

# Social Networks Conceptual Clarification Using Mining Pattern

M.Kirithika Devi<sup>1</sup>, Dr. R.Kavitha<sup>2</sup>, S.Aishwarya<sup>3</sup>

Assistant Professor, Department of Computer Science and Engineering, Chendhuran College of Engineering and Technology, Lena Vilakku, Pilivalam (po), Pudukkottai, Tamil Nadu, India<sup>1</sup>

Professor, Department of Computer Science and Engineering, Velammal College of Engineering and Technology, Madurai to Rameshwaram High Road, Viraganoor, Madurai, Tamil Nadu, India<sup>2</sup>

PG Student, Department of Computer Science and Engineering, Chendhuran College of Engineering and Technology, Lena Vilakku, Pilivalam (po), Pudukkottai, Tamil Nadu, India<sup>3</sup>

**ABSTRACT:** 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. We explain this as an outcome resulted from decentralized knowledge contributions from individuals who maximize their own utilities. Our model sheds insights on how knowledge creation, knowledge citation, and knowledge heterogeneity affect the emergence of knowledge core, in both cases of direct and indirect citations. 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.

**KEYWORDS:** Query, document ranking, web context, microblogs, Google, Twitter.

## I. INTRODUCTION

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 sunl 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 goal Nscan 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 may be 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.

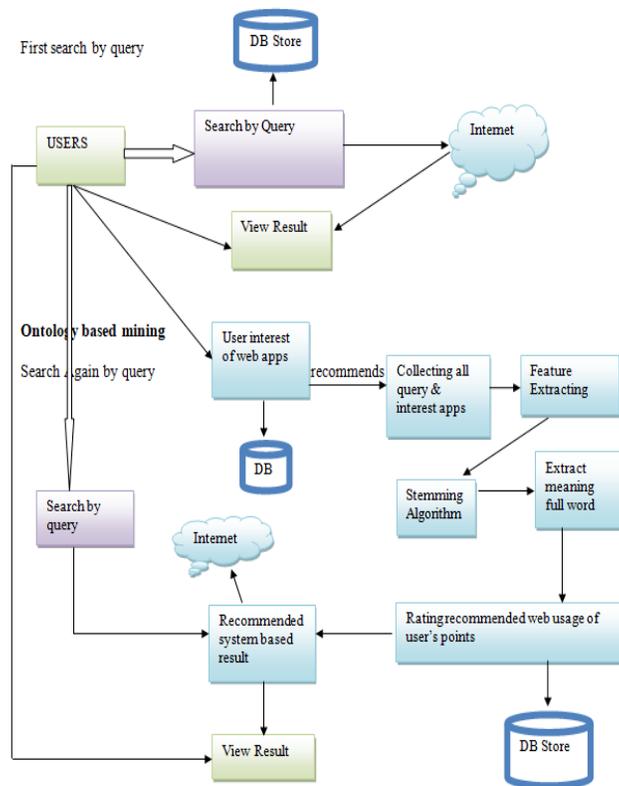


Fig.1 Proposed System Architecture

### A. Existing System Summary

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. 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.
- The bigger the training dataset size is, predicted pages are limited within the discovered Web access sequences.
- The domain ontology can be constructed manually by experts, or by automatically learning models is need to design and implement the learning models which can only be done by professionals at the beginning.

## II. RELATED WORKS

Document ranking is very important in information retrieval (IR) fields, and many studies have been processed. There are three models of general-purpose document ranking approaches: vector space model, probabilistic model, and network model [1][4]. Salton et al. introduced a vector space model [2]. The vector space model considers a document as a vector which contains both magnitude and direction. When considering each word in a document collection as one axis, each document can be represented with a vector point in the vector space. By comparing the distance or angle between two documents (or between a query and a document), it is possible to calculate how the documents are similar to queries. Robertson suggested a probabilistic model [3]. In their study, IR systems estimate the probability that a document is relevant to a query. This model seems to be similar with the vector space model but it is different from the vector space model in that it ranks documents based on the probabilistic relationships between a query and a document.

