Improve Efficient Keywords Searching Data Retrieval Process in Cloud Server

S. Saravanan
Department of CSE
M.Kumarasamy College of Engineering
Karur,Tamilnadu,India.
jeyasaraa@gmail.com

T.Abirami
Department of IT
Kongu Engineering College
Perundurai,TamilNadu,India
abi.it@kongu.edu

P. Pandiaraja
Department of CSE
M.Kumarasamy College of Engineering
Karur,Tamilnadu,India.
pandiarajap@mkce.ac.in

Abstract--**Over the past several decades there is an exceptionally large improvement in the computer technology which leads to an uncountable number of data and information emerging in and all over the world. Due to this tremendous and huge dump of data as well as web data most popular search engines are experiencing a lot of irrelevant retrieval of data. To distinguish a precise information pursuit furthermore to produce information that originates from anywhere. Too substantial to store on a search engine to such an extent that the PCs are associated with each other by the cloud search engine. Plan of search engines and its foundation grave is fundamentally focused. Search Engine Optimization is gathering of strategy and takes after by which any site can support positioning in internet searcher.**

Keywords—Browsers, Social computing, Search engige, Data collection, Page ranking.

I. INTRODUCTION

The aim of this section is to develop the accurate data search engine. In the cloud there are enormous data are available. To retrieve the query related corresponding data is challenging task in the cloud[1]. So in this chapter, the new search engine is developed. Easy and cost-effective setup, manages, and scale a search engine solution for your site or application in the cloud is overseen by cloud web search engine[2]. Search engines are encountering a great deal of unimportant recovery of information due to the gigantic and colossal dump of information and additionally web data[3]. Furthermore, the data itself may be too large to store on a single machine such that the computers are interconnected with each other by the massive internet storage technologies[4]. This approach mainly focuses on the design of search engines and its infrastructure grave[5]. By utilizing Search Engine Optimization is gathering of strategy and takes after, by which any site can help positioning in a web crawler. Moreover, examining procedure is utilized to stay away from the uneven dissemination of information. In cloud Search engine delivers low-inertness and the information appearance will expand the proficiency in its optimized search[6].

II. CLOUD SEARCH ENGINE ARCHITECTURE

In cloud metaphysics and a web interface comprises of three specializes in Cloud framework. The agent-based search engine is a cloud that counsels cloud ontology for thinking about the relations of Cloud administrations. Request from consumers is given to the cloud server, OS type, CPU speed. To fulfill consumer's requirements, the yields of the cloud are arrangements of cloud administrations requested by the rating of their appropriateness[7]. Cloud administrations are dictated by a Cloud benefit disclosure specialist that reason about cloud benefits by counsel cloud metaphysics. By following the cloud system operates, cloud administrations publicize their cloud suppliers by enlisting their administrations into the database of the cloud. Utilizing a web interface inquiries are sent to the cloud framework if the client utilizing any cloud framework as a part of presently. Three main components of cloud are explained below;

Query processor: In the cloud, the client sends a question then the query processor investigates the question and changes over the Query into a proper frame.

Reasoning: The comparison between two cloud ideas is controlled by counseling cloud metaphysics in the reasoning system.

Rating: Aggregated comparability is resolved and requested from most noteworthy of least administration utility in the rating procedure. The highest benefit utility would be chosen and different administrations which have high administration utility would be chosen to in the subsequent page in cloud benefit.
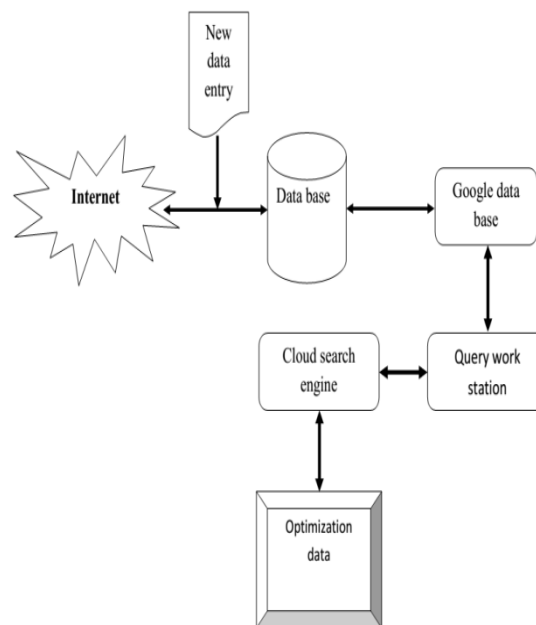

