

Article

An Analysis of Internet Surfers' Recognition Processes and its Patterns

—An Experimental Study—

Kazuko Hozumi

ABSTRACT

In this paper, we examined how internet surfers reach to the information they needed, by analyzing the internet log information in a web homepage of a university. We obtained data of their recognition processes and behavior in internet.

keywords

information driven creativity, access log statistics, recognition process, World Wide Web

1. Introduction

In these days, the spread of the use of internet is increased surprisingly. In Japan, the surfing pattern has been changing compared with last year[1]. Internet surfers' surfing pattern is changed from *an excursion pattern* to a *visiting pattern*. It means that the surfers (from here, we call surfers users) would like to visit a web site which has some useful information with some information, which is introduced in newspapers and magazines as interesting internet sites. This will be acceptable because a user might not waste their precious time to visit boring web sites.

When a user has some purposes to surf in web sites, *visiting with surfing pattern* is understandable. If we used search engines of webs to find information with a keyword in a browser, the result of the search would be only concerned with the keyword. The information which helps getting hints of ideas will be a different type of search such as matching keyword data.

As we stated in the paper[2], in the area of the *information driven creativity*, the use of communication tools, especially using internet is very efficient to find out hints of new ideas, creating new products, and new services in management. Even if we could get such kind of information from internet web sites, we should pay much effort to surf in the great information flood, because information might have been presented with structures which would be difficult to understand.

In the area of *information filtering system*, in this time, there is no practical system which could filter information with understanding meanings or contents of information[3]. The reason why this kind of system has not been developed is that it is not clear how we can decide the information which has meanings.

Practically, in getting information, we first find an adequate web site, and then surf web pages using links. This kind of search needs time, however we could find some new findings from the web. Our purpose of this paper is to get data of users' surfing behavior processes, and to utilize it to the information driven creativity.

2. Data analysis

We analyzed the log information of the web site, which was opened as one of colleges in a university's homepage, developed by author's seminar

students. The top menu of homepage has three kinds of information : the course information, the programs of study, and the programs of seminar. The precise structure of the homepage and characteristics of the homepage are shown in Fig. 1 and Table 1 respectively.

Table 1. The contents of top menu and estimated users' behavior

top menu	contents	Could users know the information from the menu?
the course information	courses opened in the college	yes
the programs of study	curriculums of classes and educational materials	yes
the programs of seminar	seminars open to the course and seminars' products	no

2.1 Contents of data analysis

We analyzed the log information about the following items.

1. the tendency of total access about what kind of information a user wants.
2. the tendency of access counts, in the case when a user could not know the contents of information from the menu.
3. the number of access counts which a user should navigate the link when a user could not know the contents of information from the menu.
4. the tendency of access counts, in the case when a user could know the contents of information from the menu.
5. the tendency of access counts and time and difference between inside the university and outside the university, in the case when a user could know the contents of information from the menu.
6. what kind of information a user takes, in the case when a user could know the contents of information from the menu.

- This home page was developed by the junior college students who belong to author's seminar in the period from April 1995 to January 1996. It was opened to public on Jan. 27, 1996.
- The address of URL is <http://commerce.b17.kanagawa-u.ac.jp>
- Educational materials, which were developed in the period from April 1996 to June 1996 by author, are linked to the menu of *the programs of study*.
- The log data was collected for six months, in the term from February 1 to July 31 1996.
- Developed environments: NeXT workstation machine, NeXTstep operating system, OmniWeb browser, HTML language.
- Total file size : the course information : 44 k bytes
the programs of study : 196 k bytes
the programs of seminar : 25,497 k bytes
- In analyzing the log information, we use UNIX command, Microsoft Excel, and Excel Statistics by SRI cooperation¹.

2.3 Outline of results

Outline of the data analysis of the log information is as follows:

- the number of total access : 33,353
(from February 1 to July 31 , 1996)

¹ We normalized the log information like that: deleting all debugging accesses and graphical data such as buttons and icons. We did not use log analysis software like Webstat, which we can download from a web page as free software, because this kind of software counts all of the accesses. There are many discussions about audience rating of access counts. It was said that seeing only the counting accesses has problems. An audience rating should be count like that : how long same users see the page, how many times the top pages have read (access count divide file numbers) , so on [4].

- statistics by monthly
 February : 4,814 March : 4,165 April : 8,423 May : 4,217
 June : 6,203 July : 5,531
- statistics by domain
 academic domain : 76%
 the other organization : 24%
- statistics by country
 Japan : 92% USA : 8 % the other : 0 %

2. 4 Results of data analysis

2. 4. 1 the tendency of total access of the web pages

How does a user behave when he/she refers a homepage? He might go out only when he looked at the top menu of the homepage, or he might navigate all the pages when he liked them. If he could understand what information in it from the top menu, and he found out the pages is useless for him, he will stop navigating the links. If he could not estimate what information in the pages, he would navigate links until he could understand the contents of the pages.

