ISSN (Print) : 0974-6846
Indian Journal of Science and Technology, *Vol 9(10), DOI: 10.17485/ijst/2016/v9i10/88899, March 2016* ISSN (Online) : 0974-5645

Data Preprocessing and Cleansing in Web Log on Ontology for Enhanced Decision Making

F. Mary Harin Fernandez[1*] and R. Ponnusamy[2]

[1]Department of Computer Science and Engineering, Sathyabama University, Jeppiaar Nagar, Chennai – 600119, Tamil Nadu, India; mary.fherin@gmail.com
[2]Rajiv Gandhi College of Engineering, Chennai – 602105, Tamil Nadu, India; r_ponnusamy@hotmail.com

Abstract

Background/Objectives: Web applications are growing at a massive promptness and its users rise at exponential speed. The deviations in technology have made it potential to capture the user's essence and interactions with web applications through web log file. **Methods/Statistical Analysis:** Due to huge amount of extraneous data in the web log, the original log file cannot be openly used in the Web Usage Mining (WUM) system. The web server log files used for mining several expedient patterns to analyze the access behavior of the user. Data pre-processing plays a dynamic role in Data Mining. **Findings:** We propose a new web log mining method for determining web access procedure from interpreted web usage logs which integrates data on user behaviors through self-rating and communication tracking. The raw data is pre-processed in order to improve the quality of data to be extracted. We discuss the significance of information pre-processing method and abundant steps elaborate in receiving the essential information effectively. This pre-processing method used to process and analysis the web log data for extraction of user search patterns. The fuzzy association rule minimizes users' exploration period and facilitate for enhance decision making. **Applications/Improvements:** The proposed data cleaning method removes the extraneous records from web log. With this progression, we create a Personalized Ontology, which assists various semantic web applications such as site perfection, business intelligence and recommendations for behavioral analysis. Finally, we illustrate the efficiency of this method by the experimental results in the framework of data pre-processing and cleansing extraneous data for personalized ontology.

Keywords: Behavioral Tracking, Data Preprocessing, Knowledge Discovery, Ontology Creation, Weblog Mining

1. Introduction

Web log mining is one of the applications of data mining technique which determine requested page patterns from web log data, in order to better comprehend and serve the essentials of web-based solicitations[1]. Numerous applications similar to personalization, e-commerce, recommender systems, web site designing, and learning from web pages are erected professionally by knowing users tracking information through web[2]. The first step in web usage mining is the data gathering from web page. It involves of collecting the pertinent web data. Data source can be gathered from server, proxy server, client or acquire from a personalized database, which encompasses business or analyzed web data[3].

Once the user log on to the system each clicks of a user is saved in a web server log. The pre-processing is a major method which extracts fields from the web log cleansed irrelevant data, apply pattern discovery and analyses those discovered data, and identify transaction and path completion for mining the web log. We track user's behavior by transaction identification and path completion phase in the preprocessing method. In file extraction phase we classify patterns from the URL and store it in a database to apply various analyses on those classified data.

* *Author for correspondence*

We develop Personalized Library Ontology as a user's reference search pages. The user formerly their pursuit on ontology, they must register in a profile on order to track user's behavior. Once the user register in a profile, before their search the relevant web pages will be displayed to the user. This process diminishes the user's search time. The advantages of proposed system is to automatically track user's every clicking behavior, retrieve relevant or exact data and view significant information in analyzed form for knowledge detection from web log and services.

1.1 An Introduction to Web Log Mining

Web log mining distillates on cleansing data, which eliminates all, searched content topics and error in HTTP and save only the URL information. The pre-processing level detects user ID and session ID which is not analyzed for further behavior of the user. The existing web mining notion determines preprocessing technique which is not applied to any application databases.

Srivastava J, Cooley R, Deshpande M and Tan P-N proposed approaches for session identification, user identification, path completion, and periodic identification they provide heuristics methods to compact with the troubles throughout data preprocessing[4]. Fang Yuan, Li-Juan Wang, Ge Yu, they concentrated on analyzing log data in order to mine usage pattern. The usage patterns are classified in to relevant page group, similar user group, and frequency accessing tracks[5]. Maristella Agosti, Giorgio Maria Di Nunzio, provide detail description of user session with the initial outcomes on diverse phases, number of requested URL session and its length and timestamp[6].

A-Nikos Koutsoupias, proposed a statistical analysation technique and a tool for analysis log file and improved kind of preprocessing the web log information[7]. Zhang Huiying, Liang Wei, proposed an intelligent process for information preprocessing in web custom mining[8]. Tsuyoshi Murata and Kota Saito defined a technique for explicatory users' comforts based on an exploration of the site-keyword diagram. They proposed extracting sub graphs from web log according to the users' interest[9].

V.V.R. Maheswara Rao and V. Vallikumari**,** proposed a learning algorithm to isolate search engine accesses logically and human user in reduced duration[10]. Martin Arlitt, well defined the characterization of user sessions with their different perceptive on web pages such as path definition and the entire detail about the sessions[11]. Tanasa, D. and Trousse B, they combine several web server logs of the similar institute and afford an elucidation to intersect entire log files and rebuild the visit. Data summarization is done only on interested data and analyzed those stored information[12].

