

## User behaviour pattern for online learning system: UiTM iLearn portal case

Siti Fairuz Nurr Sadikan<sup>1</sup>, Azizul Azhar Ramli<sup>2</sup>, Mohd Farhan Md. Fudzee<sup>3</sup>, Siti Sapura Jailani<sup>4</sup>,  
Mohd Ali Mohd Isa<sup>5</sup>, Prasanna Ramakrisnan<sup>6</sup>, Roslani Embi<sup>7</sup>

<sup>1,2,3</sup>Faculty of Computer and Information Technology, Universiti Tun Hussein Onn Malaysia, Malaysia

<sup>4,5,6,7</sup>Institut Pendidikan Neo (iNED), Universiti Teknologi MARA (UiTM), Malaysia

---

### Article Info

#### Article history:

Received Dec 15, 2018

Revised Feb 16, 2019

Accepted Mar 1, 2019

---

#### Keywords:

Classification

Navigational pattern modelling  
clustering

Server log files

Web usage mining

---

### ABSTRACT

A Web server log files contain an entire record of the user's browsing history such as referrer, date and time access, path, operating system (OS), browser and IP address. User navigation pattern discovery involves learning of user's browsing behaviour to gain the pattern from web server log file. This paper emphasizes on identifying user navigation pattern from web server log file data of iLearn portal. The study implements the framework for user navigation including phases of acquisition of weblog, log query parser, preprocessor, navigational pattern modelling, clustering, and classification. This study is conducted in the context of the actual data logs of the iLearn portal of Universiti Teknologi MARA (UiTM). This study revealed the navigational patterns of online learners which relatively related to their intake or group along the semester of 14 weeks. Besides, access patterns for students along the semester are different and can be classified into three (3) quarter, namely Q1, Q2 and Q3 based on the total of week per semester. Future work will focus on the development of prototype to improve the security of online learning especially during the assessment progress such as online quiz, test and examination.

*Copyright © 2019 Institute of Advanced Engineering and Science.  
All rights reserved.*

---

### Corresponding Author:

Siti Fairuz Nurr Binti Sadikan,  
Faculty of Computer and Information Technology,  
Universiti Tun Hussein Onn Malaysia,  
86400 Batu Pahat, Johor, Malaysia.  
Email: fairuz.sadikan@gmail.com

---

## 1. INTRODUCTION

The Industrial Revolution 4.0 (IR 4.0) has given a new impetus to educational transformation. In recent years, education through online learning is becoming more popular. The information stored mostly on internet especially website for online learning also increasing rapidly day by day. The web sites play an important role where the authenticated user including student, lecturer and other staff view, uploads, download, and browse many contents according to their need.

A web server provides a way to browse a web site by assigning IP address to identify the host, and to record every event in the form of web log file. Analyzing and modelling web navigation behaviour from web log file is helpful in understanding user behaviour activity. Web mining is the process of discovering hidden information from Web log file [1]. It can be classified into three different categories namely web content mining, web usage mining and web structure mining [7] as shown in Figure 1.

A web content mining is the discovery of contents from web documents including web search content, search page content and result page content such as image, text, audio, video etc. In the other hand, a web structure mining focus on analyzing the physical link structure of websites such as link structure,

internal structure and URL structure. Besides, web usage mining analyzes the browsing activity which including the phases of preprocessing, pattern discovery and pattern analysis.

