

Advanced Data Searching System with Spatial Database in PWS Framework

¹G.Sangeetha, ²G.Anishya,
¹Assistant Professor, ²Student M.E CSE
^{1,2} Department of Computer Science and Engineering
^{1,2} Valliammai Engineering College

Abstract - In order to improve the search quality of various web services Personalised Web Search has developed. The PWS framework can effectively generalize profiles by queries while respecting user specified requirements. The Content based ranking algorithm that automatically ranks the search engine results based on term frequency. Based on ranking technique, the required information is retrieved by the users. To enhance the stability of personalization the Advanced data search Technique is used in the PWS process. This system maintains a spatial database that provides faster access based on different selection criteria and to extract appropriate results. Then linkability assessment is done on search engine results to verify the results generated are linked with the user profile and ensures that the results are accurate thus satisfying all the users

Keywords - Personalized web search(PWS), Spatial database, Linkability Assessment, Range Search, Linkability metric

I. INTRODUCTION

The web search engine has long become the most important portal for ordinary people looking for useful information on the web. However, users might experience failure when search engines return irrelevant results that do not meet their real intentions. Such irrelevance is largely due to the enormous variety of users' contexts and backgrounds, as well as the ambiguity of texts and mismatching of keywords. To improve the search quality of various web services Personalised Web Search has developed Personalized web search (PWS) is a general search technique provides better search results, for individual user needs. PWS can generally be categorized into two types, namely click-log-based methods and profile-based ones. The click-log based methods provide results based on clicking pages on user query. The click-log based methods provide results based on clicking pages on user query. The click based strategy can only work on repeated queries from the same user. It does not work on hierarchy of users and hence irrelevant results are produced. The profile based method provides search results based on user personal information and their interests. The search engine generates a set of results comparing the user query and their profile. Hence appropriate results are provided satisfying all users. Hence in this profile based PWS is preferred over click based. The profile based PWS framework can effectively generalize profiles by queries while respecting user specified requirements. In order to improve the search engine results Content based ranking is used. The Content based ranking algorithm that automatically ranks the search engine results based on term frequency. Based on ranking technique, the required information is retrieved by the users. To enhance the stability of personalisation various methods are available. Here In order to enhance the accuracy of search results an advanced data searching system is used in PWS framework. The Advanced data searching system consists of a spatial database. The Spatial database stores many query objects. Range Searching Technique is used for extracting information from the spatial database. Linkability Assessment mechanism is used to evaluate if the user query is linkable to their user profile

RELATED WORKS

The web search engine has long become the most important portal for ordinary people looking for useful information on the web. However, users might experience failure when search engines return irrelevant results that do not meet their real intentions. Such irrelevance is largely due to the enormous variety of users' contexts and backgrounds, as well as the ambiguity of texts and mismatching of keywords. To improve the search quality of various web services Personalised Web Search has developed. Personalized web search (PWS) is a general search technique provides better search results, for individual user needs. PWS can generally be categorized into two types, namely click-log-based methods and profile-based ones. The click-log based methods provide results based on clicking pages on user query. The click-log based methods provide results based on clicking pages on user query. The click based strategy can only work on repeated queries from the same user. It does not work on hierarchy of users and hence irrelevant results are produced. The profile based method provides search results based on user personal information and their interests. The search engine generates a set of results comparing the user query and their profile. Hence appropriate results are provided satisfying all users. Hence in this profile based PWS is preferred over click based. The profile based PWS framework can effectively generalize profiles by queries while respecting user specified requirements. In order to improve the search engine results Content based ranking is used. The Content based ranking algorithm that automatically ranks the search engine results based on term frequency. Based on ranking technique, the required information is retrieved by the users. To enhance the stability of personalisation various methods are available. Here In order to enhance the accuracy of search results an advanced data searching system is used in PWS framework. The Advanced data searching system consists of a spatial database.

The Spatial database stores many query objects. Range Searching Technique is used for extracting information from the spatial database. Linkability Assessment mechanism is used to evaluate if the user query is linkable to their user profile

II. PERSONALIZED WEB SEARCH(PWS)

The main goal is to improve the quality of the search engine. The information available in the web is important not only useful to individual user and also helpful to all users from various fields such as business organization, education, and some research areas. Personalized web search(PWS) is a searching technique that provides better results for individual user needs. One of the most critical benefit of the personalized search has is to improve the quality of decisions consumers make Through PWS The information available in the online is more structured data by making search engines by generating high precision results and to make all the users possible to find anything easily. Web mining is one of the data mining techniques to automatically extract the information from web documents. The PWS is broadly classified into two types i)click based method ii)profile based method. Click based personalization is done based on repeated clicking of particular URL..The Fuzzy based clustering method automatically clusters the feedback sessions based on clicking link multiple times. But this method is not applicable for hierarchy of users.It is the main goal of personalization is to satisfy all the requirements of individual users. Click based personalization does not satisfy multiple users. And hence profile based personalization is preferred over click based methods. Profile based personalization effectively generalizes queries by their user profiles. Profile based personalization is applicable for multiple users through hierarchy of user profile. Here personalization is done based on user behavior, their interests, field etc.,Through Profile based personalization search engine generates appropriate results satisfying all the users. Web content mining involves text and document and structures. Web usage mining involves data from user registration and user transaction.. It is very helpful to generate a new page, lot of pages are added, removed and updated anytime. Data sets available in the web can be very large and occupy ten to hundreds of terabytes, need a large farm of servers. A web page contain three forms of data, structured, unstructured and semi structured data. A number of algorithms are available to make a structured data, one such algorithm is Content based ranking algorithm. An unstructured data can be analyzed using term frequency; The term frequency determines the total relevancies of each link in a page. It defines the number of times the term appears in a page, a word occurs in a document of a corpus. We have to improve searching in the web by adding structured documents by profile based personalization. It provide efficient results for a hierarchic of users.

