

Provision of Relevant Results on Web Search Based on Browsing History

¹G Rama Krishna, ²G Harika Priya, ³V Sravani, ⁴Y Sai Chandu, ⁵G Sai Srikanth

¹ Department of computer science Engineering, KL University, Andhra Pradesh

Email: ¹ramakrishnag_cse@kluniversity.in, ²gharika52@gmail.com, ³sravani.v2899@gmail.com, ⁴chanduyadama63@gmail.com, ⁵sai.gatram@gmail.com

Abstract

Different users submit a query to a web search engine with different needs. The general type of search engines follows the "one size fits all" model which is not flexible to individual users resulting in too many answers for the query. In order to overcome this drawback, in this paper, we propose a framework for personalized web search which considers individual's interest introducing intelligence into the traditional web search and producing only relevant pages of user interest. This proposed method is simple and efficient which ensures quality suggestions as well as promises for effective and relevant information retrieval. The framework for personalized web search engine is based on user past browsing history. This context is then used to make the web search more personalized. The results are encouraging.

Keywords: Personalized Web Search, Information retrieval

1. Introduction

Browsing for information has been made simple due to the development and increase in efficiency of the World Wide Web (WWW). In the past, the only way to gather information was to go to a library or to buy books that might contain the particular information. Now a day, with the increase in search engines and personalized computers, we can access information in an easier manner. The use of search engines like Google and Yahoo are used to make this process simple. However, these search engines might not give us the information that we seek in a simple way.

Generic search engines cannot identify the needs of different users. It is difficult to let the search engine know what the user actually want. Suppose we wanted to find some information about Cornell. User types the keyword "Cornell" and expects that "www.cornell.edu" would be the most relevant site to our query. As there may be millions of pages on the web using the word Cornell, all those results will be displayed to the user. But using personalized search engine, exact information can be retrieved according to user's need by using the users past browsing history.

Generally the personalized web search can be categorized into two types, one is Click based method and other is Profile based method [6]. Click based method mainly concentrates on the clicked pages of the users browsing history. Profile based method is implemented by using user interest model that is generated by user profiling [7]. This method improve the search experience with complicated user-interest models generated from user profiling techniques But this method is unstable in some situations though it performs effectively for any sort of query. There are three types of approaches for Profile based Personalized Web Search[5] systems:(1) systems that uses relevant feedback, (2) systems in which user need to register their interests or demographic information and (3) systems that recommends information to the user's based on browsing history. In first approach, users have to register their personal information such as name, e-mail id, etc., and users have to provide feedback on relevant or irrelevant judgments which is time consuming. The discovery of patterns from the usage data by itself is not sufficient for performing the personalization. In the second technique, user should give their interests and its ratings on a scale from bad to good. This technique is time consuming and users prefer easier methods. So, the third technique is better than others. In this, User's browsing history allows each user to perform more fine-grained search by capturing changes of each user's preferences without any user effort. Users can find the relevant snippet in personalized search results in less time when compared to original results. Profile based method does not support

the run time profiling and may not help to improve the quality of search for some complex queries [8]. As there are different approaches for profile based personalized web search but it does not support the runtime profiling. In profile based method, the information is gathered from browsing history, query history, click through data bookmarks and user documents. This collected data may easily reveal the users private life. Click based method is straight forward and it can perform consistently and considerably well.

There are three approaches for building the user profile to present an individual user's preference [9]. They are Rocchio method, K-Nearest method and Support Vector Machines. It was found that K-Nearest method is better than others in terms of its efficiency and robustness. The Rocchio algorithm is based on a method of relevance feedback found in information retrieval systems which stemmed from the SMART Information Retrieval System around the year 1970. Like many other retrieval systems, the Rocchio feedback approach was developed using the Vector Space Model. The algorithm is based on the assumption that most users have a general conception of which documents should be denoted as relevant or irrelevant. Therefore, the user's search query is revised to include an arbitrary percentage of relevant and non-relevant documents as a means of increasing the search engine's recall, and possibly the precision as well.