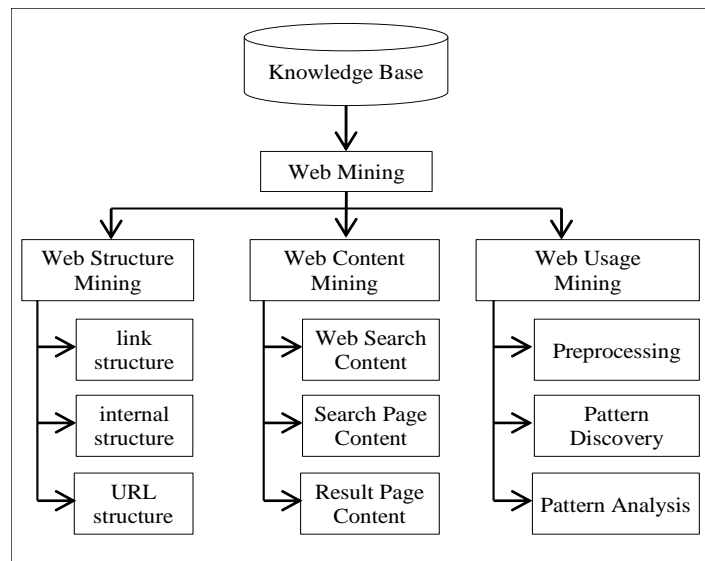


Figure 1. Web Mining Categories

The aim of web usage mining is to understand the browsing and navigation through web pages in order to enhance many things and it can be used for different purposes such as personalization, system improvement, site modification [3] and identifying the user behaviour. In web usage mining, the main data mining techniques are used including mining of association rules, extraction of consecutive patterns, and clustering in order to extract conductive patterns and offer recommendations based on them [5].

This study presents an algorithm for preprocessing of web log file, clustering of user navigation pattern for modelling user navigation pattern. The organization of the paper is as follow: Section 2 illustrates the related work, Section 3 discusses framework for user navigation which contain several phase and preprocessing step. This section also contains the algorithm for data cleaning, user identification, session identification and content retrieval. Section 4 presents the sample results and last section presents the conclusion and future work for this study.

## 2. RELATED WORK

The user navigation behaviour based on preferences can be predicted from the result of web usage mining. Web usage mining is one of the active research areas and extensive research work has been carried out in the recent years [10]. There are number of techniques have been proposed by various author including acquisition of web log, preprocessing, pattern discovery and pattern analysis.

Previous study as in [6] proposed the automatic classification of web user navigation patterns which is a novel approach for classifying user navigation patterns and predicting future requests of expected users. While [4] proposed a system for discovering user navigation patterns using a graph partitioning model where undirected graph based on connectivity between each pair of Web pages was considered and weights were assigning to edges of the graph. Besides, author [3] presented another user navigation pattern mining system based on the graph partitioning. An undirected graph based on connectivity between Referrer and URI pages was presented along with a pre-processing method to process unprocessed web log file and a formula for assigning weights to edges of the undirected graph.

Besides, [8] proposed a solution to predict user request from navigation pattern by using graph partitioned clustering algorithm to group users with similar navigation pattern. An undirected graph based on the connectivity between each pair of web pages is used. Each edge in the graph is assigned a weight, which is based on the connectivity time and frequency. Connectivity time measures the degree of visit ordering for each two pages in a session. Meanwhile, author [2] presents the Prediction of User navigation patterns using Clustering and Classification (PUCC) from web log data.

In the other hand, [3] proposed the use of weighted fuzzy probabilistic c-means algorithm for pattern discovery on web usage mining and adaptive neuro-fuzzy inference system with subtractive algorithm for user navigation pattern analysis. The researcher claims the study had improving the prediction result. Author [11] proposes a novel approach called Fuzzy C-Means Clustering-based Collaborative Filtering approach (FCM based CF) and algorithm consolidates the web services and suggests the better web services based on the user navigation. It shows that fuzzy c-mean algorithm can improve the prediction result. Therefore, the researcher extends the research by presenting framework for user navigation with the used of fuzzy c-mean algorithm to find out the user behaviour in this study.

### 3. FRAMEWORK FOR USER NAVIGATION

In this section, the framework for user navigation is presented to analysing user behaviour. This framework contains of six (6) phases which are acquisition of weblog, log query parser, preprocessor, navigational pattern modelling, clustering, and classification as shown in Figure 2.

