Computing and Informatics, Vol. 22, 2003, 337–353

AN APPLICATION OF COLLABORATIVE WEB BROWSING BASED ON ONTOLOGY LEARNING FROM USER ACTIVITIES ON THE WEB

Jason J. Jung

School of Computer and Information Engineering
Inha University
253 Yonghyun-Dong, Nam-Gu
402-751 Incheon, South Korea
e-mail: `j2jung@intelligent.pe.kr`

Manuscript received 21 September 2004; revised 25 October 2004
Communicated by Pavol Návrat

Abstract. With explosively increasing amount of information on the Web, users have been getting more bored to seek relevant information. Several studies have introduced adaptive approaches to recognizing personal interests. This paper proposes the collaborative Web browsing system that can support users to share knowledge with other users. Especially, we have focused on user interests extracted from their own activities related to bookmarks. A simple URL based bookmark is provided with semantic and structural information by 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. As a result of our experiments, about 53.1 % of the total time was saved during collaborative browsing for seeking the equivalent set of information, compared with single Web browsing. Finally, we demonstrate implementing an application of collaborative browsing system through sharing bookmark-associated activities.

Keywords: Collaborative web browsing, ontology learning, user modeling

1 INTRODUCTION

Recently the amount of information has been exponentially increasing on the Web. Navigation for searching relevant information in this Web environment is one of the most lonely and time-consuming tasks [20]. There have been many kinds of studies to deal with this problem, the so-called "*information overloading*". Most of them have been involved in user profiling through analyzing the recorded behaviors of each user. For example, the personal assistant agent systems can predict the reactions of the corresponding users like removing junk e-mails from mailbox, or while browsing, proactively prefetch and show the relevant Web pages based on user preferences [17].

Contrary to these single user-centred approaches, we assume that collaboration among many users can be another way to improve the performance of information retrieval. Since computer supported cooperative work (CSCW) was introduced by Cashman and Greif [9], many kinds of domains have been concerning collaboration interactions. In this paper, we introduce collaborative Web browsing, which is an approach whereby users share knowledge with the other like-minded neighbors while searching for relevant information on the Web space. Through communicating with the others, users can acquire many kinds of experiences (or heuristics) such as how to select and rank the searching results, how to make the sequence of queries, and how to choose searching methods, and they provide the other users with their own knowledge as well. Generally, according to [23], collaborative browsing systems can be discriminated into four classes. With respect to temporal and spatial characteristics, each system can be either synchronous or asynchronous, and either local or remote. In traditional library, collaborations must be local and synchronous. On the other hand, in digital library and our proposed system, users can communicate with others remotely and asynchronously.

This paper proposes the extended application of a BISAgent [12], which is a bookmark sharing agent system based on a modified *TF-IDF* scheme without considering user preferences [12]. This bookmark is playing a role of a pointer, primarily to URL information, built-in to the various Internet Web browsers such as Mosaic Web browser, Netscape, and Internet Explorer (more exactly, called Favourites within MS-Windows platform). Typically, a bookmark is always stored on the software clients. For example, bookmarking the Web site "Museum of Modern Art" makes a local file containing the URL information generated in client. The bookmark file is shown in Table 1.

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

Table 1. An example of bookmark file of "Museum of Modern Art"

More importantly, we have focused on pieces of information related to user interests. We assume that recognizing which a user is interested in is a very important task in collaborative Web browsing. Querying relevant information to the other users, filtering the query results, and recommending them are major tasks that have to be implicitly conducted. According to the GVU's survey [8], nowadays there is no doubt that the number of bookmarks has been increasing more than ever. This means that a set of bookmarks in the user's folder can be considered as a piece of information to infer user interests [11]. In order to uncover user interests from their own bookmark sets, we need the ontology based semantic analysis of each Web site pointed by these bookmarks. Thereby, Web directory organized as topic hierarchy is applied to semantic labeling of bookmarks. By using labeled bookmarks, we can establish tree-structured interest map for each user. In addition, we employ simple ontology learning based on hierarchical clustering method for dynamic adaptation of user interest map, as shown in Figure 1. We need to consider which activities are taken to a certain bookmark. For example, if a user has been regularly visiting a Web site, the corresponding bookmark must be more associated with his interests rather than other bookmarks which are just stored.