We counted what information a user looked up in the pages. In this paper, the word “a user” means a server’s address of log information. In the top of the menu page, we have three links to the *course information*, the *programs of study*, and the *programs of seminar*. We counted accessed pages like that. If same users looked up the categories of three links over twice, we deleted the duplication of references. We counted the restricted requested pages and top menu page that had been accessed by outside users, it means that we omitted accesses of our university users (inside users). Table 2 shows the total tendency of reference of our home pages.

Table 2. Percentage of accessed pages

only the programs of seminar	41%
only the top home page	17%
only the programs of study	15%
all of the menus	8%
both the programs of study and seminar	7%
both the course information and the programs of seminar	7%
only the course information	4%

Over 40% of users got out only referring the *programs of seminar*. The total percentage of users who referred the seminar pages is 63. It seems that a user accesses unknown information.

2. 4. 2 the tendency of accesses in referring the seminar pages

How many links should a user navigate in getting information if it would be useful or not? Here, we divided users into four types: foreign countries, Japanese organizations, Japanese academic domains, and our university domains. Users who navigated more than two pages were counted, because it is hard to understand the contents of pages if only one page is referred. We analyzed the access counts of users, whose number is almost the number of people, visited this pages. We translated the data with natural logarithm transformation and calculated chi-square value using ANOVA statistics.

There are differences among these types of accesses, as F-value is 25.21 and P-value is 1.78E-15. The least significant differences (LSD) score is significant (1%) in all combination of four types, except for between "Foreign countries" and "Japanese org". Users who belong to those types are seemed to have the different reference tendency.

Table 3. The difference among 4 types of visitor

types of visitors	Number (%)	Mean	SD.
Foreign countries	94 (13.7%)	1.63	0.90
Japanese org.	186 (27.2%)	1.93	1.33
Japanese academics	109 (15.9%)	2.43	1.45
Our university	295 (43.1%)	2.79	1.43
No. of total visitor	684 (100%)		

About 25.8% of total visitors (922) stopped accessing after they had referenced only one page.

2. 4. 3 the tendency of access counts

In this section, we analyzed users' behavior of accessing pages in referring the seminar pages. Table 4 shows the number of access counts, using median. No academic society users are seemed to access only twice.

Table 4. The median of access counts in referring the seminar pages

type of visitors	Median
Foreign countries	2
Japanese org.	2
Japanese academics	4
Our university	8

When a user could not know what information in the pages, he referred the pages accessing only a few times. When he is interesting in the contents, he tried and tried navigating pages precisely. The percentage of users from foreign countries who referred over twenty times pages is only 1 %. However, some of our university domain's users got to all over our pages. In totality, half the users are seemed to stop accessing when they understand what information in the pages.

2. 4. 4 the tendency of access counts when a user knows it

In the *course information*, a user could understand what information in it. We analyzed if this type of access is different from that of unknown information as mentioned in 2. 4. 2. Table 5 shows similar analysis of the above. There is no difference among user types. Users are seemed to stop accessing those pages after twice. Behavior of users is seemed different according to information.

Table 5. The difference among 4 types of users

type of visitors	Number (%)	Mean	SD.	Median
Foreign countries	29 (18.7%)	0.83	1.01	2
Japanese org.	39 (25.2%)	0.88	0.60	2
Japanese academics	40 (25.8%)	0.83	0.66	2
Our university	47 (30.3%)	1.11	1.01	3
No. of total visitor	155 (100%)			

(Means and SD. were translated like mentioned in 2.4.2)

2. 4. 5 the tendency of access time and counts of the pages

In the *programs of study*, there are educational materials, which were developed for students to take the *information processing* class, and it could be referred from the educational environment of the university's homepage. We linked it to the *programs of study* page to make introduction to outside users. An outside user could not know such kind of information in it. Here, we analyzed access tendency between the outside users and students who took the class.

There are no significant differences between users, both access counts and access time. Moreover, the number of users who access only two pages is over 52% and 46%, for students and for outside users, respectively. The reason why only two pages were referred is that one lesson was built of

two pages. Both type of users looked up them and stopped reading when they could have understand the contents.

Table 6. The difference between students and outside users

items	outside users	students
the number of total access	281	896
the number of access over 2 pages	41	192
Mean of access counts	2.93	3.09
SD. of access counts	1.31	1.89
Mean of access time (second)	284.07	189.41
SD. of access time (second)	501.14	332.39

2. 4. 6 the tendency of access contents in referring the educational materials

A student should access pages twice, because the educational material consists of explanation pages and example of subject pages. Table 7 shows the reference tendency what information students and outside users read.

Table 7. Access counts of students and outside users

user	explanation	example	menu	total
students	448 (50%)	310 (35%)	138 (15%)	896
outside users	174 (61%)	60 (21%)	50 (18%)	284

Students referred explanation first, then read examples. Outside users referred the explanation only, or read examples only. We thought most of the outside users look the explanation first. Contrary to our expectation, the outside users also looked up examples that are to do subjects for students. In the case, the tendency of reference is different from the user types.