Kohavi R, mines E-Commerce Data, and explains the correlated problem happens in the web log and presents measurements as good, bad, and ugly to daze the issues in the web log[13]. Show-Jane Yen, Yue-Shi Lee and Min-Chi Hsieh, presents a competent forwarding web track design mining algorithm. The mining period can be concentrated they used

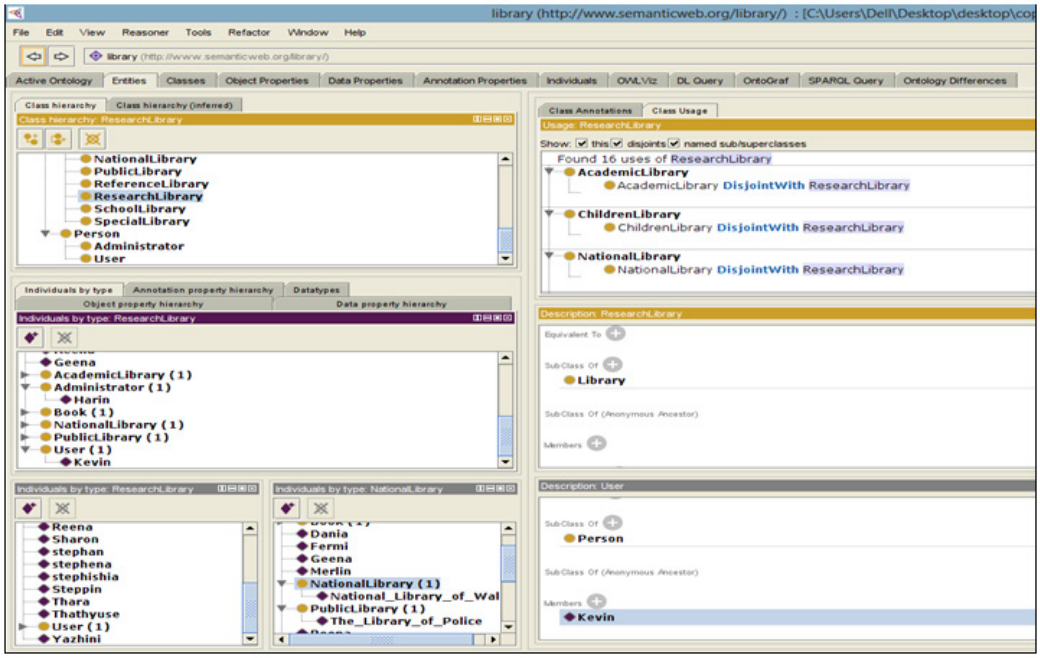

Figure 1. The personalized library ontology using protégé editor 4.3.

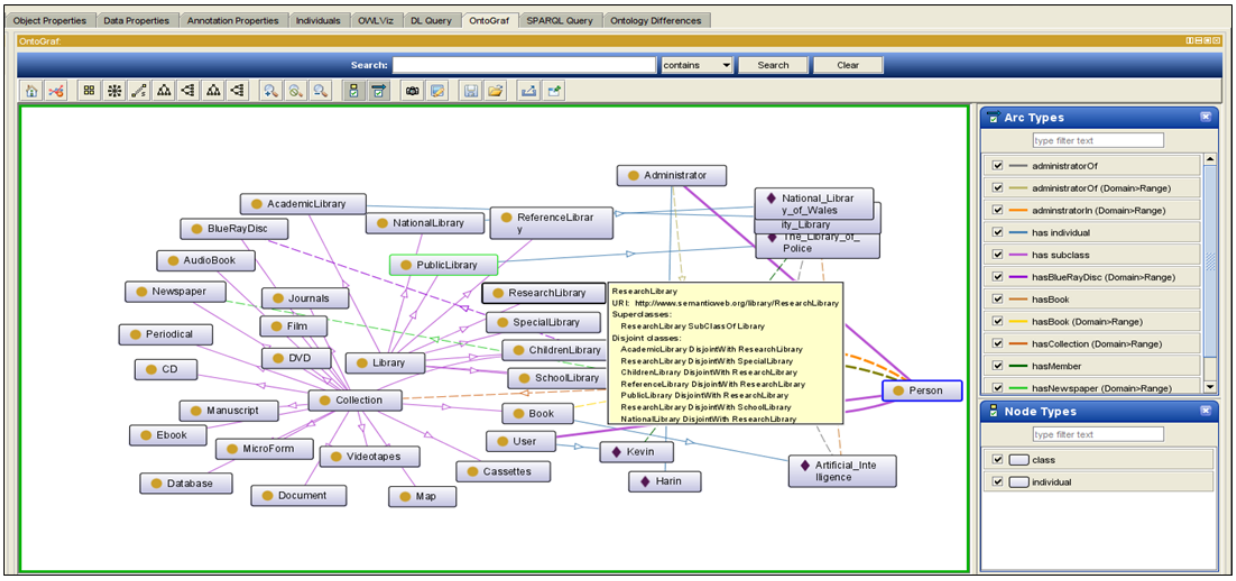

Figure 2. The graphical illustration of classes, subclasses, individuals and their relationship.

existing mining outcome and proposed a new patterns from an inserted and deleted portion of web log[14]. Bettina Berendt, Bamshad Mobasher, Myra Spiliopoulou, Jim Wiltshire, compared referrer based and time based heuristics for visit modernization. They proposed a heuristic method which depends on the Web site design and visit's length[15].

