1st International Conference on Applied Soft Computing Techniques
22 & 23.04.2017
In association with
International Journal of Scientific Research in Science and Technology

A Smart Searching Technique for Optimizing Relevant Web portal Discovery

S. Dhanasekaran[1], Vamshikrishna Bandari[2], Ravi Teja[3], Vishnu Gupta[4]

*[1] Department of CSE, Kalasalingam University, Srivilliputtur, Tamilnadu, India
srividhans@gmail.com[1]

*[2] Department of CSE, Kalasalingam University, Srivilliputtur, Tamilnadu, India
krishnavamshi.12321@gmail.com[2]

*[3]Department of CSE, Kalasalingam University, Srivilliputtur, Tamilnadu, India
teja.14raviteja@gmail.com[3]

*[4] Department of CSE, Kalasalingam University, Srivilliputtur, Tamilnadu, India
thotavishnugupta@gmail.com[4]

ABSTRACT

This Research work is mainly deals about the Minimization of search options in a search engine. In general the keyword searched in any search engine gets some millions of results in microseconds. The output is obtained by analysing and processing a bulk data, thereby obtaining all relative or most searched sites and web pages. But here we put forth an idea for getting the fixed and efficient result in a short duration of time. The keyword searched in the search box gets the top 5 related sites per page through which the user can obtain the exact and efficient output. By this method of searching the search time for user and searching load for system and the server, both gets reduced which results in the effective usage of search engine.

Keywords: Search engine, crawling, spiders, building an index, API key, ranking factors.

I. INTRODUCTION

The frequently used search engine "Google" returns a bulk data even for a small search keyword typed. The returned data may or may not be required by the users totally. It causes great dissatisfaction if not required. So, this work introduced a new search engine "**Searchin** which is a modified version of Google. This new search engine gets the top most 5 results which are very much related and efficient to the users. There by reducing the search time. It is the modification of search engine API key.

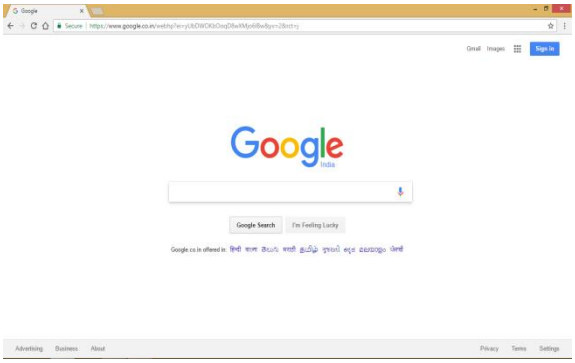
II. METHODOLOGY

For optimizing or minimizing of search options we are going to use the search engine API key in our Research work. As we know that the best and fastest search engine in world is "GOOGLE", we need to modify the source code for optimization in a way that the top 5 results are displayed. For modification firstly we need to understand the search engine, that how it works, ranking factors.

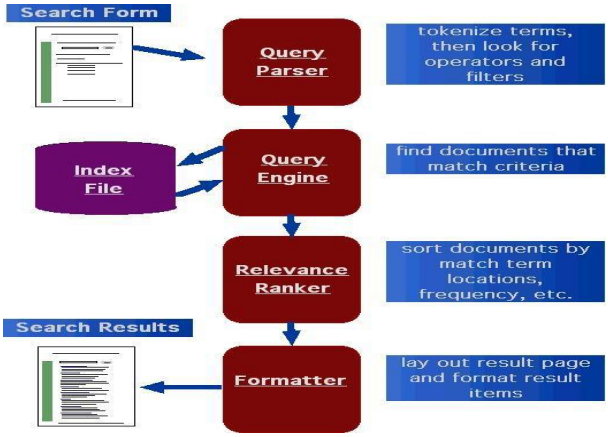
Working of search in engine:

Search engines mainly perform two functions namely crawling and building an index. The term crawling is used because the search engine while

searching crawls each and every part of the web server and finds many billions of interconnected documents on the web. The interconnected documents are analyzed and best results are found. To complete these tasks search engines have constructed database centers all over the world. These database centers hold thousands of machines processing large quantity of data very quickly. Whenever we search anything in the search engine, we expect it to get the results as fast as possible, even a small delay causes great dissatisfaction. So, these engines work hard to obtain result as fast as possible. Whenever a keyword is searched the search engine searches in two major ways and obtain the result i.e.., the most visited and the most relevant, based on their previous searches and displays all in an order depending on their priority from first to last.



When a person performs an online search, the search engine find its billions of documents and does two things: first, it returns only those results that are relevant or useful to the searcher's query; second, it ranks those results according to the popularity of the websites serving the information. It is both relevance and popularity that the process of **SEO** is meant to influence. How do search engines determine relevance and popularity? To a search engine, relevance means more than finding a page with the right words. In the early days of the web, search engines didn't go much further than this simplistic step, and search results were of limited value. Over the years, smart engineers have devised better ways to match results to searchers' queries. Today, hundreds of factors influence relevance, and we'll discuss the most important of these in this guide. search engines typically assume that the more popular a site, page, or document, the more valuable the information it contains must be. This assumption has proven fairly successful in terms of user satisfaction with search results. Popularity and relevance aren't determined manually. Instead, the engines employ mathematical algorithms to sort the relevance, and then to rank in order of quality (popularity). These algorithms often comprise hundreds of variables. In the search marketing field, we refer to them as "ranking factors."



Ranking factors:

Ranking algorithm varied your search option in billion and trillions of pages in cloud to identify how similar to each one .Revealing the Companies own ranking algorithms is kept secret.

