Ontology Based Web Page Recommendation System

A.Sivasangari, S.Poonguzhali, Immanuel Rajkumar

Abstract— The emerging web page development requires semantic applications with customized administrations. The proposed methodology presents a customized suggestion framework, which makes utilization of item representations and also client profiles created based on ontology. The domain ontology helps the recommender to improve the personalization: from one perspective, client's interests are displayed in an increasingly powerful and precise route by applying an area based derivative technique; on the other side, the stemmer algorithm derived content- based filtering approach, gives an evaluation of resemblance among a thing and a client, upgraded by applying a semantic likeliness strategy. Recommender frameworks and web personalize were assumed by Web usage mining as a critical job. The proposed strategy is s successful framework dependent on ontology and web usage mining. Extricating highlights from web reports and building applicable ideas is the initial step of the methodology. At that point manufacture metaphysics for the site exploit the ideas and huge terms separated from reports. As per the semantic similitude of web archives to bunch them into various semantic topics, the distinctive subjects suggest diverse inclinations. The proposed methodology incorporates semantic information into Web Usage Mining and personalization process

Keywords— Ontology, Filtering method, Cluster

1. INTRODUCTION

Generally questions are submitted to the search engines to speak to the data needs of client in every web search application. Be that as it may, now and then questions may not actually speak to clients' particular data needs since numerous uncertain inquiries may cover a wide point and distinctive clients might need to get data on various angles when they present the equivalent query [1] B. Liu et al streams out the use of web mining and [2] B. Mobasher brings up the significance of data mining for web personalization. For instance, consider when the inquiry "Apple" is submitted to a web index, a few clients need to find the landing page of a United Kingdom paper, while some others need to become familiar with the common learning of the Apple . Subsequently, it is fundamental and impending to catch diverse client seek objectives in data recovery. We characterize client seek objectives as the data on various parts of a question that client bunches need to acquire . Data need is a client's specific requirement to get information to fulfill his/her need. Client seek objectives can be considered as the list of data requirements for a question. The derivation and investigation of client seek objective sweep have a great deal of favorable circumstances in

improving web crawler importance and client experience. A few focal points are abridged as pursues. To begin with, we can rebuild web list items as indicated by client look objectives by gathering the list items with a similar inquiry objective; in this way, clients with various pursuit objectives can without much of a stretch find what they need. Second, client look objectives spoken to by certain watchwords can be used in question proposal subsequently, the recommended inquiries can assist clients with forming. Third, the circulations of client look objectives can likewise be valuable in applications, for example, re-ranking web indexed lists that contain diverse client seek objectives.

2. RELATED WORK

Mehrbakhsh et al [3] proposed a recommender system based on collaborative filtering using ontology. Collaborative filtering is used for solving the two issues in the recommendation system such as sparsity and scalability. Ontology is used for improving the accuracy of the system. Dimension reduction technique is used for reducing the unwanted features and enhance the scalability of the system.

Chakkrit et al [4] proposed a context aware tourism recommender system for supporting decision making in tourism. Improved naive Bayes algorithm is used for tourism web classification. Threshold limt value is fixed for all types of category. Based on the temporal ontology, it will verify the date and retrieved information related to tourism or not. System recommendation is based on time and season. Limitation of this work is users's current location parameter that is not considered as a parameter for recommendation system.

Existing web service discovery and recommendation system based on UDDI registers or keyword dominant web service engine [5]. A good recommendation system should recommend the web services based on the users's interest. In this work, collaborative filtering and content based recommendation techniques used for recommendation. Collaborative filtering considers the other's user rating for predicting the rating score for the services. Pearson correlation is used for calculating the similarities between the users. Unified approach is used to reflect the preferences of users.

Neha et al [6] searched proposed a methodology for improving the ecommerce web site ranking by applying ranking algorithm and back propagation model is used to to perform the ranking. Web page retrieval is based on dictionary module that perform the data preprocessing like data cleaning and removing stem words. Web content

------
Revised Manuscript Received on April 12, 2019.
   A.Sivasangari, Sathyabama Institute of Science and Technolgy, Chennai, T.N, India. (E-mail: sivasangarikavya@gmail.com)
   S.Poonguzhali, Sathyabama Institute of Science and Technolgy, Chennai, T.N, India (E-mail: b.poonguzhali@hotmail.com)
   Immanuel Rajkumar, Sathyabama Institute of Science and Technolgy, Chennai, T.N, India (E-mail: imman047@gmail.com)

mining is applied to check the relevant web page documents and specify the priority of the document. Wrong interpretation of search query is avoided by applying ontology method. Searched product is displayed on the priority of the e commerce web site.

