Special Feature

User Support Systems with Adaptive User Profiling Based on User Activity Histories

Tetsushi Morita, Tetsuo Hidaka, Hisanobu Matsuoka, Shinichi Okano, Akimichi Tanaka, and Yasuhisa Kato[†]

Abstract

This article introduces two systems that apply user profiling technology to Internet searching. That is, they use the activity histories of individuals as a basis for determining the unique characteristics of individuals or communities to help people find ways to use the Internet that are best suited to them.

1. What is user profiling?

Internet search services usually provide identical data to all users, but some recent net services provide and display data that suits a user's individual taste, characteristics, and situation. Search techniques that use user profiling to help users find the data they want among an enormous amount of information available on the Internet are gradually being developed. User profiling technology is intended to acquire information related to users' interests and aims. There are two main approaches.

(1) Direct method: Ask users directly to tell us about their interests etc. via a questionnaire or by active selection.

(2) Indirect method: Record a user's behavior and construct his/her activity history and then automatically identify current interests etc. from it.

The direct method is easy to implement, but the burden on the user will be heavy because it relies on the user continually entering information into the system; therefore, it does not seem practical. NTT Cyber Solutions Laboratories has developed user profiling methods with the ability to learn from the user's activity history. These methods enable us to collect user profiles using the indirect method, which greatly decrease the user's workload. In this article, we introduce two systems in which these methods have

been implemented. Memory-Retriever (MR) [1] accumulates and processes user operations on the local terminal (personal computer (PC)), while XappaLinks [2] focuses on the user's activity history in blog (Weblog) space.

2. Memory-Retriever (MR): stores and retrieves PC operations

MR is an application that extracts a user profile by automatically recording the user's behavior such as browsing websites, printing data, and sending or receiving e-mail. It also retrieves the extracted profiles and shows the results. As **Fig. 1** shows, MR consists of an action storage part, profile extraction part, and application programs. It runs on the user's PC.

The action storage part monitors the Web browser, operating system (OS), etc. and records the user's behavior as an activity history. "Text Selection", shown in **Fig. 2**, means to record the text that the user selects when browsing a website as a sign of his/her interest in the material. The action storage part consists of a core and a monitoring plug-in; more behaviors can be recorded by adding new plug-ins. For privacy protection reasons, data is encrypted before being stored. The profile extraction part has three functions:

(1) Experience score analysis

This function calculates "experience scores", which represent how important each behavior is to the user such as web page browsing. In more detail, experience scores are calculated based on the accu-

† NTT Cyber Solutions Laboratories
Musashino-shi, 180-8585 Japan
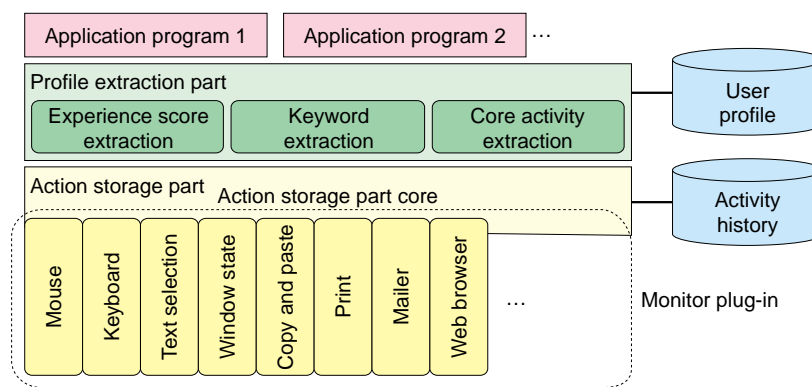E-mail: kato.yasuhisa@lab.ntt.co.jp

Fig. 1.   System configuration of MR.
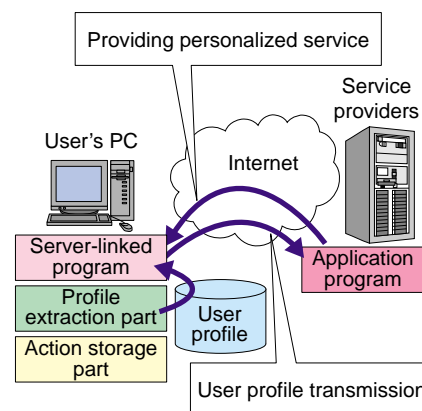

Fig. 2.   Text selection.

Fig. 3.   User profile usage example.

mulation of the times when a PC window remains active and also that of the frequency of print, copy, and paste operations performed by the user. These scores cannot be calculated from the simple access logs held in Web servers.

