



# User Centric Web Page Recommender System Based on User Profile and Geo-Location

**Abdulhamid.Y.Endris<sup>1</sup>, Meeta Kumar<sup>2</sup>**

<sup>1</sup> PG Student, Computer Science Department, Symbiosis Institute of Technology, Pune, India

<sup>2</sup> Asst. Prof., Computer Science Department, Symbiosis Institute of Technology, Pune, India

<sup>1</sup>[abdulhamid.endris@sitpune.edu.in](mailto:abdulhamid.endris@sitpune.edu.in), <sup>2</sup>[meeta.kumar@sitpune.edu.in](mailto:meeta.kumar@sitpune.edu.in)

**Abstract:** *A web recommender system assists a user to manage the vast amount of search results suggested by a search engine when surfing the Web with the help of recommender system agents. In this case the recommender system will reduce the effort a user has to spend in searching and navigating to important web pages or web documents and simultaneously increases the user ability to find more relevant documents. When same query is used by different users in finding useful information on the World Wide Web (WWW), the search engines return the same result for all users regardless of who submitted the query. But in a real time situation each user may have different information needs for the same query. Therefore, the search results should be based on the user information needs.*

*In this paper we propose a recommender model which recommends web pages to a user based on the user profile and geographic location of the user. We propose a web-recommender model based on user location and a personalized search method to improve the search strategy and the quality of the results. We first create a user profile, identify the user location and then return the result to the user based on user query considering the user profile and location of each user. After that the web is crawled moving from link to link and an index page is constructed that includes the user keywords used as a query in the search. This index references with a user profile and a user search query. The recommender system, then lists the pages that have the same keyword or related with the keyword that were written as the user's search terms and returns the ranked results in reference to the user profile and user location.*

*Trying out multiple sites while looking for information is a basic concern in web page recommendation systems. The engine looks for keywords throughout each Web page, and returns the result based on the matching with the user profile, frequency of the key word on the page and based on the appearance of key words in the title of the web page or in the URL directly.*

**Index Terms—** *User profile, Location detection, personalized search, Search query, Crawler.*

## I. Introduction

Finding useful information on the World Wide Web is a common task for almost all internet users. One of the available services on the internet through WWW can be searching information or knowledge from the web pages. According to Netcraft Internet Research[11], Anti-Phishing and PCI Security, a company which provides a monthly internet research report, the number of hosts found worldwide running internet web sites by the Web Server Survey is very large [over 180 million in October 2008]. Hence it is getting increasingly difficult for the web users to locate relevant information on the internet. The results returned to the user by the search engine / browser may not typically be the information expected by him/her. A web recommender system is thus aimed to refine web search results and address this issue. Since the volume of information on the web increasingly grows, it is difficult for searching information on WWW. Therefore, we get the feeling of being stunned in selecting of service choices. As one of the most hopeful approaches to ease this burden, recommender systems have begun in

different domains, including E-commerce, digital libraries, and in any knowledge management concerns. These recommender systems offer personalized recommendations based on user favorites. The recommender system takes the user query, relate the query with the user profile and then search the information on the web and then provide relevant detailed information to the user. Search engines recommenders have two major functions, i.e. crawling and building an index, and providing a ranked list of search results based on the user query. When a person does an online search, the engine uses a crawler and scrubs its mass of billions of documents and it returns only those results that are relevant or useful to the searcher's query with ranked results might be according to number of keywords or according to the popularity of the websites serving the information. The crawling approach is based on tree structure and can effectively represent the websites that need to be accessed. It can easily find different transition links between Web pages.

In the past web search recommender systems, search results are based on the web popularity or based on the relevance of the website rather than focusing on the user interests and needs. In previous cases, mostly the concentration focuses on the engines, itself not in the user side. Even if the user tries too many times to get the information, the engine returns the result without regard to what people/user really want, what they want to use, and how best it should fit their needs. Therefore, the user confuses in identifying or in searching relevant information, the user takes more time to get information which is more related to his or her interest of searching. This situation can be explained as follows: two different users may need to search a hotel for having lunch, let's say the first user is located in Arthur road, Mumbai, India and the second user is located in FC road, Pune, India. While they need to search a "hotel" they can get the same information, may be results of number of hotels around the world. But that will not help them, it is difficult for the user to get their pertinent content and it is also time consuming. This is also the same as while the user which is a medical professional and another user which is a software engineer by profession try to search information about "virus". In both cases the user requires different piece of data for the same keyword. Personalized web recommendation is considered as a gifted solution for solving these problems, since different information can be obtained depending upon the users profile and their location.