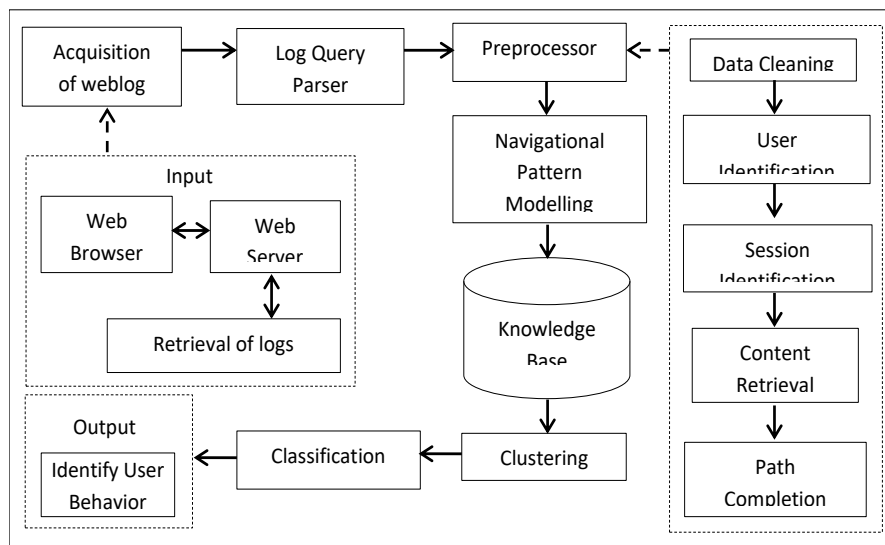


Figure 2. User Navigation Framework

#### 3.1. Phase 1: Acquisition of Weblog

Weblogs consist of history of user navigation that stored in web server. A weblog files of the iLearn portal are made up of millions of lines, each of which representing an operation performed by a particular user such as student, lecturer and staff in the online learning. The size of the log files keeps growing, due to the increase in the number of users of the online learning, and the variety of event information. Due to this limitation, only 45012 records is used in the study which consist of the following attributes such as no. pelajar, referrer, date and time access, path, operating system (OS), browser and IP address.

#### 3.2. Phase 2: Log Query Parser

The log query parser is taken to extract unstructured log to structured log based on the user interest. This parser provides universal query access to text based data such as log files, xml files and csv files.

The log file in form of .xls is converted into .csv file in this study.

#### 3.3. Phase 3: Preprocessor

The log file contains unstructured format of user navigation information, so conversion is required and can be done through data preprocessing technique. This process deals with loading of the data, performing accuracy check, putting the data together from disparate sources, transforming the data into required format and finally to structure the data as per the input requirements of some data mining algorithm [9]. In this study, the phase of data preprocessing technique consists of five (5) steps which are data cleansing, user identification, session identification, content retrieval and path completion as shown in Figure 3.



Figure 3. Data preprocessor steps

**Step 1: Data Cleansing**

Web log data file consists of many irrelevant data including image requests, erroneous requests and spider navigation requests. In this step, the irrelevant records are eliminated in order to get the traversal pattern. The unnecessary records for this research are ‘/modules/main/bottom.php’, ‘/activity/Jason’, ‘/messages/notification’, and ‘/tracker/load’. The data cleaning algorithm is used to eliminate irrelevant and unnecessary records as shown in Figure 4.

```

Data Cleansing Algorithm
Input: Web Server Log File data
Output: Log File data

Step 1: Read log file record from (Web Server Log File)
Step 2: IF ((log file record. referrer == '/modules/main/bottom.php') ||
(log file record. referrer == '/activity/Jason') || (log file
record. referrer == '/messages/notification') || (log file record.
referrer == '/tracker/load') || (log file record. referrer ==
'http://i-learn.uitm.edu.my/v3'))
{ Remove from the log file } End IF
Step 3: Repeat step 1 and step 2 until EOF (End of File)
Step 4: Stop and save file in database.
END
    
```

Figure 4. Data Cleaning Algorithm

**Step 2: User Identification**

This step requires a creating UserID table where matric no. is used to differentiate the user in order to find out the id for each user. The user identification algorithm is shown in Figure 5. This Figure 5 shows user identification algorithm which based on the matric no. of student that registered in i-learn portal. Besides, the table are also created to store UserID, Matric No., Date access, Path, Operating system, Browser, IP address.

```

User Identification Algorithm
Input: Log File data
Output: Unique User Table

Step 1: Initialization
Create Table include the following field:
(UserID, Matric No., Date access, Path, Operating system, Browser, IP
address)
Step 2: Read record from log file data
Step 3: User's matric no. sequential records are compared
Step 4: IF (matric no. is NOT IN Users Table)
THEN assign UserID to matric no.
Add both to Users Table
ELSE IF (matric no. is IN User Table)
THEN Add it with same UserID
ELSE Assign (next UserID) to matric no.
Add both to User Table
Step 5: Repeat step 2 to step 5 until EOF (Log File data)
Step 6: Stop and store result.
END
    
```

Figure 5. User Identification Algorithm

### Step 3: Session Identification

In this step it requires creating a SessionID column where the user session can be classified based on IP Address, Browser, Operating System (OS) and Date and Time. The session identification algorithm is shown in Figure 6. The Figure 6 shows the session identification algorithm which consisting of several step and based on IP address, browser and types of operating system used along the session.

