WEB PAGE RE-RANKING TECHNIQUE IN SEARCH ENGINE

Ms.S.Muthukakshmi[1], R. Surya [2], M. Umira Taj [3]
Assistant Professor, Department of Information Technology, Sri Krishna College of Technology,
Kovaipudur, Coimbatore, India[1]

U.G Scholar, Department of Information Technology, Sri Krishna College of Technology, Kovaipudur, Coimbatore, India[2,3]

ABSTRACT –
Web-page re-ranking plays an important role in Web systems. Satisfactory and Useful knowledge discovery from Web usage data representation for effective Web-page recommendations are challenging and crucial. This paper proposes a novel method to efficiently provide better Web-page recommendation through semantic-enhancement by integrating the Web usage knowledge and domain. A number of effective queries have been proposed to generate Web-page candidates. Web page Recommendation systems can take advantage of semantic network to overcome common limitations of current systems and improve the recommendations' quality.

Keywords: Stemming algorithm, Web mining, query suggestion, recommendation system.

I INTRODUCTION

Social re-commendation has attracted a lot of attention recently in the search communities of information retrieval, machine learning and data mining, due to the potential value of social relations with the increasing popularity of social media. We have many popular commercial social recommender systems. A variety of social recommendation web page are proposed in literature, which can be generally grouped in two categories: related data and un-related data. Despite the studies of social web page recommendation most traditional social web page recommendation algorithms are based on batch training techniques which assume all user logs are provided in the user-item matrix. Such assumptions are not applicable for real-world web page recommendation applications. The user logs arrive sequentially in an on-line web application. The batch recommendation algorithm has to be retrained from scratch whenever new user logs are received, the training process becomes extremely time-consuming. If the size of training data is too large, it is difficult for handling all the data in the batch mode. User preference could drift over time in real-world web application, which makes the batch learning processes fail to capture such changes on time. In on-line web page recommendation, the number of user ratings collected at each time-stamp is much smaller than the ratings in the off-line web page recommendation, which means all the items have to be recommended in a good manner. Currently, knowledge sharing and social net-working sites like Twitter and face-book are popular platforms for users to produce shared opinions for the items like summary and item review.

In today's highly competitive e-commerce environment, the success of a site depends on the site's ability to turn casual browsers and retain visitors into potential customers. Recommender system technologies and Automatic personalisation have become critical tools, because they help engage visitors in-depth and more intimate level by tailoring the site's interaction with a visitor to their needs and interests. Web personalization can be defined as any action that tailors the Web experience to a particular user, or set of users. Let's understand the additional concepts which are going to be used throughout the paper before getting deep into the main concept.
The concepts that are used in this paper are given below:

- Web Mining Process
- Semantic Network
- Single word query
- Domain Knowledge
- Web Page Recommendation
- Stemming algorithm

The objective of this technique is to find the instances of the relations and concepts from the Web mining domain, such that they can be used to perform further web mining tasks in the system.

II LITERATURE REVIEW

In Classical web page recommendation the result which is generated in the form of web pages, hence there is no semantic meaning of common navigation profile system. There are many problems which are included in the classical Systems. Now one of these problems is NEW PAGE PROBLEM in the network. The classical systems failed to recommends the newly added pages or the products to the users. since these pages or websites are not in the current common navigation profiles.

In order to overcome the new page Problem the common navigation profile can be extracted in terms of information or semantic meaning. For that particular purpose stemmer should be used. But in classical systems stemmer was not used. Clustering is another problem in the existing systems. In this clustering number of recommended pages are increased. Unrelated pages or non-useful links are recommended which user never prefers. RDF and LCS algorithms are Previous techniques. . LCS: Longest Common Subsequence RDF: Resource Description Framework. In pattern matching technique the comparison of two sequences to determine their similarity is one of the fundamental problems. A sequence or a list of recommended products are generated to the user by the Longest Common Subsequence method which is useful for the latest trend shopping that is online shopping purpose. But there are other users who do not use such online shopping websites but they still want the personalization in their web page recommendation system.

The Numbering features should be represented by single components of vectors representing items. Which hold the exact value of that feature. But unfortunately classifiers of all types tend to take a long time to construct the feature. For example, if we wish to use decision trees, we should have one tree per user. Such approaches tend to be used only for relatively small size of problems Constructing a single tree does not require the profiles that we look at all.The problem is we have to consider many different predicates that could involve complex combinations of features.