The concept of personalized web search plays a great role to improve the overall experience of the users. In this personalized search the focus is on optimization for users, instead of centering on search engines alone. Therefore, it motivates us to focus more on creating an interactive, personalized search for the benefit of users, working on getting related information for improving search engine results hastily. The focus has now diverted towards building more useful and satisfying experience for the user, and quite rightly so. We propose a personalized search engine while assuming that there could be other links that are more appropriate to your search but that will lie outside of the users circle, this could be taken as a limitation. The system could lead to a great impact in the quality of the search results. Our model focuses on the user profile, and Location based personalized search system and the recommender system through the engine side.

## II. Related Work

Web page recommendation became a major requirement in web information searching, wherein different machine learning algorithms used for the links to the recently viewed web pages are shown. Google uses ranking algorithm for searching web pages that contain the keywords to be searched, and simultaneously extracts the key words from the web pages and then based on the extracted keywords frequency in the web pages a rank to each page has been given. The best matched link with the highest frequency of searching key word comes first in the Google search recommendation. Google Search engine ranking algorithm works by importance of the website which is decided by counting the links attached to that website. The link counting may take a process of iteration and finalizes the algorithm with a ranking result. There are web page recommendation systems done using web log file and knowledge representation. User Centric based personalized web recommending is a better and automated way of providing search result of web pages concentrating on the user's information needs.

Web usage mining is a proved approach that captures user preferred navigation patterns using web usage log data [1]. Search engine optimization is a process that improves the rank of websites and web pages, has explained tips in search engine optimization to Improve Google Ranking of Website [8]. M. VenuGopalachari, Po Sammulal [2] proposed web page recommendation strategy which uses domain knowledge with the construction of ontologies and web usage data by means of the web log file. It is a prediction model that recommends web pages for an individual user based on the corresponding usage information or semantic log history. A novel method to provide a better webpage recommendation based on Web usage and domain knowledge [3]. It uses an ontology based model that represents the domain knowledge of a website, a semantic network that represents domain knowledge, and a conceptual prediction model, which is integrating Web usage and domain knowledge based on the frequently viewed Web pages. The recommended strategies are done using the domain knowledge and the prediction model to predict the next pages with probabilities for a given Web user based on his or her current Web page navigation state.

Qingyan Yang et al.[13]Propose a novel personalized Web page recommendation model called PIGEON (abbr. for PersonalIzed web paGerEcommendatiON). They employ a collaborative filtering framework for personalization and formulated a bipartite graph based method to calculate user similarities Ujwala H.Wanaskar et al.[15] propose a hybrid recommendation approach for better results. In their recommender system, they combine techniques of collaborative approaches and content based approaches to avoid some limitations and problems of pure recommender systems, like the cold-start problem.

RuimeiLian [4, 5] proposes personalized e-commerce recommendation system to construct an online recommendation model on commercial websites. The recommendation process has two parts: offline part and online part. In the online part, with the help of search engine, dynamic Web page recommendation is done on the base of what the user currently browses during the client session. The offline part consists of three phases: data pre-processing, collaborative filtering and data mining. For the online part, the main task is the online page recommendation. When the user visits a website, the recommendation system gets his cluster pattern after user identification. The next step is to recommend page sets. The address of new pages which may be interesting for the user are collected and attached to the bottom of the currently requested page. Kazunari Sugiyama et al. [16] propose several approaches that can be used to adapt search results according to each user's information need. Then compare the retrieval accuracy of proposed approaches including user's browsing history. AviralNigam[7] Compares the heuristic approach to techniques of web crawling. The comparison has done on the simulated web environment and the time to search is the time taken to reach the most relevant page from any initial page.

