Social tagging using marked strings in Web pages

Yuki Matsuoka[i ii iii], Ryuuki Sakamoto[iii], Sadanori Ito[iv iii], Hideaki Takeda[i ii iii], Kiyoshi Kogure[iii]

i The Graduate University for Advanced Studies,
Shonan Village, Hayama, Kanagawa 240-0193 Japan

ii National Institute of Informatics,
2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430 Japan

iii Advanced Telecommunications Research Institute International,
2-2-2 Hikaridai, Seikacho, Soraku-gun, Kyoto 619-0288 Japan

iv Tokyo University of Agriculture and Technology,
3-8-1 Harumi-cho, Fuchu-shi, Tokyo 183-8538 Japan

E-mail : m- yuki@grad.nii.ac.jp

ABSTRACT

In this paper, we propose a system called "Aikuchi," which enables users to mark a string on a Web page and share it with other users. If users mark a string on their Web browser using their mouse cursor, Aikuchi recommends links to other Web pages based on recommendation algorithms. When a user clicks a recommended link, Aikuchi highlights a marked string. We offered the system as part of a conference support system. According to the analysis of user logs, users preferred Web page recommendations based on strings marked by other users to those based on page similarity using TFIDF and collaborative filtering. As a result, we think that marked strings can act as user-generated metadata.

Categories and Subject Descriptors

H.3.3 [**Information Search and Retrieval**]: Search process, Selection process; H.3.5 [**Online Information Services**]: Data sharing

General Terms

Experimentation, Human Factors

Keywords

Social tagging, marking, user-generated metadata

1. INTRODUCTION

It goes without saying that there is an enormous number of pages on the Internet; therefore, we often need to use search engines to find ones that interest us. However, search results frequently include "noise" pages. While search engines do return pages matching users' query words, some pages may include words that differ from what users intended. One possible solution to this is to employ user-generated tags from Social Bookmark Services (SBS), enabling users to easily find appropriate pages. However, as users freely add tags, numerous other types of tags also accumulate on Web pages.

Because users view each page subjectively, they add tags to express the contents, their feelings about the page, the purpose of the page, and so on. When users find Web pages using such tags, though, such diversity becomes an obstacle to finding pages that are actually appropriate. In this research, we employ strings marked by users as tags to exclusively express a given page's contents. The system we have implemented, called *Aikuchi*, enables users to mark strings in Web pages and share them with other others. If a user marks a string in the Web browser using his or her mouse cursor, Aikuchi recommends links to other Web pages by applying it recommendation algorithms, and when a user clicks a recommended link, the system highlights the marked string, which is called a *footprint*. The recommendation algorithms are based on page similarity using TFIDF, collaborative filtering, and word matching using words in the footprint or another Web page. We offered the system as a part of a conference support system and analyzed whether users prefer the recommendation algorithm based on marked strings to search Web pages.

2. RELATED RESEARCH

Semantic annotation systems have previously been proposed that highlight strings in Web browsers. One example is CHOSE [1], which stores metadata as hyperlinks in a Distributed Links Service [2] and uses it to highlight ontology terms in the text strings on the Web browser. Magpie [3] highlights strings related to the ontology of the user's choice on the browser. Both systems highlight strings to make the connection between a string and an ontology. Our system links from a string to offered pages and differs from them by the lack of connection between the link and a meaning; it is instead related to the user's interest. In this paper, we reuse a selected string as metadata to search Web pages.

SAAW'06, November 6, 2006, Athens, GA, USA.

3. A SOCIAL TAGGING SYSTEM

3.1 Overview

We propose a system called *Aikuchi* that enables users to mark strings in Web pages and jump to other pages from the marked string. The system lets users search pages using marked strings from other users.

