

Study on Web Analytics Utilizing User Environment Segmentation in “Business to Business” site.

Akiyuki Sekiguchi

Graduate School of Systems and Information Engineering,
University of Tsukuba
Tokyo, Japan
sekichan2008@gmail.com

Kazuhiko Tsuda

Graduate School of Business Science,
University of Tsukuba
Tokyo, Japan
tsuda@gssm.otsuka.tsukuba.ac.jp

Abstract— In this study, we surveyed correlation of website access by user environment. Correlations can help us understand differences in user behavior that vary by time by place, or by user environment. We also found the user behavior tracking like visit numbers or page dwell time categorized by user segmentation is effective. For example, mobile device usage rate is higher in non-working hours than working hours and viewed pages are different. We confirmed user environment segments with a correlation approach can be used for web analytics for user navigation studies or even marketing use.

Keywords— web metrics; web analytics; B to B website

I. INTRODUCTION

Business-to-Business (B to B) manufacturer websites have changed in role and responsibility since use of the Internet has become widespread even among hardware engineers. The purpose of visiting a website has changed from searching for technical documents into searching for solutions or products without any face-to-face contact. In B to B manufacturer industry, traditionally sales related activity or even marketing related activity was tied to face-to-face salespersons' activity. In this period, the website's role was for searching and for providing technical documents or technical software resources. With the growth of e-commerce, users have become hesitant to meet salespeople and at the same time the manufacturer wants to track user behavior on the web and utilize analytic data for marketing and improving sales figures.

According to “The End of Solution Sales,” Harvard Business Review, July-August 2012, a recent Corporate Executive Board study of more than 1,400 B to B customers found that those customers completed, on average, nearly 60% of a typical purchasing decision—researching solutions, ranking options, setting requirements, benchmarking pricing, and so on—before even having a conversation with a supplier.“ This means many B to B customers select products/solutions and even buy them without intervention from salespeople. This means the website is becoming more important than before even from a business point of view even though it was used only for information delivery in the past. That is why we need study on dedicated B to B web site analytics.

We have been studying web metrics methodology and actual user behaviors just dedicated to B to B manufacturer

websites. In our past survey, there were two purposes for Business to Business (B to B) web analytics: (1) Improve and optimize the site for users by path analysis; and (2) Use in marketing activities. Compared to B to C web analytics, B to B web analytics have the following three characteristics:

(a) In many cases, the buyer is not the same person as the web user. So it is important to analyze all the users from the same company or organization as a single unit. That can be stated as B to B to C web analytics, not simply B to B.

(b) The goal of visitors to the website is often not only to make a purchase. Main conversions can be downloading a file, making an e-mail subscription or inquiring online.

(c) It is rare for a user to complete their goal within a single session. In most cases, users require multiple sessions spread out over a long period of time to complete their goal.

In past studies, we came up with proposals for B to B website analytics and also we studied the effectiveness of web analytics by user segmentation and especially the importance of page dwell time. In this study, we try to examine web analytics by user environment and see the effectiveness from this point of view. User environment means user access hours (time of day), connection type, and device type, etc. This study's data is based on web access data for a Japanese website from 2014 January 1st to September 30th from a global B to B manufacturing company.

Section II presents previous work. Section III shows user segmentation and focus metrics. Section IV is a study on visit times and dwell time by hours and connection type. Section V is web analysis findings by segmentation by content directory. Section VI shows analytics by web connected devices. Section VII is about major findings and we have conclusion in Section VIII.

II. PREVIOUS STUDIES

In our previous studies, we came up with a web analytics scheme for B to B websites and we defined B to B site conversion types and the importance of user registration on web. This is the first study of B to B type conversion related [16]. In another study, we checked effectiveness of page dwell time as well as traditional metrics like page views, unique users, visits and conversion rate. [15].

We found the importance of registered versus unregistered user segmentation and confirmed user behavior is different in each segment [17]. There are many studies on web metrics for e-commerce behavior [2] and a study on

personalization metrics on web [11], and studies on web metrics related to the B to B market [8]. There have also been general web analytic studies. [1][3]-[7][9][10][12]-[14].

