RESEARCH ARTICLE

Design and Implementation of Meta Search Engine Using Vector Space Model for Personalized Search

Mr. Ishwar.N.Bharambe[1], Prof. Dinesh.D.Patil[2]

[1]P.G.Student in SSGBCOET, Bhusaval, Lecturer in J.T.Mahajan Polytechnic, Faizpur & NMU, India
[2]Associate Professor & Head of Computer Science& Engg. in SSGBCOET, Bhusaval & NMU, India
[1] ishwar.bharambe@gmail.com; [2] dineshonly@gmail.com

Abstract— This In this paper, we design & implement Personalized Meta Search Engine using vector space model. The search engine is that we tell the machine to learn users' interest, so the personalized meta search engine can help users to pick up the important information for them fast by using their interest keeping in the top of the database using Google & Bing Search engine's. Personalized search engine can sort the results according to users' interest, the results that user likes will be on the beginning of the search links. It is a better to use Vector Space Model to help us implement the personalized meta search engine. We use Vector Space Model to model the user and the results' interest, then we use cosine angel to get the similarity of these interest.

Keywords— User, Search Engine, Personalized Meta Search Engine, User interest

I. INTRODUCTION

The Internet can enable people to get the information more efficiently. On other way, with today's information knowledge in enormous forms, as well as network information into the exponential grows of the trend; the search engines are more essential in our life. Because of the strong benefit of integrating information that makes the results more comprehensive, the meta- search engine is more popular in our day to day life. Because the meta-search engine can get more information from large sources, there is lots of information that users don't think about. This pros turns to cons. It makes user to use more time to deal with the information they are not interested in. Against the background, personalized meta- search engine is one way to solve the problem. The mean of personalization is search engine can help users to sort the important information for them by using user's interest. Personalized Meta Search engine will get the users' interest at the beginning of the of results, so it is very convenient for users to access useful information. In this paper we will introduce the design and implementation of meta-search engine. We model the results and users' interest according to the Vector Space Model. They can put the users' interested information at the beginning of results, so the users can get the information rapidly.

The web search engines generally provide search results without consideration of user interests or context. We propose a personalized search approach that can easily extend a conventional search engine on the client side.

The prime reason for the SEs indexes the pages on the basis of keywords. On the other hand, when we are searching the internet we quite often may not know the correct and complete set of key words that might have led us to the desired url.

This paper suggests a new approach that is based on some algorithms which considers cosine angle similarity and uses them to implement a Personalized Meta-Search Engine.

II. LITERATURE SURVEY

A. In meta-search engine

It will examine the advantage and disadvantage of various approaches. There are three main directions for implementing Meta Search Engine:

1. Growth in user-interface
2. To sort the results of query
3. Consider the algorithms for indexing of web-page.

The more concentration on user requirements is recommended in the architecture of Meta-Search Engine. Personalized Meta-Search Engine has been already proposed that provides quick response with re-ranked results after extracting user preference. It uses Naïve Bayesian Classifier for re-ranking.

Some MSEs use proxy log records for accessing user's pattern and store these patterns in the database. A relevance score is measured using some heuristic for each user and the url that she/he visited. A profile is maintained the user which contains currently visited most relevant urls. Relevance of these urls with their respective relative position is updated in profile when users visit those links further.

Current research also suggests the framework of Meta-search engine based on Agent Technology. An enhanced version of open source Helios Meta-search engine takes input keywords along with specified context or language and gives refined results as per user's need.

All the proposed solutions refine search-results up to some extent but they have a serious drawback which is that the user profile is not stationary from this it is observed that we need to consider alternative methods of re-ranking. This is provided by really statistical methods like Latent Semantic Analysis (it is also called as Latent Semantic Indexing) and the newly introduced Probabilistic Latent Semantic Analysis (it is also called Probabilistic Latent Semantic Indexing) which promises to give results that are more correct than those of Latent Semantic Analysis. Thus, the emergence of these algorithms and the need for robust meta- search engines.

Probabilistic Latent Semantic Analysis (PLSA) give robust results for Information Retrieval when the task is to search the most relevant documents from a given corpus, for a given query. As both of these methods depend on the Vector Space Model, the Vector Space Model is explained prior to both.

