Experimental Results for a Method to Discover Human Relationship Based on WWW Bookmarks

Masahiro HAMASAKI [†], Hideaki TAKEDA [‡] [†]

[†] *Graduate School of Information Science, Nara Institute of Science and Technology,*
8916-5 Takayama-cho, Ikoma-city, Nara 630-0101, JAPAN
[‡] *National Institute of Informatics,*
2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, JAPAN
[†]*masa-ha@is.aist-nara.ac.jp* [‡]*takeda@nii.ac.jp*

Abstract. In this paper, we show how human network can be found and used on WWW. We proposed a system called kMedia that can assist users to form knowledge for community by showing shared topics networks (STN) among them. kMedia uses bookmarks as users' knowledge because the structure of bookmarks and web pages registered in bookmarks reflect user's interest. We conducted an experiment to know how kMedia can support users. One result is that folder recommendation is more effective than page recommendation. The other is that recommendation is more effective for people belonging to the same real communities than those to different communities. According to these result, we propose a new measurement called "category resemblance" that is recommendation measurement based on resemblance of folder structures. This measurement shows higher than all other system generated parameters and human evaluation to detect human relationship.

1 Introduction

Human network is one of powerful and practical ways to solve information flood because we can act intelligent agents to collect, filter and associate necessary information. We have proposed a system called kMedia that can support users to form community knowledge by showing shared topic networks (STN) among them [1]. A shared topics network is a network where each user is associated to other via topics owned by other users. A preliminary test showed that this approach worked well to guide people. In this paper, we investigated why this approach can work well with more precise experiments. One interesting result is that recommendation is more effective among people whose categorization of information are similar to each other. According to this result, we propose a new measurement called "category resemblance" that is recommendation measurement based on resemblance of folder structures.

This paper is organized as follows. In Chapter 2, we show kMedia system overview. We show how human network can be found and used on WWW using kMedia. Results of an experiment of kMedia are shown in Chapter 3. We compare our system with other systems in Section 4, and conclude this paper in Chapter 5.

2   System Overview

kMedia is a client-server system where each client system is provided for a user and a server system is provided for a community. A client works to process a user's bookmark files to extract keywords and show results to the user, and the server to calculate page similarity and determine topics relations.

The system reads her/his bookmark file and extracts keywords for each URL in it by fetching pages for these URLs and analyzing their texts. It extracts words from texts except stop words and some common words, then selects some of the most occurred words in them. Then the system calculates similarity between every pair of pages in the collected bookmark files. Page relevance between two pages is measured by sums of occurrence of keywords that appeared in the both pages.

After the system calculates relevance between pages, the system calculates relevance between folders. It finds pairs of folders in which relevant pairs of pages exceed the threshold. We call the number of marked pair folder relevance.

3   An Experiment

3.1   The Evaluations of Recommended Pages, Folders and Persons

The main objective of the experiment is how our proposed method is useful to identify users' relations. We asked twelve persons to submit their bookmark files to the system, and asked subjective evaluation to each recommended page and folder, and person who recommended such pages and folders by ranking 5 to 1 (5 is the best, and 1 is the worst). Criteria for evaluation are how much suggested pages or folders are similar or useful according to their own bookmark and how much she/he want to contact or meet the recommending persons.

The left graph in Figure 1 plots users' page evaluation against page relevance, and the right plots users' folder evaluation against folder relevance. The gradient of the second graph is apparently larger than the first one. It indicates that folder recommendation is more effective than page recommendation.
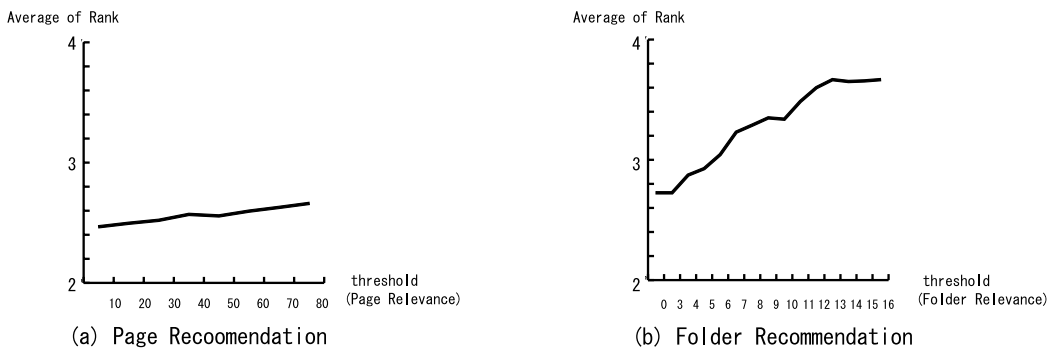