2. Personalized Library Ontology

The personalized Library ontology is generated by using Top-Down loom with the protégé ontology editor tool 4.3. The several Libraries' information is collected and the required information for developing ontology is extracted as classes, individuals and properties. Figure 1 illustrates Library ontology here the classes and individuals' hierarchy are displayed in leftmost side and the class description and their usage are shown in rightmost side of the screen shot[16,17]. We develop each class in library ontology with the corresponding properties, individuals, restrictions and associations among objects[18].

Figure 2 illustrates sample graphical representation of classes, subclasses and their individuals. The relationships among classes, the members of subclasses and individuals are shown at rightmost side of the screen shot. The main Library class has subclasses of Academic Library, Research Library, Children Library, National Library, Public Library, and etc. each subclasses has its object property, data property and individuals.

The personalized Library Ontology is updated by on model in Jena API. We use the following algorithm for automatic updates on ontology[19]. The users' information is automatically updated on Library ontology as data type property (Datatype Prop), object property (ObjectProp) and individuals. The users exploit natural language to retrieve data from library ontology. The user query in natural language is then converted into consequent SPARQL query[20]. This SPARQL Query is directly used to retrieve information from ontology[21].

Input: user data
Output: Ontology updated with user data
userUpdatorAlgo( )
1: create UserClass data
2: **UserClass**← **newUser**
3: If newUser already exist **then**
4: Report error msg
5: **Else**
6: **newUser**←**createIndividual** for user class
7: If (DatatypeProp and ObjectProp) **exist**
8: get DatatypeProp → dataPropLink, dataPropVal and ObjectProp → objPropLink, ObjPropVal
9: Else
10: newUser → **getRange** (DatatypeProp and ObjectProp)
11: dataPropLink =getDatatypeProp (datatypeProperty [k])
12:objPropLink =getObjectProp (objectProperty[k])
13: **k++**
14: **End If**

15: **set**DatatypeProperty(newUser, dataPropLink, dataPropVal)
16: set ObjectProperty(newUser, objPropLink, ObjPropVal)
17: **End If**
18: Get LibName
19: Produce **ASK** query in **SPARQL**
20: Execute query
21: End the Process

The following algorithm automatically updates Library information on ontology. The subclasses to the main library class (**Library)** are created as **newLib** (new library) information. The **newLib** subclass is formed by the data type property (datatypeProp), object property (objProp) and individuals. If the **newLib** subclass already exists then we get an error massage. Using Jena API we show the improved result to the user.

Input : Library data
Output : Ontology updated with LIB data
LibUpdatorAlgo ( )
1: create Library subclass
2: **Library ← newLib**
3: If newLib already exist **then**
4: Report error msg
5: **Else**
6: newLib ← Create subclasses, datatypeProp, objProp and individuals
7: **Get** datatypeProp (name,type,dept,address, etc) of **newLib**
8: If current datatype Prop already exists **then**
9: Add reference to that datatypeProp
10: Else create **new** datatypeProp
11: End If
12: **Get** objProp (hasbook, hasvideo, hasaudio, etc) of **newLib**
13: If current objProp already exists **then**
14: Add reference to that objProp
15: **Else** create **new** objProp
16: **End If**
17: Get individuals of **newLib**
18: **If** current individuals already exist **then**
19: Add reference to that individual
20: **Else** create **new** individual
21: End If
22: End If

3. Proposed System Architecture

Generally, the input for the proposed method is users' web log URL files shown in Figure 3. We track user's each mouse click on web page during their exploration. The web log files contain unrefined information. The irrelevant and spontaneous data from web log files are isolated by using pre-processing and cleansing method. Then we analyze those processed data for user's enhanced decision making before they commit to search on library ontology.

