

## A Novel Technique Optimization Search Engine Using Accurate Keywords Discover in Cloud

<sup>1</sup>S. Saravanan and <sup>2</sup>V. Venkatachalam

<sup>1</sup>Department of Computer Science and Engineering, Kumarasamy College of Engineering,  
Karur, Tamil Nadu, India

<sup>2</sup>The Kavery Engineering College, Salem, Tamil Nadu, India

**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. The major aspire of this proposed improved weis to identify an accurate data search and also to generate data that comes from anywhere. Furthermore, the data itself may be too large to store on a single machine such that the computers are inter connected with each other by the massive internet storage technologies. This approach mainly focuses on design of search engines and its infrastructure grave. Improved micro partitioning is a modularized approach of cloud computing mainly framed to overcome the pitfalls in the traditional search engine and also in manipulation of large information stored in a single computer. This study used search engine optimization is group of method and follows by which any website can boost ranking in search engine. SEO is two ways, first one is on page and second one is off page. Additionally in order to avoid the uneven distribution of data the data sampling technique is used. Henceforth, the search engine in cloud produces low-latency and the data materialization will increase the efficiency in its optimized search and thus outperforms the traditional approaches.

**Key words:** Search engine, data sampling, micro partitioning, page ranking, optimization, irrelevant data

---

### INTRODUCTION

The study goes for wedding the website improvement and positioning procedure with the advantages of cloud innovation. This approach mostly concentrates on plan of web crawlers and its foundation grave. Enhanced Micro dividing is a modularized approach of distributed computing fundamentally surrounded to defeat the pitfalls in the conventional web crawler furthermore in control of huge data put away in a solitary PC. Moreover, the information itself might be too expansive to store on a solitary machine to such an extent that the PCs are entomb associated with each other by the huge web stockpiling innovations. This approach mostly concentrates on outline of web crawlers and its foundation grave. Enhanced micro apportioning is a modularized approach of distributed computing mostly confined to conquer the pitfalls in the conventional web crawler furthermore in control of expansive data put away in a solitary PC. This study utilized search engine optimization is gathering of technique and takes after by

which any site can support positioning in internet searcher. SEO is two ways initial one is on page and second one is off page. Also with a specific end goal to stay away from the uneven dispersion of information the information examining method is utilized. Hereafter, the search motor in cloud delivers low-inertness and the information emergence will build the effectiveness in its upgraded inquiry and consequently beats the customary methodologies

**Literature review:** Disseminated processing is the since a long time prior envisioned vision of figuring as an utility, where cloud customers can remotely store their data into the cloud with a specific end goal to welcome the on-demand astonishing applications and organizations from a typical pool of configurable enlisting resources (Ma *et al.*, 2014). As cloud computing gets the opportunity to be dominating, more tricky information are being united into the cloud, for instance, messages, singular prosperity records, association finance data and government report set cetera. The way that data

proprietors and cloud server are not any more attracted out the same trusted zone may put the outsourced decoded data at danger (Iosup *et al.*, 2011) the cloud server may spill data information to unapproved substances (Javadi *et al.*, 2011) or even be hacked (Lai *et al.*, 2011). It takes after that sensitive data must be mixed before outsourcing for data insurance and doing combating unconstrained gets to. A champion among the most common ways to deal with do all things considered is through watchword based interest. Such catch phrase look system licenses customers to explicitly recuperate reports of interest and has been by and large associated in plaintext look for circumstances (Song *et al.*, 2014). For situated watchword based looking for process, differing investigators have prescribed various procedures. Among them an unassuming group of significant looks at are shown in this piece. Other than Zhangjie Fu have unveiled a feasible approach to manage deal with the issue of multi-catchphrase situated look for over encoded cloud data supporting proportional word questions. The standard duty of this study was packed in two points of view: multi-watchword situated interest to achieve more correct rundown things and proportional word based request to support identical word questions. Expansive trials on certifiable dataset were performed to support the approach, showing that the proposed game plan was incredibly fruitful and viable for multi-catchphrase situated looking in a cloud area. Furthermore, Meng *et al.* (2011) have elucidated the secure ranked keyword search over encrypted cloud data. Here, they describe and handle the issue of capable yet secure situated catchphrase look for over mixed cloud data. Situated look unbelievably overhauls structure usability by giving back the planning records in a situated orchestrate as for certain significance criteria (e.g., watchword repeat), appropriately making one phase closer towards down to business sending of security shielding data encouraging organizations in cloud computing. They first given a direct yet idealize improvement of situated watchword look under the bleeding edge Searchable Symmetric Encryption (SSE) security definition and demonstrate its inefficiency. To finish more sensible execution, they then exhibited a definition for situated searchable symmetric encryption and give a gainful blueprint by properly utilizing the momentum cryptographic primitive, mastermind ensuring Symmetric Encryption (OPSE) (Zhu *et al.*, 2011). Concentrated examination shows that their system course of action acknowledges “as-strong as could be normal in light of the current situation” security guarantee stood out from past SSE arranges while viably understanding the goal of situated watchword look for.