B. Vector space model

The most of the text-retrieval techniques are based on indexing keywords. Since only keywords are unable to capturing the whole documents' content, they results poor retrieval performance. But indexing of keywords is still the most applicable way to process large corpora of text. After identification of the significant index term a document can be matched to a given query by Boolean Model or Statistical Model. Boolean Model applies a match that relies on the extent. The Fig.1 represent of the documents Doc1 and Doc2 in space of three terms namely "Information", "Retrieval" and "System". Three are perpendicular dimensions for each term represents "**Term-Independence**". This independence can be of two types namely linguistic and statistical.

When the occurrence of a single term does not depend upon appearance of other term, it is called Statistical independence. In Linguistic independence; interpretation of a term does not rely on other any term an index term satisfies a Boolean expression while statistical properties are used to discover similarity between query and document in Statistical Model.

The statistically based "Vector Space Model" which is based on the theme of placing the documents in the n-dimensional space, where n is number of distinct terms or words (as- t1, t2…tn) which constitutes the whole vocabulary of the corpus or text collection. Each dimension belongs to a particular term. Each document is considered as a vector as- D1, D2…Dr; where r is the total number of documents in corpora. Document Vector can be shown as following: **D**r={d1r,d2r,d3r,……..dnr}
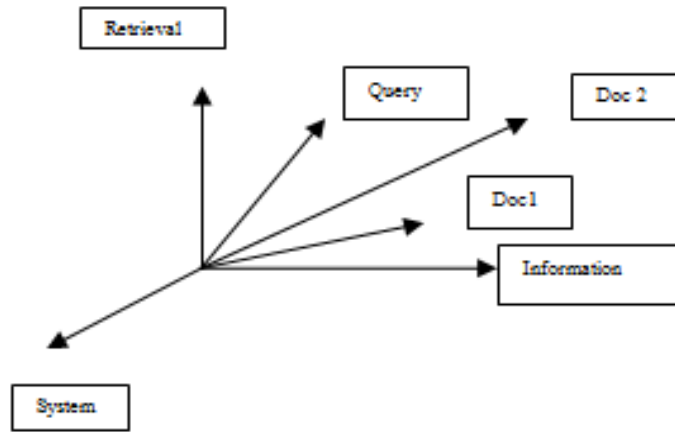
Fig.1. Document Representation in term space

Where dir is considered to be the ith component of the vector representing the rth document. There are various similarity measures that are proposed and one of them, that is very frequently used, is **Cosine Similarity**.

Cos $= \mathbf{Q} * \mathbf{D} / |\mathbf{Q}| * |\mathbf{D}|$

The above expression represents the cosine of the angle between two vectors in the term space. The relevant document will be that one which is the nearest to given query. In the same way two documents would be considered relevant if they are in neighbor-hood region of each other.

The other measure are 1) Inner Product $= Q_j * D_j$

2) Dice Coefficient $= 2 Q_j * D_j / \{ Q_j 2 + D_j 2 \}$

3) Jaccard Coefficient $= Q_j * D_j / \{ Q_j 2 + D_j 2 - Q_j * D_j \}$ Each component of document vector is always associated with some numeric-factor which is called weight of that respective term in document. This weight, wi, can be replaced by term-count or term-frequency (tfi). This assignment leads to another variation of the model that is called "**Term Count Model**".