Fig. 1. Cloud Search Engine

Google crawlers are done non-sponsored listing results about because of the postings as indicated by ranking algorithm Figure 1. The pages are sorted order to Google. At the Stack concept using and the correct corner supported posting are likewise appearing on the List In First Out(LIFO) page. A search box is used by the user query; it may be from his country or from World Wide Web. Search Engine Optimization (SEO) is mainly used for selecting higher ranking search engines. For example, Google, Yahoo, Ask, Bing, and so forth rank site pages in view of specific elements that influence its positioning all the significant web indexes, the privilege sorts of signs on the website pages producing the SEO. Both content and related coding of the site to enhance its and unmistakable quality in natural inquiries made by the web crawlers to the center procedure utilized as a part of SEO is to redesign. Ordinarily get a higher number of guests the streamlined sites acquire better positions. This examination depends on surveying diverse enhancement strategies for individual pages or the whole site to make them seek engine neighborly. Review additionally basically breaks down and compresses the center procedures proposed in the contemporary writing. A near investigation of the past research work with respect to the systems utilized as a part of SEO and pinpoints certain crevices in the known site design improvement strategies are this paper offers.

III. NEED FOR SEARCH ENGINE

The interest of search engine in our context has a similar functionality as a Google desktop for security tokens. Here, we use the methodology introduced in the Information Retrieval literature for full-text search. Then, a document relates to any form of data files, terms relate to any form of data/metadata elements, term frequencies relate to data/metadata element weights and a query is equivalent to a full-text search. It has been widely studied by the information retrieval community for decades.

The core problem is, given a collection of documents and a user query expressed as a set of terms {ti}, to retrieve the k most relevant documents that the queried is granted. In the wide majority of works, then

The Tf - IDF score is used to rank the query results:

$$tf - id\,f(d) = \sum t \in Q \log\left(fd, t + 1\right) \log\left(N / Ft\right)$$

where fd, t is the frequency of term t in document d, N is the total number of indexed documents, and Ft is the number of documents that contain t. This formula is given for an illustrative purpose, the weight between fd, t and N/Ft varying depending on the proposals. Classically, full-text search queries are evaluated efficiently using a B tree-like inverted index storing for each term t the number Ft and the list of (d, fd, t) pairs.

The Page Rank Algorithm

Google's founders Larry Page is named page rank algorithm is used for identifying the significance of web pages separated in the range 1-10, where 10 represents full score (Higher PR value) and 1 represents lower PRv value[8]. r(Pii) Denoted Page Rank of a page Pii, is the sum of the page ranks of all pages. Pii The formula of page ranking Pii and │p1│ is the number of out links from a page

p1 into pointing set of pages BP. Pages-p1, p2...pn linked to the page x, and then the page x has its PRv value as follows:

PRv(x) = (1-d)+d{PRv(P1)/C(P1)+PRv(pn)/C(pn)}

Wherever, PR=Page Rank, d=damping factor and C=count of leaving links from a page.

Figure 2 Increase the number of inbound links, improved ranking is advantages of Page Ranking. Links from sites pertaining to the same topic as your site more importance is given. Sites having are more benefits from low outbound links on the links. Links from sites towards contributing more authority to the PR of your site (Higher PR). Over a period of time links must be built progressively.
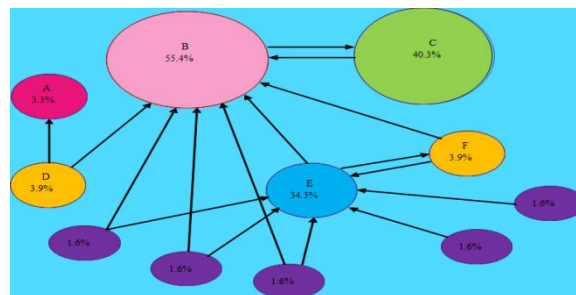

Fig. 2. The page rank Algorithm

Mathematical Page Ranks are expressed in percentages for a simple network. (Logarithmic table is used by Google.) Even though there are fewer links to C, Page E has a lesser Page Rank than Page C, hence it has high value as the one link to C comes from an important page. The web surfers will reach Page E 8.1% of the time when 15% likelihood of jumping to a page chosen at random from the entire web and random pages has an 85% likelihood of choosing a random link from the page they are currently visiting. (Corresponds to a damping factor of 85%, the 15% likelihood of jumping to an arbitrary page) Without damping other pages would have Page Rank zero rather than all web surfers would eventually end up on Pages A, B, or C. Page A effectively links to all pages in the web in the presence of damping even though there are no outgoing links of its own.

