Journal of Theoretical and Applied Information Technology
20th September 2014. Vol. 67 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

A PREDICTIVE USERS PREFERENCE IDENTIFICATION TECHNIQUE TO IMPROVE WEB USERS QUERY RELATIONAL PROCESSING

[1]D. MADHUSUBRAM, [2]DR. S.P. SHANTHARAJAH
[1] Research Scholar, BharathiarUniversity, Coimbatore, India;
[2] Professor, Sona College of Technology, Salem, India
E-mail: [1] dms_ram@yahoo.com , [2]spshantharaj@gmail.com

ABSTRACT

One of the major challenges in the web is accessing relevant information according to the needs of the user. With each user having different information needs in relation to his/her query, the search results should be personalized according to the information needs of the users. Personalized ontology model for knowledge representation and reasoning over user profiles fails to match the local instance repository queries (i.e.,) users with global information base (i.e.,) web database. Although the application of ontology model has been underway for many years and many algorithms related to ontology have been developed, it is not applicable to the majority of the existing web documents. Query planning for Weighted Additive Aggregation Queries (WAAQ) obtained optimal set of sub queries with incoherency bounds and least number of refresh messages were sent from aggregators to the client. But WAAQ is not effective in developing the cost model for complex queries. To overcome the issues related to complex queries, Predictive User Preference Identification (PUPI) technique is developed based on the relational users and stored queries. PUPI technique searches the results according to each user's need based on their relevant information with little effort from the side of the user, followed by it the effectiveness with complex queries are verified. PUPI then extends to Condition Redefine Query which redefines the queries according to the user relational profile and stored query context. In order to learn user's preference, PUPI make use of both semantic and lexical information. PUPI take into account the web queries for the current user task and engender a new query language model for redefining based on the user query model and the considering the user profile. PUPI technique performed experiment on Weka tool using the MSNBC.com Anonymous Web Data. By analyzing the results, PUPI technique is better than using the state-of-art methods. Experiments were conducted on the factors such as execution time, micro averaged accuracy, query user relationship ratio, query frequency similarity and true positive rate to reveal the efficiency of the technique, query frequency similarity and true positive rate.

Keywords: *Ontology Model, Predictive User Preference Identification, Semantic Information, Condition Redefine Query, User Relational Profile, Stored Complex Queries*

1. INTRODUCTION

World Wide Web (WWW) satisfies the needs of the individual users with the information resources on the WWW persist to develop and web search engines help the users to find useful information on the WWW. Conversely, when the similar query is submitted by different users, most search engines come back with the same results irrespective of the user who submits the query. In functionality, each user has different information needs according to his/her query. Moreover, a JAVA programmer may require information about the Java Island that may not be found in his preferences. One difficulty of automatic personalization techniques is that they cannot be applied in commonly. So, not the entire user interests are relevant all of the time, usually only a subset is active for a given situation, and the rest cannot be considered as relevant preferences.

The Web enabled people with varied goals and characteristics access an ever-growing amount of information. From the side of the web designer and web administrator, they use the conceptual design of website as illustrated in [3] by measuring the link connections on the website. Consequently, web log files are preprocessed and then path investigation technique is used to inspect the URL information concerning admission to electronic sources. The appearance of hand-held electronic devices, such as palmtops and cellular phones, has amplified the

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

potential for information access from anywhere, anytime. In context, finding the right information remains a complex problem with the growing rate at which new information becomes obtainable and the heterogeneity of people searching information.

Alternately, the Internet offers almost unlimited access to information of all kinds. As the volume of the heterogeneous resources on the web increases and the data becomes more mixed, enormous answer results are issued to user queries. Thus, large amounts of information are generated in which it is frequently difficult to distinguish between relevant information and secondary information or even noise. Current research works developed user query with the user's preferences by creating user profile information for obtaining the results of the query. However, a user profile may not be adequate for different types of user queries. A user uses any of the queries at a time to mine out the desired website. Every browser has its own stage of popularity and reputation in the market as illustrated in [7] which was modeled using the Markov chain procedure and transition probabilities were then analyzed and calculated but the search results were still not effective.

Web mining is an appropriate tool for success for those who use electronic means of operation for conducting their business. Web mining is moreover relevant to data mining techniques that determine patterns from the Web with the application of content mining, structure mining, and usage mining but search results were less accurately predicted. A-Priori algorithm identifies the association rules as explained in [19] and was applied to different applications like structure improvement by page caching, site modification, and page personalization. The work in [21] made use of a framework with ontology based testing framework for functional and non-functional testing of web services.

Adaptive Web Access Pattern Tree for FSP mining combined Suffix tree and Prefix tree for well-organized storage as illustrated in [4] that enclosed a given item for all the obtained sequences. It eliminated recursive reconstruction of intermediate WAP tree during the mining by transfer the binary codes to each node in the WAP Tree but it takes longer time to compute. Statistical Process Control (SPC) framework as described in [17] identified the root causes for performance anomalies by applying differential profiling but failed to automate the process of identifying the root causes.

