

Evaluation of Web Miming Technique for Search Personalization using semantic optimization

Mahima

Assistant Professor, Gurugram University, Gurugram, Haryana, India

ARTICLE DETAILS

Article History

Published Online: 15 March 2019

Keywords

Web Mining, Content Mining, Usage Mining, Knowledge Extraction

ABSTRACT

Web mining is a method of text and data mining strategies to consequently find and recover attractive data from the web and gave pleasing outcome to clients. It utilizes three essential systems to extricate important data that is content mining, structure mining and utilization mining to separate significant information. To find important information from the substance of web assets content mining is use as an apparatus. Structure mining is the procedure that concentrates learning from the interconnected archive on the web and utilization mining is a basic system of consequently looking and examination the client connection designs with web servers.

1. Introduction

During the past few years World Wide Web has become a main source of information acquisition. The existence of such abundance of information, in combination with the dynamic and heterogeneous nature of the web, makes web site exploration a difficult process for the user. Websites personalization is the efficient and effectual approach to meet up the requirement of efficient web navigation. The following segment of this manuscript reviews some novel techniques that use the content semantics, structural properties of a web site and usage data of the specific user in order to improve the effectiveness of web personalization.

Web Personalized Mining is a consequence of Web 2.0 (second generation of web) that is supported with programmed semantic agents for processing user data to assist the web users on providing personalized services. Web personalization is the process of personalizing the contents of websites to the need of specified users. For better understanding we do review the web personalization with special reference to the web usage mining how usage data in addition to the structural information of a web page can help in the arena of web personalization [11].

The web personalization is the customization of information or services provided by a website to users based on the pattern discovered by their navigational behavior, recorded in website's log. This information is combined with the content and the structure of the web site, as well as the interests of the user. The personalization process can generate dynamic results, which depends on behavior of user.

2. Literature Review

This section describes the work of eminent researchers and highlights the challenges, which still require to be addressed

In this paper [1], the authors presented an adapted version of simulated algorithm to build up an intelligent personal spider agent that is based on retrieving and analysis of web documents automatically. It is no less than any challenge to find the information or suitable homepages to the users of Internet. The main difficulty with the existing searching engines or spiders is that only a few spiders do have the

communication capabilities between the spiders and the users who dispatched the spiders. Since there is no communication, it is too difficult for the users to trace or to realize the progress of searching and have to tie themselves to the interface. This manuscript presents an agent which uses the combination of CGI and Java to build up the user interface. It also allows the users to keep track of the progress of searching. The users can also make changes on the searching parameters, such as, number of homepages to retrieve. There were several algorithms have been used to develop spiders, for example, best-first searching and genetic algorithms. In this paper, we will discuss the spider that we developed with the hybrid simulated annealing algorithm. We have made a comparison to compare the performance of spiders that developed with best-first search [16]. At the end, results are depicted in this paper.

In this manuscript [2], the authors have presented three different techniques of web mining, namely web content mining, web structure mining and web usage mining and presented the significance of introducing the web mining techniques in the area of web personalization. At the last they implement how Web mining techniques can be apply for the Customization i.e Web personalization.

This paper [3] presents an overview of current techniques and applications of Semantic Web, Elearning based on web mining which has already become a core component of education. The author considers that student centralized education, in other words education personalization depending students' learning approach, is the key on educational field. They also mentioned, e-learning portals allows the students to try numbers of different techniques, make mistakes without disclosing themselves, e-learning portals provides a better personalized learning mechanism for all, information is distributed at the same level of quality, students can use the service for getting knowledge out of available information on their own schedule at any time they want and it's much more easier to update content for any topic [12].

In [4] proposed a novel technique that personalizes web search result through ontology based user profiling in form of RDF and query alternation. First, pertinent search phrase for particular user from user log file is identified by analyzing web log file. These phrases are appended to user's ambiguous query. Second, the anticipated approach also recognize user

interest on retrieved links by combing the user interest value generated from VSM (Vector Space Model) and actual page rank of that link. A novel technique is used for finding the ranking of the link. Third, the approach also suggests some keywords that help to incorporate user's current interest.

1. Web Mining

Web mining is a method of text and data mining strategies to consequently find and recover attractive data from the web and gave pleasing outcome to clients. It utilizes three essential systems to extricate important data that is content mining, structure mining and utilization mining to separate significant information. To find important information from the substance of web assets content mining is use as an apparatus. Structure mining is the procedure that concentrates learning from the interconnected archive on the web and utilization mining is a basic system of consequently looking and examination the client connection designs with web servers [13].