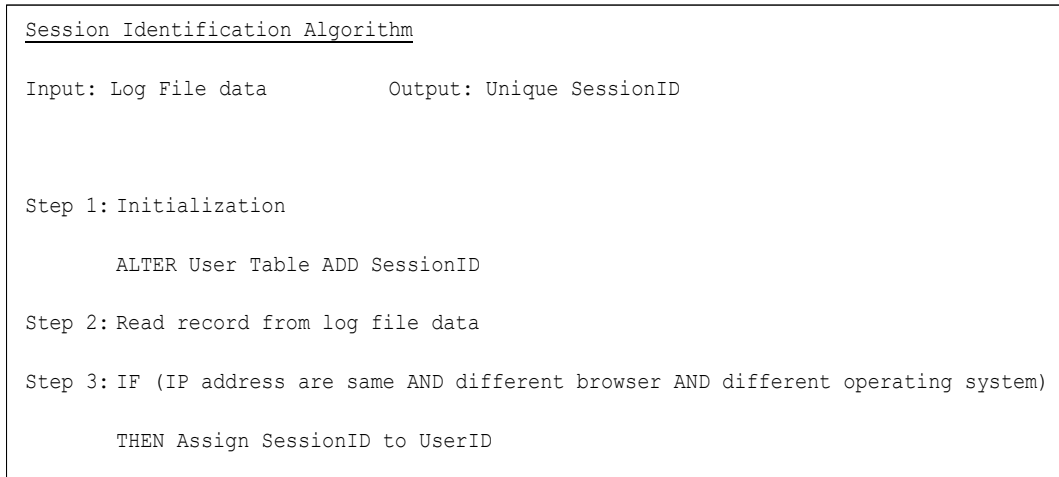


Figure 6. Session Identification Algorithm

### Step 4: Content Retrieval

This step is used to retrieve content from the referrer which helps in fast searching. Besides, the content retrieval algorithm is shown in Figure 7. The Figure 7 shows content retrieval algorithm which consisting of several steps to retrieve only necessary referrer.

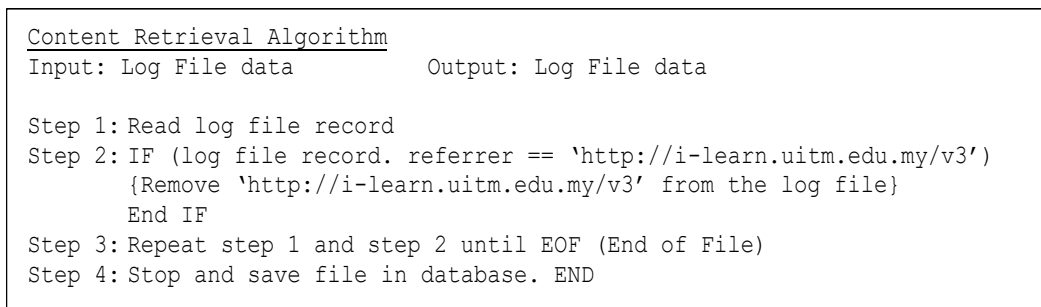


Figure 7. Content Retrieval Algorithm

### Step 5: Path Completion

Path completion should be used to acquire the complete user access path. The incomplete access path of every user session is recognized based on user session identification. The generation of PageID in sequence number like P1, P2, P3, P3...Pn are created for referrer with the activity as shown in the Table 1. The Table 1 show the sample result of path completion for PageID with the activity based on the path listed in previous step 4. This Table 1 also show the sample result which consisting of P1=Home; P2=Summary; P3=Announcement; P4=Content; P5=Assignment; P6=Entrance and Exit Survey (EES); P7=Course Glosory; P8= Course References; P9= Course Forum; P10= Assessment; P11= Member; P12=Student Feedback Online (SuFo); P13= Drawer; and P14= myCommunity.