Figure 1: Comparison between page and folder recommendation

The next result is analysis on what is influential to evaluations on human relations. We calculated correlation between evaluation on human relation and other parameters, e.g., average of similarity relations between recommended pages and average of evaluation for recommended folders. The result is shown in Table 1 and 2. Table 1 shows correlation between

evaluations on human relations and parameters on page/folder recommending, i.e., the system's performance. Table 2 shows correlation between these and evaluations on page and folder, i.e., evaluation by human.

Table 1: Correlation between evaluation on human relations and recommendation of pages and folders (calculated by the system)

	Want to Contact	Want to Meet
the number of recommended page	0.42	0.30
average of page relevance	-0.13	-0.19
the number of recommended folder	0.45	0.35
average of folder relevance	0.38	0.30

Table 2: Correlation between evaluation on human relations and evaluation on page and folder

	Want to Contact	Want to Meet
average of evaluation of recommended page	0.29	0.40
average of evaluation of recommended folder (usefulness)	0.28	0.32
average of evaluation of recommended folder (similarity)	0.09	0.20

These tables show clearly different results. In terms of human evaluation, "want to meet" is more related than "want to contact" to page/folder evaluation. Coefficient of evaluation of the parameter calculated by system is less than one of the evaluations from subjects. As a result, the simple parameter of calculated by system such as a number of similarity pages is not enough to discover relationship among people.

3.2   *The Distinctions between in-community and cross-community relationship*

Human relationship has not been clearly identified by information that are shared by people. To know the reason we draw a comparison of page and folder recommendation between in-community and cross-community relationship. The aim of this comparison is to discovery parameters to indicate human relationship. Here in-community relationship means that all the subjects belong to a community in their real life, and cross-community relationship means that the subjects belong to different communities.
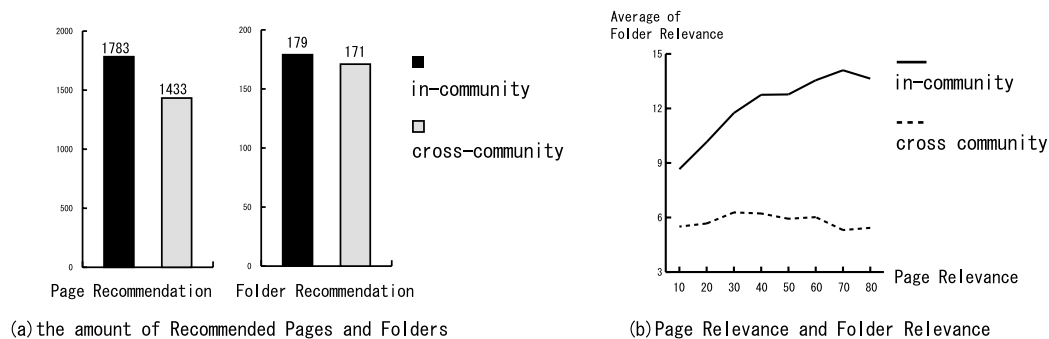

Figure 2: Comparison of in- and cross-community relationship for page and folder recommendation

Figure 2(a) shows that the amounts of recommended pages and folders for in-community and cross-community relationship. Both the numbers of recommended pages and folders are in case of in-community larger than in case of cross-community. But difference in folder recommendation is not so big in comparison with page recommendation.

Figure 2(b) is a graph in which the horizontal axis indicates relevance for recommended pages and the vertical axis one for recommended folders. There is a high correlation between folder relevance and page relevance in case of in-community relationship. However such a tendency does not appear in case of cross-community relationship.

It means that relevance of pages and folders are more related when subjects are within a community than when those are without a community.

