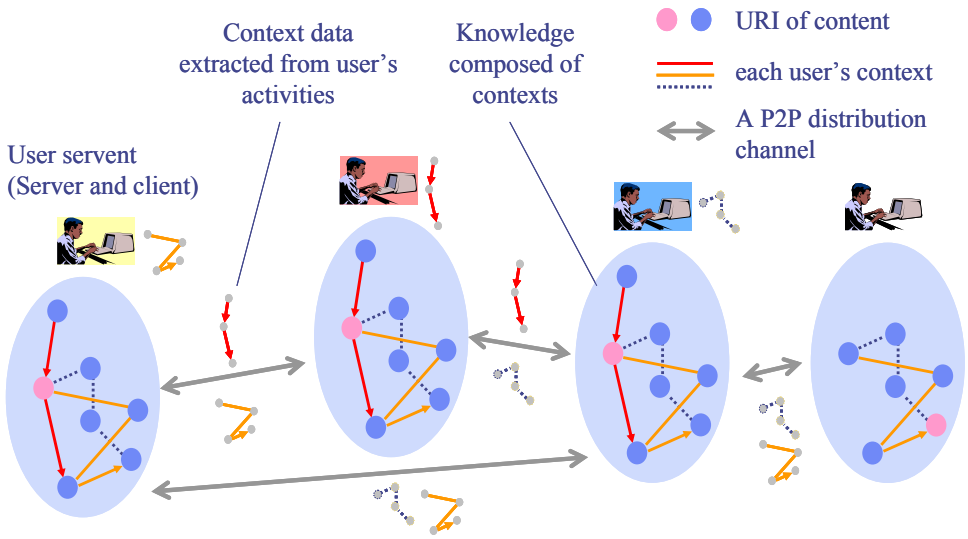

Figure 5. Concept of exchanging contexts with other users.

users. Any user can view the knowledge composed of gathered contexts, for example to create a new document.

4.1 Context Collection

Contexts can be produced by users' activities related to the handling of content by using various kinds of applications, such as a file manager or a Web browser and its bookmark editor. The knowledge sharing platform should store contexts without placing much burden on users, because users will not cooperate with the context collection if it is troublesome for them. Therefore, we provide applications with a function that monitors the relationship among contents when users use these applications, and map it to XML as shown in Figures 2 and 3. Applications that have this function are called "knowledge sharing clients". Hyperclip is one of the knowledge sharing client applications. They communicate with a server called a "context bureau", which also works on the same user PC (see Figure 6). Of course it could exist in any domain, for example an intranet or a particular PC on the Internet. It receives the context from the clients and stores it.

4.2 Architecture for sharing

If the context bureau receives contexts from applications like every user's Web browser, the volume of context information could become huge. It would be difficult to administer the contexts on one server, because different users can produce different contexts about the same content on the network, and these are dynamically changeable and may include redundancy or inconsistencies.

For that reason, we manage contexts by distributed management. We use a P2P protocol like Gnutella[10], which enables file swapping. Gnutella administers a target file for distribution by an entity called a Gnutella servent (coined from server-client, because it can act as both client and server). Files within the limits of a specified hop count are possible target files. Contexts can also be distributed by using the context bureaus as Gnutella servents. The context bureaus communicate with each other by P2P protocol over the network. A bureau that is searching for the context of a content item asks others if they have a context that is related to the same content. If the bureau finds the context, it shares it with bureaus in different domains by swapping XML files.

