EAI Endorsed Transactions
on e-Learning

Research Article ICST.ORG

Web Log Pre-processing and Analysis for Generation of Learning Profiles in Adaptive E-learning

Radhika M. Pai[1], Sucheta V. Kolekar[1,*], Manohara Pai M.M.[1]

[1]Department of Information and Communication Technology, Manipal Institute of Technology, Manipal University, Manipal, Karnataka, India

Abstract

Adaptive E-learning Systems (AESs) enhance the efficiency of online courses in education by providing personalized contents and user interfaces that changes according to learner's requirements and usage patterns. This paper presents the approach to generate learning profile of each learner which helps to identify the learning styles and provide Adaptive User Interface which includes adaptive learning components and learning material. The proposed method analyzes the captured web usage data to identify the learning profile of the learners. The learning profiles are identified by an algorithmic approach that is based on the frequency of accessing the materials and the time spent on the various learning components on the portal. The captured log data is pre-processed and converted into standard XML format to generate learners sequence data corresponding to the different sessions and time spent. The learning style model adopted in this approach is Felder-Silverman Learning Style Model (FSLSM). This paper also presents the analysis of learner's activities, pre-processed XML files and generated sequences.

Received on 29 November, 2015; accepted on 01 March, 2016; published on 11 April, 2016

Keywords: Web Log Analysis, Felder-Silverman Learning Style Model, Adaptive E-learning Systems, XML, Data Pre-processing, Sequences, Adaptive User Interface etc.

doi:10.4108/eai.11-4-2016.151151

1. Introduction

In the past decade, Adaptive E-learning Systems (AESs) have attracted much attention of the researchers in the fields of Educational Data Mining (EDM). Many AESs have been proposed and implemented on Human Computer Interface (HCI) and Data Mining. Most of these AESs have focused on addressing the requirements of learners in order to improve the interaction and efficiency of the systems. E-learning systems cannot perform to the desired level if requirements handling is not an integrated part of web-based learning systems. Educators have identified the problem with these systems is that they mainly focus on learning materials and do not consider the adaptive Learner interface requirements of the learners. While going through the online courses on e-learning portals, most of the learners are unsure of their actual needs which may lead to inaccurate requests of the learning contents. To address these issues, it is beneficial that e-learning system analyze the actual needs of learners to improve their learning performance [9].

Many educational theories suggest that the integration of learning styles into learning activities may improve learning performance. AESs mainly focus on adaptation and personalization of Learner interface along with adaptive learning contents. The integration of Learner profiles with the learning style can be useful to enhance AESs by providing a learner with personalized learning guidance that is appropriate to their potential needs.

The work proposed in this paper is useful to identify the learning profiles based on analysis and pre-processing of web logs data. The identified learning styles is classified based on Felder-Silverman Learning Style Model (FSLSM) to identify exact types of learners. Many researchers have concentrated on identifying learning styles based on FSLSM to provide learning contents to specific types of learners. The described work is integrated with web portal that can be applied with other e-learning systems.

This paper is organized as follows. Section II introduces related research work in pre-processing and learner profiles with adaptive e-learning styles. Section III describes the methodology of the system. Section IV presents algorithmic approach of analysis and pre-processing methods of web

*Corresponding author. Email: kolekar.sucheta@gmail.com

log data. Section V presents the log analysis and sequence data which is the result of pre-processing. Section VI gives conclusions and discusses future works.

2. Related Works

In teaching-learning process every learner will have different background knowledge, learning preferences and individual students may expect different approaches towards learning courses. With this main objective, adaptive e-learning systems (AEHSs, AdaLearn, iLessons) have been developed to offer personalized learning content to improve their learning outcome [20] [2] [18].

Many researchers have been concentrating on identifying the learner's learning styles to implement adaptation in online learning management systems. Automatic and dynamic detection of learning style has its attractive advantages in incorporating into portal to identify types of learners in no time [1] [15] [6]. There are many methods proposed which automatically estimates learner's learning styles with respect to the Felder-Silverman Learning Style Model based on their behaviors about the online course.

To bring about the dynamic nature into the system and to provide adaptive interface, the learning behavior of an individual has to be analyzed and modeled. In order to understand the process, it is necessary to automate the process of detecting learning style based on the learning behavior and generate personalization and adaptation for different types of learners. Considering learning and how to improve a learner's performance, these systems must know the way in which an individual student learns best [17] [3].

A learning-style model classifies students according to where they fit on a number scales pertaining to the ways they receive and process information. Mismatches exist between common learning styles of engineering students and traditional teaching styles of engineering professors. [7] [8]. The importance of applying these learning styles to different learning systems, various problems still need to be solved, such as matching teaching contents with the student's learning style [8]. Enabling teachers to know their learner's learning styles and making students aware of their own learning styles increases teacher's and learner's understanding about the learner's learning process, allows teachers to provide better support for their students, and has therefore high potential to enhance teaching and learning [10] [5].

Web Usage Mining (WUM) is a kind of data mining method that can be used to discover user access patterns from Web log data. WUM includes three phases that are called pre-processing, pattern discovery and pattern analysis. Web usage analysis or web usage mining or web log mining or click stream analysis is the process of extracting useful knowledge from web server logs, database logs, user queries, client side cookies and user profiles in order to analyze web learner's behavior. Web usage analysis requires data abstraction for pattern discovery. This data abstraction can be achieved through data pre-processing [13] [16] [11] [12].

An important research area in education and technology is how the learners use e-learning. By exploring the various factors and relationships between them, we can get an insight into the learner's behaviors for delivering tailored e-content required by them. Although many tools exist to record detailed navigational activities, they don't explore the learner's usage patterns for an adaptive e-learning site [14] [19] [4].

3. Methodology

The approach of learning profiles generation through usage log data of e-learning portal which is shown in Fig.1. The detailed analysis of log data and pre-processing of usage data is used to generate learning sequences. The implementation of e-learning portal has been done using Microsoft Visual Studio 2010 and Microsoft SQL server 2005. The portal is deployed on IIS server to provide access to all learners in the Internet. The log file option of IIS server is set to W3C extended log file format which will capture usage details of learners who are accessing the portal.

