

GLOBAL JOURNAL OF ENGINEERING SCIENCE AND RESEARCHES
Identification of User Pattern Deviation from Web Interface through User Actions Tracking

Dr. Ketan D. Patel^{*1} & Dr. Satyen M. Parikh²

^{*1}Asst. Prof, AMPICS, Ganpat University, Gujarat, India

²Dean, Faculty of Computer Applications, Ganpat University, Gujarat, India

ABSTRACT

The growth of websites as well as their popularity increases day by day. Website provides the service and solutions of at most problems people are facing in their regular life. Due to this, it is an essential job for web designers to identify the way people are using the website as well as the problems they face during their visit. Web admins as well as designers have the mindset that website should be accessed in predefined pattern by users to perform particular task. But in actual case, it might not be true and some deviations of patterns are there during the visit. Due to such deviations, extra time and efforts may require to fulfill the task on the site. Hence such deviations should be identified and addressed. This research work is an attempt to find out such deviations of user visits as well as their choice that influence their visit. To do so, user action tracking approach is implemented in which tasks are given to users to perform on the web interface and their actions are recorded and analyzed to identify deviation if any in patterns of page visits.

Keywords: *Pattern Deviation, Association Mining, Data Mining, User Action Tracking, Web Usability.*

I. INTRODUCTION

Due to the increasing popularity of websites, it is necessary to provide a web interface to the users which are usable and easy to understand. In ideal case web users prefers that the web pages they wants to access should be accessible within 3 clicks [1]. Hence it is important task for web designers to design the web interface in such a manner that web resources should be accessible with minimum efforts. For that it is important to identify the actual visit path of users while they interact with the website.

Web admins and designers have assumptions that web pages should be accessed with particular predefined ideal pattern. For example if user wants to visit Page C then he or she need to follow the ideal path ie. A → B → C. However it is not always true that end user will follow the same path to reach page C. Such deviations are possible due to the web interface design which is not user friendly and hence user is not following the ideal path or it may depends on user choice of what they want from the site. Hence it is essential to identify both user patterns and user choice during their visit.

In this research, user visit patterns and choices are identified by an approach called User Actions Tracking. With this approach user live actions are observed and stored in proper format which are then analysis using association mining technique of data mining. Association rule mining is the method of data mining to discover hidden association between data items.

Let's say, A and B are two data sets among which we want to identify hidden association. Association Rule is interpreted as the presence of item set A involves the presence of item set B [2]. In the present research work A and B item sets represents the web pages that are visited together during the visit. Support and Confidence are two important measures used to find association rule [2]. The formulas to calculate Support and Confidence are as below:

$$\text{Support} = \frac{\text{Number of page visit path containing both the pages A and B}}{\text{Total number of page visit path in the source data}}$$

$$\text{Confidence} = \frac{\text{Number of page visit path containing both the pages A and B}}{\text{Number of page visit path containing page A}}$$

Support is calculated by dividing total number of page visit path containing both the pages A and B with the total number of page visit path in the source data. It is a ratio of number of page visit path which contains both A and B pages with the total number of page visit path in the source data [3]. Higher support value indicates strong association between two pages.

Confidence value is calculated by dividing number of page visit path that includes both A and B pages with total number of page visit path which include page A only. It is ratios of number of page visit path which contains both A and B pages with the number of visit path containing page A only [3]. Confidence value 1 or 100% between page A and B indicates that every time when user visits page A he or she must visit page B in the same visit. It shows the strong association between pages A and B.

Such association findings help web designers and web admins to identify that the user follows the actual path or not? If not means some sort of redesign in the structure is required so that users have an ease to use it. It also helps to find out the choices of end users.

II. METHOD & MATERIAL

To identify and trace the user actions, group of users are identified to participate in the user action tracking approach. Participated users are asked to perform certain predefined tasks on the live educational website. While users interact with the website to perform the task assigned to them, their actions and visits are recorded by this approach.

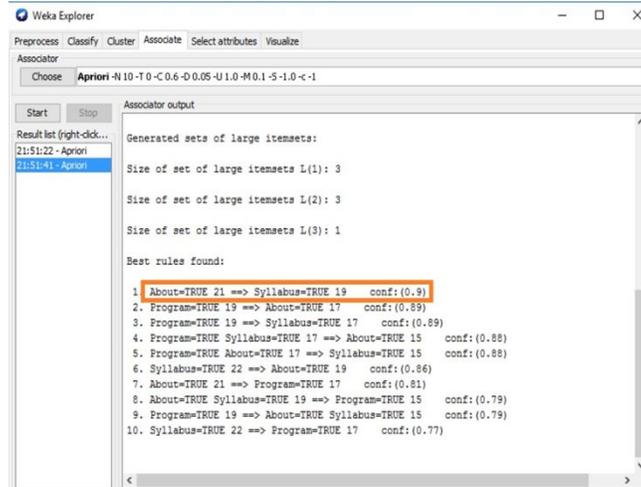
The first task given to the participants is to find out the specific course information from the educational web site and their actions on the interface were recorded and observed in proper format. This helps to identify the actual navigation path of the end users. The details of path they followed ie pages they visited are observed, recorded and then for better representation and analysis it is converted into .csv file. Association mining algorithm called Apriori is applied on the collected data using data mining tool called WEKA.