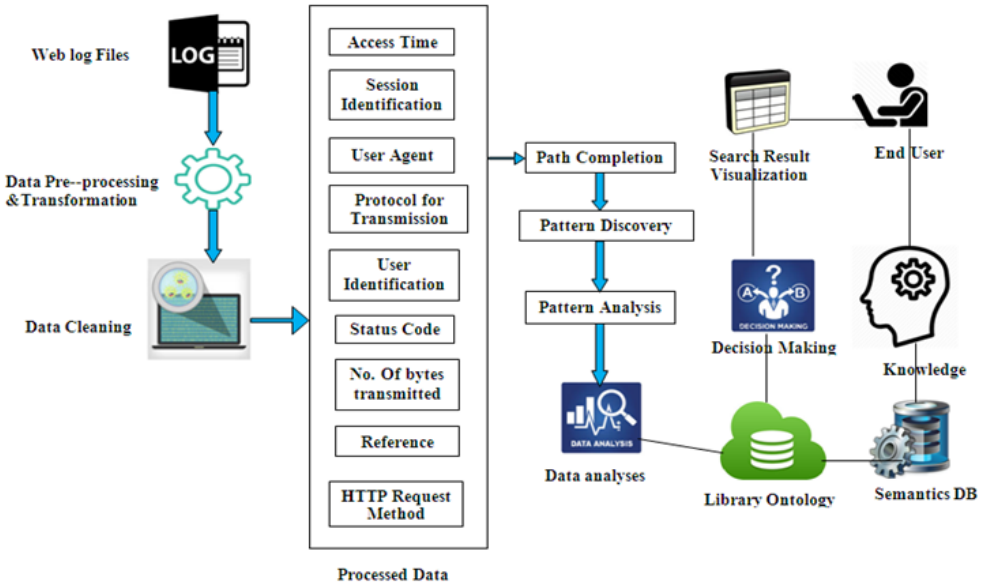

Figure 3. The proposed method architecture.

First step, pre-process raw log file, is the task of field mining, facts cleaning, user detection, session discovery, Transaction identification and path completion. Second step, we determine the patterns of mined file and apply analysis of those data for several user behavioural applications. Third step, the pertinent ranking and decision making notions are applied to each user's entries in order to diminish users' searching duration.

3.1 Web Log Mining

Web server log saves all clicks of the user's URL in the website. The mining method used to extract the URL saved in Web server log. We discover user's behaviour by mining web log URL[22]. Various analyses can be achieved in mined data. User can use various types of resources during their search. The web log file analysis algorithm identifies the type of record in weblog file. The splitter () method algorithm group the categories of data in the record for resource allocation. The mined information is saved in separate databases to analyze different type of files such as images, videos, css, txt, audio and etc.

Web Log File analysis Algorithm

1. Read log **Record R** from webLogFile
2. Split **R** by **Splitter ( )** method
3. Get **category** from **Splitter ( )** method
4. **If**(**category** !=NULL)
    Assign **group** based on **category**
    If group contains (gif or jpg or jpeg or png or bmp or txt or asp or css)
    Assign information into **imageFile**
    imageFile ←group
    Else If group contains (flv or avi or mpg or mpeg or mp4 or swf)
    Assign information into **videoFile**
    videoFile ←group
    Else If group contains (mp3 or wav or mid or rm or ram)
    Assign information into **audioFile**
    audioFile ←group
    End If

End If

5. **Concatenate** prefix and combine all divided **group** into table.
6. **End**

Splitter ( ) Method Algorithm

Input: Record R

Output: image [], video [], audio [], uniqueUser []

1. **category← Record R**
2. **If category** contains prefix
    Check prefix **category**
    If (prefix in image) **category**
    group←**imageFile**
    Else If (prefix in video) **category**
    group← **video File**
    Else If (prefix in audio) **category**
    group←**audioFile**
    End If
    End If
3. **If category** contains IP address
    Check for uniqueUser
    If operatingSystem and user's Browser are different
    Then group ← newUser
    Else user already exists.
4. **End If**

3.2 Pre-Processing and Transformation

First step, preprocess raw log file, is the task of field extraction, data cleaning, user identification, session identification, Transaction identification and path completion. Second step, we determine the patterns of extracted file and apply analysis of those data for several user behavioral applications. Third step, the pertinent ranking and decision making notions are applied to each user's entries in order to diminish users' searching duration. A Web log, encompassing Web server information, is formed as an outcome of the http method that is track on Web server log.

3.2.1 Field Mining

Field mining is the preliminary step for data pre-processing. Weblog clamps numerous fields such as access time, http request method, path of the source on the web server, status code, number of byte transmitted, user agent and referrer. To analysis user behaviour we exact data from log files. Each user entry is denoted as a single line of the web log record. We use two special characters 'comma'(,) and 'space' (" ") for mining these log record[23].

3.2.2 Data Cleaning

Web log has massive quantity of obligatory and superfluous records in the server log file[24]. The data cleaning phase used to separate extraneous and unsolicited data from the web log entries[25]. The web log has all type of user interaction

data in the record. This phase used to filter obligatory data from the unstructured log record. To analysis user behavior we store essential data in a separate database for resource allocation. The ineffective data such as status code failed data, timeout (user Id not in used) data, redundant data etc. are removed from the web log record.

3.2.3 User Identification

Identifying distinct user will help us to classify the type of user and their behaviour. Once weblog files are cleaned, next phase we identify the user's ID. In our approach users are identified by their IP address or user's registration profile. We detect distinct user by their browser and operating system. We assumed that each grouping of IP address/ Operating system/ Agent as a distinct user[26]. We also include the login facts to the log file to acquire username.

