Journal of Engineering Sciences
Vol 11, Issue 4 , April/2020
ISSN NO:0377-9254

# Ontology Based Web Mining Technology

**[1]Syed Chand Pasha,[2]Shaik Khaja Pasha,[3]Syed Sadeq,[4]Mukaramuddin**

[1, 2,3]UG Scholar, Department of Information Technology, ISL engineering college, Hyderabad, India.
[4]HOD, Dept of Humanities and sciences, ISL engineering college,
Hyderabad, India.

**OBJECTIVE**

To developed and implemented personalized-recommendation system, a system that makes use of representations of items and user-profiles based on ontology's in order to provide semantic applications with personalized services.

**SYNOPSIS**

Recommendation systems can take advantage of semantic reasoning-capabilities to overcome common limitations of current systems and improve the recommendations' quality. In this paper, present a personalized-recommendation system, a system that makes use of representations of items and user-profiles based on ontologies in order to provide semantic applications with personalized services. The recommender uses domain ontologies to enhance the personalization: on the
one hand, user's interests are modeled in a more effective and accurate way by applying a domain-based inference method; on the other hand, the stemmer algorithm used by our content-based filtering approach, which provides a measure of the affinity between an item and a user, is enhanced by applying a semantic similarity method

Web Usage Mining plays an important role in recommender systems and web personalization. In this paper, we propose an effective recommender system based on ontology and Web Usage Mining. The first step of the approach is extracting features from web documents and constructing relevant concepts. Then build ontology for the web site use the concepts and significant terms extracted from documents. According to the semantic similarity of web documents to cluster them into different semantic themes, the different themes imply different preferences. The proposed approach integrates semantic knowledge into Web Usage Mining and personalization processes.

**ABSTRACT**

In Web search applications, queries are submitted to search engines to represent the information needs of users. However, sometimes queries may not exactly represent users' specific information needs since many ambiguous queries may cover a broad topic and different users may want to get information on different aspects when they submit the same query. For example, when the query "the sun" is submitted to a search engine, some users want to locate the homepage of a United Kingdom newspaper, while some others want to learn the natural knowledge of the sun. Therefore, it is necessary and potential to capture different user search goals in information retrieval. We define user search goals as the information on different aspects of a query that user groups want to obtain. Information need is a user's particular desire to obtain information to satisfy his/her need. User search goals can be considered as the clusters of information needs for a query. The inference and analysis of user search goals can have a lot of advantages in improving search engine relevance and user experience. Some advantages are summarized as follows. First, we can restructure web search results according to user search goals by grouping the search results with the same search goal; thus, users with different search goals can easily find what they want. Second, user search goals represented by some keywords can be utilized in query recommendation thus, the suggested queries can help users to form their queries more precisely. Third, the distributions of user search goals can also be useful in applications such as re-ranking web search results that contain different user search goals.

Due to its usefulness, many works about user search goals analysis have been investigated. They can be summarized into three classes: query classification, search result reorganization, and session boundary detection. In the first class, people attempt to infer user goals and intents by predefining some specific

classes and performing query classification accordingly. Lee et al. consider user goals as "Navigational" and "Informational" and categorize queries into these two classes. Li et al. define query intents as "Product intent" and "Job intent" and they try to classify queries according to the defined intents. Other works focus on tagging queries with some predefined concepts to improve feature representation of queries. However, since what users care about varies a lot for different queries, finding suitable predefined search goal classes is very difficult and impractical. In the second class, people try to reorganize search results. Wang and Zhai learn interesting aspects of queries by analyzing the clicked URLs directly from user click-through logs to organize search results. However, this method has limitations since the number of different clicked URLs of a query maybe small. Other works analyze the search results returned by the search engine when a query is submitted. Since user feedback is not considered, many noisy search results that are not clicked by any users may be analyzed as well.

**EXISTING SYSTEM**

- Useful knowledge discovery from Web usage data and satisfactory knowledge representation for effective Web-page recommendations are crucial and challenging.
- Existing system provide method to efficiently provide better Web-page recommendation through semantic enhancement by integrating the domain and Web usage knowledge of a website. Two new models are proposed to represent the domain knowledge.
- The first model uses ontology to represent the domain knowledge. The second model uses one automatically generated semantic network to represent domain terms, Web-pages and the relations between them. Another new model, the conceptual prediction model, is proposed to automatically generate a semantic network of the semantic Web usage knowledge, which is the integration of domain knowledge and Web usage knowledge.
- A number of queries have been developed to query about these knowledge bases. Based on these queries, a set of recommendation strategies have been proposed to generate Web-page candidates. The recommendation results have been compared with the results obtained from an advanced existing Web Usage Mining (WUM)method.

