

Re-Ranking Based Personalized Web Search Using Browsing History and Domain Knowledge

¹N.P.Sasi Thev, ²N.Vivekanantha Moor

¹Post Graduate Student, ²Asst. Professor

Dept. of Information Technology, SRM University, Chennai,, India

Abstract- Different users usually have different type of information needs when they use search engines to find web information. But the current web search engine provides the same information to the every user. To overcome this problem we propose a personalized web search engine Based on the user profile and the domain knowledge the system keeps on updating the user profile and enhanced user profile. This enhanced user profile is then used for suggesting relevant URL web pages. With the use of re-ranking concept, its providing relevant information based on the user searching browsing history.

Index terms- Domain Knowledge, Enhanced user profile, Re-ranking.

I. INTRODUCTION

With the growth of World Wide Web, The Google has provided a lot in looking information from the web. They helps the user details on the web simple and fast. But there is still room for enhancement. Present web Google do not consider particular needs of user and provide each user similarly. It is challenging to let the online look for motor know what we the user actually want. The Google are following the "one dimension suits all" design which is not convenient to individual users. When different users give same question, same outcome will be came back by a common online look for motor, whichever user presented the question.

When different users give same query, same result will be returned by a typical search engine. It becomes difficult for the user to get the relevant content. We suggest a framework for personalized web search which considers individual's interest into mind and enhances the traditional web search by suggesting the relevant pages of his/her interest. We have proposed a simple and efficient model which ensures good suggestions as well as promises for effective and relevant information retrieval.

We have implemented the proposed framework for suggesting relevant web pages to the user. This framework for Personalized search engine consists of user modeling based on user past browsing history or application he/she is using etc. This make the web search more personalized. This section presents different approaches and the related work done in the field of Personalized Web search.

Personalized web search is considered as a promising solution to handle these problems, since different search results can be provided depending upon the choice and information needs of users. It exploits user information and search context to learning in which sense a query refer.

This might not be appropriate for users web need different information. While looking for the information from the web, users need information based on their interest. For the same keyword and key phrase two users might need different part of details. This fact can be described as follows: a scientist and a developer may need details on "virus" but their areas are is entirely different. Biologist is looking for the "virus" that is an organism and developer is looking for the harmful application. For this type of question, a number of records on unique subjects are returned by general Google. Hence it becomes difficult for the user to get the relevant content.

Moreover it is also time intensive. Customized web look for is considered as a promising solution to handle these problems, since different look for outcomes can be provided depending upon the choice and details needs of users. It uses user details and look for perspective to learning in which sense a question refer. In order to perform Customized Web look for it is essential design User's need/interest. Construction of details is an integral part for personalized web look for.

User profiles are constructed to design user's need depending on his/her web usage data. This document suggests a structure for building details and enhances the details using qualifications information. This Enhanced User profile will help the user to recover focused details. It can be used for indicating good Web pages to the user depending on his look for question and qualifications information.

II. RELATED WORK

Framework for Personalized search engine consists of user modeling based on user past browsing history or application he/she is using etc. And then use this context to make the web search more personalized. This section presents different approaches and the related work done in the field of Personalized Web search.

Applied a wrapper around the look for website that accumulates information about client's look for action and develops information by determining gathered information (queries or snippets). They have used these information to re-rank the on the internet look for motor results and the rank-order of the user-examined results before and after re-ranking were compared. They discovered that user information depending on concerns and user information depending on thoughts both were similarly effective and re-rank provided 34% enhancement in compare to rank-order.

Recognized that current web Google do not consider the unique needs of user or passions of user and suggests a novel technique which uses look for record of user to learn user information. This work uses user's look for record for studying of information and classification structure for studying of a general user profile and then brings together both information to classify user's question to signify user's look for objective and to disambiguate the words used in question.

Also recognizes that different users may have need of different unique information, when they use Google and techniques of customized web look for can be used to fix the problem effectively. Three techniques Rochio technique, k-Nearest technique and Support Vector Machines have been used in to build information to present an individual user's choice and discovered that k-Nearest technique is better than others in terms of its performance and sturdiness.

