

# Web Mining — Concepts and its Applications

\*Dr. S.S. Gautam, \*\*Manish Kumar Tiwari

\*Associate Professor, MGCGV, Chitrakoot, Satna, M.P.

\*\*Research Scholar, MGCGV, Chitrakoot, Satna, M.P.

---

*Abstract- Recover information from World Wide Web is a boring assignment since the expansion in the ease of use of knowledge backup supply on it. So this raises the need to utilize a clever system to recover the information from World Wide Web. The way in which Web information of getting back and Web base data warehousing are boosted with the removal of facts from the Web using web mining tools. Web usage mining is one of the best developing areas of web mining. Its notice in analyze users recital on the web after exploring right to use logs made its fame very quickly in E-services areas. Most of the e-service providers realized the fact that they can relate this tool to keep hold of their clientele. This paper tries to provide an insight into web mining and the different areas of web mining. Then it spotlight on Web usage mining, its application and collision in E-services.*

---

## PROLOGUE

Web mining is the application of data mining. It is a process to remove knowledge from web data, including web documents, hyperlinks between credentials, and practice logs of web sites. A panel prearranged at ICTAI 1997 asked the question “Is there something different about web mining?” As no eventual conclusions were reached then the marvelous notice on web mining in the past five years, and a number of noteworthy ideas that have been developed. It has certainly answered this question in the assenting in a big way. In accumulation a quite steady community of researchers paying notice in the area has been formed, mostly through the successful series of Web KDD workshops, which have been held yearly in conjunction with the ACM SIGKDD Conference since 1999 and the web analytics workshops, which have been held in conjunction with the SIAM data mining conference.<sup>1</sup>

The World Wide Web has multiplicity of knowledge service centers, like news sites, encyclopedias, education sites, e-commerce etc. So the information in WWW is extended widely in theses information hub worldwide.<sup>2</sup> To recover from these disseminated storage areas is a quite hard process and it necessary requires a competent tool to find the preferred information. Only a clever system which effectually mine for knowledge can decide these troubles. The subsequent factors made it hard for an effectual data warehousing and data mining.

- The massive size of the web
- No proper arrangement for the web documents.
- The active environment of the information source.
- The variety in usage and consumer society.

However these are confront that stand as the powerful strength for the research into competent and effectual discovery and use of property on the Internet. The basic distinctiveness of web make us have to think about to form and get longer the unfashionable research methodologies. Two dissimilar approaches were taken in originally important web mining. First was a “process-centric view,” which define web mining as a series of task.<sup>3</sup> Second was a “data-centric view,” defined web mining as a stipulations of the types of web data that was being worn in the mining process.

The second meaning has become more satisfactory, as is apparent from the approach adopted in most papers that have physically location the issue.<sup>4</sup>

WEB MINING is the function of data mining method to remove information from web data, i.e. web content, web structure, and web usage data.

## WELL-KNOWN APPLICATIONS

Enthusiasm regarding the web in the precedent few years has led to the web applications being technologically advanced at a much faster rate in the industry than examine in web related expertise. Many of application are standing on the use of web mining ideas even though the association and made-up the similar technologies did not regard as it as such. We explain some of the majority thriving applications in this section.<sup>5</sup> evidently, understand that these applications employ web mining is largely a demonstration exercise. For all application group argued below we have chosen a famous agent purely for ideal purposes. This in no means entail that all the techniques explain was developed by that group alone. On the dissimilar, in the majority cases the victorious techniques were developed by a quick “copy and improve” move toward to each other’s ideas.

## UTILITY IN E-SERVICES

World Wide Web is turn out to be a broad extend medium for the flow of information. Improvement in technology exposes that the quantity of data in web and its multifaceted structure are growing day by day. In this situation, the request of web usage mining has its own meaning, Quick development in online services called e-service functions like e-commerce, e- governance, e- market, e-finance, e-learning, e-banking etc. has made trade society and clientele face a new circumstances. Acceptance of clever marketing approaches is the barely solution for the trade society to face the dispute of business rivalry and the customers' alternative to choose from several other possibilities. The papers spotlight on function of web custom mining in three major services i.e. e-services E-commerce, E-learning and E-governance.<sup>6</sup>

### 1. E-COMMERCE

E-Commerce means two operate parties founded on Internet according to definite regulations or normal developing the entire traditional trade action in digital network approach. Exchange of goods or services through Internet E-commerce produce enormous volume of relations. This marvelous growth in the E-commerce venture, twisted to product spare. As well as they faced a general question on how to know the customer happiness and their buy trend.<sup>7</sup> the rivalry in the field raised the require of serving customers in better way also initiated so many question. In order to give improved service to the users there should be a prerequisite of an efficient promotion strategy for analyzing their happiness and usage. Finding out the usage patterns from web data is the techniques which adopted in web usage mining.

