



ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 9, September 2018

A Proposed Method to Personalize and Extract the Hidden Knowledge Through Web Usage Mining and Pattern Discovery

Ruchika Kalal, Manish Tiwari

Research Scholar, M. Tech. CS, Geetanjali Institute of Technical Studies, Udaipur, Rajasthan, India

Assistant Professor, Geetanjali Institute of Technical Studies, Udaipur, Rajasthan, India

ABSTRACT: The research on WWW and associative algorithms and models has shown that WWW has grown beyond the expectation. Moreover, the development of Internet technology the growth of World Wide Web exceeded all expectations. A lot of information is available in different formats and retrieving content has become a very difficult task. Google today, however has used the best possible approach to solve in retrieving contents as and when required. But, again Google despite of its own powerful tool Search Engine Optimization (SEO) has failed in solving the issues of problems in Web Usage Mining (WUM). Web mining is the application of data mining on web data and web usage mining is an important component of web mining. The goal of web usage mining is to understand the behaviour of web site users through the process of data mining of web data and Web usage mining is to understand the behaviour of web site users through the process of data mining of web Access data.

I. INTRODUCTION

Project Overview

Google is a very popular and interactive medium for propagating information today. Due to the vast, varied and dynamic nature of web it raises the scalability, multimedia data and temporal issues respectively. The development of the web has been rise to large quantity of data that is freely available for user accessed by different users effectively and efficiently. That is why; the number of researchers in the field of application of Data mining techniques on the web is increasing. The web mining is the set of techniques of data mining applied to extract useful knowledge and implicit information from web data. As more organizations rely on the internet to conduct daily business, the study of web mining techniques to discover useful knowledge has become increasingly important. However, with the magnitude and diversity of available information from the internet, it is not insignificant to locate the relevant information to satisfy the requirements of the people with different background. To assist Web surfers in browsing the Internet more efficiently, one of the topics that have attracted much attention is modeling the web user's browsing patterns and making recommendations. Web mining enables one to discover web pages, text documents, multimedia files, images and other types of resources from web.

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijirccce.com

Vol. 6, Issue 9, September 2018

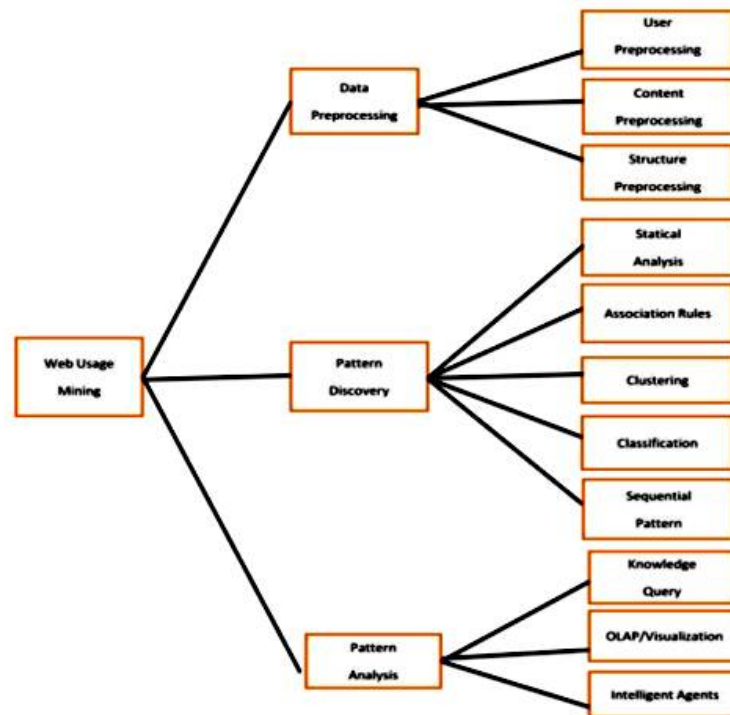


Fig.2-Web usage mining procedure and techniques

Figure 1 : Data preprocessing, pattern discovery, and pattern analysis

II. ANALYSIS OF PAST CONTEXTUAL CONSTRUCT MODEL

2.1 Background

The context of this investigation into IQ is an exploration of how high-end² information users make value-judgments of the information they encounter on the Web. This is an important point, because the thesis itself is not, strictly speaking, about human ISB, it is about human perceptions of IQ. Specifically it is an exploration of human perceptions of IQ from the point-of-view of receivers (or users) of information. More specifically, given (1) the pervasive nature of Web technologies (Lyytinen & Rose, 2003; Ford, 2004; Hinson, 2006); (2) their current status as the largest instrument for information dissemination in human history (Case *et al.*, 2004); and (3) the unique perspective – that information users who choose to gather their knowledge from the Web are required to make their own choices regarding how “correct” that information is (Ford *et al.*, 2001; 2002; Wildemuth, 2002); researchers are afforded a unique opportunity to observe human IQ related decision-making processes across a truly wide spectrum of circumstances.

