Journal of Universal Computer Science, vol. 11, no. 2 (2005), 213-228
submitted: 31/8/04, accepted: 15/12/04, appeared: 28/2/05 © J.UCS

Collaborative Web Browsing Based on Semantic Extraction of User Interests with Bookmarks

Jason J. Jung
(Intelligent E-Commerce Systems Laboratory,
School of Computer and Information Engineering,
Inha University, Korea
j2jung@intelligent.pe.kr)

Abstract: With the exponentially increasing amount of information available on the World Wide Web, users have been getting more difficult to seek relevant information. Several studies have been conducted on the concept of adaptive approaches, in which the user's personal interests are taken into account. In this paper, we propose a user-support mechanism based on the sharing of knowledge with other users through the collaborative Web browsing, focusing specifically on the user's interests extracted from his or her own bookmarks. Simple URL based bookmarks are endowed with semantic and structural information through the conceptualization based on ontology. In order to deal with the dynamic usage of bookmarks, ontology learning based on a hierarchical clustering method can be exploited. This system is composed of a facilitator agent and multiple personal agents. In experiments conducted with this system, it was found that approximately 53.1% of the total time was saved during collaborative browsing for the purpose of seeking the equivalent set of information, as compared with normal personal Web browsing.

Keywords: Web Browsing, Collaborative Works, Ontology
Categories: H.3.1, H.3.3, H.5.3, H.5.4

1 Introduction

With the development of network technologies, the amount of information available on the World Wide Web has been increasing exponentially. Navigating in a search for relevant information in this Web environment is one of the most lonely and time-consuming tasks [Maes, 94]. There have been numerous studies designed to deal with this problem of "*information overload*", most of which have involved in user profiling through analyzing the behaviors of each user. For example, the personal assistant agent system can predict the reactions of the user, thereby enabling it to perform such actions as removing junk e-mails from the mailbox, or, while browsing, to proactively prefetch and show candidate Web pages based on the user's preferences [Lieberman, 95]. In contrast to these single user-centered approaches, in this study, we make use of collaboration among multiple users as another way of improving the performance of information retrieval. In this paper, we introduce collaborative Web browsing, which is an approach whereby users share knowledge with their like-minded neighbors while searching for information on the Web. By communicating with others, users can acquire many kinds of experiences (or heuristics), such as how to select and rank the search results, how to make an appropriate sequence of queries, and how to choose the best searching method, as well as providing the other users with their own knowledge. More importantly, we focus on those items of information which are related to the user's interests. In collaborative Web browsing, we consider that recognizing the user's interests is a very important task. Moreover, asking relevant information for other users, filtering the query results, and recommending them are additional major tasks that have to be implicitly conducted.

In this paper, we introduce the extended application of a *BISAgent*, which is a bookmark sharing agent system based on the modified *TF-IDF* scheme [Jung, 00]. We extend the system proposed in this previous work, by endowing it with the capability of recognizing user preferences. Typically, a bookmark is always stored on the client's computer and contains the relevant URL information, with this function being built in to the various Internet Web browsers, such as the *Mosaic* Web browser, *Netscape* browser, and Internet Explorer (referred to as *Favorites* within the MS-Windows platform).

[DEFAULT]
BASEURL = ***http://www.moma.org/***
[InternetShortcut]
URL = ***http://www.moma.org/***
Modified = 00B19BFB5C49C401B1

Table 1: Example of bookmark file of "Museum of Modern Art"

For example, bookmarking the "Museum of Modern Art" Web site results in the creation of a local file containing the URL information generated on the client-side. The bookmark file is shown in [<see> Tab. 1]. According to the *GVU*'s survey [GVU, 97], the number of bookmarks is in a state of constant increase. In effect, the set of bookmarks in the user's folder can be regarded as a piece of information which can be used to infer the user's interests [Jung, 03]. In order to uncover the user's interests from his or her own bookmarks, we employ an ontological supervisor which can perform the semantic analysis of the Web sites pointed to by these bookmarks.

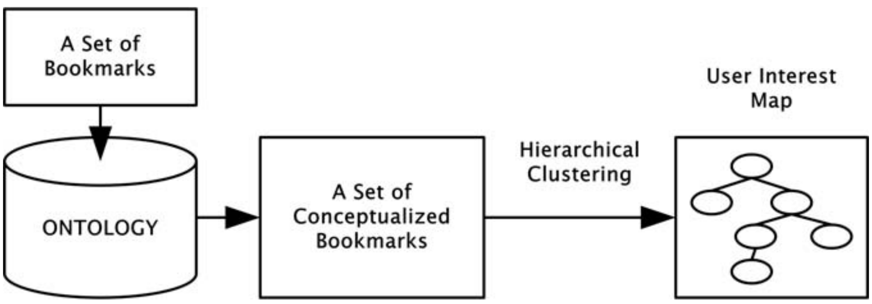

Figure 1: Establishing user interest map based on semantic learning from bookmarks

In so doing, we focused on the establishment of a Web directory organized using a topic-based hierarchy. By aggregating bookmarks labeled by Web directory, a tree-structured interest map can be established for each user. In addition, we employ a simple ontology learning scheme based on a hierarchical clustering method, in order to dynamically adapt the user's interest map, as shown in [<see> Fig. 1].

