

Generation and Expansion of Web Navigation from Browsing Log

Matsuo, Teppei

Graduate School of Information Science and Electrical Engineering, Kyushu University

Hirokawa, Sachio

Computing and Communications Center

<https://hdl.handle.net/2324/2982>

出版情報 : Proceedings of International Symposium on Information Science and Electrical Engineering. 2003, pp.605-608, 2003-11

バージョン :

権利関係 :

Generation and Expansion of Web Navigation from Browsing Log

Tepei Matsuo*, Sachio Hirokawa**

**Graduate School of Information Science and Electrical Engineering, Kyushu University*

matsuo@matu.cc.kyushu-u.ac.jp

***Computing and Communications Center, Kyushu University*

hirokawa@cc.kyushu-u.ac.jp

ABSTRACT

The WWW (World Wide Web) is a treasure of knowledge. In order to utilize such material on the WWW effectively, we propose a method to generate a navigation script from a user's browsing log. The script generated with this method is used to guide a user through the web pages that were recorded in the browsing log. The script language is defined with XML and has sequential, selective and parallel controls. The control structure is extracted from the browsing history using the timestamps and the link information of each URL. We demonstrate an example of WWW tour generated by this system.

1. INTRODUCTION

The rapid spread of the Internet has made WWW (World Wide Web) the biggest medium for people. Containing huge amount of information, it now can be called a treasure of knowledge. Web navigation is one of the most important subjects to support web user's search and understanding of information that exists enormously in WWW. Web navigation is important even in individual's web pages. For example, those who have their own web pages usually have their link collections to recommend for visitors. The link collection is sorted out based on individual's point of view. It is a kind of "static" web navigation by individuals. We aim to utilize materials on WWW not statically but "dynamically", so we proposed a script language which can be used to achieve dynamic navigation of web pages and developed an interpreter of the language. One can describe his or her scenario by browsing the pages along the way that he or she intends.

However, constructing a meaningful course is not easy. We focused on a browsing log since it can be a good material for generating a course. We developed a system that generates a navigation course effectively from the log [1]. Once a navigation course is generated from a user's browsing log, another user can follow the course and can share the experience and knowledge of the author of the course. As empirical evaluation, we applied the system for a browsing log of searching information of AIDS. We obtained a navigation course of about 100 nodes.

The structure of the rest of the paper is as follows: in Section 2 we cover related work. We discuss our algorithm in Section 3, and our navigation system in Section 4. We introduce a new function of the system in Section 5, and an application in Section 6. Future work and conclusion is described in Section 7.

2. RELATED WORK

There are some other related works for "dynamic" web navigation. Web Watcher [2], Letizia [3] and Adaptive Web Sites [4] provide navigation of specific web sites, but they do not treat general web pages like ours. Additionally, our navigation system uses a log that is made by a certain client, not the one of a specific web site.

Kazama et.al.[5] proposed a method to build up a graph to show HTML document correlation by analyzing user's log. However, the graph cannot be modified dynamically.

Ariadne [6] uses a Java application to navigate, but it does not control the time like ours navigating automatically with some advanced settings. Ariadne's navigation can step backward and forward, but it is done only by user's work. It does not provide "parallel navigation" that our system does, either. WebOFDAV [7] is a system that visualizes linking information, and the graph is modified by the changes of user's interest. The graph grows with new nodes that imply user's new interest, and the user can see the web surroundings. The navigation of this system is based on user's decision, which is different from ours in a point that our system controls from navigator's side.

Hasegawa et.al. [8] proposed navigation for education resources. The system decides the order of priority of the next resources to provide to the user. The decision is conducted with analyzing the browsing log. The user deepens understanding as the stage goes. Thus, this system has many similarities to ours.

Also, Kayama's system [9] decides the way of navigation with analyzing the browsing log. It maps the resources based on their properties additionally, and shows the map constantly. The map helps the user to see where he or she is among the whole resources, so this is also similar to our system.

3. GENERATION OF NAVIGATION COURSE FROM BROWSING LOG

A user browses a web page through the proxy server. All the requests that the browser makes are documented in a log file in the proxy server. The log file consists of the timestamp, IP address, URL and so on. They are all sorted out in chronological order. A program in the proxy server automatically generates a navigation script from the log file. Other users can visit the same web pages according to the navigation script.

