

# IMPROVING EFFICIENCY OF PERSONALIZED WEB SEARCH USING CONTENT AND KEYWORD EXTRACTION

Ms.G.ANISHYA<sup>1</sup> Mrs.G.SANGEETHA<sup>2</sup>

<sup>1</sup>M.E CSE, <sup>2</sup>Faculty of CSE dept, <sup>1,2</sup>Valliammai engineering college

<sup>1</sup>anishyarose91@gmail.com, <sup>2</sup>gsangeethavec@gmail.com

**Abstract** —. Personalized web search (PWS) has demonstrated its effectiveness in improving the quality of various search services on the Internet. In order to improve the efficiency of search engine results a PWS framework is implemented. The PWS framework can effectively generalize profiles by queries while respecting user specified requirements. To rank the search engine results content based ranking methodology is based on the user query, search engine results are retrieved. Analyse the search engine results based on keywords and content. User query is pre-processed to identify the root words. Every root words are considered for Dictionary construction. From every result page keywords and content words are pre-processed and compared against the dictionary. If a match is found then particular weight is awarded to each word. Finally, the total relevancy of the particular link against user request to determine the term frequency

**Keywords** — Personalized Web Search, Ontology, Content Based Ranking, Term Frequency.

## 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. Personalized web search (PWS) is a general category of search techniques aiming at providing better search results, which are tailored for individual user needs. The solutions to PWS can generally be categorized into two types. They are click-log-based methods and profile-based method. The click-log based methods are straightforward—they simply impose bias to clicked pages in the user's query history. Although this strategy has been shown to perform consistently and considerably well, it can only work on repeated queries from the same user, which is a strong limitation confining its applicability

In contrast, profile-based methods improve the search experience with complicated user-interest models generated from user profiling techniques[6]. Profile-based methods can be potentially effective for almost all types of queries. Although there are pros and cons for both types of PWS techniques, the profile-based PWS has proved more effectiveness in improving the quality of web search recently

with increasing usage of personal effectiveness in improving the quality of web search recently with increasing usage of personal and behavior information to profile its users which is usually gathered from query history, browsing history and click-through data bookmarks user documents.

## A. RELATED WORKS

In recent years, many works have been done to infer the so called user goals or intents of a query. But in fact, their works belong to query classification. Some works analyze the search results returned by the search engine directly to exploit different query aspects. However, query aspects with no efficiency in have limitations to improve search engine relevance. Some works take personalization through two broader categories namely i)click based methods and ii)profile based methods. The click based methods generate search engine results through clicking of particular link . But this strategy works only on repeated queries from the same user. But are not applicable on multiple user queries. On the other hand, profile based methods provide results comparing user profile and user query. It works for hierarchy of user profiles and generate better results .Content based ranking is done is proposed to rank the search engine results by analyzing content and keywords[4]. By analyzing the content and keywords, term frequency is calculated. Term frequency is the number of times a term or a document appears in a page. This determines the total relevancies of a link in a page. This ranking in personalized framework reduces complexity of users and provide better results satisfying all the users.. One application of user search goals is restructuring web search results. There are also some related works focusing on organizing the search.

## II. PROBLEMS IN INFORMATION RETREIVAL

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 irrelevant results are generated by the search due to many reasons.

In appropriate 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.

## III. DESIGN GOALS OF WEB SEARCH

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. The information available in the online must be 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 three issues in the WWW are web content mining, web structured mining, web usage mining. Web structure mining involves web structure documents and links.

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 or a corpus.

We have to improve searching in the web by adding structured documents by profile based personalization. It provide efficient results for a hierarchy of users.

#### IV. PERSONALIZED WEB SEARCH

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. They are click-log-based methods and profile-based method.

The click-log based methods provide results based on clicking pages on user query. The profile based methods provides results based on user personal information and their interests. The click-log based methods provide results based on clicking pages on user query [2]. 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.

##### A. Ranking Search Engine Results

To rank the relevant pages content based ranking algorithm is used [5][6]. Based on the user query, search engine results are retrieved. Every results are analysed based on keywords and content. User query is pre-processed to identify the root words. Every root words are considered for Dictionary construction and Dictionary is built with synonyms for the user query.

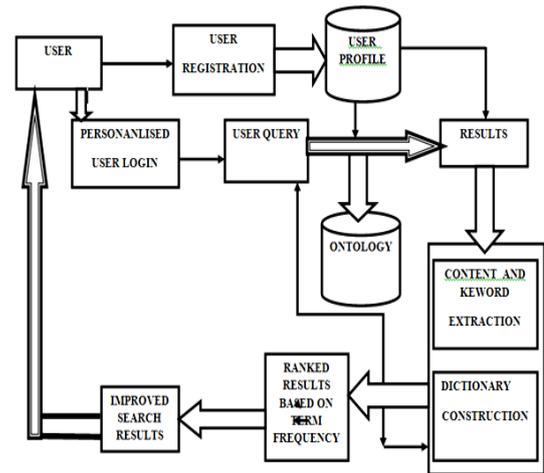


Fig1. Proposed PWS Framework

#### V. IMPLEMENTATION OF PERSONALIZED WEB SEARCH

##### A. User profile and ontology construction

In PWS process each new user registers themselves with their personal information. Already registered user login with their username and password. A hierarchy of user profiles are stored in a user profile database.