III. PROBLEMS IN PWS BASED INFORMATION RETREIVAL

Personalized Web search has become a most important portal among all the users worldwide. Though various search engines provide results based on the user queries., sometimes irreverent results are generated by the search due to many reasons. Inappropriate information retrieval occurs due to mismatching of vocabularies with the search word .In some cases irrelevant results are generated for ambiguous queries, for eg.Apple iPhones.Sometimes results generated by the search engine do not satisfy the customers through incomplete data. Authentication is another risk in search engines.Sometimes the search results are not linkable with the user preferences thus failing to satisfy all the users.

IV. PROPOSED ADVANCED DATA SEARCHING SYSTEM

This Advanced Searching system comprises of a spatial database.The Spatial database contains multidimensional data objects with their extensions.The Spatial database consists of user query data along with its location,landmarks,reviews,region names etc.,The Spatial database provides faster access to those data based on different selection criteria. Range Search preprocess the set of multidimensional data in the database.It determines the intersection between the user query data and the data objects in the spatial database.It defines the minimum distance from the query point to the object in the database.Finally the minimum distance from the object is determined. To enhance the stability of the data Linkability Assessment mechanism is used.This mechanism is used to evaluate if the user query is linkable to their user rofile.Linkability metric quantifies if the query is likely to be linked to the profile.This evaluation is based on the popularity of the query and the proximity of the query to user profile.Linkability is the average between the popularity and the proximity of the query.

V. IMPLEMENTATION OF ADVANCED DATA SEARCHING SYSTEM SPATIAL DATABASE

Spatial database is a database that manages multidimensional objects in space with explicit knowledge about objects ,their extends,location, position,region in search.Spatial data composed of a single point to rectangles to thousands of polygons.



Fig2.Spatial database

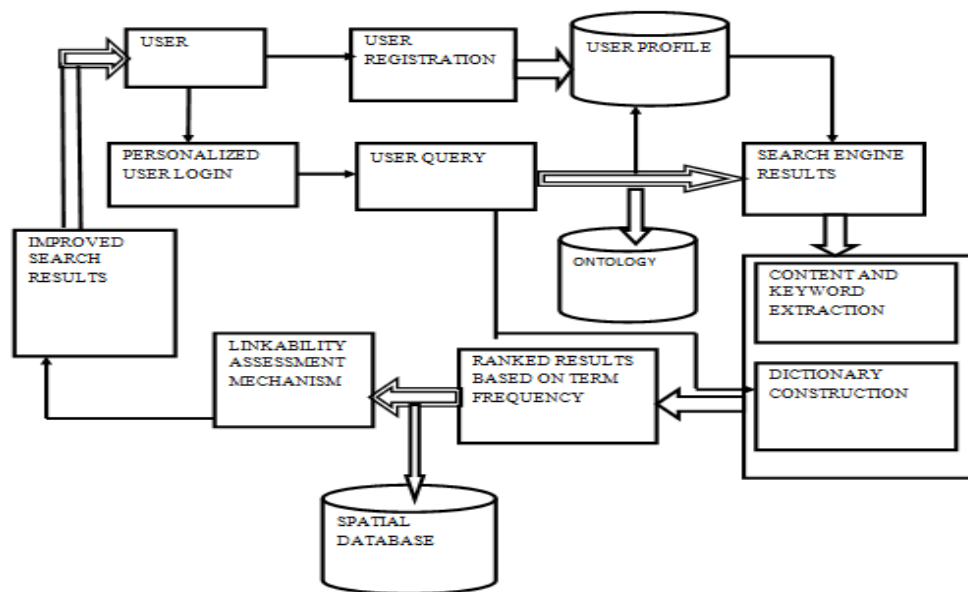


Fig1. Advanced PWS framework Architecture

RANGE SEARCH

Range Search preprocesses the set of objects in the spatial database. It finds the minimum distance S_p between the user query and various concepts in the database and the closest object to the user query is determined.

