

Design of e-Learning Applications Through Web Mining

T. Gowthami¹, P. Boopathi²

Research Department, STC College, Thippampatti^{1,2}

Abstract: The E-Learning has become matured learning paradigm with the advent of web based learning and content management tools, and shifted the focus of entire world from instructor centric learning paradigm to learner centric approach. The Web Mining plays a very important role for the E-learning systems. In personalized E-Learning system, user customizes the learning environment based on personal choices. In a general search process, a hyperlink which is having maximum number of hits will get displayed first. For making a personalized system history of every user need to be saved in the form of user logs. The proposed system provides a new approach with combination of web usage mining, HIT algorithm and web content mining. It combines hits results on user logs and web page contents with a clustering algorithm called as Lingo clustering algorithm. In this paper, we will describe an approach aiming to achieve personalization in e-Learning services using web usage and Content mining. This paper discusses about the benefits and usefulness of web mining in e-learning.

Keywords: Web usage mining; web content mining; web personalization; e-learning system, Lingo; HITS.

INTRODUCTION

World Wide Web has developed into an intense and intelligent media for the correspondence of data. Distinctive clients geographically situated at better places need to get to the unique data sorts effectively. The routes of clients with web locales produce a gigantic store called Web get to log record which can be examined to find the navigational examples of the clients. The investigation of Web get to log record is named as Web Usage Data mining. Data blast on the Internet has set levels of popularity on search engines. Individuals are a long way from being happy with the execution of the current search engines, which regularly return a large number of documents in light of a client query. A hefty portion of the returned documents are unimportant to client's need. The exactness of ebb and flow search engines is well under individuals' desires. To discover more exact responses to a query another era of search engines, question noting frameworks have showed up on the web. Dissimilar to the customary search engines that lone utilize watchwords to coordinate documents, this new era of framework tries to coordinate documents this new era of frameworks tries to comprehend the clients question and propose some comparable inquiries that other individuals have regularly approached and for which the framework has the right answers. Truth be told the right answers have been arranged or checked by human editors much of the time at that point ensure that in the event that one of the recommended questions is genuinely like that of the client, the appropriate responses gave by the framework will be applicable. The presumption behind such a framework is, to the point that many individuals are occupied with similar inquiries the much of the time inquired. Keeping in mind the end goal to beat the issues, some web search engines have actualized techniques to recommend elective questions to clients. The reason for these techniques is to enable clients to indicate elective related questions in their search procedure all together either to clear up their data needs or to reword their query plan to recover more related search comes about. The procedures utilized as a part of these restrictive business frameworks are generally secure however watch that those proposed inquiries came back from these search engines are somewhat comparative in their terms. This may infer that those recommended questions are probably going to be created by utilizing straightforward query extension strategies. For example, if the client searches for Yahoo! search motor the accompanying related questions are exhibited: messenger, best yahoo mail. Be that as it may, as we can envision, there are a decent number of different questions identified with mail yet probably not having the term yahoo expressly in their term vectors.

LITERATURE SURVEY

[1] Sandesh Jain et.al proposed that the personalization of e-Learning services using web mining and semantic web. [2] Umadevi et.al proposes the Design of e-Learning application through web mining. [4] Shi Na, Guan yong, and Liu Xumin proposed that the drawbacks of the standard K-means algorithm, such as the need to compute the distance from each data items to all cluster centroids, is eliminated by introducing two simple data structures. One data structure is used to hold the label of cluster and the other data structure is used to store the distance from every data item to the nearest cluster obtained in each step that can be used to find the distance in the next step. The main downside of K-means is to decide the number of clusters and initializing the centroids for the first iteration. [5] Parul Gupta and A.K.

Sharma have obtained aggregate usage profiles from the discovered pattern to provide effective recommendation systems for real time Web personalization systems. [6] BamshadMobasher, Honghua Dai, Tao Luo, and Miki Nakaguva have presented a clustering technique which forms clusters from the set of documents. Every document is assigned with an identifier, so that closer document identifiers are assigned to similar documents. They have proposed an improvement for this clustering algorithm to form super clusters from mega clusters which are formed using similar clusters in a hierarchical clustering process. Their work describes the search process optimization. [7] Dae-Won Kim, KiYoung Lee, Doheon Lee and Kwang H. Lee have presented a scheme by combining Spherical K-means algorithm and flock of agent based FClust algorithm. Spherical K-means algorithm is mainly used for clustering sparse and high dimensional data. FClust is mainly applicable for representing high dimensional data in a visualization plane..