3.2.4 Session Identification and Ordering

A request made by a user to web server in the log entry is known as session. A single user may have distinct or numerous sessions in a while and each session are limited with the typical clicks recognized in definite period. The following two methods are widely used to identify user sessions, 1. Highest Forward Orientation transaction identification[27]. 2. Position Measurement Scheme. We set 30 minutes as a default timeout. When the user Id is not in User set above 30 minutes then it is understood that new user-session has started. The session Ordering is a heuristic to restructure the identified sessions. The authentic sequence of clicks stream is arranged in an ascending order in the database.

3.2.5 Transaction Identification

To identify users' behaviour we use transaction identification method. User visits number of web pages in a session. We track those (from the initial web pages to content) web pages and save it in a database. The historical web page in a database is analyzed for new user references. The two exiting methods maximum Forward Track and Frequent Access Track are used to calculate total number of web pages in a session and frequent access web pages for the new users' references which minimizes the search time.

- **Maximum Forward Track [MFT]**

The web pages in a session are tracked and stored in a database. From stored databases we summaries each user's search concept on web page URL[28]. The redundant web pages in the database for same user are also eliminated.

- **Frequent Access Track [FAT]**

Often used web pages for the same content are found from the URL and save it in a database. The tracked database which contains Frequent Access Path in a session is known as Support web page path for the new user's search. The FAT is stored in a database and displays for the new user after registering their profile and before their new search made in a web page[29].

3.2.6 Path Completion

The path completion or path support is a user search web page identity. User requested pages presented either by auxiliary path or directly of demanded path. We save complete path of each session's URLs to database and organize the reference path and requested path. This way detects omitted user clicks from one web page to another web page using back and onward button during their information pursuit on website which cannot originate in web server. We track intact web path that user search on website in a specified session to detect users' behavior and demonstrate these search data content to the pertinent search user to reduce their pursuit time which leads rigorous URL path to the new user. The succeeding illustration is a part of URL with summaries users' search history and classified reference path and auxiliary path.

The Table 1 illuminates as sample complete path

Table 1. Path completion samples from UR

#	IP Address	Method	Requested Path	Referred Path	Summarized Path
1	192.168.2.1	GET	Topic1.html	-	Topic 1→-
2	192.168.2.1	GET	Topic 2.html	-	Topic 2→ -
3	192.168.2.1	GET	Topic 3.html	Topic 1.html	Topic 1→ Topic 3→ -
4	192.168.2.1	GET	Topic 2.html	Topic 1.html	Topic 1→ Topic 2→ -
5	192.168.2.1	GET	Topic 4.html	Topic 3.html	Topic 1→ Topic 3→ Topic 4 → -
6	192.168.2.1	GET	Topic 5.html	Topic 4.html	Topic 1→ Topic 3→ Topic 4 → Topic 5 → -
7	192.168.2.1	GET	Topic 6.html	Topic 4.html	Topic 1→ Topic 3→ Topic 4 → Topic 6 → -
8	192.168.2.1	GET	Topic 4.html	Topic 7.html	Topic 7→ Topic 4 → -
9	192.168.2.1	GET	Topic 2.html	Topic 1.html	Topic 1→ Topic 2→ -

detection from URL. Here we determine the widespread search of a user's behavior in a session by their searching method. In the above example we identify number of user and summaries their search behavior.

3.3 Pattern Discovery

Pattern discovery uses a range of extracting notion in a pre-processed log data. Here we use Association rule and modify knuth morris pratt algorithm for pattern discovery and similarity check. The categories of image, video, audio, text, CSS and etc. are applied for pattern identical. Primarily it finds the length of the extracted log files and evaluates each adjacent type and returns significance in Boolean.

3.4 Pattern Analysis

We relate Pattern analyses for the pertinent data from pattern discovery phase. The user's pursuit interested web data are exposed and weighed for the prospect orientations to novel user. The user typically discovers hard in searching data on the entire web pages, instead they expect relevant data navigation before their search on the website. For user's concern we analysis the pertinent pattern data from the web log and visualize those analyzed results to the relevant user before their search on the web site. The following Figure 4, Figure 5, Figure 6 and Figure 7 illustrations shows different kinds of files analyzed from the web log. This information is mined from the web usage log in order to detect the type of resources used by the user during their web page search.

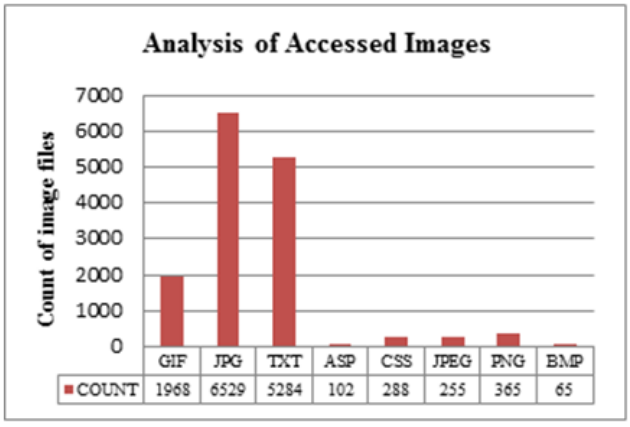

Figure 4. Image file analysis.