A local anomaly detector as explained in [8] recognized individual records with anomalous attribute values, and then noticed the patterns by recording the digit of anomalous records. But the record was higher than the expected but not evaluated on disease outbreaks. The detection of web attack as demonstrated in [9] analyzed web server log file where the intrusion detection consisted of events, system shaped and operated to detect system intrusions.

Accurate identification of intrusion of system with person name is cooperative in various web related tasks such as information retrieval. The routine removal of lexical pattern based approach as described in [5] resourcefully extracted out a larger set of candidate aliases from snippets retrieved from a web search engine. It defined the frequent ranking scores to assess the aliases of the candidate. A large-scale evaluation framework as illustrated in [13] for personalized search based on query logs was not applicable in snippet clicks. An account user's browsing behavior when viewing a page, such as dwelling time, page scrolling, mouse snap, and mouse group were not investigated.

Query personalization is based on the surveillance that different users locate different answers relevant when searching because of different preferences and goals. It generates personalized results by enhancing a query with related preferences stored in a user profile and changing the order and perhaps the size of results. The preferences were connected to a request and influenced the final answer based on the query, the profile and the personalization logic. A personalized answer is always graded according to the user's preferences, depending on the query personalization logic and included all results that matched the query or only a subset that satisfied certain related preferences.

Web database related preferences presented an unsupervised, online record matching method in [20] which efficiently recognized duplicates from query result records obtained from multiple Web databases. After removal of the same-source duplicates, the presumed non duplicate records from the same source was used as training examples to improve the burden of the users. Saturn as developed in [16] deal successfully with these query personalizing problems using a novel multiple ring order-preserving.

The use of a novel order-preserving hash function made fast query processing. Replication across and within query rings formed the foundation over

Journal of Theoretical and Applied Information Technology
20th September 2014. Vol. 67 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

which mechanisms were developed by making sure the query load balancing and fault tolerance correspondingly. The Ochiai algorithm as illustrated in [18] included all the faults but the additional similarity metrics cannot be directed for test generation and did not evaluated the effectiveness. Test generation and fault-localization techniques for client side were not developed.

DoubleGuard, an IDS system as demonstrated in [14] has localized faults in network performance of user sessions transversely on both the front-end web server and the back-end database. DoubleGuard, analysis failed to consider the potential for caching expensive requests to reduce the end-to-end latency. An automatic annotation approach as described in [12] first aligned the data units on a result page into diverse groups such that the data in the same group had the identical semantic. At the same time, decorative tag detection was not perfect in automatic annotation approach, which resulted in certain tags to be falsely detected as decorative tags, leading to mistaken merging of the values of different attributes.

The objectives of Predictive User Preference Identification (PUPI) technique include the following (i) to develop PUPI technique to obtain better search result for the users by extracting the semantic and lexical information of an object, (ii) to redefine the queries according to the user relational profile using Condition Redefine Query and stored query context using PUPI technique, (iii) to obtain the result according to the retrieval usefulness by creating more complete query model that represents better information need and (iv) to obtain effective result by integrating all the related and presumed words in the stored database query model using weka tool for web user prediction framework design and analysis.

The structure of paper is as follows. In Section 1, the Query personalization and drawbacks of the exiting user preference technique on web is presented. In Section 2, the Predictive User Preference Identification Technique based on the relational users and the stored queries is demonstrated. Section 3 explains about the experimental setup with parametric factor description. Section 4 analysis the result with the help of the table and graph values and section 5 illustrates the existing works with limitations. Section 6 finally concludes with beneficial web result for the users.

2. MATERIALS

Predictive User Preference Identification Technique provides personalized results to relevant users by exploiting the context. Consequently the web stored queries are improved to take part in the current user task. Following this a new query language is formed on the principle of query redefining based on the user task model and the measured user profile. Figure 1 presents the system architecture of PUPI technique.