3.3 Category Resemblance

As mention above, there is the distinction of effects for parameters calculated by the system between in- and cross-community relationship in recommendation. It is easy to ascribe real communities, but is there any measurement to characterize effects of real community? We propose category resemblance to measure possibility for information recommendation. We have found that relevance pages and folders are related when people are in-communities. In others words, larger relevance of folders can suggest larger longer relevance of pages.
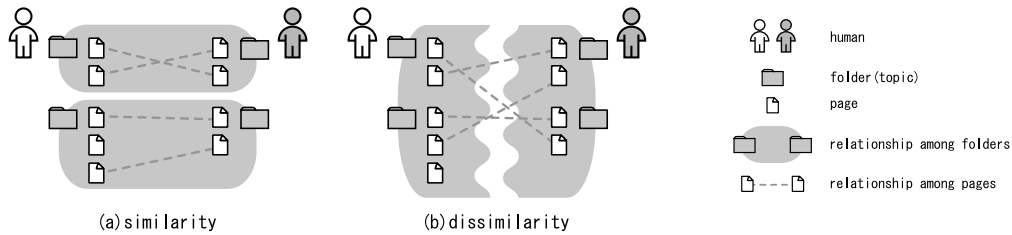

Figure 3: Similarity in method of categorization

In term of the number of recommended pairs of pages, there is no difference between Figure 3(a) and 3(b). However, similarity of methods of categorization between two persons in (a) is larger than (b), so that relationship among people in (a) seems to be stronger than the relationship in (b).

We propose a new measurement called "category resemblance" that is recommendation measurement based on resemblance of folder structures. It is defined as the following formula:

$$C_{ij} = \frac{Nf_{ij} \times Rf_{ij}}{Np_{ij}}$$

C_{ij} : $Categorize\ Resemblance$ Rf_{ij} : $Average\ of\ folder\ rational$
Nf_{ij} : $Number\ of recommended\ folder$ Np_{ij} : $Number\ of\ recommended\ page$

Table 3 shows that correlation between category resemblance and the evaluation on human relations. This correlation coefficient is higher than not only all other parameters calculated by the system, but also human evaluation on pages and folders.

Table 3: Relationship between parameters on human relations and category resemblance

	Want to Contact	Want to Meet
Categorize Resemblance	0.49	0.55

4 Related Work

Kautz et al.[2] emphasized importance of people relations for WWW and have done primary work for finding people relations, i.e., their system called Refferal Web can find people by analyzing bibliography database. Our aim is very similar to them, but we realized it differently. The benefit of our approach is to identify topics shared by users.

Community Board[3][4] is another way to explicate users' relations by using topics. This system can show dynamics of interest on topics, i.e., who and when initiates and participate discussion for topics. But no discovery of topics is supported because topics themselves in this study are already shared by participants.

Grassroots[5] proposed to use "folder structure" as a basic structure to organize information and people. Grassroots approach is excellent but imposes a heavy charge to every user because it requires unifying a folder structure of various activities from information storing to information and even folders for other users. Our system can ease this problem by finding topic relations among users automatically.

5 Conclusion

In this paper, we discuss how relation among people should be explicated to facilitate information exchanging and proposed a system called kMedia that can show shared topics networks for this purpose. Our method to identify shared topics network is simple and effective. We use folder structures as topics and identify inter-topic relations by analyzing texts associated to the folders. This combination of human knowledge and automatic discovery of relations works well. Even discovered relations between pages are not sufficient, discovered relations between topics can be acceptable. We investigated how people can be more acceptable with folders than in pages, and found that people can believe other that have similar aspects for information. By examining our experiment results, we can identify a measurement called category resemblance to calculate it. Category resemblance shows how much the methods of information categorization are similar. This measurement shows higher than all other system generated parameters and human evaluation to detect human relationship.

References

[1] H. Takeda, T. Matsuzuka, Y. Taniguchi. Discovery of shared topics networks among people – a simple approach to find community knowledge from www bookmarks. In *Proceedings of the PRICAI00*, 2000.

[2] H. Kautz, B. Selman, and M. Shah. Refferal web: Combinign social networks and collaborative filtering. *Communications of the ACM*, 40(3):63-65, 1997.

[3] F. Hattori, T. Ohguro, M. Yokoo, S. Matsubara, and S. Yoshida. Socialware: Multiagent systems for supporting network communities. *Communications of the ACM*, 42(3):55-61, 1999.

[4] S. Matsubara, T. Ohguro, and F. Hattori. Communityboard: Social meeting system able to visualize the structure of discussions. In *Proceedings of the 2nd International Conference on Knowledge-based Intelligent Electronic Systems (KES'98)*, pages 423-428, 1998.

[5] K. Kamiya, M. Róscheisen, and T. Winograd. Grassroots: A system providing a uniform framework for communicating, structuring, sharing information, and organizing people. In *Proceedings of The 6th International World Wide Web Conference (WWW-6)*, 1997.