International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2015): 6.391

Personalized and Effective Web Page Recommendation Search System based on Ontology and Semantic Enhancement

Hurma Begum[1], B. Sasidhar[2]

[1]M.Tech Student, Dept. of Computer Science & Engineering, Mahaveer Institute of Science & Technology, Telangana, Hyderabad, India

[2]Professor, Dept. of Computer Science & Engineering, Mahaveer Institute of Science & Technology, Telangana, Hyderabad, India

Abstract: ***The project proposes a personalized-recommendation search system, a system that makes use of representations of items and user-profiles based on ontologies and semantic enhancement in order to provide semantic applications with personalized services. It provides a better personalized recommendation search system by integrating domain knowledge and web usage knowledge. It uses two methods ontology and semantic enhancement. Ontology approach is written in logical based language. Here the input data is weblogs that record user sessions on a daily basis. The user session include information about users, webpage navigation activities (web snippets). Based on these facts, we aim to discover domain knowledge from the titles of visited web pages at a website and represent the discovered knowledge in domain ontology to support effective recommendation systems. Metaphysics technique is semi-automated in order that the event efforts from the developers are reduced. Second, Semantic method is achieved by using two different methods. A domain based mostly that produces illation regarding users interests, and a Taxonomy based mostly similarity want to refine item-user matching algorithmic rule up overall results. Propose recommender system also creates User profiles based on user preferences or past search histories. Password security is provided by Double Caesar cipher encryption method. The Propose system also provides the location from where the user has searched. This Project provides better personalized recommendation search system which is secure and effective.***

Keywords: Ontology, Web-snippets, Recommendation-systems, Google, Web Mining

1. Introduction

In net search applications, queries are submitted to go looking engines to represent the knowledge desires of users. However, usually queries may not exactly represent users specific information needs since many ambiguous queries would possibly crawl a broad topic and fully completely different users may want to get information on different aspects when they submit the same query. For example, when the query “the Jaguar” is submitted to a search engine, some users want to locate the Jaguar Car company, while some others want to learn the information (knowledge about the Jaguar the Cat). Therefore, it is necessary and potential to capture fully completely different user search goals in information retrieval [4]. The process tend to outline user search goals because the info on completely different aspects of a question that user teams need to get. To get Information need is a user’s specific desire to obtain information to satisfy his/her needs /wants[2].

User search goals are thought-about because the clusters of knowledge desires for a question. The illation and analysis of user search goals will have a great deal of benefits in up program conation and user expertise. Some benefits are summarized as follows. First, we are able to reconstitute net search results consistent with user search goals by grouping the search results with identical search goal; therefore, users with completely different search goals can easily find what they want. Second, user search goals delineate by some keywords is employed in question recommendation thus; the urged queries can facilitate users to make their queries additional exactly. Third, the distributions of user search goals may also be helpful in applications resembling re-ranking net search results that contain completely different user search goals [5] ,[6], [9].

Data mining :The actual data mining task is the analysis of large quantities of data to extract previously unknown, interesting patterns such as groups of data records (cluster analysis), unusual records (Anamoly detection)and dependencies (association rule mining) automatically or semi automatically.

Data mining involves six common classes of tasks:

- Anamoly Detection (Outlier/change/deviation detection) – The identification of unusual data records, that might be interesting or data errors that require further investigation.
- Association Rule Mining (Dependency modelling) – Searches for relationships between variables.
- Clustering – is the task of discovering groups and structures in the data that are in some way or another "similar", without using known structures in the data.
- Classification – is the task of generalizing known structure to apply to new data. For example, an e-mail program might attempt to classify an e-mail as "legitimate" or as "spam".
- Regression – attempts to find a function which models the data with the least error.
- Summarization – providing a more compact representation of the data set, including visualization and report generation.

The abundance of information available on web & digital libraries, in combination with their dynamic nature has fast

increasing difference in searching what users want whenever he/she needs it and in a way which meet our requirements. As a result, the role of user-modelling and personalizing information access is becoming crucial. User needs a Personalized support in moving by using large amounts of available information access to their interests and tasks. Many data sources have Recommendation systems as a way of personalizing their contents for users. Recommendation systems guides Users in a personalized way to help find their products items etc that may be interested in based on what is known about user in their profile. Recommendation systems take input about Users interests to get a list of recommend items.