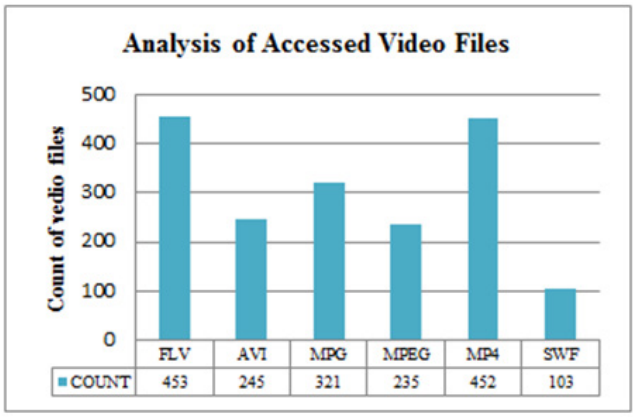

Figure 5. Video file analysis.

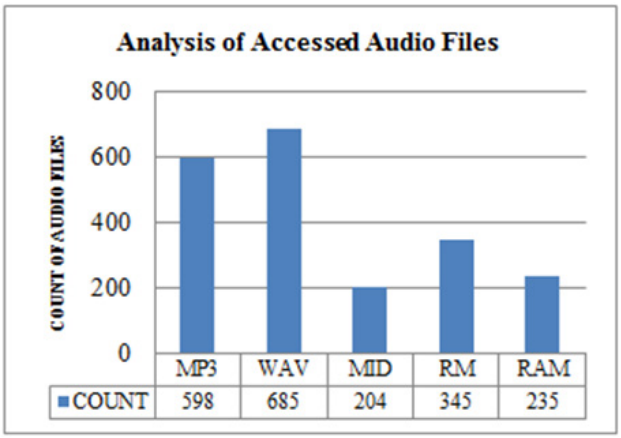

Figure 6. Audio file analysis.

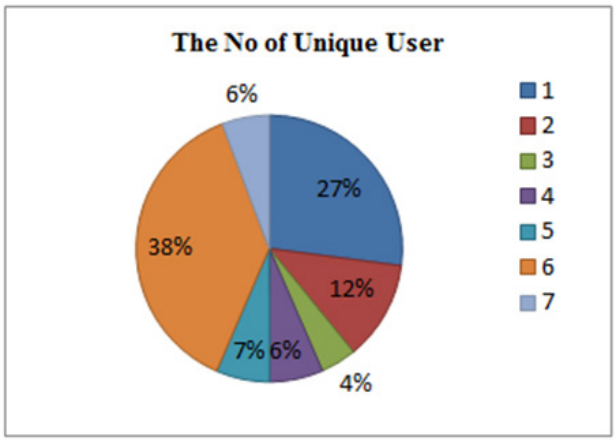

Figure 7. Unique user analysis.

3.5 Relevant Ranking and Decision Making

User ranks each web page during the search according

to their emotional impacts. This rating mechanism helps to present pertinent content to the novel user and diminishes pursuit period from huge web pages. Once the user registers their profile before their search, the requested relevant web pages automatically displayed to the user with the evaluated and analyzed data[30, 31]. The decision making concept describe episodic web access pages. The fuzzy association rule is used to detect repeatedly referred web ages[32,33].

Decision Making Algorithm

1: Identify set of Fc patternsets
2: Prompt together positive (P) and negative (N) for fuzzy association
3: Inset label M to patternsets X for every frequent access
4: Check for positive (P) and negative (N) association
If X ∪{M} ==Frequent
  If (X→M)
    P=Positive association rule
Endif
Elseif X ∪{M}! = Frequent
  If (¬X→M)
  N=Negative association rule
  Endif
Endif
5: Check any patternsets falls between X and M, then
Verify Support and Condition threshold with the following Association ¬X∪M→y and Association X∪¬M→y.

We isolate frequent access pattern sets and describe optimistic and undesirable links by relating fuzzy association rules to the pattern sets[34,35]. User renders their decision to the positive association, which describes frequent access patterns. The negative association rule describes occasionally access web patterns[36,37]. Using this fuzzy association user make a quick decision to retrieve data from the huge web pages.

4. Experimental Result

The web log data in personalized Library ontology is pre processed. The fields are extracted and saved in a table form to analyze user behaviour. We estimate the processed data from ontology and display the result in graphical form. The personalized Library ontology presents exact data for user's search on ontology. This Library ontology responded efficiently to all users' queries.

Table 2. Frequently used request and response pages

URL Path	REQUEST HITS	RESPONSE HITS
Library/AcademicLibrary	3509	3300
Library/ResearchLibrary	3009	2800
Library/NationalLibrary	878	689
Library/PublicLibrary	1057	700
Library/SchoolLibrary	1269	1100
Library/SpecialLibrary	2555	2100
Library/ChildrenLibrary	500	480

The log file generated during client request is stored in web server. This contain IP address, status code, number of byte transmitted, user agent and referrer, Timestamp, URL method, html name, user ID, session ID, http protocols, Referred web page, agent operating system, window type, window version and etc. Table 2 present frequently used pages on Library ontology and Figure 8 shows graphical representation of request and response time analysis. Figure 9 illustrates error rate analysis facts, where we can acquire information about errors that occurs during user interaction. This information resolves the errors and improves the business logic.