In order to implement collaborative Web browsing based on this concept, we designed a multi-agent system consisting of a facilitator agent and multiple personal agents. These agents can communicate with each other using ACL (Agent Communication Language), with respect to the interest maps of the users. The personal agent can predict the corresponding user's information needs during browsing, and generate queries for the purpose of obtaining accurate recommendations. The facilitator agent has to be aware of all of the participating personal agents and their interest maps generated from the local bookmarks.

In the following section, we discuss previous works related to collaborative Web browsing. In section 3, we address collaborative searching tasks on the Web. In section 4 and 5, we describe the semantic labeling of bookmarks and the extraction of the user's interests from the labeled bookmarks, respectively. In section 6, we describe the overall architecture for the proposed system and present our experimental results in section 7. Finally, in section 8, we conclude with directions for future work.

2 Related Work

Generally, collaborative browsing systems can be divided into four classes [Rodden, 91]. With respect to its temporal and spatial characteristics, each system can be either synchronous or asynchronous, and either local or remote. In a traditional library, collaborations must be local and synchronous. On the other hand, in a digital library and in our proposed system, however, users can communicate with others remotely and asynchronously. As the representative systems for collaborative browsing, the recently developed *Let's Browse* [Lieberman, 99], *ARIADNE* [Twidale, 96], and *WebWatcher* [Armstrong, 97] have some interesting features. *Let's Browse* uses the infrared sensors for the purpose of detecting the presence of users without any explicit actions, and it makes it possible to instantly exchange information between users. *ARIADNE* records the searching process in a digital library [Twidale, 98], thus allowing this information to be visualized and reused. It is particularly helpful to beginners trying to look for items in unfamiliar topics.

However, the most important difference between these different collaborative Web browsing systems is the method used to extract user preferences from personal information. While *Let's Browse* and *ARIADNE* use the *TF-IDF* scheme to analyze the keyword frequency of Web pages, both *WebWatcher* and our own system focus on incremental learning approaches based on machine learning algorithms. More exactly, our system deals with the extraction of the user's interest through the semantic learning of their activities. The concern about ontology learning has been increasing ever since the semantic Web was introduced. Through the ontology learning of information from heterogeneous sources, the semantic structure can be retrieved and applied to document management and clustering.

As a similar attempt at sharing user bookmarks, the *XBEL* (XML Bookmark Exchange Language) [Drake, 04] has been introduced. This is an interchange format, which is based on the extensible mark-up language (XML), for the hierarchical bookmark data used by current Web browsers.

3 Collaborative Searching on the Web

We can meet a group of people working together researching information on the Web about a certain topic. Group searching takes place when two or more people share a common aim and coordinate their searching efforts [Twidale, 97]. We can decompose collaborative searching tasks in three procedures. First, each participant in this collaboration has to access, process and filter by importance and relevance the information gathered from the Web. Second, they have to synthesize and present them either as a whole in the form of report or in an organized way in the form of hierarchical tree. Third, they can share and recommend certain information between each other, according to their own preferences.

In this computing environment, attaining efficiency for the collaboration requires to address the following two issues:

- The time for finding the appropriate information. This time includes the time needed to access, download and process (read) the Web page in order to decide if it is relevant to the topic being searched. Motivated by the fact that approximately 81% of all individual's URLs had been previously visited by them [Cockburn, 01], it can be hypothesized that a number of the pages to be visited by two or more users will be common.
- The organization of information spaces. About 28% of 3291 survey respondents reported the difficulty of organizing information space in using the Web [GVU, 97]. Individually, Web users can organize personal information space, as collecting relevant information.

In order to deal with these problems, this paper concentrates on user modelling based on extracting the user's interests. After modelling each user's interests is established, personal agents should be able to predict what kind of information the users will look for. We can consider that the users are potentially satisfied with information of other users in a same user group who are interested in a same topic. They can efficiently save the time for finding relevant information. More seriously, personal information spaces organized by personal information like bookmarks are semantically heterogeneous. We need to integrate and manage these spaces. Ontology can be efficiently applied to leverage removing the semantic gap between these spaces in this collaborative task.