III. USER SEGMENTATION AND FOCUS METRICS IN THIS STUDY

We categorized potential segmentation as shown in Table I. In the previous study, we saw content category segmentation and some others. In this study, we focus on connection type and the time period in which users are accessing the manufacturer’s website. We wanted to see the difference in behavior by user environment (Time and Place).

TABLE I. USER SEGMENT EXAMPLES

Segmentation category	Examples and considerations
1. By content category	Viewers of product information versus viewers of investment relations (IR)/company information User seeking to download software versus e-commerce users
2. By user environment (by time slot or by connection type)	By time hours Midnight users versus business hour users By connection type like through providers or through company network.
3. By user referrer	Users arriving through search engine, by e-mail click, or by bookmark/URL typing
4. By visit frequency	First time versus second and more frequent users
5. By user commitment level (registered or unregistered)	Registered users versus unregistered users
6. By company profile	Focus customer versus unfocused . Large customers versus small customers
7. By industry	User behavior by industry
8. By participation	Only web tracking for converted customers or unconverted customers
9. By device type	User navigation can differ by device.

IV. VISIT TIMES AND DWELL TIME BY HOURS AND CONNECTION TYPE

As a quick reference, for a B to B web site there is much difference in traffic between weekdays and weekends. However, the trends of traffic by time of day show almost the same peaks across all days. Figure 1 shows general trend in web visit number by hours.

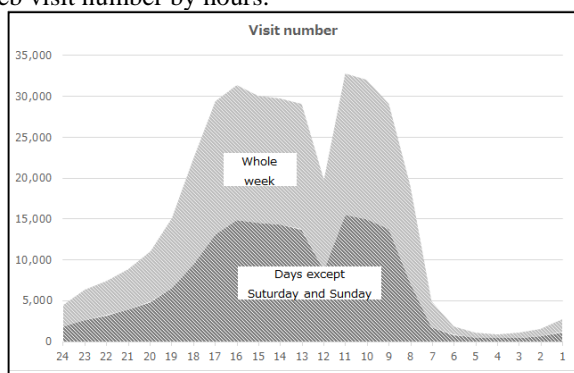


Figure 1. General trend in web visit number by hours

According to the statistics provided by the Japan Institute for Labor Policy and Training shown in [18], "hours of work per week, manufacturing" is 42.2 per week. Also, "9:00 AM to 7:00 PM" are typical working hours in Japan. We defined three time periods as "1. Home and commuting", "2. Work", and "3. Commuting and home". Firstly, we tracked user accesses by time period with consideration to company size. Normally small-sized companies or individual engineers tend to use normal internet providers and middle or large-scale companies use their own domains. We tracked them by time of day distinguishing the users who came through normal domains and users who came through internet providers (called "Providers"). Figure 2 shows a visit numbers trend.

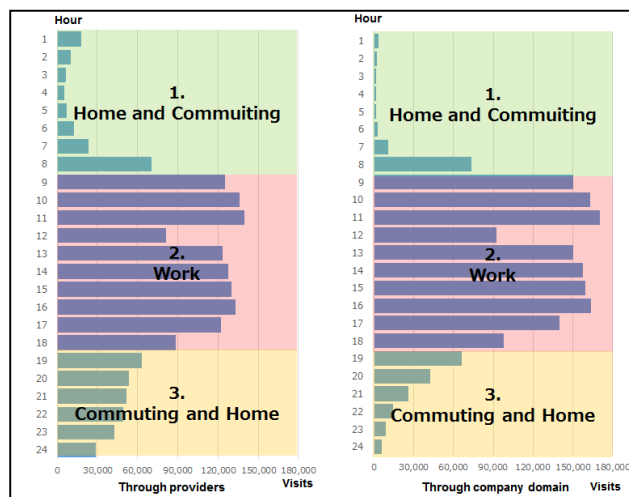


Figure 2. Visit numbers by time of day through providers and company domain

As a result the total coefficient of correlation of company and provider users is 0.59 and middle level of correlation (Similar trends) but if we omit the times from 1:00 AM to 9:00 AM, there is a strong correlation of 0.78. This means trends are similar throughout the day except for one time period. Only the time period between 1:00 AM and 9:00 AM shows some difference in numbers between "via providers" and "via company domain". To illustrate this in more detail, Figure 3 shows visit numbers between 1:00 AM and 9:00 AM by hour. From late night to morning we assume some engineers work at home or work at small companies which use connections through providers.