K Nearest Neighbour (KNN) is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure (e.g., distance functions). KNN is one of those algorithms that are very simple to understand but works incredibly well in practice. Also it is surprisingly versatile and its applications range from vision to proteins to computational geometry to graphs and so on. KNN is one of the **top 10 data mining algorithms**. KNN is a *non parametric lazy learning* algorithm. When you say a technique is non parametric, it means that it does not make any assumptions on the underlying data distribution. This is pretty useful, as in the real world, most of the practical data does not obey the typical theoretical assumptions made. It is also a lazy algorithm. This means that it does not use the training data points to do any *generalization*. In other words, there is *no explicit training phase* or it is very minimal. This means the training phase is pretty fast. Lack of generalization means that KNN keeps all the training data. More exactly, all the training data is needed during the testing phase. This is in contrast to other techniques like SVM where you can discard all non support vectors without any problem. Most of the lazy algorithms especially KNN make decision based on the entire training data set. KNN classification is one of the most fundamental and simple classification methods and should be one of the first choices for a classification study when there is little or no prior knowledge about the distribution of the data. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible which is shown in figure 1. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on.

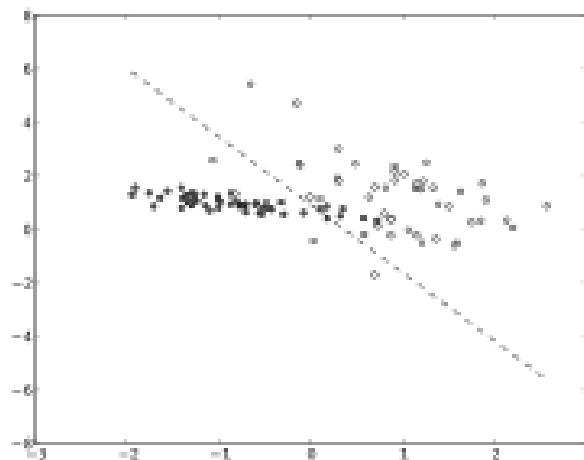


Figure 1. SVM Model

Support vector machines (SVM) are a group of supervised learning methods that can be applied to classification or regression. As there are different techniques for personalized web search but it is still unclear whether personalization is consistently effective on different queries for different users and under different search contexts. Click based personalized web search provides the relevant results to the user in less time and no need of registering the user details before. Click based method is user friendly. The rest of the paper is organized as follows: In section II, related work is presented. In section III, the existing methods for personalized web search and disadvantages of the existing system are presented. In section IV, the proposed method for personalized web search with experimental results are presented. At last, we draw some conclusions and present some future work in section V.

2. Related Work

Rakesh Kumar et al., [1] proposed framework for personalized web search, which considers user's interest to suggest the relevant pages. This framework uses user's profile and domain knowledge. User Profile is constructed by using data of web usage of the user and Domain knowledge is used to store information about different categories and different domain. Information obtained from user profile is classified according to categories and domain. DMOZ directory is used as source for categorizing the data. By using this domain knowledge, Enhanced User Profile is developed. For this, each URL in the user's profile is matched with URL present in Domain Knowledge and the most relevant URL's are added to Enhanced User Profile. Cosine similarity is used to find matching URL's from user profile and domain knowledge to update the Enhanced User Profile. Rakesh Kumar et al., [1] shown different experimental results for personalized web search using Enhanced User Profile and concluded that the performance of the system using Enhanced User Profile better than the performance by using through simple user profile. Micro Speretta et al., [2] used another approach for building user's profile based on user's interaction with particular search engine. This approach build a wrapper around search engine, which collects user's search activity information like queries submitted and snippets visited. Two separate profiles were created based on queries submitted and snippets of user selected results. These two profiles were used to calculate a new rank of results browsed by users and average of this rank compared with Google's rank. They performed various experiments and found that user profile built from queries produced improvement of 33% in the rank of selected result and that of built from snippets is of 37%. S. Vanitha [3] introduced technique which contains separate modules for user profile and click through data, which are mapped for more personalized web pages. Weights of terms are used in different categories to identify the interest of the user. It suggests the extraction of key terms to group with the weight. Then weight and group is to be compared with different categories to store in database in order to find user's interest. The other module which is Click through data module uses Bayesian classification algorithm to generate patterns to find frequently needed information. Then results of these two modules are compared to search more personalized web pages. F Liu et al., [4] presented a technique to create user's profile which can be constructed automatically from user's search history. This uses user's search history for learning user profile and category hierarchy which can be used for learning general profile and then these two profiles are combined to categorize user's query to represent user's need for searching.