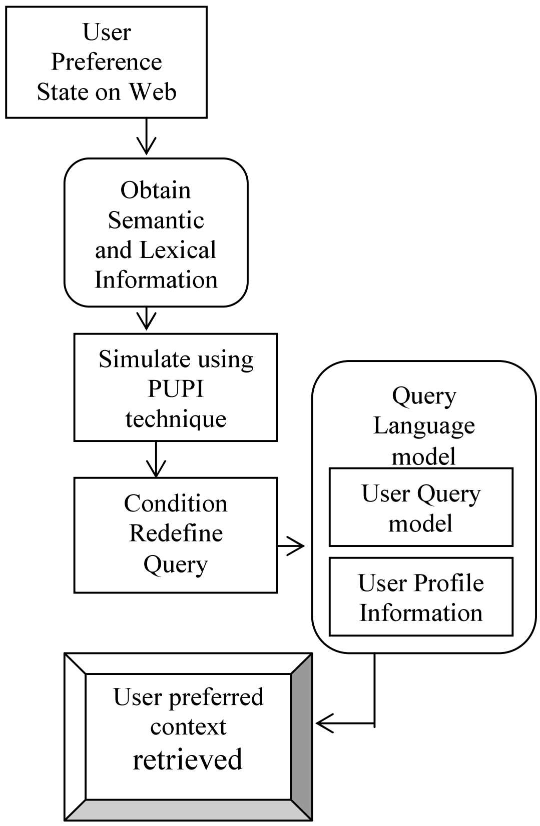

Figure 1: System Architecture Of PUPI Technique

Figure 1 illustrates the system architecture for the overall system process based on the relational users and stored queries. The state of user preference depends on the semantic and lexical information. Lexicon object with an amount of lexical entries describe the syntactic properties and is plotted via lexical sense objects in PUPI technique. User preferred semantic lexicon information in PUPI analyses the dictionary of words labeled with semantic classes and associations between the words. The user condition query is redefined for adding the relevant terms to the original user query. Depending on these relevant terms, the initial user

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

query is redefined, which are related to the current user preferences for producing Condition Redefine Query (CRQ).Query (CRQ).

User profile information consists of a collection of personal data associated to a specific user. In PUPI technique, dynamic user profile is measured as composite data in the form of attribute-value pairs where each pair represents a profile's property. The properties in PUPI technique are grouped into categories to help users to recognize the relationships between concepts. Additionally, PUPI keep away from the use of incorrect concept inside queries. The annotating of each notion in PUPI user query model is performed by giving value for each attribute based on the accumulated similarity score. A user profile is formed consisting of non-null values. The updating of query language model is performed using a relevance technique for building the restructured user query model and user profile information. Finally the user preferred context are retrieved through search engine.

2.1 Condition Redefine Query processing

Query expansion in PUPI technique refers to the process of adding relevant terms to the original query. However, in a more general sense, it also refers to methods of query redefinition. The requisites are applicable if they are associated to the query, user profile, and user preference task state in the same time and do not contain unrelated requisites. Depending on these relevant terms the initial user queries are redefined which are related to the current user task for producing Condition Redefine Query. CRQ in PUPI technique is handled by evaluating the different user preference states of the current task because each users query state relevant to the results for the same user query will be different.

The user preference states and user profiles are modeled by predicates of typical form <attribute, cost>. Each user state Uj has at least one predicate which consist of the pairs (aUj, cUj> aUj) refers to the state attributes of the user task preference at the state U_j,

AUj = ∑ aUj where j >= 1 ……… Equation (1)

Let (aPk, cPk) represent the user profile attribute cost pairs with Pk representing the relevant user profile attributes, thus:

$AP_k = \sum_1^k aP_k$ where k>=0 …… Equation (2)

Let IQ an initial query which is composed of many requisites {r1, r2... rn} and related to the user preference at hand. The user redefine query in the user preference state U_j is as follows,

$CRQ(IQ, P_k, U_j) = IQ + AU_j + AP_k$
……….. Equation (3)

The relevant results Ri for the user preference states Uj are produced by applying CRQ (IQ, Pk, Uj) using the user information retrieval context. The results Ri for user preference state Uj are more relevant than the normal results produced using the initial query IQ in Uj. In the following sections the CRQ produced depending on user profiles and user preference task is presented.

2.2. PUPI Query Language Model

Let us assume a general language model for query extension including the contextual factors and user profile information. A typical gain function for query language modeling using PUPI technique is defined as follows

Gain (IQ, R) = $\sum r_{\in} S\ P(r|\theta_{IQ}) \log P(r|\theta_{IQ})$

……….. Equation (4)

Where, θIQ is a query language model using PUPI technique, estimated by relative frequency of keywords in the query, and S is the set consisting of lexical and semantic information. P (r|θIQ) is the probability of requisite 'r' in the query model. Let us use θIQO to denote the PUPI original query model, θIQU for the contextual user preference state model, and θIQP for a user profile model.

P (r|θIQ) = ∑ i€Y aj P(r| θIQ) ….. Equation (5)

Where, Y= {0, U, P} is the set of all component models and a_j are the mixture weights of query.

The retrieval operation in PUPI technique is limited to semantic and lexical matching, based on the few words in the query. To improve retrieval effectiveness, PUPI technique creates more complete query model that represents better information required. In particular, all the related and presumed words are included in the PUPI query model. In these cases, the initial query model contained the basic requisites query is constructed and a new model CRQ containing the added requisite terms (i.e.,) complex terms.

