

# Discovering Student E-Learning Preferred Navigation Paths Using Selection Page and Time Preference Algorithm

<https://doi.org/10.3991/ijet.v12i10.7246>

K. Dharmarajan

Bharathiar University, Coimbatore, Tamil Nadu, India

Vels University, Chennai, Tamil Nadu, India.

dharmak07@gmail.com

M. A. Dorairangaswamy

St. Peter's University, Chennai, Tamil Nadu, India.

**Abstract**—In this paper, the student navigation paths and student or visitor interested page is identified. Student navigation interest pattern mining contains both the frequently navigation path based on webpage memory size and session length. Relatively comparing access proportion of viewing time and selective page size, preference can be used for mining student learning pattern instead of interested subject. In order to identify Preferred Navigation Paths, an efficient algorithm for Visitor Access Matrix (VAM) by the page to page transition probabilities statistics of all visitor behaviors is introduced in this paper. Second, we propose an efficient algorithm for Selection and Time Preference (SATP) to identify the preference of web pages by viewing time. Third, the user interested page would calculate by both memory size and session. In this way we proposed the Preference of page content size and session identifier algorithm. The performance of the proposed algorithms is evaluated and the algorithms can determine preferred navigation path efficiently. The experimental results show the accuracy and scalability of the algorithms. This approach may be helpful in E-learning, E-business, such as web personalization and website designer.

**Keywords**—Web usage mining, E-Learning, Web Mining, VAM, SATP, Preference of page content size and session identifier algorithm

## 1 Introduction

### 1.1 Web data mining

The Internet is very imported in innovative technology to store and share the information. Hence enormous growth of World Wide Web is maintained quality of website. Moreover, E-Learning and digital marketing are rapidly emerging and signif-

importance of forestalling the necessity of their website visitor. To identify the web user interests for improving the usability of E-Learning process [2].

The E-Learner navigation path is useful in practice, and the mined patterns can be used to predict and understand visitors' browsing behavior and user interest. It is useful in improving web user navigation pattern experience, and better recommendations and personalized services for their customer's efficiency.

Web data mining is more efficient technique in data mining and it is an innovative area where mining data is an important task and various methods has been proposed in order to solve the various issues related to the web mining in existing dataset[3]. The huge number of data is collected from web server in the form of web log file. This is a main source of information for understanding website user navigation behavior [1].

This paper mainly focus on identify the web user behavior and improve them to get the required information in quick manner and suitable form in less time. Every day approximately millions of web sites are hosted in the web server. With the growing of internet web based organization are keep to analyze the navigating behavior of the user interesting in their website

## 1.2 Process of Web Usage Mining

Web Usage Mining also consists of four-step process. The first step is data collection, the second step is data pre-processing, the third step is pattern discovery and the last step is pattern analysis.

The first step is **Data Collection** User web log dataset is gathering from E-Learning web portal of "www.collegenotes.org".Log files are created by web servers automatically. Web Log Files are files that contain information about website visitor activity. Each time a visitor requests any file (page, image, etc.) from the site, information of his request is appended to a current log file.

The second step is **pre-processing** phase contains four separate phases. (i) Data cleaning, (ii) User Identification, (iii) Session Identification, (iv) Path completion.

1. The data cleaning is removing the useless request from the log file. The process of minimizing data size for the data cleaning phase is to remove the jpeg, word, gif, sound, animation files.
2. The user identification is to identify users, IP addresses are used. Each user has unique IP address and each IP address represents one user. But, user identification is greatly complicated by the existence of local caches, corporate firewalls and proxy servers.
3. The Session Identification is identification of the user session is also important because it will largely affect the quality of pattern discovery result. A user session can be defined as a set of pages visited by the same user within the duration of one particular visit to a web-site. A set of pages visited by a specific user is considered as a single user session if the pages are requested at a time interval not more than a specified time period particular visit to a website [5].

4. Path completion phase is to discover user's travel pattern, the missing pages in the user access path should be appended after identifying sessions, while accessing the website, based on what user accesses, different path are formed for A path  $p = \{p_1, p_2, \dots, p_n\}$  where  $n$  is the number of pages traversed in a single session[4].