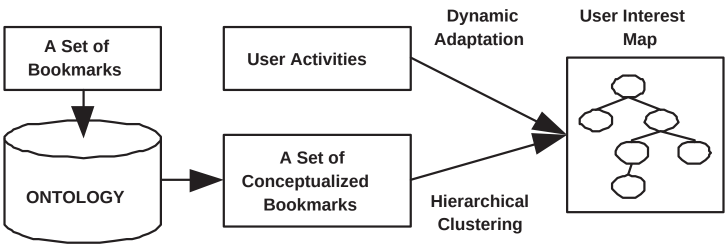

Fig. 1. Establishing user interest map based on ontology learning from bookmarks

In order to achieve the goal of this idea, according to the user interest maps, the facilitating agent and personal agents can communicate with each other. A personal agent consists of bookmark repository, inference module for extracting user interests, and user interface module. Personal agents can predict the corresponding user's needs during browsing, and automatically generate queries for accurate recommendations. All messages between personal agents are under the control of facilitating agent. Additionally, facilitating agent can broadcast bookmarks related to a certain topic.

The following section explains the previous work related to collaborative Web browsing. Sections 3 and 4 describe semantic labeling of bookmarks and extraction of user interests from labeled bookmarks, respectively. We will address the whole system architecture of our idea in Section 5 and we present the experimental result in Section 6. Finally, in Section 7, we conclude with directions for future work.

2 RELATED WORK

As the representative systems for collaborative browsing, *Let's Browse* [18], *ARIADNE* [25], and *WebWatcher* [1] have been developed. They have shown some interesting features. *Let's Browse* has infrared sensors for 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 digital library [26]. Thus, this information can be visualized and reused. It is particularly helpful to beginners trying to look for items.

However, the most important difference from them is how to extract user preferences from personal information. While *Let's Browse* and *ARIADNE* applied *TF-IDF* scheme to analyze keyword frequency of Web pages, *WebWatcher* and our system have been focusing on incremental learning approaches based on machine learning algorithms. Our system, more exactly, deals with extraction of user interest from ontology learning of their activities. The concerns about ontology learning has been increasing, since semantic Web was introduced. Through ontology learning of information from heterogeneous sources, semantic structure can be retrieved and applied to document management [6] and clustering [24].

Additionally, as a similar attempt of sharing user bookmarks, the XBEL (XML Bookmark Exchange Language) [5] was introduced. This is an interchange format, which is based on the extensible markup language (XML), for the hierarchical bookmark data used by current Web browsers.

3 SEMANTIC LABELING BASED ON ONTOLOGY

This paper assumes that a set of bookmarks implies the corresponding user's intentions reflected during Web browsing. Thereby, we have to extract features from bookmarks such as term frequencies, hyperlinks on web pages, and URLs. We employ web directories as the replacement of ontology for semantic labeling. When labeling bookmarks of users, some drawbacks of Web directories will be described, and then we introduce how to deal with these problems in this paper. Furthermore, indirect labeling based on link analysis will be proposed for bookmarks of which URLs are not registered in the Web directory yet.

3.1 Web Directory as Topic Hierarchy

Ontology, the so-called semantic categorizer, is an explicit specification of a conceptualization [10]. It means that ontology can play a role of enriching semantic or structural information to unlabeled data. We have regarded Web directory as topic-specified ontology. Such are `Yahoo.com` (`http://www.yahoo.com/`) and *Cora* (`http://cora.whizbang.com/`). These Web directories can be used to describe the content of a document in a standard and universal way as ontology [16]. Besides, these Web directories are organized as topic hierarchical structure that is an efficient way to organize, view, and explore large quantities of information that would, otherwise, be cumbersome [21]. In this paper, we assume that each bookmark of users can be labeled by referring on a Web directory.

3.2 Drawbacks of Web Directory

There are some practical obstacles to simple URL-based labeling, because most of Web directories are forced to manage a non-generic tree structure in order to avoid a waste of memory space caused by redundant information [13]. We briefly note that problems with categorizing the URL information with Web directory as ontology are the following:

The multi-attributes of a Web site. A Web site can be involved in more than a topic. The causal relationships between categories make their hierarchical structure more complicated. As shown in Figure 2 (1), the URL information can be included in some other categories, named as 'A' and 'B'.

The semantic relationships between categories. There are two kinds of semantic relationships, which are the subordination between dependent categories and the redundancy between identical categories. A category can have more than a topical path from root node. As shown in Figure 2 (2), the category 'C' can be a subcategory of two possible categories 'P'. Furthermore, some categories can be semantically identical, even if they have different labels.