III. PROPOSED SYSTEM

The proposed system, we propose a content ontology to accommodate the extracted content and location concepts as well as the relationships among the concepts. We introduce different entropies to indicate the amount of concepts associated with a query and how much a user is interested in these concepts. With the entropies, we are able to estimate the effectiveness of personalization for different users and different queries
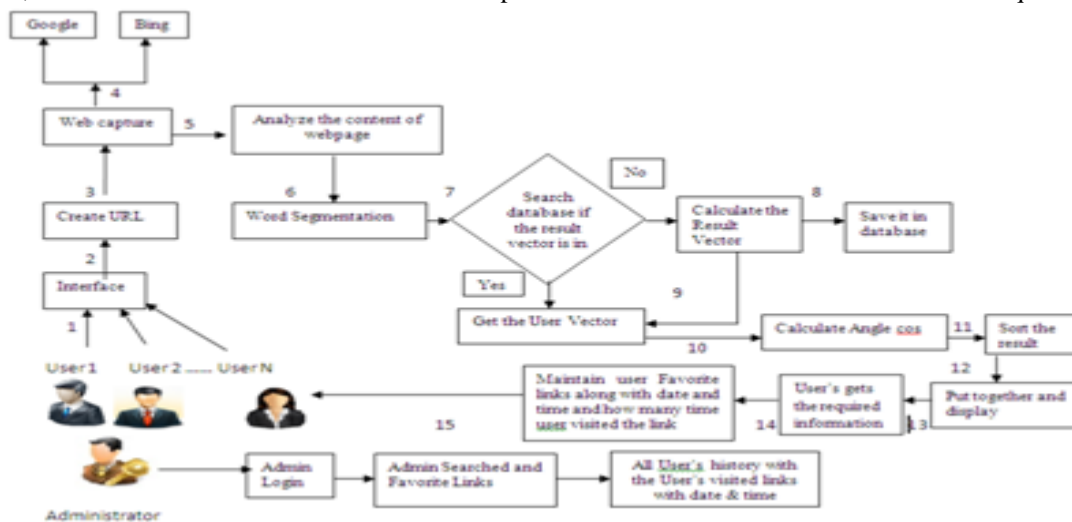

Fig. 2. Proposed System Architecture

A. DESIGN

This system consists of a JSP front with the composition of the background java program. The user interface using JSP production is used with the user interaction (Figure in step 1), the system obtains the keywords entered by the user. It then turns the query to the URL that can get results from Google, Bing (Figure in step 2). Then the page crawling module will search request processing module based on the module generated by the URL of the web pages to crawl (figure in step 4). Due to the page coming from different sources (respectively

from Google and Bing), each page is independently analyzed by engine. This is page by page analysis engine module to extract the key content, such as extracting the results of each of the page URL, title, and text descriptions. (Figure in step 6). Then word segmentation results are achieved by the page analysis module (figure in step 7). Result modeling module will use the result of English word segmentation. Result modeling module will search for the database if it contains the result of URL and its vector. If it is not contains the result, the result modeling module will calculate the result's vector (the detail will show in the model module) and put the result into the database (Figure in step 8). Otherwise, the result modeling will use the vector directly (Figure in step 9). Then the system will get the user's interest vector, this vector will use to calculate cosines of angel between result vector and user vector (Figure in step 10). The system will use these values to sort the results and feedback to users along with the date & time of the favorite link he/she searched. (Figure in step 11, 12, 13, 14, and 15). The architectural design of the personalized Meta search engine is shown in the Fig. 2.

After the sorting the result user will get the required information from the result i.e. url stored by the Search Engine such as Google and Bing . After the searching particular information which required for him/her. After some day/time he/her can access the same information by just login to the Personalized Meta Search engine the user can get the same information by clicking on it's view favorite links. And it also maintain the record of how many time user is visited that links in the Hits column.

In this paper, we put the Administrator to search required information and also maintain the record of the various user's history.

B. IMPLEMENTATION

The users will login to the Personalized MSE. It will search the information which is necessary for him/her. When the administrator will log in to the to the Personalized Meta Search Engine then it will store the information of both the users such as date of searching the information and what they have searched along with the user's favorite links of the all the user which they are visited along with number of time he/she visited a particular link in the Hits column with date and time

The Personalized Meta Search Engines don't require traversing the network, downloading web documents or building up an index. They mainly consist of member search engine selection, query forwarding, result integration and other algorithm.

