Providing The Best Search Logs to The User Using Semantic Ranking Algorithm

V. Hanumath Sushma* and R. Subash**

ABSTRACT

Searching the weblogs has been studied at query level or session level. The session is a collection of tasks made by the user while searching the query .By doing this, it cannot easily find out whether the task in a session belongs to a particular thing or not and also about the searching behavior of the user. To know the searching behavior it uses tasktrial. For predicting the user satisfaction, it uses the ranking algorithm,and it combines with semantic search. For combining the related queries, it uses the query flow graph to group the related query. SVM and MDP are using to cluster the queries.SVM uses the classifier as a query.

Keywords: Support Vector Machine, Markov Decision Process.

1. INTRODUCTION

People uses different search engines for the information which they needed and to provide the easy assess to them make them combine the queries and URLs.ing search as a key enabling technology[7].By collect two, three and more than one query, it is not able to determine the user search behavior and also the links which has been searched by the user[12] in the search history. The queries can be ranked based on the result list made by the user,and it can retrieve the related query which may be useful to the user. For knowing the user browsing behavior, frequent patterns are mined from result list[17].For defining the World Wide Web,it describes the web standards which are formal and nonprosprietary standards.

In recent years, the term which is frequently searched by the user made of endorsing a set of standardized best practices for building web pages. Search usability has defined as the effectiveness, efficiency, and satisfaction with which the authenticated users can complete user defined tasks in a particular environment. For improving the users online experience, three basic principles had been used here,i.e., structural firmness, functional convenience, and presentational delight. Structural inflexibity relates to the performance and security. The availability of convenient characteristics refers to the practical accessibility, such as a site's ease of use and ease of navigation that help user interaction with the interface. To stimulate the user senses, the presentational delight refers to the web characteristics. Usability engineering provides methods for measuring search usability and to address the usability issues.

Heuristic evaluation by experts and user-centered testing are typically used to identify the point and to ensure satisfactory usability. However, significant challenges exist, including 1) problem identification due to false alarms which was typical in the expert evaluation,2) knowing the difference between the testing environment and the user searching environment and 3) the maintenance cycles and the prolonged evolution which is typical for web applications could be increased. In general, the web brought a new paradigm of interaction, reducing the gap between the human's cognitive conception of the task and its computational representation.

*   Computer Science and Engineering SRM University Chennai, Tamilnadu, *Email: sushma.vutukuri123@gmail.com*

**   Assistant Professor (O.G) Computer Science and Engineering SRM University Chennai, Tamilnadu, *Email: subash.r@ktr.srmuniv.ac.in*

The semantic web, the main idea is to have semantically related representations of knowledge rather than static stand-alone information all over the screen.It supports explicit expressiveness in representing and relating domain information by using semantic information comprising ontologies and resource description framework (RDF) data referring to them. The model-based user interface design (MUID) paradigm is an attempt to carry through such a trade-off. The implicit idea behind MUID is to separate the conceptual level of a user interface, which leads, consequently, to the explicit specification of different aspects of the interface itself, such as domain knowledge, presentation, dialog, and behavior[8]. Click stream analysis is a new idea for measuring the web page importance. A webspider can access the serverlog files.The usage information about the user has recorded in web logs. Analyzing web log files to extract useful patterns is called the web usage mining.Web Usage mining is finding of user access patterns from one or more Web servers. A server log is an important source for performing web usage mining because it records the browsing behavior of site visitor. Clickstream is an ordered sequence of the web page viewed by a user that is a session of visited web pages; pages are presented one by one in a row one at a time. Complete click sequence contains the information of all users and analyzes how many URLs the user followed by the search result. Clickstream analysis is used for the extraction of the information from the log file. Log file generated by web server contains massive amount of web page data that is valuable for understanding the movement of web site visitor. Time spent on a particular web page is an evaluation of the user attention to that web page. The duration of each visited page in a session as metric can help in detection of crawler[19].

2. PRELIMINARIES

2.1. World Wide Web

The World Wide Web (WWW) is distributed system which consists of millions of clients and servers to access the linked documents whereas, Servers maintain collections of document and clients provide users an easy-to-use interface for presenting and accessing those documents.