User Surveys

Four (participant self-observation) user surveys were developed to investigate the three levels of investigation identified in the previous section.

Survey #1 TAM-IR: examined users’ attitudes and expectations regarding their search engine interaction.

Survey #2 TAM-IQ: examined users’ attitudes and expectations in relation to the process of using the Web to find and retrieve high quality information

Survey #3 ISB-Survey: examined various information seeking and searching behaviours, cognitive behaviours, and search engine strategies.

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 9, September 2018

Survey #4 IQ-Survey: examined users' general perceptions of IQ using sixteen identified IQ dimensions, and Web user behaviours in relation to user perceptions of IQ.

III. PROPOSED METHODOLOGY

The previous chapter introduced the contextual construct model used in the current research to help the researcher develop a valid and rigorous approach to the methodology-related aspects of the research project. Methodologies relevant to an exploratory, inductive research approach were presented and discussed in the context of how they applied to the current research. The discussion was largely theoretical, developing a picture of the multi-discipline context of the research, philosophical foundation and epistemology, and specific methodologies for data collection and analysis.

Response Time & Completion

An issue that became apparent as data-collection started was the completion-rates of respondents. Numerous participants registered to be part of the user-group, only to either not complete all four surveys, or – in some cases – not even begin the first survey.

A major contributing factor to users not completing all surveys may have been that the data collection process involved users completing multiple surveys. The decision to create multiple surveys was influenced by a number of factors;

- 1.) Three separate bodies of data/user information were being collected; that is; TAM related, ISB related, and perceptions of IQ;
- 2.) Internal validity required that some types of questions needed to be addressed before other types of questions; *and*
- 3.) The total number of questions being asked would have entailed an extensively large survey being designed. Such a survey may have been time consuming enough for some users to stop the survey before completing the whole survey, or not taking part in the survey at all.

The researcher addressed completion-rate issues by keeping a record of the “status” (which surveys had been completed) of each respondent. Given that the maximum response-time (how long a user could take to submit all surveys) was set at around 6 months, reminder emails were periodically used to prompt registered users which survey they last completed, and which survey they were required to do next. This strategy proved to be highly effective.

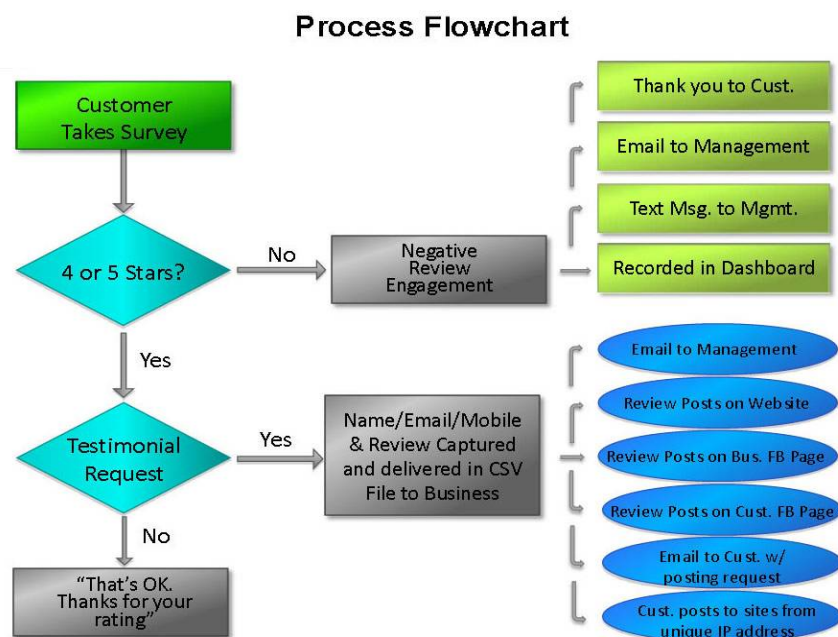


Figure 2 - The user/survey engagement occurred.

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijirccce.com

Vol. 6, Issue 9, September 2018

3.1 Data Analysis

The data analysis phase of the research, illustrated in figure 4.6, involved the synthesis and (2) analysis of user results, within a framework of (3) exploration; (4) confirmation; and finally (5) induction; processes. The following section describes the various strategies undertaken to handle and analyses the collected user results.