The Fig.7 shows the cos angle of the result vector And display the url who's cos angle is grater on the top of the page.and display the common links just once in the Personalized Meta Search Engine.
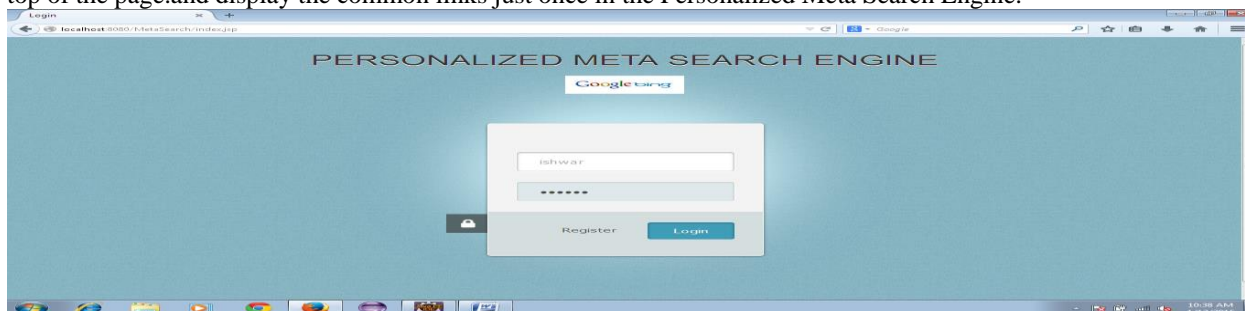

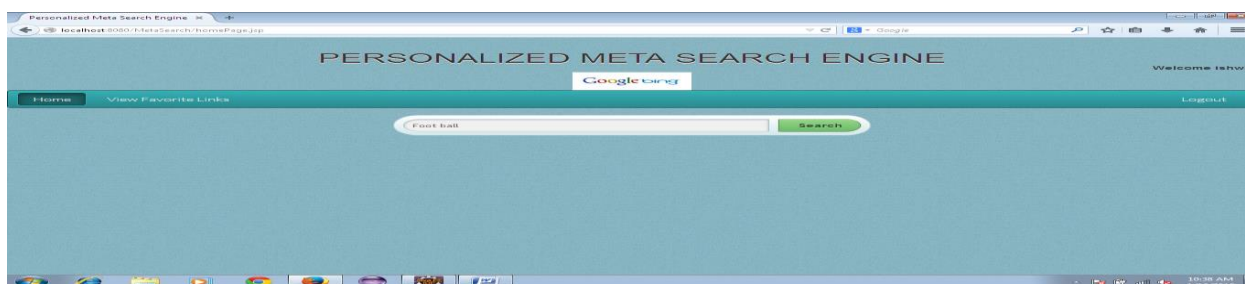Fig.3. User Login form to Personalized Meta Search Engine

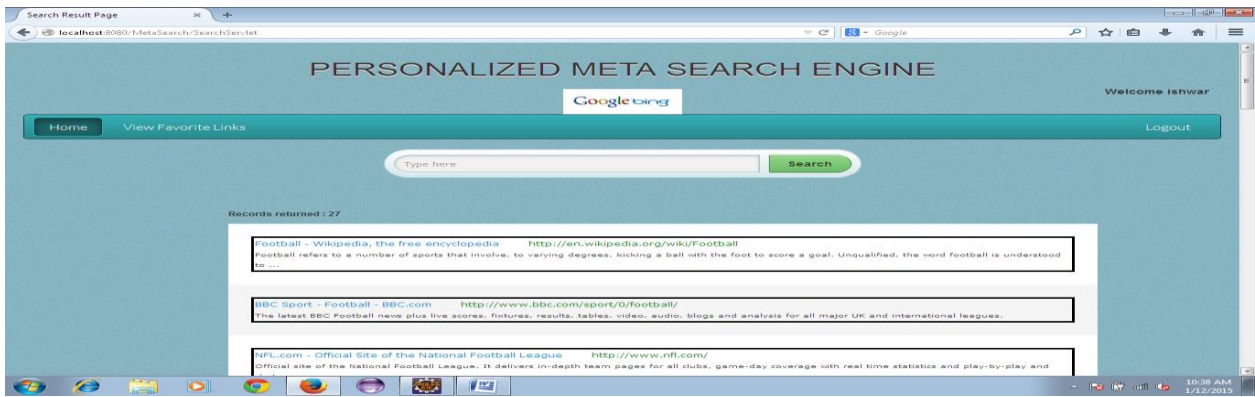Fig.4. User 1 is searching the information for Foot ball