Recommended a perspective centered customized web look for design. In this paper the writers have given an individualized web look for result which is depending on the need of user in various situations. The research of design has led to three ideas to apply the design, which is semantic listing for web sources, modeling and obtaining user perspective and semantic likeness related between web sources and user perspective. The writer has described it as perspective centered flexible customized web look for.

We have suggested a Customized Web search design with place choices. In this document the place and material idea has been divided and is structured into different ontology to make an ontology-based, multi-facet (OMF) user profile which is taken by web history and place attention. This design actually gives outcomes by describing the ideas depending on the choice of user.

By keeping the different attention of the users in mind, place entropy is presented for finding the level of attention and information related to place and question. The personalized entropies actually stabilize the appropriate outcome material and place material. At last, an SVM depending on the ontology is created which can be used for future objective for position or re- position. The tests reveal that the outcomes created by OMF information are more precise in evaluation with the ones which use guideline method.

Have suggested an individualized web look for design that brings together group centered and material centered facts depending on novel position strategy. These days, posting information on internet has become a daily action. A great deal of information is submitted in the form of web pages, news, and weblogs etc. regularly. So, it becomes very challenging for the user to look for appropriate material. Not only for users but also for Google like Google it becomes challenging. Mass confusion is the only reason behind this challenging situation. Other than this user's choice is the second problem, which is not taken into account while generating the outcomes.

III. FRAMEWORK FOR PERSONALIZED WEB SEARCH

A. Architecture Diagram

We propose a system for customized web seek which considers singular's enthusiasm into psyche and upgrades the customary web look by recommending the pertinent pages of his/her advantage. We have proposed a straightforward and proficient model which guarantees great recommendations and also guarantees for compelling and pertinent data recovery.

Notwithstanding this, we have executed the proposed structure for proposing applicable pages to the client. Our framework considers client's profile (focused around client's weblog/route perusing history) and Domain Knowledge to perform customized web seek. Utilizing a Domain Knowledge, the framework stores data about diverse space/classifications. Data got from User Profile is arranged into these defined classifications.

The learning operators takes in client's decision naturally through the examination of client route/scanning history, and makes/upgrades improved User Profile molding to the client's latest decision In the general architecture of the project is user handle the query and collect the information from the web from that user search query to re-ranking the web page and finally its display the user which web pages mostly used by user. It's very easy to collect the information by user history.

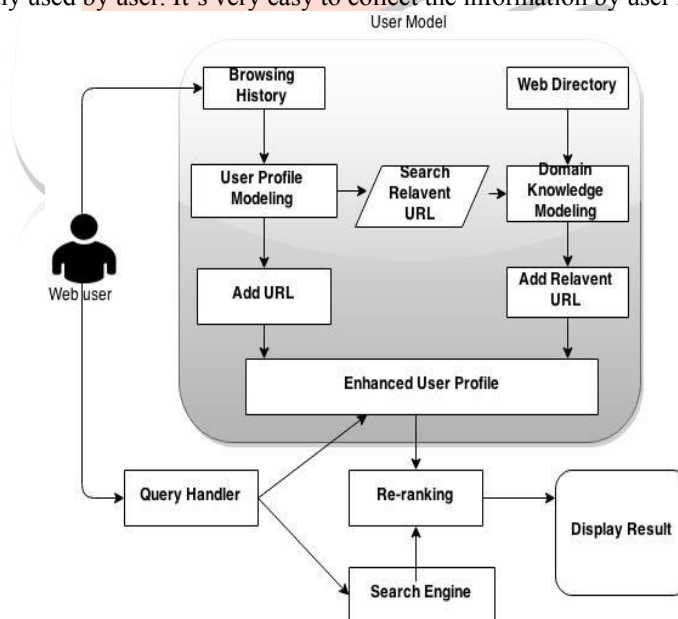


Figure 1: Architecture Diagram For Personalized Web Search

The web directory will help to the domain knowledge and it's using Add URL and Add Relevant URL pages. From the user past browsing history. Once the client inputs inquiry, the framework gives great proposals to customized web pursuit focused

around improved client profile. Further our model makes great utilization of the preferences of prominent web search tools, as it can re-rank the results acquired by the web index focused around the improved client profile.