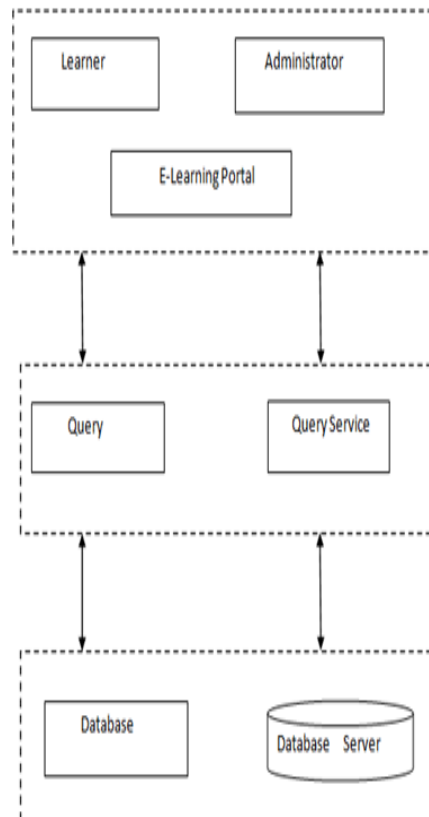


Figure 1: Search Component of Service Oriented Reference Architecture for Personalized E-Learning System.

Web Mining Applicable To E-Learning

A. Web Mining Techniques

Web mining is one of the important branches in data mining. Extraction of useful information or pattern from the web data is called web mining. It can be classified into three following categories as shown in figure (2).

1. Web structure mining
2. Web content mining
3. Web usage mining
4. Web Structure mining

- **Web Content mining**

Web content mining refers to the extraction of useful information from the Web document's content. The content of the web page consist of text, images, audio, video etc., it includes the techniques like clustering or associating web pages according to the respective branches. It also helps in discovering patterns in web pages to mine useful data.

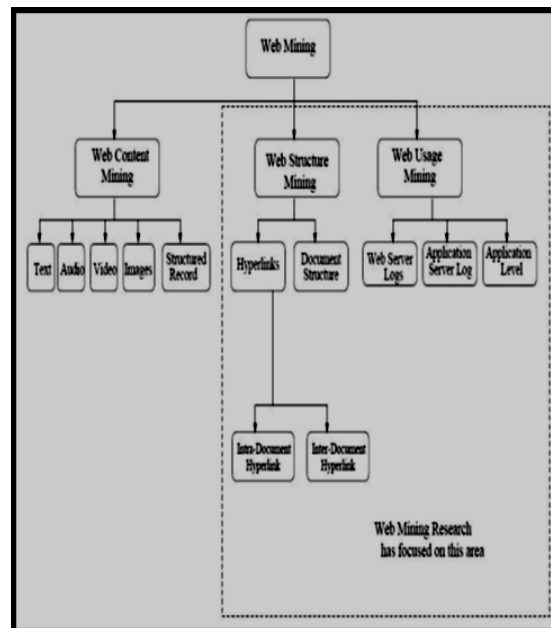


Figure 2. Structure of Web mining

- **Web Usage mining**

Web Usage Mining is the type of data mining techniques to discover interesting usage patterns from Web data, in order to understand and better serve the needs of Web based applications. Usage data captures the identity of Web users along with their browsing behavior at a Web site. Moreover, the most important research initiatives in Web usage mining and personalization area are presented. But this is not as much effective method for personalization.

- **Web structure mining**

Web structure mining targets on analysis of the web and one of its uses is to identify more preferable documents helps to discover similarities between web sites or discovering significant sites for a specific topic or branch or in discovering web communities. It is also used to reveal the schema of web pages

B. Web mining based E-learning System:

The E-learning system consist three parts. Teaching resource library, learning platform and user. Education resource library is a storage server to store different types of resource which is related to education. The learner of that web based system is the user. Web server is the Learning platform that gives web based learning platform to user.