The documents that it referred called a Uniform Resource Locator (URL). A URL,which is comparaed to an IOR in CORB and that of an contact address in Globe. It specifies where a document is located and by considering the DNS name of its associated server along with a file name by which the server can look up the document in its local file system. Furthermore, a URL specifies the application-level protocol for transferring the document across the network.

2.2. Support Vector Machine

The main purpose of a support vector machine (SVM) is to supervise the learning methods that analyze data and recognize patterns, used for classification and regression analysis. More formally, a support vector machine constructs a hyper plane or a set of hyper planes in a high- or infinite- dimensional space, which can be used for classification, regression, or other tasks. Intuitively, a good separation is achieved by the hyper plane that has the largest distance to the nearest training data points of any class (so-called functional margin), since in general the larger the margin the lower the generalization error.

2.3. RDF

RDF is a Resource Description Framework,and it is used to describe the resources on the web.RDF is designed to be read and understood by computers and not for being displayed to people and that was written in XML.RDF is one of the semantic web activity. That means it has exact meaning,and that can be understood and processed by computers which can integrate information from the web.

RDF is used to identify the things which are using URIS. And it describes three things as resource, property and property value. A Resource is a thing which has URI, such as "http://www.w3schools.com/

rdf".A Property is a Resource which has a name that may be "author". A Property value is a property which has a value such as "Jan Eglin Refsnes" or "http://www.w3schools.com".

2.4. MBUID

MBUID stand for Model -based Design of User Interfaces. The main purpose is to reuse designs and use different models for different aspects. The different models who are using are task model, presentational model, dialog model and context model. MBUID uses multiples levels of abstraction that is tasks, platform/ modality independent toolkit independent and code. It will check the availability of tools as limited in scope, research tools, not always publicly available.

2.5. MDP

MDP stands for Markov decision processes,and it provides a mathematical function for modeling decision-making in situations where outcomes are in two ways as partly random and partly under the control of a decision maker. And these are useful for studying a wide range of optimization problems which could be solved via dynamic programming.
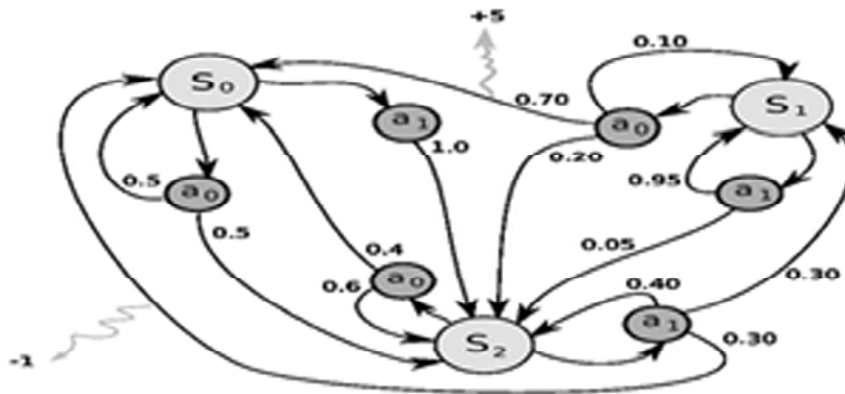

Figure 1: working procedure of Markov Decision Process

2.6. Semantic Ranking Algorithm

Semantic search is used to improve search accuracy by taking the searchers history and understanding the searcher's intent and the circumstantial meaning of terms as they appear in the searchable history of the users, which was on the Web or within a closed system, to provide the most relevant results. Semantic search systems take the various points including for comparing and it includes the text of search, location, intent, variation of words, synonyms, generalized and specialized queries, concept matching and natural language queries to provide relevant search results. Most of the web search engines like Google and Bing provides some elements of semantic search. Based on that, we are ranking the searching history of the user and the search engine provides the accurate results to the user.