Page et al. suggested a link analysis algorithm, which is known as PageRank in Google search engine [4]. In the approach, each document is assigned a numerical weight with a hyperlinked set in the document. The documents which have more weights are considered as more important, so search engines can place such documents in higher positions. Those general-purpose approaches have focused on how to retrieve relevant documents to queries.

In addition to those general-purpose approaches, some of recent studies on document ranking have been concentrated in classifying or clustering queries according to their characteristics [9][12]. In these approaches, they employed various features to extract more relevant results. The authors argued that providing different ranking results according to the characteristic of queries leads a better satisfaction. The considered features were query logs [9], [10], query term distributions and mutual information [11], and geographical locality [12], and so on. However, most of these studies classified queries just using the static information of queries, and did not consider the context of queries which is one of dynamic information related to queries.

Microblogs also can be used for portal web search engines which should support a variety of search needs, including real-time web search. Becker et al. used Twitter data to address two major challenges in addition to basic web search problems [11]. They dealt with how to quickly crawl the relevant content and how to rank documents with impoverished links and click information. In their approach, they used Twitter to and fresh URLs for queries, and ranked the fresh URLs by computing some proposing features, textual features, social network features and others. However, they did not extract the context from microblogs for adaptive search result to users' interests.

Apart from those studies, there are several attempts to use microblogs on various IR problems. Becker et al. proposed the real-world event identification [13], and Beitzel et al. recommended real-time topical news by using Twitter [14]. Those

approaches focused on the services sensitive to the context of queries, but it is hardly found applying context for adaptive document retrieving.

### III. PROPOSED SYSTEM SUMMARY

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*

- 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.

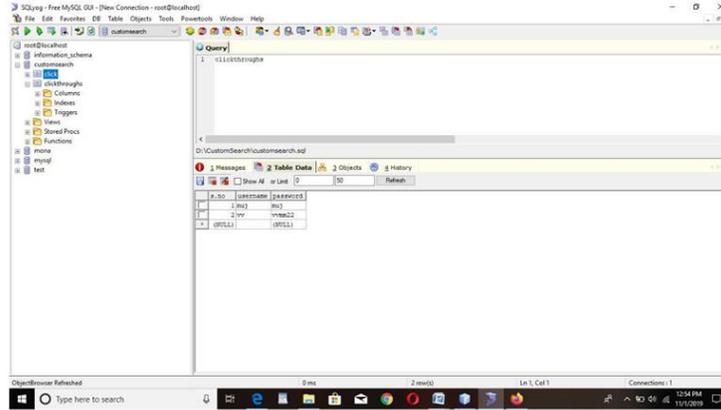
### IV. RESULTS AND DISCUSSION

The following figure, Fig.2 illustrates the Login Page view of the proposed system.



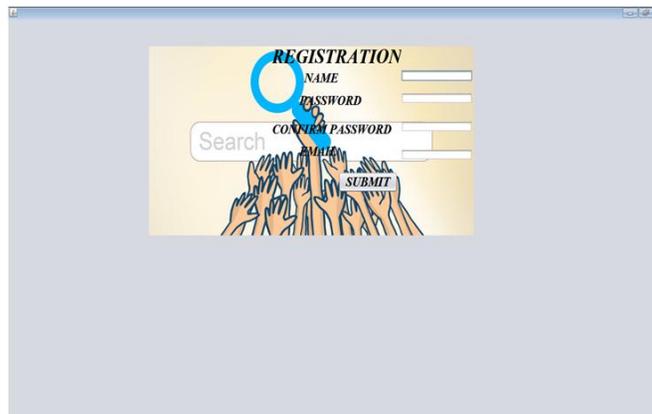
Fig.2 Login Page

The following figure, Fig.3 illustrates the Database view of the proposed system.



