*Recent Developments in Electronics and Communication Systems*
*KVS Ramachandra Murthy et al. (Eds.)*


*doi:10.3233/ATDE221229*

# Query Recommendation for Search Engine Results Using Machine Learning

UMA N[a], Dr. NIRMALA[b] , PRANAV[c] , POCHIMI REDDY HARSHITHA[d] , CHUKKALURU POOJITHA[e]

[a] *Department Of Computer Science And Engineering Department, New Horizon College Of Engineering Bangalore, India*
[b] *Department Of Computer Science And Engineering Department, New Horizon College Of Engineering Bangalore, India*
[c] *Department Of Computer Science And Engineering Department, New Horizon College Of Engineering Bangalore, India*
[d] *Department Of Computer Science And Engineering Department, New Horizon College Of Engineering Bangalore, India*
[e] *Department Of Computer Science And Engineering Department, New Horizon College Of Engineering Bangalore, India*

**Abstract.** Now a days, web content over the World Wide Web is growing fast, it has become tougher to satisfy the necessities of the client's queries results. This paper provides a technique for advising a sequence of queries that are similar with the client's input search. The related searches are based on past given searches by the clients itself. This method tells about the clustering process where syntactical or semantically similar searches in groups are found. This also proposes some queries which are similar or related to the queries submitted by the client to get the required information which is relevant and efficient. This method does not only detect the similar and related searches and queries but can also rank them considering the similarity measure. And this technique is executed using real data sets from search engine query log. It gives queries in websites like Google, yahoo, Bing etc.

**Keywords.** Clustering, Query recommendation, Search engine

## 1. Introduction

Many clients find that it's tough to get the relevant information that we require as there is a drastic increase with the popularity and amount of World Wide Web, even though when a client use a popular search engine like Google and yahoo. Generally these search engines [6] permit clients to specify the queries as simple as possible like keywords, by approaching the traditional information systems. Yet, this sequence of words given does not always provide the required information, hence it is most likely important to retrieve

client's required information and to achieve client's satisfaction of given results provided by the search engine. Clustering [7] the results which are given search engine in particular portions is a way of upgrading the search engine results and gaining the resulted information we desire through the vast amount of contents from the world wide web as the user can find the desired information in desired topics of query results. Second paragraph. When the client don't use the particular keywords or any specific search words, this can cause a problem of gaining some un-necessary outcomes so the client has to be particular about the terminology and information domain. Sometimes clients have only some knowledge about the terminology they use for searching and they doesn't get required information. To get this problem over, just clustering the search engine results is not enough as the difficulty is not present in getting huge information but it is in their terminology and the words they use for searching. The past searches are generally stored in search log files, these can be used as additional requirement to help the upcoming clients.

A query recommendation system [8] is based up on the big-scale web page archive and web access logs, these evaluate three query recommendation strategies which is based upon many feature spaces like URL, Web community etc. The proposed method is mainly helpful for our search engine clients to gather their respective information or the results very easily, this method also recommends the similar queries along with the searched query when the client searches so that the client can get a proper input query with the terminology and domain knowledge where this is essential to gather all the relevant outcomes in a search engine. We are implementing an algorithm to put forward similar queries to a query given by the client. The main aim of this process of clustering is to characterize all similar queries into categories basing the information in search log file. When a client search for anything this algorithm identifies the particular group of similar searches and queries and gives a rank to them by its relevance to the client input search and at the last it recommends all past similar queries to a client. This algorithm generally follows:

- When a client gives a query as an input, this algorithm uses Linear SVC's [9] and identifies the related and similar cluster to the given search or a query.
- At the end, queries given the related cluster will be ranked upon their support and the similarity.
- As shown in the below figure, it does the clustering process and ranks them accordingly by using the clients previous queries.