Details of the system are as follows: Users can mark a string in their Web browser using their mouse cursor when they find something of interest on a page. (Fig. 1). Then the recommendation window pops out to display recommendations from other pages called *recommendation links* (Fig. 2). If the user selects a recommendation link, the system shows the specified page, and once a recommendation link is selected, the system highlights the used string. We call this highlighted string a *footprint* (Fig. 3). It is shared; i.e., users can see the footprints added by all other users. If the user places her or his mouse cursor over the footprint, the recommendation window pops out again, with Fig. 4 Showing examples of *recommendation links* and *footprint links*. The latter includes Web pages to which some users have jumped from the marked string.

Figure 1. Marking on the Web browser

The left figure indicates title, authors and abstract for a paper. The right figure is translation of the left it.

Figure 2. Popped-out recommendation window

The right figure shows translation of the recommendation links

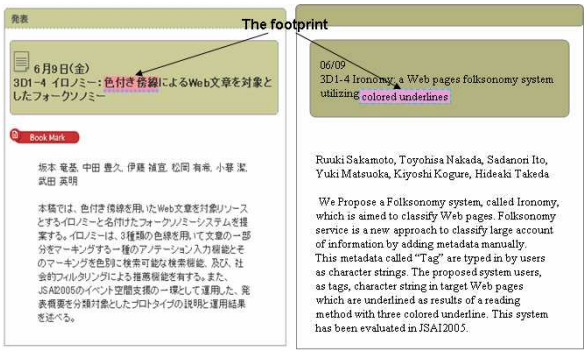

Figure 3. The footprint

The right figure is translation of the left it.

Figure 4. Recommendation window after the cursor is placed over the footprint

The right figure shows translation of the footprint link and the recommendation links

3.2 Implementation

Figure 5 shows the system's structure. The system runs as a script on the user client and on the Web server, and is implemented with JavaScript and PHP.

When a user accesses a Web page, the JavaScript engine asks the Web server if a footprint exists, and shows it if that is the case. When a user marks a string in the Web browser, the JavaScript engine obtains the marked string and sends it to the Web server, which then takes the marked string and calculates the recommendation links according to certain algorithms. Following calculation, the Web server sends recommendation links and the JavaScript engine shows the recommendation window to the user. If the user selects a recommended link, the JavaScript engine sends it and the Web server saves the following information related to the selected link as a text file in the metadata storage.