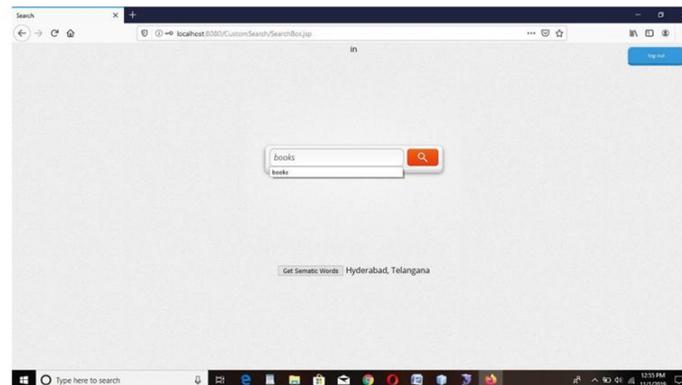
**Fig.3 Database View**

The following figure, Fig.4 illustrates the User Registration Page view of the proposed system.



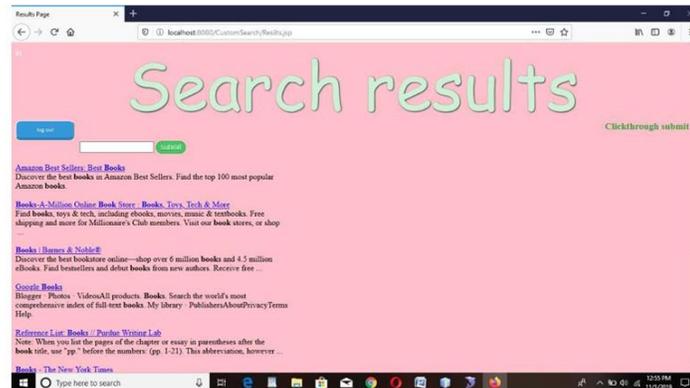
**Fig.4 User Registration**

The following figure, Fig.5 illustrates the Customer Search view of the proposed system.



**Fig.5 Customer Search View**

The following figure, Fig.6 shows the Search result view of the proposed system.



**Fig.6 Customer Search Result View**

## V. CONCLUSION AND FUTURE SCOPE

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 CAP 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.

Future work will be how system can avoid this kind of problems by considering uncommon sources, and perform with web information mining more elaborately. Since our approach is based on the current context of queries, we performed the survey as soon as possible after the web documents were crawled.