It has been a significant technology for consideration users' behaviors on the Web. Their technologies to gather client click streams, buying and traversing patterns etc. are vital information for the E-commerce activity. They help in investigating demographic data and assist them to propose cross-marketing procedure transversely products and services. It also supports e-commerce sites to keep the most gainful customers, get better the functionality of web based function provides more specially made content to visitors. In adding up with the use of Web usage mining techniques e-commerce companies can get better products quality or sales by anticipating troubles before they happen.<sup>8</sup> They also give groups with earlier unknown buying patterns and performance of their online clientele. More prominently the fast criticism the groups gain by using Web usage mining is very obliging in raising the company's profit.

### 2. E-LEARNING

E-learning is a structure of electronically supported learning which allows the community to learn several subjects at anytime and anywhere. The ease in using the apparatus to browse the resources on the web, its open-mindedness in deploying and maintaining resources ended the web as an excellent tool for delivering courses. Web is only major choice to administer and maintain learning resources and has turn into one of the foremost choice of modern advanced distance education system. As learning becomes more technologically advanced, the difficulty of available learning resources also enlarged accordingly. It is hard to evaluate the structure of the course content and its effectualness on the learning process. Track and judge all the actions performed by learners are also very hard as well as time consuming.<sup>9</sup>

This is the situation where web usage mining can add. The blueprint analysis capability of web usage mining has a significant role in web-based learning system. They can analyze the students and instructors performance and improve the educational experience. Tracking the behavior happening in the course website and pit patterns is also helpful to improve or adapt the course contents. This allows instructors to assess the access behavior, evaluate the learning activities and match up to learners. The agreement of the course contents can be enhanced by analyzing the traversal paths of the course content web pages is another benefit of Web usage mining.

### 3. E-GOVERNANCE

E-governance endow with an only web portal that incorporate all services that comprises government, non profit and private-sector bodies. In such a type of service arrangement which affords prepared admission to information. The user passage point quality is an important factor. This is one of the hard user-centric restraints since this has to offer information to wide and various users. If the appearance sub system regulates according to the individual proclivity of each user will ensure wide contribution in e-governance systems.

The blueprints of the online actions of the users can be exposed by means of Web usage mining methods. These blueprints disclose the user benefit and that can be exploiting to fine tune consumer interface and propose the most suitable browsing pathway. User necessities also are revealing in their steering behavior. Analyzed results can be seen as information to be used in clever online applications, cleansing web site maps, web based modified system. This technology also uses the knowledge of users of past congregation to provide advice to users of current session.<sup>10</sup>

## FEW EXAMPLES

### MODIFIED CONSUMER PRACTICE IN B2C E-COMMERCE

#### AMAZON.COM

Untimely on in the life of Amazon.com its prophet CEO Jeff Bezos experiential, "In a conventional shop, the main effort is in getting a buyer to the shop. Once a buyer is in the shop they are probable to make procurement since the price of going to another shop is high and thus the budget of marketing is in commonly much higher than the in shop customer experience financial plan. In the case of an online amass, getting in or out requires precisely one click, and thus the main center must be on customer experience in the shop." This basic observation has been the driving strength behind Amazon's inclusive approach to modified customer experience, based on the mantra "a personalized shop for every client". A mass of web mining techniques, such as relations between pages visited and click-path study are used to progress the customer's knowledge during a "shop visit." Information gained from web mining is the key aptitude behind Amazon's features such as "instant recommendations," "purchase circles," "wish-lists," etc.

#### GOOGLE

Google is one of the mainly popular and extensively used search engines. It offers users right to use to in order from over 2 billion web page that it has indexed on its server. The excellence and rapidity of the search facility makes it the most unbeaten search engine. Earlier search engines intense on web content only to return the pertinent pages to a query. Google was the first to bring in the significance of the link structure in mining information from the web. Page Rank, which trial the consequence of a page is the fundamental technology in all Google search products and uses structural information of the web graph to revisit high quality outcome.<sup>11</sup>

#### TRACKING DOUBLE CLICK

"Web-wide tracking," i.e. following an entity across all sites he stays is a fascinating and contentious technology. It can provide an indulgent of an individual's lifestyle and habits to a level that is unparalleled, which is obviously of marvelous interest to marketers. A thriving example of this is Double Click Inc.'s DART ad management technology. Double Click serves advertisements, which can be embattled on demographic or behavioral attributes to the end-user on behalf of the client, i.e. the web site by means of Double Click's service. Sites that use Double Click's service are part of The Double Click Network and the browsing performance of a user can be trailed across all sites in the network, using a cookie. This makes Double Click's ad target to be based on very primitive criteria. Alexa Research has recruited a panel of more than 500,000 users, who have willingly agreed to have their every click track, in return for some freebies.