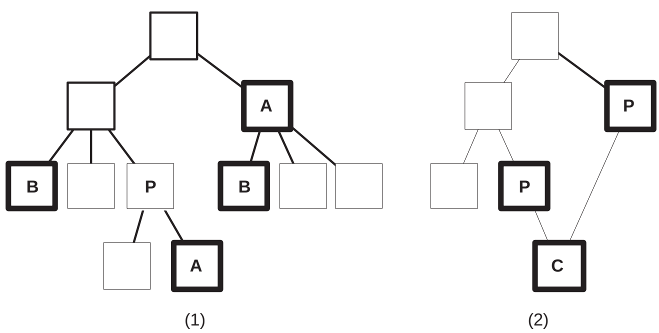

Fig. 2. (1) The multi-attributes of a Web site; (2) The semantic relationship between two categories – subordination

For example, due to the multi-attributes, a Web site related to "Artificial Intelligence" and "Database" can be labeled to these two categories. Some Web sites registered in the category "Computer Science: Artificial Intelligence: Constraint Satisfaction: Laboratory" can also be in the category "Education: Universities: Korea: Inha University: Laboratory", because these categories are semantically associated with each other. Also, all Web sites which are labeled as a particular category can be exactly the same as those labeled as the other category, because they are semantically identical with each other.

3.3 Two Ways of Semantic Labeling

In order to conduct the semantic labeling of each user's bookmarks, we extract the URL information from bookmarks and assign hierarchical topic (or categorical) paths to them by referring to Web directory. There are two kinds of labeling, which are direct and indirect labeling. It depends on whether this Web site is registered on Web directory.

Direct labeling is a simple querying process looking up the corresponding URLs from Web directory. In order to deal with the drawbacks of Web directory, we have to find out a set of labels including all possible paths as the results. On the other hand, indirect labeling is needed for unregistered Web sites. Originally, *HITS* algorithm is a task selecting a subgraph from the Web by a certain query [15, 3]. From this subgraph, two kinds of nodes are identified: *authoritative* pages to which many pages link, and *hub* pages that consist of comprehensive collections of links to valuable pages on a specific topic. We propose the modified *HITS* (Hyperlink-Induced Topic Search) algorithm searching the most similar data from already labeled dataset. It is based on link analysis for searching authoritative pages about a certain topic on the hyperlinked space like Web. As shown in Figure 3, the Web site M requested by clients is not registered yet on Web directory.

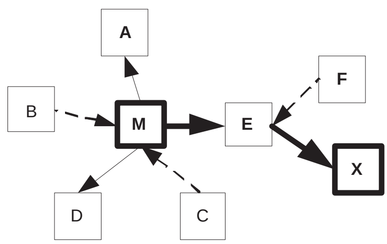

Fig. 3. Indirect labeling of unregistered Web site, M

The solid arrow lines are outgoing hyperlinks to the other Web sites, while the dashed ones are incoming hyperlinks from the others. The Web site X is the nearest neighbor category that is registered on Web directory. The hyperlinked Web pages organize a directed graph $G = (V, E)$, where V and E is the set of nodes representing Web sites and the set of hyperlinks between v_i and v_j, respectively. Practically, the set of nodes V is obtained through parsing hyperlinks in HTML documents. In order to search the most authoritative node of a particular Web site, we focus on incoming and outgoing hyperlinks of each Web site. When the unlabeled Web site M is given, we can formulate incoming and outgoing hyperlinks of graph G from M as the asymmetric adjacency matrix $O(M)^{(d)}$, where d is the number of iteratively expanded radius, which means the hyperlinked distance from M. We simply predefine d as three. If a Web site p_j is linked from p_i ($p_j \leftarrow p_i$), the matrix element $[O(M)]_{ij}$ is assigned one; otherwise, $[O(M)]_{ij} = 0$. This $O(M)^{(d)}$ is an $h \times h$ square matrix where h is the number of Webpages linked from the M.

Therefore, we can reach some labeled nodes, as repeating this matrix expansion along outgoing links, within the predefined maximum link radius d. If there are more than one labeled nodes at the same distance, we have to evaluate the incoming (authoritative) degree of those nodes by using the following equation

$$O(M)^{(d)}_{MX} = \max_{j^* \in L} \left[\sum_k O(M)^{(d)}_{kj^*} \right] \tag{1}$$

where j^*-th Web sites are labeled. It means that the Web sites referred by the other sites can be regarded as the more authoritative one.