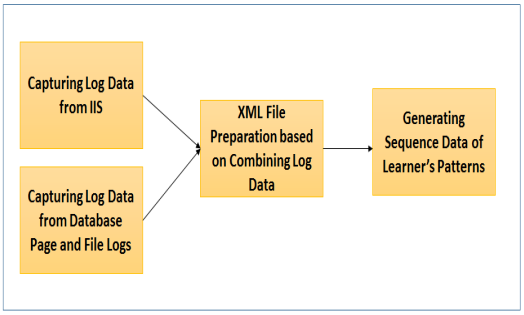


Figure 1. Web Log Analysis Approach

3.1. Evaluation Parameters based on Log File and Database

Following are the parameters used to define the log file and database.

1. **Topics (T):** Topics are related to the contents of the portal and are defined by the instructor of the course.
2. **Actions (A):** The click made by learner on particular content type like text links, video lectures, downloadable links etc. Following are the type of contents for analysis:
   - A_1=Page Request
   - A_2=Text Data: PDF
   - A_3=Demo: PPT
   - A_4= Video:MP4

   Action A_1 is default action for any other action.

3. **Accessing Parameters of Files**

   (a) **Duration (D(TFile$_i$)):** is the time spent on a specific type of file of a topic. The time is calculated based on the starting time (Start$_{Time}$) and ending time (End$_{Time}$) of accessing a file in a particular session time.
   Session Time (Sess$_{Time}$): As equation 1, time between Logged in action and Logged out action.

$$D(TFile_i) = End_{Time} - Start_{Time} \quad (1)$$

   File$_i$=file id assigned to each file for specific topic.
   n = number of files available in the portal.
   Total duration (D) spent on accessing files for specific topic is calculated in equation 2.

$$D = \sum_{i}^{n} D(TFile_i) \quad (2)$$

   (b) **Frequency (Freq$_i$):** Count of accessing a type of file during the visits on portal shown in equation 3. Frequency (Freq) indicates the learner's interest in specific type of contents which will help to identify the learning style on portal.

$$Freq_i = \sum_{i}^{n} A_i \quad (3)$$

   A$_i$= action of accessing a specific type of file during the visits on portal.

4. **Accessing Parameters of Pages**

   (a) **Duration (D(TPage$_j$)):** is the time spent on a page. The time is calculated based on the starting time (Start$_{Time}$) and ending time (End$_{Time}$) of accessing page in a particular session time.
   Session Time (Sess$_{Time}$): As in equation 4, time between Logged in action and Logged out action.

$$D(Page_j) = End_{Time} - Start_{Time} \quad (4)$$

   Page$_j$=page id assigned to each page on portal.
   m = number of pages available in the portal.
   Total duration (D) spent on accessing pages on portal is calculated in equation 5.

$$D = \sum_{j}^{m} D(Page_j) \quad (5)$$

   (b) **Frequency (Freq$_j$):** Count of accessing pages during the visits on portal shown in equation 6. Frequency (Freq) indicates the learner's interest in pages which will help to identify the learning style on portal.

$$Freq_j = \sum_{j}^{m} A_j \quad (6)$$

   A$_j$= action of accessing a page during the visits on portal.

4. Web Log Pre-processing to identify Learning Sequences

Data pre-processing consists of various steps which are tedious and time consuming to implement in real-time application. Also same steps can not be suitable in e-learning application for analysis. The proposed methodology describes about the issues to be addressed, parameters to be considered and algorithms to generate the XML files for learning sequences.

4.1. Issues to be addressed during Pre-processing:

1. Collection of Web Log Data: The web log data for usage of learners need to be captured in different formats. IIS log data of web server gives the information related to sessions of learners along with pages accessed. There is an requirement to capture data at the application level on web server which will consider the detailed usage of learner's sessions. Logs are to be generated and stored in the database along with sequence of pages and files which are accessed by specific learner on portal.

2. Pre-processing of Web Log Data: The captured web log data needs to be converted into standard format for analysis purpose. Generally, several pre-processing tasks need to be done before applying web mining algorithms on the web server logs such as Data Cleaning, Log Identification and Session Identification.

3. Pre-analysis of Web Log Data: Before converting the web log data into standard format the pre-analysis has to be done with different algorithmic approaches. This is useful to identify patterns and to check whether the captured data is correct or not.

4. Length and Order of the activities in a session: The e-learning sessions are as set of activities where learner is accessing available learning components and contents on portal. To identify the learning styles of a learner, logs should capture the activities with respect to time in each session. Such sessions will be different from other sessions according to length and order of activities. To combine common patterns and common Learner profiles sessions should be aligned based on similarity between sessions.

4.2. Assumptions/Requirements for Pre-processing of Web Log Data:

1. Idle time spent on specific file or page is 10 min. If a learner is not doing any activity on portal for 10 min then the session will terminate automatically and learner will get logged-out from the portal as per time oriented heuristic.

2. All sessions of one specific learner's will be considered upto 30 min to generate final XML logs.

3. If a learner has not logged out or left the session in between then those sessions are considered for 10 min as per idle time for final XML logs.

4. Unique session ids are maintained for every log record of learner.

5. Log records are filtered out only for GET/POST methods and for .aspx pages.

6. Log records are also filtered out for unique learners by removing Crawler/Spider/Robot.

7. Log records are sorted as per session time and different sessions of one learner are combined to understand usage patterns.

4.3. Parameters to be considered in XML generation algorithms:

1. Session: Sequence of pages and files accessed by a learner on a particular website during a specified period of time. One sequence includes number of sessions with total time of 30 mins.

2. Frequency: Number of times a specific file and a page is accessed by learners in all sessions.

3. Time Spent: Total time spent on file or page by learner's in all sessions.

4. Page Sequence: Pages and files accessed in a specific order by a learner. Page sequence is useful to identify the learning path of each learner.

The log data is generated in standard XML format based on Assumptions mentioned above. The XML file can directly be useful as input for clustering algorithms. Two different XML files are generated: File1 contains Learner session data of pages/files, time spent on page/file in each session and frequency of accessing the pages/files. File2 contains page and file sequence of each learner which are called as web sessions.