Mohamed-K HUSSEIN et al. [6] Present and analyses the currently important algorithms for ranking of web pages such as PageRank and Weighted PageRank. And then finally proposes a ranking algorithm based on Topic-Sensitive PageRank and Weighted PageRank. The proposed algorithm provides a scalable approach for search rankings using Link analysis. Zhu and Wu [17] proposed a research analysis on search engine optimization by using reverse engineering factors and built a system that automatically crawled 200 thousands web pages. After that they analyzed the page rank, URL and HTML based Google search results.

### III. Contribution and Methodology

In this Section, we first introduce some basics and preliminaries in our proposed model and then further explain the proposed methodology.

#### a. User Profile

The user profile is a record of user information obtained from user feedback and which is stored in the server to define the user's characteristics and behaviors. This helps the users to have a better access of required services from different places and different machines or systems

In searching information the users need highly depends on the profile of the user. Therefore the search recommender system should have to consider the user preference or profile in searching information to improve the users need.

#### b. Location Identification/Detection

Location based information searching in www is one of the important tasks what we expect from search engines. In location based search the main search focus is not only based on the user's search query, but also on the geographical location to where the query is related to. Therefore, in location based searching, searching information and retrieving the result is based on the user location. Location of a user can be expressed in terms of country, state, city, postal code and street name.

In personalized web search while the user login to the system the location of the user is automatically detected using the browsers 'geoAPI' capability and displayed to the search page. No need to store the location in to the database as other user profiles because the user location may vary with in a short time. Therefore, the detected location will be applicable for one client session only. Every time the user logs in to the system, the location gets automatically updated.

#### c. Web Crawler and Page Indexer

A crawler is a system which used to retrieve or capture all the links found in the web page from WWW in an automated way. Web Crawler is used by search engines to download pages from the web, and to index it for further processing in the search result to provide fast search. Without using crawlers, it is very difficult for search engines to exist. The Web is a vast directed graph, with documents as vertices and hyperlinks/ links as edges. Therefore, it is essential to explore the graph using proper graph traversal algorithms [12].The crawling and indexing process is explained in diagram as follows.

#### d. Personalized Search and Search Query

Personalized search is a mechanism used to allow individual users to search information, and providing the result based on the individual users need. In personalized search all users should have their own user profile and user account to login to the system and provide a search query. In personalized search, the user search query can be modified relating to the user profile and then the search results ranking also modified. The search query is a term or a phrase which can be entered by the user as a criterion for searching information in a web.

#### e. Keyword Construction and Keyword Frequency

Keyword is a high level summary of a document, presenting the overall content and topic of the document[10].Keyword construction is the process of extracting terms that mainly describe the matter of a document or text. Keywords are words or terms in a document which can represent relevant information and describe the main topic about the document. Keyword

construction is an important task in mining and information retrieval (like search engines). We construct a keyword from each web pages discovered by the crawler and stored it in the database for further search processing. The keywords stored in the database are stored with their corresponding links or web pages where exactly it comes from or obtained. Keyword frequency shows the number of times or how often a keyword or phrase appears on a website page. The frequency of keywords is an important factor in ranking of a search result.

All this preliminary process can be summarized in the figure.