We define the notations for two ways of semantic labeling. Let the user U_i have the set of bookmarks B_i as follows:

$$B_i = \{b^i_1, b^i_2, \ldots, b^i_m\} \tag{2}$$

where m is the total number of bookmarks. Each bookmark in this set should be labeled with all possible categories represented as the hierarchical paths by scanning Web directory database. Therefore, the set of conceptualized bookmarks C_i is the summation of the following two sets CB_i and CRB_i

$$CB_i = \{cb^i_1, cb^i_2, \ldots, cb^i_n\} \tag{3}$$

$$CRB_i = \{crb^i_1, crb^i_2, \ldots, crb^i_\alpha\} \tag{4}$$

where n is the total number of concepts including the bookmarks in B_i, and α is the size of CRB_i, which is the set of concepts subordinately related to CB_i. Because of the drawbacks of Web directory database explained in Section 3.2, we have to retrieve CRB_i, including additional concepts.

Generally, due to the drawbacks of Web directories, the variable n becomes larger than m. Here, we mention the step for conceptualizing the bookmarks by referring to Web directories as follows:

Function *Semantic Labeling* (User U_i)
 var
   $counter_1$, $counter_2$: integer;
   b: Set Of Bookmark[ ];
   cb, crb: Set Of Conceptualized Bookmark[ ];
 begin
   b := *Bookmark* (U_i);
   $counter_1$:= 1;

   repeat
     cb := cb + *Concept* ($b[count_1]$);
     repeat
       $counter_2$:= 1;

if $((isLinked(Concept(b[counter_1]))) = \text{TRUE})$ then
$crb := crb + Linked(Concept(b[counter_1]))$;
until $counter_2 = size(b[counter_1])$
$counter_1 := counter_1 + 1$;
until $counter_1 = size(b)$;
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. Then the function *isLinked* checks if the input parameter is connected from 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 have supplemented with a candidate category set. The candidate category set improves the coverage of user preferences. This means that potential preferences can be detected as well.

4 EXTRACTING USER INTERESTS FROM BOOKMARKS

In order to extract user interests, semantically labeled bookmarks are aggregated on the interest map (i-Map), and user actions involved in each of them are monitored. We assume that there exists influence propagation between topics on i-Map of each user, and Bayesian probability theorem is exploited to deal with these propagation problems. Every category of the i-Map has to be assigned the degree of interest (DOI) value.

4.1 Ontology Learning from Bookmarks

Ontology learning plays a role of integrating and maintaining many different types of data, more importantly, extracting the semantic structure of them. Such data types are lexical ontology, domain-specific ontology, semi-structured data and even free text documents. Ideally ontology learning has four main phases that are import, extract, prune, and refine [19].

From semi-structured data like bookmarks, we want to uncover the underlying semantic structure of user preferences. For the first step, we import a set of bookmarks labeled by Web directory with some shortcomings. Then, we are focusing on extracting semantic information represented as topical path from Web directory and organizing information space based on hierarchical clustering. Hierarchical clustering is the process of organizing tree structures of objects into groups whose members are similar in some ways [14]. 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 [19]. When clustering conceptualized bookmarks, the top-down algorithm is more suitable than the bottom-up approach, because directory path information is already assigned to the bookmarks during conceptualization step. Finally, for pruning and refining, user activities should be gathered. It means temporal changes of user interests have to be detected.

4.2 Bayesian Estimation with 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 [22], [7]. According to [2], the strength of causal influences between categories is simply expressed by this conditional probability

$$P(parent, children) = \sum_i \left[P(parent|child_i) \times P(child_i)\right]. \quad (5)$$

This probability means how categories reflect their causal relationship on parent nodes. In this paper, we simply apply *Bayesian networks* to estimate the user interestness about a certain topic. 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 only exploit the influence propagation function of *Bayesian* estimation process. After the structure is given by Web directory, this function can be used to dynamic adaptation of user activities. We note that bookmark activities are simply divided into four kinds of behavior classes, as follows:

Saving a bookmark. When a user is interested in a Web site, s/he will put the URL information of this Web site in the bookmark repository. This activity can be performed as an intent to visit the Web site again.

Reusing a bookmark. When a user is looking for information, s/he will generally check his/her bookmarks rather than use a web search engine. Thereby, the periodic usage patterns of certain bookmarks can be recognized. For example, because on-line magazines are published monthly or weekly, it can be noted that users interested in these magazines will periodically access the Web sites providing these magazines.

Deleting a bookmark. If a user mistakenly stores a site to bookmark repository, s/he will delete the bookmark. Additionally, the changes of users' interests can result in this activity.