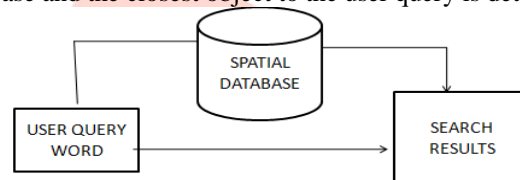


Fig 3. Preprocessing of results

LINKABILITY ASSESSMENT

Linkability Assessment Mechanism evaluates whether the user query is linkable with the user profile information. Linkability metric is used to evaluate the user query based on popularity of the query among the users in the user profile list and the proximity of the user query. Thus satisfying all the users with most relevant information.

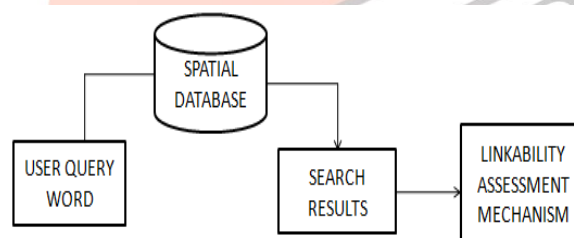


Fig 4. Linkability assessment process

LINKABILITY ASSESSMENT ALGORITHM

This mechanism is used to evaluate if the user query is linkable to their user profile. Linkability metric quantifies if the query is likely to be linked to the profile. This evaluation is based on the popularity of the query and the proximity of the query to user profile. Linkability is the average between the popularity and the proximity of the query.

Algorithm:

Require: The query Q

- 1: linkability $\leftarrow 0$
- 2: for all keyword $\in Q$ do
- 3: $M_u \leftarrow \text{ComputeUserProfileMetric}(\text{keyword})$
- 4: $M_g \leftarrow \text{ComputeGroupProfileMetric}(\text{keyword})$
- 5: linkability $\leftarrow \text{linkability} + M_u + M_g$
- 6: end for
- 7: linkability $\leftarrow \text{linkability} / |Q|$
- 8: if linkability > 0 then 9: SendToObfuscator(Q)

```

9: else
10: SendToPrivacyProxy(Q)
11: end if

```

Linkability Metric: Linkability metric defines the average between the popularity and the proximity of the query. Popularity is the number of times the query used by the PWS users. Proximity defines the closeness of the query with the user profile.

$$\text{Linkability metric} = \frac{\text{Proximity}}{\text{Popularity}}$$

Proximity Metric:

$$M_u = \frac{\sum_i \frac{a_i - b_i}{a_i + b_i}}{\#keywords}$$

M_u is the user profile similarity metric

a_i usage frequency of particular query

b_i average usage frequency of all keywords

Popularity metric

$$M_g = \frac{\sum_i \frac{a_i - c_i / d_i}{a_i + c_i / d_i}}{\#keywords}$$

where, c_i is the number of times the query used among the PWS users
 d_i is the number of users using the particular query

VI. CONCLUSION

The Advanced data Searching System presented in This paper improves the overall performance of a PWS system. PWS captures user profiles in a hierarchical manner. The framework makes use of spatial database to extract the overall user requirement as this database manages multidimensional objects in space with explicit knowledge about objects. Range searching is done to determine the closest object to the user query. Further improved efficiency and stability is achieved through linkability assessment mechanism in This framework. It allows the search engines to generate results similar to the user profile. It also provides results based on the number of times the query used among the PWS users. This advanced overcomes the drawbacks available in ranking mechanisms. This framework also allows each user to authenticate themselves with the non-trustworthy server. PWS also has Profile Updation that allows updation of user data as their requirements change. Thus this system provides the PWS users well authenticated with the server. This system helps retrieving accurate results by improving the quality of search engines through linkability mechanisms. The results also confirmed the effectiveness and efficiency of our solution using advanced data search system in personalised web search. This system satisfies all the users by providing most related data.

REFERENCES

- [1] Chen "Spatial database with location preferences", 2014
- [2] Lidan Shou, "Supporting Privacy of Personalized web search", 2014.
- [3] Zhang Lu, "A New algorithm for inferring user search goals with feedback sessions", 2013
- [4] Restructuring Web Search Results Using Parameter Estimation Algorithm
- [5] G. Poonkuzhali, "Content Based Ranking of Search Engines", 2013
- [6] Chen, "Rankbox: An Adaptive Ranking System for Mining Complex Semantic Relationships Using User Feedback", 2012.
- [7] Josno, "User Profile Creation Based On Navigation Pattern for Modeling User Behaviour With Personalised Search", 2013.
- [8] T. Joachims, "Evaluating Retrieval Performance Using Clickthrough Data," Text Mining, J. Franke, G. Nakhaeizadeh, and I. Renz, eds., pp. 79- Physica/Springer Verlag, 2003.
- [9] M. Spertta and S. Gach, "Personalizing Search Based on User Search Histories," Proc. IEEE/WIC/ACM Int'l Conf. Web Intelligence (WI), 2005.
- [10] B. Tan, X. Shen, and C. Zhai, "Mining Long-Term Search History to Improve Search Accuracy," Proc. ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD), 2006.
- [11] K. Sugiyama, K. Hatano, and M. Yoshikawa, "Adaptive Web Search Based on User Profile Constructed without any Effort from Users," Proc. 13th Int'l Conf. World Wide Web (WWW), 2004.