Fig.5. Several Links for User 1 from Personalized Meta Search Engine for query Foot ball
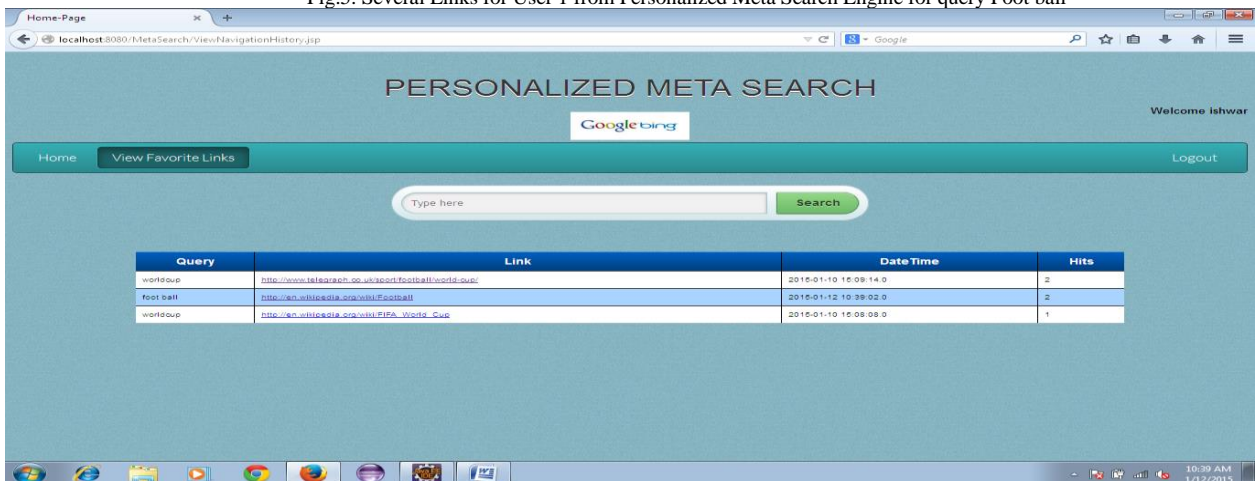

Fig.6. User 1 View Favourite links searched along with Date, Time & Hits

The demonstration of the Personalized Meta Search Engine for the user 1 is as shown in the above diagrams. Similarly the user 2 will login to the Personalized Meta Search Engine and same flow is of User 1 will be use for the user 2.Simarly for the administrator and also it's store the User History and stored how any time the user is visited that links in column of hits.
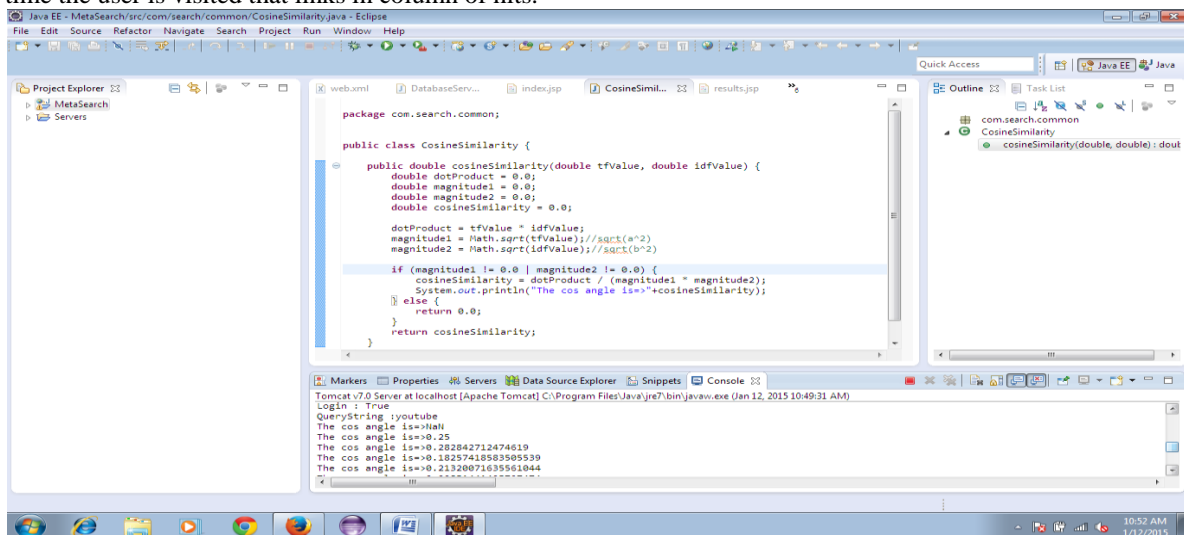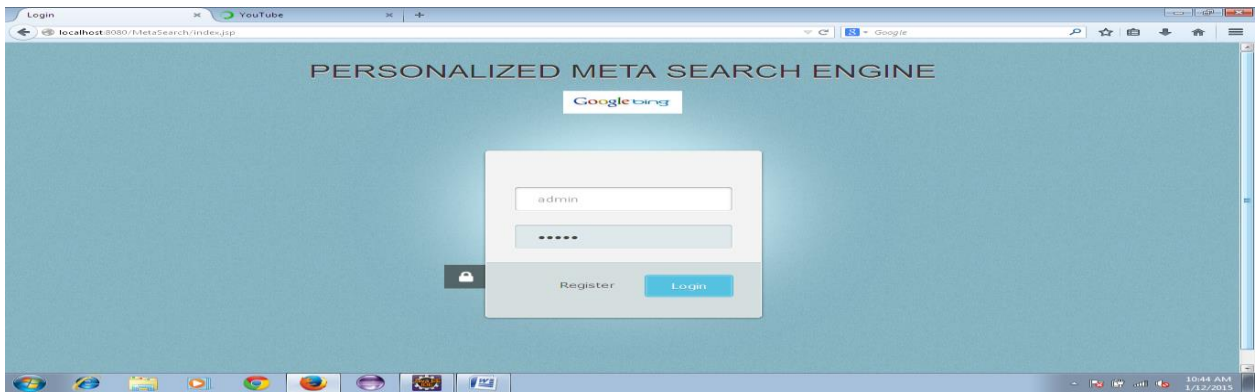