4.4. XML File1 and File2 Generation:

Session can be described as the time spent on portal by a learner from the moment he/she logged in to the moment he/she logged out. The sessions of each learner is combined and identified as the time spent on each page/file separately. Session also describes how many time each learner accessed the page/file. The total time spent on each page and file as well as frequency of accessing the page and file is converted into XML tags for the unique learner id as shown in Algorithm 1. Session can also described as the sequence of pages and files accessed by each learner in specific order as shown in Algorithm 2.

INPUT: A finite set of Learners $L = L_1, L_2, ...L_N$,
Sessions $S = S_1, S_2, ...S_Q$,
PageURL $P = P_1, P_2, ...P_R$
and FileURL $F = F_1, F_2, ...F_X$
OUTPUT: XML File

initialize $Session_{Time} \leftarrow 0$, $SessionLog_{ID} \leftarrow 0$,
$SessionPageLog_{ID} \leftarrow 0$, $SessionFileLog_{ID} \leftarrow 0$,
$PCount \leftarrow 0$, $PLogTime \leftarrow 0$, $FCount \leftarrow 0$,
$FLogTime \leftarrow 0$, $Flag \leftarrow 0$
for *each Sessions* S_j where $j \leftarrow 1$ **to** Q **do**
  compute $Session_{Time} = Start_{Session} - End_{Session}$
  if "$Session_{Time}$" $> 29min$ **then**
    get $Learner_{ID}$, $SessionLog_{ID}$ from *SessionLog* file
  end if
end for
for *each Session* S_j where $j \leftarrow 1$ **to** Q **do**
  for *each PageURL* P_k where $k \leftarrow 1$ **to** R **do**
    get $SessionPageLog_{ID}$ from *PageLog* file
    if $SessionLog_{ID} == SessionPageLog_{ID}$ **then**
      get $Learner_{ID}$, *PageURL*, $Page_{ID}$, *LogTime* from *PageLog* file
      set $Flag \leftarrow 1$
    end if
  end for
  if $Flag == 1$ **then**
    for *each PageURL* P_k where $k \leftarrow 1$ **to** R **do**
      for *each Learner* L_i where $i \leftarrow 1$ **to** N **do**
        if "*PageURL* "*is accessed* **then**
          set $PCount = PCount + 1$
          set $PLogTime = PLogTime + LogTime$
        end if
        create *XMLTag* for $Session_{ID}, Learner_{ID}, PageURL, PLogTime, PCount$
      end for
    end for
  end if
end for
for *each Session* S_j where $j \leftarrow 1$ **to** Q **do**
  for *each FileURL* F_d where $d \leftarrow 1$ **to** X **do**
    get $SessionFileLog_{ID}$ from *FileLog* file
    if $SessionLog_{ID} == SessionFileLog_{ID}$ **then**
      get $Learner_{ID}$, *FileURL*, $File_{ID}$, *LogTime* from *FileLog* file
      set $Flag \leftarrow 2$
    end if
  end for
  if $Flag == 2$ **then**
    for *each FileURL* F_d where $d \leftarrow 1$ **to** X **do**
      for *each Learner* L_i where $i \leftarrow 1$ **to** N **do**
        if "*FileURL* "*is accessed* **then**
          set $FCount = FCount + 1$
          set $FLogTime = FLogTime + LogTime$
        end if
        create *XMLTag* for $Session_{ID}, Learner_{ID}, FileURL, FLogTime, FCount$
      end for
    end for
  end if
end for

Algorithm 1: XML File1 Generation Algorithm

5. Experimentation Results and Analysis

5.1. Result and Analysis of Web Log Data

1. **Analysis done from Database: Frequency and Total time spent on specific file by learners in all of their sessions**
   Fig.2 shows the number of times file accessed and total time spent on specific file by learners in all of their sessions. Learners who have accessed the portal are spent time on files in different sessions. The graph is shows the result of most frequently accessed files by learners in specified duration.

2. **Snapshot of Report at Instructor Side: Number of times a specific type of file accessed by Learners**
   Fig.3 shows the report which an instructor can generate to get learner wise count of access of different types of files. As per the implementation, the portal is supporting only for PDF, PPT and Video files. Depending on the frequency of accessing specific types of file, learners interest in specific material can be identified.

INPUT: A finite set of Learners L = L_1 , L_2 .....L_N ,
Sessions S =S_1 ,S_2 .....S_Q ,
PageURL P=P_1 , P_2 .....P_R
and FileURL F=F_1 , F_2 .....F_X
OUTPUT: XML File

initialize $Session_{Time} \leftarrow 0$, $SessionLog_{ID} \leftarrow 0$,
$SessionPageLog_{ID} \leftarrow 0$, $SessionFileLog_{ID} \leftarrow 0$, $IsPage \leftarrow 0$
for *each Sessions* S_j where $j \leftarrow 1$ **to** Q **do**
  compute $Session_{Time} = Start_{Session} - End_{Session}$
  if "$Session_{Time}$" $> 29min$ **then**
    get $Learner_{ID}$, $SessionLog_{ID}$ from *SessionLog* file
  end if
end for
for *each Session* S_j where $j \leftarrow 1$ **to** Q **do**
  for *each PageURL* P_k where $k \leftarrow 1$ **to** R **do**
    get $SessionPageLog_{ID}$ from *PageLog* file
    if $SessionLog_{ID} == SessionPageLog_{ID}$ **then**
      get $Learner_{ID}$, *PageURL*, $Page_{ID}$, *LogTime* from *PageLog* file
      set $IsPage \leftarrow TRUE$
    end if
  end for
  if $IsPage ==' TRUE'$ **then**
    for *each PageURL* P_k where $k \leftarrow 1$ **to** R **do**
      for *each Learner* L_i where $i \leftarrow 1$ **to** N **do**
        if "*PageURL* "*is accessed* **then**
          create *XMLTag* for $Session_{ID}$,$Learner_{ID}$,*PageURL*
          set $Session_{ID}$,$Learner_{ID}$,*PageURL* in "*UserActivities* "table
        end if
      end for
    end for
  end if
end for
for *each Session* S_j where $j \leftarrow 1$ **to** Q **do**
  for *each FileURL* F_d where $d \leftarrow 1$ **to** X **do**
    get $SessionFileLog_{ID}$ from *FileLog* file
    if $SessionLog_{ID} == SessionFileLog_{ID}$ **then**
      get $Learner_{ID}$, *FileURL*, $File_{ID}$, *LogTime* from *FileLog* file
      set $IsPage \leftarrow' FALSE'$
    end if
  end for
  if $IsPage ==' FALSE$ **then**
    for *each FileURL* F_d where $d \leftarrow 1$ **to** X **do**
      for *each Learner* L_i where $i \leftarrow 1$ **to** N **do**
        if "*FileURL* "*is accessed* **then**
          create *XMLTag* for $Session_{ID}$,$Learner_{ID}$,*FileURL*
          set $Session_{ID}$,$Learner_{ID}$,*FileURL* in "*UserActivities* "table
        end if
      end for
    end for
  end if
end for
for *each Learner* L_i where $i \leftarrow 1$ **to** N **do**
  set *UserActivities Orderby Date* and *Time*
end for