Personalization refers adapting to the ones requirements, interests and preferences/Priorities of each User. It includes:

- Recommending (items)
- Filtering (information)
- Predicting (behavior)

2. Literature Survey

Here shows the various analyses and research made in the field of our interest and the results already published, taking into account the various parameters of the project .Thus the following Research papers are referred to get the knowledge or idea for the propose system we are developing. All the following papers are gathered from different sources , studied in detail and their descriptions are as follows:

a) Search Engine Personalization: Efficient Query Processing in Geographic Web Search Engines:
In this paper given by Yen-Yu Chen, TorstenSuel, Alexander Markowetz[16] , we've got a bent to review the matter of economical question method in ascendable geographic search engines. Question method can be a serious bottleneck in customary web search engines, and additionally the most reason for the thousands of machines utilized by the most engines. Geographic program question method is totally different during this it desires a combination of text and special process techniques. They propose several algorithms for economical question method in geographic search engines,- 68integrate them into Associate in nursing existing web search question processor, and choose them on huge sets of real data and question traces.

b) Mining Preference Using Spy Voting for Search Engine Personalization:
This paper given by Wilfred Ng Lin Deng and DikLun Lee addresses program personalization. we have a tendency to tend to gift a greenhorn approach to mining user's preferences on the search results from click through info and victimization the discovered preferences to adapt the search engine's ranking perform for up search quality. we have a tendency to tend to develop a greenhorn preference mining technique called SpyNB(spy NAIVE bayes)[13], that depends on the wise assumption that the search results clicked on by the user reject the user's preferences but it does not draw any conclusions regarding the results that the user did not click on.

c) Applying Co-training to Click through Data for Search Engine Adaptation:
The title Applying Co- training to click-through data for search engine adaptation presented by Qingzhao Tan, Xiao young Chai, Wilfred Ng and DikLun Lee[11] proposed a replacement algorithm, Ranking SVM during a} very Co-training Framework (RSCF). basically, the RSCF algorithm takes the click through information containing the items among the search result that are clicked on by a user as Associate in Nursing input, Associate in Nursing generates accommodative rankers as associate degree output. By analyzing the click through information, RSCF rest categorizes the information knowledge the information the data as a result of the labeled dataset that contains the items that are scanned already, and conjointly the un-labelled information set, that contains the items that haven't still been scanned. The labeled information is then augmented with un-labelled information to induce larger information set for employment the rankers.

d) Privacy-Enhancing Personalized Web Search:
In the year 2007 , Yabo Xu, Benyu Zhang, Zheng Chen and Ke Wang presented a climbable manner for users to automatically build affluent user profiles. These profiles summarize user's interests into a hierarchic organization in line with specific interests. Two parameters for specifying privacy requirements Pine Tree State sure unit of measurement unit projected to help the user to decide on the self-satisfied degree of detail of the profile data that is exposed to the program me. Experiments showed that the user profile improved search quality once place next to plain MSN(Microsoft Network) rankings[15].

e) Personalized Concept-Based Clustering of Search Engine Queries
A paper by Kenneth Wai-Ting Leung, Wilfred Ng, and DikLun Lee[8] , in the year November 2008 introduced an honest approach that captures the user's abstract preferences therefore on offer customized question suggestions. We've a bent to deliver the products this goal with two new ways. First, we've a bent to develop on-line techniques that extract ideas from the web-snippets of the search result came from a matter and use the ideas to identify connected queries for that question. Second, we've a bent to propose a greenhorn two section customized collective clump rule that's ready to generate customized question clusters.

f) Personalized Web Search with Location Preferences:
Kenneth Wai-Ting Leung ,DikLun Lee , Wang-Chien Lee proposed a greenhorn net search personalization approach that captures the user's interests and preferences among the sort of ideas by mining search results and their click through. as a result of the important role location knowledge plays in mobile search, we've a bent to separate ideas into content ideas and placement ideas, Associate in Nursing organize them into ontology's to create associate degree ontology-based, multi-facet (OMF) pro let precisely capture the user's content and placement interests and so improve the search accuracy[7]. Moreover, recognizing the particular incontrovertible fact that whole different totally different totally different completely different completely different fully different users and queries might have different

emphases on content and placement knowledge, we've a bent to introduce the notion of content and placement entropies to measure the amount of content and placement knowledge associated with a matter , and click on content and placement entropies to measure what proportion the user is fascinated by the content and placement knowledge among the result.