4 Information Conceptualization Based on Ontology

In this paper, we assume that the presence of a specific set of bookmarks provides information on the user's intentions reflected during Web browsing. Therefore, we have to extract various features from bookmarks such as the term frequencies, the hyperlinks to other Web pages and the URLs themselves. We employ Web directories as the replacement of ontology for semantic labeling. When labeling the bookmarks of users, two main drawbacks of Web directories will be described, and then, we explain how we deal with these problems in this study. Furthermore, the method of indirect labeling based on link analysis will be proposed for bookmarks whose URLs are not yet registered in the Web directory.

4.1 Web Directory as Topic Hierarchy

Ontology, the so-called semantic categorizer, is an explicit specification of a conceptualization [Gruber, 93]. This means that ontology can be used to enrich unlabeled data with semantic or structural information. We consider Web directory as a topic-specific ontology. Examples of such Web directories are *Yahoo.com* (http://www.yahoo.com/) and *Cora* (http://cora.whizbang.com/). Web directory can be used to describing the content of a Web page document in a standard and universal way as ontology [Labrou, 99]. Besides, these Web directories are organized in the form of a topic-based hierarchical structure, which is an efficient way to organize, view and explore large quantities of information that would, otherwise, be cumbersome [McCallum, 99]. In this paper, we assume that each of the user's bookmarks of users can be labeled by referring to a well-organized Web directory.

4.2 Drawbacks of the Web Directory

There are some practical obstacles to simple URL-based labeling, because most Web directories are forced to manage a non-generic tree structure, in order to avoid wasting memory caused by redundant information [Jung, 01]. We briefly note problems that arise when categorizing URL information using the Web directory as its underlying ontology:

- ***The multiple attributes of a Web page.*** A Web page can be involved in more than one topic. The causal relationships between the different categories make the associated hierarchical structure more complicated. In the example shown in [Fig. 2] (1), the URL information of a certain Web page for one category can be included in another category, where these two categories are referred to 'A' and 'B'.

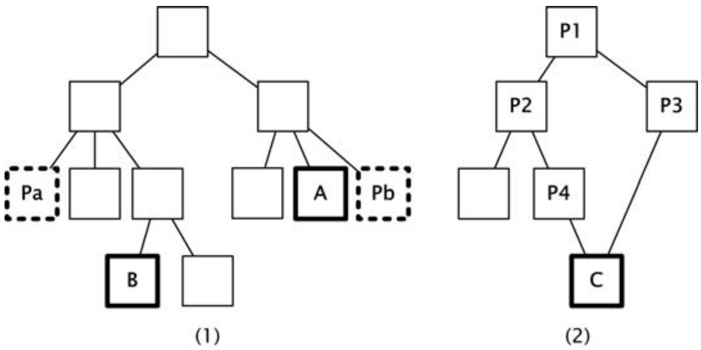

Figure 2: Drawbacks of Web directories (1) The multiple attributes of a Web page, and the semantic relationship between two category - duplication; (2) The semantic relationship between two categories - subordination

- ***The semantic relationship among categories***. There are two kinds of semantic relationships, namely those that are the duplication between identical categories and the subordination between dependent categories.

Some categories can be semantically identical, even if they have different labels. In [Fig. 2] (1), all Web pages in which category ‘Pa’ is including are the same as those in category ‘Pb’. Next, a category can have more than one topical path from the root node. As shown in [Fig. 2] (2), the category ‘C’ can be a subcategory of more than the other categories such as ‘P1: P2: P4’ and ‘P1: P3’.

For example, due to the multiple attributes, a Web site related to the topics “*Artificial Intelligence*” and “*Database*” can be labeled to both of these two categories. Some Web sites registered in the category “*Computer Science: Artificial Intelligence: Constraint Satisfaction: Laboratory*” can also be included in the category “*Education: Universities: Korea: Inha University: Laboratory*”, because these categories are themselves dependent on other categories. Also, in certain cases, all of the Web sites assigned to a particular category can be exactly the same as those found in other categories, because they are semantically identical to each other.

4.3 Two ways of Semantic Labeling

In order to label the bookmarks of users, we extract the URL information from the bookmarks and perform a labeling process that assigns hierarchical topic (or category) paths to the bookmark. There are two kinds of labeling, which are referred to as direct and indirect labeling, depending on whether the Web site in question is registered in the Web directory.

For the Web sites already registered in the Web directory, we can apply direct labeling to them. Direct labeling is a simple querying process which involves looking up the corresponding URLs in the Web directory. In order to deal with the drawbacks of the Web directory, we have to acquire a set of labels which includes all possible paths in order to obtain the desired results.