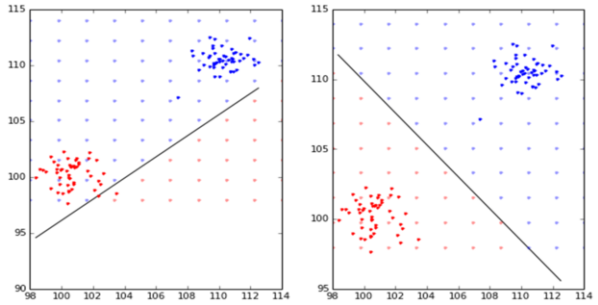


**Figure 1.** linear SVC's.

## 2. Relative Work

This article [1] proposes a clever way to deal with recover abroad nationals of a respective country by utilizing web idea mining procedures. This methodology incorporates a 3-venture process: A. key phrase piece B. Imposition of the query constraint 3. Filter of the web search. In light of the proposed approach, we foster a structure to understand a web recovery framework for looking through proficient abroad nationals. The system incorporates modules such as

i. Query Agent
ii. Web search engine
iii. Snippet Parser
iv. Page Filter
v. Metadata Generator.

Search engine optimization [2] plays a vital job in web-based advertising which can be a urgent boundary for affecting accomplishment for any business. A little retailer to enormous internet business has distinct fascination with advertising and marking through SEO and to accomplish the ideal outcomes as quickly as could really be expected. The motivation behind search engine optimization is to work on the natural positioning of a site with driving web search tools. This article presents and talks about the idea of white cap website improvement search engine optimization. This exploration chips away at the most recent calculations of web search tools particularly for Google without getting sandboxed.

In this paper [3], Optimized techniques for the webpage ranking on SEO (search engine optimization) are provided. The SEO (search engine optimization) is a technique which is used to improve the network traffic by increasing the visible locations in the results of search engines. This optimization technique can be categorized as a method to elevate the site by keeping the end goal to have a high outcome or a top rank as a main motive. Pages from these broad network databases of the search engines shows them in a sequential order to the user. Sometimes, It can happen when a URL might not totally tally with the searches done by the user, but it can provide three or four match words that may be under the necessities of the user..

In this paper [4], the Persian search engine's query has been presented to know the search behaviour of a huge users. They have described and separated the respective analyses into three classes: i) Analyses based on stats ii) Analyses based o temporal iii) Analyses which are based on topics. The overall statistics of queries posed by the clients comes under first category. Analyses of that time has the most significant role falls under second category. And, the last category shows the analyses which has done on the content queries. The first category which is analyses based stats presents that the length of average of the queries which are in Persian are bigger than the English ones. And, The queries occurring more than 7 times is lesser than the queries which occurs only once. Through the analyses it is proved that the population is proportionate to the posted queries from every province

In this paper [5], the most updated developments in traditional web search engines such as Google search engine from which Search engine optimization comes. User interaction will increase to particular website through various search engine and usability

of search engines and one of the main aspects is to improve the performance. Web page rank will be improved by various algorithms those results depends on search engine. Rank is calculated by using factor like content, the number of incoming link and the outgoing link etc. by using all of these rank of web page is displayed accordingly.

## 3. Design Implementation

When a user gives any search to our query recommending search engine, then this search will be taken and gets divides into categories accordingly using NLP and formed correctly into a sentence or takes the keywords as tokens and be divided into categories accordingly. These categories are again divided by using linear SVC'S using machine learning and produce the recommended results to the user.