The Proposed System Architecture

The objective of this approach is to provide users with the information they want or need, without expecting from them to ask for explicitly. Figure (3) represents proposed architecture for personalized E-learning system combining HITS algorithm on user logs and Lingo clustering algorithm.

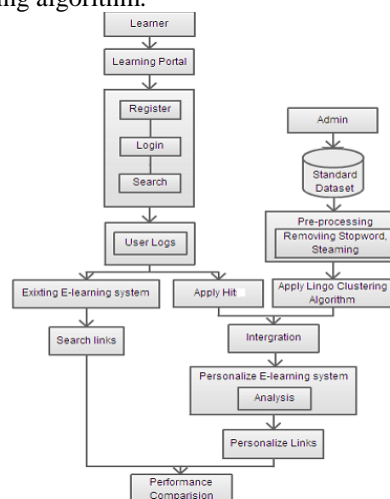


Figure 3. Architecture of proposed approach

Admin is responsible for forming a standard dataset including learning objects. This dataset is preprocessed by removing stop words and stemming. A new user first makes registration to the learning portal. When user login to the system using his own user name and password and search for a particular topic, at server side user logs are stored for that search.

Then HIT algorithm is applied to those logs to increase the weight of that log. For proposed approach, preprocessed data is given to content mining using Lingo clustering algorithm. Clusters are formed from preprocessed data. Final results are calculated by combing user logs, Hits algorithm and clustering results.

Algorithm of the Proposed System

A.HIT Algorithm

The Hypertext Induced Topic Search algorithm is an algorithm used to find documents relevant to a particular keyword topic. Created by Krishna Bharat while he was at Compaq Systems Research Center and when you enter a query or keyword into the Google search engine, the Hypertext Induced Topic Search algorithm helps to find relevant keywords whose results are more informative about the query or keyword.

The algorithm operates on a special index of expert documents. These are pages that are about a specific topic and have links to many non-affiliated pages on that topic. Pages are defined as non-affiliated if they are authored by people from non-affiliated organizations. Results are ranked based on the match between the query and relevant descriptive text for hyperlinks on expert pages pointing to a given result page.

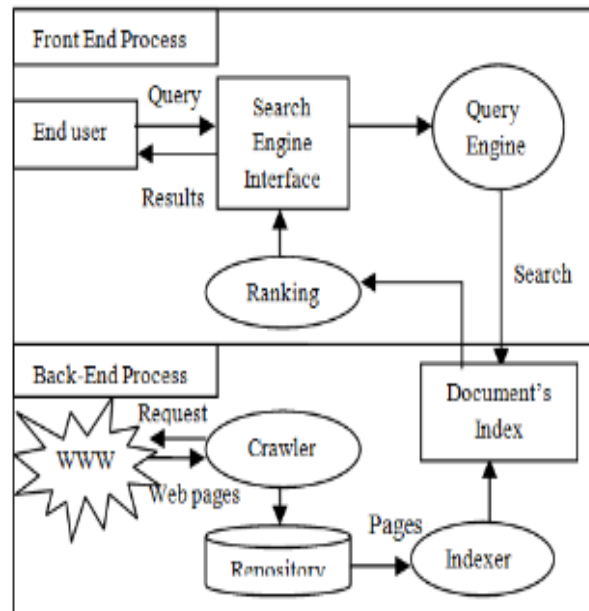


Figure 4. Search Engine Process.

Comparison of Page Rank and HITS Algorithm

Criteria	PageRank	HITS
Mining Techniques	Web Structure	Web Structure and Web Content
Working Process	Computes Rank values at index time and results are stored on the priority of pages.	'n' highly relevant pages rank are computed.
Input Parameters	Inlinks to a page.	Inlinks, outlinks and content
Relevancy	Less (as this algorithms ranks the page at indexing time)	More (as this uses hyperlink structure and also consider the content of the page.
Quality of Result obtained	Medium	Less than PageRank Algorithm
Advantages	➔ Query-time cost of incorporating precomputed PageRank importance score	➔ HITS is a general algorithm used for calculating the