(2) Keyword detection

This function extracts keywords from the recorded behaviors to index them. The most frequent keywords are helpful in determining the content in which the user is highly interested.

(3) Key action sequence integration

This function extracts the chronological sequence of user's actions that include common keywords, as a "key action sequence". A key action sequence demonstrates that a user acted for the same purpose for a certain period of time, which is important to the user.

An application program has been developed for each user profiling service. As shown in **Fig. 3**, user profiles are first stored in the user's PC. However, we may use the user profiles in different ways; that is, the

information needed to support the desired service will be sent to the service provider through the Internet, and personalized service will be provided based on the sent information.

MR has a function for retrieving the user profiles and displaying search results. The "key action sequences" when the keyword "NTT" was used are shown in **Fig. 4**. The results show that information, site, and laboratories were also extracted as the most frequently appearing keywords. From those results, we can estimate the purpose of an action sequence. The details of a key action sequence are shown in **Fig. 5**. You can see the individual actions forming the key action sequence. The figure also shows that individual actions are displayed here as thumbnails of website pages browsed. Those surrounded by a color frame have high experience scores. At the same time, we can see the link anchor text clicked by user and the text selected by the "Text Selection" function, which reveals the user's interest or purpose.
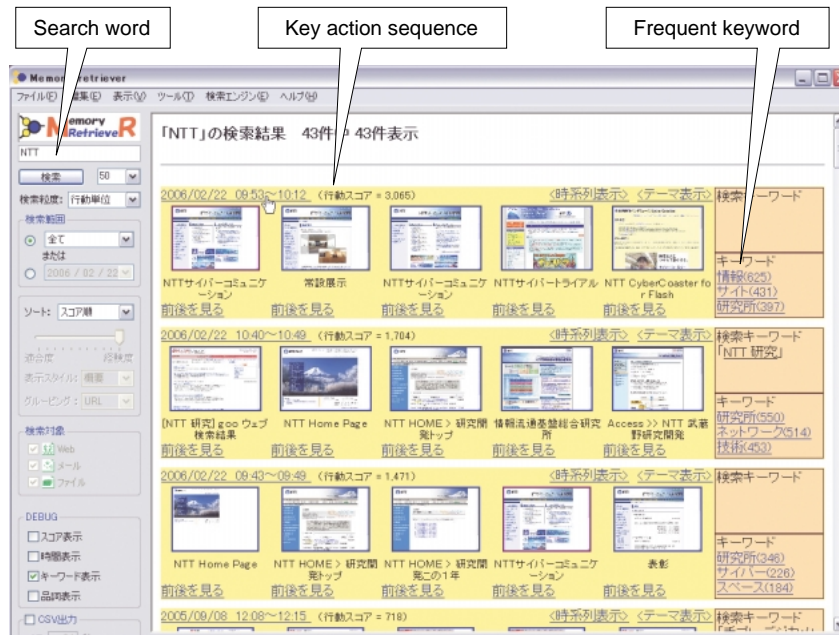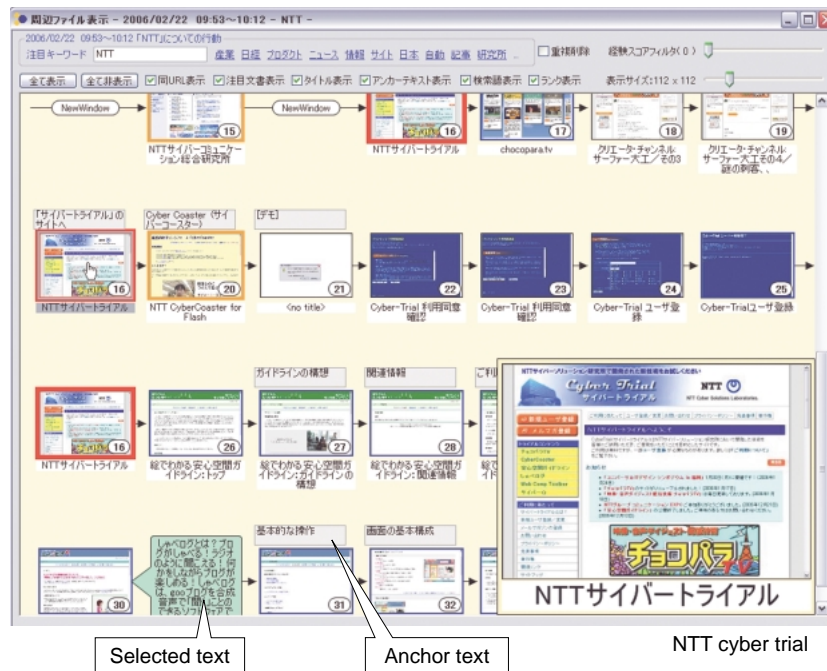
Fig. 4.   User profile search results.