This kind of data is useful for deciding which content should be shown or targeted to customers accessing the website at each time of day. Also, it can be assumed that provider users are made up of not just customers but also a general audience who are looking for IR information, company information or even some news through the sites. In fact during this time period numbers for these contents are relatively higher than during business hours. The proportion of IR/press release visits are 33% higher compared to normal working hours.

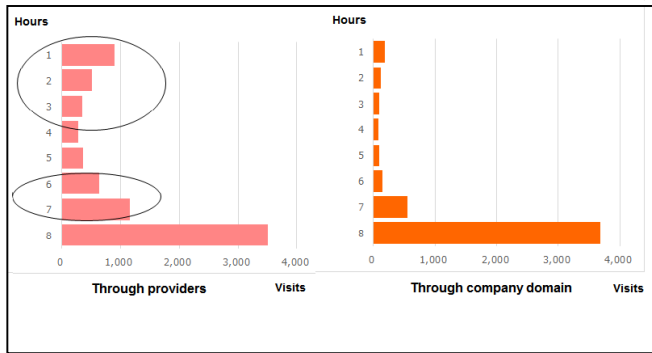


Figure 3. Visit numbers by time of day through providers and company domain from 1:00 AM to 9:00 AM

Next we looked at the page dwell time. Please refer to Figure 4. This is average of page dwell time by each hour for providers and company domain. The total coefficient of correlation is 0.68 for access via providers and via company domains. However, if we calculate the correlation for outside working hours, i.e. 7:00 PM to 1:00 AM, the correlation is 0.94. We can see some difference between the connection types for this hours. This shows the possibility of differences in usage between provider users and company users. The types of pages that are actually viewed are different. Generally, company users view purchasing information more and provider users view more press releases or IR information. We will investigate which information is accessed more in our next study.

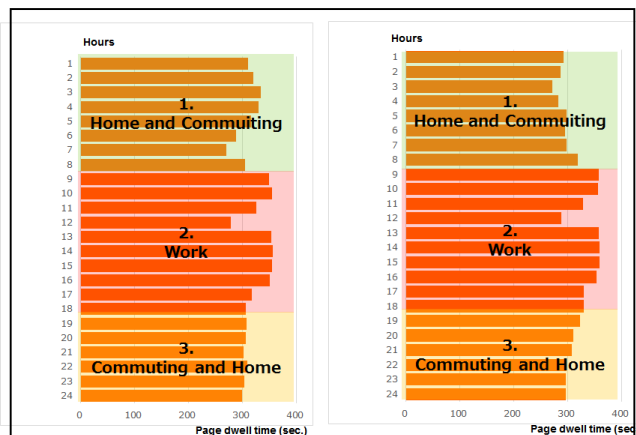


Figure 4. Page dwell time by time of day through providers and company domain

Please refer to Figure 5 for page dwell time by time of day by connection type. Page dwell time for small-sized customers who use providers peak at 10:00 PM and they probably work from home or on trains while commuting in Japan. This can be related to the fact that trains are the most common way of commuting in Japan. Also, for company

domain users this could indicate engineers working on development overnight.

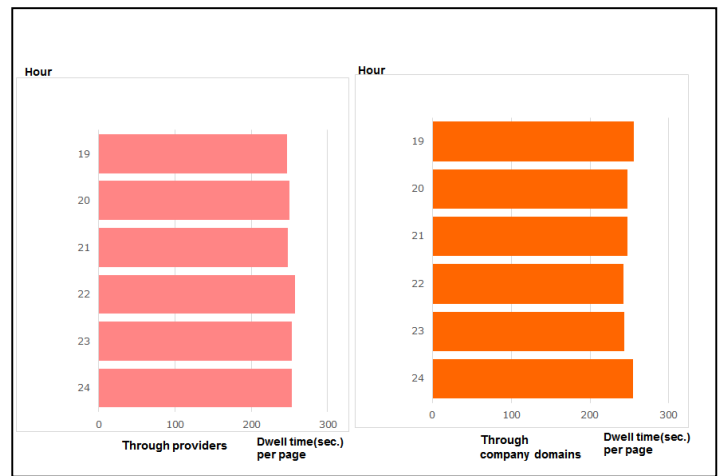


Figure 5. Page dwell time by time of day through providers and company domain from 7:00 PM to 1:00 AM

