Extracting User Behavior by Web Communities Technology on Global Web Logs

Shingo Otsuka†, Masashi Toyoda†, Jun Hirai‡, and Masaru Kitsuregawa†

†Institute of Industrial Science, The University of Tokyo
4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan
{otsuka,toyoda,kitsure}@tkl.iis.u-tokyo.ac.jp

‡Systems Integration Technology Center, Toshiba Solutions Corporation
3-22, Katamachi, Fuchu-Shi, Tokyo 183-8512, Japan
Hirai.Jun@toshiba-sol.co.jp

Abstract. A lot of work has been done on extracting the model of web user behavior. Most of them target server-side logs that cannot track user behavior outside of the server. Recently, a novel way has been developed to collect web browsing histories, using the same method for determining TV audience ratings; i.e., by collecting data from randomly selected users called panels. The logs collected from panels(called panel logs) cover an extremely broad URL-space, and it is difficult to capture the global behaviors of the users. Here we utilize mining results of web community to group those URLs into easily understandable topics. We also use search keywords in search engine sites because user behavior is deeply related to search keyword according to preliminary experiments on panel logs. We develop a prototype system to extract user access patterns from the panel logs and to capture the global behavior based on web communities.

1 Introduction

Web user behavior analysis is a very important research area, with a wide area of applications such as one-to-one marketing, and user behavior identification. Most research uses access logs on server-side(so called these server logs). On the other hand, a new kind of web business similar to the survey method on TV audience rating has emerged. It collects histories of URLs visited by users(called panels) who are randomly selected without statistical deviation. We call those URLs as *panel logs.*

The panel logs include *panel ID which is assigned to each panel, access time of web pages, reference second of web pages, URLs of accessed web page and so on.* Therefore we know that when or where did each panel access URLs. Moreover panel logs also include search keywords submitted to search engines. However, it is difficult to capture the user behavior based on URL-level analysis because panel logs cover an extremely broad URL-space.

Here we apply web community mining techniques[1] to give a better understanding of user global behavior. A web community is a collection of web pages created by individuals or any kind of associations that have a common interest on a specific topic[1]. We use the results of web community mining to map an

[1] In this paper, ‘community’ means ‘web community’

URL to an easy-to-understand topic. We also statistically analyze the importance of search keywords that appear in the panel sessions and their relation with the pages in the web communities. We propose a system to interactively support the analysis of global user behavior from panel logs and demonstrate the effectiveness of our proposed system.

The rest of the paper is organized as follows. Section 2 will review related work. In section 3 we will explain panel logs and web communities. Our system will be discussed in section 4. Section 5 will show example of using the system and discuss effectiveness of our system, while section 6 will give the conclusion.

2 Related works

- Extracting user behavior
  This field is a hot topic because it is directly connected to e-commerce business. [2, 3] discussed how to extract user behavior patterns. A method to cluster users with the same behavior patterns is presented in [4]. As for users' grouping, it is discussed in [5].
  These research focus on user behavior on a certain web server because these logs are limited only to web pages in a web server.
- Extracting relevance of web communities
  Most of the works adopt web page link analysis. Recently, some approaches which use access logs have been proposed. [6] proposed web pages clustering using access logs and [7] proposed extracting relevance of web communities from users' access patterns. [8] discussed usability of using OLAP for web access logs analysis.
- Analysis of search keywords
  Search engine vendors are doing analysis of the search keywords. Clustering of search keywords using Lycos's server logs were presented in [9]. [10] showed the result of clustering of Encarta's server logs which is an online encyclopedia provided by Microsoft. The research to improve search precision was discussed in [11]. These works analyze user behavior related to search keywords.
- Visualization of access logs
  To easily understand the results of access logs analysis, [12] proposed a method to visualize access logs. [13] discussed visualization of user behavior at an online yellowpage site.
- Others
  [14] has some similarities with our study. This paper discusses clustering users and web pages, and its purpose is to extract correlation between clustered users and web pages. This research also uses proxy logs which are client side logs. These logs record all URLs which are viewed by users and are similar to panel logs because it is easy to identify users using their private IP. Recently, they provide the logs in the internet open to the public, but few researches have been done using these logs. Some other works focus on different aspects of web log mining, such as indexing method for large web access logs[15], and analyzing web access logs in mobile environment[16].

As mentioned above, most of the researches focus on the analysis of user behavior in a web site. [14] uses proxy logs which are similar to our study.