Wide test comes to fruition display the adequacy of their solution (Pathan and Monowar, 2012). Suman and Chempavathy have illuminated the efficient and secure retrieval of encrypted cloud data based on top-K multi catchphrases. In disseminated fig+uring on account of far reaching number of data customers and chronicles, it was fundamental for the request organization to allow multi-watchword question and give result closeness positioning to meet the viable information recovery require. The catchphrase word reference was static and can't be widened easily when the amount of watch words augmentations. Also, it doesn't take the customer direct and watchword get to repeat into record. The key duty of this study packed in two points of view: multi-catchphrase situated request to fulfill more exact rundown things and equal word based interest to reinforce identical word queries (Huang *et al.*, 2013). Here, the energy multi-watchword situated look approach; the catchphrase dictionary was static and can't be extended viably when the amount of watchwords increases (Rodrigues *et al.*, 2012). They build up a phenomenal tree-based record structure and propose an “unquenchable depth-first search” figuring to give profitable multi-watchword situated look for. The secured kNN figuring was utilized to scramble the record and question vectors and meanwhile ensure exact essentialness score number between encoded rundown and request vectors (Rodrigues *et al.*, 2012):

$$tf-idf(d) = \sum_{t \in Q} \log(fd, t + 1) \log(N/Ft)$$

Where:

$fd, t$  = The repeat of term  $t$  in report  $d$

$N$  = The total number of recorded records

$Ft$  = The amount of documents that contain  $t$

This formula is given for illustrative reason, the weight among  $fd, t$  and  $N/Ft$  fluctuating depending upon the proposals. Generally, full-content request are evaluated capably using a B-tree-like switched record securing for each term  $t$  the number  $Ft$  and the summary of  $(d, fd, t)$  sets.

**The page rank calculation:** The page rank figuring which is named after one of google's creators larry page is used to recognize the noteworthiness of site pages isolated in the range 1-10 where 10 addresses full score (Higher PR regard) and 1 addresses bring down PR regard. The page rank of a page  $P_i$ , implied  $r(P_i)$  is the total of the page ranks of all pages demonstrating into  $p_i$  formula of page ranking where BP is the course of action of pages

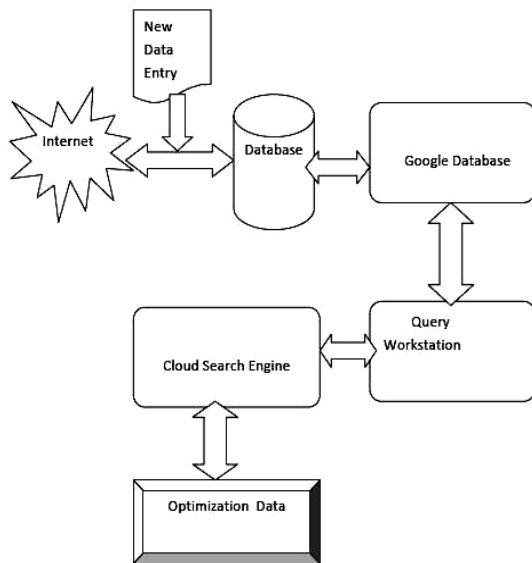


Fig. 1: Cloud search engine architecture

showing into  $P_i$  and  $|P_j|$  is the amount of out associations from page  $P_j$ . Accept page A has its pages- $p_1, p_2, \dots, p_n$  associated with it, then page A has its PR regard as takes after:

$$PR(A) = (1-d) + d\{PR(p_1) / C(p_1) + \dots + PR(p_n) / C(p_n)\}$$

Where:

