ijecs open access www.ijecs.in
International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 2 Issue 12, Dec.2013 Page No. 3340-3344

Extraction of Ontological Information based on Semantic Analysis

Nataraj J, Mural Krishnan P, Praveen Godfrey, Arun S.N.
School of Information Technology, VIT University,
Vellore, Tamil Nadu, India.
E-mail id: natarajlogin@gmail.com

Abstract: ***In the rapid growth of communication field and updating of data rate that make the task to find relevant information retrieval is more difficult in the modern world. Probably User always wants to access information from the web search log, to gain the context-based data for their required information. Many researchers worked on the process of content-based information retrieval process in the web search engine. The model addressed inability to describe relations between search terms. In this paper, present an ontological based pattern description technique is the extraction of semantic data from search log through the query processing system. An experiment shows that this model can increase a usage scenario of the person.***

Keywords: **Semantic analysis, Pattern log, Query Evaluation, Web Search log**

1. Introduction

Information Retrieval is the process of gathering data from communication network. If the user want to access the data without technical knowledge, then it will become difficult to access the user required data in the web. Based on the analysis of the information retrieval problem, some other researchers worked on the area to implement the context based information retrieval system. To reduce the user's complexity, it could analyze and produce the result as the user requirement on the search log. But there is no guarantee for information correctness and lots of conflicting data are retrieved by the search engines and the quality of provided information also varies from low quality to high quality. It's given only related information of the user's concept in the web. From the several aspect of user's interruption could be analyzed and construct the retrieval based Semantic analysis. In this concept explain the properties of information retrieval process. One special feature is added to provide some security while in the time of retrieving the data.

Retrieval of data from the communication system is the huge problematic to analysis it. Here semantic ontology used to extract information from the web resource efficiently. To create pattern log, Query evaluation and semantic analysis in the onto-frame node. Through the onto-frame node, every data can be accessed at the users effectively. In this model, get some optimized search result and also increases accuracy, efficiency and frequent to retrieve the data. Extract an updated result is evaluated using pattern log whenever server can update.

The rest of the section is discussed as section II explain the related work, section III describe the semantic extraction and pattern analysis, section IV detail about the result evaluation, section V conclude the theme of work and section VI reference of varies author about their concern work.

2. Related Works

Based on the analysis of information retrieval from the search log, some researchers tried to increase the performance of gathering data. They deployed the concept of probabilistic query expansion from search log. From this concept, got some query reformulation and expansion based on the users requirement. In this model, A Context-sensitive substitution algorithm used to perform content-based searching process.

Some other researchers like Erkowitz et al. investigate the problem of index page synthesis, which is the automatic creation of pages that facilitate a visitor's navigation of a website. By analyzing the web log, their cluster mining algorithm finds collections of pages that tend to co-occur in visits and puts them under one topic. They then generate index pages consisting of links to pages pertaining to a particular topic.

Nakayama et al.[1] also try to discover the gap between the website designer's expectations and visitor behavior. Their

approach uses the inter-page conceptual relevance to estimate the former, and the inter-page access co-occurrence to estimate the latter. They focus on website design improvement by using multiple regressions to predict hyperlink traversal frequency from page layout features.

Spiliopoulou[10] et al propose a "web utilization miner" (WUM) to find interesting navigation patterns. The interestingness criteria for navigation patterns are dynamically specified by the human expert using WUM's mining language which supports the specification of statistical, structural and textual criteria.

Alexandre Viejo [3] user profiles by submitting *fake* queries to the WSE. The system locally profiles users according to their social network accounts. This strategy is assumed to provide better personalized service than other schemes in the literature. The privacy dimension is addressed by building *fake* queries that keep the *macro-interests* of the user while hiding her micro-interests.

Eugene Agichtein [2] describe relevance feedback is used in the positive relevance feedback and negative relevance feedback are combined to form a new query method, using negative relevance feedback of information to improve the accuracy of retrieval, similar to the vector space model feedback algorithm for query expansion. These degrees of relevance are still dependent on the mindset of the person who actually needs these documents. Fox et al used an instrumented browser to determine whether there was an association between explicit ratings of satisfaction and implicit measures of searcher interest and identified the measures that were most strongly associated with user satisfaction. They found that there was a link between user activity and satisfaction ratings, and that click through, dwell time, and session termination activity combined to make good predictors of satisfaction for Web pages. Fox et al. found that short dwell times and clicking many (four or more) search results for a query were both indicators of dissatisfaction. Behavioral patterns were also used to predict user satisfaction for search sessions.