Fig.7.Diplay cos angle of the result vector

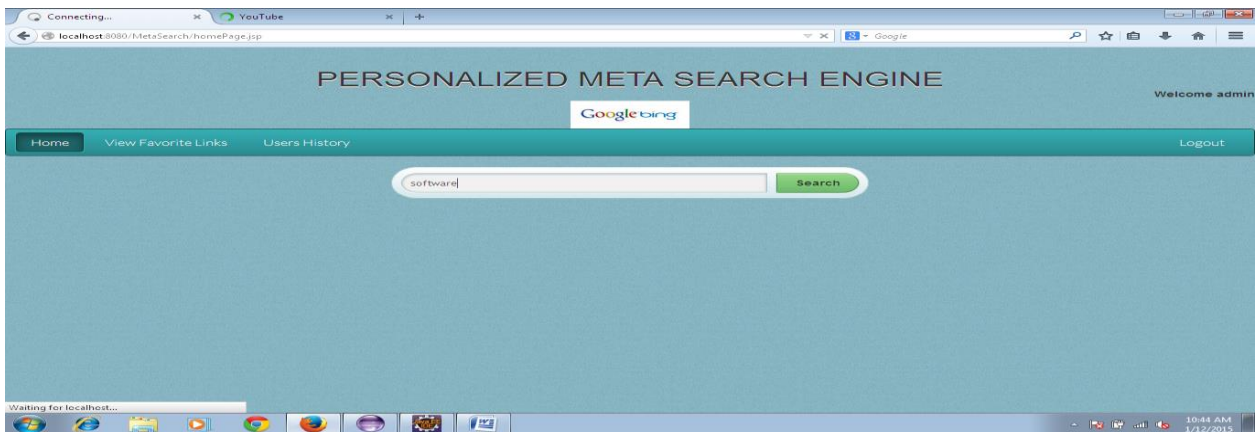Fig.8. Adminstration login to the *Personalized Meta Search Engine*
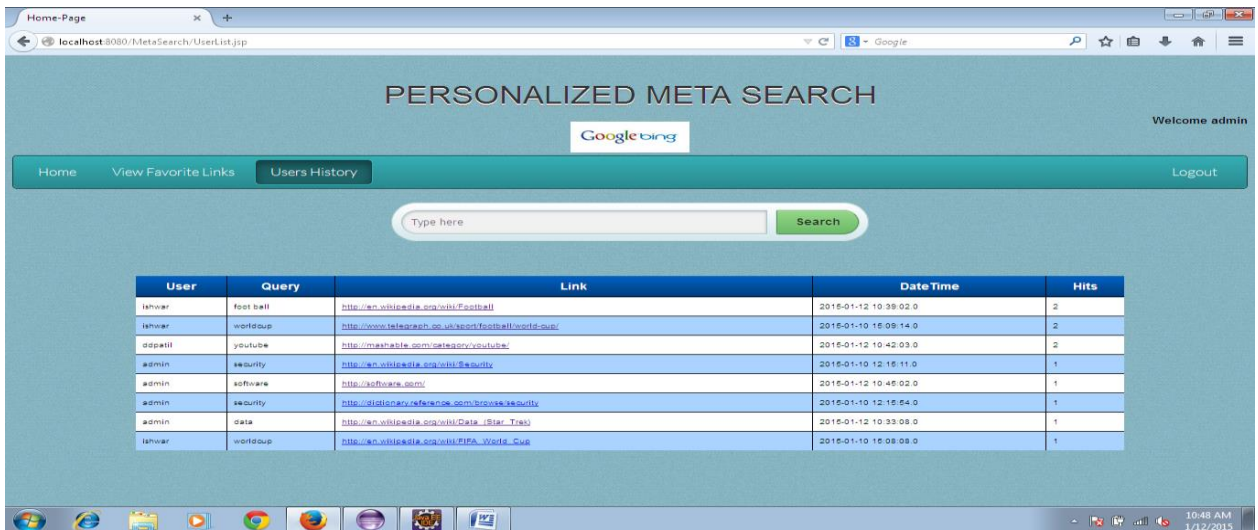

