Personalized Web Search Using Browsing History and Domain Knowledge Based on Enhanced User Profile

G Rohit [1*], V Ramakrishna [2], L Satya Rajiv Krishna [1]
[1*]Post Graduate Scholar, Dept. of Computer Science Engineering, K L University,
Vaddeswaram, Guntur District, India
[2]Assosciate Professor, Dept. of Computer Science Engineering, K L University, Vaddeswaram,
Guntur District, India

Email: gajulrohit@gmail.com

ABSTRACT

Generic search engines are important for retrieving relevant information from web. However these engines follow the "one size fits all" model which is not adaptable to individual users. Personalized web search is an important field for tuning the traditional IR system for focused information retrieval. This paper is an attempt to improve personalized web search. User's Profile provides an important input for performing personalized web search. This paper proposes a framework for constructing an Enhanced User Profile by using user's browsing history and enriching it using domain knowledge. This Enhanced User Profile can be used for improving the performance of personalized web search. In this paper we have used the Enhanced User Profile specifically for suggesting relevant pages to the user. The experimental results show that the suggestions provided to the user using Enhanced User Profile ae better than those obtained by using a User Profile.

Keywords: Data pre-processing, user modelling, Page ranking strategies, personalization.

1. Introduction

With the help of the World Wide Web, every user has a specific goal in searching a lot of information. From web, as it means that it seems as a platform for retrieving the vast information from the web, that are associated with education,marketing,sports etc., search engines plays an important role in searching for information. But there are still some changes should be made according to the usage of the internet. As the usage of the internet is increasing day by day. Search engines must contain huge amount of data, and hence it is impossible for retrieving the meta data in time and searching accuracy may reduce because it becomes it sometimes may highly important for retrieving exact meta data, from huge amount of data. To handle such situation we should develop a personalized web search engine. It is highly difficult to let the search engine to known what uses actually want.

When different users we are trying to execute the some query. Typically, the searching results will be same, i.e., in the sense what will about is difficult ideas and also uses may have different type of results is based upon their interest. Providing some search results for different users might not appropriate for the users which require difficult information. For the same search results two or more users may require different pieces of information. So let us explain the above sentence in a brief i.e., let's take two persons of different streams who wants as a biochemist and a software engines. They want to search the some keyword called virus .Biochemist needs the information related to malicious software, but unfortunately they all are getting some results. So current existing should be explored i.e., we should built a generic search engine. For this type of query it should project results are different topis and must submit distract information that must dynamically, presented the generic search engine.

Disadvantages of existing system
- ➢ When different users, wants to search for some keyword, some results will be returned by the typical search engine for the working different stream.
- ➢ As the internal usage was increasing day by day, it highly difficult for the user to get the accurate results.
- ➢ Extracting the Meta data from big data will be highly difficult.
- ➢ Search result will be some as if they are hard coded.
- ➢ No accurate search results will be generated.
- ➢ They will be no word for secure browsing.

2. Related work

As the web content (data storage) is growing maximum and more and also necessary methods should be provide in delivering the accurate results for individuals users. The most searchers for the keyword are,

- ➢ Problems that are faced by data itself.
- ➢ Problems that are faced by different users in search for some keywords.
- ➢ Problems regarding the content of search requests.
- ➢ Problems in understanding the dynamic changes of users request.

Google Architecture Overview

Google architecture users mainly based on web crawling and it was done by different distributed web crawlers. The execution of this architecture start from URL server. As soon as when we request an URL from the URL search engine, it directly goes and checks in URL server. This URL server will build the crawler dynamical based upon user request. This crawler will search in the store serves and dynamically based, Anchor's based on crawler. This Anchor consists of user related links. This generated link are build dynamically based upon page ranking. And these links are the search results for the user.