On the other hand, indirect labeling is used for unregistered Web sites. This method is based on link analysis, and involves searching “authoritative” pages about a certain topic on the hyperlinked information space like Web pages [Ding, 02; Kleinberg, 99]. We propose a modified *HITS* algorithm which allows the most similar data to be obtained from the already labeled dataset. The hyperlinked Web pages are organized into a directed graph $G = (V, E)$, where V is the set of nodes representing the Web sites, and E is the set of hyperlinks between v_i and v_j. In order to search the most authoritative node of a particular Web site, we focus on the outgoing links of that Web site. For the given unlabeled Web page w, the outgoing and incoming links of graph G can be formulated as the asymmetric adjacency matrix, $O(w)_x^{(d)}$, where $[O(w)]_{ij} = 1$ if $v_i \rightarrow v_j$ and $[O(w)]_{ij} = 0$, otherwise. Also, the variable, d, is the number of iterated expansions, which means the distance from node w. This $O(w)$ is a $|V|\times|V|$ square matrix, where V is the set of nodes within the distance d. Therefore, we can reach some labeled nodes, by repeating this iteration along the outgoing links. If there are more than one labeled node at the same distance, we have to evaluate the incoming degree of these nodes by using the following equation $L_{indirect}$

$$L_{indirect}\left(O(w)_x^{(d)}\right) = \max_{j^* \in j}\left[\sum_k [O(w)]_{kj^*}\right],$$

where the j^*-th Web sites are labeled. This means that the Web sites can be regarded as more authoritative ones, since they are referred to by a larger number of other Web sites. In the example shown in [Fig. 3], the Web site, *m*, which is requested by the clients, is not yet registered in the Web directory. The solid arrow lines are outgoing links to other Web sites, while the dotted lines are incoming links from other Web sites. The Web site, *x*, belongs to the nearest neighbor category that is registered in the Web directory.

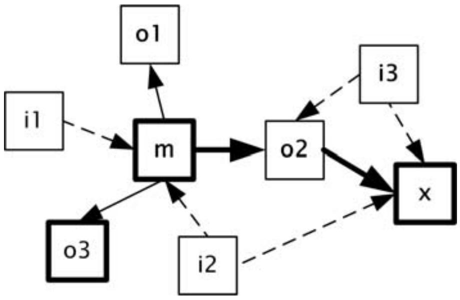

Figure 3: Indirect labeling of unregistered Web site, m

The link matrix of a graph in [Fig. 3] is given by

$$O(m)_x^{(2)} = \begin{array}{c} \\ m \\ o1 \\ o2 \\ \ldots \\ x \end{array} \begin{array}{c} \begin{array}{ccccc} m & o1 & o2 & .. & x \end{array} \\ \begin{bmatrix} 0 & 1 & 1 & \ldots & o \\ 0 & 0 & 0 & \ldots & 0 \\ 0 & 0 & 0 & \ldots & 1 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & 0 \end{bmatrix} \end{array},$$

where the distance threshold *d* is predefined as two. Let the Web pages ‘o3’and ‘x’ be registered in the Web directory. By using $L_{indirect}$, the maximum authoritative Web page ‘x’ can be obtained.

Next, we define the notations used for semantic labeling. Let the user, U_i, have the set of bookmarks, B_i, as follows:

$$B_i = \{b_1^i, b_2^i, \ldots, b_t^i\},$$

where *t* is the total number of bookmarks. Each bookmark in this set is labeled with the corresponding categories represented by the directory paths. Therefore, the set of conceptualized bookmarks, C_i, is given by

$$C_i = CB_i + CRB_i,$$

where

$$CB_i = \{cb_1^i, cb_2^i, ..., cb_n^i\} \text{ and } CB_i = \{crb_1^i, crb_2^i, ..., crb_\alpha^i\}.$$

The variable n is the total number of concepts, including the bookmarks in B_i. Also, α is the number of additional concepts subordinately related to CB_i. This is caused by the drawbacks of Web directories which are mentioned in section 4.2. Generally, the variable, n, becomes larger than t. Here, we mention the step used for conceptualizing the bookmarks by referring to the Web directories as follows:

```
Function Semantic_Labeling ( User )
var
  counter1, counter2: integer;    B: set_bookmark[];
  CB, CRB: set_conceptualized_bookmark[];
begin
  B := Bookmark ( User );
  counter1 := 1;
  repeat
    CB := CB + Concept ( B[count1] );
    repeat
      counter2 := 1;
      if ( ( isLinked( Concept( B[counter1] ) ) ) = TRUE ) then
        CRB := CRB + Linked( Concept( B[counter1] ) );
    until counter2 = size( B[counter1] )
    counter1 := counter1 + 1;
  until counter1 = size( B );
  return ( CB, CRB );
end.
```