Algorithm 2: XML File2 Generation Algorithm

3. **Analysis done from Database: Number of times Learners Accessed Different Topics**
   Fig.4 shows the number of times the learners accessed the topics. This analysis gives the interest in specific topic and requirement of providing good material on different topics. The analysis can be further be captured into accessing the topics as per learner's requirement. E.g. how many learners accessed previous concept or advanced concept topic after accessing main topic.

4. **Number of times Learners Accessed Different Pages on Portal**
   Fig.5 shows that learners have not only accessed TopicSearch to access different files but also accessed other pages such as exercise and announcement pages. This analysis is important to understand the behavior of learner's on the portal.

5. **Time spent on Portal by Learners**
   Fig.6 shows that how much time is spent by learners on the portal in all their sessions. This analysis is important to identify the sequences of each learners based on their session ids and session time for pre-processing.

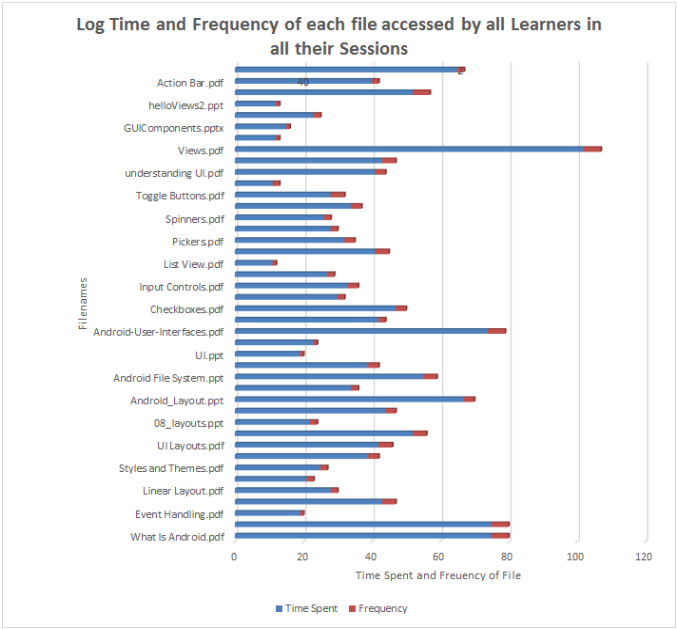


Figure 2. Frequency and Time Spent on Specific File by Learners

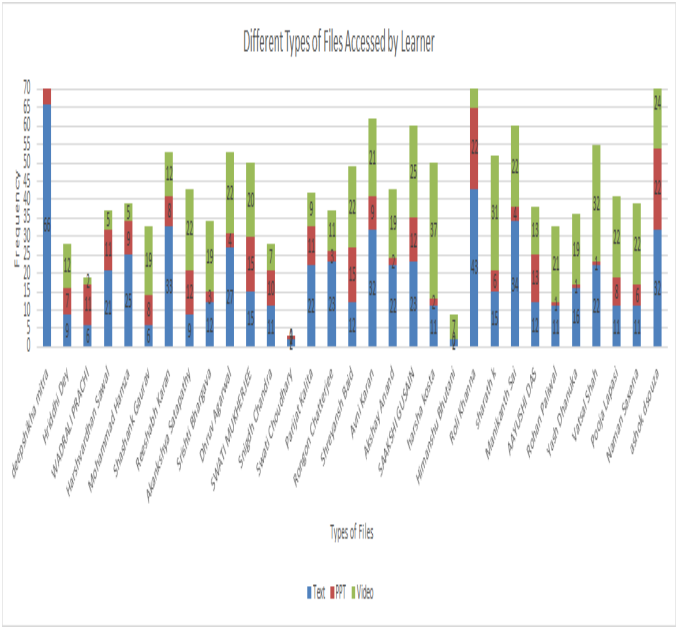


Figure 3. Number of Times Learners Accessed Specific Type of File

5.2. Pre-processing of Captured Usage Data

The application which generates both XML files is implemented using Microsoft Visual Studio 2010 and Microsoft SQL server 2005. The application fetches session data of each learner from database and IIS log files The screen-shot of page table and file table is shown in Fig.7