3. Existing System

With the advancement of World Wide Web, web search tools have contributed a considerable measure in seeking data from the web. They help in discovering data on the web fast and simple. In any case there is still opportunity to get better. Normal web search tools don't consider particular needs of user and serve every user just as. It is hard to tell the web index what we the user really need. Nonexclusive web crawlers are taking after the "one size fits all" model which is not versatile to individual users.

When different users give same query, same result will be retrieved by a typical search engine, doesn't consider which user submitted the query. This might not be suitable for users which require different information. While searching the data from the web, users need the data based on their interest. For the same keyword the different users might require different piece of information. For this type of query, a number of documents on distinct topics are returned by

generic search engines. Hence it becomes hard for the user to get the relevant information and it is time consuming also.

Disadvantage:

- Typical search engines will return same result for different users when same query is given.
- It becomes hard for the user to get the relevant content.

4. Proposed System

We propose a methodology for customized web look which considers singular's enthusiasm and improves the user web seek by recommending the applicable pages of his/her advantage. The personalized web search utilizes user browsing history and it makes good use of the advantages of popular search engines, it can re-rank the results obtained by the search engine based on the user hits to the link. Our personalized web search engine provides search facility to the users in two different ways depending upon the type of search.

The two different ways through the searching is done are

- Customized local search
- Normal web Search

User should select the search type beforehand. For each search type, the user can select the mode of search. There are four different modes in each search type. They are Text, Image, News and Videos. If the user selects the text mode, the information is retrieved which is related to text only. Customized local search provides the relevant results to the user based on the user's interest. Whereas, the Normal web search is similar to the general search engines.

4.1 Customized Local Search

Search Local is one type of search in Personalized Web Search Engine. In this Searching the information is done using the local database i.e., it considers all the users in the Local Area Network (LAN). Local search type provides ranking to the each URL, based on the user hits the ranking of each URL changes. For example, let us consider the word "apple iphone" which we are going to search in the local search type then the results will be displayed to the user as figure 2 (a). The search result consists of Rank, Title of the page and URL of the page for the given keyword.

Search Results

Rank	Title	URL
1	iPhone - Shop iPhone Unlocked or Select Carrier - Apple Store (U.S.)	http://store.apple.com/us/iphone
2	Apple iPhone 5 News, Features & Release Date PCMag.com	http://www.pcmag.com/apple-iphone/
3	Apple iPhone Economist - World News, Politics, Economics ...	http://www.economist.com/topics/apple-iphone
4	Shop Apple iPhones including iPhone 6, 6 Plus, 5s & 5c from AT&T	http://www.att.com/shop/wireless/devices/apple/iphone.html
5	Apple - iPhone - Compare Models	https://www.apple.com/iphone/compare/
6	Apple iPhone 5c Verizon Wireless	http://www.verizonwireless.com/smartphones/iphone-5c/
7	Apple iPhone - Radio Shack	http://www.radioshack.com/cell-phones/apple-iphone
8	iPhone 7 release date rumours, features and images - News ...	http://www.macworld.co.uk/news/iphone/iphone-7-rumours-roundup-release-date-new-features-images-3530504/
9	iPhone - Wikipedia, the free encyclopedia	http://en.wikipedia.org/wiki/IPhone
10	iPhone - Apple Support	https://www.apple.com/support/iphone/
11	Apple iPhone 6 Plus - T-Mobile	http://www.t-mobile.com/cell-phones/apple-iphone-6-plus.html
12	All Apple phones - GSMARena.com	http://www.gsmarena.com/apple-phones-48.php
13	Apple - iPhone	https://www.apple.com/iphone/
14	Apple iPhone 6 Plus - Best Buy	http://www.bestbuy.com/site/iphone/iphone-6-plus/pcmcat342300050018.c%3Fid%3Dpcmcacat342300050018
15	Activation Lock - iCloud	https://www.icloud.com/activationlock/

Figure 2(a). Local search type results

As the rank of each URL changes based on the user hits, if user clicks the URL (<https://www.apple.com/iphone/compare/>) which was ranked as five for the first search as shown in figure 2(a) then from the next search that page will be searched first with rank one as shown in figure 2(b). This enables the user to open the interested link in less time without searching for the viewed link. It returns the results in less time and there is no need to register the user interests before.