#### WEB COMMUNITIES AOL

One of the main successes of America Online (AOL) has been its large and trustworthy customer base. A large segment of this client base contributes in a variety of AOL societies, which are gatherings of users with similar benefit. In accumulation to provide that a opportunity for each such community to interrelate between themselves, AOL offer them with practical information and services. Over time these society have grown to be well-visited water holes for AOL users with common interests. Relating web mining to the data collected from society communications presents AOL with a very good understanding of its communities, which it has used for targeted marketing through advertisements and e-mail intransitive. Freshly, it has happening the concept of "community sponsorship," where by an organization, say Nike, may sponsor a society called "Young Athletic Twenty incredible." In return, customer survey and new product development specialists of the funding organization get to contribute in the community, possibly without the information to other member. The idea is to take care of the society as a highly specialized focus group, recognize its requires and opinion on new and accessible products, and also test strategies for influence opinions.

#### AUCTION BEHAVIOR EBAY

As person in the world where we have a lot of things than we require, the attraction of replacing our ineffective work for a few cash, no matter how small, is fairly powerful. This is explicable from the success of flea bazaars, garage sales and estate sales. The concentration of eBay's creators was to create an infrastructure that gave this advocate a worldwide reach, with the ease of doing it from one's home PC. In adding together, it popularized public sale as a merchandise selling, buying method and provides the excitement of gamble with no the difficulty of having to go to Las Vegas. All of this has complete eBay as one of the most triumphant businesses of the internet era. Regrettably, the secrecy of the web has also created a major problem for eBay public sales, as it is not possible to differentiate real bids from fake ones. eBay is now using web mining method to analyze arrange behavior to resolve if a bid is deceitful. Current efforts are geared towards considerate participants' bidding patterns to create a more proficient auction market.

#### PROCESS MINING

Mining of market storage bin data, scatterbrained at the point-of-sale in any shop, has been one of the noticeable successes of data mining.

Yet, this data offers only the end result of the process, and that too judgments that ended up in product purchase. Click-stream data make available the opportunity for a complete look at the decision making process itself, and information extracted from it can be used for optimizing, manipulate the process, etc.<sup>12</sup> has decisively established the value of process information in recognizing users' performance in conventional shops. Research requires to be carried out in

- (1) Remove process models from usage data,
- (2) Accepting how different parts of the process model collision various web metrics of interest,
- (3) How the process models change in reply to various changes that are made, i.e. changing stimuli to the user. Figure 1 shows an approach of modeling online shopping as a state evolution figure.

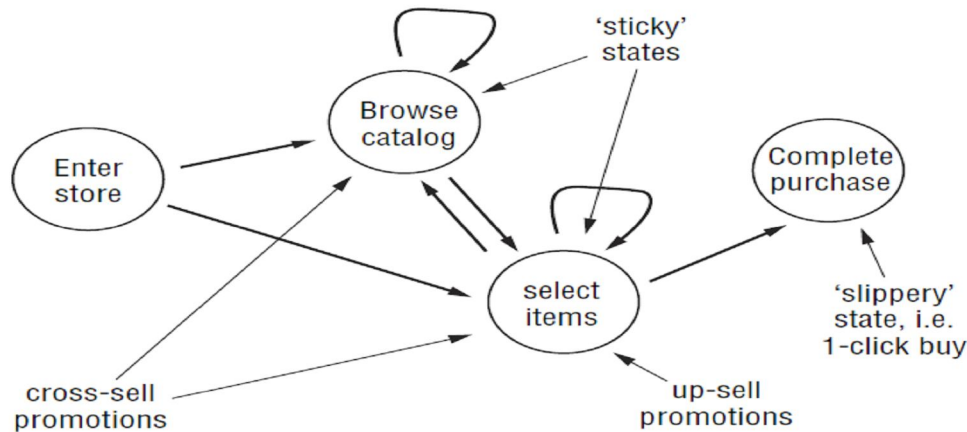


Figure 1 Shopping pipeline model as state evolution figure.