$d$  = Damping variable

PR = Page rank

$C$  = Count of active connections from a page  
advantages of page ranking are higher the quantity of inbound connections, better to rank

Joins from locales relating to an indistinguishable subject from your site have more weight-age. Links from destinations having low outbound connections are more valuable. Joins from destinations having power (High PR) contribute more to the PR of your site. Links ought to be assembled bit by bit over a timeframe.

**Cloud search engine architecture:** There are four courses in a web search tool is watched. Non supported posting comes about because of the postings that are finished by Google crawlers as per their positioning calculation. For these posting we require not to pay anything to Google. The pages are positioned concurring their quality. Supported listing are additionally appeared in the outcome page at the top most and the right corner, for these posting we have to pay to Google. Seek box is utilized by the client for his inquiry; it might be from his

nation or from world wide web. Google moment begin writing out our question in Google, it begins showing our outcome dissecting every word. This component relies on upon the speed of association, numerous a times it doesn't deal with moderate association Fig.1.

## MATERIALS AND METHODS

**Components of search engine:** Components of search engine are shown in Fig. 2.

**Abounding part:** The web document gathering lives in an advanced circulation focus a virtual component that is not confined by topographical impediments and can create unbounded. Web files must do the data collection and request assignments in solitude. Accordingly, all web searchers have a crawler module. This module contains the item that accumulates and requests the web's documents. The proliferating programming makes virtual robots, called bugs that dependably scour the Web amassing new information and webpage pages and returning to store them in a central vault.

**Page repository segment:** The frightening little creatures return with new webpage pages which are by chance set away as full, aggregate site pages in the page storage facility. The new pages remain in the vault until they are sent to the requesting module where their key information, for instance, catchphrases and meta names is stripped to make a compacted type of the page. Indexing portion. The requesting module takes each new uncompressed page and focuses only the key descriptors, making a stuffed depiction of the page that is secured in various records. The uncompressed page is then removed from the page storage facility.

**Records segment:** The records hold the huge pressed information for each site page. In the substance record; substance, for instance, catchphrase, title and remain content for each page is secured in a compacted outline using a revamped report structure. Advance productive information concerning the hyperlink structure of pages is secured in compacted outline in the structure record. The proliferating from time to time gets to the structure record to find unordered pages. Novel reason records, for instance, records on pdf reports or ppt records are kept up to satisfy specific task questions

**Web search tool ranking parameters:** The raking module stations appropriate pages and positions all of them together of the general score designated to them. The situating parameters are those essential spotlights on

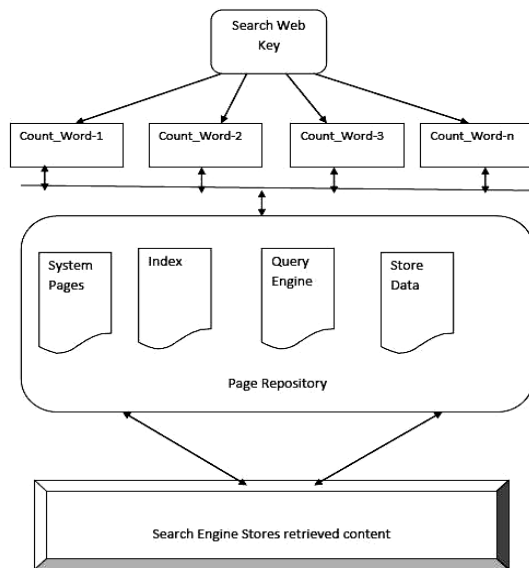


Fig. 2: Cloud search engine

which the situating score is doled out. Situating parameters are widely appointed close-by and off-site parameters. This study consider as a two head systems.

