A Proposal of Community-based Folksonomy with RDF Metadata

Ikki Ohmukai[1], Masahiro Hamasaki[2], and Hideaki Takeda[1]

[1] National Institute of Informatics (NII)
2-1-2, Hitotsubashi, Chiyoda-ku, Tokyo, Japan
[2] National Institute of Advanced Industrial Science and Technology, Japan

Abstract. In this paper we propose a social bookmark system using several metadata and personal network to construct a community-based ontology. Our system allows users to browse friends' bookmarks on his/her personal network. The user can map their own tags onto more than one tag from different friends so that they are linked to the user. The matchmaker-based recommendation and network expansion method will work efficiently because it is derived from personal interest and trust. We also design an RDF-based metadata framework to support the open and distributed model.

1 Introduction

We propose a social bookmark system using several metadata and personal network to construct a community-based ontology. Recently a lot of social software has been released along with " Blog explosion "[1], especially tag-based systems providing a new concept called folksonomy. However, current folksonomy systems do not work efficiently because the tags are arbitrarily attached and are not consistent. It is also difficult to separate the meaning of a homonym with a context.

These problems occur because the systems are designed to manage all tags at one time, so we propose a new system that constructs a desirable tag network by maintaining personal relationships and small communities.

To realize our goal, we should consider that most community-based software does not perform suitably because of a lack of participants. In this paper we designed our system based on the "double-loop gratification principle", which is our proposition[2].

2 Double-loop Gratification Principle

We can enumerate many benefits for community-based information sharing, while there are also dissemination of information sharing hurdles. One of the hurdles are the privacy and security issues related to sociological point of view.

Another issue is the feedback issue related to the cognitive point of view. Feedback on contribution to information sharing is rarely visible. One of the reasons why people do not wish to use information sharing tools is that their effort looks to be in vain because of a lack of feedback. McDowell et al. [3] pointed out this issue as *instant gratification*. They said that instant gratification is needed to involve people in Semantic Web applications, and their application, called Mangrove, has succeeded because of the realization of instant gratification.

We agree with the importance of instant gratification, but instant gratification should differ in information sharing applications. In Mangrove, user contribution is quickly reflected in the information sharing results and services because the system collects and revises them as fast as it can. It is a nice feature but it sacrifices variety and scalability of information sharing, because information sharing naturally takes time.

We think that information sharing applications should simultaneously have two types of gratification, i.e., instant gratification that can be obtained even without information sharing, and delayed gratification that can be obtained through information sharing. It always takes effort for users to become accustomed to new applications. Instant gratification can be an anchor to keep users using applications. While users keep using them, delayed gratification are the real benefits of the information sharing that arrives in them. The balance of these two types of gratification is important than the quantity of them. As I previously mentioned, benefits from information sharing tend to take time, and it is too strict a restriction to require instant gratification by information sharing.

The Web has both types of gratification. Authoring hypertexts gives people instant gratification. It is a new fascinating method for people to organize their own information that is difficult to write down as a stable and well-organized form, like word processing documents. Since authoring hypertexts and publishing them are so closely connected on the Web, people are publishing their information with almost no extra effort. Then they will receive delayed gratification as feedback from other users who read their published information arrives.

The problem is how to design such systems with two types of gratification. Through our observation of other systems and our experience with information sharing, we propose *translucence strategy* to assist people in shifting from instant gratification to delayed gratification receivers. The strategy is simple: just put people in a situation where they can feel possible delayed gratification within kissing distance. Then they shift to the next step where they can receive delayed gratification. The step should be minimized, i.e., it should take a very small amount of extra effort to join information sharing, in addition to the ordinary effort used to obtain instant gratification.

In the following section, we explain the community-based folksonomy system we built and how the above strategy works in it.