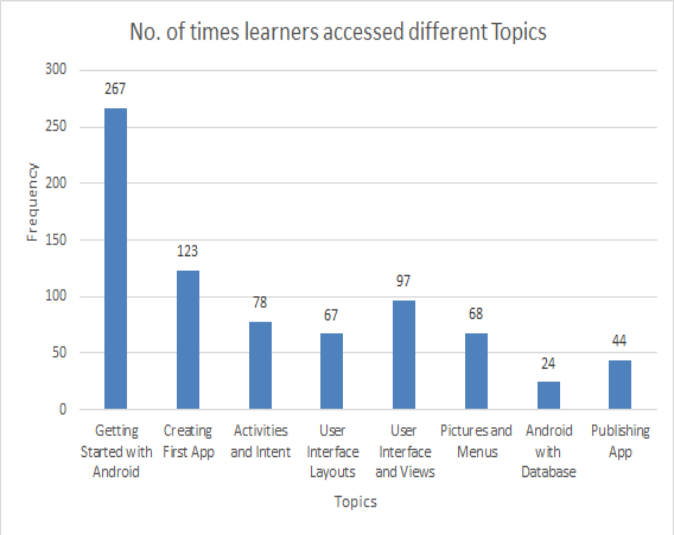


Figure 4. Analysis of Number of Times Learners Accessed Different Topics

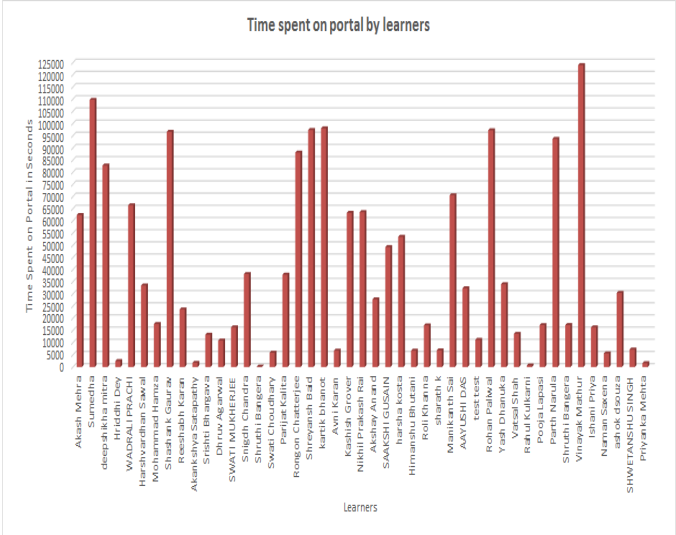


Figure 6. Analysis of Number of Times Learners Accessed Different Pages

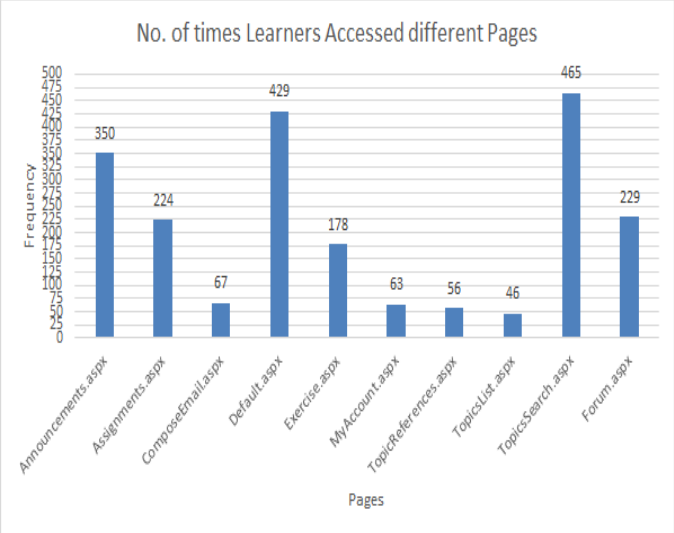


Figure 5. Analysis of Number of Times Learners Accessed Different Pages

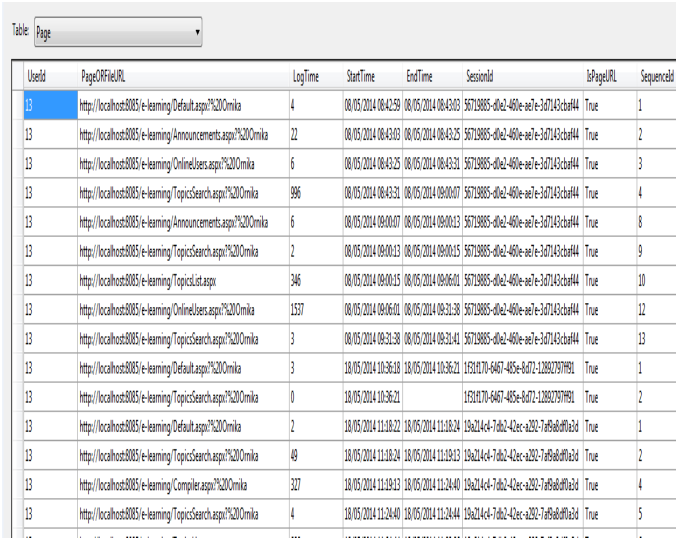


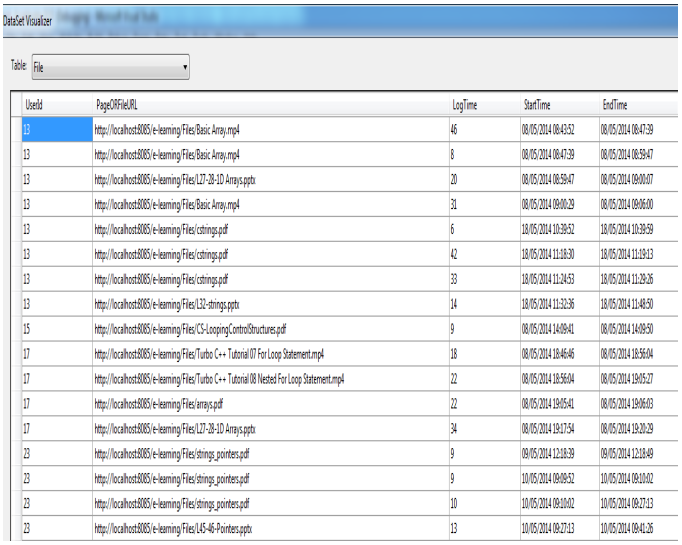
Figure 7. Page Table

and 8 respectively which is used to generate XML files for Sequence Data. The resultant XML files can be saved on the disk. The XML file generates different tags for each record which will be easy for understanding. The XML file 1 and file 2 shown in Fig.9 and 10 respectively. In second XML file each learner is having unique SessionId with SequenceId (it is sequence of activity number) and each learner can have multiple SessionIds with SequenceId.