If we observes the above architecture these are some bugs that are need to be fixed i.e., Here the search results were only based on keywords but not upon the area of interest of that particular person, i.e., Here this web search engine is not personalized.
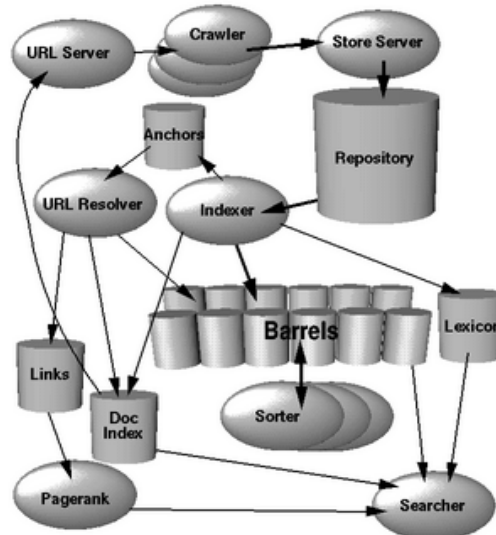

Fig: 1 A High Level Google Architecture

Also the content of the search results might not related to the user.

3. Indexing

After fetching the web pages to store serves as per above figure. The store server then compares the web pages into respotany. We known that every webpage has a URL ID called as document ID, and a new URL i.e., parsed out of a web page. At this stage every page should be indexed and it was done by the indexed and sorted. And it is sorted based upon the hils and occurrences of the WebPages. If does not bother about the content of the user.

Literature Review

Here comes the picture of personalization. Personalization is the process of customizing the webpage to the individual users i.e., filtering out the irrelevant results and identified the results based on that keywords.

Our literature review point not only of personalization but also secure personalization. Before discussing secure personalized we should know the key steps for the process of personalization.

➢ Adding extra data for webpages.
➢ Personalized user modelling.
➢ Recommending personalized page ranking.

1. Adding Extra Information

It is the process of add the metadata for the web pages as the can be sorted according to user profiles. Adding of information is be web data pre-processing. The addition of data to the web pages can be done in one or many ways:

➢ By providing some semantic to the WebPages.
➢ Sorting the web pages based on page ranking.
➢ Dividing the WebPages based on each domain knowledge.

For Example

```
<xml version= "1.0"?>
<rdf:RDF xmlns:rdf=:http://www.w3.org/1999/02/-rdf syntax"
Xmlns/;camera="http;??www.yuchen.net/photography/camera.owl3:>
<rdf:description
Rdf:about="http"//www.yuchen.net/package.rdf">
<rdf:label>Nikon-D70</rdfs:label>
</xml>
```
If we observe carefully we are adding some extra information to the webpage so that the result will be same up to the context by the tags
<Camera: photographer>

Some other pre-processing methods

One step of personalization for data pre-processing can be achieved through data clearing, identification of visited pages, path identification, session identification feature identification and user identification. If we combine all these words together, initially we should route a session for the user to access the web pages based on user content before that data should be cleaned and modify by providing the entire information for WebPages. As when the webpage was hit it automatically must update the ranking of that page can be based upon the content of the user rather than content.

ATR (Automatic Term Recognition)

It is a process of refining the estimation values based on personalized user phase content. The re extraction process can be done using verb algorithm.

NLP (Natural Language Process)

These tools are mainly used for the extraction of noun phrases. This can be done by filtering or the process of removes infrequent words (or) the word that are not listed in the World Wide Web directory.

DHC (Divisive hierarchal clustering)

DHC algorithm determines the clustering correlation values between the rain of week and strong correlation values.
Secure framework for proposed system:
In this proposed secure user personalization comes into the picture, because it is an essential port of a personalized web search. Of an individual users interest but also secures the web browsing history and users interest data. So, it is the process of developing the personal interest and content preferences of the users in terms of browsing history, user knowledge, about the world, likes and dislikes etc...
This framework will mainly consists of two parts. They are:
1. Secure modelling
2. User modelling

1. Secure modelling

Universal fact is we can browse web page only through browser "Internet Explorer". So let's enhance the current system to some extent, before browsers the web pages in the browser. Initially we must get register into the browser with a trusted email and password. This registration process follows with the area of interest in browsing. Area of interest can change dynamically based upon the user interest. So that the entire browsing history will be authorised to this particular person and there will be no chance of misusing this frame work. Why because here user profile comes into the picture based on the web search using user id and user password according to the cloud storage for storing the history dynamically in cloud in order extract metadata from big data.