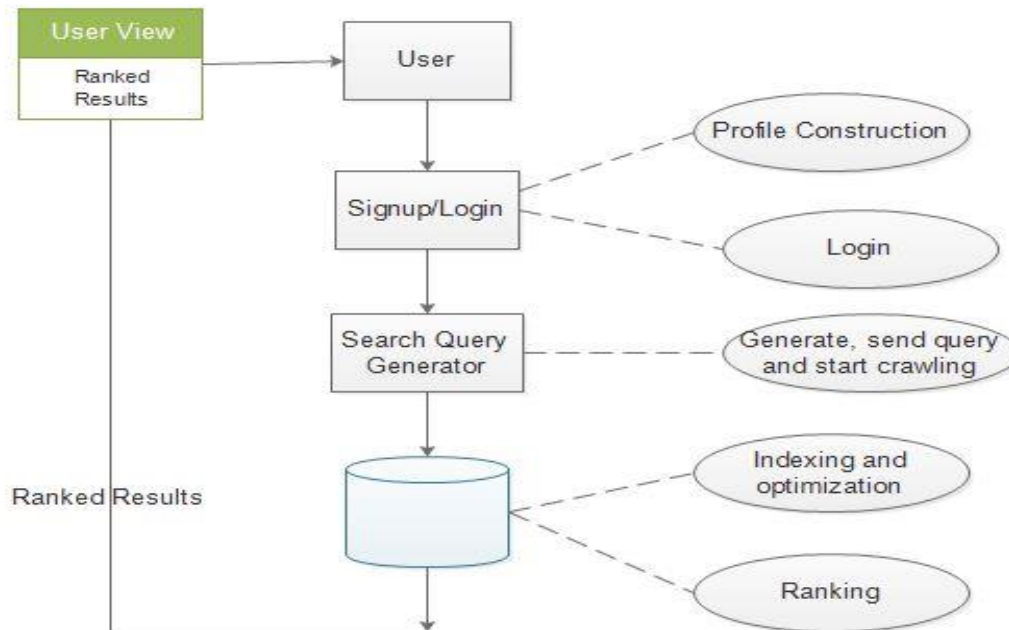


Figure 1 Process Overview

#### IV. Recommendation Process

##### *Personalized Web Page Recommendation Steps*

In this personalized web page recommendation system, the users first have to construct their profile, and then they need to log into the system using their username or email and password. When a user sends a log in request to web server, session is created for the user. During session when user logs in to the system the user profile and location of the user is detected automatically and stored as a session data. A user session can be either as a single transaction of many page references, or a set of many transactions each consisting of a single page reference[9,14]. The user location which is detected automatically displayed in the redirected page. On the same page, the user search interface is available and the user can use any search query for searching the information with available user profile options to be included in the search. Then the recommender system returns the result based on the selected user profile or location. This can be described in pseudo code as follows.

##### *Pseudo Code and Algorithm*

- i. Initialize Google Hosted Libraries // to serve users directly from Google's network of datacenters.
- ii. Make the document ready to manipulate a page safely
- iii. Focus on the Search Field and take search keywords entered by the user
- iv. Consider the user who wrote the search query
- v. Extend the configuration and users search query
- vi. Set the URL of Google's AJAX search API and send the query
- vii. If response data / result available, append it to the results list
- viii. Display the result until all values are shown or up to the maximum length.

The basic outline of the algorithm is as follows:

```

config = { type, results per page, page number};

$('#id'). Focus ();

search.term = search.term || $('#id').val();

Search.term = search.term + "profile data";

send API URL+search type+ Search.term;

results = response.results;

if results.length

    for i=0;i<results.length;i++

        append new result(results[i])

    end for

    results.display() { Tiltle, Description/ publisher, url}

end if

else

    results.empty() { no results found for the search term/query}

end else

```

Figure 2: Outline of Algorithm for Recommender System

## V. Discussion

Since we have designed a user centric based personalized web page recommendation system, we have a user friendly interface for using the system. In this section, we discuss the results obtained using personalized web page recommendation techniques compared with normal search engine search results. We have done a sample search experiments using many search terms and here we present using two search terms *i.e* "Hotel" and "Football Club". We have used this search terms in both Google search technique and personalized web page recommendation technique. In all search cases both techniques returns different result for the same keyword.

Figure 3 shows the overall Search Interface. Whereas Figure 4 and figure 5 Shows results while searching for "hotel" with no considering user profile and with considering user profile or Location respectively. In figure 4 the result shows similar with obtained in searching for hotel in Google search. But figure 5 show results focuses on the user's location, since the user profile is chosen as a consideration for the recommender system in searching the available information.

Figure 6 and figure 7 Show results while searching for "Football club" with no considering user profile and with considering user profile (Gender, Location) respectively. In figure 6 the result obtained is same as the result obtained while searching in Google. But in the case of figure 7 we have chosen two criteria's *i.e* Gender and Location for the recommender system to recommend or search the available information in the web and the result is definitely different from the results obtained in the first search (*i.e* in figure 6).



