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. 5, Issue 11, November 2017

Reverse Engineering Static Content and Dynamic Behaviour of E-Commerce Websites: A Survey

Sudhanshu Chourey [1] , Prof .Prashant Kumar Koshta[2]

Research Scholar, Department of Computer Science, Gyan Ganga College of Technology, India [1]

Professor, Department of Computer Science, Gyan Ganga College of Technology, India [2]

ABSTRACT: E-commerce websites rely heavily on summarizing and analysing the behaviour of customers, making an effort to influence user actions towards the optimisation of success metrics such as CTR (Click through Rate), CPC (Cost per Conversion), Basket and Lifetime Value and User Engagement. Knowledge extraction from the existing ecommerce websites datasets, using data mining and machine learning techniques, has been greatly influencing the Internet marketing activities. When faced with a new e-commerce website, the machine learning practitioner starts a web mining process by collecting historical and real-time data of the website and analysing/transforming this data in order to be capable of extracting information about the website structure and content and its users' behaviour. Only after this process the data scientists are able to build relevant models and algorithms to enhance marketing activities. This is an expensive process in resources and time since it will always depend on the condition. We may not know a priori that a visit to a Delivery Conditions page is relevant to the prediction of a user's willingness to buy and therefore would not enable tracking on those pages.

I. INTRODUCTION

E-commerce is one of the most disruptive innovations in trading. Marketing and advertising techniques are used to in enhance costumers' behaviour, trying to increase sales and products. Recommendation systems are one of the used techniques. Data mining and machine learning techniques had been applied to e-commerce as a way to improve e-metrics. Customer retention and engagement, click-through rate, conversion rate, shopping cart abandonment rate, customer lifetime value.

II. LITERATURE SURVEYS

The Paper [1] proposed with the objective of demonstrating the feasibility and applicability of the developed process, In this way, the implementation consists of a chain of processes arranged such as the output of each element of the chain is the input of the next one., we firstly have our data sources, consisting of the e-commerce website itself and the usage logs associated with it. On one hand, we have the website that is generally built using common web technologies and other resources. On the other hand, we got the logs associated with the website. These logs contain information about the requests or events associated with the interaction of the user with the website. If the data is captured at the server layer, this logs contains information about the HTTP requests of the user as he navigates between website's pages. If the data is captured at the application layer, this logs contains richer data about the interaction of the user, containing not only information about the pages he visits but also information about user interactions (i.e. clicks).

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. 5, Issue 11, November 2017