3. Semantic Evaluation process

In the approach, User- supplied data can be retrieval effectively based on the semantic analysis. To make it in an effective way, construct the onto-frame node in the client side. Whenever clients want to search their data, onto-frame enable its intra-component communication of the system. Each component has updating log and also behavior instance of the clients. Then it can connect to the server through the communication medium.

Server

Server is the centralized part of the web service. It can maintain the huge database and response to the clients who are request the service of their requirements. Server can update the data whenever it receives the latest information. It can connect with data log.

Data Log

It consists of semantic representation of data stored in the log and also based on the ranking relationship to extract the data. It allows local data sources to be supplemented with data tables which have been extracted from the Web documents, the domain specific part of the log can manually built by ontologisms taking into account 1) the vocabulary used in the preexisting local databases in order to index the data and 2) the domain information available within the databases schema. This data log connected with the server to analyze the information.

Onto-frame node

It consist of 3 communicating model to elaborate the descriptive phase,

Query evaluation

This blog process the data in the form of queries under the two ways of operation. In the first operation, gathering the data from the search blog and send it to the server. Another operation is the extraction of ontological information from the data log and response to the semantic log. In each time the information are updating in the data log. This blog can continuous monitoring to access the updating information and send it. The user's request is evaluated in the semantic log. The queries can be compared with the semantic information to the pattern log and transform into search log whatever the user could approach it.

Pattern log

In this log, users profile updated data can be stored as pattern. Each time user's data can be updated whenever they are searching in the search log. The data can update from the communication service at every time based on the newly added documents in the Server. The pattern log gets input from the semantic log and stored as behavioral pattern when every time semantic log produces an updating result. The pattern log validates the user's requested query to the behavioral pattern and response to the search log.

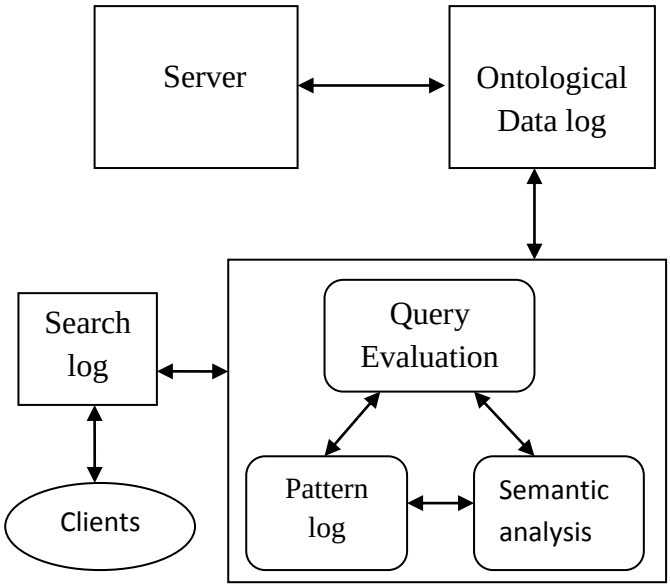

Fig.1. Semantic retrieval framework

This log operation works under the analysis of behavioral pattern algorithm to describe their functionality and produce the result. From the below algorithm explain the validation process of each semantic data. If the data reject it go to search until it match the behavior pattern. In other end, the extracted data can be processed in the semantic log and update the behavioral pattern in the pattern log.

Algorithm: Analysis of behavior pattern

Procedure

input: Semantic data s, Query r

output: validated result to the user

initialization

String data ;

Store the semantic behavior from the semantic log in the form of pattern s

Evaluate requested data

for i=1 to s

if(r.data == s.data)

return r.data;

else

Fail to response the data

Go to next until it validated.

endif

endfor

end

Semantic Log

The main originalities of querying subsystem of the semantic analyzes are: 1) to retrieve not only exact answers compared with the selection criteria but also semantically close answers; 2) to compare the selection criteria expressed as fuzzy sets representing preferences with the fuzzy annotations of data tables

.