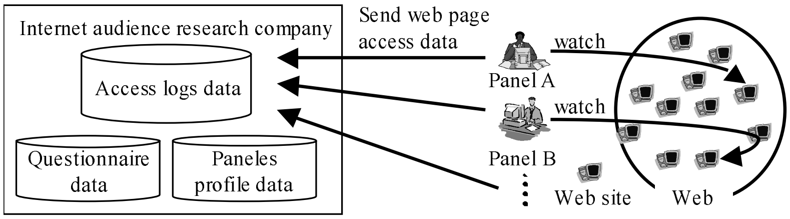


Fig. 1. Collection method of panel logs

Table 1. The details of the panel logs

An amount of data	about 10(Giga byte)
A term of collecting data	45(weeks)
A number of access	55,415,473(access)
A number of session	1,148,093(session)
A number of panels	about 10,000(persons)
A kind of search keyword	334,232(variety)

However, its purpose is to cluster web pages only, while our purpose is extended further to help understanding of the user behavior.

3 Panel logs and web communities

3.1 Panel logs

In our experiments, we use panel logs built by Video Research Interactive, Inc. that is one of internet rating companies. Panels are randomly selected based on RDD(Random Digit Dialing), and are requested to install a software that automatically reports web access log to the server of Video Research Interactive. Details of the data is shown in Figure 1 and Table 1.

We do not use the questionnaire data and the panels profile data in Figure 1 due to privacy reasons. The panel logs consist of *panel ID, access time of web pages, reference second of web pages, URLs of accessed web page and so on*, and the data size is 10GB and all panels are in Japanese. Panel ID is a unique ID which is assigned to each panel, and it is specific to an individual panel. Notice that panel logs also include search keywords submitted to search engines.

Usually, analysis of access logs uses the concept of *session* which is a sequence of web accesses. A session is defined as a set of URLs visited by a panel during a web browsing activity. We employed a well-known 30 minutes threshold for the maximum interval[17], such that two continuous accesses within 30 minutes interval are regarded as in a same session.

3.2 Web communities

We use the notion of web community to capture the global behavior of panels. A web community is a set of web pages created by individuals or any kind of associations that have a common interest on a specific topic. By examining how panels navigate through web communities, we can see more abstract behavior of panels than URL-level analysis, such as, what kinds of pages are visited by panels with the same interest.

In this paper, we define a web community as ‘a set of relating web page which are combined by hyperlinks’[18]. Most researches on web communities can be loosely classified into two methods. One method is extracting dense subgraphs[19] and the other is extracting complete bipartite graphs[20]. The former method determines the borderline between inside and outside of the web community using the theorem of ”Maximum Flow Minimum Cut” based on network theory. The latter method extracts complete bipartite graphs in the web snapshot as the hyperlinks between web pages which include common interest topics represented by complete bipartite graphs.

In our previous work, we created a web community chart[21] which based on the complete bipartite graphs, and extracted communities automatically from a large amount of web pages. We crawled 4.5 million web pages at February 2002 and automatically created 17 hundred thousand communities from one million selected pages. In this paper, we analyze panel logs using web communities extracted by our proposed technique. Moreover, though it is not very accurate, we automatically extract community labels which are word lists expressing the community contents, by analyzing anchor tags of the links to the pages belonging to the communities. Therefore, one can have a summary of communities without actually browsing them.

Since the time of the web page crawling for the web communities is in between the duration of panel logs collection, there are some web pages which are not covered by the crawling due to the change and deletion of pages which were accessed by the panels. Thus we define *matching factor* as follows to examine matching ratio between the URLs belonging to web-communities and the URLs included in panel logs.

$$matching\ factor = \frac{\begin{array}{c} the\ matching\ number\ of\ URLs\ belong\ to\ communities \\ and\ included\ in\ panel\ logs \end{array}}{the\ number\ of\ URLs\ included\ in\ panel\ logs}$$

We measured the *matching factor* and the result was only about 19%. We enhanced the matching factor by softening the matching condition using follow processes.

- Deleting directory or file part when the URLs included in panel logs do not match the URLs belonging to web-communities[2].
- Deleting site(domain) part when the URLs deleted directory part do not match[3].

The result is shown in Table 2. If we delete directory(file) part in URLs, the matching factor increases about 40% and when we delete ‘subdomain part’, the matching factor improves further about 8%. By modifying URLs, about

[2] When ‘http://xxx.yyy.com/aaa/bbb/ccc.html’ does not match, delete ‘ccc.html’ and check the remaining ‘http://xxx.yyy.com/aaa/bbb/’. This process is repeated on the URL includes directory or file part.