The functions *Bookmark* and *Concept* return the set of bookmarks of an input user and the set of concepts matched with an input bookmark by looking up the ontology, respectively. The function *Linked* retrieves the additional concepts related to the input concept, once the function *isLinked* has checked if the input parameter is connected to more than one parent concept on the ontology. As a result, the size of each user's category set becomes larger than that of his bookmark set, because of the incomplete properties of the category structure mentioned in the previous section. Therefore, we supplemented the user's category set with a candidate category set. The candidate category set improves the coverage of the user's preferences. This means that potential preferences can be detected as well.

5 Semantic Extraction of User Interests from Bookmarks

In order to extract the user's interests, the semantically labeled bookmarks are aggregated on the interest map (*i*-Map). We assume that there exists influence propagation between the different topics on the *i*-Map of each user, and the *Bayesian* probability theorem is exploited to deal with these propagation problems. Every category of the *i*-Map has to be assigned a *DI* (Degree of Interest) value.

5.1 Semantic Learning from Bookmarks

Ontology learning has four main phases, namely importing, extracting, pruning, and refining [Maedche, 02]. We focus on extracting the semantic information from bookmarks based on hierarchical clustering, which is the process of organizing tree structures of objects into groups whose members are similar in certain ways [Kaufman, 90]. The tree of hierarchical clusters can be produced either bottom-up, by starting with individual objects and grouping the most similar ones or top-down, whereby one starts with all the objects and divides them into groups [Maedche, 02].

When clustering conceptualized bookmarks, the top-down algorithm is more suitable than the bottom-up approach, because the directory path information is already assigned to the bookmarks during the conceptualization step.

5.2 Bayesian Estimation of User Interests Based on Influence Propagation

Basically, *Bayesian* networks are probabilistic models that allow the structured representation of a cognitive or decision process and are commonly used for decision tree analysis in business and the social sciences [Pearl, 88; Giarratano, 94]. According to [Baeza-Yates, 99], the strength of causal influences between categories is simply expressed by this conditional probability

$$P(parent, Children) = \sum_i \left[P(parent \mid child_i) \times P(child_i)\right].$$

This probability refers to the issue of how categories reflect their causal relationship on parent nodes. The degree of user preference for the parent node is the summation of the evidential supports of the child nodes linked to the parent node. We assume that each category is assigned the corresponding *DI* value, according to the following axioms:

1. The initial *DI* of a concept is the number of times that this concept is matched with the set of bookmarks through the function *Semantic_Labeling*. The larger the *DI* of a concept is, the more interested the corresponding user is in this concept. In other words, the number of times that a concept is matched with the set of bookmarks is linearly proportional to the user preference for this concept.

$$\textit{Number of times the concept is matched} \propto DI(C_i)$$

2. The *DI* of a concept is propagated from its subconcepts using this influence propagation equation

$$propagate[DI(C_i)] = \frac{\log_k (DI(C_i)+1)}{N},$$

where N is the total number of siblings of a concept C_i and k is given by

$$k = \text{variance}(DI(subc(C_i)) + bias = \sigma^2 + bias,$$

where $subc(C_i)$ is the set of subconcepts of C_i, and *bias* is used for the exceptional case such as the variance σ^2 is zero. We note two important characteristics of influence propagation between concepts.

- *The dispersion of DI.* As the number of concepts of a parent is increased, each of them has less influence on its parent concepts.
- *The distance between concepts.* The closer the concepts are, the more tightly related they are to each other. In other words, the influence propagation increases exponentially, as the distance between the concepts decreases.

3. The *DI* of a concept is measured from the propagations of all subconcepts, and all concepts have influence on the root node.

$$DI(C_i) = \sum_j \left(propagate(DI(subc(C_i)_j)) \times DI(subc(C_i)_j) \right)$$

4. Those concepts whose *DI*'s are over a predefined threshold value after normalization step are taken to represent the user's interests.
5. The user's interests can change. Therefore, we have to consider newly incoming bookmarks. This means that every time he or she inserts a bookmark, the *i*-Map of the user should be updated.

5.3 Tree Representation of User Interests and Example

In [Fig. 4], we show an example of the process of mining a user's interests from his or her bookmarks.