WEKA (Waikato Environment for Knowledge Analysis) is one of the popular open source data mining tools which is developed using java language. WEKA uses the collection of different machine learning algorithms. Free availability and simplicity makes WEKA a powerful tool for effective data mining techniques like association mining, classification and clustering, visualisation [4]. WEKA supports .csv file format and hence user actions data are converted into .csv format.

III. RESULT & DISCUSSION

The user action data recorded, collected and converted into .csv format was directly imported into WEKA tool. After importing the data, association mining algorithm called apriori is applied on the collected data. Results are shown in Figure 1.

Figure:1



Association Mining on User Actions Data

Figure 1 shows the results of association mining algorithm applied on the collected data using WEKA tool. Results show 0.9 (90%) confidence values between pages called *aboutcourse* and *syllabus*. Confidence value 0.9 shows that 90% time when participants visited the page *about* they have also visited the page *syllabus*. It simply means that page *about* and page *syllabus* is most likely pages of participants when they are looking for course details in the website. It gives us the actual visit path as well as choice of web users. If web admin and designers wants that, to get the course detail user should visit *about* page and then *syllabus* page. Here findings say that 90% of participants followed the path which is as per expectations. Only 10% users have deviated from the actual path.

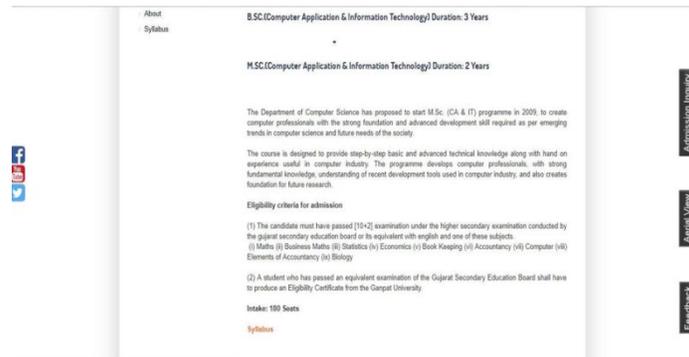
Another interesting finding which is come to know from the findings indicates that majority of the visitors who visit page *about* also visit the page *syllabus*. Hence web designers can put direct and visible link of *syllabus* page from *about* page if not exist.

Second task was given to the participants to download syllabus of a particular course from the educational website. Participants live actions were observed and found that most of the participants have directly visited the *syllabus* page without visiting the *about* page in this case. This shows the choice and actual needs of website visitors while they interact with the website. It indicates that when users are searching for syllabus there is no need to redirect them or force them to visit about course page.

It is essential to implement the findings provided by the user tracking approach to the educational website taken in this research to enhance overall usability of the site. Usability is a measure that is associated with an ease to use an object by its users. Here object can be a product or service, web page or website, information, software, various tools etc. A web interface is required to be easy, attractive and enjoyable to use [5]. To do so, findings are provided to the web admin of the education website and following modification is done to make web navigation simple and easy to use.

Implementation based on the findings

Findings says that people who visit *about* page strongly prefer to visit *syllabus* page and hence direct and visible link of *syllabus* page is required on *about* page so that users can directly visit *syllabus* page when they are on *about* page. Such link is putted on the *about* page. Figure 2 shows the same.



Implementation based on findings

IV. CONCLUSION

Popularity of website needs to address the serious concerns regarding the usage of the website. Web admins and web designers must have too aware about the usage patterns of users and user choice for their visit. It is crucial to know about the deviation between actual and expected page visit pattern. To do so, user action tracking approach is implemented which collects and observed user visit data and based on that identified that when users are looking for details regarding particular program, they mostly prefer to visit *about* and *syllabus* pages together. When they are looking for syllabus page, they do not prefer to visit *about* page and they directly visit *syllabus* page. Findings of this research helps web administrator to change the user interface as well as navigation path so that the interface becomes simple, clear and easy to use which satisfies the needs of the users as per their choice.

REFERENCES

1. R. Ahmad, Li. Zhang, F. Azam, "Measuring 'Navigational Burden'", *Proceedings of the fourth International Conference on Software Engineering Research, Management and Applications, IEEE, 2006.*
2. N. Gopalan, B. Sivaselvan, "Data Mining Techniques and Trends".
3. M.H. Dunham, X. Yongqiao, Le. Gruenwald, Z. Hossain, "A Survey of Association Rules", 2000.
4. *An Introduction to Weka*, Online Available At: <http://opensourceforu.com/2017/01/an-introduction-to-weka/>
5. Online Available at: <http://usabilitygeek.com/an-introduction-to-website-usability-testing/>