3. Existing System

In absolute there are no re baronial of abstracts and no aegis for the user seek result. The Search results with recommendations will be resultant of thousands of users. Here in existing systems like google and Yahoo search engines what the User searched (everything including YouTube searched history watched history and all the histories of the google account) will be visible to the google industry and the search is not personalized. Recommendation is based on billions of users interests. However, users have concerns about how their personal information is used. Privacy, as opposition security or confidentiality, extremely depends on the person concerned and the way that person might have the benefit of sharing personal info. The question here is whether or not an answer is found wherever users themselves are able to set their own privacy levels for user profiles to boost the search quality.

3.1 Limitations

- Existing recommendation systems are: cold-start, over-specialization and Domain-dependent.
- The performance of existing system depends on the sizes of coaching datasets. The bigger the coaching dataset size is, foretold pages are restricted inside the discovered net access sequences.
- These existing systems are less secure as what the users searched (all Searched History)are shown(visible) to everyone.

4. Our Approach

In proposed system, we presents a personalized-recommendation search system, a system that makes use of representations of items and user-profiles supported ontologies so as to produce linguistics applications with personalized services. projected system uses 2 techniques. The semantics method achieved by using two different methods. A domain-based technique makes inferences regarding user's interests and a taxonomy-based similarity technique is employed to refine the item-user matching algorithmic rule, improving overall results. This method can be easily incorporated into a webpage recommendation process because of its fully automated .Proposed recommender system supported metaphysics and net Usage Mining. The primary step of the approach is extracting options from net documents and constructing relevant ideas. Then build metaphysics for the net web site, use the ideas and important terms extracted from documents. From documents According to the linguistics similarity of net documents to cluster them into completely different linguistics themes. This is semi-automated so that development efforts from developers are reduced. Creates user-profiles based on user preferences or past search history. Provides location from where the user has searched. Thus we are proposing a personalized recommendation search system (engine) which is secure. Password encryption is done by double ceasor cipher method. Propose system also provide authentication so it is secure and efficient.

4.1 Advantages

- Group action domain data with net usage data enhances the performance of recommender systems mistreatment ontology-based net mining techniques.
- This model is semi-automated so as that the event efforts from developers are reduced. Propose search engine (system) is secure and efficient as it is personal and secure.
- The filtering algorithmic rule, that follows a stemming approach, makes use of a linguistics similarity technique supported the data structure of the metaphysics to refine the item-user matching score calculation.
- This proposes personalized web search recommendation system can be used by more than one person as it gives authentication. So it is secure.

The below figure shows the Block diagram for the propose system i.e., personalized mobile search system. Diagram shows how the client sends the request to the server, the Glassfish server process the request by sending to all of its components at its side. When the Client enters the query, Server will first check in DB, if requested query is available then it search using mined word, recommend and return to the client. If query is not present then server will search using query only and then submit to the client. Fig also shows semantics that gives the related meanings of the query.

4.2 Explanation

MINEDWORD: Mined Word association rules between words for measuring the similarity between documents to enhance the text retrieval. For each document relevant to a query it formed a group of documents having atleast one common frequent set of words with the answer document. Then we measured precision of documents in same group as answer set to corresponding query. Thus frequent sets of words mined from the test Database are useful in ranking of query result sets to improve the precision of retrieval.