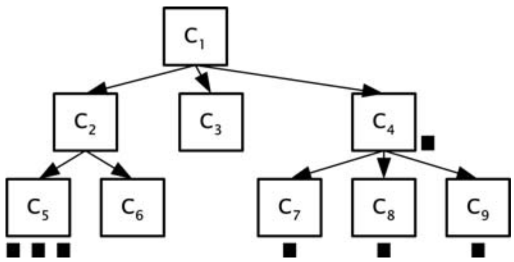

Figure 4: Example of the conceptualized bookmarks of a user

The black squares indicate the bookmarks of user U_i, for which the initial states are assigned in the following equations:

$$DI(C_4) = 1,\ DI(C_5) = 3,\ DI(C_6) = 0,$$
$$DI(C_7) = 1,\ DI(C_8) = 1,\ DI(C_9) = 1$$

According to the influence propagation equations, all of the *DI*'s of the other concepts can be computed. The *DI*'s of C_2 and C_4 are as follows.

$$DI(C_2) = \sum_{k-1}^{2} \left(propagate[DI(C_k)] \times DI(C_k) \right) = 1.11$$

$$DI(C_4) = 1 + \left(\log_2 2/3\right) \times 1 \times 3 = 2.0$$

The mean of all *DI*'s is 1.44 and the *DI* of each concept is assigned after normalization. If the threshold value is 0.2, only C_4 and C_5 are extracted as the concepts the corresponding user is most interested in.

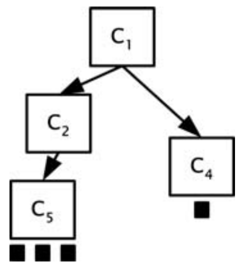

Figure 5: Tree structured representation of i-Map for the high ranked concepts

In [Fig. 5], the *i*-Map of a particular user is represented in the form of a tree. Each node refers to the high ranked categories, which are considered to be those topics that the user is most interested in.

6 Collaborative Web Browsing with Recommendation

The collaborative Web browsing system proposed in this paper is remote and asynchronous, because it is based on the Web environment and the information available about a participant's interests, which are extracted from his or her own bookmarks and ontology. More importantly, all communications between agents are conducted without requiring any user intervention. Also, while browsing to search for information on a particular topic, "implicit" recommendations can be made to the user by the facilitator in the following two ways:

- *By querying specific information for the facilitator.* After the information about a particular concept has been requested, the facilitator can determine who has the maximum *DI*'s for this particular concept by scanning its yellow pages.
- *By the facilitator's broadcasting the new bookmarks of like-minded users.* Every time a user inserts a new bookmark, after conceptualization this fact is sent to the facilitator. In this way, users can obtain information which is related to common concepts from their neighbors, and store it in their own *i*-Map.

As shown in [Fig. 6], the overall system architecture consists of two main parts, namely the facilitator, which is located between the users, and the client-side Web browser that communicates with the facilitator.

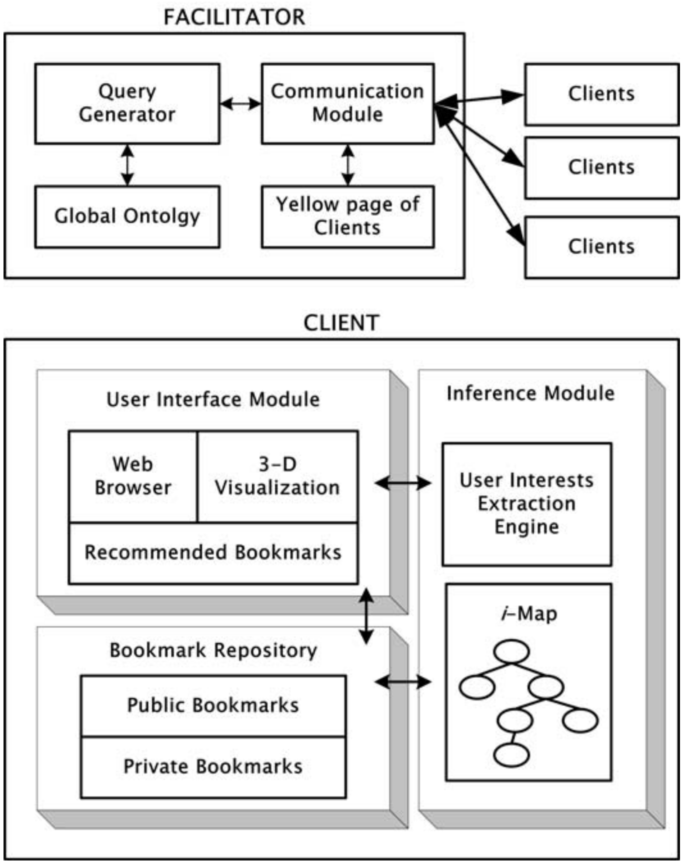

Figure 6: System architecture

Each client needs a personal agent consisting of user interface module, inference module and bookmark repository. This agent initializes and manages the corresponding user's ***i***-Map based on his or her bookmark repository. Therefore, it has to be able to communicate with the facilitator agent, and refer to the global ontology e.g., the Web directory.