In general goal:

- Make the most of probability of achieving final state
- Maximize predictable sales from every visit

#### CHRONOLOGICAL TRANSITION OF THE WEB

People's interface with the web is altering the web as well as the way public cooperate with each other. As storing the history all of this communication in one place is obviously too amazing a task, at least the altering to the web is being evidence by the pioneering internet archive project. Research requires to be carried out in remove temporal models of how web contented, web structures, web communities, establishment, hubs, etc. evolve over time. Large organizations usually archive usage data from their web sites.<sup>13</sup> with these sources of data accessible; there is a large scope of research to expand techniques for scrutinizing of how the web changes over time

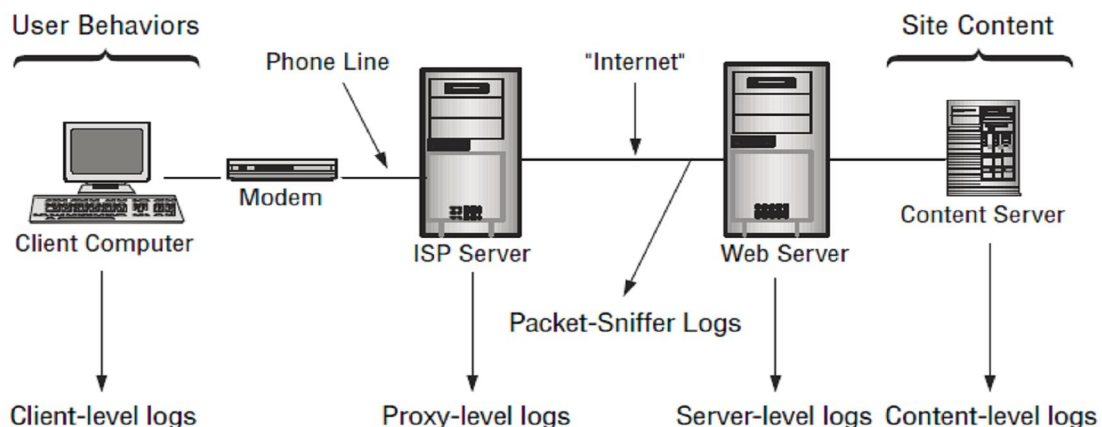


Figure 2: Elevated level architecture of dissimilar web logs.



#### FRAUD AND RISK STUDY

The secrecy provided by the web has lead to a important enlarge in attempted fraud, from unlawful use of individual credit cards to lacerate into credit card databases for blackmail function. However another example is online sale fraud, which has been increasing on famous sites like eBay. Since all these frauds are being carry out through the internet, web mining is the ideal analysis technique for perceiving and preventing them. Research issues comprise developing techniques to recognize known frauds, characterize them and identify emerging frauds. The issues in cyber danger analysis and interruption finding are quite similar.<sup>14</sup>

#### PRIVACY AND WEB MINING

Although there are so many benefits to be expanded from web mining. A clear disadvantage is the possible for severe infringement of privacy. Public approach towards privacy give the feeling to be nearly schizophrenic, i.e.<sup>15</sup> people say one thing and do quite the differing have confirmed that people were eager to provide fairly personal information about themselves, which was completely inappropriate to the task at hands if provided the right motivation to do so. Also, clearly bringing notice to information privacy policies had virtually no effect. One explanation of this apparently contradictory attitude towards privacy may be that we have a bi-modal view of privacy, namely that "I'd be willing to share information about myself as long as I get some (tangible or intangible) benefits from it, and as long as there is an implicit guarantee that the information will not be abused." The examining issue generated by this attitude is the need to develop move toward, methodologies and tools that can be used to verify and authenticate that a web service is certainly using user's information in a manner reliable with its stated policies.<sup>16</sup>

#### CONCLUSIONS

Web usage mining is suitable an active concerning field of research because of its potential marketable benefits. It is further possible to analyze the visitor's performance by linking the Web logs with cookies and forms, and which could help e-services site to address several business questions. Its awareness in analyzing user's actions on the web after discovering access logs made its fame very rapidly specially in E-services areas. Details like user log files, request for resources etc. are uphold in web servers, which is the core mining area of web usage. The study of these gives the user browsing case and that can be utilized for target advertisement, improvement of web design, satisfaction of clientele and making market analysis. Most of the e-service suppliers understand the fact that they can apply this tool to keep their clientele

As the web and its usage continues to grow, so too grows the chance to analyze web data and remove all manner of useful knowledge from it. The past five years have seen the appearance of web mining as a fast increasing area, due to the labors of the research society as well as different organizations that are working it. Our hope is that this overview provides a starting point for fruitful debate.