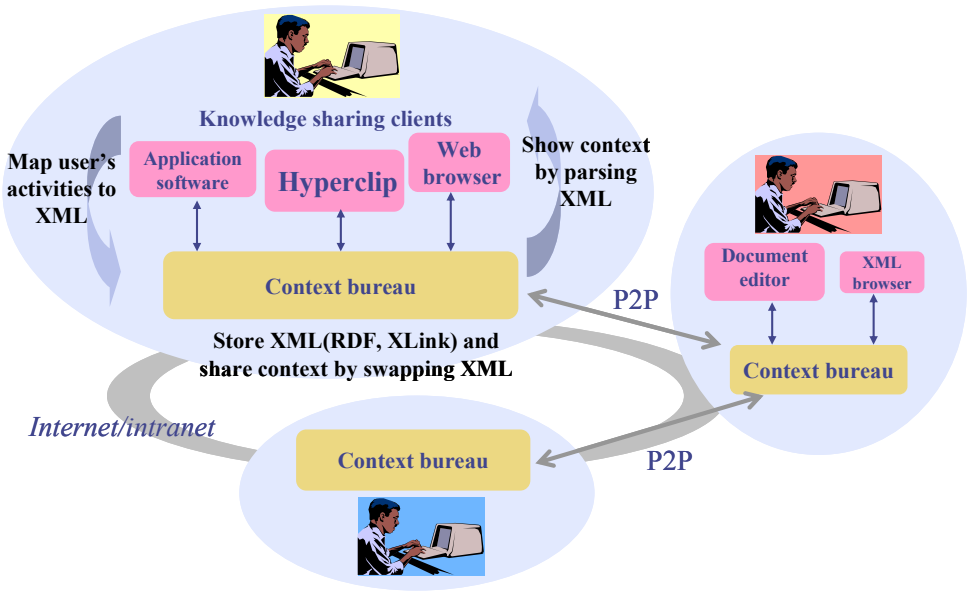

Figure 6. Overview of architecture for knowledge sharing platform.

5. HYPERCLIP

We propose a tool called Hyperclip for clipping one content to another while allowing users to describe the semantic link between them. The user action for registering a content item with the system like a bookmark is called "clipping" in this research. Current bookmarks are only categorized in a tree structure and mainly handle Web pages, but Hyperclip can handle various resources and context that can be described by using RDF or XLink.

5.1 Basic user interface

Hyperclip is designed to be always available on every user's PC desktop. Imagine we are studying documents about RDF. We can clip a content item found on the Internet or in an intranet to another content item stored in Hyperclip by a simple drag&drop operation.

Hyperclip handles various kinds of contents as resources that have URIs, for example not only Web pages using html, but also MS-Word, MS-Powerpoint, and PDF files, and e-mail, and so on.

5.1.1 Semantic Icon

Figure 7 shows the graphical user interface of Hyperclip. Resource titles are shown in a tree structure. And each title has two icons. One shows the content type, which indicates related application software for viewing or editing this resource. If the user double-clicks the resource title, he/she can view this content using that application. The other icon is a "semantic icon" that shows the relationship between this resource and the next one on the left.

Semantic icons are provided according to the specific RDF Schema[11] or vocabulary, for example "office work". "Office work", as we use it here, has some properties of RDF. Figure 7 shows that the "modified version of RDF [PR] (PR: proposed recommendation) is RDF [REC] (REC: recommendation)". It also shows that the "sequential document (which the user is expected to look at next) of RDF [REC] is RDF Schema Specification", "translation (Japanese version) of RDF [REC] is RDF in Japanese", and so on.

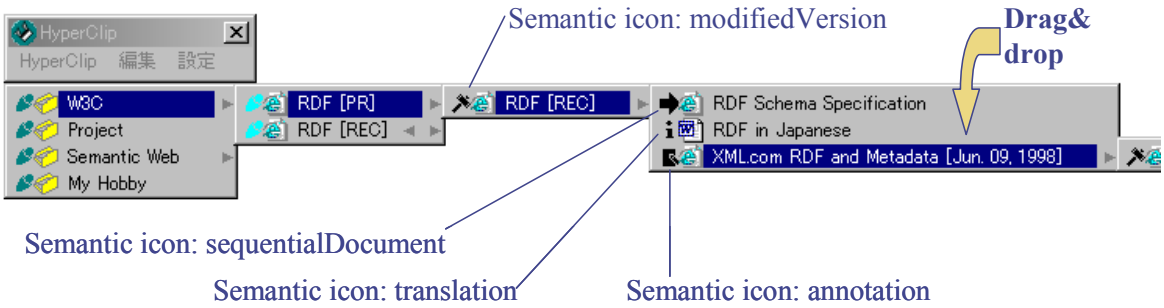

Figure 7. GUI of Hyperclip showing semantic links between resources using semantic icons.

Figure 8. A dialog window for adding meanings of relationships when content is clipped

5.1.2 Adding meanings of relationships when content is clipped

When a user clips a resource to one in Hyperclip, a dialog window appears (see Figure 8). The window asks the user to give the relationship for the RDF property and the user can easily select it by using a pull-down menu.

5.1.3 Example of RDF data created by Hyperclip

Figure 9 represents the context in Figure 7 as RDF data.

5.2 User interface for searching for remote context

The way to return the context, which is produced by applications like Hyperclip and stored in a distributed context bureau, to users in a timely manner is important. Users can access the context bureau and receive a context or can ask it to search for another context in different domains.