- Date
- User ID
- Marked string
- Position of the marked string on the Web page

- URL of the selected recommended link
- Selected recommendation algorithm

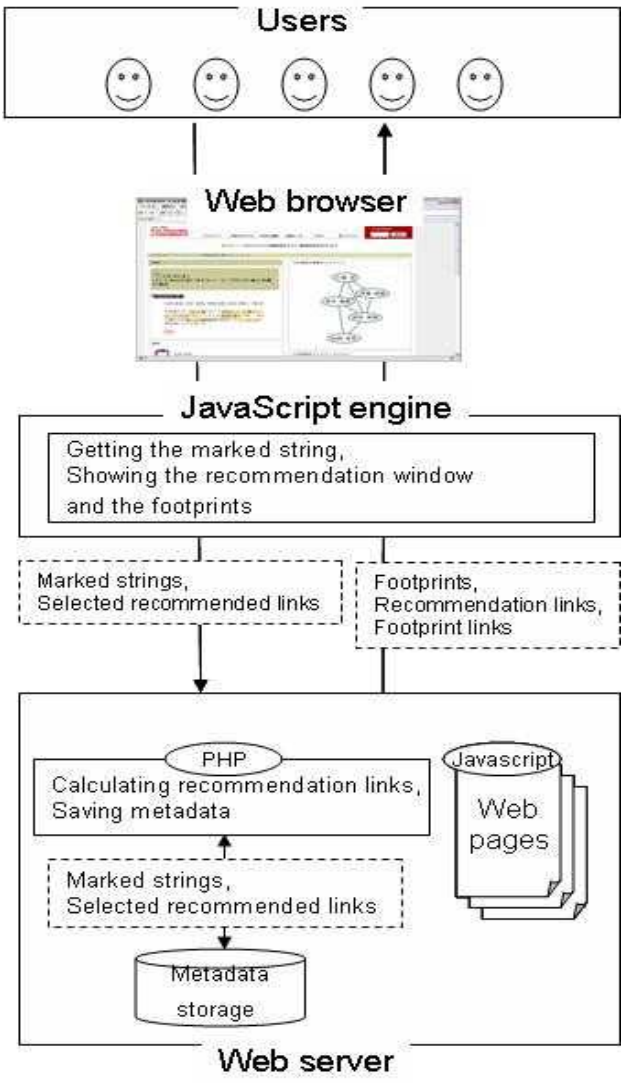

Figure 5. System structure

When a user places his or her mouse cursor over a footprint, the JavaScript engine takes the string of the footprint and sends it to the Web server. The Web server gets the footprint's string, calculates recommendation links, and obtains footprint links from the metadata storage. The Web server then sends both footprint links and recommendation links. Next, the JavaScript engine again displays the recommendation window to the user and if the user selects a footprint link or a recommended link, the JavaScript engine sends it and the Web server saves information about it.

3.3 Recommendation algorithms

When a user marks a string or places the mouse cursor over a footprint on a Web page, the system shows recommendation links using four types of algorithms that explore Web pages to find useful pages based on the string or footprints. To make recommendations, we use the system in a closed environment where the available Web pages are fixed[1].

The algorithms are as follows:

A) Page similarity using TFIDF
B) Collaborative filtering using the number of footprints on a Web page
C) Word matching between a marked string and footprint's strings
D) Word matching between a marked string and the Web page's strings.

In algorithm A, the system employs calculated page similarity based on TFIDF [4]. We calculate the TFIDF value of words on Web pages and cosine similarity using them. If a user marks a certain string on a Web page, the system recommends high-similarity pages to it regardless of the marked string and footprints.

In algorithm B, the system recommends pages with collaborative filtering [5], using the number of footprints on a Web page as the users' evaluation of that page. We adopt not incoming links but outgoing links, because selected recommendation links are not valuable for users. Therefore, the system recommends Web pages with a high predictive value using the number of footprints.

In algorithm C, if a word in the marked string matches one in the footprint's string on other pages, the system recommends the matched page.

In algorithm D, if a word in the marked string matches one in another page's string, the system recommends the matched page.

Table 1 shows a comparison of recommendation algorithms with respect to whether they use a marked string and footprints.

Table 1. Comparison of recommendation algorithms

Recommendation algorithm	A marked string	Footprints
A	×	×
B	×	○
C	○	○
D	○	×

Algorithm A makes recommendations based on page similarity between Web pages without using a marked string or footprints. This is assumed to be ordinary recommendation. Algorithm B, meanwhile, uses the number of footprints on a Web page to look for neighborhood users, and algorithms C and D use a marked string as a query to search Web pages with matching words. The difference is the search target: a footprint's strings or a Web page's strings.

We investigated which algorithms are preferred by users. The system recommends up to two Web pages for each algorithm and shows them in random order. Algorithm priority was applied so that the same link was calculated in order of C, D, B, and A. The system did not inform users about these algorithms. When users clicked a recommended page link, we determined that users preferred its recommendation algorithm.

[1] Aikuchi is not limited to closed environments, but we restricted the system to test recommendations in the following test cases.

4. ANALYSIS

We offered Aikuchi as a part of a support system at the Japanese Society for Artificial Intelligence 2006, which was held from June 7th to 9th, 2006. The support system was operated as a Web system, and every conference participant could access it using a user ID and password. The target Web pages comprised 276 pages that included authors and abstracts for papers. We prepared recommendation pages using algorithm A. When users marked a string, the system took a few seconds to display recommendation links. After the completing the experiment, we obtained 324 footprints and 172 links from them. Analysis of the results revealed that there were 45 users who marked strings one or more times, 28 users who jumped from marked strings to other pages, 88 users who placed their mouse cursors over footprints one or several times, and 33 users who jumped from footprints to other pages. In this section, we describe which algorithms were preferred by users.