[3] When ‘http://xxx.yyy.com/’ does not match, delete ‘xxx’ and check the remaining ‘http://yyy.com/’. Note we do not check top/second domain name like ‘.com’, ‘co.jp’ and so on.

Table 2. The matching factor between the URLs belonging to web-communities and the URLs included in panel logs

no modification	18.8%
matching when deleting directory(file) part	36.3%
matching when deleting site part	7.7%
no matching	37.2%

65% of the URLs included in panel logs are covered by the URLs in the web communities.

3.3 Other results of preliminary experiments

According to preliminary experiments on panel logs, user behavior is deeply related to search engine sites and search keywords. We omit the details due to the space limitation. The analysis system also focuses on search keywords as well as the mapping with web communities.

The preliminary experiments also reveal that many panel logs include ”Yahoo! shopping”, ”Yahoo! auctions” and ”Rakuten [4] ”. We can easily infer the contents of these sites without the labeling from web communities. So we define ”Yahoo! shopping” and ”Rakuten” as ”Shopping sites”, ”Yahoo! auctions” and ”Rakuten auctions” as ”Auction sites” and ”Search engine or portal sites” as ”the search engine group”.

4 Panel logs analysis system

Our proposed system does not only provide both analysis features based on the community mapping and search keywords, but also supports the synergy between them. Since the search keywords represent the user purposes of web browsing, we can infer the reason behind a user visit to a certain community from the relation between the search keywords and the web community. Even when the search keyword is absent, we can understand some details of the page visit from the relation between pages in the community. Since the results from conventional URL analysis is hard to digest, the relation between web communities will help us to figure out the global user behavior.

The web communities in our system are identified by serial numbers (called *community IDs*). The system has a function FindIDByURL() to find community ID by analyst specified URL, as well as following functions :

– *ListIDByKeyword() : A function to show a list of web communities visited by users following some search keywords*
  If we input some search keywords, we can see the list of communities which are visited using these keywords.

– *ListKeywordByID() : A function to show lists of search keywords used for visiting the communities*
  If any community ID specified, the list of search keywords used for visiting the communities are showed.

[4] Rakuten is the most popular shopping site in Japan.

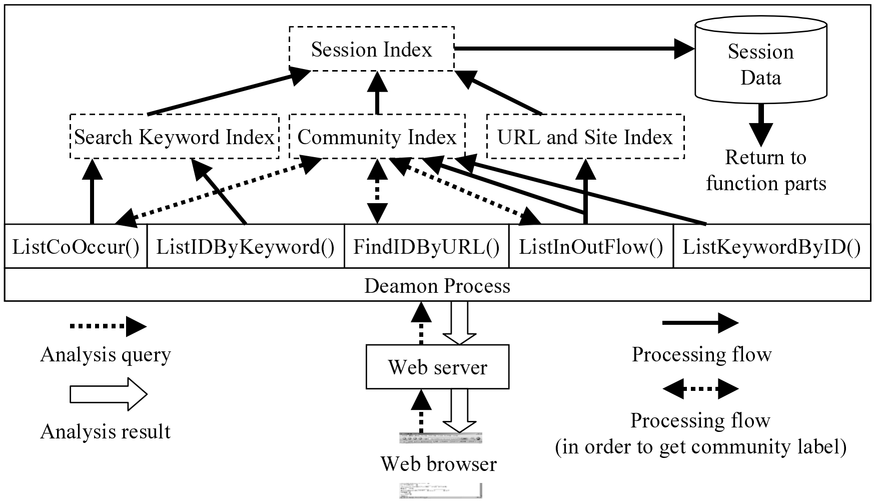
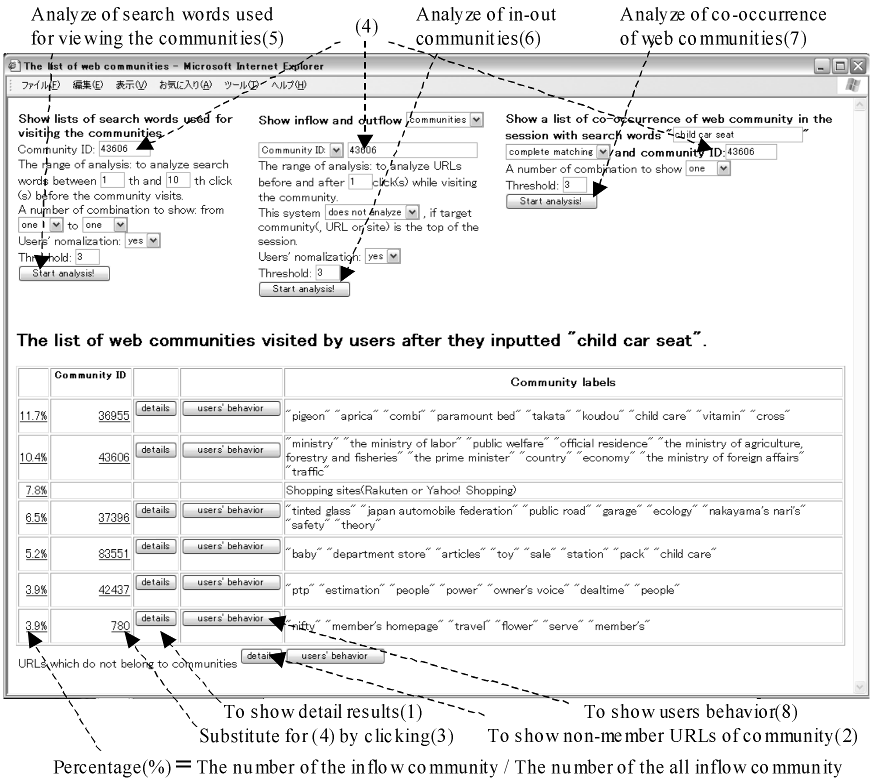


