NIET Journal of Engineering & Technology (NIETJET)
Volume 10, Issue Winter 2021 ISSN: 2229-5828 (Print)

A Web Mining and Optimization Approach for Improving Data Retrieval Performance in Web Search Engine Outcomes

Mohammad Shahid[1]
[1]Department of computer science &Engineering, GH Raisoni College of Engineering,Nagpur India

Abstract: ***The web is a collection of documents (information, photographs, audios, videos, and so on) uploaded and published by a huge number of individuals, and at the same time, a great number of people are using online search engines to find their relevant documents. When users use various search engines to do searches, they will receive a vast number of relevant and irrelevant sites in answer to their queries. People are obtaining more irrelevant sites against the query supplied by users, thus new approaches are used to the search results to aid users in navigating the result list. To sort the results to be shown to users, search engines utilise several methods of query optimization and query categorization. As a result, optimal data retrieval is the process of selecting the most relevant information resources from a large collection of data resources. As a result, a method to optimising and integrating online content, web mining, and approaches for boosting a search service's knowledge of user search queries is presented in this work. The focus of the research will be on improving the performance of relevant data retrieval in web search engine results.***

Introduction

As we all know, the online includes a massive amount of material that is continually expanding at a rapid rate since most users use the internet to locate relevant and fascinating content, and search engines have become one of the most popular tools for web users to find relevant information. And, most of the time, consumers lose patience after receiving a slew of unsolicited papers after clicking on various links. Thus, providing a user-friendly tool for extracting relevant material without having to examine the entire data set at the beginning has become a major priority among web mining research communities. Searching is regarded as one of the most significant aspects of the World Wide Web due to the use of queries.

Now, in the age of Yahoo!, Bing, Google, and others, each is attempting to outperform the other in terms of search engine performance. Many search engines are accessible nowadays, however some are more popular due to their crawling and ranking algorithms, such as Google, Yahoo!, and Bing, for example. When a user looks for information using these search engines, he generally has a notion of what he wants but is unable to formalise the question. Hundreds of millions of online pages are downloaded, indexed, and stored by the search engine. Every day, they respond to tens of millions of questions.

As a result, determining the nature of the information needs underlying Web users' searches has become a significant research challenge. As a result, web mining, web categorization, web optimization, and ranking mechanisms become more important for successful retrieval and searching, which often entails scanning through vast amounts of web information. Because the quantity of data available on the internet now exceeds millions of gigabits, it is critical to employ effective search strategies in order to index and rank such vast amounts of data. Crawling the Internet for all data, indexing all data, applying query classification algorithms and query optimization techniques, ranking these indexed documents to give a clear separation between the documents that are more frequently viewed and the ones that are not, and displaying the best results are all steps involved in implementing a successful and more efficient search engine. The three primary forms of information are content of data, structure of data, and log data. Web mining research has been separated into three areas based on these three types of information: web content mining, web structure mining, and web use mining. The goal of online content mining is to extract usable information or knowledge from the contents of web pages. Content on the internet.

Publisher: Noida Institute of Engineering & Technology,
19, Knowledge Park-II, Institutional Area, Greater Noida (UP), India.

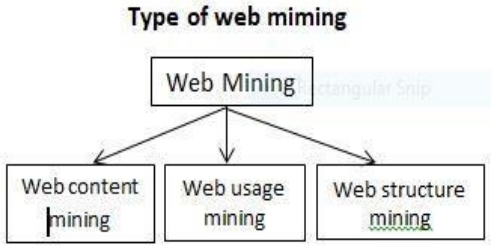

Fig 1.0

An Approach to Web Data Mining

The web is a strong platform for uploading and retrieving data as well as mining important data. Web Data research has faced several obstacles as a result of the vast, dynamic, diversified, and unstructured character of web data.

Web mining is a prominent issue in study because it combines two active research areas: data mining and the World Wide Web. Database, information retrieval, and artificial intelligence all intersect in web mining. Web mining is a technique for extracting the most interesting and valuable patterns and implicit data in terms of information from World Wide Web activity. In comparison to other types of data mining and retrieval, web mining is wwwShown in fig 1.1.