5.3. Sequence file generation from XML Data

The generated XML files are used to create two sequence files for Clustering Algorithm. The first sequence file is as shown in Fig.11 is the patterns of files and pages accessed by each learner in each session. The second sequence file is as shown in Fig.12 is used to store the time spent on each file and page which is accessed by each learner in the session. The formats of sequence files are as below:

1. Sequence id, Session id, Learner Name, Page Name/File Name



DataSet Visualizer

Table: File

UserId	PageORFileURL	LogTime	StartTime	EndTime
13	http://localhost:8085/e-learning/Files/Basic Array.mp4	46	08/05/2014 08:43:52	08/05/2014 08:47:39
13	http://localhost:8085/e-learning/Files/Basic Array.mp4	8	08/05/2014 08:47:39	08/05/2014 08:59:47
13	http://localhost:8085/e-learning/Files/L27-2D-1D Arrays.pptx	20	08/05/2014 08:59:47	08/05/2014 09:00:07
13	http://localhost:8085/e-learning/Files/Basic Array.mp4	31	08/05/2014 09:00:39	08/05/2014 09:06:00
13	http://localhost:8085/e-learning/Files/cstrings.pdf	6	18/05/2014 10:39:52	18/05/2014 10:39:59
13	http://localhost:8085/e-learning/Files/cstrings.pdf	42	18/05/2014 11:18:30	18/05/2014 11:19:13
13	http://localhost:8085/e-learning/Files/cstrings.pdf	33	18/05/2014 11:24:53	18/05/2014 11:29:26
13	http://localhost:8085/e-learning/Files/L32-strings.pptx	14	18/05/2014 11:32:36	18/05/2014 11:48:50
15	http://localhost:8085/e-learning/Files/CS-LoopingControlStructures.pdf	9	08/05/2014 14:09:41	08/05/2014 14:09:50
17	http://localhost:8085/e-learning/Files/Turbo C++ Tutorial 07 For Loop Statement.mp4	18	08/05/2014 18:46:46	08/05/2014 18:56:04
17	http://localhost:8085/e-learning/Files/Turbo C++ Tutorial 08 Nested For Loop Statement.mp4	22	08/05/2014 18:56:04	08/05/2014 19:05:27
17	http://localhost:8085/e-learning/Files/arrays.pdf	22	08/05/2014 19:05:41	08/05/2014 19:06:03
17	http://localhost:8085/e-learning/Files/L27-2D-1D Arrays.pptx	34	08/05/2014 19:17:54	08/05/2014 19:20:29
23	http://localhost:8085/e-learning/Files/strings_pointers.pdf	9	09/05/2014 12:18:39	09/05/2014 12:18:49
23	http://localhost:8085/e-learning/Files/strings_pointers.pdf	9	10/05/2014 09:09:52	10/05/2014 09:10:02
23	http://localhost:8085/e-learning/Files/strings_pointers.pdf	10	10/05/2014 09:10:02	10/05/2014 09:27:13
23	http://localhost:8085/e-learning/Files/L45-46-Pointers.pptx	13	10/05/2014 09:27:13	10/05/2014 09:41:26

Figure 8. File Table

```
<?xml version="1.0"?>
  <File>
    <UserId>7</UserId>
    <FileURL>http://mitelearning.com/Files/Android Installation & Configuration.pdf</FileURL>
    <LogTime>2</LogTime>
    <Count>1</Count>
  </File>
  <File>
    <UserId>7</UserId>
    <FileURL>http://mitelearning.com/Files/1 Hello World Using Android Adt Tools.m4v</FileURL>
    <LogTime>8</LogTime>
    <Count>2</Count>
  </File>
  <File>
    <UserId>7</UserId>
    <FileURL>http://mitelearning.com/Files/Android Overview.ppt</FileURL>
    <LogTime>3</LogTime>
    <Count>2</Count>
  </File>
  <Page>
    <UserId>7</UserId>
    <PageURL>http://mitelearning.com/Default.aspx?%20deepshikha</PageURL>
    <LogTime>6</LogTime>
    <Count>3</Count>
  </Page>
  <Page>
    <UserId>7</UserId>
    <PageURL>http://mitelearning.com/TopicsSearch.aspx?%20deepshikha</PageURL>
    <LogTime>34</LogTime>
    <Count>5</Count>
  </Page>
  <Page>
```

Figure 9. XML File 1

```
<?xml version="1.0"?>
<UserLogs>
  <Page>
    <UserId>1</UserId>
    <PageORFile>/Default.aspx</PageORFile>
    <LogTime>1</LogTime>
    <SessionId>56719885-d0e2-460e-ae7e-3d7143cbaf44</SessionId>
    <IsPageURL>True</IsPageURL>
    <SequenceId>1</SequenceId>
  </Page>
  <Page>
    <UserId>1</UserId>
    <PageORFile>/TopicsSearch.aspx</PageORFile>
    <LogTime>2</LogTime>
    <StartTime>08/05/2014 08:43:03</StartTime>
    <EndTime>08/05/2014 08:43:25</EndTime>
    <SessionId>56719885-d0e2-460e-ae7e-3d7143cbaf44</SessionId>
    <IsPageURL>True</IsPageURL>
    <SequenceId>1</SequenceId>
  </Page>
  <Page>
    <UserId>1</UserId>
    <PageORFile>/Announcements.aspx</PageORFile>
    <LogTime>5</LogTime>
    <SessionId>56719885-d0e2-460e-ae7e-3d7143cbaf44</SessionId>
    <IsPageURL>True</IsPageURL>
    <SequenceId>1</SequenceId>
  </Page>
  <File>
    <UserId>1</UserId>
    <PageORFile>Android Installation & Configuration.pdf</PageORFile>
    <LogTime>7</LogTime>
```