2.1 Web Content Mining

Web Content Mining (WCM) is the method of retrieving information or descriptions from the content of web elements. The resource discovery in the WCM is depending on ideas such as searching and browsing on agents methods [14]. In addition, WCM is an regular method that work beyond extracting keywords as there is no machine-readable semantic in the content of a content information. In a depiction that could be utilized by computers, some methods suggested restructuring the document content. Using covering to map files to several data model is the normal method to exploiting known configuration in papers [16]. Methods used lexicons for content explanation are taken as difficult task by the investigators. These types of mining techniques work to determine all sub-links of the hyperlinks in a document so as to produce the structural description on a web page. There are two gatherings of web a content mining technique that mine the substance of reports legitimately and recover the substance inquiry of different apparatuses like web search tools. Web Content Mining includes multimedia contents such as text, video, image and audio data. Web Content Mining also works with the mining of concept hierarchies from the Web pages and their usual classification. The key benefit of web content mining is to extract required information on the web.

2.2 Web Structure Mining

Web Structure Mining (WSM) is the internet task of filtering information that discovers links between web pages means how the web documents are interlinked with one another in the Web. The Web can show additional data than the document data [17]. For instance, joins indicating an archive show the notoriety of the report while joins leaving a record demonstrate the lavishness or maybe the assortment of themes canvassed in the report. Although mining the formation of the data, a tree like structure is utilized for evaluating and defining the XHTML or the HTML tags in the web link.

2.3 Web Usage Mining

The method of retrieving required methods in internet entrance records is Web Usage Mining (WUM), also known as Web Log Mining. Web links verification and gather user interaction information when receiving resource requests. Text mining methods are used here to find data and information in the visitor's website's browsing nature [14]. Analyzing internet

access logs from separate websites can assist user conduct and internet structure to improve the design of colossal resource collection. In Web Usage Mining there are two primary trends driven by apps, namely common access method accessing and formalized usage techniques [9]. The common access design following dissects the web logs to comprehend get to examples and patterns. These examinations can reveal insight into improved formation and gathering of asset suppliers. In order to take out an efficient web usage mining, the submission of efficient mining and information renovation methods are required before study is carried out.

3. Inconceivable correlation among semantic web and web mining

The purposed of this examine work is to obtain a elevated considerate of user performance at the point in instance of Web browsing in order to improved technique for the users on the Web content.

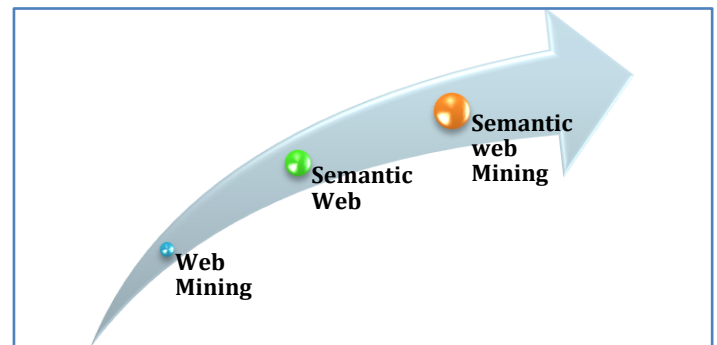


Figure.1: Description of Semantic Web Mining

The term Semantic Web infers an intellectual system i.e. a significant web. It plans to influence PCs to comprehend the significance of data on the pages instead of simply introducing them to clients [6]. The thought is to construct Web astute and computer meaningful by giving apparatuses to discover trade and decipher data to a restricted degree by including data.

Despite the fact that the Semantic Web (SW) has been the vision for the up and coming age of the Web, where data is wanted to be valuable for the general population as well as for the computers but one noteworthy snag in executing SW is that computer don't have the sort of memory that individuals have. Since individuals utilize dialect from premature times to their works so it make simple for them to formulate associations among various keywords and ideas and to formulate deductions in light of settings. Yet, this isn't the situation with PCs. To make PCs ready to comprehend importance of words and connection between various words, they should have reports portraying every one of the words and rationale to make fundamental associations. With the assistance of ontologies SW accomplishes this objective, where philosophy is a record or document that formally characterizes terms and their relations [4]. We can state that, it is kind of vocabulary that depicts objects and their connections in an area.