**DISADVANTAGES OF EXISTING SYSTEM:**

- Existing recommendation systems are: cold-start, sparsely, overspecialization and domain-dependency.
- The performance of existing system depends on the sizes of training datasets. Thebigger the training dataset size is, predicted pagesare limited within the discovered Web access sequences.
- The domain ontology can be constructed manually byexperts, or by automatically learning models is need to design and implement the learning models which can only be done by professionals atthe beginning.

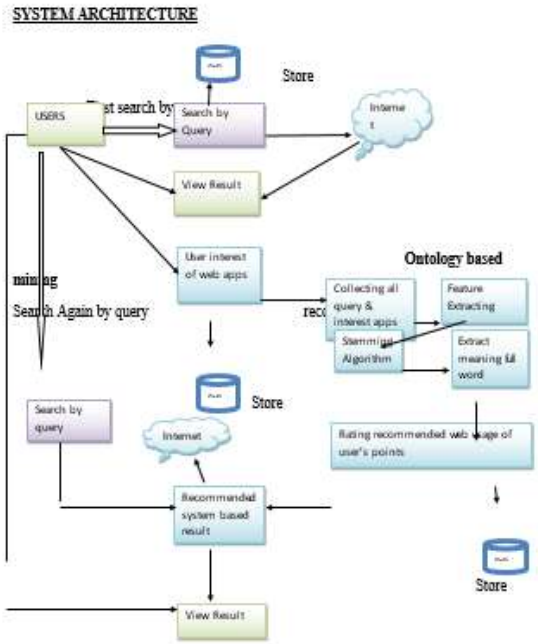
**PROPOSED SYSTEM**

- In proposed system present a personalized-recommendation system, a system that makes use of representations of items and user-profiles based on ontologies in order to provide semantic applications with personalized services.
- The semantics method achieved by using two different methods. A domain-based method makes inferences about user's interests and a taxonomy-based similarity method is used to refine the item-user matching algorithm, improving overall results. The recommender proposed is domain-independent, is implemented as a Web service, and uses both explicit and implicit feedback-collection methods to obtain information on user's interests.
- Proposed recommender system based on ontology and Web Usage Mining. The first step of the approach is extracting features from web documents and constructing relevant concepts. Then build ontology for the web site use the concepts and significant terms extracted from documents. According to the semantic similarity of web documents to cluster them into different semantic themes, the different themes imply different preferences.

**ADVANTAGES OF PROPOSED SYSTEM:**

Integrating domain knowledge with Web usage knowledge enhances the performance of recommender systems using ontology-based Web mining techniques.

- The construction of this model is semi-automated so that the development efforts from developers can be reduced.
- The user-profile learning algorithm, responsible for expanding and maintaining up-to-date the long-term user's interests, employs a domain-based inference method in combination with other relevance feedback methods to populate more quickly the user profile and therefore reduce the typical cold-start problem.
- The filtering algorithm, which follows a stemming approach, makes use of a semantic similarity method based on the hierarchical structure of the ontology to refine the item-user matching score calculation.




**DATAFLOWDIAGRAM:**

Data flow diagrams illustrate how data is processed by a system in terms of inputs and outputs. Data flow diagrams can be used to provide a clear representation of any business function. The technique starts with an overall picture of the business and continues by analyzing each of the functional areas of interest. This analysis can be carried out to precisely the level of detail required. The technique exploits a methodcalled top-down expansion to conduct the analysis in a targeted way.

As the name suggests, Data Flow Diagram (DFD) is an illustration that explicates the passage of information in a process. A DFD can be easily drawn using simple symbols. Additionally, complicated processes can be easily automated by creating DFDs using easy-to-use, free downloadable diagramming tools. A DFD is a model for constructing and analyzing information processes. DFD illustrates the flow of information in a process depending upon the inputs and outputs. A DFD can also be referred to as a Process Model. A DFD demonstrates business or technical process with the support of the outside data saved, plus the data flowing from the process to another and the end results.