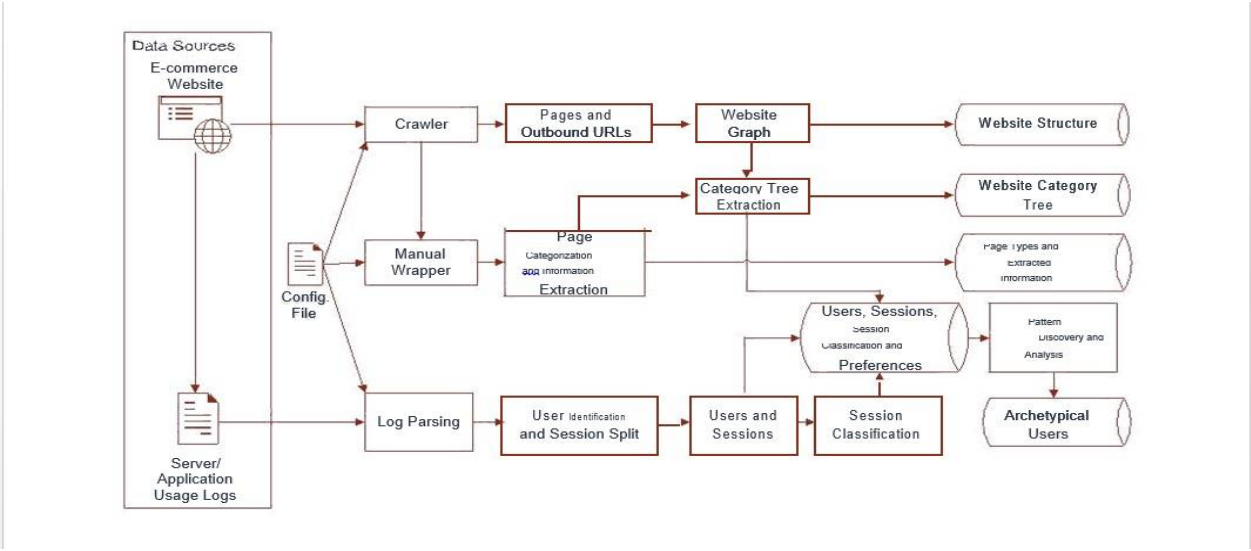

Fig 1.0 **system overview**

In This paper [2], The deep web contents are the information content that cannot be indexed by search engines as they stay behind searchable web interfaces. Current deep web directories mostly have less coverage of relevant web resources which degrade their ability. A Crawler goes across a variety of web pages during a crawling process.
Hence to achieve efficient crawling and wide coverage, ranking and prioritizing links of different sites is necessary. The objective of this system is to extract deep web information with wide coverage for hidden web resource and uphold efficient crawling for focused web crawler.

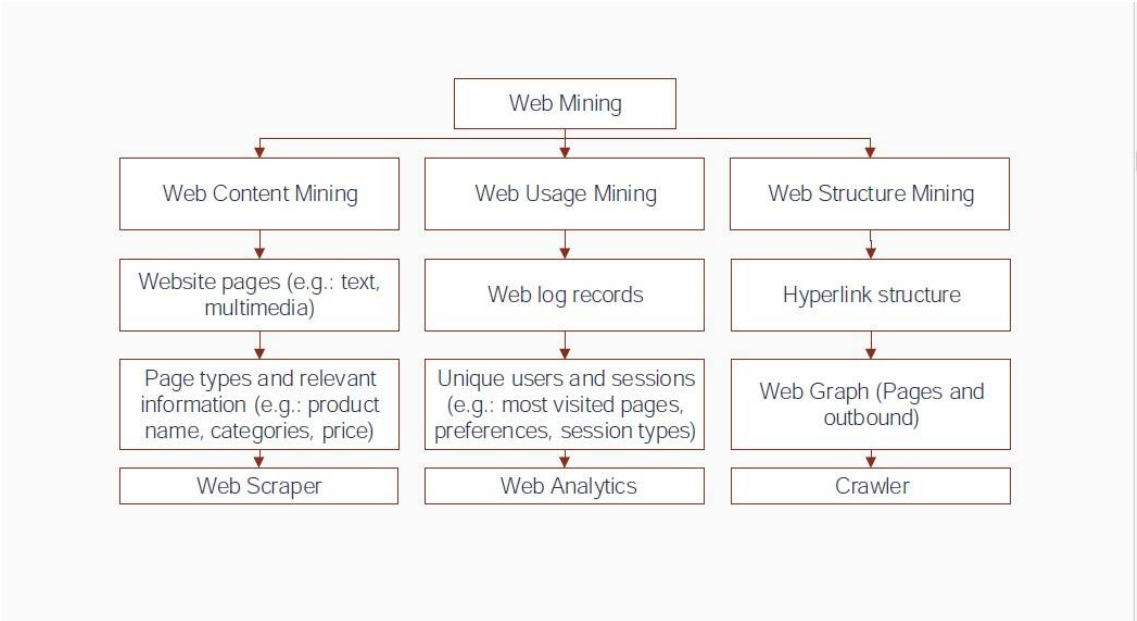

Fig 2.0

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. 5, Issue 11, November 2017

In This Paper [3] The more the number of search engines accessing a website, the more will be its visibility when searching for a particular web site. The observed results show that all search engine crawlers are not visiting all the websites. In this experiment the data set 1 was accessed by more number of search engines compared to data set 2. Certain search engines were consistent in the number of visits and number of pages crawled while a few were not consistent or irregular in their visits and pages crawled. It is found that data set 1 is more visible to search engine crawlers as it is crawled by more number of search engines compared to data set 2. The results also showed a positive correlation between the number of visits and number of pages crawled. A better search engine optimization policy can be followed to make the websites visible to different search engines so that the websites will be listed top in the search engine rankings.