Remembering a bookmark. Sometimes users place a bookmark in their own bookmark sets more than once. They show patterns such as rearranging bookmarks, renaming bookmarks, and making new directories.

Each behavior implies how much users are interested or disinterested in a particular category. According to these bookmarking activities, we have to update user preferences. We assume that every behavior should be manually assigned the causal rate (CR) which is the numeric value [-1, 1] by users, compared with [11] using *Hebbian* learning algorithm. Especially, the CR is used for elaborating the propagation of the causal influence among categories when updating user preferences adaptively. This demonstrates the high coverage of the user's potential preferences. Thus, in this paper, the CR's of two activities 'Saving' and 'Deleting' are trivially determined to 1 and -1, respectively. Activity 'Remembering' is not considered in this paper.

We propose that each category is assigned the corresponding DOI value, according to the following axioms:

1. The initial DOI 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 DOI of a concept means that the corresponding user is more interested in this concept.

$$\text{Number of matched times of concepts} \propto DOI(C_i)$$

   This means that this number of times is in linear proportion to user preference for the corresponding concepts.

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

$$Propagate[DOI(C_i)] = (\log_k(DOI(C_i) + 1))/N \tag{6}$$

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

$$k = Variance(DOI(subc(C_i))) + bias = \sigma^2 + bias \tag{7}$$

   where $subc(C_i)$ is the set of subconcepts of C_i, and $bias$ is for the exceptional cases like the variance σ^2 is zero. We usually predefine $bias = 2$.

   The dispersion of DOI. As the number of subconcepts of a parent is increased, each of them has less influence on its parent concepts.

   The distance between concepts. The closer concepts are more tightly related to each other. In other words, the influence propagation is exponentially increasing, as the distance between concepts becomes closer.

3. The DOI of a concept can be measured from the propagation of all subconcepts, and all concepts have influence on the root node.

$$DOI(C_i) = \sum_j [Propagate(DOI(subc(C_i)_j)) \times DOI(subc(C_i)_j)] \tag{8}$$

4. Concepts of which DOI's are larger than the predefined threshold value finally represent user interests, after normalization step.

5. User interests can be changed by user activities. From Equation (6), influence caused by user actions taken to a bookmark c_i can be modified as

$$Propagate[Action(C_i)] = Propagate[CR_{Action} \times DOI(C_i)] \tag{9}$$

where CR_{Action} is the causal rate of the corresponding action.

These heuristic axioms can deal with not only the initialization of i-Map but also with dynamic adaptation from user activities.

4.3 Tree Representation of User Interests and Example

We show a simple example for mining user interests from bookmarks, as shown in Figure 4.

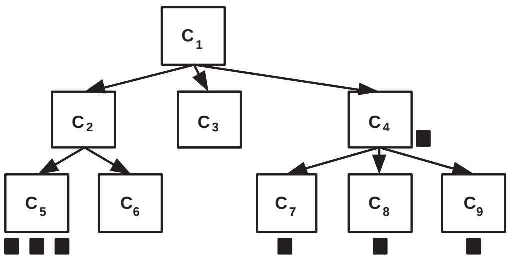

Fig. 4. Example of the conceptualized bookmarks of a user

The black squares indicate the bookmarks of a user U_i, and all categories are assigned the initial DOI's like $DOI(C_4) = 1$, $DOI(C_5) = 3$, $DOI(C_9) = 1$ and so on. The influence propagation equations can be applied to computing the semantic relationships between categories. Then, the DOI's of C_2 and C_4 are as follows.

$$DOI(C_2) = \sum [propagate[subc(DOI(C_2))] \times DOI(subc(C_2))] \tag{10}$$

$$= \frac{\log_{6.5}(DOI(C_5) + 1)}{2} \times DOI(C_5) = 1.11 \tag{11}$$

$$DOI(C_4) = 1 + \frac{\log_2 2}{3} \times 1 \times 3 = 2.0 \tag{12}$$

From the variances of subconcepts of C_2 and C_4, the values 6.5 and 2 can be calculated with adding *bias*, respectively. The mean of all DOI's is 1.44 and the DOI of each concept is assigned after normalization. If the threshold value is 0.2, only C_4 and C_5 are extracted as the most interested concepts for the corresponding user.

In Figure 5, the i-Map of a particular user is represented in the form of a tree. Each node means the high ranked categories estimated as topics that the user is mostly interested in.