The log consists of semantic behavior (O2) and user behavior (O1). The user's search result can be mapped and stored in O1. Example: If user searches a keyword "Sun", it can search the user's behavior of the conceptual keyword. In the user's behavior have the updated result about many user's can try to retrieve the answer x. This kind of information assigns a priority O2.Another behavior is the semantic feature of the text which user can search. The sim(a,b) is calculated by using probabilistic distribution model.

$$P(Sim(a,b)) = P(comm(a,b)) - P(diff(a,b)) + P(wrinkle(a,b))$$

Where comm(a,b)indicates the commonality between a and b. diff(a,b)stands for the difference and winkler(a,b) for the improvement of the result using the method introduced by Winkler. Two property values a and b are equal if their string similarity is greater or equal to the predefined threshold δ. Therefore, we define an equality function as follows:

$$P(sim(a,b)) = \begin{cases} 1, & sim(a,b) >= \delta \\ 0, & otherwise \end{cases}$$

Where E(sim(a,b)) be the semantic feature of the text. The collection of text can be analyzed through the semantic feature and stored in the semantic behavior O2.Semantic information is extracted from the different ontology's in the web service. That means, independent data can manipulation from the web resource into preferable information to the user. Let us take O1 and O2 be the independent data and find the similarity between them

$$Sem(O1,O2) = \frac{\Sigma (O1 \wedge O2)}{O1}$$

Here O1 is the user profile recorded data and O2 is the semantic meaning of the data. Result of the sem(O1,O2) is finally produce the response of the semantic log.

Search log

Search log is used to processing of text information and integrate with relevant hyperlink to the data log. It contains the two communications between client and server. In the clients are accessed their view of text information by using the search log itself. The onto-frame produces the output to the search log. This log can store the data and arranges the data based on the precedence basis in the search log.

Clients

Users who are all sharing the data from the communication medium. Each client wants to access the web resource with different thought in different manner to retrieve the data. Based on the perspective behavior, the client can search the information in the search log.

4. Implementation

The model can be observed that the usage of web pages increase the interaction of customer to the communication channel. It is divided into two behaviors such as user's behavior and semantic behavior. In the user behavior update the usage scenario of the searching text. For example, user types the text "apple" in the search log. The data logs extract the ontological data from the server and send it to the onto-frame node. Whenever type the text "apple", the relevant

text can be extracted from the user profile and give priority to the data. The user profile consists of similarity of the text what users can search for the text. For example, if user tries to access the color of the apple, then the user profile contains updated information of apple is red color. Another behavior is to extract the meaning and functionality of the word. For example, the semantic behavior has the probability of the properties, functionalities and historical data about the description of apple fruit. Semantic log separate the users profile oriented data and semantic data from the ontological information. Then pattern log as,

Example: $Q = \pi e,att,v(\sigma v:"Hotel"\ (R)) \cap \pi e,att,v\ (\sigma v :"Istria"(R))$

User required getting information about the hotels in the Istria by using pattern analysis.

$Q=\pi e,att,v(\sigma v:"Hotel" \wedge v:"Istria"(R))$

Finally, get the response from the requested query to the user.

5. Evaluation of Result

The behavior instance is the effective concept used to retrieve the result from the uncertainty information web resource. Many techniques have been proposed in the field of information science. But there is no fulfillment of user's requirement while search and retrieve information from the server.

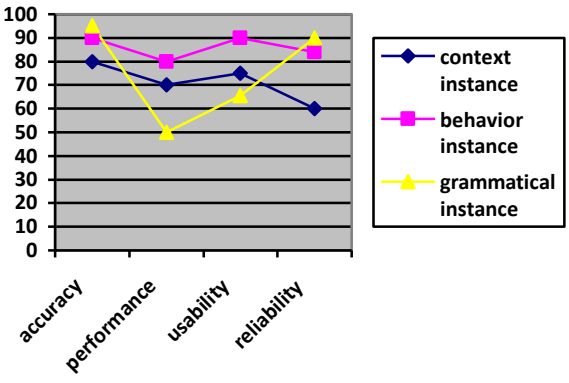

Fig-2: Comparison of behavior instance, context based instance and grammatical instance.

The graph explain the context instance is the property of matching instances are computing by considering the weight factors of the properties associated to them, better outcomes in terms of precision and recall are achieved. Next, behavior instance is the validation of quality pattern to extract the information and produce the response to the user. In this model, become more efficient to handle the searching process. In above graph, describe the advantage of using this model when compare to other model. Finally, discuss the grammatical pattern analysis. In this model is not suitable for the large searching application. It performs the searching process with low usage of retrieve the data but more accuracy of the model.

