Recommendation of Web Pages Based on User Profile and Information

Priyanka R. D.[1]Karthikeyan C.[2]Saravanan R.[3]
[1,2]Assistant Professor[3]Student
[1, 2,3]Department of Computer Science &Engineering
[1,3]Adithya Institute of Technology, Coimbatore, Tamilnadu, [2]United Institute of Technology, Coimbatore, Tamilnadu, India

Abstract**—***We intend to apply a recommendation system which can be personalized that makes use of user profiles and data derived from ontologies. This aims to provide semantic applications with personalized services. Quality of the suggestions can be improved by means of capabilities of semantic analysis and it overcomes the problems of current systems. Domain ontologies are used to enhance the personalization. Here a domain-based inference method is used to represent the user's interests in an accurate and effective manner. Also the content-based filtering method uses the stemmer algorithm which gives a measure of the similarity between an item and a user, and a semantic similarity method is applied for enhancement. An effective recommender system is proposed based on ontology and Web Usage Mining which plays a key role in web personalization. Different preferences were implied by different semantic themes due to the semantic similarity of web documents.*

Keywords—Domain Ontology, Semantic Analysis, Web Usage Mining, Web Personalization, Semantic Similarity

I. INTRODUCTION

In applications, that use web searching, we submit queries to search engines to represent our information needs. But it is not always possible to exactly specify the information needs through queries. Also many queries are ambiguous and they tend to cover wide area of topics and different users like to obtain different information on different aspects for the same query. For example, when the query "the apple" is submitted to a search engine, some users want to search for the company apple and their products, while some others want to learn the benefits of the fruit apple. Hence, it is very essential and necessary to identify the search goals of different users in retrieval of information. Here the user search goals are defined as the information on different views of a query that user groups need to acquire. Information need is a user's particular desire to acquire information to satisfy his/her need. User search goals can be considered as the clusters of information needs for a query. The assumption and study of user search objectives can have large number of benefits in improving user experience and search engine application. Some advantages are: Web search results can be restructured according to user search objective by combining the search results with the same search objective; thus, different search objectives can be easily identified by the users. Keywords representing the user search objectives can be used in query suggestion thus; the users can form their queries precisely with the help of the recommended queries. User search objectives are distributed and they are useful in applications such as web search results re-ranking.

II. RELATED WORK

Due to its effectiveness, many works concerning user search objective have been examined. They can be outlined into three classes: query classification, search result reorganization, and session boundary detection. In query classification, user objectives and intents are attempted to understand by people by describing some classes and performing query classification. To improve feature representation of queries, the classification of queries with pre describing concepts are focused. Finding suitable pre described search objective classes are very hard and unfeasible since users think about changes a lot for various queries. People seek to restructure search outcome. Wang and Zhai examining the clicked URLs straightforwardly from user click-through logs to arrange search outcome and learn motivating aspects of queries. However, the numbers of unusual clicked URLs of a query are tiny, so this method has restrictions. When a query is submitted the search outcome is returned by the search engine. Many noisy search outcomes that are not clicked by any users may be examined as well.

III. SYSTEM IMPLEMENTATION

A. Creating Search History

The browsing history on user's computer can be treated as the data basis for the outline of the user. The users' attention can be easily identified by the focused terms and dimensions of the document set. This module is the greatest tool to understand the users' session and probability. Search engines can have a better illustration of the search framework behind the present query once query groups have been recognized.

B. Query Clustering

Different query clusters can be created based on the users' queries. The individuality effect can also be achieved based on the grouping process by occupying concept-based user profiles. Every user submits the query and it is considered as unique node and each query as user identifier. Grouping of nodes is performed in such a way that the present query and the clicks can be identified and formed as a single group.

C. Query Reformulation

It is mandatory that the present query groups should be closely related and to have the relevance between them. It is assumed that only few queries and clicks are issued by the users within a short period of time. The query relevance, in specific which queries are intended to be issued like information are maintained by the search history. This captures the relationship between queries frequently leading to clicks on similar URLs as shown in Fig. 1. The search logs contains information regarding Query reformulation graph and the query click graph and also how to use them to

determine relevance between queries or query groups within a user's history.