2. User modelling

User modelling consists of mainly three phrases:
  ➢ Browsing history.
  ➢ User profile modelling.
  ➢ Domain knowledge modelling.
At this stage of user profile modelling was based on user's blog/navigation/browsing history to main knowledge in order to personalize the secure web search. In this above frame work of user modelling domain knowledge plays a key role, why because it sorts call retrieves the information based on different domains and categories. Here domain knowledge phrase fetches the search results from the web directory based on domain knowledge and places, these results in an enhanced user profile.
        If the content results of the enhanced user profile matches with the query handler, then this search will also sort the web pages based upon the page ranking and displays the results. Hence it is possible to say that based upon this frame work can easily build a generic search engine based secure personalized web search.

Brief Discussion:

Domain Knowledge Modelling

Domain Knowledge Modelling was mainly based upon the user's interest. It is said to be the background knowledge that the user wants to enhance in the user profile. So to enhance the results of the domain knowledge we need to first crawl the web pages from web directory based on the user interest and a search query and a domain knowledge.

Now we form a table called as category matrix where weight of an each term is said to be category, weight is represented by the frequency term category and will represent the matrix where number of times the term is present in the category.

2. User modelling

User profile is used to project the search results based upon the user interest and further products their intension for new queries. It also helps us to deal with ambiguous queries.

In order to build an effective user profile, we need to first classify browsing history of the user that were accessed based on particular category. Here category table was construction as mentioned above in the table. So that now we come to know that construction of user profile was based on user browsing history.

Term Category	Category 1	Category 2	Category 3	-	Category n
T1	W11	W12	W13	-	W1n
T2	W21	W22	W23	-	W2n
T3	W31	W32	W33	-	W3n
|	|	|	|	-	|
Tm	Wm1	Wm2	Wm3	-	Wmn

Table: Term Category matrix

When number of web pages was browsed by the user and that fast the threshold of the user profile will be increased and it was represented by the fined number of category weights, it was denoted by

$$U= \{CW1, CW2, CW3, - - - - - -, CWm\}$$

Where W will be the number of WebPages of that particular category i which are visited by the user.

Enhanced User Profile

Enhanced user profile play a major role and it is an important part of the frame work. Enhanced user profile improves the threshold of the user profile by using the domain knowledge.
EUP consists of each unique URL of the webpage of the user profile, match it with the domain knowledge URL's. This process can be briefly explained in the following steps:

1. Select unique URL from user profile.
2. Add the URL to the enhanced user profile.
3. Find the consecutive simulates between enhanced user profile based on the category and domain knowledge.
4. Rank URL according to the cosine similarly.
5. Calculate the average of cosine similarly for top 20 URL.
6. Display the URL in the search results.

