Information Acquisition Support by *Information Map*

Koichiro Kawase

School of Information Science,
Japan Advanced Institute of Science and Technology

February 13, 1998

Keywords: collaborative filtering, visualization, computer network environment.

1 Background and Purpose

A computer user can get a variety of information in large quantities through the networks owing to remarkable development about WWW, Netnews, and e-mail, etc. However, there has been tremendous growth in the quantity and the diversity of information available. As a result, there is now a need for the user to exclude unnecessary information and obtain new and useful information.

"Collaborative Filtering" is a powerful technology to deal with the above need. Collaborative filtering is summed up as follows: When the user acquires information, he/she does according to not only his/her own knowledge and values but also others'.
Collaborative filtering is based on the point of view that people have good friends as filter to select the useful information. For instance as a typical filtering system, there is something which the user can select information referring to the annotations and the evaluations which are put by other. Alternatively there is the system which ranks information by voting and displays that ranked information with high scores priority.

However, there are the problems in an existing collaborative filtering system. For example, when the annotation is required by the natural-language, the user gets tired of annotating. And there are a lot of filtering systems which cannot do filter functionally in case of not using by many users. Moreover, there are the problems which should be considered when the system presents information to the user. For instance, the user cannot grasp the significance of the information only by enumerating useful information, or the categorized expression is not always suitable for user's thought.

We propose a new collaborative filtering system based on the above background and the consideration of the information acquisition activity of the person. Then we evaluate our prototype system and examine the effect of our proposal.

2 Design of the System

In this research, to propose the policy for the design of our new system, we survey the activity when people acquire information. Hence we gave the questionnaire survey. Now, we express the basic policy of our system based on the results of our survey.

First, people try to get information with proper combination of various information media. A person gets information from WWW finally, but he/she might get to know the URL of this content from Netnews. Therefore our system treats uniformly various types of information on the article of Netnews and in the contents of WWW etc.

Second, people have various types of information, and this information are not considered independently but relations are made among these by individual reasons. People will acquire new and useful information referring to organized and related information by themselves. Therefore when we consider our new collaborative filtering system, we propose the policy that organized and related information which each user owns is visualized and user can refer to the visualized object with each other. Information organized by each user is expressed by the graph. In this research, we define the graph as *Information Map* and the edge as "link". The node in the information map shows the whereabouts of information and the link shows the presence of the relations between information. Others' knowledge is able to be referred to by merging the information maps among users. It is planned for each user to acquire new and useful information by this function.

3 Implementation of the Prototype System

We implemented the prototype system with the perl language and the java language according to above-mentioned design policies. The information map is displayed on a Web browser as java applet. The user can access the entities of information by clicking the nodes in this applet.

Each node in the information map is linked by two kinds of link. One kind of link is actually put between information in accordance with the format of HTML. This is generated based on the access log of the user whom httpd records. Another kind of link is added automatically by analyzing information. When a user refer to others' knowledge by using information map system, the automatic link can represent the relations between some information which he/she has known already and new information which the system indicates. So such function can give the users some suggestions for priority of access to new information or relevance between own interests and new information.

This applet also has text area which indicates title, URL, and key words included in each node. By using this function, the users can guess the priorities and the relevances of information.

4 Evaluation of the Prototype System

We performed the experiment to examine the efficiency of the prototype system. By analyzing the conduct of testees, we obtained the following observations.

- Using this prototype system enables one to acquire new information efficiently and usefully in comparison with the effort by oneself.
- There are problems about link as follows.
  - A lot of generated links in the information map may be displayed in complex.
  - The key word group displayed in the text area may not become an effective reference for the testee.

5 Future Works

It has been recognized that the method proposed by this research makes it easier to obtain new and useful information for the user of the computer network environment. We think that the support of the information acquisition allows one certain further more by dealing with the problem clarified by the evaluation experiment and advancing the verification in addition.

The tasks in the future are shown as follows;

- When the information map will be made, by analyzing information more accurately,
  - the more meaningful key words should be extracted.
  - the precision of links should be improved, and it would be given more significance to the existence of the links.
- Verifying the scalability of a number of users and information.
- Verifying the performance in the long term.