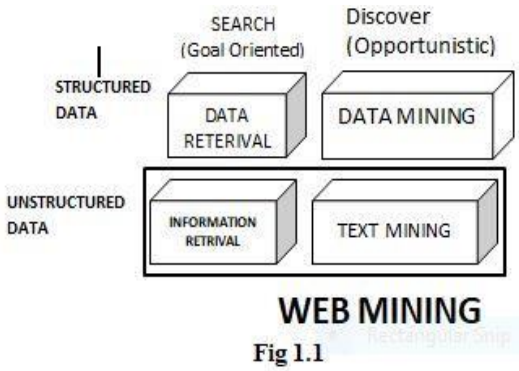

Fig 1.1

Web Content Mining

Web content mining is the technique of extracting complete meaning from web documents' content. Because the majority of web information is text, web content mining is similar to text mining. Online content mining, on the other hand, is distinct from data mining in that web data is typically semi-structured and/or unstructured, whereas data mining focuses on structured data. Because of the semi-structured nature of the web, web content mining differs from text mining, which focuses on unstructured materials. Images, music, video, text, and structured information such as tables and lists make up these online pages. Web content mining is a procedure that goes beyond keyword extraction since web pages do not have machine-readable semantics.

1) Structured text mining. 2) Unstructured text mining. 3) Semi structured text mining .4) Multimedia mining.

Web Structure Mining

Web structure mining seeks to discover the link structure of hyperlinks at the inter-document level, resulting in organised summaries of material on web sites. Based on the topology of hyperlinks, online structure mining classifies web pages and produces information such as similarity and connection between diverse websites. Web structure mining may also be used to determine the structure of a Web document. This kind structure mining may be used to expose the schema of web pages, which is useful for navigation and allows you to compare and integrate web page schema. If a web page is directly connected to another online page, or if the web sites are neighbours, we want to know what the relationship between those web pages is. The relationships may be of one

Publisher: Noida Institute of Engineering & Technology,
19, Knowledge Park-II, Institutional Area, Greater Noida (UP), India.

of two types: they could be connected by synonyms or ontology, or they could have similar contents since they are both hosted on the same web server and hence generated by the same individual. Mining the structure of the web Identifies links between online sites and concentrates on resolving issues.

- Reducing the number of irrelevant search results
- Assisting in the indexing of material on the internet.

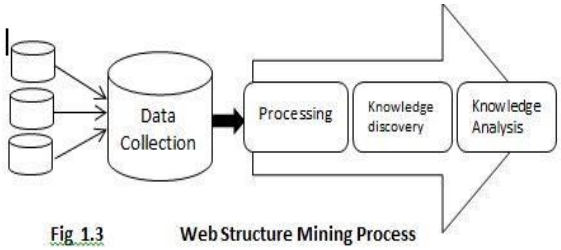

Fig 1.3 Web Structure Mining Process

Analysis of Web Usage

The method for obtaining valuable usage patterns from online data. Patterns in online users' browsing and navigation data are discovered. Web use mining has long been a useful tool for gaining a better understanding of how people use the internet. Most online use mining research nowadays focuses on the web server side, with the primary goal of the study being to improve a website's service and server performance. The principal web server logs are a data source for web use mining. The practise of discovering browsing trends by studying a user's navigational activity is known as web use mining. This information takes as input use data, which is data stored in web server logs that records user visits to a website.

Web use mining is concerned with the identification of economically valuable information based on internet users' interactions with websites in order to create customised web pages or provide improved search engines. Meaningful data may be extracted from online usage statistics. The method of collecting information patterns from internet log data using data mining techniques is outlined. A collection of methods for generating patterns and learning from online usage data.