$$Cosine\ (dj,u)= <dj*u>/||dj||*||u||$$
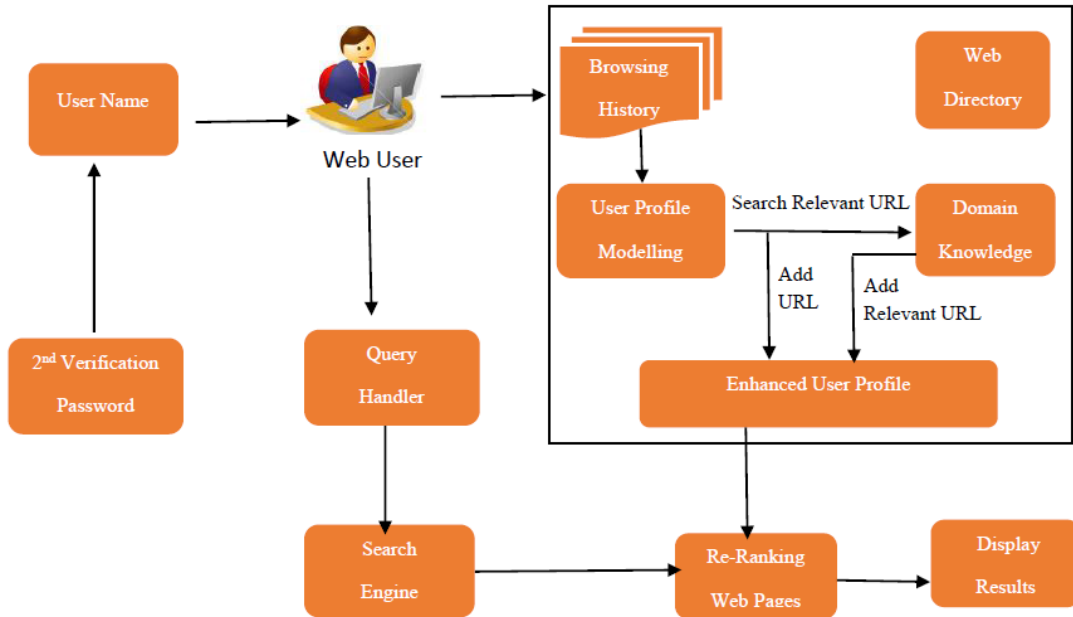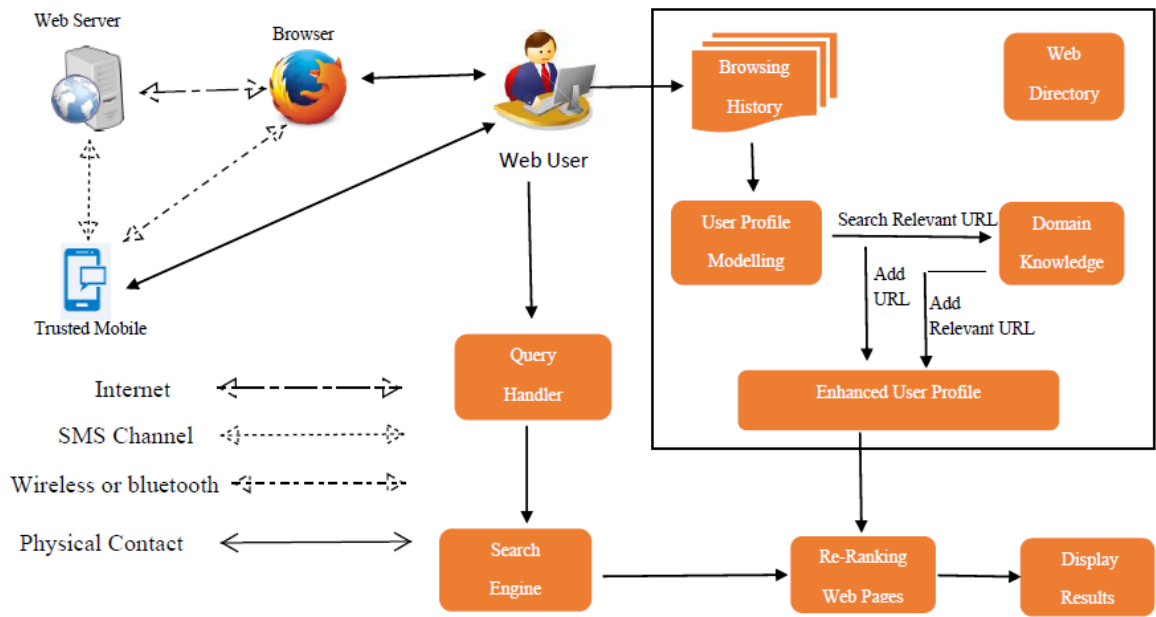