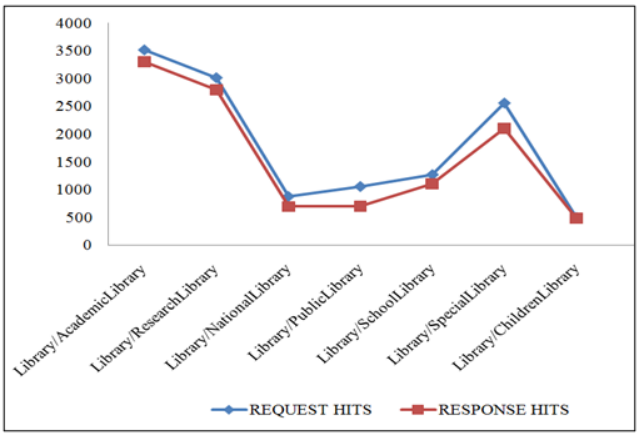

Figure 8. Graphical representation of frequently used pages.

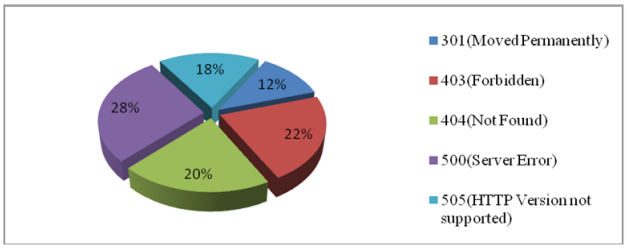

Figure 9. Error rate analysis.

5. Conclusion and Future Enhancements

Our proposed method captures user's essence and communications with Library ontology through web log files. First step, we use data preprocessing and cleansing method to remove superfluous data from web log. Second step, we establish the patterns of mined file to analyze the processed data for several user behavioural applications. Third step, the relevant grade and decision making concept are applied to each user's entries in order to shrink users' search period. The advantages of proposed method are trend analysis of data, which develop the business decisions, site improvement, increase resources and reduce users' search duration on ontology. We proposed a personalized Library ontology that automatically updates and analyze data in the requested page. In future we enhance the data transformation task as fully automated and improve the quality of mining web log files. We can also increase the volume of requests resources in personalized Library ontology.

6. References

1. Chitraa V, Selvdoss DA. A survey on preprocessing methods for web usage data. International Journal of Computer Science and Information Security (IJCSIS). 2010; 7(3):78-83.
2. Domenech JM, Lorenzo J. A tool for web usage mining. 8th International Conference on Intelligent Data Engineering and Automated Learning-IDEAL'07, Springer-Verlag: Berlin Heidelberg. 2007. p.695-704.
3. Fernandez FMH, Ponnusamy R. User behavior framework for personalized library ontology. Proceedings of the National Conference on Man Machine Interaction (NCMMI 2014), India. 2014. p.62-8.
4. Srivastava J, Cooley R, Deshpande M, Tan P-N. Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data. ACM SIGKDD Explorations Newsletter, 2000; 1(2):12-23.
5. Yuan F, Wang L-J, Yu G. Study on data preprocessing algorithm in Web Log Mining. Proceedings of the 2nd International Conferences on Machine Learning and Cybernetics. Xi'an, China. 2003. p.28-32.
6. Agosti M, Di Nunzio GM. Web Log Mining: A Study of User Sessions. 10th DELOS Thematic Workshop on Personalized Access, Profile Management, and Context Awareness in Digital Libraries. Corfu, Greece. 2007. p.1-5.
7. Koutsoupias N. Exploring web access logs with correspondence analysis. 2nd Hellenic Conferences on AI. SETN-2002, Thessaloniki, Greece. 2002. p.229-36.
8. Huiying Z, Wei L. An intelligent algorithm of data pre-processing in Web usage mining. 5th World Congress on Intelligent Control and Automation, WCICA'04. 2004; 4:3119-23.
9. Murata T, Saito K. Extracting users interests from Web Log Data. Proceedings of the 2006 IEEE/WIC/ACM International Conference of Web Intelligence, Hong Kong. 2006. p.343-46.
10. Maheswara Rao VVR, Valli KV. An Enhanced Pre-Processing Research Framework For Web Log Data using A Learning Algorithm. Nabendu Chaki et al. (Eds.), NeTCoM 2010, CSCP 01. 2011. p.01–15.
11. Arlitt M. Characterizing web user sessions. Internet and Mobile Systems Laboratory HP Laboratories: Palo Alto. 2000. p.2000-43.
12. Tanasa D, Trousse B. Advanced data preprocessing for inter sites Web usage mining. IEEE Intelligent Systems. 2004; 19 (2):59-65.
13. ohavi R. Mining E-Commerce Data: The Good, the Bad, and the Ugly. Proceedings of the 7th ACM SIGKDD international Conference on Knowldge discovery and datamining. San Francisco, CA. 2001. p.8-13.
14. Yen S-J, Lee Y-S, Hsieh M-C. An efficient incremental algorithm for mining web traversal patterns. Proceedings of IEEE International Conference on e-Business Engineering (ICEBE'05). Beijing. 2005. p.274-81.
15. Berendt B, Spiliopoulou M. Analyzing navigation behavior in Web sites integrating multiple information systems. VLDB Journal, Special Issue on Databases and the Web. 2000; 9(1). P.56-75.
16. Maedche A, Staab S. Ontology learning for the semantic web. IEEE Intelligent Systems. 2005; 16 (2). P.72-9.
17. Sure Y, Erdmann M, Angele J, Staab S, Studer R, Wenke D. OntoEdit: Collaborative ontology development for the semantic web. In Proceedings of the 1st International Semantic Web Conference on the Semantic Web. Sardinia, Italy, Springer-Verlag: Berlin Heidelberg, ISWC '02. 2002. p.221-35.
18. Chitra S, Kalpana B. Optimum session interval based on particle swarm optimization for generating personalized ontology. Indian Journal of Science and Technology. 2014; 7(8):1137-43.
19. Harin Fernandez FM, Ponnusamy R. Automated populates and updates personalized ontology with analysis result. 2014 IEEE International Conference on Advanced Communication Control and Computing Technologies (ICACCCT). Ramanathapuram, India. 2014. p.580-5.
20. Harin Fernandez FM, Ponnusamy R. Decision making and analyzing ontology from ontology log data using description logic. IEEE International Conference on Advanced Communication Control and Computing Technologies (ICACCCT), Ramanathapuram, India. 2014. p.629-33.
21. Khamparia A. Performance analysis of SPARQL and DL-Query on electromyography ontology. Indian Journal of Science and Technology. Aug 2015; 8(17):1-7.
22. Suneetha KR, Krishnamoorthi R. Identifying user behav-