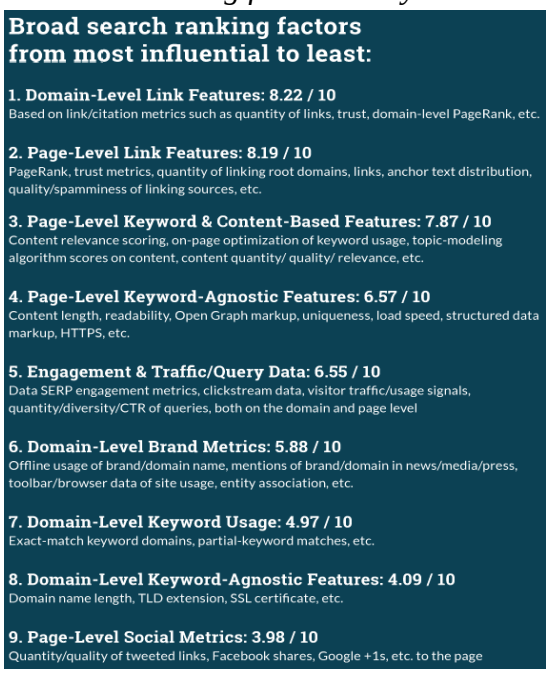
In past 3 years Google made a difficult to achieve it .earlier ,sites or web protocols are ranked on how much times a specific keyword /option was searched. These leads to "keyword stuffing ".where these pages or sites are most consists of meaningless matter as long as it provide the keyword or option in them.

Then the concept of importance based on linking was introduced more popular sites would be more linked to, obviously – but this led to a proliferation of spammed links all over the web. Now each link is determined to have a different value, depending

on the "authority" of the site in question. If a high-level government agency links to you, it's worth far more than a link found in a free-for-all "link directory".

III. EXPERIMENTAL RESULT

Ranking factor survey



Ranking factor Analysis

IV. CONCLUSION AND FUTURE SCOPE

On analyzing the working and ranking results of search engine we create a new search engine named as **"Searchin"** which function as Google because we used Google API key with slight modification as we want.

Smart searchin engine

The results come out by searching a keyword is based on ranking factors; cache stored in system and recently viewed pages. By this Searchin we can access data as fast as (exactly 10 times faster than **Google**). Based on the observation the following benefits has been achieved by our smart searching techniques Faster than Google in some aspect, Google filters are also accepted. Time consumption on searching is optimized compared to any other search engines, Results per page is five which allow the user to find best one. Data consumption is less, Algorithm is easy to design.

However this technique is going to find most relevant result with minimum time and data consumption. In future this searching techniques will be implemented in regional languages.

V. References

[1] Feng Zhao, Jingyu Zhou, Chang Nie, Heqing Huang, Hai Jin, "SmartCrawler: A Two- Stage Crawler forEfficiently Harvesting Deep-Web Interfaces", IEEE Transaction on services computing, Vol-99, 2015.

[2] M. Burner, "Crawling towards Eternity: Building an Archive of the World Wide Web," Web Techniques Magazine, vol. 2, pp. 37-40, 1997.

[3] Allan Heydon and Marc Najork, Mercator: A scalable,extensible webcrawler. World Wide Web Conference,2(4):219–229, April 1999.

[4] Jenny Edwards, Kevin S. McCurley, and John A. Tomlin. An adaptive model for optimizing performance of an incremental web crawler. In Proceedings of the Tenth Conference on World Wide Web, pages 106–113, Hong Kong, May 2001. Elsevier Science.

[5] Luciano Barbosa and Juliana Freire. Searching for hidden-web databases. In WebDB, pages 1– 6, 2005.

[6] Soumen Chakrabarti, Martin Van den Berg, and Byron Dom. Focused crawling: a new approach to topic-specific web re-source discovery. Computer Networks, 31(11):1623–1640, 1999.

[7] Jayant Madhavan, David Ko, Łucja Kot, Vignesh Ganapathy, Alex Rasmussen, and Alon Halevy. Google's deep web crawl. Proceedings of the VLDB Endowment, 1(2):1241–1252, 2008.

[8] Olston Christopher and Najork Marc. Web crawling. Foundations and Trends in Information Retrieval, 4(3):175–246, 2010.

[9] Balakrishnan Raju and Kambhampati Subbarao. Sourcerank:Relevance and trust assessment for deep web sources based on inter-source agreement. In Proceedings of the 20th international conference on World Wide Web, pages 227–236, 2011.

[10] Denis Shestakov and Tapio Salakoski. On estimating the scale of national deep web. InDatabase and Expert Systems Applications, pages 780–789. Springer, 2007M.K.Sherbiny El "Efficient fuzzy logic load frequency controller"Energy Conversion and Management 43(2002)1853-1863 Elsevier

Author's Profile

DHANASEKARAN SUBBIAH working as Assistant Professor, at Kalasalingam University Srivilliputtur, Tamilnadu, India. His Native place is Srivilliputtur. He received the B.E degree in Computer Science and Engineering from Madurai Kamaraj University, Tamilnadu, India in 2004 and the M.E degree in Computer Science and Engineering from Annamalai University, Tamilnadu, India in 2007. He is pursuing Ph.D., in Kalasalingam University, Krishnankoil, and Tamilnadu, India under the Guidance of DR.V.VASUDEVAN, Eminent Professor and Registrar of Kalasalingam University. He has published many Research papers in Reputed International Journal and International conferences. He is a Life time member of ISTE (Indian Society for Technical Education).He is currently engaged in research in cloud computing and artificial intelligence.