IV. RE-RANKING ALGORITHM

The clicking patterns of a user can be represented as a frequency-based distribution. We re-ranking the distribution according to the clicking frequency for the calculation of rank correlation coefficient. To do that, a threshold value is needed. The threshold is a tolerance for re-ranking URLs having close number of clicking frequencies. Using the confidence interval method, it can be obtained. The confidence interval can be estimated from the clicking histories of M users.

Average clicking frequency for the URL U_k by the users is given by

$$f_{avg, k=1/M} \sum_{p=1}^M f_{p,k} \quad \text{Eq.1:}$$

Where $f_{p,k}$ represents the clicking frequency of URL U_k by the p -th user. If the average clicking frequency for U_i and U_j is almost same, it is hard to say which one is more preferable between U_i and U_j to the users. In this case, to estimate the threshold, we assume that clicking frequency difference collected from users is normally distributed. Hence, confidence interval of clicking frequency difference can be established. The average clicking frequency difference $\Delta f_{(k,k+1)}$ is computed by

$$\text{Eq.2: } \Delta f_{(k, k+1)} = f_{avg, k} - f_{avg, k+1}$$

To find the confidence interval of clicking frequency difference, the clicking frequency difference's standard error SE is calculated by

$$\text{Eq.3: } SE = s / \sqrt{n} \quad (3)$$

Where s is the standard deviation of the clicking frequency difference and n is the number of clicking differences. It can be applied when the data of clicking information is huge. In real situation, the data is sparse, so it is needed to estimate the confidence interval using a statistical distribution method. We use t-distribution to calculate the confidence interval in this paper, because t-distribution is closely related to the normal distribution and t-distribution with an infinite number of degrees of freedom is same as normal distribution.

$$\text{Eq.4: } C_{\Delta f} = \mu_{\Delta f} \pm |t_{(n-1), \alpha/2} \times SE| \quad (4)$$

$C_{\Delta f}$ is the confidence interval of clicking frequency difference, $\mu_{\Delta f}$ is the mean of average clicking frequency difference, and t is the confidence coefficient with $n-1$ degree of freedom. The value of α depends on level of confidence (LOC) such as 95% and 99%. We calculate the value of α using

$$\text{Eq.5: } LOC = (1-\alpha) \times 100 \quad (5)$$

According to the value of α , the value of t to be used in the calculation of confidence interval can be looked up in the standard chart of the t-distribution. By using the confidence interval, we generate the re-ranking distribution by the following re-ranking algorithm.

Input:

Confidence interval ($C_{\Delta f}$), Number of URLs (N), clicking frequency of the URL at position q in the list ($f[q]$), Original rank of the URL at position q in the list ($R[q]$).

Output:

```

Re-ranking distribution of URLs
q = 1;
while q ≤ N do
  If  $f[q] - f[q + 1] \leq C_{\Delta f}$  and  $R[q] > R[q + 1]$  then
    Exchange ( $f[q], f[q + 1]$ );
  end
  else q = q + 1;
end

```

Figure 2: Re-ranking algorithm

The algorithm can return the corrected positions the URLs, according to their ranks provided by the search engine if their clicking frequency difference is less than or equal to the confidence interval. It means that we respect the original rank provided by the search engine in re-ranking the URLs.