Through the personal agents' reporting the bookmarking activities of their clients, the facilitators can automatically generate queries and recommendations. Most importantly, the facilitator agent has to create the yellow pages for information about all participants. Then, each bookmarking activity can be automatically transmitted to the facilitator.

7 Experiments and Implementation

We constructed a hierarchical tree structure for use as a test bed using the information contained in the section "*Home: Science: Computer Science*" at *www.yahoo.com*.

This tree consists of about 1300 categories and the maximum depth was eight. In order to gather bookmarks for this information, 30 users explored the directory pages of *www.yahoo.com* for 28 days. Whenever the users visited a Web site related to their own interests, they stored the URL information in their bookmark repositories. Finally, 2718 bookmarks were collected. In order to evaluate this collaborative Web browsing process, based on the extraction of the user's interests, we adopted the measurements *recall* and *precision*.

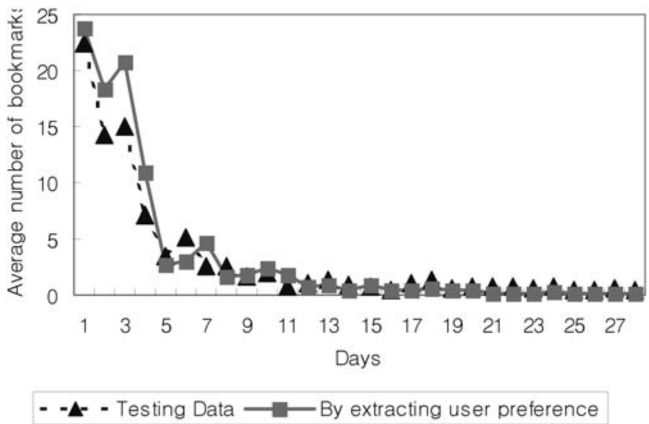

Figure 7: Experimental result in terms of recall with recommendation

After all of the bookmark sets of the users were reset, the users began to gather bookmarks again while receiving the system's recommendations according to their own preferences. During this time, the users were being recommended relevant information retrieved from the test bed based on their interests, as extracted up to that moment. In case of browsing with recommendations, altogether 80% of the bookmarks were collected in only 3.8 days, representing a saving of about 53.1% of the time spent in the case of normal single browsing, as shown in [Fig. 7].

The *precision* was measured by evaluating the ratio of the inserted bookmarks among the recommended information set. In other words, this was a measurement of the accuracy of predictability. As the number of recorded bookmarks increased, the user's preferences gradually converged so as to become more stable. Figure [Fig. 8] shows the experimental result concerning to the *precision* of the recommendation based on the user's preferences. In the beginning, the *precision* was low, because the user's preferences had not yet been determined. While the user's interests were being extracted during the first 6 days, the *precision* of the recommended information was tracked and compared with that of the testing dataset.

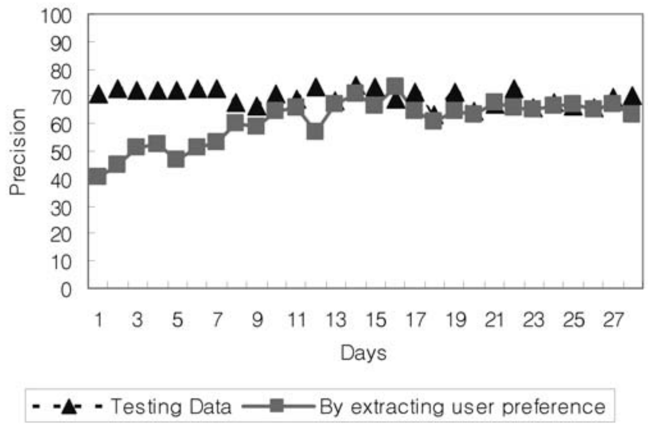

Figure 8: Experimental result in terms of precision with recommendation

During the remaining part of the experiment, the *precision* stayed at the same level as that of the testing dataset.

8 Conclusions and Future Work

In this paper, we assume that bookmarks are the most important indication of the user's interests. However, due to the lack of semantic information that can be obtained from simple URL based bookmarks, we focused on developing a way of conceptualizing them by referring to Web directories. Once the semantic and structural information for the users' bookmarks has been provided, not only the precision but also the reliability of the extraction of the user's preferences was improved. Then, by establishing *i*-Maps of the corresponding users and *DI*'s of the concepts contained in these maps, we made it much easier to generate queries for relevant information and to share bookmarks among like-minded users. In this way, we implemented a collaborative Web browsing system sharing conceptualized bookmarks.