IV. SEO TECHNIQUES

Off-page Optimization: Off-Page SEO techniques to improve your search engine optimization includes social media blogging, blog marketing, forum marketing, etc., The techniques are Directory Submission, Social media, Video, Sitemap, Article Submissions, Profile posting, Forum posting, the steps for off page optimization Figure 4.
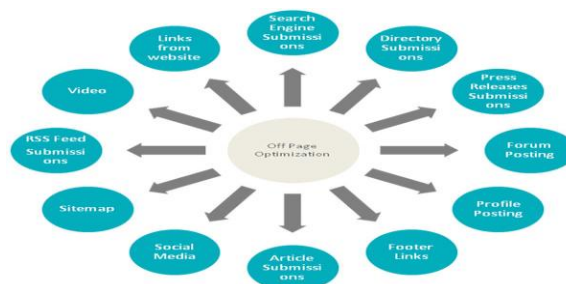

Fig. 4 .Off-page optimization

Back connecting: On other all around presumed sites Place your site interfaces. This makes a difference upgrade page rank. Choose presumed web facilitating organization choose .Have its own particular static IP address of the website. Article Distribution to different sites. Include Forum and Blog Commenting. Your Blog and Really Simple Syndication (RSS) Feed Submit. Site in catalogs List ". gov" and ".edu" joins. Social bookmarking are include. Check shares of your site content for various media occasionally. Use clear content for connections that ought to be found nearby. Informal organization: the long-range informal communication sites connections to the place. Multiply your site's substance informal communication locales by making your web-joins/site pages on these destinations on multiply your site's substance .to add to your records and eventually visit your site This makes a difference another individual. Bunches: Join gatherings, Yahoo, MSN, Google and post remarks on these bunches. To dependably have a reference/connection to your site or the particular website page. Too canvassed in this technique is Sending an email to masses. Place remarks containing joins of your site and join online groups and examination sheets. To share your connections furthermore, embed connections of the rumored destinations in yours.Website pages Utilize trusted sites .Where your site is transferred Check area.

On-Page Optimization: On-page optimization measures taken in the website to improve its search rankings. It is a factor effect on your Web site or Web page listing in natural search results.It incorporates those strategies that should be possible on the pages of a site it incorporates those methods. The web index crawlers read the site content this procedure makes a difference. Exhibits quality and will achieve higher situated pages with the eradicable site. The techniques are the location of the Keyword Density, Page loading, meta tags, meta description, Body Tags, Friendly URL, Image URL and outgoing link, for On page optimization Figure .5.

Fig. 5. On Page SEO

Area name and title of the site/pages ought to contain the watchword for which the site is required to be streamlined. For the site, substance choose spellbinding content. Figure out watchword states through which site is intended to be enhanced and put these catchphrases in headings, stay tag and Alternative labels and also in the primary substance/content. Utilize most recent apparatuses and create a flawless design of the site that can be effortlessly slithered also, programming for site advancement (e.g., WordPress). The sites substance ought to follow the constantly changing internet patterns with a regular redesign site with new and one of a kind substance. The most recent watchwords/popular expressions that are more apropos to their site substance in the site pages contain.Catchphrases utilized for streamlining ought to likewise reflect in all the

connecting structures set on the site pages. Video, pictures, sound and different substance ought to be upgraded on the same lines. Select watchwords in the wake of dissecting its thickness by utilizing SEO Tools. By the Google, bug utilize watchwords that can be slithered effortlessly.In the title tag of the page put watchword. URL for the sites use brief and succinct.

V. PARTS OF SEARCH ENGINE

There are several parts of search engine and they are listed as follows;

Teeming component: Web archive accumulation that is not restricted by topographical requirements and can develop unbounded that comprises the digital distribution center. All web search engines have a flatterer module to information gathering and order undertakings. Crawler module programming is done information accumulation and order of the web's records.

Page Repository component: In the page repository which site pages are briefly put away, bugs give the outcome with new website pages are put away in page store. The new pages until they are sent to the ordering module by store. Catchphrases and meta labels of the indexing part are secured to make a compacted form of the page. Making a portrayal of the page that is put away in different files in a packed way, the new uncompressed page will take from ordering module. Page storehouse is expelled after the uncompressed page is accessible.

Indexes component: Profitable data accessible in indexes for each webpage. In the structure file staying profitable data with respect to the hyperlink structure of pages is put away. Some of the time abounding gets to discover structure list unordered pages. To fulfill specific undertaking questions special reason files on pdf records or ppt documents are kept up.

Search Engine Ranking Parameters: Ranking module is done separating of applicable pages and positioning of them all together in view of the general score. Basic focuses on which the positioning score is distribute are considered as positional parameters.

