Improved Searching Method Based on Genetic and Page Rank Revised Algorithms

Khandve Dhanashri,Sudrik Snehal,Wagh Sonali,Yewale Savita,Borse P.T.
Dr.D.Y.Patil School Of Engineering,Lohegaon.Pune
Computer Engineering Department

Abstract— **Now a days, To get relevant search is big challenge because of data patterns and different large terms. In this paper we studied about search engine which improve the efficiency of system. This paper introduces a search engine developed using GA, Page Rank Revised algorithm and K-NN algorithm. Search Engine is a program which find specific page which provides relevant result of user query. Many web search engines are available like Google, yahoo, Bing, ask etc. this paper presents a new rank Based system by using page Ranker Reviser algorithm with data mining techniques and also studied result of K-Nearest Neighbor and page rank revised algorithm.**

Search Terms:-Search Engine, Genetic Algorithm, K-Nearest Neighbor Algorithm.

I. INTRODUCTION

Web Search could be the Best Source of information .Web page ranking is a main part of information rescue system. search engine produce so many results in response to user query if user send s the query for particular topic then system can have thousands of result for that query. Ranking system gives more appropriate result for user query.Follwing figure 1 Shows the working of search engine[1]
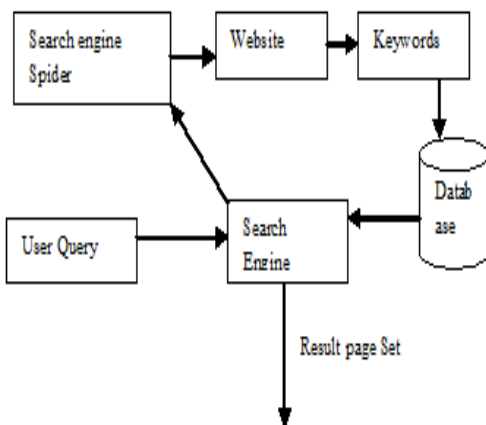

Figure1: working Of Search Engine.[4]

Page ranking system ranks` the pages based on weight of the query. There are following figure2 describe working of ranking system.
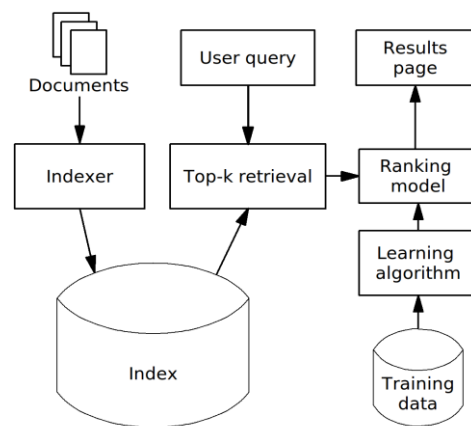

Figure2:Working Of Ranking System.

This ranking depends upon different factors like number of incidence of keyword.

II. Literature survey

[1] Michael E. Houle and Michael Nett, "Rank-Based Similarity Search: Reducing the Dimensional Dependence" , IEEE Transactions On Pattern Analysis And Machine Intelligence, Vol. 37, No. 1, January 2015.
This paper includes searching technique which makes user search data quite efficient and a structure for parallel search, the Rank Cover Tree, whose ordinal pruning policy make sure only of direct comparisons between distance values. This technique gives a relationship or similarity between search document and user query.

[2] Yuefeng Li, Abdulmohsen Algarni, Mubarak Albathan, Yan Shen, and Moch Arif Bijaksana, "Relevance Feature Discovery For Text Mining",

IEEE Transactions On Knowledge And Data Engineering,
Vol. 27, No. 6, June 2015.
Authors of this paper focus on relevant feature selection in text documents. It introduces a method to select unrelated documents for premium features.
It provides a capable methodology for developing effective text mining models for relevance feature detection.

[3] Fabrizio Lamberti, Member, IEEE, Andrea Sanna, and Claudio Demartini, Member, IEEE , "A Relation-Based Page Rank Algorithm for Semantic Web Search Engines" IEEE Transactions On Knowledge And Data Engineering, Vol. 21, No. 1, January 2009

This paper mainly use page relevance principles based on information that has to be derived

from the whole knowledge base,making their application often unachievable in huge semantic environments .authors of this paper represents a new ranking strategy, that is capable of providing a weight score for a Web page into an annotated result set by simply considering the user input query, the page explanation, and the primary ontology.

[4] Nisha Bansal And Dr. Paramjeet Singh," Improved Web Page Ranking Algorithm Using Semantic Similarity And HITS Algorithm", *International Journal Of Emerging Trends & Technology In Computer Science (IJETTCS)*,Volume 3, Issue 4 July-August 2014
This paper contains a technique which is based on ranking system with improved HITS and Semantic Similarity techniques. This technique is used to rates the web pages. This technique is used to rank a web page from a set of given web pages.