We had thought that the pages were not useful for the outside users, because there were no meanings if users did not take the course. Moreover,

the pages were opened in very short period, so there were very few chances to access the pages compared with the other home pages. However, one third of the users accessed the pages. It seemed the possibility of the remote education using Web pages.

2. 4. 7 the frequency of reference

A user wants to look up the pages again if the information is valuable. Our pages did not change except adding educational materials. The followings are the users' access count frequency when a same user visits again (In a strict sense, the number is not the number of people).

frequency count	1	2	3	4	5	6
percentage of user	83.0%	9.0%	5.0%	0.6%	1.0%	0.6%

3. Summary

The followings are some findings through our experimental study:

- 1 What kind of information a user wants to get? A user is apt to get information which is not clear from the menu and which could not be estimated.
- 2 We divided users into four types : foreign countries, Japanese organizations, Japanese academic domains, and our university domains. Under this circumstances, is there any difference among four types of users? There exist some differences of a visiting style among those types.
- 3 How does a user find out information he needed? When information is not clear from the menu, a user is apt to try to find out if it is useful or not by checking two web pages. Half of visitors stopped navigating when they found it is not useful. The other visitors navigated from

four to eight times when they were interested in the information.

4. Is there any different behavior among four types of users when information is clear from the menu? There is no difference among the user types. Users who need information navigated links until they had accomplished their purpose.
5. Is there any difference' between users who have some clear purpose to access pages and users who do not have? The former type of users are students who are taking the *information processing* course, and they should access the web page to do their homework or review lessons. There is no difference between the users. Students and outside users are the same about access time and access count of this search.
6. Is a user's information access path sequential? No, it seems that a user was not necessary to navigate sequentially.
7. Does a user visit to a domain again? No. It seems that if the information is not renewed, a user would not visit same domains again.

4. Conclusion

Internet surfing is apt to be said that it is only amusement. However, even in a small web site opened as a part of a university, there are so many visitors to find information they could be interested in. Visitors' category of this web site is not only academics but also companies and the other organizations. Moreover there are also visitors from foreign countries, as USA, Canada, Germany, Singapore, Brazil, New Zealand, Republic of Korea, Mauritania, and so on, though half of the seminar page are written in Japanese. We put on a mail address in the pages so that a user could present his or her opinions and requests to our pages. There are some

replies not only from Japan but from other countries. It seems that we could use WWW not only to send information outside the world, but to communicate with unknown people interactively.

In this study, we did not discuss the way of presenting educational materials and its effects. The educational material pages were developed to reduce teachers' load: to reduce time to pass out papers to students, to reduce time of frequently asked questions from students, and to be convenient for students to ask questions anytime, and so on. In this experience, we could agree to the problems in education using WWW, mentioned in Satoh[5], 1996. He mentioned in the article that education using WWW is apt to cause excess of information. Considering the problems, we should think about the way of offering educational materials with total educational environment.

In this study, we obtained many experimental data from the web log information. Users seem to behave according to characteristics of information, and there exists different behavior among user types. In order to users' convenience, the understandability of web pages is very important in the menus, especially for an *excursion pattern user*. If we could offer such pages, it will reduce a user's troubles and unnecessary traffic load. Though it is a matter of course, web pages should be always improved with fresh information. At the last, though this evaluation of data is obtained only from one junior college, we could not generalize users' behavior in this. However we could mention the tendency of internet surfers' behavior which was ambiguously thought. We will renew our pages and will continue analyzing log information from now on. As the future research, we will make a user's behavior model such that a user could find out ideas or hints to be emergent evolution or creative evolution in the area of information driven creativity.

This research article is revised presentation paper of "the 2nd International Conference on OA & Information Management-The Current Information Technology and Office Automation-," which was held on November 1996 at Makuhari, of Japan Society for the Study of Office Automation (JSSOA).

Acknowledgments

We thank Prof. Akira Tohyama of Chuo University for his advice and suggestions of this research. We also thank Prof. Tsutomu Nishikori of Yokohama College of Commerce Prof. Hiroyuki Nishimura of Toho Gakuen Junior College, and Prof. Atsuteru Kawabata of Junior College of Kanagawa University for their kind suggestions.

References

- [1] "Second internet active user investigation: Mite Asobu Jidai ha Owattuta," *Nikkei Multimedia*, July 1996, pp. 43-71.
- [2] Kazuko HOZUMI, "Information Driven Creativity Supported by Information Technology," *the Review of Economics and Commerce*, Society of Economics, Kanagawa University, Vol. 29, No. 3, pp. 79-115, 1994.
- [3] Masahiro MORITA, et al, "Information Filtering System : Prescriptions for Information Flood," *Journal of Information Processing*, Vol. 37, No. 8, 1996, pp. 751-758.
- [4] "Hit suu kyousou ha furui- internet shicho-ritsu chosa no subete", *INTERNET magazine* 1996/8, pp. 251-257.
- [5] Osamu SATOH, "Impact of Multimedia on Education-Toward Virtual Calssroom," *Journal of Office Automation*, Vol. 17, No. 1., May 1996, pp. 59-64.