## 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.
- [9] M. O'Mahony, N. Hurley, N. Kushmerick, and G. Silvestre, "Collaborative recommendation: A robustness analysis," *ACM Transactions on Internet Technology*, vol. 4, pp. 344-377, 2004.
- [10] G. Stumme, A. Hotho, and B. Berendt, "Semantic Web Mining: State of the art and future directions," *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 4, pp. 124-143, 2006.
- [11] B. Zhou, S. C. Hui, and A. C. M. Fong, "CS-Mine: An Efficient WAP-Tree Mining for Web Access Patterns," in *Advanced Web Technologies and Applications*. vol. 3007: Springer Berlin /Heidelberg, 2004, pp. 523-532.
- [12] J. Borges and M. Levene, "Generating Dynamic Higher-Order Markov Models in Web Usage Mining," in *Knowledge Discovery in Databases: PKDD 2005*. vol. 3721: Springer Berlin / Heidelberg, 2005, pp. 34-45.
- [13] C. I. Ezeife and Y. Lu, "Mining Web Log Sequential Patterns with Position Coded Pre-Order Linked WAP-Tree," *Data Mining and Knowledge Discovery*, vol. 10, pp. 5-38, 2005.
- [14] B. Zhou, "Intelligent Web Usage Mining," Nanyang Technological University 2004.
- [15] C. Ezeife and Y. Liu, "Fast Incremental Mining of Web Sequential Patterns with PLWAP Tree," *Data Mining and Knowledge Discovery*, vol. 19, pp. 376-416, 2009.
- [16] T. T. S. Nguyen, H. Lu, T. P. Tran, and J. Lu, "Investigation of Sequential Pattern Mining Techniques for Web Recommendation," *International Journal of Information and Decision Sciences (IJIDS)*, pp. x-x, 2012.
- [17] S. T. T. Nguyen, "Efficient Web Usage Mining Process for Sequential Patterns," in *Proceedings of the 11th International Conference on Information Integration and Web-based Applications & Services*, Kuala Lumpur, Malaysia 2009, pp. 465-469.
- [18] L. Wei and S. Lei, "Integrated Recommender Systems Based on Ontology and Usage Mining," in *Active Media Technology*. vol. 5820, J. Liu, J. Wu, Y. Yao, and T. Nishida, Eds.: Springer-Verlag Berlin Heidelberg, 2009, pp. 114-125.
- [19] A. Loizou and S. Dasmahapatra, "Recommender Systems for the Semantic Web," in *ECAI 2006 Recommender Systems Workshop Trento, Italy, 2006*.
- [20] D. Dzemydiene and L. Tankeleviciene, "On the Development of Domain Ontology for Distance Learning Course," in *20th EUROMini Conference "Continuous Optimization and Knowledge-Based Technologies"* Neringa, LITHUANIA, 2008, pp. 474-479.
- [21] S. Boyce and C. Pahl, "Developing Domain Ontologies for Course Content," *Educational Technology & Society*, vol. 10, pp. 275-288, 2007.
- [22] J. M. Gascuena, A. Fernandez-Caballero, and P. Gonzalez, "Domain Ontology for Personalized E-Learning in Educational Systems," in *Proceedings of the Sixth IEEE International Conference on Advanced Learning Technologies*, 2006, pp. 456-458.
- [23] Y. Yalan, Z. Jinlong, and Y. Mi, "Ontology Modeling for Contract: Using OWL to Express Semantic Relations," in *Enterprise Distributed Object Computing Conference, 2006. EDOC '06. 10th IEEE International*, 2006, pp. 409-412.
- [24] D. Oberle, S. Grimm, and S. Staab, "An Ontology for Software," in *Handbook on Ontologies*. vol. 2, S. Staab and R. Studer, Eds. Berlin, Heidelberg: Springer, 2009, pp. 383-402.
- [25] M. Eirinaki, D. Mavroidis, G. Tsatsaronis, and M. Vazirgiannis, "Introducing Semantics in Web Personalization: The Role of Ontologies," in *Semantics, Web, and Mining*, M. A. e. al., Ed.: Springer-Verlag Berlin Heidelberg, 2006, pp. 147-162.
- [26] G. Antoniou and F. v. Harmelen, *A Semantic Web Primer*: MIT Press, 2008.
- [27] A. Harth, M. Janik, and S. Staab, "Semantic Web Architecture," in *Handbook of Semantic Web Technologies*, J. Domingue, D. Fensel, and J. A. Hendler, Eds.: Springer-Verlag Berlin Heidelberg, 2011, pp. 43-75.
- [28] S. Grimm, A. Abecker, J. Völker, and R. Studer, "Ontologies and the Semantic Web," in *Handbook of Semantic Web Technologies*, J. Domingue, D. Fensel, and J. A. Hendler, Eds.: Springer-Verlag Berlin Heidelberg, 2011, pp. 507-580.



- [29] B. Liu, "Information Retrieval and Web Search," in Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data, B. Liu, Ed.: Springer-Verlag Berlin Heidelberg, 2011, pp. 183-236.
- [30] Z. Markov and D. T. Larose, "Chapter 1: Information Retrieval and Web Search," in Data Mining the Web: Uncovering Patterns in Web Content, Structure, and Usage, Z. Markov and D. T. Larose, Eds. New Britain: John Wiley & Sons, Inc, 2007, pp. 3-46.
- [31] M. Uschold and M. Gruninger, "Ontologies: Principles, Methods and Applications," Knowledge Engineering Review, vol. 11, pp. 93-36, 1996.
- [32] J. Borges and M. Levene, "A Dynamic Clustering-Based Markov Model for Web Usage Mining," Available online at <http://xxx.arxiv.org/abs/cs.IR/0406032>, Technical Report May 26 2004.