	<p>for a page is low.</p> <p>→ PageRank generated using the entire Web graph, rather than a small subset, it is Less susceptible to localized link spam.</p> <p>→ PageRank may be used as a methodology to measure the impact of a community like the blogosphere on the overall Web itself.</p>	<p>authority and hubs in order to rank the retrieved data.</p> <p>→ The basic aim of that algorithm is to induce the Web graph by finding set of pages with a search on a given topic. (query)</p> <p>→ Results demonstrate that it is good in calculating the authority nodes and hubness</p>
Disadvantages	<p>→ Rank Sink</p> <p>→ Spider Traps</p> <p>→ Dangling Links</p> <p>→ Dead Ends</p> <p>→ Circular References Effect of additional pages</p>	<p>→ Irrelevant authorities</p> <p>→ Irrelevant Hubs problem</p> <p>→ Mutually reinforcing relations between hosts problematic</p>
Search Engine	Used in Google	Used in IBM search Engine

CONCLUSION

The performance of the system is evaluated under different settings and in comparison with the previous method which is based only on the usage mining. The application can be used for personalized recommendation to give personalized recommendation based on users browsing history. Throughout this paper we have discussed many aspects of research for personalized E-learning system. Based on existing limitations, in this paper a new web mining approach based on combination of web usage mining and web content mining (HITS algorithm) is presented which is showing the better performance improvement as compared to the existing method.

FUTURE WORK

In future we discover the learner's time distribution pattern to realize personalized curriculum organization, discover the learning behavior pattern to build up a series of feedback and motivation system and we will give different training according to different learners' levels. For future work, research can be done in developing integration strategies for approaches that can accurately predict student performance in courses and approaches that help a select a subject or courses based on student interests.

REFERENCES

- [1] Sandesh Jain, Dhanander K. Jain, Harihar Bhojak, Ankit Bhilwar, and Mamatha J., "Personalization of e-Learning Services using Web Mining and Semantic Web", International Journal of Machine Learning and Computing, Vol. 2, No. 5, October 2012.
- [2] Umadevi, B.Uma Maheswari, P.Nithya, "Design of E-Learning Application through Web Mining", International Journal of Innovative Research in Computer and Communication Engineering (An ISO 3297: 2007 Certified Organization) Vol. 2, Issue 8, August 2014.
- [3] Manasi Chakurkar1, "A Web Mining Approach for Personalized E-Learning System", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 5, No. 3, 2014 Page 51, www.ijacsa.thesai.org.
- [4] Shi Na, Guan yong, and Liu Xumin, "Research on Kmeans clustering algorithm" Third International Symposium on Intelligent Information Technology and Security Informatics, 2010 IEEE.
- [5] Parul Gupta and A.K. Sharma, "A Framework for Hierarchical Clustering Based Indexing in Search Engines", BIJIT - BVICAM's International Journal of Information Technology, July – December, 2011; Vol. 3 No. 2; ISSN 0973 – 5658.
- [6] Asur, S., and Huberman, B.: "Predicting the future with social network." Web Intelligence and Intelligent Agent Technology (WIIAT), 2010 IEEE/WIC/ACM International Conference on. Vol. 1. IEEE, 2010.
- [7] Au Yeung, C. M., and Iwata, T.: Strength of social influence in trust networks in product review sites. In Proceedings of the fourth ACM international conference on Web search and data mining (pp. 495-504). ACM, 2011.
- [8] Bakshy, E., Hofman, J. M., Mason, W. A., Watts, D. J.: Identifying influencers on twitter. In Fourth ACM International Conference on Web Search and Data Mining (WSDM), 2011.
- [9] Becker, H., Naaman, M., Gravano, L.: Beyond Trending Topics: Real-World Event Identification on Twitter. ICWSM, 11, 438-441, 2011.
- [10] Becker, H., Chen, F., Iyer, D., Naaman, M., Gravano, L.: Automatic Identification and Presentation of Twitter Content for Planned Events. In ICWSM, 2011.
- [11] Becker, H., Iyer, D., Naaman, M., Gravano, L.: Identifying content for planned events across social media sites. In Proceedings of the fifth ACM international conference on Web search and data mining (pp. 533-542). ACM, 2012.