**Inquiry irrelevant method:** Request irrelevant method involves flooding indexer and repository. Address irrelevant method does not depend on the customer yet rather is an advancing perpetual process. Each component of query irrelevant method works inside the cloud controller. Beside storage facility the different parts of query irrelevant method exists in the cloud controller. Each one of the data which flooding secures escapes the storage facility which is accessible in the essential stockpiling. Requesting module will be accessible in cloud controller while indexes will be accessible in main storage

**Question pertinent method:** Request material method involves query module and a ranking module. Each component on request subordinate module outlines a part of cluster. Bundle controller contains request module with all the situating estimations and parameters. While situated pages are secured in sub-stockpiling as showed by the personalization of data present on a basic level stockpiling. At whatever point customer fires a request its goes to a particular gathering dependent upon the sort and data set required by the question and that is managed by cloud controller. At whatever point a data match is found then the request result is returned in a flash to cloud controller.

## RESULTS AND DISCUSSION

**Micro partitioning cloud search engine method:** Make pages and records, New pages, Edit and delete. Create website joins same as the title page. SEO methods include two imperative methodology: First is close-by streamlining second is off-site page progression site page the region optimization-site hand upgrade identifies with the practices which are used in the midst of the change of a webpage. Recorded underneath are the basic perspectives that need particular thought in the midst of the change of a webpage in order to make it web searcher very much arranged. The going with Steps are used. Territory name and title of the site/pages should contain the watchword for which the site is required to be progressed. Pick entrancing substance for the site substance. Understand catchphrase expressions through which site is proposed to be enhanced and put these watchwords in headings, hook tag and Alternative (Alt) names and moreover in the essential substance/content. Make a perfect plan of the site that can be adequately crawled. Use latest instruments and programming for website progression. Regularly redesign webpage with new and remarkable substance so that the destinations substance should fit in with the relentlessly changing Internet designs. Ensure that the site pages contain the latest catchphrases/prevalent expressions that are more pertinent to their site substance.

The catchphrases used for progression should in like manner reflect in all the associating structures set on the site pages. So additionally, video, pictures, sound and diverse substance should be upgraded on comparable lines. Select watchwords in the wake of analyzing its thickness by using SEO Tools. Use catchphrases that can be crawled easily by the Google creepy crawly. Placed catchphrase in title tag of the page. Substance of the page should contain such watchwords that can be easily found on pages in the midst of the change. Off-website page optimization-ff-page change relates to the practices through which the webpage and its substance are caused/duplicated over the internet to extend its action which in this manner helps updating Google situating of the website. The going with steps are used. Back associating: Place your site interfaces on other especially assumed locales. This updates page rank. Some other methodology consolidates. Pick an assumed web encouraging association. Site should have its own static IP address. Ensure article distribution to various destinations. Consolidate forum and blog commenting. Introduce your blog and RSS feed. List your site in indexes. Get “gov” and “edu” joins. Casual people

Table 1: Comparison of existing search engine

Existing search engine	Optimization cloud search engine
Lack of standards	Modularity
Platform dependency	Scalability
Security issues	Flexibility and security
High cost	Low cost
Slowly data search	Quickly data search
Inter cloud operations	Allows you to modify a component without disturbing the next one
Not accurate data search	Accurate data search

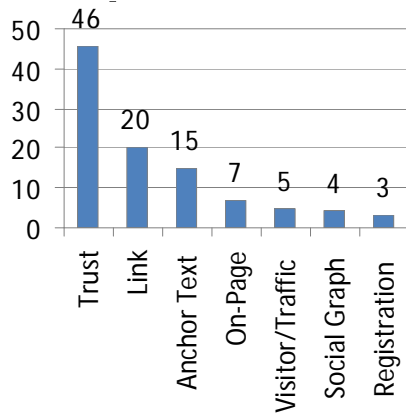


Fig. 3: Proposed results

group: Place associations of the interpersonal collaboration locales, for instance, Twitter, Facebook and Google+ on your site. Duplicate your webpage's substance onto these long range casual correspondence goals by making your web-joins/site pages on these regions. This helps different people to add to your records and finally visit your site. Groups: Join social events, for instance, Yahoo, MSN, Google and post comments on these get-togethers. Your comments should constantly have a reference/association of your site or the specific site page that you wish perusers to scrutinize. Sending email to masses is similarly secured in this framework.5.1