Fig. 5.   Display of user profile details.

3.   XappaLinks: supports blog communication

Weblogs (blogs), which enable individuals to easily and continuously exchange information over the Internet, provide unique communication functions such as comments and trackbacks[*1]. Therefore, blogs are becoming a dominant information medium fol-lowing bulletin board services (BBS), websites, and mailing lists. However, it is hard to acquire all the communication activities among blogs, and users must put in a lot of effort to determine the relationship

---

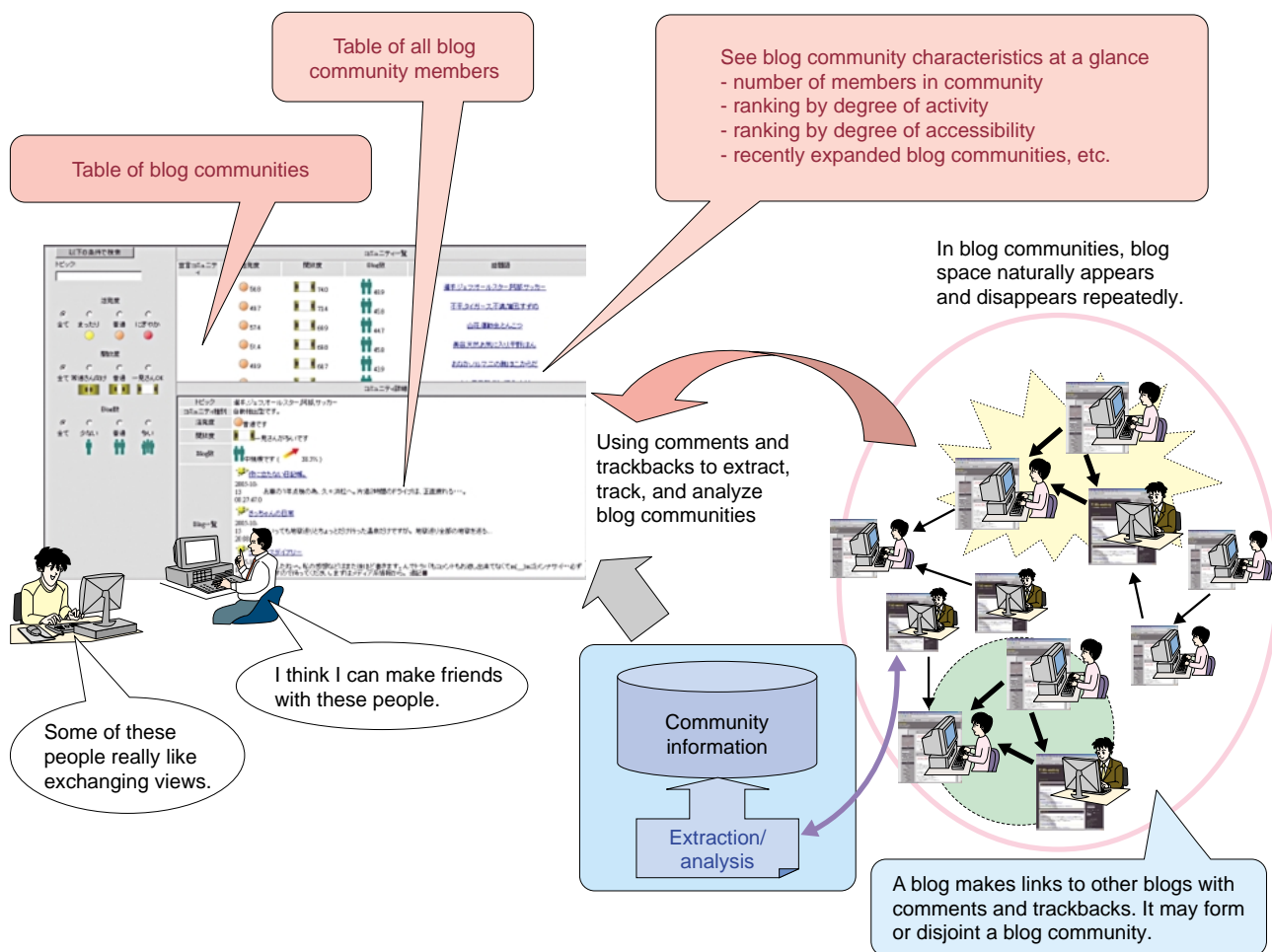*1   Trackback is a mechanism for communication among blogs.