When the user selects a resource title and clicks the right button of the mouse, he/she enters search mode. If another user on the network has clipped resources that are related to the selected one, the user can see them with the same GUI, as shown in Figure 10. Resources that are found on a remote

```
<?xml version="1.0"?>
<rdf:RDF
 xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
 xmlns:context="http://www.ntt.co.jp/2000/context#">
  <rdf:Description rdf:about="http://www.w3.org/TR/1999/PR-rdf-syntax-19990105/">
    <context:title>RDF [PR]</context:title>
    <context:modifiedVersion>
     <rdf:Description rdf:about="http://www.w3.org/TR/1999/REC-rdf-syntax-19990222/">
       <context:title>RDF [REC]</context:title>
       <context:sequentialDocument
          rdf:resource="http://www.w3.org/TR/2000/CR-rdf-schema-20000327/" />
       <context:translation rdf:resource="http://www.ntt.co.jp/sato/RDFinJapanese.doc" />
     </rdf:Description>
    </context:modifiedVersion>
  </rdf:Description>
</rdf:RDF>
```

Figure 9. Description of context shown in Figure 7 by using RDF.

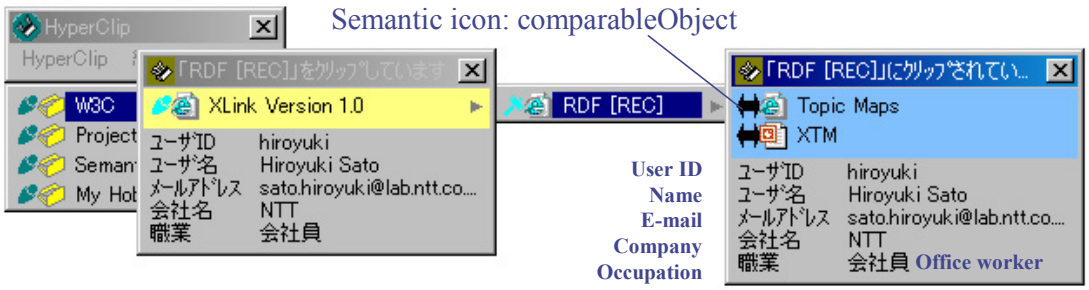

Figure 10. Showing context of remote user on the same GUI.

PC are put next to the selected one on the right and left. And Hyperclip shows who created the relationships below the found resources by indicating the user profiles. Users can limit access to their profiles to the persons looking for the context, for example to anyone or to a specific group, and so on.

If there are too many search results for Hyperclip to display, then the user can select some of them by using RDF properties in a context. Figure 10 shows that the user got the system to display resources linked by the RDF property "comparable object". Therefore, the GUI shows that "RDF [REC] has a comparable object called Topic Maps", and so on.

6. DISCUSSION ABOUT APPLICATIONS

Here we examine the merits of Hyperclip and the knowledge sharing architecture described in section 4 when they are applied to various situations.

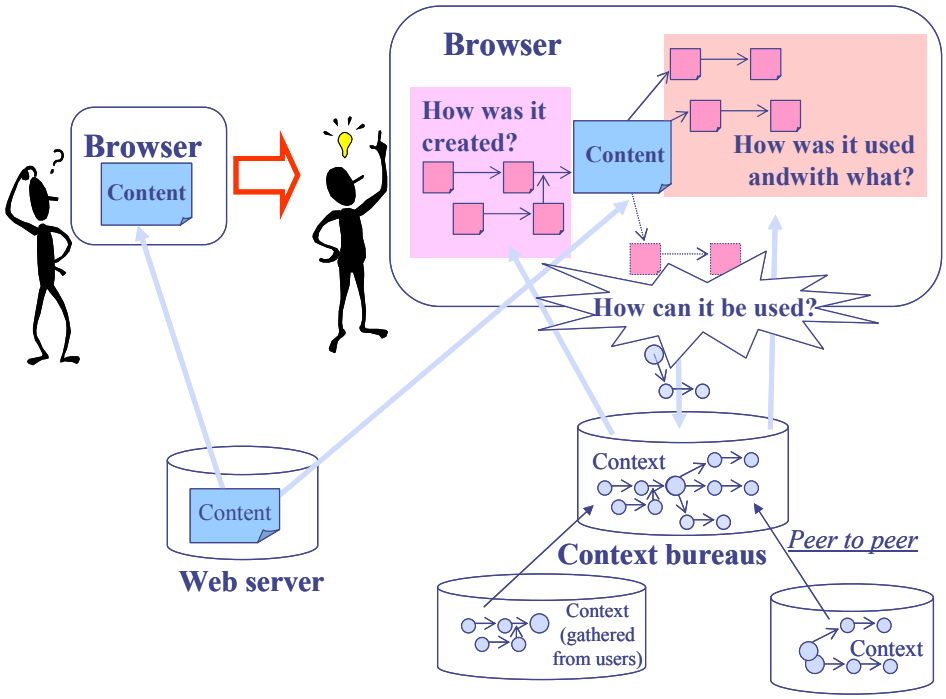

Figure 11. Service concept of an application based on a knowledge sharing platform.