Salam et al [7] proposed a framework for semantic recommender system which assists users to find the relevant objects based on their interest. Ontology model is developed and used to classify the key words. Indexing process is repeated for improving the quality of the index of the repository. It is based on semantic relations and query key words.

3. PROPOSED SYSTEM

The proposed system provides a personalized-recommendation system, which is employing the depiction of items and user-profiles on the basis of ontology to facilitate semantic applications with personalized services of mining in semantic web applications. The following fig 1 shows the architecture.

The semantics accomplished by utilizing two diverse methods additionally feature the web utilization mining applications. A space based technique makes deductions about client's interests and a scientific categorization based comparability strategy is utilized to refine the thing client coordinating calculation, improving generally speaking outcomes. The recommender proposed is space autonomous, is actualized as a Web administration, and utilizations both express and verifiable criticism accumulation techniques to acquire data on client's interests.
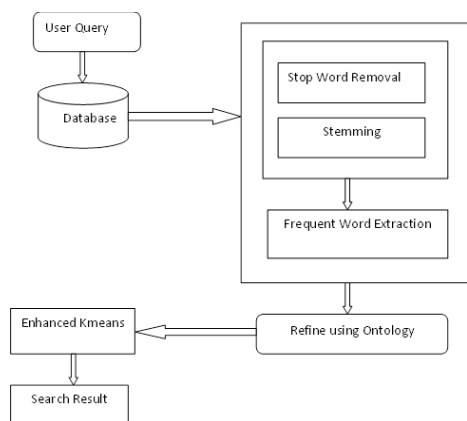
Fig. 1. Proposed Architecture

3.1 Creating Search History

The proposed work is based on web usage mining with ontology. Features are extracted from the web documents. Relevant concepts are constructed from these features. This work includes the web page recommendation on functional basis such that the search results. Result of searched query is stored in the database. The searched done in the ontology order such that the user interest result is stored in the data base by using the stemming algorithm and in which the further the proposes the hybrid clustering algorithm in which the related result is clustered. The clustered data is done to the stored the data base. In this work, the search engine understand the user's needs by their searching experience. Query groups are identified for search engine.

3.2 Query clustering

User queries can be classified in to different clusters by using divisive clustering method with concept based user profiles. Query submitted by the users is considered as objects. All the objects in the query are put in one group and in succession. It partition the objects in different group at every move. The procedure is repeated for all the objects until it is classified into any one of the clusters. This method increase the productivity of the searching by supporting multiple partitions. Search result pages can be grouped according to the categories. Queries submitted by the users are very short and ambiguous meaning. Query classification is very difficult than the traditional document classification. The following table 1 shows the query topic classification examples.

Table I Examples of Query clustering

| Query	Categories
Raspberry	Computers/Hardware/Foods/Cooking
Orange	Fruit/Color/Tele communication
Aloe Vera	Cooking/Medicine/Health/ Beauty
Song	Humanities/Living/Enjoyment
Olympic 2019	Sports/Schedule/Tickets/Entertainment

3.3 Query reformulation

It is important to verify the relevance between the query clusters. Query reformulation is considered an indication of unsatisfied with the previous query. Query reformulation graph and query click graph can be used to determine the relevance between queries in the history of users. An undirected bipartite graph is used for query reformulation.

$$E_{ql} = (q_i, l_j)$$

The Edge E_{ql} exists only if the URL link l_j is clicked by user for the query q_i

URL query weight is calculated by total number of times query is used and total number of times the document is clicked by the users. Weight of the edge is calculated by counting common URLs. Relevant queries can be obtained by using queries clicked on the same set of documents by the users in the search logs. Query normalization is performed by replacing all letters with their corresponding lower case, removing trailing spaces and replace whitespace characters with a single space. Fig 2 shows URL query map representation.
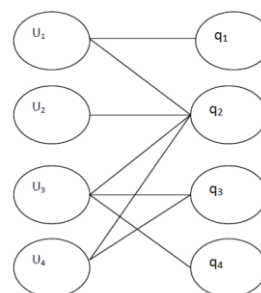
Fig2 URL Query Map

3.4 Pseudo document

There is a need to do mapping of feedback session with pseudo documents. The procedure for pseudo documents preparation consists of two steps. First one represent the URLS in the feedback session. Textual process is implemented by transforming letters to lowercase, remove stop words. Second step combine both clicked and unclicked URLs in the feedback session. Search result described in fig3.
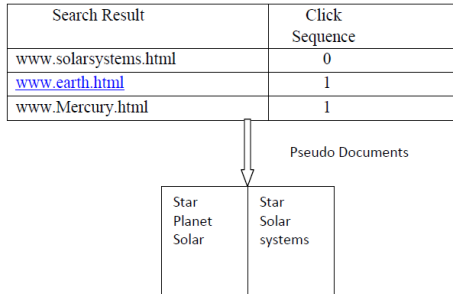
Search Result	Click Sequence
www.solarsystems.html	0
www.earth.html	1
www.Mercury.html	1