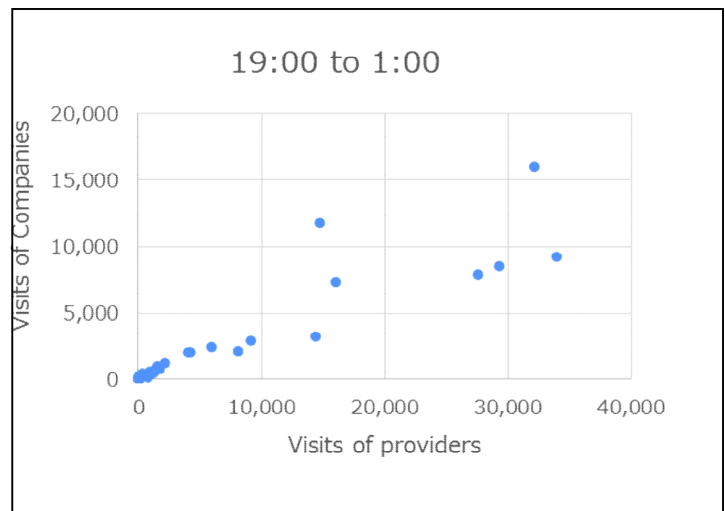


Figure 6. Correlation of visits from 19:00 to 1:00

Figure 6 shows correlation of visits from 19:00 to 1:00. Visit number is correlated between them and not su much difference.

V. ANALYSIS BY DIRECTORY

We surveyed the correlation between provider users and company users in terms of several segments. Firstly, we looked at correlation by content category (directory). There is a strong correlation for dwell time between providers and companies shown in TABLE II. TABLE II shows correlation coefficient in page dwell time for each hour and connection type. However, there is some different correlation just for some directories. Referring to TABLE III, the search function is one area of differentiation, especially in the 1:00

to 9:00 zone. Correlation here is lower than other periods. The number of searches performed by provider users is lower than those performed by company users. It is assumed that normally mobile access is through providers and these users are viewing websites during their train commute and do not search for any solutions or products but do view press releases or events during this time.

TABLE II. DWELL TIME CORRELATION TOTALLY

Directory Dwell Time		1:00 to 9:00		9:00-19:00		19:00to24:00	
		Providers	Companies	Providers	Companies	Providers	Companies
1:00 to 9:00	Providers	1.00					
	Company	0.84	1.00				
9:00-19:00	Providers	0.98	0.83	1.00			
	Company	0.88	0.92	0.88	1.00		
19:00to24:00	Providers	0.92	0.80	0.91	0.87	1.00	
	Company	0.91	0.84	0.90	0.86	0.91	1.00

Actually the company in this study tries to provide a different user interface to different customers depending on the time period. For 19:00 to 1:00, some navigation elements are changed with A/B testing, and the conversion rate for downloads is 125 times higher for time targeting. This type of analysis can be used for marketing purposes. For this purpose we will keep studying for further details.

VI. ANALYSIS BY DEVICE

We also looked at the relationship by device. We cannot see the actual device that a user owns but we can see information on OS (Operating System). There is much less correlation between providers and company trends in the time periods “1:00 AM to 9:00 AM” and “7:00 PM to 1:00 PM”. Most likely the mobile device usage rate is higher in non-working hours than working hours as shown in TABLE IV and TABLE V. Both table show correlation coefficient in page dwell time for each hour and connection type. Currently, unlike B to C sites the layout of most B to B sites is not mobile device compliant. However, B to B sites need to think about mobile device compliance especially for users who access through providers.