V. CONSTRUCTION METHOD AND MODULES

A. k-Nearest Neighbors Method

k-Nearest Neighbors (kNN) method is trying to find the nearest neighbor vectors of a web page and then use the average distance between the page vector and its neighbor vectors to measure the user interest relevance of the page. In order to establish

user profile model in kNN method, every document about the user interest should transform into a corresponding profile vector in the vector space. After turning into a vector, a web page can find the nearest neighbor vectors using the cosine value of the angle between the page vector and the user profile vectors as the following function shows:

$$\cos (p, u) = \frac{p \cdot u}{\|p\| \times \|u\|}$$

Where, p and u respectively represent the web page vector and the user profile vector. According to the cosine value, the k nearest user profile vectors of the web page are gotten. Then, the probabilities of relevance between the web page and user profiles are calculated by the following function:

$$P_u(p) = \frac{\sum_{i \in k, y_i = U} \cos (p, u_i)}{\sum_{i \in k} \cos (p, u_i)}$$

Where, p u (p) denotes the probability of relevance between the web page p and one of user profile U. u_i is one of the k nearest user profile which is belong to the user profile U. y_i denotes the user profile that u_i belongs to. kNN method is a more intuitive and simple user profile modeling method than others. As a passive machine learning method, kNN method does not require training in advance. It can directly use the vector space consisting of user profile vectors to test the user interest correlation of the web page.

B. Support Vector Machines Method

Support vector machines (SVM) method is trying to determine the interest correlation between a web page and the user profiles by finding an optimal hyper-plane, which can best distinguish the different user profiles. In the process of using SVM method to establish user profile model, the optimal hyper-plane is found following the corresponding machine. The describes the key point of optimal hyper-plane in SVM. Then, the user interest correlation of the test page is gotten by using the probability statistics method based on the optimal hyper- plane.

SVM method is a machine learning algorithm with good performance, however it requires more complex training process and parameter settings. Moreover, as a binary classifier, SVM needs positive cases and negative cases to find the optimal hyper-plane of a user profile. For a user profile, whether submitted by the user or automatically collected by the system, the user interest information are all associated with the user profile, i.e., positive examples of user profile. Therefore, to use the SVM method to create user profile model, we have to first construct negative examples of user profile for training.

C. Domain Knowledge Modeling

Domain knowledge is the background knowledge that we used to enhance the user profile. The source which we have used for preparing Domain Knowledge is DMOZ directory. For preparing Domain Knowledge, first we have crawled the web pages from DMOZ directory for some specified categories, where each category is represented by collection of URL's present in that category.

After crawling, we have extracted the keywords from the crawled web pages. The collections of keywords form the vocabulary for the crawled pages. Now we form a term- category matrix, which specifies weight of each term in each category. The weight may be represented by frequency of the term in that category. Here W_{ij} represents number of times the term t_j is present in Category $Cate_j$.

The matrix may be represented as follows:

Terms/ Category	Cate1	Cate2	Cate3	----- ---	Cate n
T1	W11	W12	W13	----	W1n
T2	W21	W22	W23	----	W2n
T3	W31	W32	W33	----	W3n
'	'	'	'	----	'
'	'	'	'	----	'
Tn	Wm1	Wm2	Wm3	----	Wmn

Table 1 Terms-Category Matrix (TMC)

D. User Profile Modeling

Information is used to indicate user's attention and estimate their objectives for new concerns. User User profile also helps to deal with uncertain concerns. To create the consumer profile, we need to categorize the websites utilized by a person into particular classification. AlchemyAPI has been used for identifying websites. AlchemyAPI categorizes a website by providing it a particular classification along with assurance (numerical value) which reveals its possibility of that belongs to that particular classification. If the website is categorized with assurance above the specified limit level then only we have consider that web page to play a role for that classification .