3 Community-based Folksonomy

Folksonomy is a bottom-up taxonomy constructed on a socialware[4]. The users of the socialware append several keywords, called tags, onto a bookmark or a photo when they upload that resource on the web. The folksonomy system manages these resources with the tags and provides a list of resources whose tags are the same, so that users can obtain various information in specific contexts. The system also provides a view of tags that are annotated to a single resource by multiple people, and this means how the people read that resource. Besides increasing the number of users, we can find which topics are interesting to ordinary people with a services called "tag cloud"[5].

Currently the scope of the folksonomy system is expanding from bookmarks and photos to music, movies, and spatial information. Compared to a conventional top-down category systems, the folksonomy offers more flexible expressions with less maintenance cost.

However, no one guarantees a quality of tags and the vocabulary in the folksonomy system, since they are generated by the bottom-up process. Each tag is arbitrarily made by separate users and this causes taxonomy inconsistency, such as "weblog", "blog", and "blogging". This problem will seriously affect the quality of a tag-based search.

Another problem is that the meanings of a homonym cannot be separated, since a tag is represented as a text string. For example, a user wants to browse the contents of the word "apple", the system will produce pictures of fruits and articles about computers.

These problems are caused because a folksonomy system uses simple keyword matching of the entire collected data. It is natural that the context of a word disappears when the system tries to generalize the meaning of the word.

To resolve these problems, we do not use system-wide data but introduce personal networks as an infrastructure for the community-based information distribution. If there is an inconsistency in the tags among several different people, they can implicitly appreciate what their friends on the personal network think about them. The homonym problem would hardly ever occur, since it is rare that people in the same community ambiguously interpret a word. As a result, precise word networks will be created in every community.

On the other hand, profit of the community-based system depends on the number of participants who are in the users' personal network. Therefore, we should design our system to provide some benefits for standalone users and encourage recruitment of their friends into the system.

4 Proposed System

4.1 Basic Functions

Our proposed system provides three basic functions: 1) personal contents management, including blog entries and bookmarks, 2) personal network management, and 3) contents recommendations. Figure 1 shows a snapshot of our system. At first the user can register bookmarks, add tags, and maintain a personal network as follows.

- Bookmarking

Fig. 1. Snapshot of our system

When the user finds favorite contents, he/she registers it as a bookmark by calling up bookmarklet, which is a tiny application described by JavaScript. The user can add a comment and tags in the register interface.

– Tagging
  The tagging interface provides an edit function of the tags for bookmarked entries. This interface communicates with the user's own blog tool to synchronize tags lists, so that the user can tag onto his/her contents via a blog tool.
– Tag Surfing
  The users of our system can play tag surfing and view tag cloud, similar to conventional social bookmark services.
– Social Networking
  The user can create personal links with his/her friends by FOAF TrackBack, described below.

4.2 FOAF TrackBack

Our system identifies each user using FOAF[6], which is an RDF-based metadata format for describing human relationships. In the FOAF specification, a "knows" statement in two separate people 's sections means bidirectional links between them. To easily generate bidirectional links, we propose a "FOAF TrackBack" procedure with our system[7].

First, the user X enters user Y's URI in his/her own FOAF manager. The manager X asks manager Y to acquire the FOAF data of Y, and writes "X knows Y" link in its FOAF. Manager Y records "Y is known by X" link in its FOAF and notifies to user Y. If user Y agrees, his/her manager registers the "Y knows X" link. Repeating this process, the personal network of the user is constructed. The following recommendation methods are performed in the network.

5 Mapping and Recommendation

In our system, the user can browse friends' bookmarks on his/her personal network. All of their bookmarks are available, but he/she can choose contents with selected tags. When the user selects a friend's tag, he/she should choose from the following options: 1) map with own tag, 2) import friend's tag, and 3) classify into "untagged", as shown in Figure 2.

The user can map their own tag onto more than one tag from different friends, so that they are linked by the user. Then the system delivers one's tags and bookmarks as recommendations to the others, and the receiver will choose from the map/import/untagged or deny these contents. If he/she chooses the former options, the system makes FOAF relationships between the sender and the receiver. The matchmaker-based recommendation and network expansion method will work efficiently because it is derived from personal interest and trust.