Fig. 6.   Blog communication profiling technology.

among bloggers by viewing the contents of each blog. XappaLinks solves such problems by putting blog communication profiling technology and blog profiling technology to practical use.

Blog communication profiling technology (**Fig. 6**) includes a blog community extraction engine and a blog community analysis engine. The blog community extraction engine tracks communication activities between blogs such as comments or trackbacks to extract frequently communicating blogs. The blog community analysis engine analyzes communication activities in the extracted blog community to quantify the characteristics of the community such as the degree of activity and openness to outside opinion. It also extracts topics that frequently appear in the community.

Such extraction or analysis results are ranked on the screen of a community browser which runs over a WWW (World Wide Web) browser so that users can see the characteristics of each blog community or a

list of blogs in which they participate. Moreover, to see the characteristics of blogs, users can use XappaLinks to search not only by keyword as usual, but also by blog characteristics.

4.   Community badges

Each blog community tends to be aloof from real society connections, just like a social networking service (SNS), so it can be said that the communities favor looser connection on the network. Therefore, it is difficult to recognize to what blog community a user belongs, what kind of people there are in the community, and to what extent his/her field of activities has expanded.

A community badge is part of the XappaLinks system's mechanism for building a new blog community for a blogger who registers an image as a symbol of the blog community. A badge made from the registered image will be put on the members' blogs (**Fig.
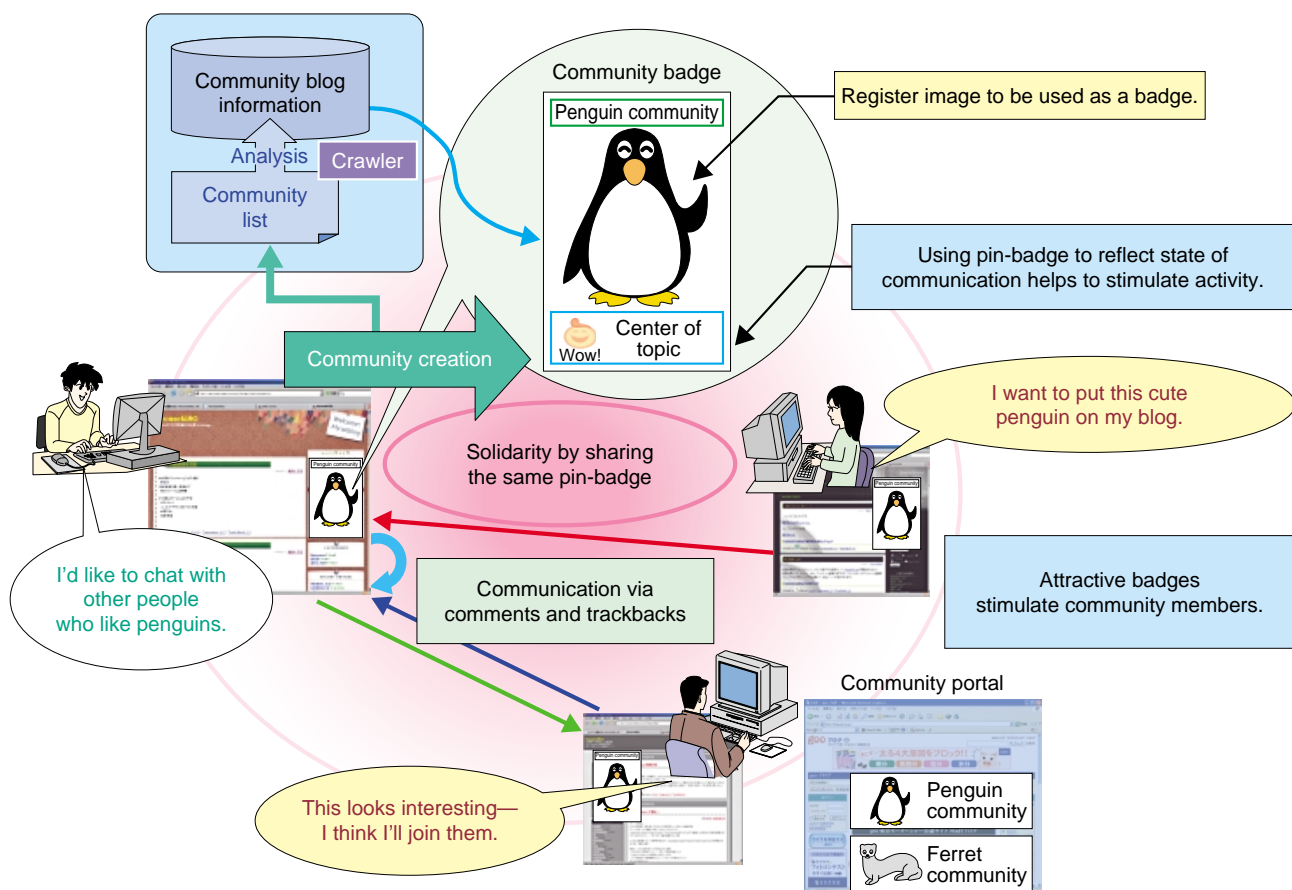
Fig. 7. Community badge.