Fig. 2. Architecture of the system

- *ListInOutFlow() : A function to show inflow and outflow communities*
  If any community ID, URL or site name specified then the lists of communities visited before and after the specified community are shown.
- *ListCoOccur() : A function to show co-occurrence of communities*
  If we input some search keywords and specify any community ID then the lists of communities in the sessions which include the specified search keywords and community ID.
  This function allows analysts to specify parameters such as co-occurrence number N_{CO} and it can find communities which were visited together with at least other N_{CO} communities in the sessions when users input specified keywords and visit specified community.

Our system also supports *continuous analysis*. We can start new analysis immediately using the current results whenever we find them interesting, Therefore, it is easy to analyze user behavior based on a new perspective of search keywords and visits to the communities.

The architecture of the system is shown in Figure 2. In our system, the panel logs are maintained in session units (labeled as session data) and stored in secondary device. We used a session index to access the session data. The system has indices for communities, search keywords, URLs and site names to find appropriate sessions. Each index holds session IDs, which can be used to access session data through session index. The community index also contains community labels. It is used to search community ID from an URL or to obtain a community label that corresponds to a community ID.


Analyze of search words used for viewing the communities(5)
(4)
Analyze of in-out communities(6)
Analyze of co-occurrence of web communities(7)
The list of web communities - Microsoft Internet Explorer
ファイル(F) 編集(E) 表示(V) お気に入り(A) ツール(T) ヘルプ(H)
To show detail results(1)
To show users behavior(8)
Substitute for (4) by clicking(3)
To show non-member URLs of community(2)
Percentage(%) = The number of the inflow community / The number of the all inflow community

Show lists of search words used for visiting the communities
Community ID: 43606
The range of analysis: to analyze search words between 1 th and 10 th click (s) before the community visits.
A number of combination to show: from one to one
Users' nomalization: yes
Threshold: 3
Start analysis!

Show inflow and outflow communities
Community ID: 43606
The range of analysis: to analyze URLs before and after 1 click(s) while visiting the community.
This system does not analyze , if target community(, URL or site) is the top of the session.
Users' nomalization: yes
Threshold: 3
Start analysis!

Show a list of co-occurrence of web community in the session with search words "child car seat"
complete matching and community ID: 43606
A number of combination to show one
Threshold: 3
Start analysis!

The list of web communities visited by users after they inputted "child car seat".