Fig: Proposed Framework

Future work

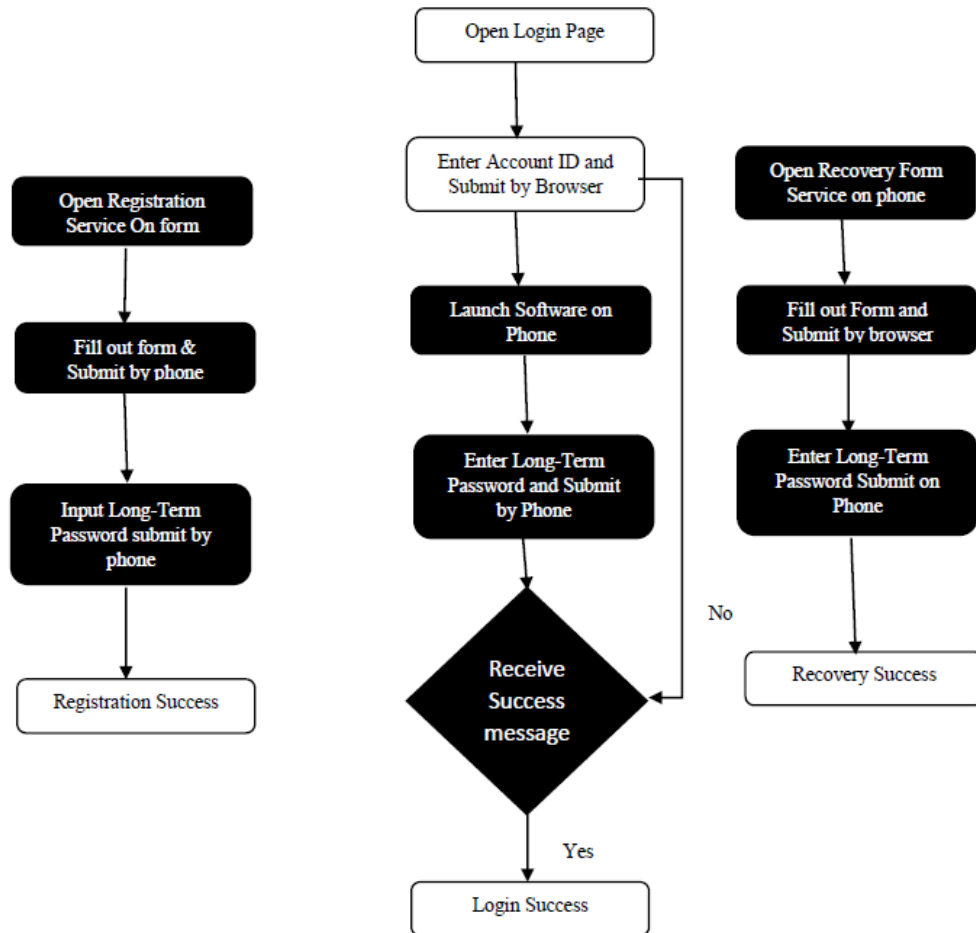Fig: A Frame work for Highly Secured Personalized Web search

Fig: Activity Diagram for Secure Browsing

Conclusion

In this paper we have proposed a frame work for secure personalized web search for browsing history with the help of secure modelling, User profile and domain knowledge. Based on the user profile and the domain knowledge, system will keep on updating the user profile and builds on enhanced secure personalized web search. Here we had applied a collaborative filtering for personalized web search in our frame work.

References

[1]. K.W.T. Leung, D.L. Lee and Wang-Chien Lee, "Personalized Web search with location preferences", IEEE 26th International Conference on Data Engineering, pp. 70 I - 712, 2010

[2]. M Speretta and S Gauch, "Personalized Search Based on User Search Histories", Proceeding Of International Conference on Web Intelligence ,pp. 622-628,2005.

[3]. C Liang, "User Profile for Personalized Web Search", International Conference on Fuzzy Systems And Knowledge Discovery, pp. 1847-1850,2011.

[4]. F Liu, C Yu and W Meng, "Personalized Web Search for Improving Retrieval Effectiveness", IEEE Transactions On Knowledge And Data Engineering, pp. 28-40, Volume 16,2004.

[5]. X Pan, Z Wang and X Gu, "Context-Based Adaptive Personalized Web Search for Improving Information Retrieval Effectiveness", International Conference on Wireless Communications, Networking and Mobile Computing, pp. 5427 - 5430, 2007.

[6]. O. Shafiq, R. Alhajj and 1. G. Rokne, "Community Aware Personalized Web search", International Conference on Advances in Social Networks Analysis and Mining, pp. 3351 - 355,2010

G.Rohit is currently working toward   the Post-Graduation degree in computer science and engineering at the KL University. His area of interest is knowledge engineering and Web Services.

Dr.V.Ramakrishna is Assoc Professor at KL University. His area of interest is software engineering.

Loya Satya RajivKrishna is currently working toward   the Post-Graduation degree in computer science and engineering at the KL University. His area of interest is knowledge engineering and Web Services.