Fig.9. Admin is seraching information for software

Fig.10. User Histroy in the admin login

IV. CONCLUSIONS

The personalized search provide a common interface and conducts searches in many search engines simultaneously and return results in a uniform format. In present scenario search-engines are really useful devices to extract needed information from Internet. The personalized Meta-Search engines solve the same purpose with big span of coverage and advanced features like maintaining user's profile, filtering results etc. We Proposed MSE is based on refining the results using query expansion while next keywords are suggested by MSE itself without using any dictionary.

REFERENCES

[1]   I.N.Bharambe and R.K.Makhijani,*"Design of Search Engine using Vector Space Model for Personalized Search"*, International Journal of Computer Science and Mobile Computing(IJCSMC), ISSN 2320-088X, Vol.3,Issue. 1,January 2014,pg.118-124.

[2]   I.N.Bharambe and R.K.Makhijani,, *," Design and Implementation of Search Engine Using Vector Space Model for Personalized Search "*,International Journal of Advanced Research in Computer Science and Software Engineering, ISSN: 2277 128X, Volume 3, Issue 6, June 2013,page 1019-1023.

[3]   I.N.Bharambe and R.K.Makhijani,*"Design of Search Engine using Vector Space Model for Personalized Search"*, International Journal of Computer Science and Mobile Computing(IJCSMC), ISSN 2320-088X, Vol.2,Issue. 3,March 2013,pg.63-66.

[4]   Jiandong Cao, Yang Tang and Binbin Lou*," Personalized Meta-search Engine Design and Implementation"* Software College Northeast University (NEU) Shenyang, China, 978-1-4244-5540-9/10/$26.00 IEEE 2010.

[5]   Abawajy, J.H.; Hu, M.J., *"A new Internet meta-search engine and implementation"*, The 3rd ACS/IEEE International Conference on Computer Systems and Applications, Page(s):103, 2005

[6]   Shanmukha Rao, B.; Rao, S.V.; Sajith, G.; *"A user-profile assisted meta search engine",* TENCON 2003 Conference on Convergent Technologies for Asia-Pacific Region Volume 2, Page(s):713 - 717 , 15-17 Oct. 2003.

[7]   Spink, A.; Jansen, B.J.; Blakely, C.; Koshman, S.; *"Overlap Among Major Web Search engines"*, ITNG 2006 Third International Conference on Information Technology: New Generations, 2006. Page(s):370 – 374, 10-12 April 2006.