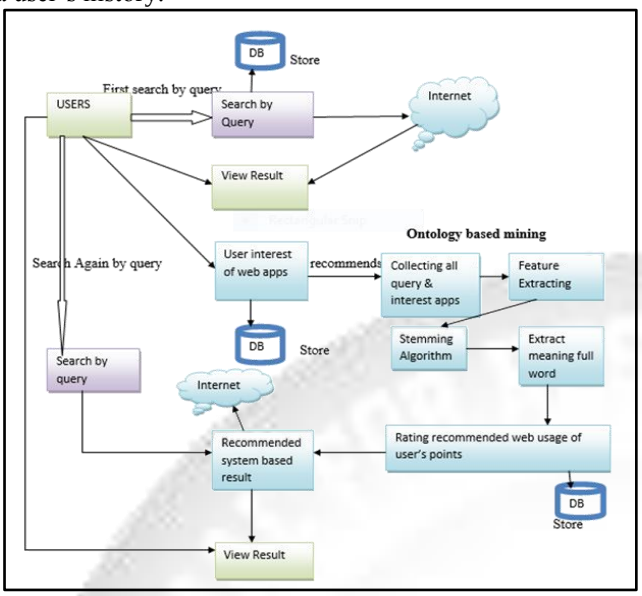

Fig. 1: Query Reformulation

D. History Grouping

All queries in a query group are identified and then it is combined in an iterative fashion (K – means). But it is difficult for two reasons. First,User's existing query groups will be affected since the association of browsing history takes place.Next, the cost is high because of the computation.

E. Pseudo-document

In this paper, feedback session is mapped to pseudocode documents user search objectives.

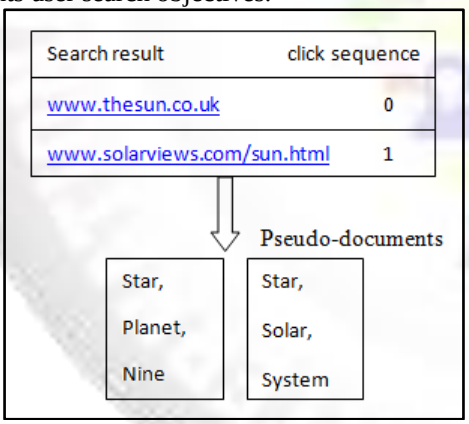

Fig. 2: Pseudo-document

The construction of a Pseudo-document contains two steps: First is representing the URLs in the feedback session. Session that contains URL is represented by its title and snippet. Then, the transformation of all the characters to small cases, stemming and stop words take place. Second, is the formation of pseudo-document based on URL representation.

F. User Search Objective

Pseudo-documents are clustered by an easy and efficient algorithm called K-means clustering. Since the number of user search objectives for a query is not known, we set K to be five unlike values. After grouping all the pseudo-documents, each cluster can be merged and treated as one user search objective.

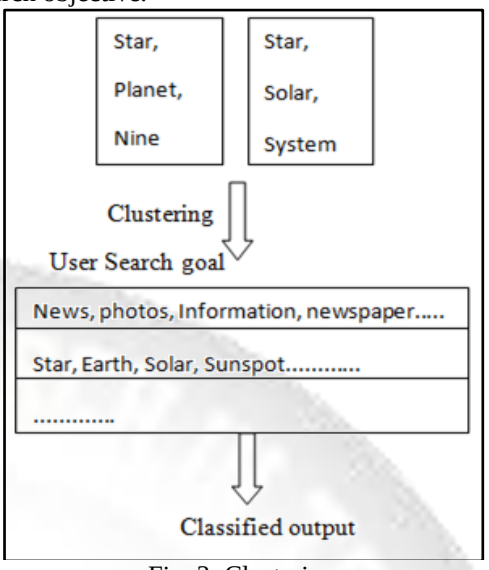

Fig. 3: Clustering

IV. FRAMEWORK

Framework contains two parts: In the first section, the query's feedback sessions are extracted and mapped to pseudo-documents. User search objectives can be identified by grouping of pseudo-documents with reserved words. Since user search objectives are not known, few inputs are given and the best possible value will be considered by the feedback from the second section. In the second section, the search results are reorganized based on the user search objectives inferred from the upper section. Then, we evaluate the performance of reorganizing search results by our proposed evaluation criterion.

A. Feedback Sessions

The session of activity that a user with a unique IP address spends on a Web site during a specified period of time is the feedback session. The number of user sessions on a site is used in measuring the amount of traffic a Web site gets. In this paper, the understanding of user search objectives is focused for a particular query.