Table 1. Sample Result for Path Completion

Page ID	Path	Activity
P1	/users/profile	Home
P2	/courses/summary/ITS432	Summary
P3	/announcements/index/c/ITS432/all /announcements/index/cg/2009373/all	Announcement
P4	• /contents/index/t:c/cid:ITS432 • /contents/index/5a97641e-424c-4ea7-8683-4c7ac0a80109/cid:ITS432 • /contents/index/cid:2009373	Content (view or download notes etc.)
P5	/assignments/dashboard/home/ITT400	Assignment
P6	/ess/dashboard/home/ITT400	EES (Entrance and Exit survey)
P7	/course_glossaries/index/ITT400	Course Glossary
P8	/course_references/index/cid:ITT400	Course References
P9	• /forumsv2/lobby/index/ITT400 • /forumsv2/lobby/index/ITS432/2009373	Course Forum
P10	/gradebook/index/2009373	Assessment
P11	/course_group_users/members/2009101	Member
P12	/sufo/surveys/index	SuFo
P13	/drawers/drawer	Drawer
P14	/groups/index, /groups/add/, /groups/join_group/, /groups/leave_group/	myCommunity

**3.4. Phase 4: Navigational Pattern Modelling**

After the preprocessing of web server log file, data mining technique are then applied. The sequence of pattern is improved from pre-processor technique, it contains the forward reference. The sub sequences can be generated by the maximum forward algorithm where it contains both forward and backward reference. The web pages accessed by the user are modelled as directed graph which N nodes represent N web pages as shown in Figure 8. This figure shows decision tree for online learning user for i-learn portal. The decision tree is generated from the weblog of i-Learn portal. Then, data of web server logs can be transformed into knowledge to uncover the potential patterns underneath the preprocessed log data and involves analyses of these patterns [9].

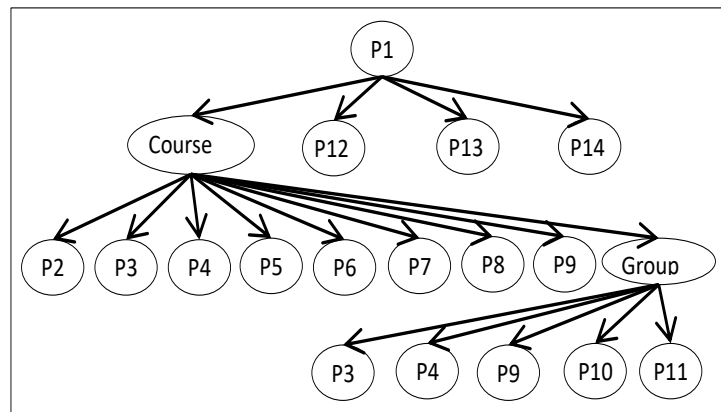


Figure 8. Decision Tree for Online Learning User

**3.5. Phase 5: Clustering**

This phase used to cluster the user behavior and the navigational pattern. Clustering plays an important role in data analysis and understanding behavior of user in the website. For this study, Fuzzy c-mean algorithm is used to find out the user behavior. There are two types of cluster which are the user cluster and the page cluster. Web page clustering is performed by grouping pages having similar content while user clustering is performed by grouping users by their similarity in navigational behavior [9]. Table 2 shows sample result of the browsing pattern of user behaviour for userID=2.

Table 2. Sample Result of the Browsing Pattern of User Behaviour for Userid=2

No	Pattern No	Browsing Pattern
1	Pattern1	P12, P13, P5, P1, P6, P3, P11, P4, P9
2	Pattern2	P6, P1, P4, P9, P8
3	Pattern3	P12, P4, P8, P1, P9, P11
4	Pattern4	P4, P1, P5
5	Pattern5	P12, P4, P11
6	Pattern6	P9, P7, P5, P12, P4, P3, P1
7	Pattern7	P1, P9, P11, P4
8	Pattern8	P1, P4, P9, P8
9	Pattern9	P1, P4, P11
10	Pattern10	P4, P2, P1, P12, P6, P5

### 3.6. Phase 6: Classification

This phase utilizes the browsing pattern based on previous analysis. The user behaviour is classified into three categories of semester classification along 14 weeks per semester as shown in the Figure 9 based on the activity interaction such as entrance and exit survey (EES), content and Students' Feedback Online (SuFO). An early semester is categorized between week 1 to week 2, where middle semester is between week 3 to week 9 and late semester is between weeks 10 to week 14. The pattern supports organization with frequent pattern for user profiling.

