



**RESEARCH ARTICLE**

# Design of Search Engine using Vector Space Model for Personalized Search

**R.K.Makhijani<sup>1</sup>, I.N.Bharambe<sup>2</sup>**

<sup>1</sup>Lecturer in SSGBCOET, Bhusaval, India

<sup>2</sup>Student in SSGBCOET, Bhusaval, India

<sup>1</sup>*richa\_makhijani@yahoo.co.in*, <sup>2</sup>*ishwar.bhaarmbe@gmail.com*

---

*Abstract— In this paper, we design & implement search Engine using vector space model for personalized search is the search engine 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. 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 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.*

**Key Terms:** - User, Search Engine, Meta-search engine, Personalized, 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. 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 key-words. 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.

We need to look into the semantics of the key words. This paper suggests a new approach that is based on some algorithms which considers semantic aspects and uses them to implement a Meta-Search Engine (MSE).

## 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. The 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.

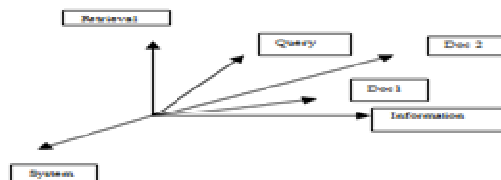
Probabilistic Latent Semantic Analysis (PLSA) gives robust results for Information Retrieval when the task is to search the most relevant documents from a given corpus, for a given query. As both of these methods depend on the Vector Space Model, the Vector Space Model is explained prior to both.

### b) Vector space model:-

A retina- 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. Fig.1 represents 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-  $t_1, t_2 \dots t_n$ ) 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-  $D_1, D_2 \dots D_r$ , where r is the total number of documents in corpora. Document Vector can be shown as following:  $D_r = \{d_{1r}, d_{2r}, d_{3r}, \dots, d_{nr}\}$



**Fig 1- Document Representation in term space**

where  $d_{ir}$  is considered to be the  $i^{th}$  component of the vector representing the  $r^{th}$  document. There are various similarity measures that are proposed and one of them, that is very frequently used, is **Cosine Similarity**.

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

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.

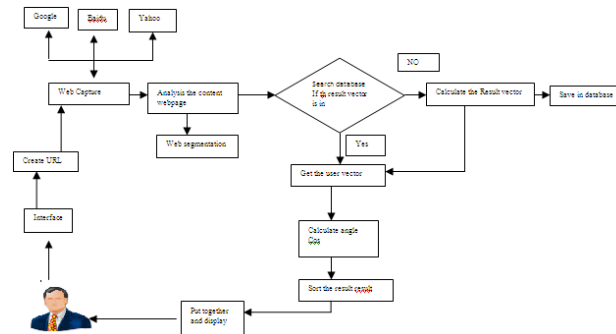
The other measures are

- 1) Inner Product =  $\sum Q_j * D_j$
- 2) Dice Coefficient =  $2 \sum Q_j * D_j / \{ \sum Q_j^2 + \sum D_j^2 \}$
- 3) Jaccard Coefficient =  $\sum Q_j * D_j / \{ \sum Q_j^2 + \sum D_j^2 - \sum Q_j * D_j \}$

Each component of document vector is always associated with some numeric-factor which is called weight of that respective term in document. This weight,  $w_i$ , can be replaced by term-count or term-frequency ( $tf_i$ ). This assignment leads to another variation of the model that is called “**Term Count Model**”.

### III. PROPOSED SYSTEM

In the proposed system, we propose a content ontology and location 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. Based on the proposed ontologies and entropies.



**Fig 2- System Architecture design**

We adopt an SVM to learn personalized ranking functions for content and location preferences. We use the personalization effectiveness to integrate the learned ranking functions into a coherent profile for personalized reranking. We implement a working prototype to validate the proposed ideas. It consists of a middleware for capturing user click throughs, performing personalization, and interfacing with commercial search engines at the backend. Empirical results show that OMF can successfully capture users' content and location preferences and utilize the preferences to produce relevant results for the users. Finally, it significantly out-performs strategies which use either content or location preference only.

The personalized Meta search engines don't require traversing the network, downloading web documents or building up an index. They are mainly consisted of member search engine selection, query forwarding, result integration and other algorithms. So, compared to robot based search engines or directory based search engines, the personalized Meta search engines have much lower technical doorsill and threshold in development and maintenance. This forces users to manually submit their queries to multiple search engines one after another until they find the information they need or give up their retrieval desire. The architectural design of the personalized Meta search engine is shown in the fig. 2.

### IV. CONCLUSION

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.

#### ACKNOWLEDGEMENT

It gives me immense pleasure in project on the topic “The Design of search Engine using vector space model for personalized search”. I acknowledge the enormous assistance and excellent co-operation extended to by my respected guide Prof.R.K.Makhijani.

I would also like to thanks the staff members for their valuable support. Lastly I would like to express my heartfelt indebtedness towards to all those who have help me directly or indirectly for the success of the project.

#### REFERENCES

- [1] Zheng Li, Yuanqiong Wang, Vincent Oria, “A New Architecture to Web Meta-Search Engine”, CIS Department, New Jersey Institute of Tech., Seventh Americas Conference on Information Systems, 2001
- [2] 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
- [3] 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
- [4] 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.
- [5] A. Gulli, A. Signorini, “Building an opensource Meta-Search Engine”, Special interest tracks and posters of the 14th international conference on World Wide Web WWW '05, ACM Press, May 2005