Figure 3 Search page interface

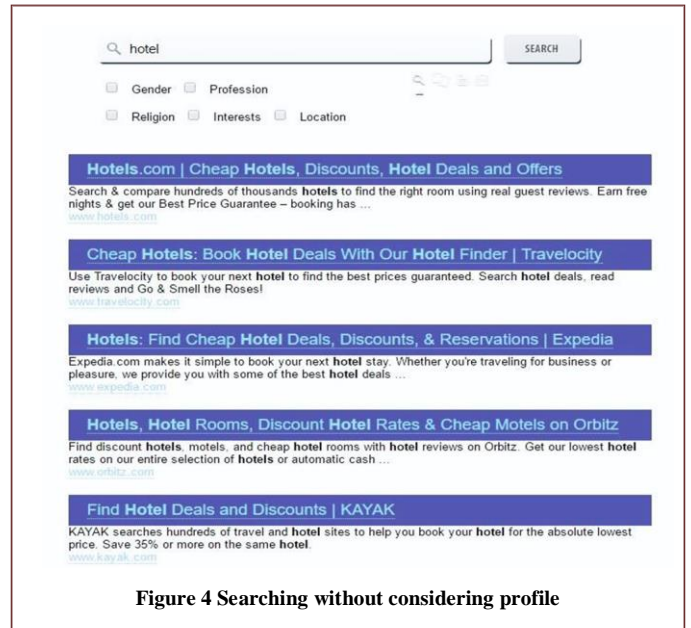


Figure 4 Searching without considering profile

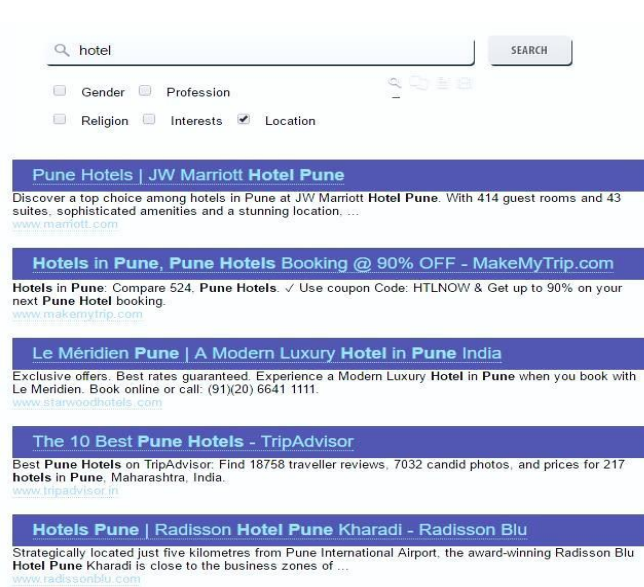


Figure 5 Searching hotel with considering profile (location)