There are some symbols that are used in the drawing of business process diagrams (data flow diagrams). These are now explained, together with the rules that apply to them.

**SYSTEM IMPLEMENTATION**

**MODULES**

1. Creating Search history
2. Query clustering
3. Query reformulation
4. History grouping

**1. CREATING SEARCH HISTORY**

Any personal documents such as browsing history and emails on a user's computer could be the data source for user profiles. This focus on frequent terms limits the dimensionality of the

document set, which further provides a clear description of users' interest. This module allows the search engine to better understand a user's session and potentially tailor that user's search experience according to her needs. Once query groups have been identified, search engines can have a good representation of the search context behind the current query using queries and clicks in the corresponding query group.

## 2. QUERY CLUSTERING

User's queries can be classified into different query clusters. Concept-based user profiles are employed in the clustering process to achieve personalization effect. The most similar pair of concept nodes, and then, merge the most similar pair of query nodes, and so on. Each individual query submitted by each user is treated as an individual node and each query with a user identifier. We perform the grouping in a similar dynamic fashion, whereby we first place the current query and clicks into a query group

## 3.QUERY REFORMULATION

To ensure that each query group contains closely related and relevant queries and clicks, it is important to have a suitable relevance between the current query groups. We assume that users generally issue very similar queries and clicks within a short period of time. The search history of a large number of users contains signals regarding query relevance, such as which queries tend to be issued closely together. This captures the relationship between queries frequently leading to clicks on similar URLs. Query reformulation graph and the query click graph from search logs, and how to use them to determine relevance between queries or query groups within a user's history.

## 4.HISTORY GROUPING

Query groups is to first treat every query in a user's history as a query group, and then merge these query groups in an iterative fashion (in a k-means). However, this is impractical in our scenario for two reasons. First, it may have the undesirable effect of changing a user's existing query groups, potentially undoing the user's own manual efforts in organizing her history. Second, it involves a high-computational cost, since we would have to repeat a large number of query group similarity computations for every new query.

## RESULTS



## CONCLUSION

In this paper, a novel approach has been proposed to infer user search goals for a query by clustering its feedback sessions represented by pseudo-documents. First, we introduce feedback sessions to be analyzed to infer user search goals rather than search results or clicked URLs. Both the clicked URLs and the unclicked ones before the last click are considered as user implicit feedbacks and taken into account to construct feedback sessions. Therefore, feedback sessions can reflect user information needs more efficiently. Second, we map feedback sessions to pseudo documents to approximate goal texts in user minds. The pseudo-documents can enrich the URLs with additional textual contents including the titles and snippets. Based on these pseudo-documents, user search goals can then be discovered and depicted with some keywords. Finally, a new criterion Stemming is formulated to evaluate the performance of user search goal inference. Experimental results on user click-through logs from a commercial search engine demonstrate the effectiveness of our proposed methods.

The complexity of our approach is low and our approach can be used in reality easily. For each query, the running time depends on the number of feedback sessions. However, the dimension of Ffs in (3) and (5) is not very high. Therefore, the running time is usually short. In reality, our approach can discover user search goals for some popular queries offline at first. Then, when users submit one of the queries, the search engine can return the results that are categorized into different groups according to user search goals online. Thus, users can find what they want conveniently.

## 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] G. Stumme, A. Hotho, and B. Berendt, "Usage Mining for and on the Semantic Web," AAAI/MIT Press, 2004, pp. 461-480.

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

[5] S. A. Rios and J. D. Velasquez, "Semantic Web Usage Mining by a Concept-Based Approach for Off-line Web Site Enhancements," in Web Intelligence and Intelligent Agent Technology, 2008. WI-IAT '08. IEEE/WIC/ACM International Conference on, 2008, pp. 234- 241.

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

[7] A. Bose, K. Beemanapalli, J. Srivastava, and S. Sahar, "Incorporating Concept Hierarchies into Usage Mining Based Recommendations," in Proceedings of the 8th Knowledge discovery on the web international conference on Advances in web mining and web usage analysis Philadelphia, PA, USA: Springer-Verlag, 2007, pp. 110-126.

[8] N. R. Mabroukeh and C. I. Ezeife, "Semantic-Rich Markov Models for Web Prefetching," in 2009 IEEE International Conference on Data Mining Workshops Miami, Florida, USA, 2009, pp. 465-470.