In This Paper [4] The k-means clustering algorithm is one of the most commonly used data partitioning Algorithms. Despite its wide use the algorithm suffers from serious drawbacks. In this paper, an improved k-means partitioning algorithm, named SkM, is proposed dealing with the selection of the initial cluster centers. In the improved algorithm, the initial centers are selected based on a factor that utilizes the standard deviation and the max value of all the data points found in the data set, as compared to the traditional k-means that performs a random selection.

In This Paper [5] Based on the thought of K-means algorithm, the object sets of e-commerce transaction data of 300 phones can be deemed as input to be clustered, in order to get clustering center and object sets of clustering data. objects can be randomly selected from data sets as the center of initial group, then assign each object to the most similar group according to the object's mean value of the group, and update the group's mean value, calculating the object's mean value of each group. Repeat the above steps until there is no change for the number of group.

In This Paper [6] does an improved K-means clustering algorithm for identifying internet user behavior. Web data analysis includes the transformation and interpretation of web log data find out the information, patterns and knowledge discovery. The efficiency of the algorithm is analyzed by considering certain parameters. The parameters are date, time, S_id, CS_method, C_IP, User agent and time taken. The research done by using more than 2 years of real data set collected from two different group of institutions web server .this dataset provides a better analysis of Log data to identify internet user behavior.

IN This Paper [7] In this experiment checked on E-commerce application where database used for non-hypertext links i.e. Deep Web. Non hypertext content indexed quicker by Google crawler with the help of sitemap metadata and robot.txt. Most of the Web's information is buried far down on dynamically generated sites, and general search engines do not find it. These traditional search engines cannot "see" or retrieve content in the Deep Web. Today wealth of information that is great source in Deep Web and therefore missed. General Search engine is crawl but not showing indexed resources in result web search engine
Page. It is a great achievement in www if every Deep Web on all major sectors likes - Education, Business, Governance, Media, Career, Multimedia etc. implements aforesaid technique for the hidden public access content. This sitemap and metadata also similar useful for other general search engine like Yahoo, Bing, Ask etc. but only Robot.txt page change according targeted search engine.

In This Paper [8] we have presented vector analysis and KMeans based algorithms for mining user clusters. We have also applied the proposed algorithms to the real world data and our experimental results show the proposed algorithm is feasible, and have scalability.

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. 5, Issue 11, November 2017

III. FUTURE WORK

To enhance the process and, mostly, the proof-of-concept, we can consider the following improvements:
Improve the crawler implementing parallelism and/or prioritisation of the frontier ; Identify and dier entiate static from dynamic hyperlinks; Carry experiments with another kind of web scrapers (e.g. wrapper induction); Increase the data crossings (e.g. cross content and usage data to get to know the favourite user brands or range of prices); Apply di erent algorithms to nding and understand the archetypical website's users (e.g. other cluster algorithms or other pattern discovery techniques);Analyse the possibility of expanding this methodology beyond e-commerce websites, nding other user cases.

REFERENCES

[1] Automating the Extraction of Static Content and Dynamic Behaviour from e-Commerce Websites By Joao Pedro Diasa,, Hugo Sereno Ferreirab
[2] Enhancing Crawler Performance for Deep Web Information Extraction byParigha V. Suryawanshi, D. V. Patil
[3] Mining Web Logs to Identify Search Engine Behaviour at Websites by Jeeva Jose(Informatica)
[4] Extending the k-means Clustering Algorithm to Improve the Compactness of the Clusters by Antonia Nasiakou (JPRR)
[5] Clustering Analysis on E-commerce Transaction Based on K-means Clustering by Xuan HUANG 2014 ACADEMY PUBLISHER
[6] Clustering of User Behaviour based on Web Log data using Improved K-Means Clustering Algorithm by S.Padmaja (IJET)
[7] Deep Web Performance Enhance on Search Engine by Deepak Kumar (ICSCTI)
[8] Web User Clustering Analysis based on KMeans Algorithm by linHuaXu (ICINA)