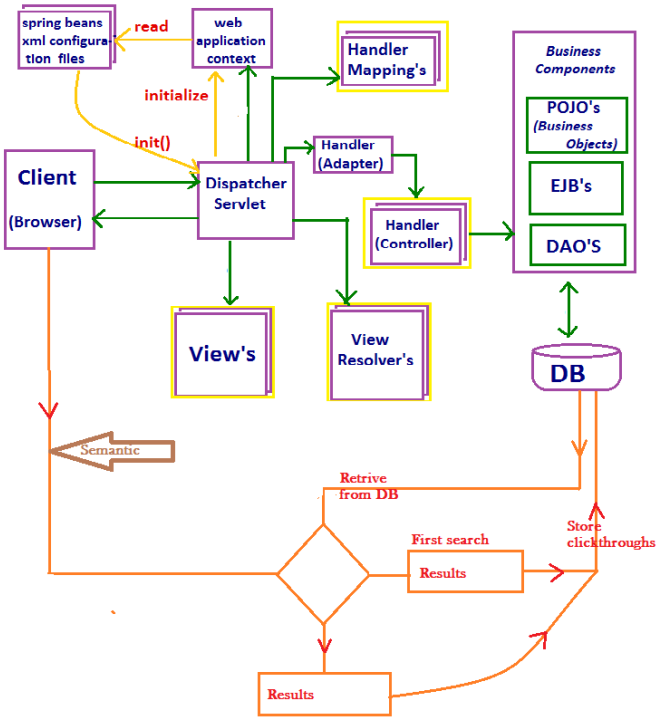

Figure 1: Block Diagram of PMSE (Personalized Mobile Search Engine)

4.3 Phases

a) Click through assortment at PMSE shopper
b) Re-ranking the search results at PMSE
c) User Interest identification
d) Diversity and idea Entropy
e) Semantics and Submit click-through

a) Click Through Assortment At PMSE Shopper:
Ontology can be derived online at the personalized mobile search engine (PMSE) server, another system style is for the user to pass solely the press through knowledge to the PMSE server, and to perform each feature extraction and RSVM coaching on the PMSE server to train the weight vectors for reran king. However, if all click-troughs are exposed to the PMSE server, the server would recognize precisely what the user has clicked. To handle privacy problems, click-through are keep on the PMSE consumer, and therefore the user might modify the privacy parameters to manage the number of private info to be enclosed within the feature vectors, that are forwarded to the PMSE server for RSVM coaching to adapt personalized ranking functions for content and site preferences.

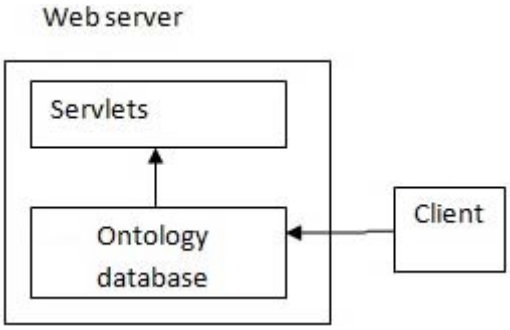

Figure 2: Click-through Collection at Client

b) Re-Ranking The Search Results At PMSE:
The second module Re-ranking the search results at Personalized Mobile search engine (PMSE) server, once a user submits a question on the PMSE consumer, the question forwarded to the PMSE server .It obtains the search results from the back-end search engine .The content and site ideas are extracted from the search results and arranged into metaphysics to capture the relationships between the ideas. The search results are then re-ranked consistent with the load vectors obtained from the RSVM (Reduced Support vector Machines) coaching. Finally, the re-ranked results and therefore the extracted metaphysics for the personalization of future queries are come to the consumer.

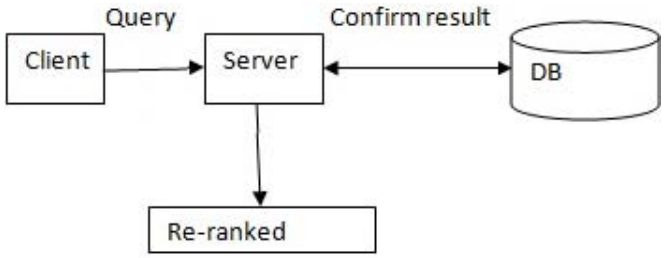

Figure 3: Re-ranking module at PMSE server.

c) Creating User Profiles/User Identification:
This is the third module which is User Interest Profiling where PMSE uses "concepts" to model the interests and preferences of a user. The ideas are any classified into 2 differing kinds, namely, content ideas and site ideas. The ontologism indicates a potential construct area arising from a user's queries that are maintained together with the press through knowledge for future preference adaptation.

The location ontology/location concept is used to associate location information with the search results. To adapt user mobility, we incorporated PS location in the personalization framework.