Journal of Theoretical and Applied Information Technology
20th September 2014. Vol. 67 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

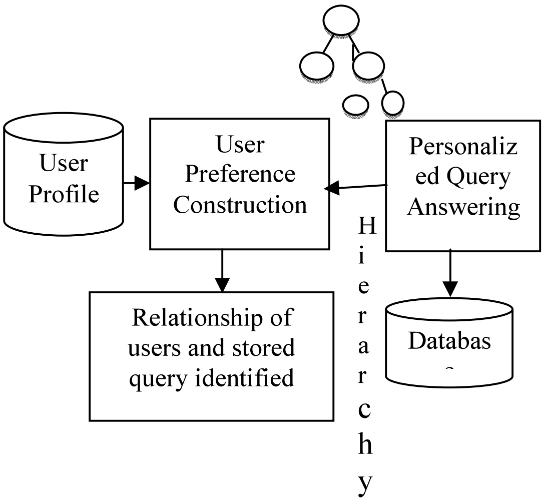

Figure2: User Profile Relationship With PUPI Technique

User Profile Relationship with PUPI Technique is illustrated in Figure 2. The user preferences in PUPI technique is associated with a query that are mined from the user profile. Given that user query comprises of a mixture of generic and more explicit preferences, which routinely identify the relationships between the generic and explicit preferences and then are organized them into a network that captures these relationships. Subsequently, PUPI personalized query answering applies different traversal strategies to process the preference results accordingly. The PUPI query system extracts the user predictive preference from the user profile information and relates with stored query database. User Preference system integration is responsible for systematizing them in a network that captures the relationships between them and determines the association between the pair of preferences.

2.3 Predictive User Preference Identification

Finally the actual process involved in PUPI technique is to map the features with the user preferences. The general implicit feedback analysis strategy relies on hierarchical form and train a classifier to induce weights. The weight of each user behavior features accordingly obtain a predictive identification model of user preferences. The training is performed using PUPI technique by comparing a broad choice of implicit performance measures with explicit user judgments for a set of queries.

FAS (Miss) = Set of results for a query and a relate result at location l, all unrelated results are ranked above l are predicted to be less relevant than the result at l.

FAS (Miss Above + Miss Next) = Strategy predicts all unrelated results immediately following a related result as less relevant than the related result, and combines these predictions with those of the FAS strategy.

For PUPI, a large random sample of queries using the search query log of web search engine produce effective results for each of the queries and explicit relevance judgments are available for each result pair. The user behavior for all the instances where these queries were submitted to the search engine is analyzed. To learn the mapping from features to relevance user preferences, a scalable Feedback Analysis Strategy (FAS) is implemented and demonstrated as given below.

In addition to information about results, user profiling is also examined. Miss Next in FAS strategy uses user profile information to predict the result following the related result at 'l' which is less relevant than the related result, with accuracy comparable to the FAS strategy. For better accuracy result, combine the FAS strategy with extension of Miss Above + Miss Next. The algorithmic step for PUPI technique is described below,

Predictive User Preference Identification

Input: Query 'IQ', User Preference 'U', User Profile Information 'P'

Output: Preferred User Result retrieved

Begin

Step 1: User Preference State on Web search engine is initialized

Step 2: While (Uj > 1)

2.1: User preference sensed based on

2.2: Syntactic properties and Lexical sense objects

2.3: End While

Journal of Theoretical and Applied Information Technology
20th September 2014. Vol. 67 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Step 3: Condition Redefine Query

3.1: Predict <attribute, cost> <aUj, cUj>
3.2:Compute

$$CRQ(IQ, P_k, U_j) = IQ + AU_j + AP_k$$

Step 4: Query Language Model

4.1: User Query model ∑ i€Y aj P(r| θIQ) computation

4.2: User Preference system integration for user profile information relate with stored database Query

Step 5: FAS combines the Miss Above + Miss Next strategy for micro averaged accuracy in query result

End

PUPI algorithm returns the user preferences from the information in the user profile that are related with a query IQ. The preferences are organized in a network, which is the output of the algorithm. User preferences that are associated to a given query comprise not only the explicit preferences stored in the user profile but also the inherent ones that consequent by composing stored preferences. In order to select the user preferences, the PUPI algorithm starts from the preferences that are stored in the user profile and are related to the query iteratively from the database. The set of preferences that are related to the query is kept ordered with increased query frequency similarity interest.

3. PUPI EXPERIMENTAL SETUP

Predictive User Preference Identification (PUPI) technique is developed in JAVA using weka tool. The PUPI technique uses the MSNBC.com Anonymous Web Dataset for experimental work which is based on the relational users and stored queries. The data describes the page visits of users who visited msnbc.com and also the recorded visits at the point of URL category in time order. MSNBC.com Anonymous Web contains the 989818 users with 5.7 average number of visit per user. The URL per category ranges from 10 to 5000. Each group is connected in order with an integer starting with "1". For example, "frontpage" is connected with 1, "news" with 2, and "tech" with 3. Each row below "% Sequences:" explain the hits in order of a single user.