Figure 10. XML File 2

2. Sequence id, Session id, Learner Name, Page Name, Time Spent on Page / File Name, Time Spent on file

Sequence id= { S_1,S_2,S_3, *upto* M } M= No. of sequences
Session id= { G_1,G_2,G_3, *upto* K } K=No. of sessions per Learner
Learner id= { u_1,u_2,u_3, *upto* N } N=No. of Learners
Page id= { p_1,p_2, *upto* P } P=No. of Pages
File id= { 1,2,3,4, *upto* F } F=No. of Files
Time Spent is in Minutes.

```
S1  G1  u1  p5  p7  p5  p10 p1  p4  p3  p5  p3  p3  p5  p10 p5  p4  p1  p10 p5  p10
S2  G1  u2  p5  p10 p9  p10 p10 p1  p10 p8  p9  p10 p5  p7  p10 p9  9   p9  8   p9
S3  G1  u7  p5  p7  p1  p10 p4  p2  p7  p10 p5  p1  p7  p10 p7  p1  p10 p2  p10 13
S4  G1  u8  p5  p10 13  p7  p2  p10 p2  p7  p10 21  26  p2  p7
S5  G1  u9  p5  p10 p1  p4  p2  p7  p7  p1  p10 p6  p10 p4  p5  p1  p1  p5  p1  p5
S6  G1  u10 p5  p10 13  p10 20  124 21  22  23  24  25  125 126
S7  G1  u11 p5  p1  p5  p7  p10 p1  p4  p2  p5  p10 13  p5  p5  p7  p10 p10 14  p5
S8  G1  u12 p5  p7  p10 p2  p4  p1  p5  p5  p5  p10 p10 p5  p10 123
S9  G1  u13 p5  p7  p10 p1  p4  p2  p10 13  13  14  20
S10 G2  u1  p5  p1  p3  p4  p3  p5  p4  p5  p5  p4  p5  p10 9   123 p8  p1  p4  p5
S11 G1  u14 p5  p10 p1  p4  p2  p7  p5  p10 13  13  14  19  p2
S12 G1  u15 p5  p2  p10 p5  p10 p5  p1  p10 p1  p5  p10 p5  p10 p1  p10 20  p5  p1
S13 G1  u18 p5  p10 p5  p1  p10 13  14  15  16  17  18  19  20  21  p5  p10 123 20
S14 G1  u20 p5  p10 13  20  p5  p5  p7  p10 p1
S15 G1  u23 p5  p7  p10 p1  p4  p5  p5  p10 p6  p8  p10 13  13  p5  p10 15  16  18
S16 G1  u16 p5  p7  p10 p1  p4  p2  p10 13  p8  p5  p1  p7  p10 14  14  123 p5  p10
S17 G2  u11 p5  p10 20  21  123 125 22  126 129 p5  p7  p10
S18 G1  u24 p5  p10 p6  p8  p5  p1  p10 p4  p10 p5
S19 G1  u17 p5  p2  p10 p5  p10 13  13  14  14  15  16  17  18  18  19  20  21  22
S20 G2  u2  p5  p10 p7  p4  p1  p2  p4  p5  p5  p10 p5  p10 p4  p5  p1  p2  p10 p6
S21 G1  u19 p5  p1  p5  p7  p10 p9  p2
S22 G1  u21 p5  p10 13  20  21  22  p5  p7  p1  p10 p2  p5  p10 p1  p10 22  123 123
S23 G1  u22 p5  p10 13  p8  p10 p2  p10 p8  p7  p1  p5  p2  p10 p5  p1  p4  p1  p10
S24 G2  u9  p5  p1  p10 p5  p1  p2  p10 13  p5  p10 p5  p10 p5  p10 p1  p2  p7  p5
S25 G1  u25 p5  p5  p10 p6  p10 p9  36  p9  p10 p9  p1  p5  p7  p7  p10 p9  p5  p5
S26 G1  u27 p5  p5  p7  p10 n1  p4  p2  p10
```

Figure 11. Sequence File 1

```
S1  G1  u1  1   2   1   1   1   1   2   1   7   1   1   1   1   3   4   1   1   1
S2  G1  u2  1   1   1   1   1   1   1   1   1   1   3   1   1   12  2   6   2
S3  G1  u7  1   3   4   1   1   2   2   1   1   4   4   1   4   3   1   8   1   8
S4  G1  u8  1   1   5   2   2   1   4   2   1   7   7   7   4
S5  G1  u9  1   1   3   3   3   1   6   3   1   6   2   1   1   2   4   1   6   1
S6  G1  u10 1   1   4   1   8   12  14  9   8   9   8   9   8
S7  G1  u11 1   3   1   2   1   2   2   5   1   1   8   1   1   4   1   1   8   1
S8  G1  u12 1   2   1   2   1   3   1   1   1   1   1   1   9
S9  G1  u13 1   2   1   3   1   4   1   9   8   7   9
S10 G2  u1  1   2   2   1   1   1   2   1   1   1   1   1   9   10  8   10  4   1
S11 G1  u14 1   1   3   1   2   1   1   1   6   10  8   8   5
S12 G1  u15 1   2   1   1   1   1   2   1   3   1   1   1   1   4   1   11  1   3
S13 G1  u18 1   1   1   4   1   5   10  9   12  7   9   9   6   8   1   1   6   6
S14 G1  u20 1   1   5   6   1   1   2   2   3
S15 G1  u23 1   2   1   2   1   1   1   1   6   1   1   10  8   1   1   14  7   6
S16 G1  u16 1   2   1   3   1   6   1   9   1   1   3   2   1   9   12  12  1   1
S17 G2  u11 1   1   6   6   6   8   13  8   10  1   2   1
S18 G1  u24 1   2   1   1   1   2   1   2   1   1
S19 G1  u17 1   3   1   1   1   4   15  9   13  12  9   17  9   12  16  11  6   8
S20 G2  u2  1   1   2   1   2   1   2   1   1   1   1   1   3   1   5   6   1   3
S21 G1  u19 1   4   1   2   1   1   3
S22 G1  u21 1   1   7   7   8   7   1   2   2   1   4   1   1   3   1   7   8   8
S23 G1  u22 1   1   7   1   1   2   2   1   1   4   1   4   1   1   3   2   3   2
S24 G2  u9  1   4   1   1   2   2   1   9   1   1   1   1   1   5   3   4   1
S25 G1  u25 1   1   1   3   1   1   13  1   1   1   3   1   3   2   1   4   1   1
S26 G1  u27 2   1   2   1   2   1   6   1
```