	Community ID			Community labels
11.7%	36955	details	users' behavior	"pigeon" "aprica" "combi" "paramount bed" "takata" "koudou" "child care" "vitamin" "cross"
10.4%	43606	details	users' behavior	"ministry" "the ministry of labor" "public welfare" "official residence" "the ministry of agriculture, forestry and fisheries" "the prime minister" "country" "economy" "the ministry of foreign affairs" "traffic"
7.8%				Shopping sites(Rakuten or Yahoo! Shopping)
6.5%	37396	details	users' behavior	"tinted glass" "japan automobile federation" "public road" "garage" "ecology" "nakayama's nari's" "safety" "theory"
5.2%	83551	details	users' behavior	"baby" "department store" "articles" "toy" "sale" "station" "pack" "child care"
3.9%	42437	details	users' behavior	"ptp" "estimation" "people" "power" "owner's voice" "dealtime" "people"
3.9%	780	details	users' behavior	"nifty" "member's homepage" "travel" "flower" "serve" "member's"

URLs which do not belong to communities details users' behavior

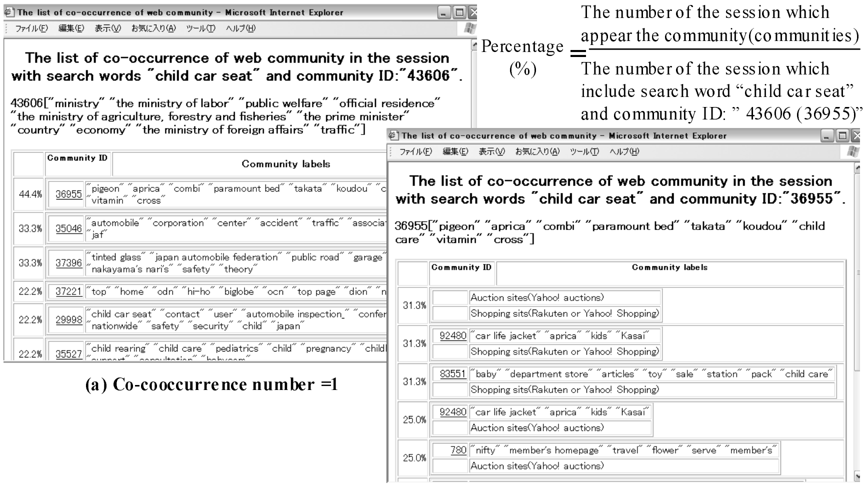
Fig. 3. A list of web communities visited by users following search keyword 'child car seat'

5 Example of extracting global user behavior

Here we describe some examples of the results obtained by our system to show its effectiveness. In this paper, we used panel logs which are collected from Japanese people. Therefore, all results have been translated from Japanese vocabulary items. We can specify search keywords, community IDs, URLs or site names to begin the analysis. Further queries can be performed based on the previous results.

5.1 Examples of analysis results using our system

Figure 3 shows a list of web communities visited by users looking for 'child car seat'(by function *ListIDByKeyword()*). The results are sorted according to order of frequency. Using community labels, we infer that community ID of 36955 relates to child car seat vendors. Similarly, we also suppose that community ID of 43606 relates to administrative agencies. We can see a more detailed results on the community by pushing button (1) in Figure 3. As we mentioned in section


The list of co-occurrence of web community - Microsoft Internet Explorer
ファイル(F) 編集(E) 表示(V) お気に入り(A) ツール(T) ヘルプ(H)

The list of co-occurrence of web community in the session with search words "child car seat" and community ID:"43606".

43606["ministry" "the ministry of labor" "public welfare" "official residence" "the ministry of agriculture, forestry and fisheries" "the prime minister" "country" "economy" "the ministry of foreign affairs" "traffic"]

	Community ID	Community labels
44.4%	36955	"pigeon" "aprica" "combi" "paramount bed" "takata" "koudou" "c "vitamin" "cross"
33.3%	35046	"automobile" "corporation" "center" "accident" "traffic" "associa "jaf"
33.3%	37396	"tinted glass" "japan automobile federation" "public road" "garage" "nakayama's nari's" "safety" "theory"
22.2%	37221	"top" "home" "odn" "hi-ho" "biglobe" "ocn" "top page" "dion" "n
22.2%	29998	"child car seat" "contact" "user" "automobile inspection_" "confer "nationwide" "safety" "security" "child" "japan"
22.2%	35527	"child rearing" "child care" "pediatrics" "child" "pregnancy" "child

(a) Co-cooccurrence number =1

Percentage (%) = The number of the session which appear the community(communities) / The number of the session which include search word "child car seat" and community ID: " 43606 (36955)"

The list of co-occurrence of web community - Microsoft Internet Explorer
ファイル(F) 編集(E) 表示(V) お気に入り(A) ツール(T) ヘルプ(H)

The list of co-occurrence of web community in the session with search words "child car seat" and community ID:"36955".