The data is obtained from the Internet Information Server (IIS) logs for msnbc.com and news-related portions of msn.com for the complete day. Each series in the dataset communicates with the page views of a user throughout the twenty-four hour period. Each event in the sequence corresponds to the user's request for a page. Requests are not recorded at the premium level of feature (i.e.,) but rather, they are recorded at the level of page category. The categories are "frontpage", "news", "tech", "limited", "view", "on-air", "misc", "weather", "health", "living", "business", "sports", "outline", "bbs" (bulletin board service), "travel", "msn-news", and "msn-sports". Any page requests served via a caching mechanism in web were not proofed in the server logs and, hence, not present in the data.

Execution time on Predictive User Preference Identification technique using MSNBC.com Anonymous Web Dataset is the amount of time taken to relate the user preferences task with stored database information, measured in terms of seconds (sec). Averaged accuracy parameter in PUPI indicates the extent to which it is able to accurately relate the user profile information with user preferred query result. PUPI effectively obtained the averaged accuracy and measured in terms of percentage against the existing personalized ontology model and Query planning for WAAQ method.

Query user relationship factor illustrates the effectiveness in relating the user query with user profile information Pj, measured in terms of percentage (%). Query frequency similarity refers to the frequency amount in relating the user query result according to the users need, measured in terms of Hertz (Hz). True positive rate factor measures the proportion of actual positives which are correctly identified. The correctly identified result is measured through sensitivity ratio. Power consumption factor is also measured in PUPI, describe as the amount of power consumed while experiment the user query preferences in web search engine, measured in terms of Watt (W).

4.DISCUSSION OF PUPI TECHNIQUE

Predictive User Preference Identification (PUPI) technique is compared against the existing personalized ontology model and Query planning for Weighted Additive Aggregation Queries (WAAQ). Table 1 given below evaluates the value and graph describes the PUPI technique improvements with

Journal of Theoretical and Applied Information Technology
20th September 2014. Vol. 67 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

beneficial user preference identification results when compared to the two existing models.

Search Log Size (KB)	Execution Time (sec)		
	Personalized Ontology Model	WAAQ method	PUPI Technique
25	223	215	192
50	254	242	212
75	341	310	285
100	400	396	326
125	541	526	451
150	649	640	528
175	745	740	645

Table 1: Tabulation Of Execution Time

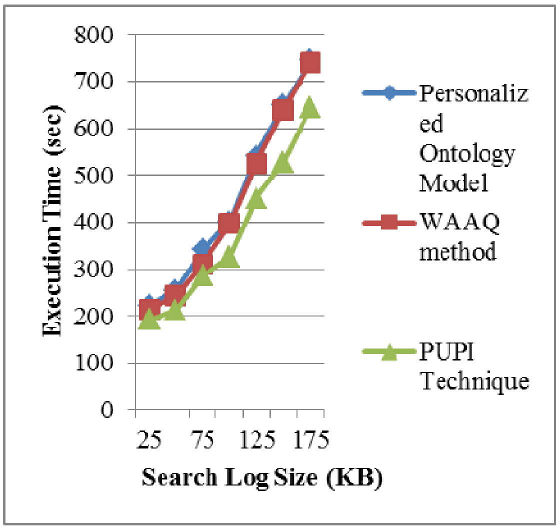

Figure 3: Measure of Execution Time

Figure 3 describes the execution time based on the search log size of Personalized Ontology Model, WAAQ method and PUPI technique. The notion in PUPI user query model gives value for each attribute based on an accumulated similarity score. The updating of query language model is performed using the relevance technique which builds the restructured user query model with minimal execution time. PUPI technique consumes 13 – 18 % lesser execution time when compared with the Personalized Ontology Model [1] and 8 – 17 % lesser time when compared with the WAAQ method [2].

Table 2 describes the micro averaged accuracy values based on the number of test queries made using PUPI technique. The value of the proposed PUPI technique is compared with the existing Personalized Ontology Model and WAAQ method on web.

Table 2: Tabulation For Micro Averaged Accuracy

No. of Test Queries	Micro Averaged Accuracy (%)		
	Personalized Ontology Model	WAAQ method	PUPI Technique
2	75	68	80
4	77	69	81
6	78	70	81
8	80	71	82
10	81	72	83
12	82	73	85
14	83	73	88

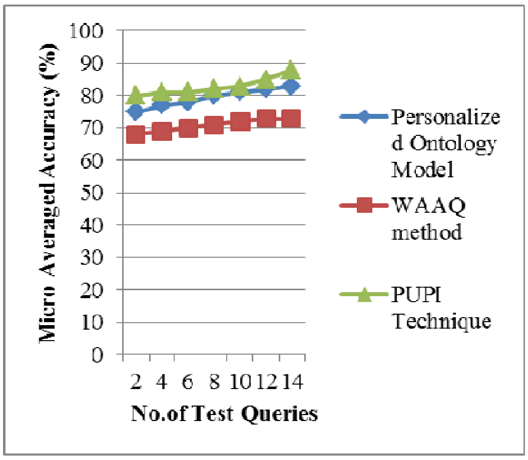

Figure 4 Micro Averaged Accuracy Measure

Figure 4 illustrates the measure of micro averaged accuracy with respect to the number of test queries made on web. The query is measured in PUPI technique and attains the better accuracy result with automatic implicit feedback analysis strategy