#### REFERENCES

- [1]. E. Kim, W. Kim, Y. Lee. Purchase propensity prediction of EC customer by combining multiple classifiers based on GA. International Conference on Electronic Commerce 2000: 274~280.
- [2]. J. B. Schafer, J. A. Konstan, J. Riedl. E-commerce recommendation applications. Data Mining and Knowledge Discovery, 2001(5):115~153.
- [3]. S. W. Changchien, T. Lu. Mining association rules procedure to support on-line recommendation by customers and products fragmentation. Expert Systems with Applications, 2001(20): 325~335.
- [4]. B. Sarwar, G. Karypis, J. Konstan, J. Riedl. Analysis of recommendation algorithms for e-commerce. Proceedings of ACM Ecommerce 2000 Conference, 2001:158~167.
- [5]. S. Yuan, W. Chang. Mixed-initiative synthesized learning approach for Web-based CRM. Expert Systems with Applications, 2001(20):187~200.
- [6]. [http://revistaie.ase.ro/content/51/104%20-20 SIVA RAMA KRISHNAN, %20BALAKRISHNAN.pdf](http://revistaie.ase.ro/content/51/104%20-20%20SIVA%20RAMA%20KRISHNAN,%20BALAKRISHNAN.pdf)
- [7]. J. g. Liu, h. h. Huang. Web Ming for Electronic Business Application, Proceedings of the Fourth International Conference on Parallel and Distributed Computing, Applications and Technologies, Chengdu, China, 2003:872~876.
- [8]. <http://www.ceng.metu.edu.tr/~nihan/ceng553/StudentPapers/01635156>
- [9]. Jaydeep Srivastava, Robert Cooley, Mukund Deshpande, Pang-Ning Tan; Web Usage mining: Discovery and Applications of Usage Patterns from Web Data; ACM SIGKDD; Jan 2000; Volume 1; Issue 2.
- [10]. Ning Bin, Lei Yuan; Research on Application of Web Mining in E-commerce; Advanced Materials Research - Scientific. Net; Volume 403 – 408; Pages 1830 – 1833; Nov 2011.
- [11]. Romero C. Ventura S, Pechenizky M , Baker R. S ; Handbook of educational data mining; 2010; CRC Press.
- [12]. Bart C Palmer; Web Usage Mining: Application to an online educational digital library service; Digital Commons@USU; 2012

- [13]. Zakareya Ebrahim and Zahir Irani; —E-government adoption: architecture and barriers| Emerald Business Process Management journal, vol.II, No.5 2005, pp589-611, 2005
- [14]. A. S.Chakraverty 1, B. G.Rani, C. B.Singla and D. D.Anand; Experience based recommendations system for e-governance; 2012.
- [15]. G.Rani; S.Chakraverty, —Boosting Interactivity of EGovernancel, International Conference on Communication Languages and Signal Processing- with Preference to 4 G Technologies|, ICCLSP 4G, January2012.
- [16]. Nasraoui, O.; Soliman, M.; Saka, E.; Badia, A.; Germain, R.; "A Web Usage Mining Framework for Mining Evolving User Profiles in Dynamic Web Sites,"Knowledge and Data Engineering, IEEE Transactions on , vol.20, no.2, pp.202-215, Feb. 2008
- [17]. Cooley, R., Mobasher, B., Srivastava. J., —Web Mining: Information and Pattern Discovery on the World Wide Web|, Proceedings of the 9th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'97), November 1997.
- [18]. H. Lieberman. Letizia: An agent that assists web browsing. In Proc. of the 1995 International Joint Conference on Artificial Intelligence, Montreal, Canada, 1995.
- [19]. Tao Huachuca, Jiang Lingyan. Web-based Data Mining Behavior Analysis and Research. Fujian Computer. 2004 No. 3
- [20]. Jin Fengrong. Study of Web Usage Mining and Discovery of Browse Interest. Master's Degree Thesis of Beijing Science and Technology University. February 2004
- [21]. Chu Hue Lee, Yo Lung Lo, Yu Hsiang Fu; A novel prediction model based on hierarchical characteristic of web site; Elsevier; Volume 38 Issue 4 , April 2011, Pages 3422 – 3430
- [22]. <http://www.amazon.com>.
- [23]. <http://www.google.com>.
- [24]. <http://www.aol.com>.
- [25]. <http://www.ebay.com>.
- [26]. <http://www.google.com/press/pressrel/b2b.html>
- [27]. <http://news.google.com>
- [28]. <http://www.doubleclick.com/dartinfo/>
- [29]. <http://www.alexa.com>.
- [30]. <http://www.wired.com/news/business/0,1367,36434,00.html>.
- [31]. <http://www.microsoft.com/netserVICES/passport/>.
- [32]. <http://www.aol.com>.
- [33]. <http://www.ebay.com>.
- [34]. <http://www.yahoo.com>.
- [35]. See <http://www.yodlee.com>.