A detailed explanation of our algorithm is shown in [1,10]. The point of the algorithm is to sort out nodes in chronological order trying to keep the linking information as much as possible, and to unify two trees into one with an optional arrow called “semantic link”.

Imagine that a user browsed web pages a, b, c, d, e, f, g, h, i, c, b, h and i. Each alphabet corresponds to a URL. Figure 1 shows a directed graph that draws a node as a web page, and an arrow as a link. This graph is called a web graph. The numbers over the nodes are the order of browsing.

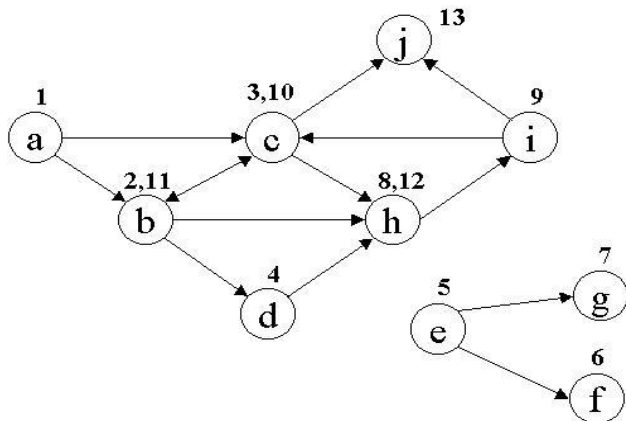


Figure 1. Web Graph

Our algorithm makes a navigation course shown in Figure 2 from Figure 1. Firstly, it generates a spanning tree from the web graph sorting nodes in chronological order. In this case, two spanning trees having node A and E on each roots are made. Secondly, it unifies these trees into one graph by connecting D and E. This edge does not correspond to hyperlink, and is called a “semantic link”. Nodes that browsed several times such as node B, C and H are distinguished. For example, B appears twice at time 2 and 11 in Figure 2. Continuous-lined arrows show hyperlinks, and arrows with broken-lined shows semantic links. A web page that corresponds to a node is displayed on the browser. A new window comes out to allow the user to select a route when it comes to a ramification point.

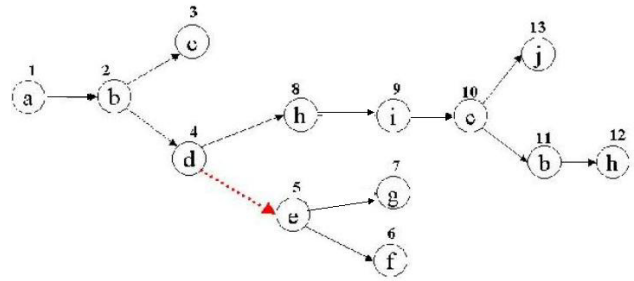


Figure 2. Navigation Course

4. NAVIGATION SYSTEM

The navigation system interprets a navigation script that tells how to navigate (Figure 3).

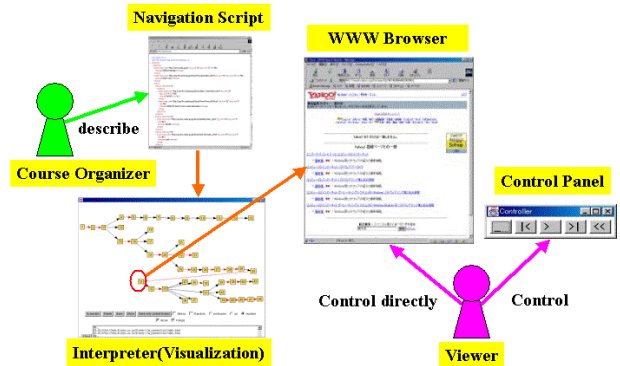


Figure 3. Navigation System

The script contains the information to navigate WWW, and is written in XML (eXtensible Markup Language) which enables editing and modifying of the system easier. The basic function of the navigation system is to navigate web pages in specified order. Besides, the navigation course occasionally has some routes so the user can select. Furthermore, the system does parallel navigation, which means that it shows a web page through the browser at the same time as it launches multimedia files such as sound and video. More details of the navigation script and how our system works are explained in [1].