36955["pigeon" "aprica" "combi" "paramount bed" "takata" "koudou" "child care" "vitamin" "cross"]

	Community ID	Community labels
31.3%	 	Auction sites(Yahoo! auctions) Shopping sits(Rakuten or Yahoo! Shopping)
31.3%	92480	"car life jacket" "aprica" "kids" "Kasai" Shopping sits(Rakuten or Yahoo! Shopping)
31.3%	83551	"baby" "department store" "articles" "toy" "sale" "station" "pack" "child care" Shopping sits(Rakuten or Yahoo! Shopping)
25.0%	92480	"car life jacket" "aprica" "kids" "Kasai" Auction sites(Yahoo! auctions)
25.0%	780	"nifty" "member's homepage" "travel" "flower" "serve" "member's" Auction sites(Yahoo! auctions)

(b) Co-cooccurrence number =2

Fig. 4. Lists of co-occurrence of web communities following search keywords 'child car seat' and visiting community 'child car seat vendors' or 'administrative organs'

3.2, there are 35% of URLs which do not belong to any communities in panel logs. One can also see these URLs by pushing button (2) in Figure 3.

When an analyst is interested in the co-occurrence of web communities in a result, the analyst can analyze further from that point. Figure 4 shows a result of pushing button (7) in Figure 3. Figure 4(a) represents the co-occurrence of web communities in the sessions which include search keywords "child car seat" and administrative agencies' community(community ID is 43606). The result indicates that there is a high possibility that users of those sessions also visit child car seat vendors' community or automobile association's community. Further study to each community reveals that community ID "35046" includes only "National Consumer Affairs Center" and community ID "37396" includes only "Japan Automobile federation".

In Figure 4(b), we show the result when the co-occurrence number is two. The result indicates frequent patterns, which represent visits to three communities in the same session because the analyst has already specified community ID 36955(which relates to child car seat vendors). The result shows that those users have a great potential for visiting auction sites *or* shopping sites in the same session. Note that it is easy to understand that community ID "83551" relates to shopping sites and community ID "92480" relates to vendors by using the community labels.

Analysis with the proposed system can clearly depict the global user behavior. For example, the user behavior with search keywords *child car seat* can be classified into user access patterns as shown in Figure 5. Notice that a significant access pattern emerges because the child car seat is used only for a short period. After the child grows up, the child car seat is no longer needed. Thus many owners put the unused seats for sale on the auction sites. The access pat-

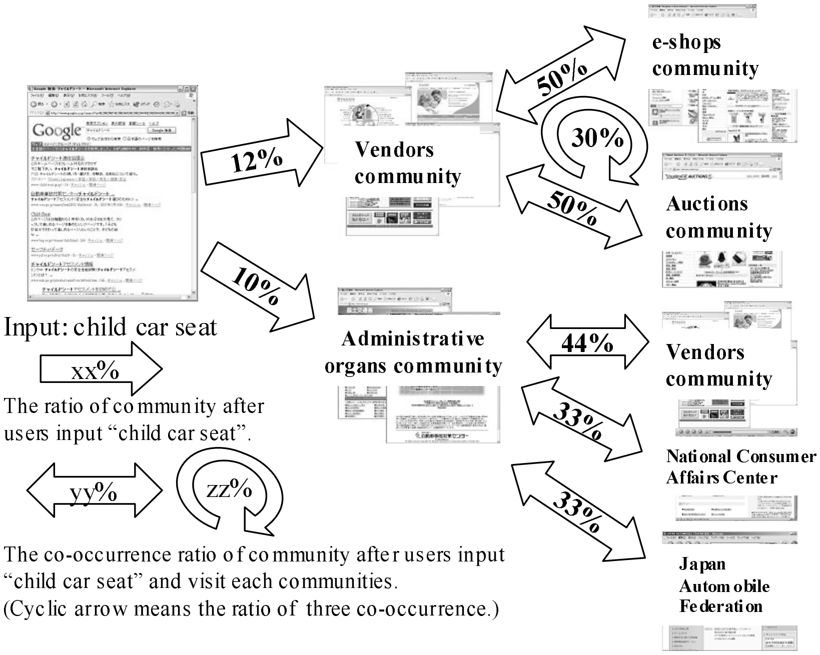

e-shops community
50%
30%
12%
Vendors community
50%
Auctions community
10%
Input: child car seat
Administrative organs community
44%
Vendors community
xx%
The ratio of community after users input "child car seat".
33%
National Consumer Affairs Center
yy%
zz%
33%
Japan Automobile Federation
The co-occurrence ratio of community after users input "child car seat" and visit each communities.
(Cyclic arrow means the ratio of three co-occurrence.)