Pseudo Documents

Star Planet Solar	Star Solar systems

Fig 3 Search Result

Input for the algorithm: Without stop word key word set ={k1,k2,…kn}

Step 1: Select ending characters of word

Step 2 : Examine the final letter

Step 3 :Verify the applicability of stemming word

Step 4: Ending characters are not matched by the stemming rule that can not be replaced by language.

Step 5: Apply the rule

Step 6 : End character is deleted and replaced by appropriate word

Step 7 : Return the stem word.

Classification result is displayed in fig 4. Pseudo documents are clustered by using K means clustering algorithm. The k means algorithm is summarized below:

1. Pseudo documents set S={s1,s2,,,sn} where every si considered as attribute
2. Define K number of clusters
3. Initialize the centroid matrix c_{ij} and sort the pseudo documents matrix
4. Calculating the centroid of the clusters by split the peudo documents matrix into k parts
5. Each cluster can be considered as one user search goal.
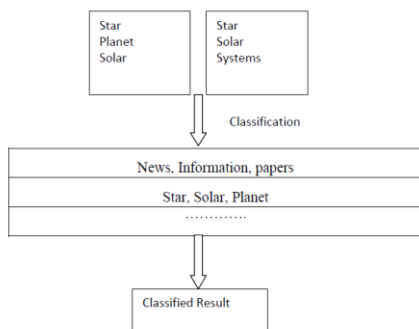6. Classify the documents into different themes.

Fig. 4 Classification Result

4. PERFORMANCE ANALYSIS

The proposed input session comprises of both clicked and unclicked URLs and finishes with the last URL that was clicked in a solitary session. It is inspired that before the last snap, every one of the URLs have been examined and assessed by clients. Along these lines, other than the clicked URLs, the unclicked ones preceding the last snap ought to be a piece of the client inputs. It ought to be noticed that the unclicked URLs after the last clicked URL ought not be incorporated into the criticism sessions since it isn't sure whether they were examined or not. Every criticism session can tell what a client requires and what he/she couldn't care less about. Additionally, there are a lot of various input sessions in client navigate logs. In this way, for deducing client look objectives, it is more effective to break down the input sessions than to investigate the indexed lists or clicked URLs straightforwardly.

There can be numerous sorts of highlight portrayals of criticism sessions. For instance, list items are the URLs returned by the web crawler when the inquiry "the sun" is submitted, and "0" speaks to "unclicked" in the snap arrangement. The parallel vector [0110001] can be utilized to speak to the criticism session, where "1" speaks to clicked and "0" speaks to "unclicked." However, since various input sessions have distinctive quantities of URLs the twofold vectors of various criticism sessions may have diverse measurements. Also, twofold vector portrayal isn't sufficiently instructive to tell the substance of client seek objectives.

The performance of the result is evaluated based on the following three measures : (i) precision, (ii) applicability and (iii) hit ratio.

$$Precision = CR/(CR + IR) \qquad (1)$$

$$Applicability = (CR + IR)/[N] \qquad (2)$$

$$Hit\ Ratio = Precision * Applicability \qquad (3)$$

Where , CR- Number of correct recommendations
IR- Number of incorrect recommendations
N- Total number of given requests

Table II Performance of stemming algorithm with Apriori algorithm

Algorithm	No of Candidate Region	Precision	applicability	Hit Ratio
Apriori Algorithm	20	70.52	81.2	83.4
Stemming Algorithm	12	85.53	92.5	94.5

Fig 5 demonstrates the time utilization of stemming calculation and Apriori Algorithm for different edge estimate. X pivot demonstrates edge size and Y hub indicates time in ms. The stemming calculation takes less time when limit gets expanded.
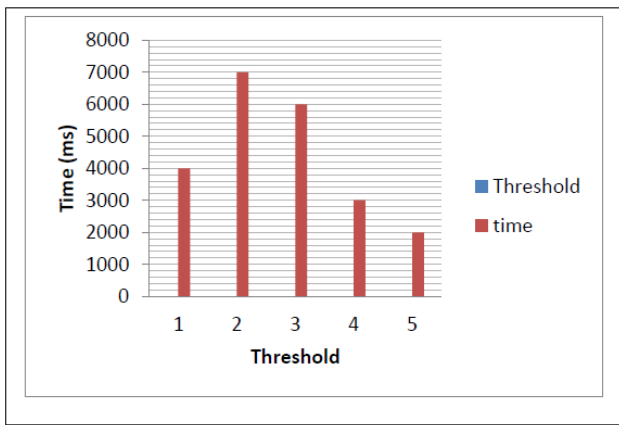
Fig 5 Threshold Analysis