5. EXPANSION OF NAVIGATION COURSE

A course generated from a browsing log is not a trivial one in the sense of size and meaning. It can be a rough sketch but is not an ideal navigation course unless the author elaborates on completing the scenario. It often happened that there was a node, i.e., an interesting web page, next to the end node of a course. If we knew the node and the node had been added, the navigation course would have been more valued. Also, we found the case that the course would be more interesting if there were

another ramification point. Therefore, we implemented a function to expand the course interactively according to user's request. In conducting expansion, the system works as follows. Firstly, a user chooses a node to expand by clicking the node on the navigation graph. Secondly, the system retrieves the HTML file of the node and extracts links from the HTML file. Thirdly, the system retrieves the HTML files of the URLs and adds nodes in the navigation graph. This function was implemented with Perl that provides a smooth interaction to treat the Internet data. The number of the nodes to add can be chosen by a user. It enables a user continue to add new nodes watching the web context in a browser. At last, a simple edit function was added to the system, which enabled to re-make and improve the navigation course.

6. WWW TOUR

The first author made rounds in WWW to get information about AIDS which is one of the biggest issues in the world. The navigation course made from the browsing log is shown in Figure 4. He kept doing the work for 84 minutes long continuously. The course consists of 90 nodes, and each node corresponds to one URL. Each node has the number which tells the order the web page was watched.

Firstly, it starts from Google (No.1), and the next (No.2) is the search result of AIDS. From the result page, it goes to the web pages of the Clinical Study Group for AIDS Drugs in Ministry of Health, Labor and Welfare¹ (from No.5 to No.31), and to the web pages of Life AIDS Project² (from No.32 to No.52), and to the web pages of Japanese Foundation for AIDS Prevention³ (from No.53 to No.73), and to the web pages of AIDS Pure Education⁴ (from No.74 to No.90). The reason why the top page of JFAP does not come from Google's search result is because a web page of LAP (No.50) has a link to JFAP top page.

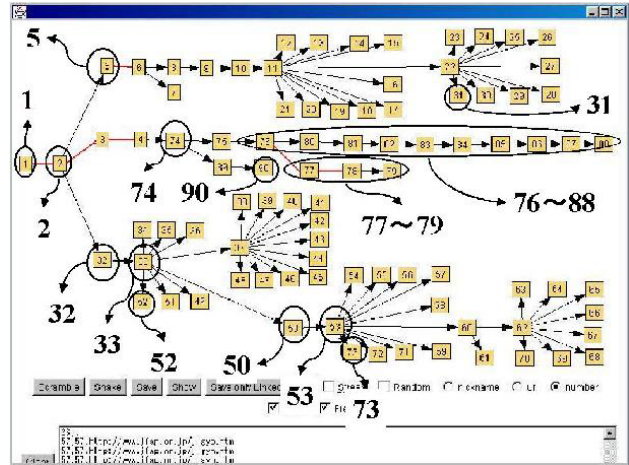


Figure 4. Rapid Prototype of Navigation Course

The following is the explanation of the figure. A set of the nodes from No.76 to No.88 is made by clicking the "Next" button continuously on a web page. The web page No.76 has a detailed explanation of what AIDS is, but it does not tell about the origin. Because of this, the author opened a new window (No.77) and searched with queries "AIDS" and "origin" (No.78), which led the author to the web page telling the origin of AIDS (No.79). Thus, the navigation course has a ramification point (No.76) that has a route explained above (from No.76 to No.88) and the one for those who would like to know about the origin.

That is one of the examples that a navigation course was made from the browsing log without labors such as "collecting web resources to construct a navigation course", "sorting the resources" and "constructing the navigation course". However, as a matter of fact, there were some unnecessary nodes in the course such as the web pages reached by mistake. Also, there were some nodes to be deleted to make the course easier to understand. Therefore, we modified some points in the course and made Figure 5. It consists of the Clinical Study Group for AIDS Drugs in Ministry of Health, Labor and Welfare (from No.5 to No.31), and AIDS Pure Education (from No.74 to No.90), and Life AIDS Project (from No.32 to No.52), and Japanese Foundation for AIDS Prevention (from No.53 to No.73), and the ramification point (No.2) that is connected to all the routes. The changes from Figure 4 to 5 are as follows. One is unifying duplicated nodes. Another is that we unified a ramification into one route if the branch nodes are very close in the sense of their contexts. Making Figure 5 was easy because Figure 4 already had main routes to construct the navigation course.