3.   METHODOLOGIES

3.1. Task Definition and Extraction

Weblogs contain multiple users and their multiple results, here system collect the entire information about the user such as his/her behavior and web pages he/she searched. Task definition is the concept will allocate the task in term of user behavioural model. It contains entire session of the user searched on the web. After collecting the task,the system will extract the information based on the level of User necessary. The following figure explains the providing the search result to the client which will be collected from the server. After that, if the users are not satisfied with the webpage the user clicks another web link which could be provided

Figure 2: How the clicking information stored in to the database

in the web page and that the web link can be directly stored into the database. For that, the Java Servlets Program is using.

3.2. Clustering Queries

When a user submits a query on the client the query forwarded to the server. It obtains the search results from the back-end search engine .The content and location concepts are extracted from the search results and organized and capture the relationships between the concepts. The search grouped in a cluster. Finally, the extracted information for the personalization of future queries is returned to the client. It stores the queries based on the user. Under the user, it stores the queries based on their searching. The Semantic web is a technology which is used for structuring and finding the relationship between the queries.And ranking is given based on their searching.

3.3. User Interest Profiling

Task trail uses "concepts" to model the interests and preferences of a user. The concepts are further classified into two different types,and they are location and content concepts. The location concepts for navigating and the content concepts is to know the context verification. Here it uses the context verification. The system indicates the possible concept space arising from a user's queries, which can be done through the click through data for future preference adaptation. Based on that, the clicks which had been made by the user that can be stored into the database, from that it pedicts the user preference and interest by using location and content concepts which had been mentioned above.

3.4. Suggesting Relevant Information

Finally,the system will provide the information based MDP. Markov decision process is a process gives content for making finaldecision. MDP always predict the things based on state, action and transition. I will suggest the information based on decision which generated by MDP. Also user profiles are maintaining by a privacy secured mechanism, which involves cryptosystem for hiding user profile in front of other users. The cryptosystem techniques are used here to protect the user profiles as well as the searching content made by the user at the time time of the searching the query in it to get the relevant information. The Markov Decision Process will store the queries and the links in the form of state and actions and finally transiting the information based on state and actions. States can be taken as queries searched by the user where as links between the web pages can be taken as the actions. It that is nothing but the transition of the information and it can be get by using the search engines of ours.

4.   SYSTEM ARCHITECTURE

The User will enter their query into the Task trail. Task trail to understand the user search behavior based on the user search history. And the query will be analyzed based on the user's interest &classified into the new query or an executed query. It will be re-ranked based on the Reduced Support Vector machine. it will collect the relevant information from the online server. Scent Trail to combine searching and browsing activities into a single interface, and they found it can help users in finding information faster than by only
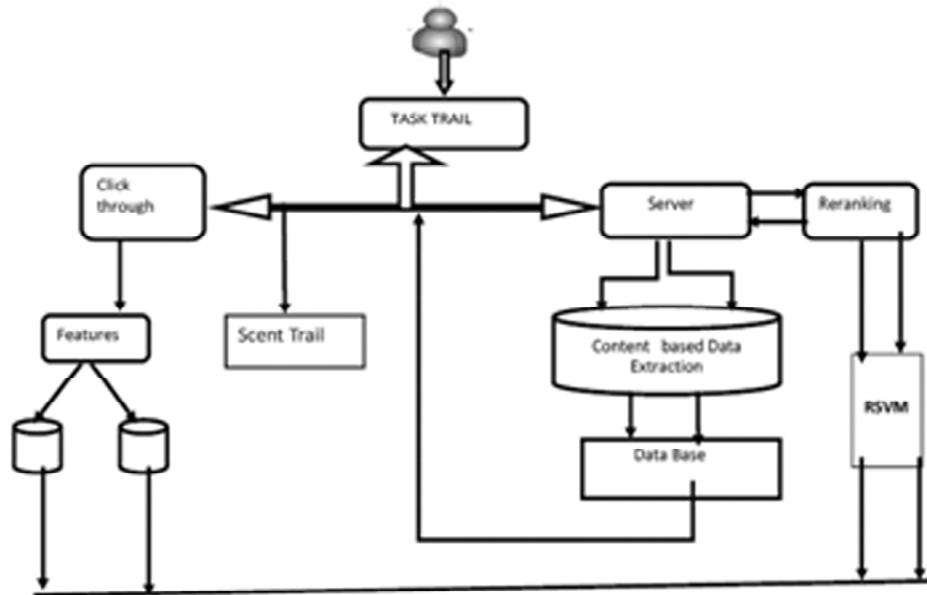
Figure 3: System architecture of how the task trial works.