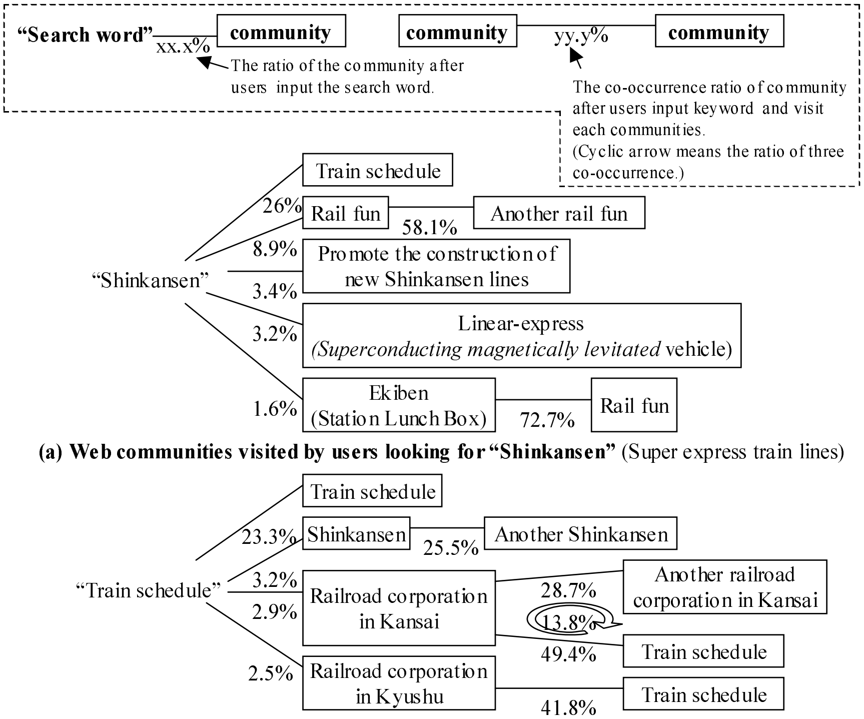
Fig. 5. The user behavior for search keyword 'child car seat'

tern shows that many users are trying to find secondhand goods at auction sites while simultaneously visiting the child car seat vendors and e-shops to compare the performance and the price. On the other hand, the aim of users which visit the community concerned with administrative organs is to acquire knowledge of the method of installation, safety standards and so on. These users also tend to visit the communities of child car seat vendors, non-profit organizations and other administrative organs. Most of the these users do not go to the auctions community.

5.2 Users' behavior on other topics

We show users' behavior found using our system in Figure 6. Figure 6(a) indicates users' behavior accompanied search keywords "Shinkansen" (the high speed bullet train). We found interesting transitions between communities such as "Promote the construction of new Shinkansen lines", "Rail fun" and "Ekiben" as well as ordinary results like "Train schedule".

Figure 6(b) shows another example for search keywords "Train schedule". Most of the users visit "Train schedule communities" but we also found some behaviors related to certain regions as the results of co-occurrence community analysis.


"Search word"
xx.x%
community
community
yy.y%
community
The ratio of the community after users input the search word.
The co-occurrence ratio of community after users input keyword and visit each communities.
(Cyclic arrow means the ratio of three co-occurrence.)
"Shinkansen"
Train schedule
26%
Rail fun
58.1%
Another rail fun
8.9%
Promote the construction of new Shinkansen lines
3.4%
3.2%
Linear-express (Superconducting magnetically levitated vehicle)
1.6%
Ekiben (Station Lunch Box)
72.7%
Rail fun
"Train schedule"
Train schedule
23.3%
Shinkansen
25.5%
Another Shinkansen
3.2%
2.9%
Railroad corporation in Kansai
28.7%
Another railroad corporation in Kansai
13.8%
49.4%
Train schedule
2.5%
Railroad corporation in Kyushu
41.8%
Train schedule

(a) Web communities visited by users looking for "Shinkansen" (Super express train lines)

(b) Web communities visited by users looking for "Train schedule"

Fig. 6. The other examples of users' behavior

6 Conclusion

It is difficult to grasp the user behavior from panel logs because these kind of logs cover an extremely broad URL-space. We proposed some methods to analyze panel logs using web community mapping and also showed the importance of search keywords from panel logs preliminary analysis.