The proposed feedback session contains the two: Clicked and unclicked URLs. It is provoked that users need to evaluate and scan all the URLs before the last click. In addition to the clicked URLs, the unclicked URLs should be a part of the user feedbacks before the last click. The left section consists of the results of the query "The sun". The right section consists of the sequence of user click, here "0" denotes "Unclicked". Consider Fig 3. There are three clicked URLs and four unclicked URLs in this example. From top to bottom, the user will examine the URLs, The user considers all the clicked and unclicked URLs and it is evaluated by the user. Each feedback session clearly reveals that what a user's interest what he/she does not want.

B. Map Feedback Sessions to Pseudo-documents

Since there is a variation in the click-throughs and queries, it is inappropriate to use feedback sessions for deriving the user search objectives.

Some illustration method is required to illustrate feedback sessions in a more proficient and consistent way.

Consider an example,When the user submits the query "the sun" and unclicked is mentioned as "0" in the click sequence. The binary vector [0101101] is representing the feedback session, where "clicked" is symbolized as "1" and "unclicked" is symbolized as "0". Moreover, different feedback sessions have different numbers of URLs, the binary vectors of different feedback sessions may have different dimensions.

Therefore, it is not suitable to use methods such as the binary vectors and new methods are required to symbolize feedback sessions.

C. Restructuring web search results

Search engines returns expected results, it is important to arrange them in a specified order. Reorganizing web search results is an application of understanding user search objectives. We need to describe the reformation of web search results. The derived user search objectives are symbolized by the vectors in and the feature depiction of each URL in the search results can be figured. Each URL is classified into a cluster centered by the derived search objective. In this paper, the classification by selecting the least space between the URL vector and user search objective vectors is performed.

V. CONCLUSION

In this paper, we proposed an efficient approach to understand user search objectives by grouping its feedback sessions characterized by pseudo-documents. Feedback sessions are introduced and to be analyzed for deriving the user search objectives relatively than search results or clicked URLs. Considering the URLs of two kinds before the last click as user hidden feedbacks and it is considered as valuable remark to construct feedback sessions. The URLs can be elevating with pseudo-documents including the titles and snippets. These pseudo-documents help the user to search objectives that can be revealed and illustrated with some keywords. In the end a new condition Stemming is originated to appraise the performance of user search goal assumption. The proposed method's effectiveness is demonstrated by results on user click-through logs from a commercial search engine. For each query, the feedback session decides the running time. In reality, our approach can determine user search objectives for a number of queries offline. Then, when queries submitted by the users, the search engine can return the results that are considered into different groups according to user search goals online.

REFERENCES

[1] H. Dai and B. Mobasher, Integrating Semantic Knowledge with Web Usage Mining forPersonalization, in Web Mining:Applications and Techniques, A. Scime, Ed. Hershey, PA, USA:IGI Global, 2005, pp. 276 - 306.

[2] B. Liu, B. Mobasher, and O. Nasraoui, Web UsageMining, in Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data, B. Liu, Ed.: Springer-Verlag Berlin Heidelberg, 2011, pp.527-603.

[3] B. Mobasher, Data Mining for Web Personalization, inThe Adaptive Web. vol. 4321, P. Brusilovsky, A. Kobsa, and W. Nejdl, Eds.: Springer-Verlag Berlin, Heidelberg, 2007, pp. 90-135.

[4] S. A. Rios and J. D. Velasquez, Semantic Web UsageMining by a Concept-Based Approach for Off-line Web SiteEnhancements, in Web Intelligence and Intelligent AgentTechnology, 2008. WI-IAT '08. IEEE/WIC/ACMInternational Conference on, 2008, pp. 234- 241.

[5] S. Salin and P. Senkul, Using Semantic Information forWeb Usage Mining based Recommendation, in 24th International Symposium on Computer and Information Sciences, 2009, pp. 236-241.

[6] B. Zhou, S. C. Hui, and A. C. M. Fong, CS-Mine: AnEfficient WAP-Tree Mining for Web Access Patterns," inAdvanced WebTechnologies and Applications. vol. 3007: Springer Berlin /Heidelberg, 2004, pp. 523-532.

[7] B.Zhou, Intelligent Web Usage Mining, NanyangTechnological University 2004.