searching in the browser .these search history will be stored as a user interest profile in tothe database. The new query provides the information from the backend search engine.

5. RELATED WORK

Web search logs document user activities on search engines, user activities on search engines. Search trails document the footprints left by users in their search process. In the information of studying task trials, the problem of efficient query processing in extensible geographic search engines, present a new approach to mining user's preferences on the search results from click through data, Ranking Support Vector Machine in a Co-training Framework,preferences in the form of concepts by mining search results.

5.1. Problem of efficient query processing

Query processing is a major barrier in standard web search engines, and the main reason for the thousands of machines used by the major engines. Geological search engine query processing is different in that it requires a combination of text and spatial data processing techniques. They propose several algorithms for efficient query processing in geographic search engines, combine them into an existing web search query processor, and assess them on large sets of real data and query traces.

5.2. Mining user's preferences on the search result from click-through data

It presents a new approach to mining user's option on the search results from click through data and using the exposed options to adapt the search engine's ranking function for improving search quality. It develops a new option mining technique called Spy NB, which is based on the practical guess that the search results clicked on by the user,reject the user's options, but it does not draw any conclusions about the results that the user did not click on.

5.3. Ranking Support Vector Machine in a Co-training Framework

It proposes a new algorithm, Ranking Support Vector Machine in a Co-training Frame work. Substantially, the RS-CF algorithm takes the click through data containing the elements in the search result that have been clicked on by a user as an input, and generates flexible ranker as an output. By considering the click

through data, RS-CF rest categorizes the data as the labeled data set, which contains the items that have been scanned already, and the undefined data set, which contains the items that have not yet been scanned. The defined data is then augmented with undefined data to obtain larger data set for training the musty.

5.4. Mining search results

Due to the important role location information shows in mobile search, we separate concepts into content concepts and location concepts and organize them into ideas to create an ontology-based multi-facet periled precisely capture the user's content and location interests and hence develop the search accuracy. Moreover, recognizing the fact that different users and queries may have different signification on content and location information, we introduce the notion of content and location decline to measure the amount of content and location information associated with a query, and click content and location decline to part how much the user is interested in the content and location information in the result.

6.   CONCLUSION

In this paper we introduced the Task trail to understand the user search behavior based on the user search history. Task trail represents all user activities within that particular task, such as query reformulations, URL clicks proposed Scent Trail to combine searching and browsing activities into a single interface, and they found it can help users in finding information faster than by only searching in the browser. The user search history URL click will be submitted to the database to analysis the user search history generated as User Interest Profile. The user's search privacy can be maintained using the cryptography application such as key generation &personalized user profile.

REFERENCES

[1]   Rui Wu, Wansheng Tang and Ruiqing Zhao" An Efficient Algorithm for Fuzzy Web-Mining".

[2]   Peng Huang, Dehua Chen, Jiajin Le, "An Improved Referrer-Based Session Identification Algorithm Using Map Reduce".

[3]   Anupma Surya, Dilip Kumar Sharma, "AN APPROACH FOR WEBPAGE ORDERING USING USER SESSION".

[4]   WANG Wenyu, GUO Gelin, JIANG Bin, SHI Yu-fang, "Automatic Classification for WDMS with Isomap and SVM".

[5]   Ali Idris and Alain Abran , "A Fuzzy Logic Based Set of Measures for Software Project Similarity: Validationand Possible Improvements".

[6]   Traugot Koch, Anders Ardö, KoraljkaGolub,"Browsing and Searching Behavior in the Renardus Web Service: A Study Based on Log Analysis".