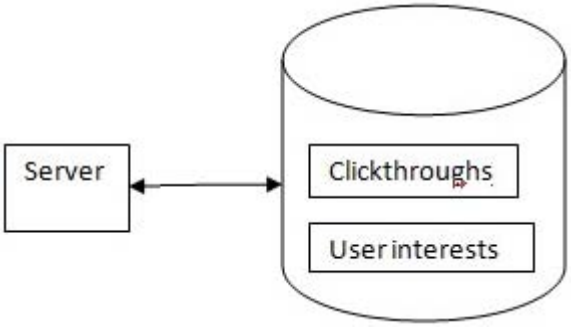

Figure 4: Creating user Profile Module.

d) Diversity and Idea Entropy
This is the module where the useful and meaningful information the requested query obtained. PMSE consists of a content aspect and a location aspect. In order to seamlessly integrate the preferences in these 2 aspects into one coherent personalization framework. In this, weights of content preference and site preference supported their effectiveness within the personalization method. The notion of personalization effectiveness is derived based on the diversity of the content and location information in the search results.

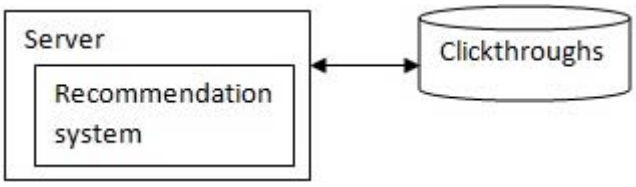

Figure 5: Extract meaningful and useful Information.

f) Semantics and Submit click-through:

A user can search the web with exactly the same word entered or they can use semantics for similar meaning of that query. The semantic of all words will be stored on server hard drive and will be retrieved when ever user wants similar meaning. When the results show up, the user can select any web site and he can submit click through for only web site he wants recommendations for.

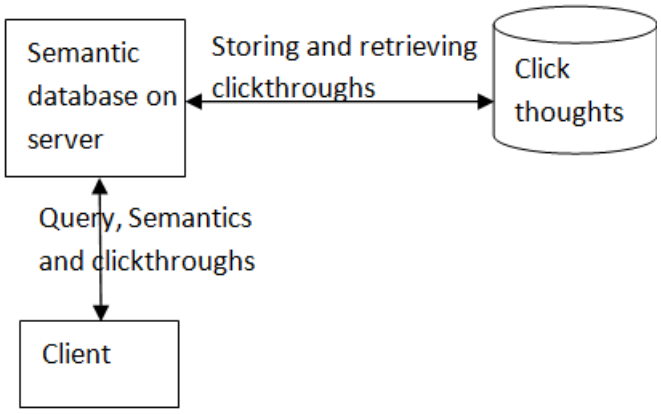

Figure 6: Click-through submission module.

5. Results & Discussions

Here, in this chapter we obtained the results after performed methodology approach. Here experimental study of the proposed system is obtained .In existing system search is based on millions of users but in our propose system search is personalized .In table 01 shows the experimental test case study at client and server side.

Table 1: Experimental Test

Test Case	Check Field	Objective	Expected Result
TC-001	Client	Not open the application	Should check s/w installation
TC-002	Client	User does not exist	Register
TC-003	Client	Wrong user name and password	Should Enter correct user name and password
TC-004	Server	Read location	should enable GPS
TC-005	Server	No Provide services using classifier	Should check algorithm
TC-006	Client	Now search any data	Should bring professional related data only

6. Conclusion

We propose a web page Recommendation Search System for recommending user click-through by learning the user preferences (Past Search Histories) .We discussed a framework for ranking the Search results and recommending it. We also incorporated location for user mobility. We have also added semantics to our framework so it will give the related meanings of a particular query. For Password encryption we used Double ceasor cipher method for security.so the proposed system is secure as it gives security and provide authentication. Experimental results show that project considerably improves the accuracy scrutiny to the baseline.