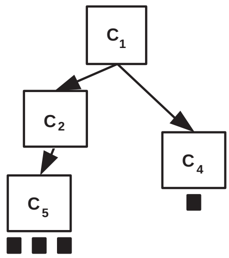

Fig. 5. Tree structured representation of *i*-Map for the high ranked concepts

5 COLLABORATIVE WEB BROWSING SYSTEM BASED ON RECOMMENDATION

The collaborative Web browsing system proposed in this paper is remote and asynchronous because this is based on Web environment and information about a participant's interests extracted from his own bookmarks and ontology. As shown in Figure 6, the whole system architecture consists of two main parts, which are a facilitator located between the users and the client-side Web browser that communicates with the facilitator.

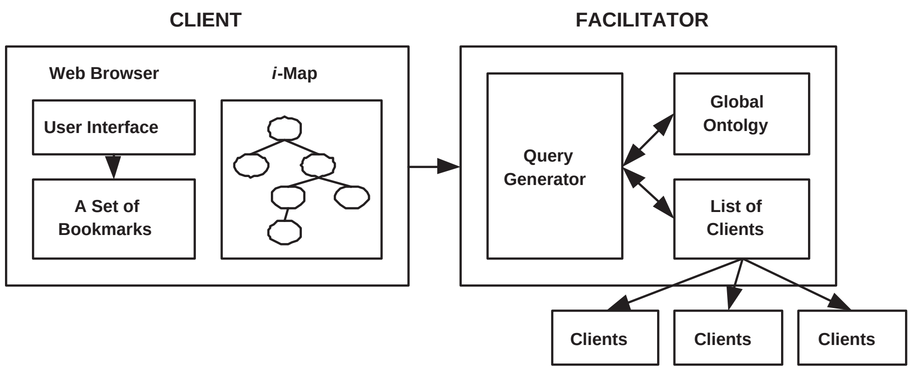

Fig. 6. System architecture

We embed autonomous and proactive agent module into this system. Every communication between agents is conducted regardless of user intervention. Also, while browsing to search information, users can be "implicitly" recommended from the facilitator in the following two ways:

- By querying specific information for the facilitator. After the information about a particular concept is requested, the facilitator can determine who has the maximum *DOI* for that concept by scanning his/her yellow pages.

- By broadcasting new bookmarks of like-minded users from the facilitator. Every time a user inserts a new bookmark, this fact, after conceptualization, is sent to the facilitator. Users thereby can obtain information related to the common concepts in their own i-Maps from neighbors.

Each client needs personal agent module. This agent initializes and manages the i-Map of the corresponding user based on bookmark repository. Thereby, it has to be able to communicate with facilitator agent, and refers to global ontology, e.g., Web directory for semantic labeling process. Through personal agents' reporting bookmarking activities of clients, the facilitator agent can automatically generate queries and recommendations.

6 EXPERIMENTS

We make this system executable on the Microsoft Windows platform. Personal Web browser is implemented by using *Borland Delphi* 6.0. Facilitator agent on server is implemented by using *Java* 2. Message generation is based on KQML (Knowledge Query and Manipulation Language) format, which is one of the most famous languages and protocols for exchanging information and knowledge [27].

In order to conduct experiments, we made up a hierarchical tree structure as a test bed for "Home: Science: Computer Science" from `Yahoo.com` and Open Directory Project (ODP) [4]. This tree consists of about 1300 categories and the maximum depth was eight. For gathering bookmarks, 30 users explored `Yahoo.com` and ODP directory pages during 28 days. Whenever users visit a Web site related to their own interests, they stored the URL information in their bookmark repositories. Finally 2718 bookmarks were collected.

We evaluated this collaborative Web browsing based on extracting user interests, as compared with single Web browsing. We adopted the measurements *recall* and *precision*. After all bookmark sets of the users were made empty, these users began to gather bookmarks again.

During fulfilling this task, users were being recommended relevant information from facilitating agent. Personal agents can retrieve the user interests information extracted from bookmarks up to that moment, and send facilitating agent queries for valuable information saved in the other users' repositories.

Figure 7 shows the number of saved bookmarks during browsing the Web. While the dashed line (testing data) means the number of bookmarks firstly collected for testing bed, the solid line is the number of bookmarks saved during collaborative Web browsing. As a result, in case of browsing with recommendations, users needed only 3.8 days for collecting 80 % of the total bookmarks. More exactly, it took 7.6 hours to collect these bookmarks. Users saved about 53.1 % of the total time spent for single browsing (16.2 hours).