Google: -

2011	2012	2013	2014	2015	2016	2017	2018
49.9	441.4	152.3	592.8	950	1184.9	881.398	342.4
53.1	605.2	149.4	509.2	7000	605.713	717.73	800.2
54.9	450.2	108.9	350	11380	511.719	893.173	3000
61.8	379	80	224.4	4669	685.202	1850	3958.83
54.1	432.2	75.3	195.2	1564	840.6	1224.905	2293.75
25	348.8	45.3	414.1	1147	1176.904	1035.235	1146.481
187.8	358.7	88	1267	742.4	790.559	806.649	717.405
183.9	230.9	224.5	1206	560	721.405	555.796	3050

The current chapter has addressed the various strategies of data engagement of the research, illustrated in figure . Included have been some of the common considerations of survey-driven data collection, and a more detailed discussion of the analysis process. The contextual construct model, has been described in terms of how it philosophically drives the methods and strategies chosen. Finally, the strategies for data analysis and presentation of research findings have been presented, and are now used as a framework to guide the results chapters of the Paper.

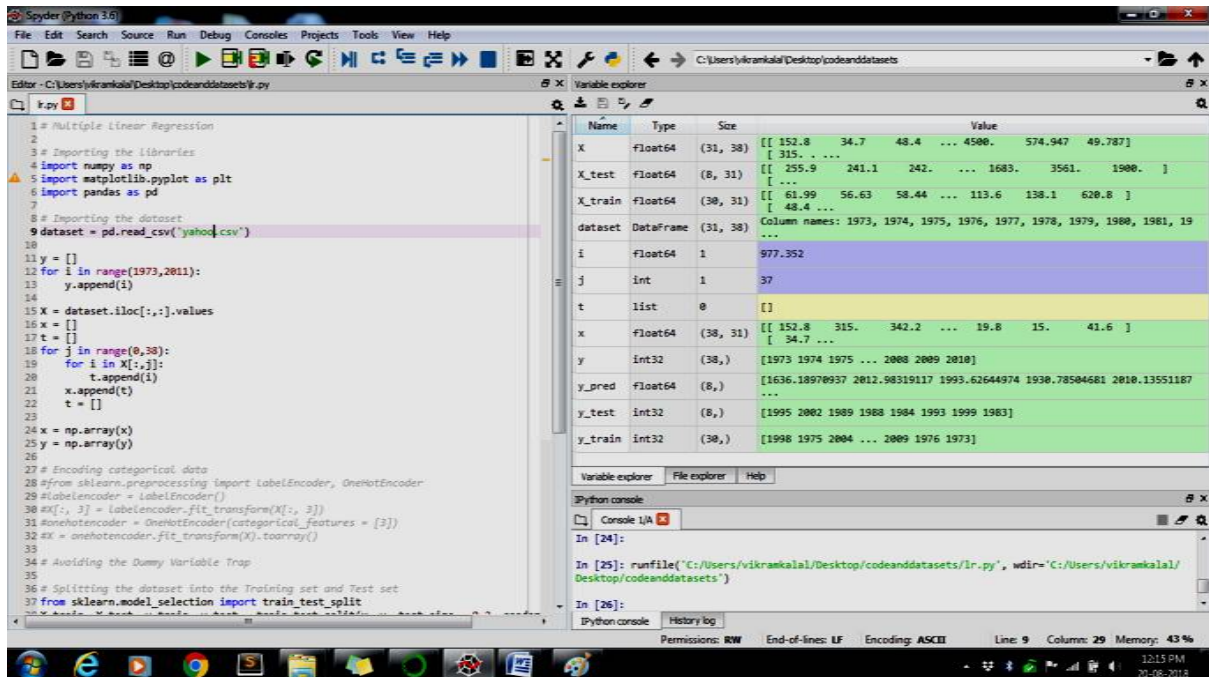


Figure 3 – Result of Yahoo Datasets

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijirccce.com

Vol. 6, Issue 9, September 2018

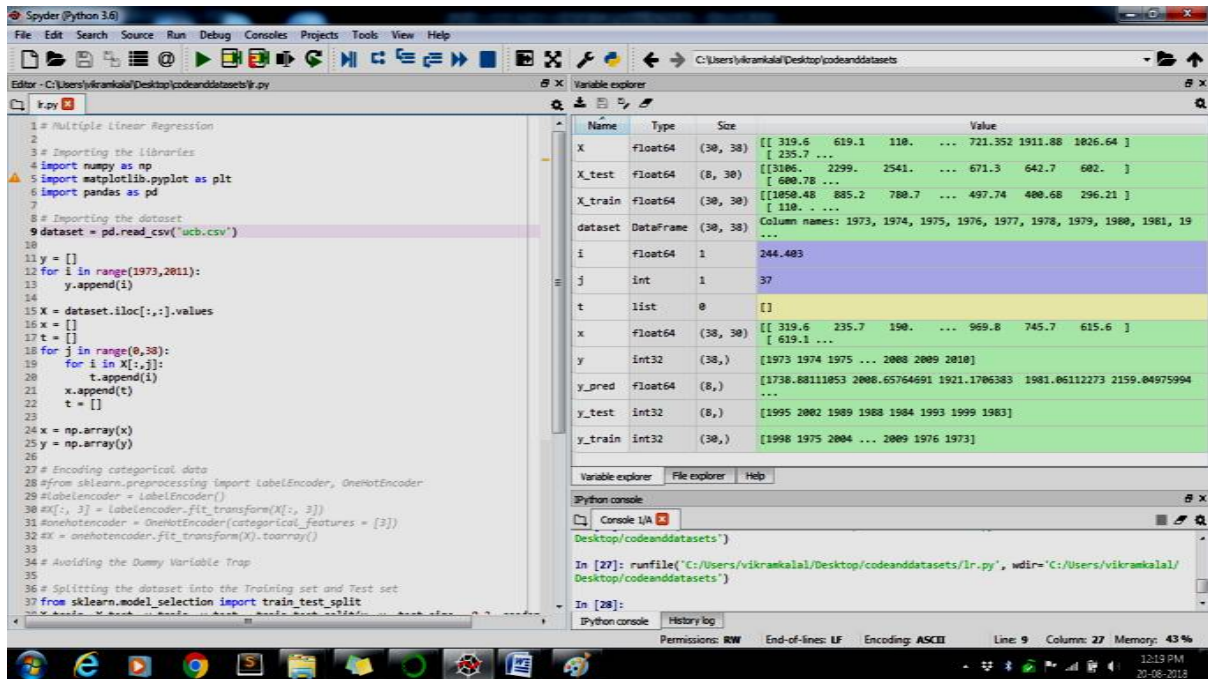


Figure 4 – Result of UCB Datasets

IV. CONCLUSION

4.1 Conclusion

This section presents some initial observations regarding the general characteristics of the user-group in this research study. It also presents the various constructed “group-cases” associated with the research, created from clustering sub-groups of users who possess similar characteristics, which will be used, as part of the research analysis framework (figure 4.6), to compare and cross analyze user results.

User Profile

The target user-group for the current research needed to be users who demand a high level of quality in the information they retrieve from the World Wide Web. To this end, it was determined that such a group could be found amongst “academic” users of the Web. A call for participation was sent out to appropriate university organizations, on-line academic community groups and list-servers, asking for users who met the following criteria.