Journal of Theoretical and Applied Information Technology
20th September 2014. Vol. 67 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

analysis. The weight of each user behavior features accordingly obtain a predictive identification model of user preferences with 2 – 6 % improved accuracy rate when compared with the Personalized Ontology Model [1]. PUPI technique is 15 – 20 % improved in accuracy when compared with WAAQ method [2].

The Query User Relationship Ratio to request the query and obtain the result on the web is measured based on the user density. The average query user relationship ratio is tabulated in table 3. The value of the proposed PUPI Technique is compared with the existing Personalized Ontology Model and WAAQ method.

Table 3 Tabulation For Query User Relationship Ratio

User Density	Query User Relationship Ratio (%)		
	Persona-lized Ontology Model	WAAQ method	PUPI Technique
50	80	82	90
100	81	83	91
150	82	84	92
200	82	85	93
250	82	87	93
300	83	88	94
350	85	89	95

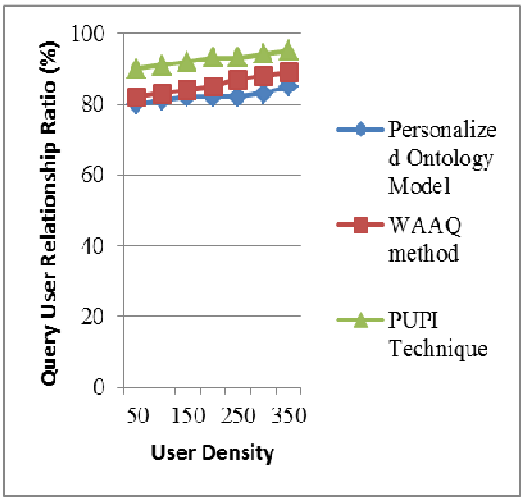

Figure 5: Query User Relationship Ratio Measure

The query user relationship ratio is measured in Figure 5 using the Personalized Ontology Model, WAAQ method and PUPI Technique. The user density ranges from 50 to 350. Query user relationship ratio is improved using the user preferred semantic lexicon information in PUPI analyses. PUPI analyses the labeled semantic classes and associations between the words. The user condition query is redefined for adding the relevant terms to the original user query in PUPI technique and improved by 11- 13 % when compared with the Personalized Ontology Model [1]. A PUPI analysis is also improved by 6 – 9 % when compared with the WAAQ method [2].

The table (Table 4) given below and the graph describe the performance of the query frequency similarity against Personalized Ontology Model and WAAQ method.

Table 4 Tabulation Of Query Frequency Similarity

User Group Size	Query Frequency Similarity (Hz)		
	Persona-lized Ontology Model	WAAQ method	PUPI Technique
5	122	127	137
10	124	132	139
15	133	138	150
20	142	153	164
25	149	152	167
30	167	175	189
35	197	217	231

Journal of Theoretical and Applied Information Technology
20th September 2014. Vol. 67 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

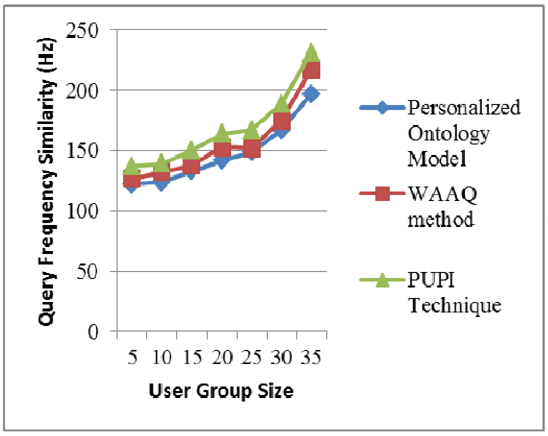

Figure 6 Performance Of Query Frequency Similarity

The query frequency similarity is measured based on the user group size. The group size ranges from 5, 10, 15, 20 up to 35. As the group count varies, query frequency similarity is improved by 12 -17 % when compared with the Personalized Ontology Model [1].

Figure 6 illustrates the query frequency similarity based on the group size. In PUPI technique, dynamic user profile is measured as composite data in the form of attribute-value pairs where each pair represents a profile's property and improves the query frequency. The query frequency similarity is measured based on the user group size. The group size ranges from 5, 10, 15, 20 upto 35. As the group size varies, query frequency similarity is improved by 12 -17 % when compared with the Personalized Ontology Model [1]. The properties in PUPI technique are grouped into categories to help users to recognize the relationships between the concepts and improved by 5 – 9 % when compared with the WAAQ method [2].

