IJASEM INTERNATIONAL JOURNAL OF APPLIED
SCIENCE ENGINEERING AND MANAGEMENT

Building Search Engine Using Machine Learning Technique

¹ K VENKATESH, ² N.DEVI SAI SESHAKSHITHA

(Assistant Professor), MSC, **DANTULURI NARAYANA RAJU COLLEGE(A) PG COURSES, BHIMAVARAM ANDHRA PRADESH**

²MSC, scholar, **DANTULURI NARAYANA RAJU COLLEGE(A) PG COURSES, BHIMAVARAM ANDHRA PRADESH**

ABSTRACT

The web is the huge and most extravagant wellspring of data. To recover the information from the World Wide Web, Search Engines are commonly utilized. Search engines provide a simple interface for searching for user query and displaying results in the form of the web address of the relevant web page, but using traditional search engines has become very challenging to obtain suitable information. This paper proposed a search engine using Machine Learning technique that will give more relevant web pages at top for user queries.

1.INTRODUCTION

World Wide Web is actually a web of individual systems and servers which are connected with different technology and methods. Every site comprises the heaps of site pages that are being made and sent on the server. So if a user needs something, then he or she needs to type a keyword. Keyword is a set of words extracted from user search input. Search input given by a user may be syntactically incorrect. Here comes the actual need for search engines. Search engines provide you a simple interface to search user queries and display the results.

1) Web crawler Web crawlers help in collecting data about a website and the links related to them. We are only using web crawlers for collecting data and information from WWW and storing it in our database.

2) Indexer Indexer which arranges each term on each web page and stores the subsequent list of terms in a tremendous repository.

3) Query Engine It is mainly used to reply to the user's keyword and show the effective outcome for their keyword. In the query

engine, the Page ranking algorithm ranks the URL by using different algorithms in the query engine.

4)This paper utilizes Machine Learning Techniques to discover the utmost suitable web address for the given keyword. The output of the PageRank algorithm is given as input to the machine learning algorithm.

2. EXISTING SYSTEM

• Information retrieval is to retrieve the information resources that we are interested in or extract whatever information we need.

• Information Retrieval (IR) may deal with the organization, storage, retrieval and evaluation of information from documents, particularly textual information.

• But we cannot give the ranks to those documents

DISADVANTAGES OF EXISTING SYSTEM:

• Information retrieval will be very difficult in large numbers of texts in a document.

• Difficult to identify the important concepts or topic in a collection of documents.

• The explicit rankings are always difficult to obtain or even not available in many documents.

3. PROPOSED SYSTEM

• The proposed search engine is very useful for finding out more relevant URLs for given keywords.

• Anyone can easily identify the important documents in a collection of documents and retrieve the related data.

• It proposes a novel model, named LDA (Linear Discriminant Analysis), easy for clustering the related documents based on that ranking

ADVANTAGES OF PROPOSED SYSTEM:

• We will build a search engine which gives the web address of the most relevant web page at the top of the search result, according to user queries.

• The main focus of our system is to build a search engine to discover the utmost suitable web address for the given keyword by using machine learning techniques for increasing accuracy compared to available search engines.

4. OUTPUTSCREENS
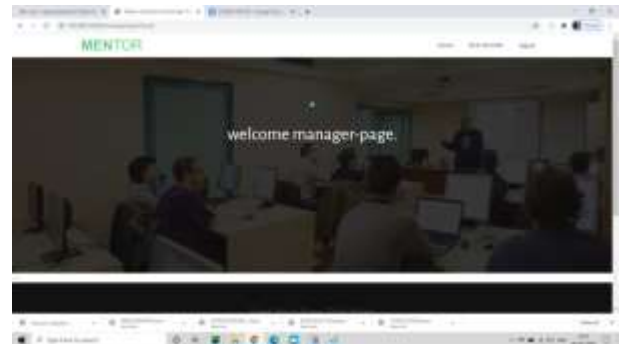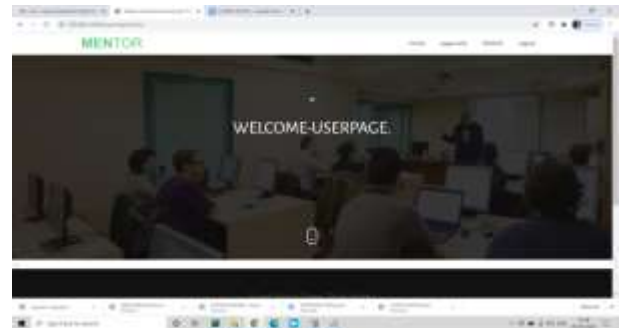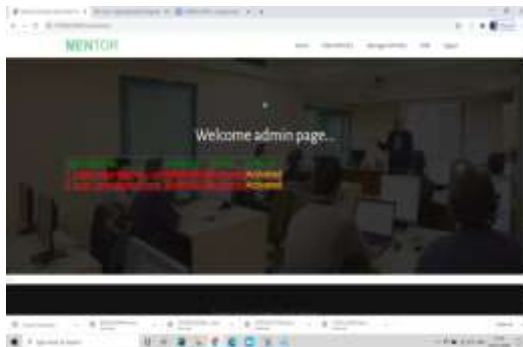
5. CONCLUSION

Search engines are very useful for finding out more relevant URLs for given keywords. Due to this, user time is reduced for searching the relevant web page. For this, Accuracy is a very important factor. From the above observation, it can be concluded

that XGBoost is better in terms of accuracy than SVM and ANN. Thus, Search engines built using XGBoost and PageRank algorithms will give better accuracy.

6. REFERENCE

[1] Manika Dutta, K. L. Bansal, "A Review Paper on Various Search Engines (Google, Yahoo, Altavista, Ask and Bing)", International Journal on Recent and Innovation Trends in Computing and Communication, 2016.

[2] Gunjan H. Agre, Nikita V.Mahajan, "Keyword Focused Web Crawler", International Conference on Electronic and Communication Systems, IEEE, 2015.

[3] Tuhena Sen, Dev Kumar Chaudhary, "Contrastive Study of Simple PageRank, HITS and Weighted PageRank Algorithms: Review", International Conference on Cloud Computing, Data Science & Engineering, IEEE, 2017.

[4] Michael Chau, Hsinchun Chen, "A machine learning approach to web page filtering using content and structure analysis", Decision Support Systems 44 (2008) 482–494,scienceDirect,2008.

[5] Taruna Kumari, Ashlesha Gupta, Ashutosh Dixit, "Comparative Study of Page Rank and Weighted Page Rank Algorithm", International Journal of Innovative Research in Computer and Communication Engineering, February 2014.

[6] K. R. Srinath, "Page Ranking Algorithms – A Comparison", International Research Journal of Engineering and Technology (IRJET), Dec2017.

[7] S. Prabha, K. Duraiswamy, J. Indhumathi, "Comparative Analysis of Different Page Ranking Algorithms", International Journal of Computer and Information Engineering, 2014.

[8] Dilip Kumar Sharma, A. K. Sharma, "A Comparative Analysis of Web Page Ranking Algorithms", International Journal on Computer Science and Engineering, 2010.

[9] Vijay Chauhan, Arunima Jaiswal, Junaid Khalid Khan, "Web Page Ranking Using Machine Learning Approach", International Conference on Advanced Computing Communication Technologies, 2015.

[10] Amanjot Kaur Sandhu, Tiewei s. Liu., "Wikipedia Search Engine: Interactive Information Retrieval Interface Design", International Conference on Industrial and Information Systems, 2014.