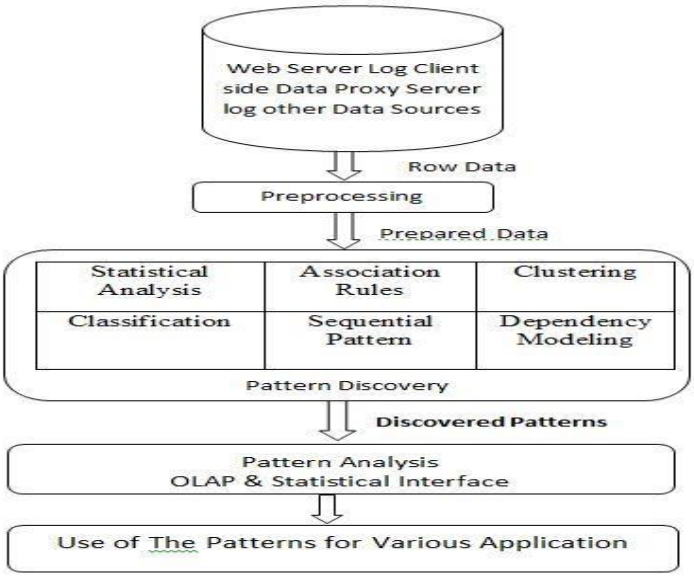

Fig : 1.4 Web Usage Mining Process

Publisher: Noida Institute of Engineering & Technology,
19, Knowledge Park-II, Institutional Area, Greater Noida (UP), India.

CONCLUSION

Web mining is a burgeoning study field in the mining industry. Finding relevant stuff on the internet is a regular challenge. However, the majority of search engines do not always give the best possible results that correspond to the user's demands. The three areas of web mining, Web content mining, Web Structure Mining, and Web Usage Mining, all play an important role in extracting particular data from the web. The paper's suggested technique focuses on an integrated strategy of web content mining-free text, web structure-hyperlinks, and web usage-web log data to enhance the performance of information (Set of DATA) retrieval in web search engine results.

REFERENCES

[1] I Mele , "Web Usage Mining for Enhancing Search – Result Delivery and Helping Users to Find Interesting Web Content ,"ACM SIGIR Conf. Research and Development in Information retrial (SIGIR' 13), PP. 765-769, 2013.
[2] Raymond Kosala, Hendrik Blockeel, Web Mining Research: A Survey, ACM SIGKDD, July 2000.
[3] J. Srivatava, R. Cooley, M. Deshpande , P.-N Tan Web Usage mining : Discovery and Application of usage patterns from Web Data. SIGKDD Explor. News1(2):12 ,2000
[4] P. Sudhakar , G Poonkuzhali, R. Kishor Kumar , "Content Based Ranking for Search Engine ," Proc. International multi Conference of Engineers and Computers Scientists(IMECS 12) ,2012.
[5] Ramakrishna M.T Gowdar , L.K Havanur ,M.S Swamy (2000) ," Web Mining : Key Accomplishment Application and future Directions", International conference on Data storage and Data Engineering(DSDE)pp 187-191, 2010.
[6] WangBin and LiuZhijing , "Web Mining Research ", in Proceeding of the 5th International Conference on Computational Intelligence and Multimedia Applications(ICCIMA'03) 2003.
[7] Hua-Jun Zeng , Qi-Cai He Zheng Chen , Wei –Ying Ma ,"Learning to cluster Web Search results ", ACM ,2004.
[8] Duhan , N., A.K Sharma , and KK Bhatia. Page Ranking Algorithms: A Survey. In Advance Computing Conference, 2009 .IACC 2009 . IEEE International .2009.
[9] Xu, J. and Li, H. "Ada Rank: A Boosting Algorithm for Information Retrieval, Proceeding of the 30th Annual International ACM SIGIR Conferene , Amster –dam, Netherlands, 2011.
[10] Laxmi Choudhary and B. Shankar Burdak , "Role of Ranking Algorithms for International Retrieval " International journal of Artificial Inteligence and Application (IJAIA) . VOL.3 No 4 July 2012
[11] Hao Chen and Susan Dumais ,"Brining order on the web :Automatically Categorizing Search Results",ACM,2012.
[12] T. Munibalaji, C. Balamurugan –Analysis of Link Algorithms for Web Miniing , International Journal of Engineering and Innovative Technology (IJEIT) ISSN 2277-3754 Volume 1 Issue 2 February 2012 , pp-81-86.
[13] Johnson , F., Gupta, S.K Web Content Minings Techniques : A Survey , International Journal of Computer Application. Volume 47 – No 11 , p44 June(2012)
[14] Wenpu Xing and Ali Ghornbani , "Weighted PageRank Algorithm ", IEEE,2004.
[15] RonGiles, How Search Engines Work, Available: http://www.website-consultant.co.nz/website/Top+10+Search+Engine+Ranking+Factors/How + Search +Engines+work.html.
[16] Nicholas O. Andrews and Edward A. Fox," Recent Development in Document Clustering Techniques", Dept of Computer Science, Virgina Tech 2007.
[17] C.D . Mining, P. Raghavan , and H. Schtze. "Introduction to Information Retrieval" Cambridge University Press.
[18] Garza Villarreal, S.E Martinez Elizalde , L., and Canseco Viveros ,A. Clustering hyperlinks for topic extraction: An exploratory analysis. In proceding of the 2009 Eighth Mexican International Conference on Artificial Intelligence, MICAI '09, pages 128-133,
[19] Washington, DC , USA ,2009 IEEE Computer Society. ISBN 978-0-7695-3933-1,2009
[20] IBM. Knowledge discovery and Data Mining. Available online http://researcher.watson.ibm.com/research/view_group.php?id=144(accessed on 8 June-2018)