This method also resolves the inconsistency in the notation of the tags. Because the mapping procedure mentioned above does not consider the text string of tags and just transforms bookmarks from one's tag to the other's tag. Consequently, the system may aggregate some bookmarks into "blog" tags for user A and "blogging" for user B, however, the system knows the equivalence between the "blog" for A and the "blogging" for B and can make reasoning on these tags.

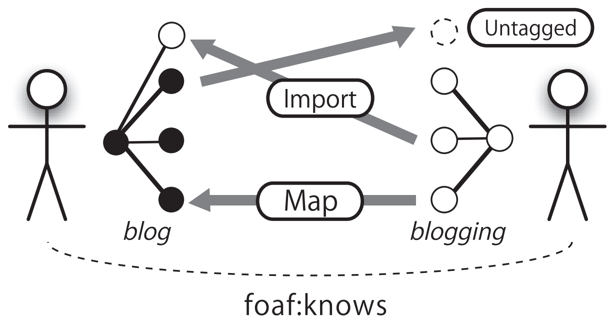

Fig. 2. Tag Mapping

6 Metadata Design

Most of the socialware is designed with a centralized server model to hold its users, so that a data structure of the system does not have to be based on XML or RDF. However, we designed the RDF-based metadata framework to support open and distributed models.

We enhanced our "personal ontology framework" suitable for community-based folksonomy. The personal ontology framework was proposed to represent personal knowledge hierarchy using FOAF, RSS, and simple RDFS. It has interoperability with conventional blog tools.

As shown in Figure 3, personal information is described in FOAF, and a knowledge structure is depicted in RDFS, and the contents RSS shows blog entries and bookmarks by the user.

We added two elements to the basic FOAF model shown in Figure 4. One is `<foaf:interest>`, which is to point out the contents RSS, and the other is `<rs:personalontology>`, that is originally defined by our Rough Semantics project [3] to indicate the RDFS tag ontology.

The RDFS tag ontology is described with the form of Open Directory RDFS format shown in Figure 4(b). Each node has a fragment ID.

The contents RSS is similar to a conventional RSS. Our RSS uses `<foaf:topic>` to point out a category on the RDFS ontology, while the conventional model applies `<dc:subject>` to express a thesis of a content. This feature makes our RSS use a backward compatibility. An example of this RSS is shown in Figure 4(c).

FOAF, RDFS tag ontology, and the contents RSS are described in separate files so that we can keep a compatibility with existing applications on these formats. It is a great benefit that our system can cope with such existing applications via these files.

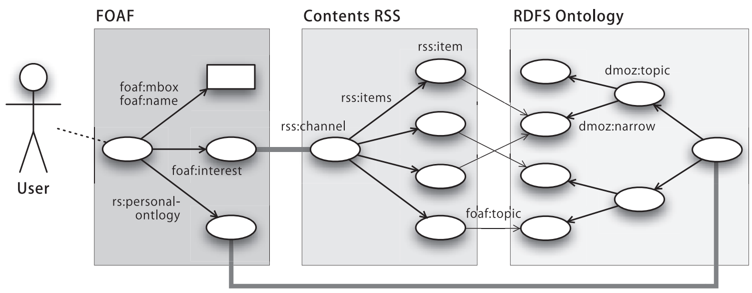

Fig. 3. Personal Ontology Framework

7 Conclusion

In this paper we propose a social bookmark system using several metadata and personal network to construct a community-based ontology. Our system allows users to browse friends' bookmarks on his/her personal network. The user can map their own tag onto more than one tag from different friends, so that they