6.1 Search engine

In addition to Web search engines based on keyword searching, we are trying to utilize the meanings of links in context to improve the precision of search results.

6.2 Knowledge management

Knowledge management [12], which should improve the efficiency of business and the quality of products, has been receiving a lot of attention. There are two main approaches to it. One is to reveal implicit know-how held by individuals and share it among groups. The other is to make users aware of knowledge buried amid the huge amount of content, such as documents used in an earlier project.

Hyperclip will enable office workers to discover the background of a content item, such as "who used it and for what project in which situation". Simply by traversing a link, they may share individual knowledge about work that has no fixed workflow or description in a business manual.

6.3 Community

If other users share their contexts via the context bureau, a user can see them. The user can give feedback to these other users by adding a new context to an existing one. As a result, the bureaus provide users with a communal knowledge space that is formed by dynamically collecting existing contexts from many people.

7. CONCLUSIONS

We proposed a way of using XML (RDF or XLink) to represent background information about an item of content (i.e., its context) and sharing the context, which is composed of links among contents, by using P2P technology. We also developed a knowledge sharing platform and a tool called "Hyperclip" to enhance intellectual creative work. They let users easy create metadata from

user activity and understand the background of contents, for example “how the existing content can be reused”, through sharing other users’ contexts as knowledge.

In the near future we will share various kinds of metadata created by other applications on our developed platform. We are now trying to handle both RDF and XLink on the same platform because, as the W3C note [13] shows, it is not difficult to transform between them.

We will continue our work to determine how effective our system is for creative work by applying it to the situations described in section 6. To do this, we may have to implement XPointer to point more precisely to a resource in a content item, as is possible in the Anotea project [14]. We will also try to understand the vocabulary or schema required for the context in each situation more specifically.

8. REFERENCES

[1] Ora Lassila and Ralph R. Swick, Resource Description Framework (RDF) Model and Syntax Specification, http://www.w3.org/TR/1999/REC-rdf-syntax-19990222", 1999.

[2] Steve DeRose and Eve Maler and David Orchard, XML Linking Language (XLink) Version 1.0, http://www.w3.org/TR/2000/REC-xlink-20010627/, 2001.

[3] Diane Hillmann, Using Dublin Core, http://dublincore.org/documents/2001/04/12/usageguide/, 2001.

[4] Amit Sheth, Vipul Kashyap, Media-independent correlation of Information: What? How?, Proceedings of the First IEEE Metadata Conference, 1996

[5] Harri Oinas-Kukkonen, What is inside a link?, Communications of the ACM 41, 7, 1998.

[6] Kenji Takahashi, Metalevel Links: More Power to Your Links, Communications of the ACM 41, 7, 1998.

[7] Platform for Internet Content Selection (PICS), http://www.w3.org/PICS/

[8] Tim Bray, XML.com: What is RDF?, http://www.xml.com/pub/a/2001/01/24/rdf.html

[9] Hiroyuki Sato, Kenji Otomo, and Tsuyoshi Masuo, A Knowledge Sharing System using XML Linking Language and Peer-to-Peer Technology, Proceedings SAINT2002, IEEE Computer Society, 2002.

[10] Gnutella Protocol Specification v0.4, http://dss.clip2.com/GnutellaProtocol04.pdf

[11] Dan Brickley and R.V. Guha, Resource Description Framework (RDF) Schema Specification 1.0, http://www.w3.org/TR/2000/CR-rdf-schema-20000327/, 2000.

[12] Thomas H. Davenport and Laurence Prusak, Working Knowledge: How Organizations Manage What They Know, Harvard School Business Press, 1998.

[13] Ron Daniel Jr., Harvesting RDF Statements from XLinks, http://www.w3.org/TR/2000/NOTE-xlink2rdf-20000929/, 2000.

[14] Annotea Project, http://www.w3.org/2001/Annotea/