TABLE III. CORRELATION BY CONTENT DIRECTORY

Directory	1:00 to 9:00		9:00 to 18:00		19:00 to 1:00	
	Providers	Company	Providers	Company	Providers	Company
products	106,475	78,981	556,082	595,534	128,806	76,145
Search	21,377	30,217	167,138	193,837	35,397	22,751
support	26,514	19,350	120,251	127,775	32,221	15,962
press	20,321	11,382	89,733	77,086	34,035	9,186
comp	18,817	9,361	89,042	74,004	29,369	8,441
gur	16,049	16,175	84,245	115,158	14,808	11,761
edge_ol	15,972	6,116	64,111	52,709	27,620	7,830
applications	9,994	6,134	48,677	52,147	16,080	7,286
career	7,877	2,579	31,870	20,476	14,428	3,193
ir	6,894	3,747	25,322	21,894	9,203	2,854
disclaimers	4,852	3,736	17,557	17,190	4,247	2,021
company_info	4,164	1,663	18,134	14,629	8,118	2,064
event	4,131	3,638	18,706	22,214	4,125	1,997
partner	3,777	2,322	22,313	19,933	6,018	2,426
contact	2,238	1,821	12,022	11,829	2,163	1,257
public	1,602	1,758	9,243	12,234	1,627	960
buy	1,601	957	6,625	5,950	1,737	757
myrenesas	1,060	990	4,766	6,171	1,209	562
cmn	1,009	632	3,907	3,514	1,175	408
purposes	952	548	3,241	2,672	1,058	326
secret	906	908	6,164	7,962	973	587
videoclip	776	437	3,253	2,966	1,036	387
redirect	750	376	3,515	3,436	1,359	528
chat	640	439	3,580	3,678	613	338
Inquiry	557	495	3,023	3,145	436	371
search	459	564	2,396	3,644	406	433
user	436	133	1,981	829	833	133
edge	424	321	2,022	2,730	812	458
_print_this_page_	389	403	1,669	2,022	310	251
smart	311	171	1,399	1,485	453	173
devcon_jpn_2014	238	204	1,362	1,455	357	147
prod	147	135	959	1,192	196	167
ecology	131	69	461	613	198	103
media	128	182	681	1,413	153	271
facebook	103	41	548	277	187	46
sitemap	93	50	380	339	87	29
legal	88	75	383	461	104	46
tech	84	52	324	401	172	76
guidance	75	84	338	404	53	31
csr	63	20	231	157	115	25
privacy	44	95	212	270	57	18
campaign	34	24	151	137	55	14
lib	33	30	180	192	30	19
registration	32	25	116	139	18	15
rss	29	20	134	176	43	15
tool	29	23	116	220	24	22
C:	18	5	106	50	16	4
supp	15	8	40	86	30	8
r_video	13	9	80	102	30	6
manga	8	1	10	2	10	2

TABLE IV. CORRELATION BY OS TYPE

OS Type Dwell Time		1:00 to 9:00		9:00-19:00		19:00to1:00	
		Providers	Companies	Providers	Companies	Providers	Companies
1:00 to 9:00	Providers	1.00					
	Company	0.28	1.00				
9:00-19:00	Providers	0.97	0.27	1.00			
	Company	0.40	0.89	0.31	1.00		
19:00to1:00	Providers	0.91	0.33	0.91	0.39	1.00	
	Company	-0.03	-0.08	0.09	-0.15	-0.18	1.00