Then a user enters his/her search word in our ontology search, a set of ontologies are developed. An ontology is nothing but the terms, concepts, relations related to the user query word.

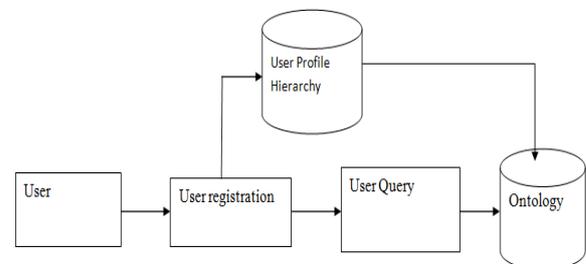


Fig2. User profile and ontology construction

##### B. Query Mapping And Search Results

Each query is processed and the results are generated by the search engine. These results are generated by the search engine based on ontologies of query word and the user profile information. The user query is mapped with user personal information for analyzing each user behavior, interests to identify the exact requirement of each individual user and results are produced.

##### C. Content And Keyword Extraction

Based on the user query, search engine results are retrieved. Individually analyse the results based on keywords and content. User Query is pre-processed to identify the root words. Every root words are considered for Dictionary construction and Dictionary is built.

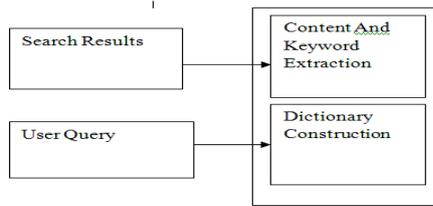


Fig 2. Content And Keyword Extraction Process

#### D. Ranking

Every result page keywords and content words are pre-processed and compared against the dictionary. If a match is found against the content or keyword and the rootword then particular weight is awarded to each word. Finally, the total relevancy of the particular link against user request is determined through the term frequency. The page which contains total relevancy value nearest to 1 are ranked as first page and 0 are ranked as last page.

#### Term frequency

The number of times the terms in the query appear in the document can help determine how relevant the document is,

$$tf_{ij} = \frac{f_{ij}}{\max\{f_{1j}, f_{2j}, \dots, f_{Vj}\}}$$

The weight of a term  $t_i$  in document  $d_j$  is the number of times that  $t_i$  appears in  $d_j$ , denoted by  $f_{ij}$

#### VI. Efficiency Evaluation Of Search Engine Results

The ranked results are automatically generated by the search engine to enhance the efficiency of the search quality.



Fig3.NonLocation Based

Fig3.illustrates the search engine results generated irrespective of user location.For example,a user enters a query word “Travels” he retrieves related information irrespective of his location in the user profile database.

Fig4.illustrates the search engine results generated respective of user location. These results helps user getting more exact

and related information.



Fig4.Location Based

#### CONCLUSION

This paper presented a client-side efficiency in personalized web search. PWS captures user profiles in a hierarchical manner. The framework allowed users to specify customized user requirements via the hierarchical profiles. The results also confirmed the effectiveness and efficiency of our solution using content based ranking in personalised web search, automatically efficient information is retrieved by the users. This system satisfies all the users by providing most related data The results also confirmed the effectiveness and efficiency of our solution.

#### REFERENCES

- [1]Lidan Shou, "Supporting Privacy of Personalized web search", 2014.
- [2]Zhang Lu, "A New algorithm for inferring user search goals with feedback sessions", 2013
- [3]Restructuring Web Search Results Using Parameter Estimation Algorithm
- [4] G.Poonkuzhali, "Content Based Ranking of Search Engines", 2013
- [5], Chen, "Rankbox: An Adaptive Ranking System for Mining Complex Semantic Relationships Using User Feedback" 2012.
- [6]Josno, "User Profile Creation Based On Navigation Patern for Modeling User Behaviour With Personalised Search", 2013.
- [7]T. Joachims, "Evaluating Retrieval Performance Using ClickthroughData," Text Mining, J. Franke, G. Nakhaeizadeh, and I. Renz, eds., pp. 79- Physica/Springer Verlag, 2003.
- [8]M. Spertta and S. Gach, "Personalizing Search Based on User Search Histories," Proc. IEEE/WIC/ACM Int'l Conf. Web Intelligence (WI), 2005.
- [9] 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.
- [10] 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.