The *precision* was measured by the rate of the inserted bookmarks among the recommended information set. In other words, this was the measurement for the accuracy of predictability. As the number of recorded bookmarks was increased,

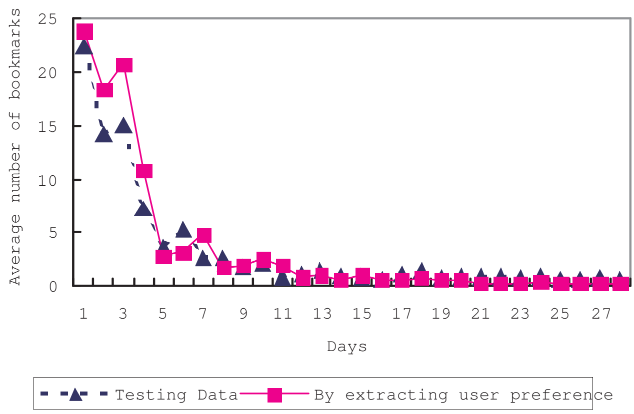

Fig. 7. Experimental result in the aspect of *recall*

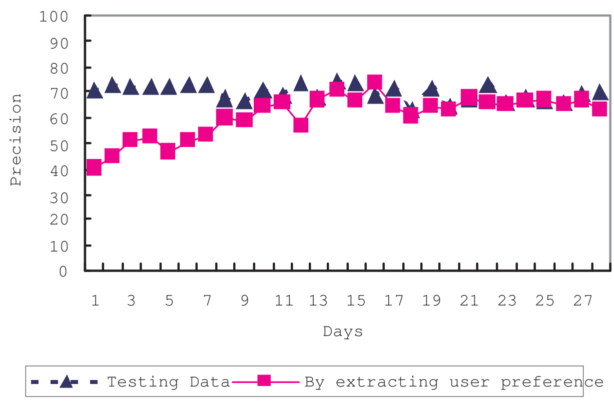

Fig. 8. Experimental result of in the aspect of *precision*

the estimation of user preferences was converged to be stable. Figure 8 depicts the experiment result of *precision* of recommendation based on user preferences. At the beginning, the *precision* of collaborative browsing was slightly lower because the user preferences were not set up yet. While user interests were extracted during the first six days, the *precision* of recommended information quickly tracked, compared with that of the testing dataset. For the rest of the experiment time, the *precision* was maintained in the same level with that of testing dataset.

7 CONCLUSIONS

This paper proposes that bookmarks are the most important user activity to support the extraction of user interests. However, due to the lack of semantic information from simple URL based bookmarks, we have been focusing on a way of conceptualizing them by referring to Web directories. When the semantic and structural information for users' bookmarks is properly provided, not only the precision but also the reliability of the extraction of user preferences was improved. More importantly, dynamic adaptation from user activities was efficiently implemented with sophisticated influence propagation between categories. Then, by establishing *i*-Maps of the corresponding users and *DOI*'s of the concepts on those map, we made it much easier to generate queries for relevant information and to share bookmarks among like-minded users. We have implemented a collaborative Web browsing system sharing conceptualized bookmarks. Based on the information recommendation on this system, we saved about 53 % of the searching time as compared with single Web browsing. Moreover, a beginner in a certain field can be efficiently helped by finding out valuable hidden information from some experts about that domain.

As future work, we are considering the privacy problems associated with sharing personal information such as age, gender, and preferences. The visualization of *i*-Map is also the next target of this work, in order to increase users' intuition recognizing their own preferences quantitatively with regard to each topic.

REFERENCES

[1] Armstrong, R.—Freitag, D.—Joachims, T.—Mitchell, T.: WebWatcher: A Learning Apprentice for the World Wide Web. Proceedings AAAI Spring Symposium on Information Gathering from Heterogeneous and Distributed Environments, 1997, pp. 6–12.

[2] Baeza-Yates, R.—Ribeiro-Neto, B.: Modern Information Retrieval. ACM Press, NY, 1999.

[3] Ding, C.—He, X.—Husbands, P.—Zha, H.—Simon, H. D.: PageRank, HITS and a Unified Framework for Link Analysis. Lawrence Berkeley National Laboratory Technical Report, No. 49372, 2002.

[4] DMOZ.Org: Open Directory Project (ODP).: Available on: `http://dmoz.org/`, 2001.

[5] Drake, F. L.: The XML Bookmark Exchange Language (XBEL). Available on: `http://pyxml.sourceforge.net/topics/xbel/`, 2004.