The exponential progressions in web advancements have empowered clients to encounter upgraded conveyance of customized administrations and data through the reconciliation of different existing innovations. Be that as it may, the current concentrated stage and normally an expansive server can't guarantee the adaptability, adaptability, unwavering quality,

non-repetition of data gave to clients. Along these lines the requirement for explore exercises in web administration and upgrades by building up a standard, adaptable however savvy, versatile and disseminated system for the help of varied framework is evident. The current situation requests the designation of knowledge of web to a littler however new shrewd group of parts known as insightful operators. The up and coming segments give an understanding into the present situations and fur

4. Conclusion

In recent research, the web recommendation community has spotlighted on categorize standards that are able to hold heterogeneity and fetch user recommendation problem on web. However issue still remains because various established techniques are no longer spirited in dynamic environment. By handling this issue, we have proposed a Web Personalization Recommendation System (WPRS), which is competent to handle heterogeneity issue and provide meaningful information in non-redundant way.

References

1. Yang, C. C., Yen, J., & Chen, H. (2000). Intelligent internet searching agent based on hybrid simulated annealing. *Decision Support Systems*, 28(3), 269-277.
2. Ratnakumar, A. J. (2010). An implementation of web personalization using web mining techniques. *Journal of Theoretical and applied information technology*, 18(1), 67-73.]
3. Mustapaşa, O., Karahoca, D., Karahoca, A., Yücel, A., & Uzunboylu, H. (2010). Implementation of semantic web mining on e-learning. *Procedia-Social and Behavioral Sciences*, 2(2), 5820-5823.
4. Makvana Kamlesh, Pinal Shah and Parth Shah. "A novel approach to personalize web search through user profiling and query reformulation" *Data Mining and Intelligent Computing (ICDMIC)*, 2014 International Conference on IEEE, 2014.
5. UTKALUNIVERSITY, R. R. (2010). Integration of web mining and web crawler: Relevance and state of art. *Integration*, 2(03), 772-776.
6. Pannu, Mandeep, RachidAnane, and Anne James. "Hybrid profiling in information retrieval." *Computer Supported Cooperative Work in Design (CSCWD)*, 2013 IEEE 17th International Conference on. IEEE, 2013.
7. Joshi, C., & Singh, U. K. (2015). A Novel Approach towards Integration of Semantic Web Mining with Link Analysis to Improve the Effectiveness of the Personalized Web. *International Journal of Computer Applications*, 128(11), 1-5.
8. Cooley, R., Mobasher, B., & Srivastava, J. (1997, November). Web mining: Information and pattern discovery on the world wide web. In *Tools with Artificial Intelligence*, 1997. Proceedings., Ninth IEEE International Conference on (pp. 558-567). IEEE.
9. Liu, T. (2012). *Data Mining over Hidden Data Sources*. The Ohio State University.
10. Pierrakos, D., Paliouras, G., Papatheodorou, C., & Spyropoulos, C. D. (2001, December). KOINOTITES: A web usage mining tool for personalization. In *Proceedings of the Panhellenic Conference on Human Computer Interaction*.
11. Sunita, Rana, (2020), Web URLs retrieval with least execution time using MPV clustering approach, *International Journal of Information Technology-Springer*, pp 1-9.
12. Mamta Sharma, Vijay, Repossession of recurrent web document with an advent of clustering algorithms in the Semantic Web mining, *JOURNAL OF ARCHITECTURE & TECHNOLOGY*, Vol 12, no 1, pp 163-168.
13. Sharma, Rana, (2020), Evaluation of query in Semantic interoperability, *PAIDEUMA JOURNAL*, Vol .13, No.1, pp 78-84.
14. Sharma and Vijay, (2020), A Systematic Literature Review Of Web Search Personalization, *Recent Advances in Computer Science and Communications*, Vol 13, No 6.
15. Sunny Sharma and Vijay, (2020), Web Search Personalization Using Query Optimization, *International Journal of Future Generation Communication and Networking*, Vol. 13, No. 1, pp. 125-136.
16. Sunita, Vijay Rana, (2020), Impact Of The Clustering Algorithms And Its Applications: Machine Learning, *International Journal of Future Generation Communication and Networking*, Vol. 13, No. 1, pp. 115-124.
17. Sunita, Rana V. (2020) A Web Extraction Browsing Scheme for Time-Critical Specific URLs Fetching. In: Singh P., Kar A., Singh Y., Kolekar M., Tanwar S. (eds) *Proceedings of ICRIC 2019. Lecture Notes in Electrical Engineering*, vol 597. Springer, pp 617-626.