Query Irrelevant Method: Overflowing, indexer, and repository comprise of query irrelevant method. Progressing constant process is a query irrelevant method. Question Irrelevant Method of every component works inside the cloud controller. Abounding acquires gets put away in the vault, which is available in the principle stockpiling. In main, storage lists will be available and in cloud controller ordering module will be available.

Query-relevant Method: Query Module and a Ranking Module are comprised of Query applicable method. Inquiry on each component subordinate module frame a part of the cluster. In cluster controller all the positioning calculations and parameters of a question module accessible.According to the personalization of information present in primary stockpiling positioned pages are put away in sub-stockpiling. Contingent upon the sort and information set essential by the inquiry and that is dictated by cloud manager. The inquiry result is returned quickly to cloud controller if information match is found.

VI. NUMERICAL RESULTS

The experimental results obtained from proposed approach are explained in this section. The experimental result is compared with Google, Bing, and Yahoo.
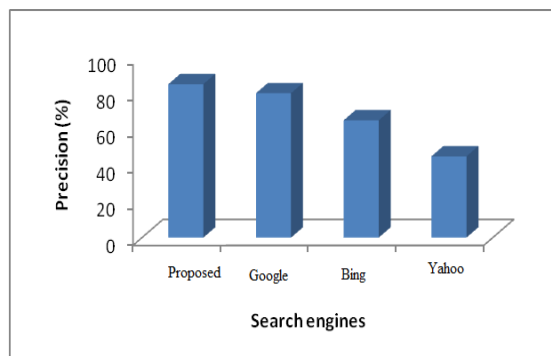

Fig. 6 .Performance analysis of different search engine using precision measure

To evaluate the performance of test queries was picked subjectively in each of the spaces are academic, entertainment and sports. The explanation for selecting the request in these three spaces is to cover a broad assortment of inquiries on the web. These queries were issued to the web index to perceive the total arrangement of significant records for the given question from the recouped search results.The Figure 6 shows that optimization results for various search engines such as Google, Bing, and Yahoo. By comparing these search engines with the proposed system with help of Stack Data Structure. The proposed system is varying much higher than other search engines. The Google is varying up to 80, the Bing vary up to 60 and the yahoo vary up to 45 and the proposed method achieves a maximum of 81.
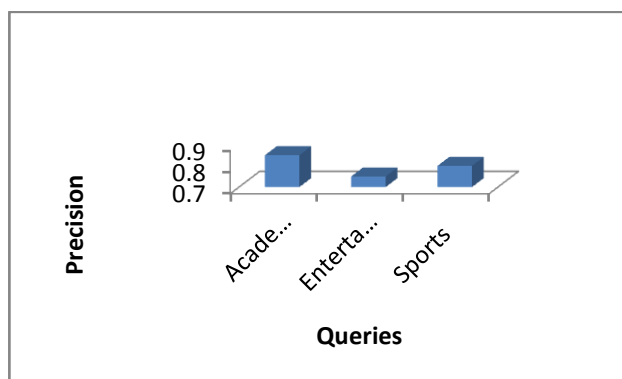

Fig. 7. Performance analysis of proposed methodology based on precision measure

Figure7 represents the performance of proposed approach using a precision measure for different domain quires. The domains used for queries are academic, entertainment and sports. By searching these queries academic give better performance than other queries by using the proposed method. Academic, entertainment and sports are the queries for checking the search engines. The academic is variable up

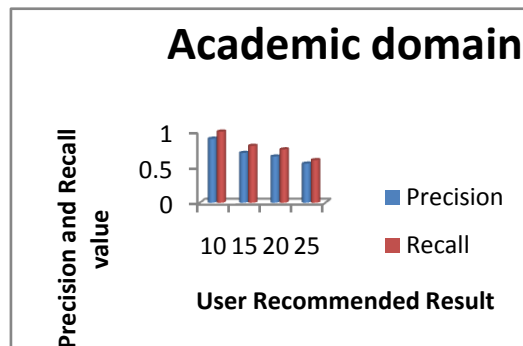to 0.85, entertainment is varied up to 0.75 and sports vary up to 0.8.