Based on the information recommendation provided by this system, a saving of about 53% was made in the search time, as compared with normal single Web browsing. Moreover, this method can enable a beginner in a certain field to be helped by obtaining valuable information from experts in this particular domain.

In a future work, we will consider the privacy problems associated with sharing personal information, such as age, gender and preferences. However, the visualization of the *i*-Map is the next target of this research, in order to enable the user to recognize his or her own preferences quantitatively with regard to each topic. Additionally, we also have to concentrate on the representation of semantic labeling using XML.

References

[Armstrong, 97] R. Armstrong, et al, *WebWatcher*: a learning apprentice for the world wide web, In Proc. AAAI Spring Sym. on Information Gathering from Heterogeneous, Distributed Environments, 1997, 6-12.

[Baeza-Yates, 99] R. Baeza-Yates, B. Ribeiro-Neto, Modern information retrieval, ACM Press, New York, 1999, 48-60.

[Cockburn, 01] A, Cockburn, B. McKenzie, What do web users do? An empirical analysis of web use, International Journal of Human-Computer Studies, 54, 2001, 903-922.

[Ding, 02] C. Ding, et al, PageRank, HITS and a unified framework for link analysis, Technical Report, No. 49372, Lawrence Berkeley National Laboratory, 2002.

[Drake, 04] F. Drake, The XML Bookmark Exchange Language (XBEL), http://pyxml.sourceforge.net/topics/xbel/, 2004.

[Giarratano, 94] J. Giarratano, G. Riley, Expert systems principles and programming, PWS publishing company, 1994, 203-205.

[Gruber, 93] T.R. Gruber, A translation approach to portable ontologies, Knowledge acquisition, 5(2), 1993, 199-220.

[GVU, 97] GVU's 8th WWW User Survey, http://gvu.gatech.edu/user_surveys/survey-1997-10/, 1997.

[Jung, 00] J.J. Jung, et al, *BISAgent*: collaborative web browsing through sharing of bookmark information, In Proc. IIP2000, IFIP, 2000.

[Jung, 01] J.J. Jung, et al, Collaborative information filtering by using categorized bookmarks on the web, In Proc. of the 14th Int. Conf. on Applications of Prolog, 2001, 343-357.

[Jung, 03] J.J. Jung, G.-S. Jo, Extracting user interests from bookmarks on the web, In Proc. Int. Pacific-Asia Conf. on Knowledge Discovery and Data mining, 2003, 203-208.

[Kaufman, 90] L. Kaufman, P. Rousseeuw, Finding groups in data: An introduction to cluster analysis, John Wiley, 1990, 37-49.

[Kleinberg, 99] J.M. Kleinberg, Authoritative sources in a hyperlinked environment, Journal of the ACM, 46(5), 1999, 604-632.

[Labrou, 99] Y. Labrou, T. Finin, Yahoo! as an ontology: using Yahoo! categories to describe documents, In Proc. of 8th Int. Conf. on Information Knowledge Management, 1999, 180-187.

[Lieberman, 95] H. Lieberman, Letizia: an agent that assists web browsing agent, In Proc. Int. Joint Conf. on Artificial Intelligence, 1995, 924-929.

[Lieberman, 99] H. Lieberman, et al, Let's Browse: a collaborative web browsing agent, In Proc. Int. Conf. on IUI, 1999, 65-68.

[Maedche, 02] A. Maedche, Ontology learning for the semantic web, Kluwer Academic Publishers, 2002, 59-61.

[Maes, 94] P. Maes, Agents that reduce work and information overload, Communication of ACM, 37(7), 1994, 31-40.

[McCallum, 99] A. McCallum, et al, Building domain-specific search engines with machine learning techniques, In Proc. of AAAI-99 Spring Sym. on Intelligent Agents in Cyberspace, 1999.

[Pearl, 88] J. Pearl, Probabilistic reasoning in intelligent systems, Morgan Kauffman Publisher, 1988.

[Rodden, 91] T. Rodden, A survey of CSCW systems, Interacting with Computers, 3(3), 1991, 319-354.

[Twidale, 96] M. Twidale, D. Nichols, Collaborative browsing and visualization of the search process, Electronic library and visual information research, 1996, 51-60.

[Twidale, 97] M. Twidale, D. Nichols, C. Paice, Browsing is a collaborative process, Information Processing and Management, 33(6), 1997, 761-783.

[Twidale, 98] M. Twidale, D. Nichols, Computer supported cooperative work in information search and retrieval, Annual Review of Information Science and Technology, 33, 1998, 259-319.