[6] Faatz1, A.—Kamps, T.—Steinmetz, R.: Background Knowledge, Indexing and Matching-Interdependencies of Document Management and Ontology-Maintenance. Proceedings of the 1st Workshop on Ontology Learning, 2000.

[7] Giarratano, J.—Riley, G.: Expert Systems Principles and Programming. PWS publishing company, 1994.

[8] Graphic, Visualization, and Usability Center: GVU's 10th WWW User Survey. Available on: http://www.gvu.gatech.edu/user_surveys/survey-1998-10/, 1998.

[9] Greif, I.: Computer-Supported Cooperative Work: A Book of Readings. San Mateo, CA, Morgan Kaufmann Publishers Inc., 1988.

[10] Gruber, T. R.: A Translation Approach to Portable Ontologies. Knowledge Acquisition, Vol. 5, 1993, No. 2, pp. 199–220.

[11] Jung, J. J.—Jo, G.-S.: Extracting User Interests from Bookmarks on the Web. Proceedings of the 7th Pacific-Asia Conference on Knowledge Discovery and Data mining, 2003, pp. 203–208.

[12] Jung, J. J.—Yoon, J.-S—Jo, G.-S.: BISAgent: Collaborative Web Browsing Through Sharing of Bookmark Information. Proceedings of the International Conference on Intelligent Information Processing (IIP), IFIP, 2000.

[13] Jung, J. J.—Yoon, J.-S—Jo, G.-S.: Collaborative Information Filtering by Using Categorized Bookmarks on the Web. Proceedings of the 14th International Conference on Applications of Prolog, 2001, pp. 343–357.

[14] Kaufman, L.—Rousseeuw, P. J.: Finding Groups in Data: An Introduction to Cluster Analysis. John Wiley, 1990.

[15] Kleinberg, J. M.: Authoritative Sources in a Hyperlinked Environment. Journal of the ACM, Vol. 46, 1999, No. 5, pp. 604–632.

[16] Labrou, Y.—Finin, T.: Yahoo! as an Ontology: Using Yahoo! Categories to Describe Documents. Proceedings of the 8th International Conference on Information Knowledge Management (CIKM), 1999, pp. 180–187.

[17] Lieberman, H.: Letizia: an Agent That Assists Web Browsing Agent. Proceedings of the 14th International Joint Conference on Artificial Intelligence, 1995, pp. 924–929.

[18] Lieberman, H.—Dyke, N.—Vivacqua, A.: Let's Browse: A Collaborative Web Browsing Agent. Proceedings of the International Conference on Intelligence User Interface (IUI), 1999, pp. 65–68.

[19] Maedche, A.: Ontology Learning for the Semantic Web. Kluwer Academic Publishers, 2002.

[20] Maes, P.: Agents That Reduce Work and Information Overload. Communication of ACM, Vol. 37, 1994, No. 7, pp. 31–40.

[21] McCallum, A.—Nigam, K.—Rennie, J.—Seymore, K.: Building Domain-Specific Search Engines with Machine Learning Techniques. Proceedings of AAAI-99 Spring Symposium on Intelligent Agents in Cyberspace, 1999.

[22] Pearl, J.: Probabilistic Reasoning in Intelligent Systems. Morgan Kauffman Publisher, 1988.

[23] Rodden, T.: A Survey of CSCW Systems. Interacting with Computers, Vol. 3, 1991, No. 3, pp. 319–354.

[24] Termier, A.—Rousset, M.-C.—Sebag, M.: Combining Statistics and Semantics forWord and Document Clustering. Proceedings of the 2nd Workshop on Ontology Learning (OL '2001), 2001.

[25] Twidale, M.—Nichols, D.: Collaborative Browsing and Visualization of the Search Process. Electronic Library and Visual Information Research, 1996, pp. 51–60.

[26] Twidale, M. B.—Nichols, D. M.: Computer Supported Cooperative Work in Information Search and Retrieval. Annual Review of Information Science and Technology, Vol. 33, 1998, pp. 259–319.

[27] UMBC KQML Web: Knowledge Query and Manipulation Language (KQML). Available on: `http://www.cs.umbc.edu/kqml/`, 1997.

Jason J. Jung is a Ph.D candidate in Intelligent E-Commerce Systems Laboratory, Inha University. His research topics include intelligent data analysis based on machine learning, semantic Web mining, and ambient intelligence. He was a visiting scientist in Fraunhofer Institute (FIRST) in Berlin, Germany in 2004.