Purpose of **Pattern discovery** is to produce meaningful patterns from the data stored after cleaning and reforming the data. The key-matching techniques, statistics, artificial neural network, data mining and clustering algorithms. This proposed process identifies user's navigational sequential patterns. The emerging pattern discovery includes data mining methods like statistical analysis association rule, clustering, classification, sequential patterns, path analysis, and dependency modeling on preprocessed log data.

**Pattern analysis** is the offline component of web usage mining, which includes the conversion and understanding of the web log data mined pattern. It contains pattern knowledge discovery technique is used to identify the hidden in sequence or predictive emerging pattern and filtering, aggregation, characterization from web log files. It help to identify the interesting pages and uninteresting pages.

Therefore, it is necessary to propose a new concept that can replace the conventional concept of confidence to reflect user interest accurately under Web environment [1]. In this paper, we investigate the issues related to mining interesting navigation patterns efficiently and accurately from Web logs. The main contributions are as follows. First, the new concept of preference considering page viewing time and user selective intention is proposed. Second, the new algorithm SATP (Selection and Time preference) is introduced [1]. Finally, the user Navigation interested page algorithm used. Considering the structure of Website, User Navigation Interested Page Algorithm can get user access preferred path of the page–page transition probabilities statistics of all users 'behaviors.

## 2 Problem Formulation

The main aim of the research work is extracting of E-Learner interested access page using web log data, mostly time spent which is based on user session time, IP addresses, page memory size ,browser details, operating system, top user.

### 2.1 Navigation Path

The sequence of web pages visited by a user at a particular time is called web navigation path. It is represented as a graph, each page in the navigation path of the user is termed as an element and navigation path is termed as order. The user may visit  $n$  number of pages in a website but the frequently visited pages and amount of time spent in a web page by the user is called as user preferred page.

### 2.2 User Preferred Page

Let  $S$  be the set of URL of Web sites and  $P$  be the set of user navigation paths. Given  $p \subset P, \forall e \in p$  (each  $e$  is an element in the user navigation path  $\forall s \in S$ ), if all  $e$  have same  $n$  preceding elements and  $x$  different pages at order  $(n+1)$ , we call there are  $x$  different selection on  $n^{\text{th}}$  elements [1].

### 2.3 User page visiting

Multiple webpage: Multiple page views during a visit then the time calculation is based on the start time between the current page and the start time of the next page hit[.]. The user navigation Sequence and time represented in Fig 1. When there is a multiple page views during a visit then the time calculation is based on the start time between the current page and the start time of the next page hit and memory usage over time with Timeline recordings show in Table 1.

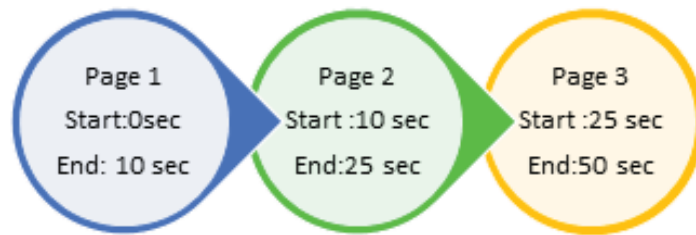


Fig. 1. User Page Navigation

Table 1. Web Page visitor time calculation

Total Time Spent (Sec)	Sequences	Average time Spent / Visit	No. of pages visited	Page Memory Size	Maximum Time spent page
50	1	$50/3 = 16.66$	3	13936 k	25 Sec

The user viewing the webpage amount of time spent and average calculation we follow the formulas.

$$A = ((S/(x-y))/86400)*1440 \tag{1}$$

A Average Time Spent

S Sum of Time Spent

$x$  visits

$y$  one page viewed

A web user visits the web page the average Time Spent is calculated in seconds. The web users have spent on your site across all visits. 86400 is represent the number of seconds in a day. 1440 is represent the number of minutes in a day.