Search Results		
Rank	Title	URL
1	Apple - iPhone - Compare Models	https://www.apple.com/iphone/compare/
2	iPhone - Wikipedia, the free encyclopedia	http://en.wikipedia.org/wiki/IPhone
3	iPhone - Shop iPhone Unlocked or Select Carrier - Apple Store (U.S.)	http://store.apple.com/us/iphone
4	Apple iPhone 5 News, Features & Release Date PCMag.com	http://www.pcmag.com/apple-iphone/
5	Apple iPhone Economist - World News, Politics, Economics ...	http://www.economist.com/topics/apple-iphone
6	Shop Apple iPhones including iPhone 6, 6 Plus, 5s & 5c from AT&T	http://www.att.com/shop/wireless/devices/apple/iphone.html
7	Apple iPhone 5c Verizon Wireless	http://www.verizonwireless.com/smartphones/iphone-5c/
8	Apple iPhone - Radio Shack	http://www.radioshack.com/cell-phones/apple-iphone
9	iPhone 7 release date rumours, features and images - News ...	http://www.macworld.co.uk/news/iphone/iphone-7-rumours-roundup-release-date-new-features-images-3530504/
10	iPhone - Apple Support	https://www.apple.com/support/iphone/
11	Apple iPhone 6 Plus - T-Mobile	http://www.t-mobile.com/cell-phones/apple-iphone-6-plus.html
12	All Apple phones - GSMarena.com	http://www.gsmarena.com/apple-phones-48.php
13	Apple - iPhone	https://www.apple.com/iphone/
14	Apple iPhone 6 Plus - Best Buy	http://www.bestbuy.com/site/iphone/iphone-6-plus/pcmcat342300050018.c%3Fid%3Dpcmcat342300050018
15	Activation Lock - iCloud	https://www.icloud.com/activationlock/

Figure 2(b). Local search results after re-ranking

The customized local search provides the user four facilities:

- Search text
- Search image
- Search news
- Search video

After typing the keyword in the search box, user can select the type of the search results. That is, user can select the text icon below the search box to get the results that are related to text. If the user selects the image icon then the image URL's only retrieved to the user. Similarly, the videos and news icons. The customized web search provides the relevant results for the keyword based on the user's interest.

4.2 Normal Web Search

Normal web search is another search type in personalized web search engine. It is similar to the general search engines. Normal web search type provides URL's to the user along with the description of the keyword. The most relevant results will be retrieved to the user. Normal web search also provides four search facilities to the user. They are search text, search image, search video, search news. This search type retrieves the most relevant results within the less time. If user selects the text icon or an image icon the results will be retrieved quickly which are in the format text or image only, same as the Google search engine. If user selects the video icon, only the video format results are retrieved with URL below the video which is shown in figure 3. So, the user can easily open the links without searching for the video format results.