Fig 5 shows the average CTR value for top 5 recommendations based on WEBPMI, WEBJACCARD, WEBDICE, WEBOVERLAP. Stemming algorithm, clicked-RSR and RSR values are shown.
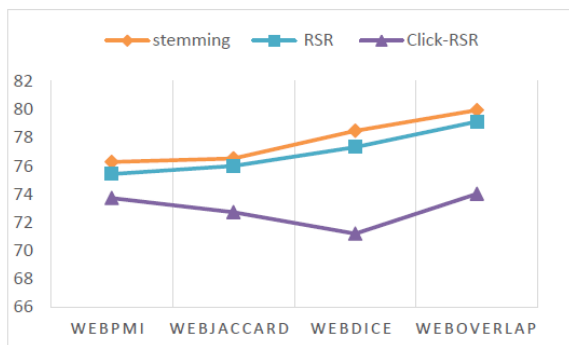

Fig 6 Average CTR value for top-5 recommendations
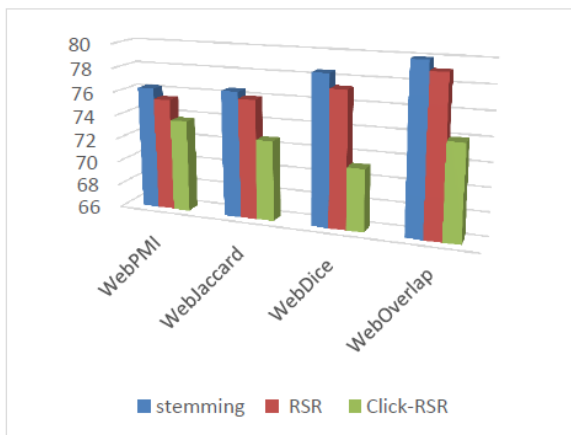

Fig 7 Comparison of average CTR value

Fig 7 shows the comparison chart of stemming algorithm with other two RSR methods.

It is clearly visible that the performance values are quite appreciable higher in stemming algorithm

5. CONCLUSION

In this paper, a novel methodology has been proposed to derive client look objectives for a question by grouping its criticism sessions spoken to by pseudo-archives. In the first place, we acquaint criticism sessions with be broke down to deduce client look objectives as opposed to indexed lists or clicked URLs. Both the clicked URLs and the unclicked ones preceding the last snap are considered as client verifiable criticisms and considered to build input sessions. In this manner, input sessions can reflect client data needs more effectively. Second, we map input sessions to pseudo archives to rough objective messages in client minds. The pseudo-records can advance the URLs with extra literary substance including the titles and pieces. In view of these pseudo-archives, client seek objectives would then be able to be found and delineated with a few.

REFERENCES

1. B. Liu, B. Mobasher, and O. Nasraoui, "Web Usage Mining," in Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data, B. Liu, Ed.: Springer-Verlag Berlin Heidelberg, 2011, pp. 527-603.
2. B. Mobasher, "Data Mining for Web Personalization," in The Adaptive Web. vol. 4321, P. Brusilovsky, A. Kobsa, and W. Nejdl, Eds.: Springer-Verlag Berlin, Heidelberg, 2007, pp. 90-135.
3. Mehrbakhsh Nilashi, Othman Ibrahima , Karamollah Bagherifard," A recommender system based on collaborative filtering using ontology and dimensionality reduction techniques", Elsevier, Expert Systems With Applications 92, 2018, pp.507–520.
4. Chakkrit Snae Namahoot1, Michael Brückner , and Naruepon Panawong," Context-Aware Tourism Recommender System Using Temporal Ontology and Naïve Bayes" , Springer,Advances in Intelligent Systems and Computing 361, pp.183-194.
5. Lina Yao, Quan Z. Sheng, Member, Anne. H. H. Ngu, Jian Yu, and Aviv Segev," Unified Collaborative and Content-Based Web Service Recommendation", IEEE Transactions on Services Computing , Volume: 8 , Issue: 3 , 2015,pp. 453-466.
6. Neha varma,Dheeraj,Monica malhothra and Jatinder," Ecommerce web site ranking using web mining and neural computing",45, 2015, pp.42-51.
7. Salam and Qusai," A framework for semantic recommender system using E learning", Journal of software, Vol 10,pp.317-330.