[5] *Seema Rani***1***, Upasana Garg,"*A Ranking Of Web Documents Using Semantic Similarity And Artificial Intelligence Based Search Engine*"*
This paper includes the system which rank the web page. This system is based on three techniques which are Semantic Similarity approach, HITS and on the basis of AI technique ,AI technique is used to access the user history to rank the webpage according to the user query.

III.RELATED WORK

In this paper ranked the user query according to three techniques, which are genetic algorithm, page rank revised as well as k-nearest neighbor method which access the each recording data set based on a combination of the classes of k records most similar to it in historical data set and application is developed using java is use as an programming language tool and SQL as an query processing of an data set. The result of the system shown with the help of graphic user interface[1][2].

K-NN algorithm:

The K-nearest-neighbor algorithm statement the distance between a query circumstances and a set of

scenarios in the data set[1].K nearest neighbors is a simple algorithm that stores all accessible cases and classifies new cases based on a correspondence quantify[1].

K-NN algorithm steps:

Step 1:- Store the output values M of the nearest neighbors to query q scenario in vector r=r(r1,…,rm) by repeating the following loop M times:
a. Go to the next state in the data set, where is the present iteration within the field {1,…..,p}.
b. Loop until we reach the end of the data set (i.e.)
c. Store into vector and into vector

Step2: Calculate the arithmetic mean output across

follows:

$$\bar{r} = \frac{1}{M} \sum_{i=1}^{M} r_i$$

Step3:Return as the output value for the query Scenario[7].

Working of Genetic Algorithm

A genetic algorithm is a search method used in computing to find true solutions to optimization and search problem. Genetic algorithm are categorized as global search heuristics[6].
Steps for Genetic Algorithm in Search engine:
Step 1:-Selection.
Step 2:-Crossover.
Step 3:-Mutation.

Working of page Rank Reviser Algorithm

Page Rank is an algorithm used by search techniques to rank pages in their search engine results .We explain the ranking method using example [4][5].
Example:
If we search "Tiger" on internet then we find the result with different images of tiger with another image in without ranking process at show in figure3.
Using ranking process server shows the result as show in figure4.

Figure3: Before re-ranking

Repository. [Online]. Available: http://www.ics.uci.edu/ Mlearn/MLRepository.html

Figure4: After re-ranking

CONCLUSION AND FUTURE SCOPE

In this paper, Improve page ranking technique to rank a page from a set of given pages. System is using genetic algorithm with data mining technique to rank the data. System can be further improved by implementing on the servers and by using multithreading techniques to improve the time efficiency and performance of the system.

REFERENCES:-

[1] Michael E. Houle and Michael Nett, "Rank-Based Similarity Search: Reducing the Dimensional Dependence" , IEEE Transactions On Pattern Analysis And Machine Intelligence, Vol. 37, No. 1, January 2015.

[2] Yuefeng Li, Abdulmohsen Algarni, Mubarak Albathan, Yan Shen, and Moch Arif Bijaksana,"Relevance Feature Discovery For Text Mining", IEEE Transactions On Knowledge And Data Engineering, Vol. 27, No. 6, June 2015.

[3] Fabrizio Lamberti, Member, IEEE, Andrea Sanna, and Claudio Demartini, Member, IEEE, "A Relation-Based Page Rank Algorithm for Semantic Web Search Engines" IEEE Transactions On Knowledge And Data Engineering, Vol. 21, No. 1, January 2009

[4] Narayan L. Bhamidipati and Sankar K. Pal, Fellow, IEEE, "Comparing Scores Intended for Ranking" IEEE Transactions On Knowledge And Data Engineering, Vol. 21, No. 1, January 2009

[5] Nisha Bansal And Dr. Paramjeet Singh," Improved Web Page Ranking Algorithm Using Semantic Similarity And HITS Algorithm", *International Journal Of Emerging Trends & Technology In Computer Science (IJETTCS)*,Volume 3, Issue 4 July-August 2014

[6] *Seema Rani***1***, Upasana Garg,"*A Ranking Of Web Documents Using Semantic Similarity And Artificial Intelligence Based Search Engine*",* International Journal Of Science, Engineering And Technology Research (Ijsetr), Volume 3, Issue 12, December 2014

[7] M. Ankerst, M. M. Breunig, H.-P. Kriegel, and J. Sander, "OPTICS:Ordering points to identify the clustering structure," in Proc. ACMSIGMOD Int. Conf. Manag. Data, 1999, pp. 49–60.

[8] A. Asuncion and D. J. Newman. (2007). UCI machine learning