As we are using DMOZ for qualifications knowledge, we have to map these Alchemy groups to DMOZ groups. Thus in our design, a User User profile is showed as a classification choice vector, where weight of each classification symbolizes user's attention in that classification As proven in the Determine 3.1, users surfing around record is used to build user profile. When the variety of websites searched by the consumer develops above the specified limit, the learning broker up-dates user profile. User attention will thus be showed by fix variety of group's loads. It can be denoted by

$$U = \{cw_1, cw_2, cw_3, \dots, cw_m\}$$

Table 2: Alchemy API to DOMZ Category Mapping

Alchemy Categories	DOMZ Categories
Arts & Entertainment	Arts
Business	Business
Computer&Internet	Computer
Culture & politics	Regional

Where, CW_j will be the variety of websites of classification i frequented by that user, stabilized by most of web page trips among all groups. For demonstrating Client Profile we have utilized Vector Space Model (VSM). We consider all the pages introduce in perusing history of specific client. Each one page relates to a particular archive. The result of vector space model is a term report framework (TDM) which speaks to every site page/record as a peculiarity vector of terms. Here we consider each one archive as a URL.

Table 3 Terms-Document Matrix (TDM)

Terms/ Category	Cate1	Cate2	Cate3	-----	Cate n
T1	W11	W12	W13	-----	W1n
T2	W21	W22	W23	-----	W2n
T3	W31	W32	W33	-----	W3n
⋮	⋮	⋮	⋮	-----	⋮
Tn	Wm1	Wm2	Wm3	-----	Wmn

E. Enhanced User Profile

Upgraded Client Profile is an essential part in our structure. An Upgraded Client Profile enhances the Client Profile by utilizing the Space Information. For setting up the Improved Client Profile we have considered every URL of the Client Profile, match it with Area Learning Urls and add most applicable Urls to the Upgraded Client Profile. Emulating steps clarify the procedure of setting up the Upgraded Client Profile. Perform the accompanying steps for each one archive (URL) in client profile.

- Select the URL from the Client Profile.
- Add the URL to the Improved Client Profile.
- Find the cosine closeness of this URL with the Urls introduce in client particular classifications from the Area Knowledgebase.
- Rank the Urls on slipping request of cosine likeness.
- Recover beat 20 Urls.
- Ascertain the normal of the cosine similitude of these main 20urls.
- From the main 20 Urls add just those Urls to the upgraded client profile whose closeness quality is over the normal worth.

To abridge the procedure, for every URL (structure client profile) most pertinent Urls from the client particular Space Information class are added to plan improved client profile. The cosine equation utilized for the likeness of the URL u in Client Profile to each one pages d_j in Area Learning is as per the following:

$$\text{Cosine}(d_j, u) = \frac{\langle d_j * u \rangle}{\|d_j\| * \|u\|}$$

A cosine closeness measure is the plot between the page in Client Profile u and the report vector d_j.

VI. CONCLUSION

In this paper, we have implemented a framework for personalized web search using User Profile and Domain Knowledge. Based on the User Profile and the Domain Knowledge, the system keeps on updating the user profile and thus builds an enhanced user profile. This Enhanced user profile is then used for suggesting relevant web pages to the user. The proposed framework has been implemented by performing some experiments. These experiments shows that the performance of the system using enhanced user profile is better than those which are obtained through the simple user profile. Our work is significant as it improves the overall search efficiency, catering to the personal interest of the user's. Thus it may be a small step in the field of personalized web search. The user can search the information from different domain. We have implemented to suggest URL from the user searching domains. In this framework we applied re-ranking concept for providing the relevant URL to the user browsing history. And the

user browsing history is very personalized. For use of this web search engine less time & bandwidth and easily get relevant information.

REFERENCES

- [1] M Speretta and S Gauch, "Personalized Search Based on User Search Histories", Proceeding Of International Conference on Web Intelligence, pp.622-628, 2005.
- [2] 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.
- [3] C Liang, "User Profile for Personalized Web Search", International Conference on Fuzzy Systems and Knowledge Discovery, pp. 1847- 1850, 2011.
- [4] 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.
- [5] 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. 701 -712, 2010 .
- [6] O. Shafiq, R. Alhajj and I. G. Rokne, "Community Aware Personalized Web search", International Conference on Advances in Social Networks Analysis and Mining, pp. 3351 -355, 2011
- [7] Rakesh Kumar and Aditi Sharan "Personalized Web Search Using Browsing History and Domain Knowledge", International Conference on Intelligent Computing , pp.493-497, 2014