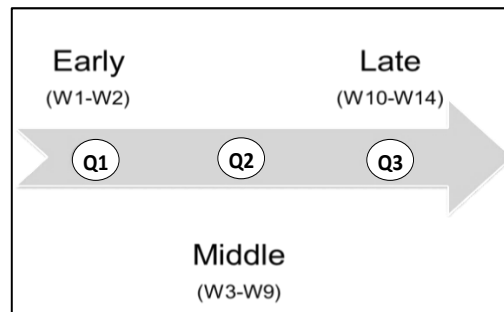


Figure 9. Decision Tree for Online Learning User

## 4. RESUL AND DISCUSSION

The result of this study revealed the navigational patterns of online learners in i-Learn portal. From this study, navigational patterns of user relatively related to their intake or group along the semester of 14 weeks is also revealed. Besides, access patterns for students along the semester are different and can be classified based on the total of week per semester. Then, the 14 week were divided into three (3) quarters, namely Q1, Q2 and Q3 as discussed in the previous section. Table 3 presented the sample result of classification of user pattern from this study.

Table 3. Sample Result of Classification User Pattern Prediction

Pattern ID	Pattern	Semester Classification
Pattern1	P1, P6, P11, P4	Q1
Pattern2	P1, P6, P11	
Pattern3	P1, P6, P4	
Pattern4	P1, P4, P11	Q2
Pattern5	P1, P4, P5	
Pattern6	P1, P9, P11, P4	
Pattern7	P4, P2, P1, P12, P6, P5	Q3
Pattern8	P12, P13, P5, P1, P6, P3, P11, P4, P9	
Pattern9	P9, P7, P5, P12, P4, P3, P1	
Pattern10	P1, P12, P6	

(Q1=Early semester, Q2=Middle semester, Q3=Late semester)

## 5. CONCLUSION AND FUTURE WORK

This paper shows the analysis of the web server log file data of iLearn portal in order to gain the navigational pattern of online learning user. The study of online user behaviour while navigating online sites is an important issue that can help to improve the security of online learning especially during the assessment progress such as online quiz, test and examination. The framework for user navigation is implemented in this research for clustering purposes in order to get pattern of learner's activity. Future work will focus on the development of prototype to accompany this work.

## ACKNOWLEDGEMENTS

The authors would like to thank the Institut of Neo Education (iNED), Universiti Teknologi MARA (UiTM) for their assistance with the collection of data for this research. This work is sponsored by Universiti Tun Hussein Onn Malaysia under TIER1/2007, UTHM research grant and Gates IT Sdn. Bhd.