ior by analyzing web server access log file. IJCSNS International Journal of Computer Science and Network Security. 2009; 9(4):327-32.
23. Buchner AG, Mulvenna MD. Discovering internet marketing intelligence through online analytical web usage mining. ACM SIGMOD Record. 2015; 27(4):54-61.
24. Cooley R, Mobasher B, Srivastava J. Data preparation for mining world wide web browsing patterns. knowledge and information systems. 1999; 1(1):5-32.
25. Cooley R, Mobasher B, Srivatsava J. Web mining: Information and pattern discovery on the World Wide Web. 9th IEEE Inernational Conference on Tools with Artificial Intelligence. Newport Beach, CA. 1997. p.558-67.
26. Kherwa P, Nigam J. Data Preprocessing: A milestone of web usage mining. International Journal of Engineering Science and Innovative Technology (IJESIT). 2015; 4(2):281-89.
27. Li Y, Feng B, Mao Q. Research on path completion technique in web usage mining. International Symposium on Computer Science and Computational Technology, ISCSCT'08, Shanghai. 2008; 1:554-9.
28. Castellano G, Fanelli AM, Torsello MA. Log data preparation for mining web usage pattern. IADIS International Conference Applied Computing. 2007. p.371-8.
29. Sumathi CP, Padmaja VR, Santhanam T. An overview of preprocessing of web log files for web usage mining. Journal of Theoretical and Applied Information Technology. December 2011; 34(1). P.88-5.
30. De Maio C, Fenza G, Loia V, Senatore S. Towards an Automatic Fuzzy Ontology Generation. Proc. IEEE International Conference on Fuzzy Systems, Jeju Island. 2009. p.1044-9.
31. Tho QT, Hui SC, Fong AC M, Cao TH. Automatic fuzzy ontology generation for semantic web. IEEE Transactions on Knowledge and Data Engineering. 2006; 18(6). P.842-56.
32. Cimiano P, Staab S, Tane J. deriving concept hierarchies from text by smooth formal concept analysis. Proc. GI Workshop Lehren-Lernen-Wissen-Adaptivit. 2003. p.1-8.
33. Ma ZM, Lv Y, Yan L. A fuzzy ontology generation framework from fuzzy relational databases. International Journal on Semantic Web and Information Systems (IJSWIS). 2008; 4(3). P.1-15.
34. Agrawal R, Srikant R. Mining sequential patterns. In ICDE'95. Proceedings of the 11th International Conference on Data Engineering. Taipei, Taiwan. 1995. p.1-12.
35. Zadeh LA. Fuzzy logic and approximate reasoning. Synthese. 1975; 30(3):407-28.
36. Moreno MN, Garcia FJ, Polo MJ, Lopez VF. Using association analysis of web data in recommender systems. In Proceedings of the 5th International Conference on E-Commerce and Web Technologies, Zaragoza, Spain, EC-Web '04. 2004. p.11-20.
37. Lin W, Alvarez SA, Ruiz C. Efficient adaptive-support association rule mining for recommender systems. Data Mining and Knowledge Discovery. 2002; 6(1):83-105.