4.1 Preference of the algorithms

Figure 6 shows a comparison of the number of selected recommendation algorithms before and during the conference. Before the conference, algorithm A was most commonly selected and the number of selections of B and C were few since the footprints were few at the beginning the conference. The number of selections for D was fewer than that of A before the conference. This means users selected a recommended link based on similarity without relation to marked strings or footprints. During the conference, the number of selections of C and D were more than before it and the number of selections of B was fewer than before it, since if Aikuchi recommended the same page links over time, then users would be tired of selecting the links recommended by B. The number of selections for algorithm A was also lower during the conference. This means users selected the recommendation algorithms based on the word in the marked string. Therefore, we found that users prefer recommendations based on words rather than similarity over time. As the number of selections for algorithm C increased, we also found that footprints are effective for recommending links.

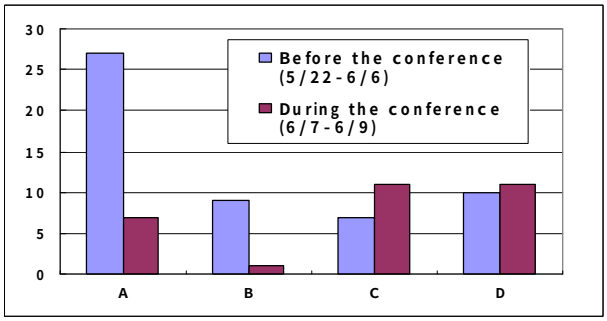

Figure 6. Number of selected recommendation algorithms, when users marked a string on a Web page.

4.2 Footprints

When users place their mouse cursor over a footprint, they can obtain not only recommendations by the algorithms described previously but also jumped pages. We investigated which algorithms are preferred by users in such cases (Fig. 7). The algorithm Z denotes the selection of jumped pages by users, and the figure shows that Z was selected in an overwhelming number of cases. Therefore, we found that users prefer to select links by using footprints.

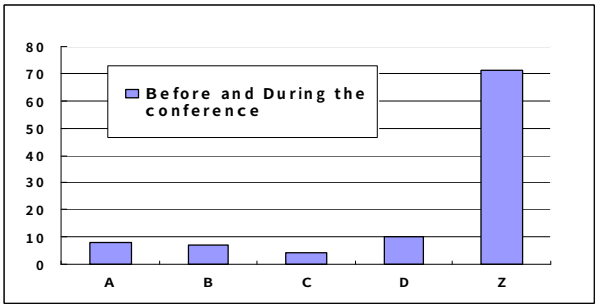

Figure 7. Number of selected recommendation algorithms, when users placed their mouse cursor over a footprint.

5. CONCLUSION

We proposed to use strings marked by users on Web pages as tags, and developed a system called Aikuchi with which users can mark strings and share them as footprints. Aikuchi recommends links based on a variety of algorithms when a user marks a string on a Web page.

Based on the analysis of user logs, we found that users preferred recommendations based on words in marked strings rather than page similarity. We consider that footprints are useful for users to select links because links based on footprints are frequently selected by users. Consequently, we believe that strings marked by users on Web pages can be used as user-generated metadata for searching pages.

6. ACKNOWLEDGMENTS

This research was supported by the National Institute of Information and Communications Technology.

7. REFERENCES

[1] L. Carr, S. Bechhofer, C. Goble, and W. Hall. Conceptual Linking: Ontology-based Open Hypermedia. In WWW10, Tenth World Wide Web Conference, May 2001.

[2] L. Carr and D. de Roure, W. Hall, G. Hill, The distributed link service: a tool for publishers, authors and readers, World Wide Web J. 1 (1) (1995) 647–656.

[3] M. Dzbor, J. B. Domingue, and E. Motta.: Magpie - towards a Semantic Web browser, In Proc. of the 2nd Intl. Semantic Web Conf., October 2003, Florida US

[4] G. Salton: Developments in automatic text retrieval, Science, Vol. 253, pp. 974-980 (1991).

[5] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl, "GroupLens: An Open Architecture for Collaborative Filtering of Netnews," In *Proceedings of the 1994 Computer Supported Collaborative Work Conference*, pp. 175–186, 1994.