Figure 12. Sequence File 2

6. Conclusion and Future Work

The use of personalized e-learning and adaptive e-learning system has become increasingly important in recent years with extensive research being devoted to finding different ways of tailoring the learning experience for individual students. The proposed work in this paper describes the framework for AES which is useful to provide Adaptive User Interface to learners based on their learning styles. The proposed work has been achieved by developing prototype of e-learning portal with different types of learning components. The portal has been used to capture the usage data of learners through log files and database. The method of capturing the learning styles comprises not only IIS log files but also database entries where important factors of learning

styles are captured. The usage data is useful to identify the learning styles of the learners according to FSLSM. Four dimensions such as Pre-processing, Perception, Input and Understanding and eight categories of mentioned dimensions like Active/Reflective, Sensing/Intuitive, Visual/Verbal and Sequential/Global are used. The proposed work focuses on grouping the session details obtained from different log files. Different algorithms have been implemented to analyze the session log details for different types of learners. The method of capturing the learning styles comprises IIS log files and database entries where important parameters of learning styles are captured. The learner's session has been considered as the total number of learning objects accessed by that learner. The captured database log data consist of the details related to pages and files as per unique session identifier allotted to a learner. To get the clear analysis of pre-processed log data the different algorithms are implemented and results are obtained. Web log pre-processing is one of the major concern in Web Usage Mining. In this work pre-processing of web log data is done based on different constraints that help to identify proper learner's sessions. The factors considered are the Log Time and Frequency for identifying the interest of a leaner. The web log data has been analyzed based on the FSLSM dimensions and mapped into learning objects. In pre-processing, the log data has been converted into the standard XML format for creating sequence files of usage patterns. These sequence files can be directly used for clustering of learner's profiles in order to understand types of learners. In future, to incorporate the use of learning styles into online courses, we plan to encapsulate the defined approach into any web based e-learning systems. The encapsulated approach will be useful to generate adaptive user interface components in order to achieve personalization and adaptation for learners.

References

[1] AHMAD, N.B.H. and SHAMSUDDIN, S.M. (2010) A comparative analysis of mining techniques for automatic detection of student's learning style. In *Intelligent Systems Design and Applications (ISDA), 2010 10th International Conference on* (IEEE): 877–882.

[2] ALIAN, M. and AL-AKHRAS, M. (2010) Adalearn: an adaptive e-learning environment. In *Proceedings of the 1st International Conference on Intelligent Semantic Web-Services and Applications* (ACM): 21.

[3] BOUSBIA, N., REBAÏ, I., LABAT, J.M. and BALLA, A. (2010) Analysing the relationship between learning styles and navigation behaviour in web-based educational system. *Knowledge Management & E-Learning: An International Journal (KM&EL)* **2**(4): 400–421.

[4] CHITRAA, V. and THANAMANI, A.S. (2013) Recommendation of web pages for online users using web log data. *International Journal of Science and Research (IJSR), India ISSN* : 2319–7064.

[5] DORÇA, F.A., LIMA, L.V., FERNANDES, M.A. and LOPES, C.R. (2013) Comparing strategies for modeling students learning styles through reinforcement learning in adaptive and intelligent educational systems: An experimental analysis. *Expert Systems with Applications* **40**(6): 2092–2101.

[6] DORÇA, F.A., LIMA, L.V., FERNANDES, M.A. and LOPES, C.R. (2013) A new approach to discover students learning styles in adaptive educational systems. *Revista Brasileira de Informática na Educação* **21**(01): 76.

[7] FELDER, R.M. and SILVERMAN, L.K. (1988) Learning and teaching styles in engineering education. *Engineering education* **78**(7): 674–681.

[8] FRANZONI, A.L., ASSAR, S., DEFUDE, B. and ROJAS, J. (2008) Student learning styles adaptation method based on teaching strategies and electronic media. In *Advanced Learning Technologies, 2008. ICALT'08. Eighth IEEE International Conference on* (IEEE): 778–782.

[9] GRAF, S., LIU, T.C., CHEN, N.S., YANG, S.J. *et al.* (2009) Learning styles and cognitive traits–their relationship and its benefits in web-based educational systems. *Computers in Human Behavior* **25**(6): 1280–1289.

[10] GRAF, S., LIU, T.C. *et al.* (2009) Supporting teachers in identifying students' learning styles in learning management systems: an automatic student modelling approach. *Educational Technology & Society* **12**(4): 3.

[11] JAFARI, M., SOLEYMANISABZCHI, F. and JAMALI, S. (2013) Extracting user's navigational behavior from web log data: a survey. *Journal of Computer Sciences and Applications* **1**(3): 39–45.

[12] KHRIBI, M.K., JEMNI, M., NASRAOUI, O., GRAF, S. *et al.* (2013) Toward a fully automatic learner modeling based on web usage mining with respect to educational preferences and learning styles. In *Advanced Learning Technologies (ICALT), 2013 IEEE 13th International Conference on* (IEEE): 403–407.

[13] LAKSHMI, N., RAO, R.S. and REDDY, S.S. (2013) An overview of preprocessing on web log data for web usage analysis. *International Journal of Computer Applications. India* **2**(4): 274–279.

[14] MAHAJAN, R., SODHI, J. and MAHAJAN, V. (2014) Usage patterns discovery from a web log in an indian e-learning site: A case study. *Education and Information Technologies* : 1–26.

[15] PHAM, Q.D. and FLOREA, A.M. () A method for detection of learning styles in learning management systems .

[16] PIERRAKOS, D., PALIOURAS, G., PAPATHEODOROU, C. and SPYROPOULOS, C.D. (2003) Web usage mining as a tool for personalization: A survey. *User modeling and user-adapted interaction* **13**(4): 311–372.

[17] POPESCU, E., BADICA, C. and MORARET, L. (2010) Accommodating learning styles in an adaptive educational system. *Informatica (Slovenia)* **34**(4): 451–462.

[18] SANDERS, D.A. and BERGASA-SUSO, J. (2010) Inferring learning style from the way students interact with a computer user interface and the www. *Education, IEEE Transactions on* **53**(4): 613–620.

[19] SUNEETHA, K. and KRISHNAMOORTHI, R. (2009) Identifying user behavior by analyzing web server access log file. *IJCSNS International Journal of Computer Science and Network Security* **9**(4): 327–332.

[20] WATSON, C., LI, F.W. and LAU, R.W. (2010) A pedagogical interface for authoring adaptive e-learning courses. In *Proceedings of the second ACM international workshop on Multimedia technologies for distance leaning* (ACM): 13–18.