We implemented the methods in a system to analyze user global behavior from panel logs. The system facilitates the mapping with web communities as well as the synergy with the search keyword analysis. We have confirmed the effectiveness of the system. The system can discover the reasons behind some web browsing patterns as well as the relation between them.

acknowledgment

This research was partially supported by Grant-in-Aid for Scientific Research on Priority Areas(c)(No. 13224014) funded by the Japanese Ministry of Education, Culture, Sports, Science and Technology(MEXT). We also wish to thank Video Research Interactive, Inc. for providing the panel logs.

References

[1] Kleinberg, J.: Authoritative sources in a hyperlinked environment. In Proceedings of the ACM-SIAM Symposium on Discrete Algorithms (1998)

[2] Shahabi, C., Zarkesh, A., Adibi, J., Shah, V.: Knowledge discovery from users web-page navigation. In Proceedings of the IEEE RIDE97 Workshop (1997)

[3] Batista, P., Silva, M.: Mining on-line newspaper web access logs. 12th International Meeting of the Euro Working Group on Decision Support Systems (EWG-DSS 2001) (2001)

[4] Fu, Y., Sandhu, K., Shih, M.: Clustering of web users based on access patterns. In Proceedings of the 1999 KDD Workshop on Web Mining(WEBKDD'99) (1999)

[5] Ungar, L., Foster, D.: Clustering methods for collaborative filtering. AAAI Workshop on Recommendation Systems (1998)

[6] Su, Z., Yang, Q., Zhang, H., Xu, X., Hu, Y.: Correlation-based document clustering using web logs. 34th Hawaii International Conference on System Sciences (HICSS-34) (2001)

[7] Tan, P., Kumar, V.: Mining association patterns in web usage data. International Conference on Advances in Infrastructure for e-Business, e-Education, e-Science, and e-Medicine on the Internet (2002)

[8] Zaiane, O., Xin, M., Han, J.: Discovering web access patterns and trends by applying olap and data mining technology on web logs. in Proc. Advances in Digital Libraries (ADL'98) (1998)

[9] Beeferman, D., Berger, A.: Agglomerative clustering of s earch engine query log. The 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD2000) (2000)

[10] Wen, J., Nie, J., Zhang, H.: Query clustering using user logs. ACM Transactions on Information Systems (ACM TOIS) **20** (2002) 59–81

[11] Ohura, Y., Takahashi, K., Pramudiono, I., Kitsuregawa, M.: Experiments on query expansion for internet yellow page services using web log mining. The 28th International Conference on Very Large Data Bases (VLDB2002) (2002)

[12] Koutsoupias, N.: Exploring web access logs with correspondence analysis. Methods and Applications of Artificial Intelligence, Second Hellenic (2002)

[13] Prasetyo, B., Pramudiono, I., Takahashi, K., Kitsuregawa, M.: Naviz: Website navigational behavior visualizer. Advances in Knowledge Discovery and Data Mining 6th Pacific-Asia Conference (PAKDD2002) (2002)

[14] Zeng, H., Chen, Z., Ma, W.: A unified framework for clustering heterogeneous web objects. The Third International Conference on Web Information Systems Engineering (WISE2002) (2002)

[15] Nanopoulos, A., Manolopoulos, Y., Zakrzewicz, M., Morzy, T.: Indexing web access-logs for pattern queries. 4th ACM CIKM Nternational Workshop on Web Information and Data Management (WIDM2002) (2002) 63–68

[16] Pramudiono, I., Shintani, T., Takahashi, K., Kitsuregawa, M.: User behavior analysis of location aware search engine. Proceedings of International Conference On Mobile Data Management (MDM'02) (2002) 139–145

[17] Catledge, L., Pitkow, J.: Characterizing browsing behaviors on the world-wide web. Computer Networks and ISDN Systems (1995)

[18] Murata, T.: Web community. IPSJ Magazine **44** (2003) 702–706

[19] Flake, G., Lawrence, S., Giles, C.L., Coetzee, F.: Self-organization and identification of web communities. IEEE Computer **35** (2002) 66–71

[20] Kumar, R., Raghavan, P., Rajagopalan, S., Tomkins, A.: Trawling the web for emerging cyber-communities. Proc. of the 8th WWW conference (1999) 403–416

[21] Toyoda, M., Kitsuregawa, M.: Creating a web community chart for navigating related communities. Conference Proceedings of Hypertext 2001 (2001) 103–112