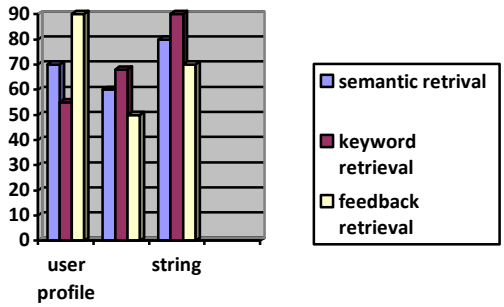

Fig-3: comparison of retrieval model

Above graph describe the effective retrieval model, differentiate the user profile, sting comparison and instance behavior. How effectively match the attribute values to the retrieval process. Based on the analysis, the searching retrieval model can be used to get effective searching process.

6. Conclusion

Information retrieval is the process of gathering data from the web service. But user might not satisfy their search result from the integrated search engine. In this paper, provide the semantic information to the user's uncertain knowledge based search information. The semantic data analyzed by using onto-frame node. This node contains pattern log, semantic log and evaluation log to extract the instance behavioral of user information. Each user has different behavior to search their required information in the search engine. To get their interactive information from the web, this onto-frame node is useful to validate both user's behavior instance and semantic instance of the ontological information. Finally, get the user interface search based result. In this process increase the usage scenario of an interactive result. In the future work to improve performance of the retrieval model, analyze the dynamic instance behavior of the user's search information.

7. Acknowledgement

This research is supported by the VIT University, Vellore under the area of information science. We thank to professor K.Karthikeyan for their valuable comments and suggestions on the ontology design

8. Reference

[1] Maleerat Sodanil and Hathairat Ketmaneechairat **,"**Information Retrieval Experiment on Subjective Words Query Expansion", International conference of information and communication Technology, pp.161-165, 2013

[2] Zheng Lu, Student Member, IEEE, Hongyuan Zha, Xiaokang Yang, Senior Member, IEEE, Weiyao Lin, Member, IEEE, and Zhaohui Zheng,"A New Algorithm for Inferring User Search Goals with Feedback Sessions", ieee transactions on knowledge and data engineering, vol. 25, no. 3, pp. 502-513,march 2013.

[3] Alexandre Viejo, David S´anchez, ”Providing Useful and Private Web Search by Means of Social Network Profiling”, Eleventh Annual Conference on Privacy, Security and Trust (PST),pp.358-361,2013

[4] S.Vigneshwari and Dr. M. Aramudhan, “A Novel Approach for Personalizing the Web Using User Profiling Ontologies”, IEEE- Fourth International Conference on Advanced Computing, ICoAC 2012

[5] Robin Sharma, Ankita Kandpa and Priyanka Bhakuni Rashmi Chauhan 1, R.H. Goudar and Asit Tyagi,”Web Page Indexing through Page Ranking for Effective Semantic Search”, 2012 IEEE,pp. 389-392

[6] Chowdhury Farhan Ahmed, Syed Khairuzzaman Tanbeer, Byeong-Soo Jeong, “Mining High UtilityWeb Access Sequences in DynamicWeb Log Data”, 11th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing,pp.76-81,2011

[7] Eugene Agichtein, Eric Brill and Susan Dumais,”Improving Web Search Ranking by Incorporating User Behavior Information”, SIGIR’06, pp.19-26,2006

[8] Krzysztof Pancerz ,“Semantic Relationships and Approximations of Sets: An Ontological Graph Based Approach”, IEEE Sopot, Poland, June 06-08, 2013,pp. 62-69

[9] Btihal El Ghali, Abderrahim El Qadi, Mohamed Ouadou and Driss Aboutajdine,”*Probabilistic Query Expansion Method Using Recommended Past User Queries”,* 2012 IEEE,pp.406-411

[10] Olfa Nasraoui, Myra Spiliopoulou, Jaideep Srivastava, Bamshold mobasher, “Advances in Web Mining and Web Usage Analysis, 8th International Workshop on Knowledge Discovery on the web,2006 Springer.

[11] Rudra Pratap Deb Nath, Hanif Seddiqui, Masaki Aono,”Resolving Scalability Issue to Ontology Instance Matching in Semantic Web”, 2012 IEEE,pp.396-404