Publisher: Noida Institute of Engineering & Technology,
19, Knowledge Park-II, Institutional Area, Greater Noida (UP), India.

[21] X Wang &C- X Zhai , "Learn from Web Search Logs to Organize Search Results", Proc. 30th Ann. Int. ACM SIGIR Conference Research and Development in Information Retrieval (SIGIR ' 07) , pp 87-94, 2007.
[22] AnHai Doan, Jayant Madhaven Pedro Domingos and Alon Halvey , "Learning to map between Ontologies on the Semantics Web ", in proceeding of the ACM WWW Conference 2002.
[23] M. Sanderson ,"Retrieving with good sense", Information Retrieval. 2(1):49-69,2000.
[24] S.M Beitzed , E. C. Jensen , D. D. Lewis, A Chowdhury,&O. Frieder ,"Automatic classification of web queries using very large unlabeled query logs" ACM Transactions on information system , 25(2) (Article no 9) 2007.
[25] Dr. Mohammad Shahid "Knowledge Discovery On The Internet (Web Mining Tooland Technique" INDIAN JOURNAL OF RESEARCH(2012)6,ANVIKSHIKI ISSN 0973-9777 Advance Access publication 20 July. 2012
[26] Dr. Mohammad Shahid" Taxonomies, Challenge And Approaches To Automotive Web Query Classification "2 ND INTERNATIONAL COFFERENCE ON COMPUTER APPLICATION 2012 ICCA'12 PONDICHERRY, INDIA.
[27] Verma, Garima; 'Agile Software Development : An Alternative Approach to Software Development',Volume No.2,Issue No.2,2014,PP.020-023,ISSN :2229-5828
[28] H.S. Pali, N.kumar; 'Biodiesal Production from Sal (Shorea Robusta) Seed Oil',Volume No.2,Issue No.2,2014,PP.024-029,ISSN :2229-5828
[29] Maneesh Kumar , Rajdev Tiwari, Rajeev; 'Automated Test Case Genration on the Basis of Branch Coverage Using Teaching Learning Based Optimization',Volume No.2,Issue No.2,2014,PP.030-036,ISSN :2229-5828
[30] Meenakshi Saini,Nitin Kathuria, V.K.Pandey; 'L Slot Circularly Polarized Broadband Antenna',Volume No.2,Issue No.2,2014,PP.037-041,ISSN :2229-5828
[31] Patro, B.Narasingh; 'Content Based Image Retrival Thro-ugh Features Like Color ,Texture and Shape',Volume No.2,Issue No.2,2014,PP.042-046,ISSN :2229-5828

Publisher: Noida Institute of Engineering & Technology,
19, Knowledge Park-II, Institutional Area, Greater Noida (UP), India.