## REFERENCES

- [1] Al-Asadi, T. A., & Obaid, A. J. (2016). Discovering Similar User Navigation Behavior in Web Log Data. *International Journal of Applied Engineering Research*, 11(16), 8797–8805.
- [2] Bharat, V. M., Atmaram, M. P., & Nirgude, V. N. (2015). Efficient User Navigation Pattern Prediction Technique. *International Journal of Advance Research and Innovative Ideas in Education*, 1(14), 424–427.
- [3] Dixit, D., & Gadge, J. (2010). A New Approach for Clustering of Navigation Patterns of Online Users. *International Journal of Engineering Science and Technology* 2(6), 1670–1676.
- [4] Jalali, M., Mustapha, N., Mamat, A., & Sulaiman, M. N. (2008). A New Clustering Approach Based on Graph Partitioning for Navigation Patterns Mining. In *International Conference on Pattern Recognition* (Vol. 4, pp. 1–4). IEEE. <https://doi.org/10.1109/ICPR.2008.4761808>
- [5] Liraki, Z., Harounabadi, A., & Mirabedini, J. (2015). Predicting the Users' Navigation Patterns in Web, using Weighted Association Rules and Users' Navigation Information. *International Journal of Computer Applications*, 110(12), 975–8887.
- [6] Liu, H., & Kešelj, V. (2007). Combined Mining of Web Server Logs and Web Contents for Classifying User Navigation Patterns and Predicting Users' Future Requests. *Data and Knowledge Engineering*, 61(2), 304–330. <https://doi.org/10.1016/j.datak.2006.06.001>
- [7] Prakash, P. G. O. (2016). Analyzing and Predicting User Behavior Pattern from Weblogs. *International Journal of Applied Engineering Research*, 11(9), 6278–6283. Retrieved from <http://www.ripublication.com>
- [8] Sujatha, V., & Punithavalli. (2012). Improved User Navigation Pattern Prediction Technique From Web Log Data. In *International Conference on Communication Technology and System Design* (Vol. 30, pp. 92–99). Elsevier Ltd. <https://doi.org/10.1016/j.proeng.2012.01.838>
- [9] Sundari, M. R., Srinivas, Y., & Reddy, P. P. (2014). A Review on Pattern Discovery Techniques of Web Usage Mining. *International Journal of Engineering Research and Applications*, 4(9), 131–136. Retrieved from [www.ijera.com](http://www.ijera.com)
- [10] Vellingiri, J., Kaliraj, S., Satheshkumar, S., & Parthiban, T. (2015). A novel approach for user navigation pattern discovery and analysis for web usage mining. *Journal of Computer Science*, 11(2), 372–382. <https://doi.org/10.3844/jcssp.2015.372.382>
- [11] Vimali, J. S., & Taj, Z. S. (2016). FCM based CF: An Efficient Approach for Consolidating Big Data Applications. In *Proceedings 2015 - IEEE International Conference on Innovation, Information in Computing Technologies, ICICT 2015*. Chennai, India: IEEE. <https://doi.org/10.1109/ICICT.2015.7396090>

## BIOGRAPHIES OF AUTHORS



Siti Fairuz Nurr Sadikan is a Doctor Philosophy student in the Faculty of Computer and Information Technology, Universiti Tun Hussein Onn Malaysia, Malaysia. Her research interests are authentication, behavioural biometric, keystroke dynamic, neural networks, fuzzy logic etc.





Azizul Azhar Ramli (Non-Member) received the B.Sc. degree in computer science from Universiti Teknologi Malaysia, Malaysia, and the M.Sc. degree in intelligence system from Universiti Utara Malaysia, Malaysia in 2002, and 2004, respectively. He finished his Ph. D degree at Graduate School of Information, production and Systems (IPS), Waseda University, Japan. His research interests include management information system, data mining, expert system, fuzzy modeling, and real-time



Mohd Farhan Md Fudzee is currently working as an associate professor at Multimedia department, Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia (UTHM). He has published a book in IT programming, 2 academic modules, 2 international book chapters, 55 technical papers in reviewed international conferences and indexed journals as well as technical reports. His research interest includes Multimedia Computing (Content Adaptation), Green Computing (Energy-related), Decision Making algorithms, Web Technology and Creative Applications.



Siti Sapura received her Bsc degree in Information Technology from University Technology MARA (UiTM). Currently she is working at University Technology MARA as Senior IT Officer.



Mohd Ali Mohd Isa is a senior lecturer at MARA University of Technology (UiTM). Previously he was the Head of Department for System Management at the Center of e-Learning UiTM from 2005 until 2015. He was responsible for the setting up of the first Learning Management System for UiTM and oversees the working of the system for 10 years. Currently he is teaching the subjects of Web Technology and Information Security, which are 2 of his research areas.



Prasanna Ramakrisnan received her Doctor of Philosophy in Visual Informatics from National University of Malaysia. Currently she is the Head of Technology Unit at Institute of Neo Education (iNED), Universiti Teknologi MARA (UiTM). Her research area includes e-learning, gamification and human computer interface.



Roslani Embi received his Doctor of Philosophy in Accounting from Virginia Poly Institute & State University, Blackburg, USA. Currently he is the Director at Institute of Neo Education (iNED), Universiti Teknologi MARA (UiTM). His research area includes Accounting and Accounting Education.