The variety of time attribute into breaks according to user navigation interest. Interval labels can then be used to replace actual time values [1].

$$T = \left\{ \begin{array}{l} 1, 0 < t \leq T_{\text{Very Low}} \\ 2, T_{\text{Very Low}} < t \leq T_{\text{Low}} \\ 3, T_{\text{Low}} < t \leq \text{Average} \\ 4, T_{\text{Average}} < t \leq \text{High} \\ 5, T_{\text{Very High}} \leq t \end{array} \right\} \quad (2)$$

Time Calculation according to Visitor navigation interest according to Visitor navigation interest, time is partitioned as follows **Fig. 2**.

Visits duration		
Number of visits: 13,135 - Average: 421 s	Number of visits	Percent
0s-30s	10,580	80.5 %
30s-2mn	338	2.5 %
2mn-5mn	185	1.4 %
5mn-15mn	254	1.9 %
15mn-30mn	292	2.2 %
30mn-1h	485	3.6 %
1h+	956	7.2 %
Unknown	45	0.3 %

**Fig. 2.** Visitor Navigation Interest

For example, if we assign 30s, 2m, 5m and more than 15m to  $T_{\text{VeryLow}}$ ,  $T_{\text{Low}}$ ,  $T_{\text{Average}}$ ,  $T_{\text{High}}$ , and  $T_{\text{VeryHigh}}$  respectively, the time could be divided as follows:

$$T = \left\{ \begin{array}{l} 1; 0 < t \leq 30s \\ 2; 30s < t \leq 2m \\ 3; 2m < t \leq 5m \\ 4; 5m < t \leq 15m \\ 5; 15m < t \end{array} \right\} \quad (3)$$

#### 2.4 Time Preference.

Let  $S$  denotes URL group of Website and  $P$  denotes group of Visitor navigation path. Let us assumed,  $p \subset P, \forall e \in p$ ,  $e$  denotes Visitor navigation path element  $\forall s \in S$ , if all  $e$  have similar  $n$  processing elements and  $x$  dissimilar web pages at order  $(n+1)$ , we call there are  $x$  dissimilar selection on  $n^{\text{th}}$  elements. Selection time preference of the  $j^{\text{th}}$  selection ( $j=1, 2 \dots x$ ) is

$$T_j / \left( \left( \sum_{i=1}^x T_i \right) / x \right) \tag{4}$$

$T_i$  denotes  $i^{\text{th}}$  selection that visitors view the other webpage through the  $i^{\text{th}}$  selection. This methodology to identify the favorite of Webpages by visiting time. If one is utilize the link of a webpage to another page, they may spend a less time on the webpage. Although, this page have many viewed times, the time preference can less the visited preference [1].

### 2.5 Selection And Time Preference Algorithm (SATP)

Let  $S$  denotes URL group of Website and  $P$  denotes group of Visitor navigation path. Let us assumed,  $P \subset P, \forall e \in P, e$  denotes Visitor navigation path element  $\forall s \in S$ , if all  $e$  have similar  $n$  processing elements and  $x$  dissimilar web pages at order  $(n+1)$ , we call there are  $x$  dissimilar selection on  $n$ th elements. Selection time preference of the  $j^{\text{th}}$  selection ( $j=1, 2 \dots x$ ) is

$$(A_j T_j) / \left( \left( \sum_{i=1}^x A_i \right) \left( \sum_{i=1}^x T_i \right) / x^2 \right) \tag{5}$$

Where  $A_i$  denotes the  $i^{\text{th}}$  selection that Visitors view the other page through the  $i^{\text{th}}$  selection;  $T_i$  is the sum of time discretization of the other page through the  $i^{\text{th}}$  selection.

