A Review of Search Engine Techniques Based On Machine Learning

Anuradha T¹, Tayyaba Nousheen Javed²

Professor¹, PG Student²

Department of CSE

Poojya Doddappa Appa College of Engineering, Kalaburagi, Kamataka, India

Abstract:

The Search engine plays a significance role in the success of web. Search engine helps the user to find the relevant information. To recuperate the information from WWW, Search Engines are commonly utilized. Search engines provide an interface for searching for user query and displaying results in the form of the web address of the relevant web page, this paper proposes a review of various search engine technique using Machine Learning.

Keywords: World Wide Web (WWW), Search Engine, Page Rank, Machine Learning.

I. INTRODUCTION

World Wide Web is actually a web of individual systems and servers which are connected with different technology and methods. Every site comprises of 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. Search input given by user may be syntactically incorrect. Here comes the actual need for search engines. Search engines provide you a simple interface to search user query and display the results in the form of the web address of the relevant web page. Figure 1 focuses on three main components of search engine.

1) Web crawler

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

2) 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 the user’s keyword and show the effective outcome for their keyword. In query engine, Page ranking algorithm ranks the URL by using different algorithms in the query engine.

Fig1: system diagram

II. Working of search engine

Search engines works by comparing hundreds of Pages using their own crawlers. These web crawler are commonly referred bots or spiders. Search engine allow user’s to search the internet for content using keywords. Currently the market is dominated by some search engines one such is Google, there are many search engines that can used. Whenever user writes a query in a search engine, the search engine result page (SERP) is returned, ranking the found pages in order of their relevance. The search engine is classified to record the information and algorithms that compute the result to reappearance and rank for a given query. In the case of web search engine like Google and Bing, the database consists of Billions of Web Pages and the algorithm look at thousands of factors to deliver the most related results. Search engine is interlinked with several mechanisms that work together to identify fragments of web content like images, videos, website pages, etc. Based on the keywords you type into a search bar. Site admins use Search Engine Optimization to enhance the content on their websites to show up relevant data in search results.

Fig2: working steps of search engine. Source: click here

Search engines use three basic mechanisms:

- Web crawlers: Bots that frequently browse the web for new web pages. Crawler collects the information that is needed to index page correctly and use hyperlinks to hop to other pages and index them too.
- Search index: A record of all web pages are organized in a way that allows association between keyword terms and page content. Search engines have the ways to categorizing the quality of content in their indexes.
- Search algorithms: They Calculate the quality of web pages, figure out how relevant that page is to a search term, and determine how the results are ranked based on quality and popularity. The Search engines try to deliver the furthermore useful results for the user to keep large numbers of users coming back time and again. This makes corporate sense, as most search engines make money by advertising. Google made an impressive $116B in 2018.

Research Article                                                                                                                           Volume 11 Issue No. 03
IJE SC, March 2021                                                                        27784                           Volume 11 Issue No. 03
III. LITERATURE REVIEW

Numerous endeavors have been made by data experts and researchers in the field of search engine. Dutta and Bansal [1] discuss various type of search engine and they conclude the crawler based search engine is best among them and also Google uses it. It gives a user more relevant web address for user query. A Web crawler is a program that navigates the web by following the regularly changing, thick and circulated hyperlinked structure and from there on putting away downloaded pages in a vast database which is after indexed for productive execution of user queries. In [2], author discuss that major benefit of using keyword focused web crawler over traditional web crawler is that it works intelligently, efficiently. The search engine uses a page ranking algorithm to give more relevant web page at the top of result, according to user need. It eases the searching method and user gets required information very easily. Initially just an idea has been developed as user were facing problem in searching data so simple algorithm introduced which works on link structure, then further modification came as the web is also expanding so weighted Page Rank and HITS came into the scenario. In [3], author compares various Page Rank algorithms and among all, Weighted Page Rank algorithm is best suited for our system. Michael Chau and Hsinchun Chen [4] discuss a system which is based on a machine learning approach for web page filtering. The machine learning result is compared with traditional algorithm and found that machine learning result are more useful. The proposed approach is also effective for building a search engine. In [5] this the author discuss the Comparative Study of the page ranking algorithms and provide the accurate result of the user search keywords. In [6], author discuss a system of creating a domain specific search engine are growing in popularity because they offer increased accuracy and extra functionality not possible with the general Web-wide search engines, that allows multifaceted queries by age-group, size, location and cost. The usage of machine learning techniques to automate the formation and maintenance Domain Specific Search Engines. In [7], author research discuss the development of a Meta search engine, called Serelec that provide an interface for filtering and classifying the search engines results as to narrow down the search results in the form of sequentially linked manner resulting in extreme reduction of number of pages. In [8], author discuss a system is to provide personalized results to the users based on their individual interests. It re-ranks the results for a given query obtained from existing search engine based on the user need. In [9] author discusses a system which compared the information retrieval performances of some popular search engines (namely, Google, Yahoo, AlltheWeb, Gigablast, Zworks and AltaVista and Bing/MSN) in response to a list of queries, varying in complexity. The queries are run on each search engine and the precision and response time to retrieve the results were recorded. In [10] this author discusses the surveys about how the search engines reduce the number of unwanted search results in the searching process. A search engine is a tool that allows a user to type the keywords and retrieve information in the form of websites that are contained in the database.