Each time while user is searching she actually aims at finding some information concerning a particular or different subject. There are several methods to extract keywords that characterize the web content which should be the exact matching between the terms which determines the similarity between the documents. Classical systems use this approach to find the similarity between the documents. But by using this system, only the binary similarity is achieved, no actual semantic similarity is considered. Semantic similarity is very important as far as similarity between documents is considered. Several number of research approaches integrate other information sources. We require more abstract representation which enables a more flexible and uniform document matching process. It uses the semantic web structures, such as stemmer. Early systems possess a very common problem of caching of web pages. When a web user searches for an already cached page, this action is not recorded in the web site's log.

III PROBLEM STATEMENT

The problem in the existing systems is that, there was no personalization based on history knowledge or stemming and domain for the personal recommendation or personalized recommendation system. This system uses domain and history knowledge to enhance the personalization for web page recommendation purpose .This system presents a page rank method that makes use of representations of item. The Search method is achieved by using two different methods. A taxonomy-based similarity method is used to refine the item-user matching algorithm and a domain-based method makes inferences about user's interests, improving overall results. Then build the ontology for the web site using the significant terms and concepts extracted from documents. According to the data similarity of web documents which cluster them into different domain themes, the different themes implies different preferences..

IV PROPOSED MODEL

- Query processing
- Query clustering
- classification
- Re-ranking

These modules are explained in detail as follows:

- **Query processing**

Any personal documents such as user's browsing history and emails on an individual's system could be the data source for user profiles. This concentrate on recurrent terms limits the dimensionality of the document set, which further provides a clear understanding of individual's. These modules allows the search engine for better understanding of a user's session and potentially tailor that user's search experience according to their needs. Once query groups are identified, that the search engines can have a very good representation of the search content behind the present query using queries and clicks in the corresponding query group.

- **Query clustering**

User's queries can be divided into various query clusters. The Concept-based user profiles are engaged in the clustering process to reach personalization effect. There is more similar pair of concept nodes and after that, combine the most similar pair of query nodes, and so on. Each individual query are submitted by each user is treated as a separated node and each query with a user identifier. We carry out the clustering in a similar dynamic fashion, and thereby we first place the present query and clicks into a query group.

- **Classification**

To make sure that each query group contains similarly related and relevant queries and clicks, so it is essential to have a suitable relevance between the current query groups. We can assume that the users generally put out very similar queries and clicks within a short period of time. Then search history of a more number of users contains signals respecting query relevance, like which queries tend to be performed closely together. This method captures the relationship between queries generally leading to clicks on similar URLs. The query click graph and query reformulation graph from search logs, as well as how to use them to determine relevance between query groups or queries within a user's history.

- **Re-ranking**

Query groups has to be first treated in every query in a user's history as a query group, and then merge these query groups in an iterative fashion (in a k-means). However, this is not applicable in our concept for two reasons. The first is, it may have the repellent effect of changing a user's existing query groups, which are used for undoing the user's own manual efforts in constructing the history. Second, it involves a high-computational cost, for that we would have to repeat a more number of query group similarity computations for each new query.

V SYSTEM DESIGN AND IMPLEMENTATION

The proposed model is for personalized web page recommendation based on stemming, Domain Knowledge and semantic Network .The semantic web usage knowledge is an integration of both web usage knowledge and domain knowledge. The previous implementation of this model will demonstrate the result which produces significantly higher performances than the previous web page recommendation and web usage mining techniques. This paper proposes a framework which is formal for integrating full domain specific with personalization process based on Web usage mining. The classical