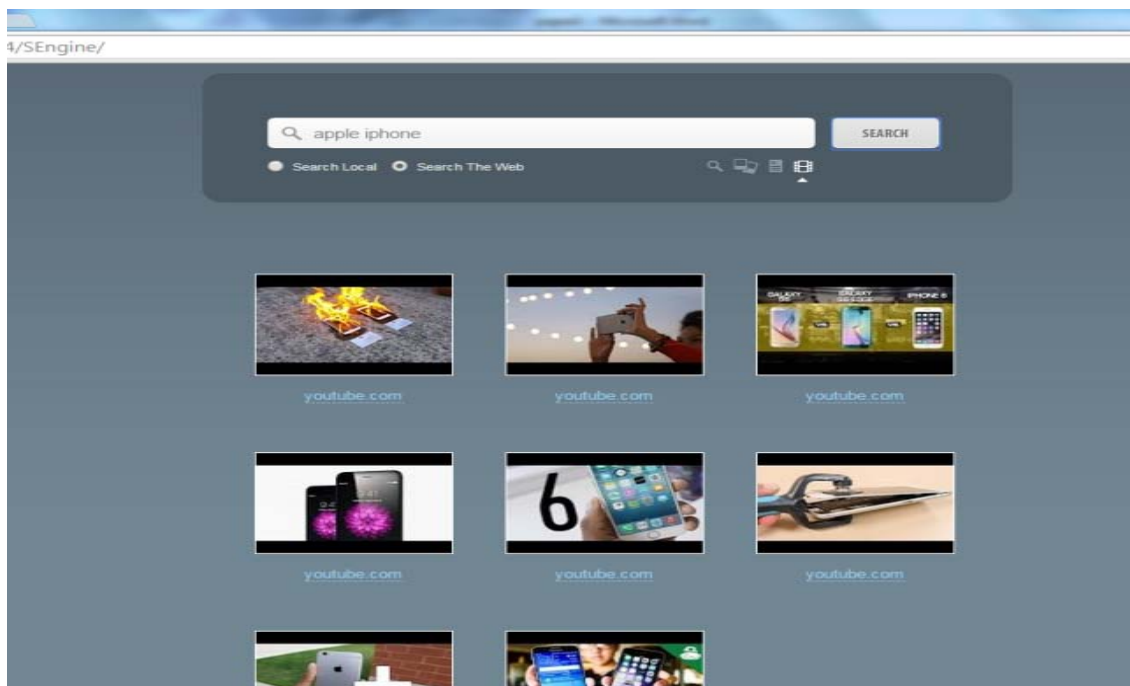


Figure 3. Normal web search results for video type

Advantage:

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.

5. Conclusion

In this paper, we propose a model for personalized web search which will consider individual's past browsing history. Personalized web search exploits a number of techniques to upgrade search quality in addition to page rank, anchor text, and proximity information. The proposed structure has been executed by performing a few tests. Our work is critical as it enhances the general pursuit effectiveness, coddling the individual enthusiasm of the user's. In this way, it perhaps a little venture in the field of customized web seeks. In future this schema may be petitioned re positioning the site pages recovered via web crawlers on the premise of user needs.

References

- [1] Rakesh Kumar & Aditi Sharan. "Personalized Web Search using Browsing History and Domain Knowledge". *Issues and Challenges in Intelligent Computing Techniques (ICICT), 2014 International Conference*. 2014: 493-497.
- [2] M Speretta and S Gauch. "Personalized Search Based on User Search Histories". *Proceeding Of International Conference on Web Intelligence*. 2005: 622-628.
- [3] S Vanitha. "A Personalized Web Search Based on User Profile and User Clicks". *International Journal of Latest Research in Science and Technology*. 2013; 2(5): 78-82.
- [4] F Liu, C Yu and W Meng. "Personalized Web Search for Improving Retrieval Effectiveness". *IEEE Transactions on Knowledge and Data Engineering*. 2004; 16: 28-40.
- [5] Sandesh Pare, Bharati Vasgi. "Personalization of the Web Search". 2014; 4(10).
- [6] Sandesh S Pare, Bharati P Vasgi. "A Short Review on Different Personalization Schemes". *International Journal of Engineering Research and Development*. 2014; 10(3).

-
- [7] Snigdha Gupta, Saral Jain, Mohammad Kazi, Bharat Deshpande, Mangesh Bedekar, Komal Kapoor. "Personalization of Web Search Results Based on User Profiling". *First International Conference on Emerging Trends in Engineering and Technology*, 978-0-7695-3267-7/08. 2008 IEEE.
- [8] Indu Chawla. "An overview of personalization in web search". 978-1-4244-8679-3/11. 2011 IEEE.
- [9] C Liang. "User Profile for Personalized Web Search". *International Conference on Fuzzy Systems and Knowledge Discovery*. 2011: 1847-1850.