<table>
<thead>
<tr>
<th>SL NO.</th>
<th>Author and title name</th>
<th>Journal name</th>
<th>Year</th>
<th>Algorithm used</th>
<th>Remark</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Manika Dutta and K.L. Bansal “A Review Paper on Various Search Engines (Google, Yahoo, Altavista, Ask and Bing)”</td>
<td>BRITCC</td>
<td>2016</td>
<td>Comparison algorithm</td>
<td>In this paper, the authors have discussed to compare the various search engine based on the accuracy, speed, search video, images, doc etc. And provided the data in the form of comparison of various search engines.</td>
</tr>
<tr>
<td>2</td>
<td>Gunjan H. Agre, Nikita V Mahajan, “Keyword Focused Web Crawler”</td>
<td>IEEE</td>
<td>2015</td>
<td>Naive bays algorithm</td>
<td>In this author have discussed the system which will extract the selected WebPages from the user searched keywords. The user keywords are automatically corrected by the search engine and then they provide the accurate results.</td>
</tr>
<tr>
<td>3</td>
<td>Tuhena Sen, Dev Kumar Chaudhary, “Contrastive Study of Simple Page Rank, HITS and Weighted Page Rank Algorithms: Review”</td>
<td>IEEE</td>
<td>2017</td>
<td>Hyperlink induced topic search algorithm, Page ranking algorithm, weighted page ranking algorithm</td>
<td>In this paper, the authors have discussed to build a search engine which is useful for finding out more relevant URL for given keyword. Based on the user queries the search engine will generate the urls for the related WebPages.</td>
</tr>
<tr>
<td>No.</td>
<td>Authors</td>
<td>Conference</td>
<td>Year</td>
<td>Title</td>
<td>Journal</td>
</tr>
<tr>
<td>-----</td>
<td>---------</td>
<td>------------</td>
<td>------</td>
<td>-------</td>
<td>---------</td>
</tr>
<tr>
<td>4</td>
<td>Michael Chau, Hsinchun Chen</td>
<td>IJEDR</td>
<td>2008</td>
<td>“A machine learning approach to web page filtering using content and structure analysis”</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Taruna Kumari, Ashlesha Gupta, Ashutosh Dixit</td>
<td>IJRITCC</td>
<td>2017</td>
<td>“Comparative Study of Page Rank and Weighted Page Rank Algorithm”</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>K R Srinath</td>
<td>IRJET</td>
<td>2017</td>
<td>“Page ranking algorithm”</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>S. Prabha, K. Duraiswamy, J. Indhumathi</td>
<td>IJCIE</td>
<td>2014</td>
<td>“Comparative Analysis of Different Page Ranking Algorithms”</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Neha Sharma, Rashi Agarwal, Narendra Kohli</td>
<td>International Conference on Advanced Computing Communication Technologies</td>
<td>2016</td>
<td>“Review of features and machine learning techniques for web searching”</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Vijay Chauhan, Arunima Jaiswal, Junaid Khalid Khan</td>
<td>International Conference on Advanced Computing Communication Technologies</td>
<td>2015</td>
<td>“Web Page Ranking Using Machine Learning Approach”</td>
<td></td>
</tr>
</tbody>
</table>
IV. Importance of Search Engine
Search engines essentially act as riddles for the wealth of information available on the internet. They allow users to quickly and easily access the information that is of genuine interest and value, without the need to splash through numerous extraneous web pages. Search engines provide users with search results that lead to appropriate information on first-rate websites. Search engine matter because they increasingly determine the information about the brands, products and the services that clients access online. Being informal to find on Google, Yahoo and MSN is as much of the marketing necessity to have a strong presence in print and broadcast media, and an effective traditional direct marketing program. To achieve and hold market share in online searches, search engines need to make a sure appearance as they deliver results that are relevant to their user search keywords. They do this by preserving records of web pages, which they develop by using computerized programs known as "spiders" or "robots" to collect information.

V. Conclusion
The volume of information available on the Internet continues to grow exponentially. As this trend continues, we say that, not only the public need powerful tools to help them to sort the information, but the creators of these tools will need intelligent techniques to help them build and continue these tools. Search engine is very useful for finding out most relevant web pages for the given keyword. Due to this, user time is reduced for searching the relevant web page.

References