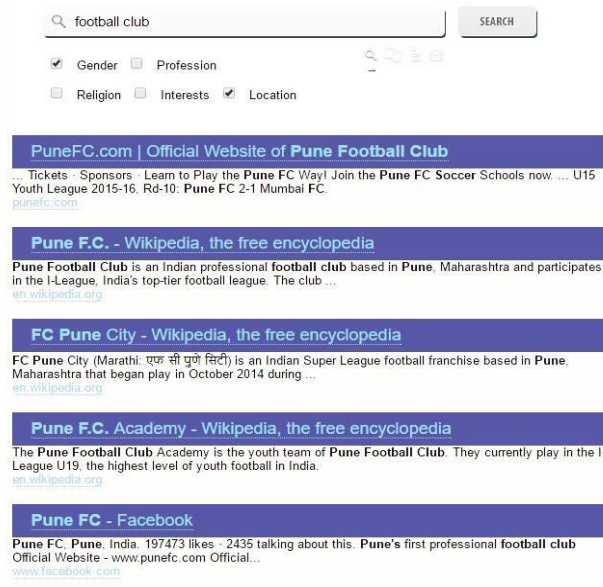


Figure 6 Searching football club without considering profile

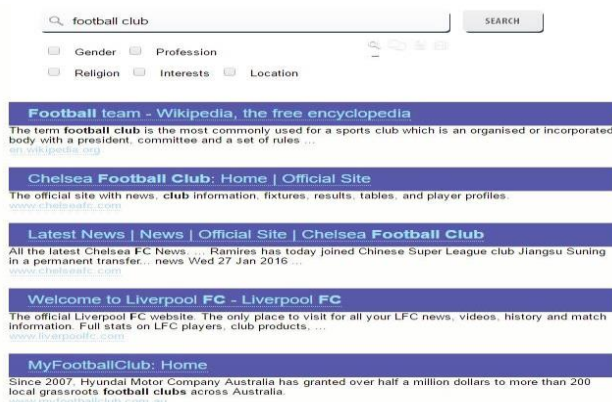


Figure 7 Searching football club with considering profile

## VI. Conclusion & Future Work

This paper proposes an enhanced web page recommendation strategy using a user profile and location. The user is suggested a list of ranked web page results while he/she enters a search key word or while he/she initiates a search query. Results suggested shows better search results are obtained in recommending web pages than the normal search, which makes the user satisfied with the web page recommendations.

As a future work, since there are profiles already constructed in other social media the user profile construction must be taken from those social medias. Therefore, only one profile will work for all systems that needs profile construction. Using a hybrid model combining the proposed methodology with other effective recommendation system may boost the result reporting time. Web pages that do not have geographic information may not be recommended to a certain user.

## REFERENCES

- [1] J Srivastava, R Cooley, M Deshpande, and P.-N. Tan, "Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data," ACM SIGKDD Explorations, vol. I, no. 2, pp. 12- 23, 2000.
- [2] M. VenuGopalachari ,PoSammulal ,”IEEE International Conference on Advanced Communication Control and Computing Technologies (ICACCCT) “ 2014.
- [3] ThiThanh Sang Nguyen, Hai Yan Lu, and Jie Lu, IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 26, NO. 10, OCTOBER 2014.
- [4] RuimeiLian ,”The Construction of Personalized Web Page Recommendation System in E-commerce” 2011.
- [5] XianchenHao, Degan Zhang and Guocheng Yin. Researches of Data Mining Based on E- commerce [J]. Mini-Micro Systems, 2001.
- [6] Mohamed-K HUSSEIN et al.” An Effective Web Mining Algorithm using Link Analysis”, Vol. 1 (3) , 2010, 190-197.
- [7] AviralNigam , “Web Crawling Algorithms”, Computer Science and Engineering Department, National Institute of Technology - Calicut, Kozhikode, Kerala 673601, India.
- [8] Khalil urRehman et al. “The Foremost Guidelines for Achieving Higher Ranking in Search Results through Search Engine Optimization”, Institute of Science and Technology,Islamabad, Pakistan, 2013.
- [9] ModrajBhavsar ,” Web Page Recommendation Using Web Mining”,Int. Journal of Engineering Research and Applications ISSN : 2248-9622, Vol. 4, Issue 7( Version 2), July 2014, pp.201-206.
- [10] Teng PAN et al. “Design and Implementation of Keywords Extraction and Management System Based on Java Platform “, 51th Jinguang Road. Langfang. Hebei. China, 2013.
- [11] [<http://news.netcraft.com/page/10/>] Site operator survey
- [12] Andrei Broder et al [<http://www9.org/w9cdrom/160/160.html>] “Graph structure in the web”]
- [13] Qingyan Yang et al.” Personalizing Web Page Recommendation via Collaborative Filtering and Topic-Aware Markov Model”, IEEE International Conference on Data Mining, 2010.
- [14] BamshadMobasher, Jaideep Srivastava and Robert Cooley “Automatic Personalization Based on Web Usage Mining” ARTICLE in COMMUNICATIONS OF THE ACM.
- [15] UjjwalaH.Wanaskar, SheetalR.Vij, DebajyotiMukhopadhyay “A Hybrid Web Recommendation System based on the Improved Association Rule Mining Algorithm”
- [16] Kazunari Sugiyama, Kenji Hatano and Masatoshi Y oshikawa “Adaptive Web Search Based on User Profile Constructed without Any Effort from Users”
- [17] C. Zhu and G. Wu, “Research and Analysis of Search Engine Optimization Factors Based on Reverse Engineering”, MINES Sch. of Software Eng. USTC, China, (2011), pp. 225-228