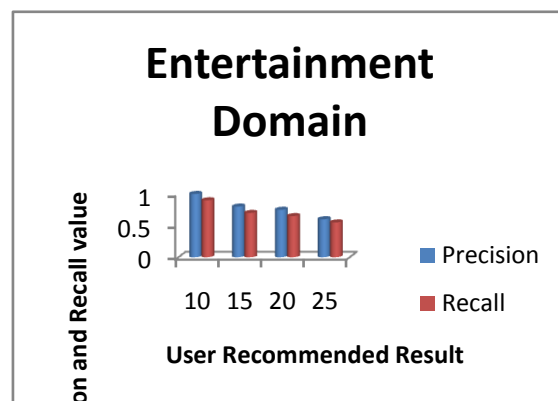Fig. 8. Performance analysis of proposed methodology by varying number of retrieved result

Fig. 9. Performance analysis of proposed methodology by varying number of retrieved result
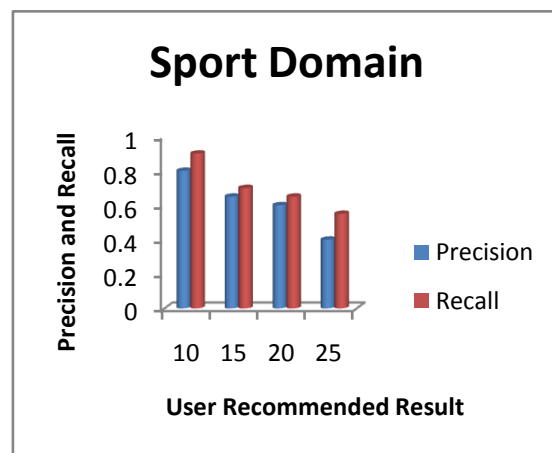

Fig. 10. Performance of proposed methodology by varying number of retrieved result

In Figure 8, 9, 10 show the performance of the proposed methodology by varying the number of retrieved results. In the academic domain precision and recall is analysis, the recall will be higher than precision, in the entertainment domain performance of precision and recall analysis. The precision is higher than the recall will analyze in the graph. Sports domain also shows the performance of precision and recall, the precision is higher than recall. The accuracy based investigation of the proposed approach with three noteworthy web indexes, Google, yippee and Bing. The examination demonstrated that the proposed web crawler has better accuracy over the other web indexes. Utilizing the web

crawler acquired the pertinent URLs for Google, Yahoo, and Bing. At long last, in view of the retrieved URLs check we rank the web index. From comprehending, among the three web index "Google" is better other two internet searcher. Additionally, demonstrates the recovery and positioning, execution of diversion space. Here, we demonstrate the outcome for five questions in view of the amusement area. Every question we get the significant URLs for a various search engine, Google, hurray and Bing. We rank the web indexes in light of the recovered significant URLs check.

VII. CONCLUSION

This examination was required to plan an internet searcher fit for returning web list items considering client's interests. The introduced search engine uses a Search optimization methodology in positioning the web assets and web crawlers. This research examines various domains like academics, entertainment, sports by looking into Google, Yahoo, and Bing. This proposed methodology gives better execution on web index optimization. An inquiry interface is intended to send client's pursuit questions to the different web crawlers and to show results to the client. The outcome area demonstrates the execution of URLs positioning and web index positioning.

REFERENCES

[1] [1].M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, "A view of cloud computing," in Communications of the acm, vol. 53, no. 4, April 2010, pp. 50-58.

[2] D. Karaboga, and B. Basturk, "On the performance of artificial bee colony (ABC) algorithm," in Applied Soft Computing, 2008,pp. 687–697.

[3] J. Hu, J. Gu, G. Sun, and T. Zhao, "A scheduling strategy on load balancing of virtual machine resources in cloud computing environment," in 3rd Int. Symp. on Parallel Architectures, Algorithms and Programming(PAAP), 2010, 18- 20 Dec. 2010.

[4] R. N. Calheiros, R.Ranjan, C. A. F. D. Rose, and R. Buyya, "CloudSim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms," Software : Practice and Experience, Vol. 41, No.1, Jan. 2011, pp. 23-50.

[5] Thilagamani, S. and N. Shanthi, 2010. Literature survey on enhancing cluster quality. Int. J. Comput. Sci. Eng.., 2: 1999-2002.

[6] S Saravanan, V Venkatachalam, Improving map reduce task scheduling and micro-partitioning mechanism for mobile cloud multimedia services [J]. Int J of Adv Intell Paradigms 8(2), 156–167 (2016)

[7] S Saravanan, V Venkatachalam,"Advance Map Reduce Task Scheduling algorithm using mobile cloud multimedia services architecture" IEEE Digital Explore,pp21-25,2014.

[8] S. Saravanan, V. Venkatachalam. "A New Method for Acquiring Relevant Data Partitioning by Optimization Techniques" , 2017 Second International Conference on Recent Trends and Challenges in Computational Models (ICRTCCM), 2017