Selection and Time preference is a proportion comparing relatively access proportion and the average of relatively access proportion of visiting time and selective intention. Preference can preclude from only mining common visited paths. It can solve the aforesaid problems. While some Web pages have more viewed times due to their area, we could reduce this affected area by selection preference [7]. While some webpages have more visited times due to more paths, we could reduce this affected time preference because those pages just have limited visit time for Visitor to pass. The value of selection and time preference is between  $(0, x^2)$ , where  $x$  is the number of the selections. After the preference of a webpage exceeds 1, it means that Visitors choose to browse the page after viewing its home page. Certain rules would not be discovered when a threshold is excessively high. At the same time, the system would slow down due to redundant rules when a threshold is too low. Consequently, an appropriate threshold is decisive to mine Visitor preferred webpage navigation patterns. Generally, preference threshold is set to be 1.

On existing Method User Navigation path would only calculate based on time suppose the page is less content. That is time and selection preference of the visitor visit the page. This method involved user spent less time visited a web page based on time and selection preference. Our Proposed algorithm would calculate the user spending time based on both memory and session (time).

## 2.6 Preference of Page Content Size and Session Identifier Algorithm

```

Input: Preprocessed Weblog file (W3C)
Output: User interested page identification (based on Memory and Session)
(1) Load Logfile using StreamReader
    Stream LogFileS = File.OpenRead(FilePath);
    String ImportFileName = Path.GetDirectoryName(Application.ExecutablePath) +
@ "logfile.csv";
(2) User path identification purpose initialize the variables
String Line = LogFileSR.ReadLine().Trim();
int IPAddress = M.Result("${IPAddr}");
IPAddress = DecodeIPAddress(IPAddress);
int HoursFromInt = Int32.Parse(M.Result("${TZInt}"));
float HoursFromFrac = Int32.Parse(M.Result("${TZFrac}")) / 100.0f;
DateTime Session = new DateTime(Year, Month, Day, Hour, Minute, Second);
int Bytes= Int32.Parse(M.Result("${Bytes}"));
(3) identify the User viewing time and memory
while (LogFileSR.Peek() > -1)
{
String TextLine = IPAddress+ Session + Bytes+ ReferringURL ;
Extract (ReferringURL, Bytes, Session);
}
if (p.WaitForExit(86400))
{
long memory = GC.GetTotalMemory(true);
memory = new PerformanceCounter("Memory", "Available Bytes", true);
var s = p.StandardOutput.ReadToEnd().Split("\n");
}
(4) Calculated user interested pages
for (int i=1;i<= ReferringURL.Length; i++)
{
String Interestedpage =(Bytes+Session)/TotalNumberofvistedPages;
}
(5) End

```

## 3 Experimental Result

This method produces accurate user interested pages based on user interested website admin to identify the user navigation pattern. Although the structure of the website would improve based on user interest. This way to help with web user navigation. This method would improve for e-business. Fig. 3 shows the webpages, Memory, IP-address and session would present.

web pages	Memory	IP-Address	Session
Tab: New Tab	30,696K	64.233.173.243	0:38
Tab: Collegenotes.org – End o...	42,900K	64.233.173.248	1:53
Tab: Collegenotes.org – End o...	43,012K	50.18.94.121	1:56
Tab: E-Books – Collegenotes....	38,396K	50.151.42.38	1:40
Tab: Events – Collegenotes.org	33,852K	54.241.198.78	2:40
Tab: Projects – Collegenotes....	36,468K	66.249.82.172	3:33
Tab: Forum – Collegenotes.org	35,284K	91.121.222.130	0:25
Tab: My Account – Collegeno...	44,268K	52.6.165.159	3:45
Tab: My Account – Collegeno...	45,668K	37.187.77.44	2:45
Tab: Web Technology Record...	58,004K	66.248.83.172	3:02
Tab: Data Structures – Colleg...	56,660K	101.222.161.85	2:18
Tab: OOAD – Collegenotes.org	60,232K	101.222.161.85	4:15

Fig. 3. User Visiting webpage time and memory

In following method used to calculate the web user most visited or more spent time in a web page. And then the memory space added the session time, then divided by the total number of visited pages in User navigation pattern and following below

$$X = \frac{\sum_{i=1}^n (M + S)}{n} \tag{6}$$

Where X denotes User spent more time on a web page, M denotes the size of the memory in user visit the web page, S denotes session or time spent in user visit the web page and n denotes total number of visited pages in all users.