Table 5 Tabulation of True Positive Rate

No. of Instances	**True Positive Rate (Sensitivity)**		
	Personalized Ontology Model	WAAQ method	PUPI Technique
20	83	88	91
40	84	90	93
60	85	91	95
80	85	91	96
100	86	92	96
120	87	93	97
140	88	93	98

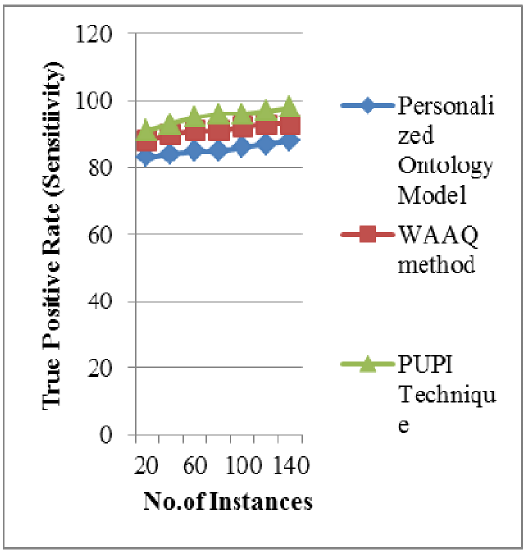

Figure 7 Measure Of True Positive Rate

Table 5 and figure 7 illustrates the measure of true positive rate based on the instances. As the instances range varies, the true positive rate attains the maximum sensitivity. The true positive rate is increased by 9 – 12 % when compared with the Personalized Ontology Model [1] and 3 - 5 % increase when compared with WAAQ method [2] The relevant results (i.e.,) true positive rate is attained by applying CRQ (IQ, Pk, Uj)on user information retrieval context. The true positive rate of PUPI technique is more efficient than when compared to the WAAQ method and Personalized Ontology Model.

5. RELATED WORK

Three significant human factors that have been examined in existing empirical studies include counting gender dissimilarity, obtaining previous information and generating cognitive styles but there remains certain amount of limitations with the present research. The research and development communities need to be conscious and which should be investigated during future research. To investigate on web clustered usage data with overcoming the above limitations, a technique named feed forward back propagation algorithm was designed in [15] to discover the web usage pattern

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

and profiles by reporting the accurate semantic search results.

Semantic Web Mining as demonstrated in [11] finds out and recovers useful and interesting patterns from huge set of web data. The web data consists of different types of information, including web structure data, web log data and user profiles data. Semantic Web Mining is a comparatively new area, usually interdisciplinary, pulling the researcher from computer science, information retrieval specialist and specialist from business studies fields to web. The conceptual prediction model as explained in [10] automatically generates a semantic network of the semantic web usage knowledge. Web usage knowledge is the integration of domain knowledge and web usage knowledge but the extreme comparisons on semantic query web-page recommendation systems were not performed.

Query planning for continuous aggregation queries as explained in [2] offered a technique for obtaining the optimal set of sub-queries with their incoherency bounds. Incoherency bound gratify client query's coherency situation with smallest amount number of refresh messages send from aggregators to the client. Bt the query cost model were not used for other factors such as load balancing of various aggregators, multi query execution.

Online Social Networks (OSN) as demonstrated in [6] give users the ability to control multiple messages posted on their own private space to avoid that unwanted content is displayed. But the model OSN failed to integrate contextual information related to the name of all the groups in which the user participates, properly weighted by the participation level. A deeper investigation on two interdependent personalized tasks concerns the extraction and selection of related features which has shown a high discriminative power.

Personalized ontology model for knowledge representation and reasoning over user profiles as shown in [1] discover ontological user profiles from regularly world knowledge base and user confined instance repositories. The research contributes to knowledge engineering that has the potential to improve the design of personalized web information gathering systems but failed to generate user local instance repositories for matching the representation of a global knowledge base. These strategies were not extended to apply for ontology model with most of the existing web documents.

6. CONCLUSION

Predictive User Preference Identification technique develops a system to relate the user query and stored database queries. The user context is measured as the authentic user state of the preferences that is undertaken when the information retrieval process is performed. User retrieval process constructed a general design that merges several models for query expansion. The user contextual model, the user profile information model and CRQ model are used in PUPI technique for effective query result. A query language model for query development includes the suitable factors and user profile information. The experimental study results obtained by CRQ queries are more relevant than those obtained with the initial user queries in the same user preference task state. MSNBC.com Anonymous Web Data usage in experimental setup attains the minimal execution time and power consumption. PUPI technique also achieves the micro averaged accuracy, averagely 14.814 % higher query user relationship ratio, true positive rate and query frequency similarity than compared to the existing work. Different models have different advantages and limitations. A promising direction we will explore in the future is to automatically anticipate which algorithm can be used for given a set of query to increase the efficiency of the model in web.