References

[1] D. Xing, G.-R. Xue, Q. Yang, and Y. Yu, "Deep Classifier: Automatically Categorizing Search Results into Large-Scale Hierarchies," Proc. Int'l Conf. Web Search and Data Mining (WSDM), pp. 139-148, 2008.R. Caves, Multinational Enterprise and Economic Analysis, Cambridge University Press, Cambridge, 1982. (book style)

[2] E. Agichtein, E. Brill, S. Dumais, and R. Ragno, "Learning User Interaction Models for Predicting Web Search Result Preferences," Proc. Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR), 2006

[3] F. Qiu, and J. Cho. Automatic identification of user interest for personalized search. In Proc. of the 12th International World Wide Web Conference (WWW), Edinburgh, Scotland, May 2006

[4] G. Chen, H. Bai, L. Shou, K. Chen, and Y. Gao, "Ups: Efficient Privacy Protection in Personalized Web Search," Proc. 34th Int'l ACM SIGIR Conf. Research and Development in Information, pp. 615-624, 2011.

[5] Joachim's, T.: Optimizing search engines using click-through data. In: Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining, Edmonton, Alberta, Canada (2002) 133–142.

[6] J. Sun, H. Zeng, H. Liu, Y. Lu, and Z. Chen. Cube SVD: A Novel Approach to Personalized Web Search. In Proc. of the 14th International World Wide Web Conference (WWW), Chiba, Japan, May 2005.

[7] Kennith wai-Ting leung, Diklun Lee,Wang Chien Lee, Personalized Web Search with location Preferences,in proceedings of the IEEE Transactions on knowledge and data Engineering , Vol. 20,No.11,Nov 2008.

[8] Kennith wai -Ting leung, Diklun Lee, Wang Chein Lee, Personalized concept based clustering of Search engine queries in proceedings of the IEEE Transactions on knowledge and data Engineering ,Vol. 20,No.11,Nov 2008.

[9] K. Sugiyama, K. Hatano and M. Yoshikawa. Adaptive Web search based on user profile constructed without any effort from users, In Proc. of the 13th International World Wide Web Conference (WWW), New York, New York, May 2004.

[10] L. Deng, W. Ng, X. Chai, and D.L. Lee, "Spying Out Accurate User Preferences for Search Engine Adaptation," Advances in Web Mining and Web Usage Analysis, LNCS 3932, pp. 87-103, 2006.

[11] Qing zhao tan ,Xiao young Chai ,Wilfred Ng , diklun Lee, Applying Co-training to Click through Data for Search Engine Adaptation in Proc.Of the Research Grant

Council of Hong kongGrant no HKUST6079/01E , DAG 01/02.EG05 and HKUST6185/02E.

[12] V.W. Chan, K.W. Leung, and D.L. Lee, "Clustering Search Engine Query Log Containing Noisy Click-throughs," Proc. Sump. Applications and the Internet (SAINT), 2004.

[13] Wilfred Ng , Lindeng , Diklun Lee , Mining Preference Using Spy Voting for Search Engine Personalization , in proc. ACM Transactions on Internet Technologies , Vol.7 , No.3.

[14] W. Ng, L. Deng, and D.L. Lee, "Mining User Preference Using Spy Voting for Search Engine Personalization," ACM Trans. Internet Technology, vol. 7, no. 4, article 19, 2007.

[15] Yabo Xu , Benyu Zhang , Zheng Chen , Ke Wang, Privacy Enhanced personalized web search , in Proceeding WWW 2007/Track:search Session: Personalization.

[16] Yen-yu chen , Torsten suel , Alexander Markowetz, Search Engine Personalization: Efficient Query Processing in Geographic Web Search Engines: in Proc. SIGMOD 2006 ,June 27-29 , Chicago , Illinois ,USA .Copyright 2006 ACM 1-59593-256-9/06/0006….$500.

[17] Y. Xu, B. Zhang, Z. Chen, and K. Wang, "Privacy-Enhancing Personalized Web Search," Proc. 16th Int'l World Wide Web Conf. (WWW), 2007.

[18] Z. Zhang and O. Nasraoui, "Mining Search Engine Query Logs for Query Recommendation," Proc. 15th Int'l World Wide Web Conf. (WWW), 2006.

[19] Z. Dou, R. Song, and J.-R. Wen, "A Large-Scale Evaluation and Analysis of Personalized Search Strategies," Proc. Int'l Conf. World Wide Web (WWW), pp. 581-590, 2007.