process was not acceptable by the user for the personalization purpose. The Semantic Web Mining Process will provide an infrastructure that enables not just web pages but programs, databases, sensors, services, personal devices and even household appliances to both consume and produce data on the web. The proposed system is used to represent the user profile and items based on stemming to provide applications with personalized services. The recommender system is proposed as Web Service and it is implemented as Domain Independent system. This kind of web service system uses both Explicit Feedback method and Implicit Feedback Collection Method which is used to obtain the individual's interest in particular search area. Feature Extraction is the first step of this proposed system. Feature extraction technique approach constructs relevant concepts by extracting features from web documents. Then it builds the root word for the website usage by extracting features from web documents. To cluster the web documents into different Semantic Themes, the semantic similarity of web documents is taken into consideration.
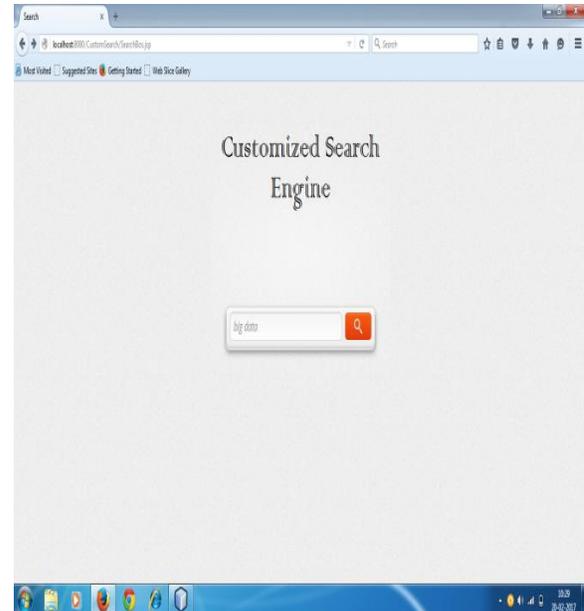

Fig 1.Search Page

Fig 2. Before Re-Ranking

Fig 3.After Re-Ranking

VI CONCLUSION

In conclusion, we can assure that this paper proposes a new method to offer a personalized Web Page Recommendation System by using stemming and Semantic Enhancements, which are called as Knowledge Representation Models for Recommendation Purpose. First model is for Domain specific Knowledge, second model is for personalization. This system is a conceptual prediction model which is used to integrate the domain knowledge and web usage. It is used to form a weighted semantic network of frequently viewed terms or pages. This system is proposed to predict next Web Page recommended by the users through gaining query knowledge.

VII REFERENCES

[1] H. Dai and B. Mobasher, "Integrating Semantic Knowledge with Web Usage Mining for Personalization", in Web Mining:
Applications and Techniques, A. Scime, Ed. Hershey, PA, SA: IGI Global, 2005, pp. 276 – 306.B
[2] B. Liu "Information Retrieval and Web Search", in Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data, B.Liu, Ed.: Springer-Verlag Berlin Heidelberg, 2011, pp.183-236.
[3]. Liu, B. Mobasher, and O. Nasraoui, "Web Usage Mining, in Web Data Mining: Exploring Hyperlinks, Contents, andUsage Data", Ed.: Springer-Verlag Berlin Heidelberg, 2011,
pp. 527-603.
[4] ThiThanh Sang Nguyen, Hai Yan Lu, Jie Lu, "Web-page Recommendation based on Web Usage and Domain Knowledge",Knowledge and Data Engineering IEEE Transaction on September 2014.
[5] .S. Grimm, A. Abecker, J. Völker, and R. Studer, "Ontologies and the Semantic Web" , in Handbook of Semantic Web echnologies,J. Domingue, D. Fensel, and J. A. Hendler, Eds.: Springer-Verlag Berlin Heidelberg, 2011, pp. 507-580.
[6] Brindha.S and Sabarinathan.P "An effective search on web log from most popular downloaded content",International Journal of Distributed and Parallel Systems (IJDPS) Vol.5, No.1/2/3, May 2014.
[7] Sneha Y. S, G Mahadevan, and Madhura Prakash, "A Personalized Product Based Recommendation System Using Web Usage Mining and Semantic Web", International Journal of Computer Theory and Engineering Vol. 4, No. 2, April 2012.
[8] Brindha.S and Sabarinathan.P "An effective search on web log from most popular downloaded content",International Journal of Distributed and Parallel Systems (IJDPS) Vol.5, No.1/2/3, May 2014.
[9] V. SithaRamulu, Ch. N. Santhosh Kumar, K. Sudheer Reddy,"A Study of Semantic Web Mining: Integrating Domain Knowledge into Web Mining",

International Journal of Soft Computing and Engineering (IJSCE)ISSN: 22312307, Volume-2, Issue-3, July 2012.

[10] C.Ramesh1, Dr. K. V. Chalapati Rao, Dr. A.Goverdhan, "A Semantically Enriched Web Usage Based Recommendation Model", International Journal of Computer Science & Information Technology (IJCSIT) Vol 3, No 5, Oct 2011.

[11] Honghua Dai and Bamshad Mobasher, "Integrating Semantic Knowledge with Web Usage Mining for Personalization", School of Computer Science, Telecommunication, and Information Systems.