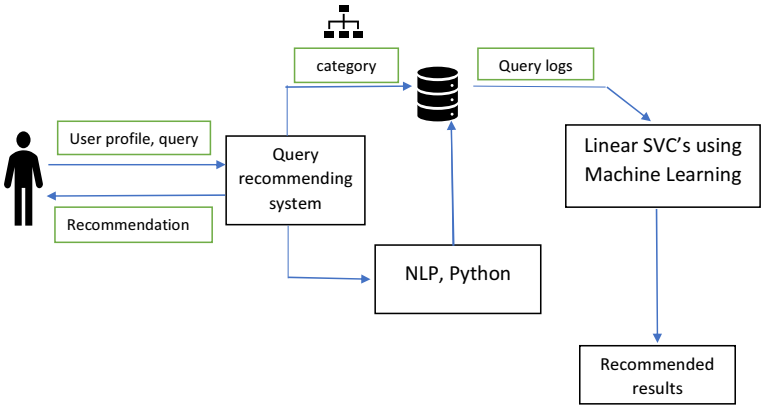


**Figure 2.** Query recommending system.

## 4. Results

When a client want to search something he needs or he require some information, he gives a search in our search engine. When the user gives a search, our search engine gives the user the information that is required. We developed a search engine in such a way that, when the user gives a search in form of keywords or any ordinary sentences, it provides a vast amount of information in many websites instead of one. Our search engine [10] gives results based on what the user requires. After the search is given, it takes user to a page where the queries are recommended in websites like Google, Yahoo, and Bing and so on. Now a days Google is the most used and efficient search engine in our day to day lives. But using our search engine helps us to gather information in websites besides Google.

## 5. Future work

There are a few various ways of improving the search engine like - Image Search Engine. The applications of this image search engine [16] make our lives essentially simpler in various business areas and in our daily routines. Its execution is when rather than text, pictures are utilized as search input, and compared or relevant images are returned as the outcome to help clients of a framework to find what they are searching for the given content. This process is based on content based image retrieval. In this process which empowers the developers and content researcher to provide situations based on keyword search in a given scenario. A comparable class in this engine allows to utilize on sites to really look at any unique or comparative photograph for various purpose. Indeed, even face-recognition in a few security frameworks utilizes similar idea to learn the identity of the individual. For exact and accurate similarity searches are done by feature extraction scaling and the enormous image-database. Image search activity ought to analyse the items in the image in highlighted vector by its pixel resolution ratio. One methodology is to compare the patches of regions between two images. In this paper Reverse Image Search by Google and large numbers of similar methods are examined and improved by executing the Deep Learning algorithm and image classifiers. Deep learning method is to remove or collect metadata from an image so that, it can be scaled, ordered and utilized in a common text query based search. Cosine similarity is utilized to calculate the angles of images in the high dimensional feature space. We can also enhance the project by implementing text to image using modified GAN's [17].

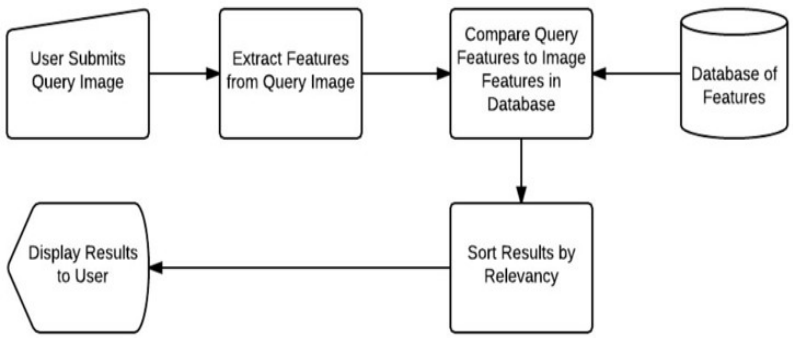


**Figure 3.** Image Search Engine.

## 6. Conclusion

Based upon the clustering process and using Machine learning techniques over the web queries that are retrieved from a search query log file, we have proposed a method for related query recommendations to input. The experiments which are done by using logs that are larger and approach to considering more searches or queries to improve the efficiency. Related keywords of cluster and expanding the queries are trying to be done.

By considering the query click answer of the documents that are similar with input query we also tried to gather information from many websites like Google, yahoo, Bing etc. In this way, we can find information in many websites rather than one. Which helps us to look up on vast amount of information where as these days we can be able to look only required amount of information in one site like Google which is the most used search engine in our daily lives. But our client can look information besides Google which makes our search engine unique comparatively with the other search engines.