**Pseudocode;****Read keyword:**

```

Countrows = No.of.relevant pages retrieved from DB
Loop_count = 0
WHILE Loop_count < countrows
    READ contents of relevant pages
    Word_count [Loop_count] = occurrences of keyword in retrieved pages
    URL [Loop_count] = url of current page
    IF Word_count[Loop-count] > Word_count[Loop_count -1]
        WRITE URL [Loop-count] at position = position - 1
        WRITE URL [Loop-count -1] at position = position + 1
    ELSE
        WRITE URL[Loop_count] at position = position+1
    END IF
    ADD 1 to Loop_count
END WHILE

```

**Comparison with existing search engine and optimization cloud search engine:** The following table shows that comparison between the existing search engine and optimization cloud search engine (Table 1).

**Result of optimization cloud search engine:** Figure. 3 demonstrates that improvement comes about different traits of website page, for example, Trust, Link popularity, Anchor text on-page keyword, Visitor/traffic, Social graph papers, Registration

**CONCLUSION**

This approach consider a useful quickly information look cloud environment where the client holding up time and server benefit time are obscure. Our principle commitment is to outline a visually impaired internet planning plan by together considering postponement and vitality among the servers. In particular, its dole out accessible sight and sound servers in view of the last time-space data of the client's solicitations and course the heterogeneous interactive media streams as per the principal start things out served run the show. Moreover, applies the visually impaired planning plan to a substance suggestion framework and give the point by point execution steps. Broad reproduction comes about show that the proposed plan can proficiently plan heterogeneous mixed media streams to fulfill dynamic QoS prerequisites in a handy cloud environment .

**ACKNOWLEDGEMENTS**

Researcher would like to thank my colleagues and my guide who have supported me towards to completed this study successfully. I am thankful to my friends and colleague members of CSE department for their continuous direction and guidance and timely support extended for me to carry out this paper successfully.

**REFERENCES**

- Huang, L., H.S. Chen and T.T. Hu, 2013. Survey on resource allocation policy and job scheduling algorithms of cloud computing 1. J. Software, 8: 480-487.
- Iosup, A., S. Ostermann, M.N. Yigitbasi, R. Prodan, T. Fahringer and D.H. Epema, 2011. Performance analysis of cloud computing services for many-tasks scientific computing. IEEE Trans. Parallel Distrib. Syst., 22: 931-945.
- Javadi, B., D. Kondo, J.M. Vincent and D.P. Anderson, 2011. Discovering statistical models of availability in large distributed systems: An empirical study of seti@home. IEEE. Trans. Parallel Distrib. Syst., 22: 1896-1903.

- Lai, C.F., J.H. Chang, C.C. Hu, Y.M. Huang and H.C. Chao, 2011. CPRS: A cloud-based program recommendation system for digital TV platforms. *Future Gener. Comput. Syst.*, 27: 823-835.
- Ma, T., Y. Chu, L. Zhao and O. Ankhbayar, 2014. Resource allocation and scheduling in cloud computing: policy and algorithm. *IETE Tech. Rev.*, 31: 4-16.
- Meng, S., L. Liu and T. Wang, 2011. State monitoring in cloud datacenters. *IEEE. Trans. Knowl. Data Eng.*, 23: 1328-1344.
- Pathan, A.S.K. and M.M. Monowar, 2012. Future trends in security issues in internet and web applications. *Int. J. Comput. Sci. Eng.*, 9: 275-277.
- Rodrigues, J.J., L. Zhou, L.D. Mendes, K. Lin and J. Lloret, 2012. Distributed media-aware flow scheduling in cloud computing environment. *Comput. Commun.*, 35: 1819-1827.
- Song, J., X. Liu, Z. Zhu, D. Zhao and G. Yu, 2014. A novel task scheduling approach for reducing energy consumption of MapReduce cluster. *IETE. Tech. Rev.*, 31: 65-74.
- Zhu, W., C. Luo, J. Wang and S. Li, 2011. Multimedia cloud computing. *IEEE. Signal Process. Mag.*, 28: 59-69.