7). Affixing the same badge creates a sense of unity among the members and not only gives them a sense of belonging to the community all the time but also encourages new members to participate in the community.

Community badges can also be used to present the role of each blog in the blog community based on the estimation extracted by the blog profiling technology, which extracts the characteristics of each blog based on observations of blogger communication activities such as comments or trackbacks and the frequency of blog entry updates. Community badges categorize the role of each blog into three typical types based on the calculated indexes or characteristic values such as how many trackbacks have been collected and what percentage of comments/trackbacks from others have been replied to. In this way, assigning community badges to present the role of each blog is expected to make bloggers quickly realize their roles and to maintain and improve motivation for updating blog entries or building up communication activities.

5. Future development

User profiling technology can be used not only for a personal tool that automatically provides information to support the user, but also for services such as advertising, marketing, recommendation service in electronic commerce sites, and social services that share profile information. To develop more accurate user profiles, we are committed to expanding the range of activity history and analyzing the information in more detail. We intend to verify the effectiveness of the user profiling technology by applying it to practical services.

References

[1] T. Morita, T. Hidaka, T. Kura, K. Ooura, and Y. Kato, "Desktop search system based on the Action-Oriented algorism," (sic) Proc. APSITT2005 (6th Asia-Pacific Symposium on Information and Telecommunication Technologies), pp. 204-207, 2005.

[2] A. Miyata, H. Matsuoka, S. Okano, S. Ishiuchi, N. Arakawa, and Y. Kato, "XappaLinks: Blog Community Support System," IEICE General Conference D-13-2, 2006 (in Japanese).

Tetsushi Morita

Research Engineer, Media Computing Project, NTT Cyber Solutions Laboratories.

He received the B.E. and M.E. degrees in engineering science from Kyoto University, Kyoto, in 1996 and 1998, respectively. Since joining NTT Software Laboratories in 1998, he has been engaged in research on e-commerce systems. His current research interests include information technology for personal data management and information retrieval. He is a member of the Information Processing Society of Japan (IPSJ).

Shinichi Okano

Engineer, Media Computing Project, NTT Cyber Solutions Laboratories.

He received the B.E. and M.E. degrees in computer science from Keio University, Kanagawa, in 1996 and 1998, respectively. He joined NTT Software Laboratories, Tokyo, in 1998. He is currently engaged in R&D of blog community systems and blog profiling systems. He is a member of IEICE.

Tetsuo Hidaka

Research Engineer, Media Computing Project, NTT Cyber Solutions Laboratories.

He received the B.E. and M.E. degrees in instrumentation engineering from Keio University, Kanagawa, in 1993 and 1995, respectively. He joined NTT Telecommunication Networks Laboratories in 1995. His current interests include human-computer interaction and information retrieval systems. He is a member of the Institute of Electronics, Information and Communication Engineers (IEICE) of Japan and IPSJ.

Akimichi Tanaka

Senior Research Engineer, NTT Cyber Solutions Laboratories.

He received the B.E and M.E. degrees in precision machinery engineering from the University of Tokyo, Tokyo, in 1985 and 1987, respectively. He joined NTT Information Processing Laboratories, Tokyo, in 1987. He has been engaged in R&D of artificial intelligence, pattern recognition, and educational systems. He is a member of IEICE and IPSJ.

Hisanobu Matsuoka

Research Engineer, Media Computing Project, NTT Cyber Solutions Laboratories.

He received the B.S. and M.S. degrees in physics from Sophia University, Tokyo, in 1992 and 1994, respectively. He joined NTT Software Laboratories, Tokyo, in 1994. His current research area is analysis of communication activities in blog space. He is a member of IEICE.

Yasuhisa Kato

Senior Research Engineer, NTT Cyber Solutions Laboratories.

He received the B.E and M.E. degrees in applied systems science from Kyoto University, Kyoto, in 1988 and 1990, respectively. He joined NTT Information Processing Laboratories, Tokyo, in 1990. He has been engaged in R&D of voice processing, educational systems, and search systems. He is a member of the Japanese Society for Artificial Intelligence and IPSJ.