[3] `http://www.roughsemantics.org/`

```
<rdf:RDF
 xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
 xmlns:foaf="http://xmlns.com/foaf/0.1/"
 xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
 xmlns:rs="http://www.roughsemantics.org/rs/0.1/"
>

<foaf:Person>
<foaf:name>Ikki Ohmukai</foaf:name>
<foaf:nick>i2k</foaf:nick>
<foaf:mbox rdf:resource="mailto:i2k@grad.nii.ac.jp" />
<foaf:weblog rdf:resource="http://www.semblog.org/i2k/" />
<rdfs:seeAlso rdf:resource="http://www-kasm.nii.ac.jp/~i2k/foaf.rdf" />
<foaf:interest rdf:resource="http://www-kasm.nii.ac.jp/~i2k/index.rdf" />
<rs:personalontology
            rdf:resource="http://www-kasm.nii.ac.jp/~i2k/ontology.rdf" />
....

</foaf:Person>
</rdf:RDF>
```

(a) Extended FOAF

```
<RDF xmlns:rdf="http://www.w3.org/TR/RDF/"
     xmlns:dc="http://purl.org/dc/elements/1.0/"
     xmlns="http://directory.mozilla.org/rdf">

<Topic rdf:id="Top">
  <tag catid="1"/>
  <dc:Title>Top</dc:Title>
  <narrow rdf:resource="Top/Arts"/>
  <narrow rdf:resource="Top/Business"/>
  <narrow rdf:resource="Top/Economy"/>
  <narrow rdf:resource="Top/Tech"/>
  ....
</Topic>

<Topic rdf:id="Top/Arts">
  <tag catid="2"/>
  <dc:Title>Top/Arts</dc:Title>
  <narrow rdf:resource="Top/Arts/Fine"/>
  ....
</Topic>
```

(b) RDFS Ontology

```
<item rdf:about="http://www.semblog.org/i2k/archives/000304.html">
<title>Blog Hacks</title>
<link>http://www.semblog.org/i2k/archives/000304.html</link>
<description>
Monday's child is fair of face, Tuesday's child is full of grace,
Wednesday's child is full of woe, Thursday's child has far to go,
Friday's child is loving and giving, Saturday's child works hard for his living,
And the child that is born on the Sabbath day is bonny and blithe, and good and gay. ...
</description>
<dc:subject>trivia</dc:subject>
<foaf:topic rdf:resource="http://www-kasm.nii.ac.jp/~i2k/ontology.rdf#Top/Arts">
<dc:creator>i2k</dc:creator>
<dc:date>2004-04-09T01:24:16+09:00</dc:date>
</item>
```

(c) Contents RSS

Fig. 4. Personal Ontology Metadata

are linked by the user. The matchmaker-based recommendation and network expansion method will work more efficiently because it is derived from personal interest and trust. We also design an RDF-based metadata framework to support the open and distributed model.

References

1. R.Blood: The Weblog Handbook. Perseus Publishing (2002)
2. H.Takeda, I.Ohmukai: Building Semantic Web Applications as Information/Knowledge Sharing Systems. Workshop on End User Aspects of the Semantic Web, Colocated with European Semantic Web Conference (ESWC2005) (2005)
3. L.McDowell, O.Etzioni, S.D.Gribble, A.Halevy, H.Levy, W.Pentney, D.Verma, S.Vlasseva: Mangrove: Enticing Ordinary People onto the Semantic Web via Instant Gratification. Proceedings of the Second International Semantic Web Conference (ISWC2003) (2003)
4. A.Mathes: Folksonomies - cooperative classification and communication through shared metadata. Technical report, University of Illinois Urbana-Champaign (2004)
5. Sifry Consulting: Technorati. `http://www.technorati.com/` (2002)
6. D.Brickley and L.Miller: FOAF Vocabulary Specification. `http://xmlns.com/foaf/0.1/` (2004)

7. I.Ohmukai, H.Takeda, K.Numa, M.Hamasaki, S.Adachi: Metadata-driven Personal Knowledge Publishing. Proceedings of the Third International Semantic Web Conference (ISWC2004) (2004)