## References


[1] Cheng-Yi Hsu1 , Yuh Tzong Liu2 , Jen-Shin Hong3 Department of Computer Science and Information Engineering National Chi Nan University Nantou (2018).

[2] Mayank Kumar Mittal Dept. of Information Technology Delhi Technological University, Delhi, India mkmittalofficial@gmail.com (2018).

[3] Davide Taibi1 , Giovanni Fulantelli1 , Ivana Marenzi2 , Wolfgang Nejdl2 , Richard Rogers3 , Asim Ijaz2 1 Consiglio Nazionale delle Ricerche, Istituto per le Tecnologie Didattiche, Palermo, Italy (2017)

[4] Foram S. Manek Department of Information Technology K.J.Somaiya Institute of Engineering and Information Technology, Sion. Mumbai, Maharashtra, India fmanek3@gmail.com (2017)

[5] Yingli Yao Library of Nanyang Medical College, 473061,China (2017)

[6] Kumar N, Sneha Y.S., Mungara J., Raghavendra Prasad S.G, A Survey on data mining methods for recommendation system, 2nd International conference, scopus (2018)

[7] Patil S., Jayadharmarajan A.R, Clustering With Modified mutation strategy, pertanika journal,scopus(2020)

[8] Deshmukh V.M., Shukla S., Content0-Restricted Botlzmann Machines for Diet Recommendation, Lecture notes in Networks and Systems, scopus (2021)

[9] Patil S., Anandhi R.J., Diversity Based Self-Adaptive Clusters using PSO Clustering for Crime data, International Journal of Information Technology, article in Scopus (2020)

[10] Venugopal S., Nagraj G., a proficient web recommender system using hybrid possiblistic fuzzy clustering and Bayesian Model app, International Journal of Intelligent Engineering systems, article in scopus (2018)

[11] Venugopal, S., Nagraj, G., A proficient web recommender system using hybrid possiblistic fuzzy clustering and Bayesian model approach, International Journal of Intelligent Engineering and Systems, Scopus, (2018)

[12] Geluvaraj, B., Santhosh, K., Prabhu, N., Akshay Reddy, V., Sandhya, T., Bhaskar, S.V., A Hybrid Approach for Predicting Diseases using Clustering and Classification Techniques, IEEE International Conference on Advances in Computing, Communication and Applied Informatics, Scopus (2022)

[13] Senthil Kumar, R., Arulanandham, A., Arumugam, S., Dinesh, G., Thirukkumaran, R., Subashmoorthy, R., Analysis of classification and clustering techniques for ambient AQI using machine learning algorithms, 4th International Conference on Smart Systems and Inventive Technology, Scopus (2022)

[14] Sowmya, H.K., Anandhi, R.J., Semantic Based Weighted Web Session Clustering Using Adapted K-Means and Hierarchical Agglomerative Algorithms, Journal of Web Engineering, Scopus (2021)

[15] Kumar, N., Sneha, Y.S., Mungara, J., Raghavendra Prasad, S.G., A Survey on Data Mining Methods Available for Recommendation System, 2nd International Conference on Computational Systems and Information Technology for Sustainable Solutions, Scopus (2018)

[16] Bitirim, Yiltan; Bitirim, Selin; Celik Ertugrul, Duygu; Toygar, Onsen , [IEEE 2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC) - Madrid, Spain (2020.7.13-2020.7.17)] 2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC) - An Evaluation of Reverse Image Search Performance of Google. (2020)

[17] Hanne, L.S., Kundana, R., Thirukkumaran, R., Parvatikar, Y.V., Madhura, K., Text-To-Image Synthesis Using Modified GANs, IEEE International Conference on Advances in Computing, Communication and Applied Informatics, ACCAI Scopus (2022)