TABLE V. CORRELATION BY OS TYPE IN DETAILS

Item	1:00 to 9:00				9:00 to 19:00				19:00 to 1:00			
	Provider	%	Company	%	Provider	%	Company	%	ISP	%	Company	%
GNU/Linux	1,662	0.78%	632	0.43%	5,257	0.48%	4,152	0.35%	2,689	0.89%	1,121	0.77%
Microsoft Window:	162,070	75.83%	139,460	95.45%	947,855	85.98%	1,145,688	97.66%	211,999	69.86%	134,006	92.22%
Others	208	0.10%	19	0.01%	358	0.03%	91	0.01%	340	0.11%	46	0.03%
UNIX	38	0.02%	7	0.00%	104	0.01%	76	0.01%	47	0.02%	14	0.01%
Apple Macintosh	5,937	2.78%	1,710	1.17%	20,218	1.83%	9,893	0.84%	12,009	3.96%	3,028	2.08%
Unspecified	63	0.03%	16	0.01%	269	0.02%	96	0.01%	177	0.06%	34	0.02%
Google Android	19,312	9.04%	2,156	1.48%	55,271	5.01%	6,543	0.56%	33,457	11.03%	3,550	2.44%
Apple iOS	24,395	11.41%	2,091	1.43%	72,954	6.62%	6,504	0.55%	42,705	14.07%	3,493	2.40%
Microsoft Window:	24	0.01%	8	0.01%	52	0.00%	32	0.00%	18	0.01%	7	0.00%
Blackberry	7	0.00%	7	0.00%	20	0.00%	12	0.00%	14	0.00%	4	0.00%
Symbian	9	0.00%	3	0.00%	7	0.00%	0	0.00%	4	0.00%	1	0.00%
WebOS	0	0.00%	0	0.00%	1	0.00%	0	0.00%	0	0.00%	0	0.00%
Adobe	0	0.00%	0	0.00%	1	0.00%	1	0.00%	1	0.00%	0	0.00%

Other examples of no correlation are “viewed page numbers” and search usage time shown in TABLE VI and TABLE VII.

TABLE VI. VIEWED PAGE NUMBER CORRELATION

Viewed page numbers		1:00 to 9:00		9:00-19:00		19:00to1:00	
Dwell Time		Providers	Companies	Providers	Companies	Providers	Companies
1:00 to 9:00	Providers	1.00					
	Company	-0.23	1.00				
9:00-19:00	Providers	0.78	-0.33	1.00			
	Company	-0.30	0.48	-0.37	1.00		
19:00to1:00	Providers	0.68	-0.22	0.68	-0.30	1.00	
	Company	-0.17	0.34	-0.24	0.35	-0.14	1.00

TABLE VII. CORRELATION WITH SEARCH USAGE TIME

req (trend daily)		1:00 to 9:00		9:00-19:00		19:00to1:00	
Visit		Providers	Companies	Providers	Companies	Providers	Companies
1:00 to 9:00	Providers	1.00					
	Company	0.91	1.00				
9:00-19:00	Providers	0.70	0.87	1.00			
	Company	0.65	0.88	0.96	1.00		
19:00to 24:00	Providers	0.44	0.52	0.76	0.61	1.00	
	Company	0.54	0.79	0.92	0.94	0.70	1.00

There is much difference between time and connection type and there is possibility navigation can be further optimized according to time of day or connection type.

VII. SUMMARY OF FINDINGS

We found the user behavior tracking like visit numbers or page dwell time categorized by user segmentation is effective. Especially the accesses like time and place (or connection type) have different trend by each segments. For example the time period between 1:00 AM and 9:00 AM shows some level of difference in numbers between “via providers” and “via company domain”. Also, for page dwell

time 7:00 PM to 1:00 AM time period has differentiations between providers and companies. Also, depending on content type we found some difference. For example in the 1:00 to 9:00 zone user behavior is different and the number of searches performed by provider users is lower than those performed by company users. It is assumed that normally mobile access is through providers and these users are viewing websites during their train commute and do not search for any solutions or products but do view press releases or events during this time. We also looked at the relationship by device. We found that mobile device usage rate is higher in non-working hours than working hours and also viewed pages are different between them.

VIII. CONCLUSIONS

We have been trying to study the effectiveness of web analytics for a B to B manufacturer site with several studies. We defined some of the segment models and examined web access using some segments. In this study, we surveyed correlation of access by user environment. There are correlations between time of day or correlation between connection types such as connecting through a provider or through a company network. We used some key web metrics such as visits and page dwell time for our correlation survey. We noticed user environment segments with a correlation approach can be used for web analytics for user navigation studies or even marketing use. With the results of this study we will keep testing and doing analytics for targeted pages or targeted navigation by user environment as our next study.