¹ <http://www.ijjnet.or.jp/aidsdrugmw>

² <http://www.lapjp.org>

³ <http://www.jfap.or.jp>

⁴ <http://www.aidscamp.com>

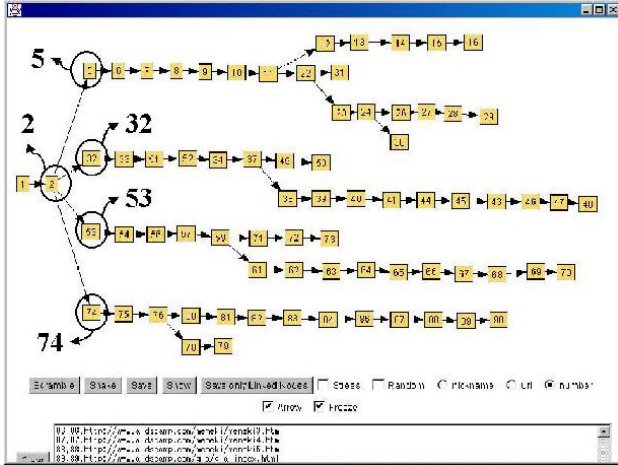


Figure 5. Completed Navigation Course

We have presented a WWW tour in this paragraph. Our system enabled making a tour that has about 100 nodes semi-automatically. Although it partly depends on the way of browsing, the nodes in the course made from the system already have its story in some way and are sorted out by the content. In other words, the order that a user looked around becomes a good source to make WWW tour effectively.

7. CONCLUSION

We have presented a navigation system which provides a WWW tour. Also, we have described a method to generate a tour automatically from a browsing log. It is an important issue how to share or express the knowledge that can be obtained from WWW. WWW provides a lot of information that people need. If they can share their knowledge easily, it would surely help each person's search action. Our navigation system realizes such help by navigating a user. From another point of view, it can be said that a user uses the system to sort out the information that was obtained through browsing. Also, the system can be applied to analysis of a user's action if the web pages are limited within one site.

To apply the system for a huge navigation course is on our future work. An editor will be necessary to make up such navigation course. By applying the concept of expanding, the navigation can be more dynamic. Evaluation of system's effectiveness is another future work.

8. REFERENCES

[1] K. Nishino, and S. Hirokawa, "Generation of Navigation Script from Log and Link", WebNet 2001, 534-539, 2001

[2] T. Joachims, D. Freitag, and T. Mitchell, "Web-Watcher: A Tour Guide for the World Wide Web", 6th IJCAI, 770-775, 1997.

[3] H. Lieberman, "Letizia: An Agent That Assists Web Browsing", Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence (IJCAI '95), 1995.

[4] M. Perkwitz, O. Etzioni, "Adaptive Web sites", CACM 43(8), 152-158, 2000.

[5] K. Kazama, S. Sato, S. Shimizu, T. Kambayashi, "HTML Document Correlation Analysis by User's Behavior in World Wide Web Navigation", Transactions of Information Proceedings Society of Japan, Vol.40, No.5, 2450-2459, 1999.

[6] J. John, A.T. Jensen, K. Geronbak, "Ariadne: a Java-based guided tour system for the World Wide Web", WWW7, 1998.

[7] M.L. Huang, P. Eades, R.F. Cohen, "WebOFDAV: navigating and visualizing the Web on-line with animated context swapping", WWW7, 1998.

[8] S. Hasegawa, A. Kashihara, and J. Toyoda, "Reorganizing Learning Resources on WWW and Its Application to an Adaptive Recommendation for Navigation Support", The transactions of the Institute of Electronics, information and Communication Engineers, D-I-Vol. J83-D-I, No.6, 671-681, 2000.

[9] M. Kayama, and T. Okamoto, "A Navigation System Based on Self Organizing Feature Map for Exploratory Learning in Hyperspace", The transactions of the Institute of Electronics, information and Communication Engineers, D-I-Vol. J83-D-I, No.6, 561-568, 2000.

[10] K. Nishino, and S. Hirokawa, "Rapid Prototyping of WWW Tour from Browsing History and Link", ICCE/ShoolNet2001, 1545-1548, 2001.