REFRENCES:

[1] Xiaohui Tao, Yuefeng Li, and Ning Zhong,"A Personalized Ontology Model for Web Information Gathering," IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 23, NO. 4, APRIL 2011

[2] Rajeev Gupta and Krithi Ramamritham,"Query Planning for Continuous Aggregation Queries over a Network of Data Aggregators," IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 24, NO. 6, JUNE 2012

[3] Resul Das., Ibrahim Turkoglu., "Creating meaningful data from web logs for improving the impressiveness of a website by using path analysis method," Expert Systems with Applications, Elsevier Journal., 2009

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[4] S.Vijayalakshmi, V.Mohan, S.Suresh Raja, "MINING OF USERS' ACCESS BEHAVIOUR FOR FREQUENT SEQUENTIAL PATTERN FROM WEB LOGS" International Journal of Database Management Systems (IJDMS) Vol.2, No.3, August 2010 DOI : 10.5121/ijdms.2010.2304

[5] Danushka Bollegala, Yutaka Matsuo and Mitsuru Ishizuka,"Automatic Discovery of Personal Name Aliases from the Web," IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 23, NO. 6, JUNE 2011

[6] Marco Anisetti, Claudio A Ardagna, Valerio Bellandi, Ernesto Damiani and Salvatore Reale,"Map-Based Location and Tracking in Multipath Outdoor Mobile Networks," IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, VOL. 10, NO. 3, MARCH 2011

[7] Diwakar Shukla, Rahul Singhai, "Analysis of Users Web Browsing Behavior Using Markov chain Model," Int. J. Advanced Networking and Applications 824 Volume: 02, Issue: 05, Pages: 824-830 (2011)

[8] Kaustav Das, Jeff Schneider, Daniel B. Neill, "Anomaly Pattern Detection in Categorical Datasets," ACM journal, 2008

[9]Priyanka Patil, and Ujwala Patil, "Preprocessing of web server log file for web mining," World Journal of Science and Technology, 2(3):14-18, ISSN: 2231 – 2587, 2012

[10] Thi Thanh Sang Nguyen, Hai Yan Lu, Jie Lu, "Web-page Recommendation based on Web Usage and Domain Knowledge," IEEE, 2013

[11] Anil Sharma, Suresh Kumar, Manjeet Singh, "SEMANTIC WEB MINING FOR INTELLIGENT WEB PERSONALIZATION," Journal of Global Research in Computer Science, Volume 2, No. 6, June 2011.

[12] Yiyao Lu, Hai He, Hongkun Zhao, Weiyi Meng, and Clement Yu, "Annotating Search Results from Web Databases," IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 25, NO. 3, MARCH 2013

[13] Zhicheng Dou, Ruihua Song, Ji-Rong Wen and Xiaojie Yuan,"Evaluating the Effectiveness of Personalized Web Search," IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 21, NO. 8, AUGUST 2009

[14] Meixing Le, Angelos Stavrou, and Brent ByungHoon Kang, "DoubleGuard: Detecting Intrusions in Multitier Web Applications," IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING, VOL. 9, NO. 4, JULY/AUGUST 2012

[15] Sherry Y. Chena, Robert Macrediea, "Web-based interaction: A review of three important human factors," International Journal of Information Management., Elsevier Journal., 2010

[16] Theoni Pitoura, Nikos Ntarmos, and Peter Triantafillou, "Saturn: Range Queries, Load Balancing and Fault Tolerance in DHT Data Systems," IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 24, NO. 7, JULY 2012

[17] Donghun Lee, Sang K. Cha, and Arthur H. Lee, "A Performance Anomaly Detection and Analysis Framework for DBMS Development," IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 24, NO. 8, AUGUST 2012

[18] Shay Artzi, Julian Dolby, Frank Tip, and Marco Pistoia, "Fault Localization for Dynamic Web Applications," IEEE TRANSACTIONS ON SOFTWARE ENGINEERING, VOL. 38, NO. 2, MARCH/APRIL 2012

[19] Preeti Sharma, Sanjay Kumar, "An Approach for Customer Behavior Analysis Using Web Mining," International Journal of Internet Computing (IJIC), ISSN No: 2231 – 6965, Volume-1, Issue-2, 2011

[20] Weifeng Su, Jiying Wang, and Frederick H. Lochovsky, "Record Matching over Query Results from Multiple Web Databases," IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 22, NO. 4, APRIL 2010

[21] N. Danapaquiame, E. Ilavarasan and Neeraj Kumar," A FRAMEWORK TO MAGNITUDE THE PERFORMANCE AND BEHAVIOR OF WEB SERVICES USING ONTOLOGY", Journal of Computer Science 10 (4): 640-646, 2014.