REFERENCES

- [1] B. Adamson, M. Dixon, and N. Toman, “The End of Solution Sales Harvard Business Review” July-August 2012, pp102-106.
- [2] Y.Ichikawa, M Nakamura, Y.Kishimoto, and T.Kobayashi, “A Proposal of Extracting Innovative Users with Web Access Log of an E-Commerce site”, IPSJ SIG Notes 2012-GN-83(2); 2012, pp.1-8.

- [3] C. J. Aivalis, A. C. Boucouvalas, , Log File Analysis of E-commerce Systems in Rich Internet Web 2.0 Applications: This paper appears in: Informatics (PCI), 2011 15th Panhellenic Conference; 2011.p. 222-226.
- [4] J. Park, K. Jung, Y. Lee, G. Cho, J. Kim, and J. Koh, "The Continuous Service Usage Intention in the Web Analytics Services", System Sciences, 2009. HICSS '09. 42nd Hawaii International Conference on; 2009.pp. 1-7.
- [5] T. Ejiri, "Web Analytics and Web Marketing: Access Log Analytics Realized Web Marketing" Kaizen Cycle Management systems: a journal of Japan Industrial Management Association 18(1); 2008.pp. 8-43.
- [6] P. Sampath Dept. of Comput. Sci. & Eng., Bannari Amman Inst. of Technol., Sathyamangalam, India, "An efficient weighted rule mining for web logs using systolic tree", Advances in Engineering, Science and Management (ICAESM), 2012 International Conference on; 2012.pp. 432-436.
- [7] S. Otsuka, M.Toyoda, and M. Kitsuregawa, "A Study for Analysis of Web Access Logs with Web Community", Information Processing Society of Japan (IPSJ), Database44; 2003.pp. 432-436.
- [8] A. Phippen, L. Sheppard, S. Furnell, "A practical evaluation of Web analytics", Internet Research, Vol. 14 Iss: 4; 2004.pp. 284-293.
- [9] K. Rebecca, P. Justin and P. "Graeme, Ethical considerations and guidelines in web analytics and digital marketing", a retail case study, Proceedings of the 6th Australian Institute of Computer Ethics conference 2012, Australian Institute of Computer Ethics, Melbourne, Vic; 2012.pp. 5-12.
- [10] X. Wang, D. Shen, H. Chen, L. Wedman, "Applying web analytics" in a K-12 resource inventory, Electronic Library, The, Vol. 29 Iss: 1; 2011. pp. 20-35.
- [11] W. Xiao-Gang, "Web mining based on user access patterns for web personalization", Computing, Communication, Control, and Management, 2009. CCCM 2009. ISECS International Colloquium; 2009.pp. 20-35.
- [12] G. K.New Mexico Inst. of Min. & Technol., Socorro, NM, USA Colbaugh, R., "Web Analytics for Security Informatics" , Intelligence and Security Informatics Conference (EISIC); 2011.pp. 20-35.
- [13] Pascual-Cid, V.Web Res. Group, Univ. Pompeu Fabra & Fundacio Barcelona Media, Barcelona, "An information visualisation system for the understanding of web data", Visual Analytics Science and Technology, VAST '08. IEEE Symposium; 2008.pp. 183-184.
- [14] N. O., L. Soliman, M., Saka, E., Badia, A, Germain, R., "A Web Usage Mining Framework for Mining Evolving User Profiles in Dynamic Web Sites" , Knowledge and Data Engineering, IEEE Transactions; 2008, pp. 202-215.
- [15] A. Sekiguchi, K. Tsuda, "Consideration on page dwell time in B to B industry web analytics", ACIS; 2013. pp. 902-904.
- [16] A. Sekiguchi, T. Katsunuma, I. Hokao, Y. Yamada, K. Tsuda, "Web analytics for B to B marketing in semiconductor industry", International Journal of e-Education, e-Business, e-Management and e-Learning, Vol.2, No. 5; 2012.pp.413-418.
- [17] A. Sekiguchi and K. Tsuda, "Study on web analytics utilizing segmentation knowledge in business to business manufacturer site", KES 2014, pp.902-909.
- [18] "The Japan Institute for Labour Policy and Training", "Databook of International Labour Statistics", 2014, pp.1.