- 1.) Users who were “academics”, or “postgraduate level” students – including users who fell into both these categories

REFERENCES

1. Abels, E. G., Liebscher, P., & Denman, D. W. (1996). Factors That Influence the Use of Electronic Networks by Science and Engineering Faculty at Small Institutions. Part I. Queries. *Journal of the American Society for Information Science*, 47(2), 146-158
2. Ford, N., Miller, D., & Moss, N. (2001). The role of individual differences in Internet searching: an empirical study. *Journal of the American Society for Information Science and Technology*, 52(12).
3. Ford, N., Miller, D., & Moss, N. (2002). Web search strategies and retrieval effectiveness: an empirical study. *Journal of Documentation*, 58(1), 30-48.
4. Ford, N., Miller, D., & Moss, N. (2005). Web search strategies and human individual differences: Cognitive and demographic factors, Internet attitudes, and approaches. *Journal of the American Society for Information Science and Technology*, 56(7), 741-756.
5. Ford, N. (2004). Modeling cognitive processes in information seeking: From Popper to Pask. *Journal of the American Society for Information Science and Technology*, 55(9), p.769-782.



ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 9, September 2018

6. Forslund, H. (2007). Measuring information quality in the order fulfilment process. *International Journal of Quality & Reliability Management*, 24(5), 515-524.
7. Foster, A. (2004). A nonlinear model of information-seeking behavior. *Journal of the American Society for Information Science and Technology*, 55(3), p.228-237.
8. Fugmann, R. (1973), "On the role of subjectivity in establishing, using, operating and evaluating information retrieval systems" *Information Storage and Retrieval*, Vol.9 No.7, p353-72.