Table 2 shows the user viewing the webpage navigation interest would calculate based on the time (session) and Memory space.

Table 2. User page navigation interest calculation based on web page memory

Viewing Timing / Memory	Very Low	Low	Average	High	Very High
Very Low	Passing	Simple Viewing	Preferred Viewing	Preferred Viewing	Preferred Viewing
Low	Passing	Passing	Normal Viewing	Preferred Viewing	Preferred Viewing
Average	Passing	Passing	Preferred Viewing	Preferred Viewing	Preferred Viewing
High	Passing	Passing	Simple Viewing	Normal Viewing	Preferred Viewing
Very High	Passing	Passing	Passing	Preferred Viewing	Preferred Viewing

The above Table.2 is a user page navigation Interest would calculated after the viewing the webpage of the user. Once the user enters the webpage the time and memory space are very low just the user passing the page, that user not interested in the web page. Simple viewing means the user visit page time is low, but the memory is very low. That is also the user not interested the web page. Normal viewing means the user visiting time and memory are high. The user interested the web page but not view more content of the webpage. Preferred viewing means the user visiting the



webpage time and memory are very high. Although the user most interested viewer of the website. The Interested User view more spent time and view more content of the webpage.

## 4 Conclusion

This research paper focus on current technology of Internet E-learning, E-business, Digital marketing development of user pattern behavior identified accurately .In this article we propose a new algorithm for preference of page content size and session identifier algorithm, which signifies user navigation interest according to webpage memory size and viewing time . This methodology is appropriate for applications in E-Learning, E-Business, such as to enhance Web site to personalized facility. This work focused on more features to the preference to analyze user navigation interest more accurately. The proposed algorithm to perform personalization in e-learning platforms for “www.collegenotes.org” by resorting to a recommender system relying on web mining techniques.

## 5 References

- [1] Xing, Dongshan, and Junyi Shen. "Efficient data mining for web navigation patterns." *Information and Software Technology* 46.1 (2004): 55-63.
- [2] Nithya, P., and P. Sumathi. "An enhanced pre-processing technique for web log mining by removing web robots." *Computational Intelligence & Computing Research (ICCC)*, 2012 IEEE International Conference on. IEEE, 2012.
- [3] Dharmarajan, K., and M. A. Dorairangaswamy. "Discovering User Pattern Analysis from Web Log Data using Weblog Expert." *Indian Journal of Science and Technology* 9.42 (2016).
- [4] Honest, Nirali, Atul Patel, and Bankim Patel. "A study of Path Completion Techniques in Web Usage Mining." *Computational Intelligence & Communication Technology (CICT)*, 2015 IEEE International Conference on. IEEE, 2015.
- [5] Liu, Haibin, and Vlado Kešelj. "Combined mining of Web server logs and web contents for classifying user navigation patterns and predicting users' future requests." *Data & Knowledge Engineering* 61.2 (2007): 304-330.
- [6] Chen, Min, and Young U. Ryu. "Facilitating effective user navigation through website structure improvement." *IEEE Transactions on Knowledge and Data Engineering* 25.3 (2013): 571-588.
- [7] Wang, Yao-Te, and Anthony JT Lee. "Mining Web navigation patterns with a path traversal graph." *Expert Systems with Applications* 38.6 (2011): 7112-7122.
- [8] Pinanta Chatwattana, Prachyanun Nilsook. " A Web-based Learning System using Project-based Learning and Imagineering." *International Journal of Emerging Technologies in Learning (iJET)* Vol 12, No 05 (2017): pp. 4-22

## **6 Authors**

**K. Dharmarajan** is with the Research and Development Centre, Bharathiar University, Coimbatore, Tamil Nadu, India and Department of Information Technology, Vels University, Chennai, Tamil Nadu, India (dharmak07@gmail.com).

**Dr. M. A. Dorairangaswamy** is with St. Peter's University, Chennai, Tamil Nadu, India.

Article submitted 03 June 2017. Published as resubmitted by the authors 07 August 2017.