International Journal of Scientific Research in Engineering & Technology
Volume4, Issue3 (May-June 2024), PP: 43-46.
www.ijsreat.com ISSN No: 2583-1240

Enhancing Search Precision Through Dictionary Lexical Analysis

A. Joyal Jones[1], Dr. V. Vaidehi[2]
[1]PG Student, Department of Computer Applications, DR MGR Educational and Research institute Chennai, Tamil Nadu, India.
[2] Professor, Department of Computer Applications, DR MGR Educational and Research institute Chennai, Tamil Nadu, India.

To Cite this Article: *A. Joyal Jones[1], Dr. V. Vaidehi[2], "Enhancing Search Precision Through Dictionary Lexical Analysis", International Journal of Scientific Research in Engineering & Technology Volume 04, Issue 03 (May-June 2024), PP: 43-46.*

Abstract: *Efficient and accurate search precision is crucial in information retrieval systems, especially as the volume of digital content continues to expand exponentially. This paper introduces a novel approach, "Enhancing Search Precision through Dictionary Lexical Analysis," designed to improve the precision of search queries by leveraging advanced lexical analysis techniques. The proposed system incorporates dictionary-based lexical analysis to enhance the understanding of user queries and refine search results. By mapping words to a comprehensive dictionary, the system identifies synonyms, acronyms, and related terms, creating an enriched semantic representation of the query. This process aids in capturing the nuanced context and intent behind user search queries, leading to more accurate and context-aware search results. Key components of the system include the construction and maintenance of a dynamic dictionary, lexical analysis algorithms, and integration with existing search engines. The dictionary is continuously updated to adapt to evolving language patterns and user preferences, ensuring the system remains effective over time. Through practical demonstrations and performance evaluations, this research showcases the impact of "Enhancing Search Precision through Dictionary Lexical Analysis" on search result relevance and user satisfaction. Comparative analyses against traditional search algorithms highlight the system's ability to provide more contextually relevant results, especially in ambiguous or specialized query contexts.*

Key Word: *Data Sharing, Searchable Encryption, Binary Search, Data Mining.*

I.INTRODUCTION

The use of cloud computing for processing and storing data is growing in popularity. Since the emergence of cloud computing, data owners are being encouraged to move their intricate data management systems from on-site locations to commercial public clouds in order to take advantage of the increased flexibility and cost benefits. Data must be stored in an encrypted format in the cloud to prevent access by unwanted users, likely including cloud service providers. Meanwhile, there needs to be an effective search access control in place for material that is designed to be shared and retrieved. Searching for keywords in the data is a typical process. Access control for in-cloud data is typically enforced by users, and search operations on encrypted cloud data are carried out at the cloud servers. With the help of this framework, we present innovative schemes like the multi-keyword dictionary with Boolean search, fuzzy search, and wildcard keyword search, which make use of the latest primitive encryption technology, "AES," in order to support the derivation of the search capability and enforce fine-grained access control. [1].

In the age of cloud computing, there is a lot of interest in the technology of keyword search across encrypted data. Before being outsourced to cloud servers, sensitive data must be secured to protect user privacy. Creating a safe and effective search plan for encrypted data requires the application of methods from several fields, including information retrieval for index representation, search efficiency algorithms, and appropriate cryptographic protocol design for system security and privacy. During the first part of the project, techniques for ranking and multi-keyword dictionary-based searches over an encrypted cloud were designed. [2].

Finding an effective solution to the issue of customers efficiently looking through encrypted cloud data stored on cloud servers is a top priority. The integration of many keyword searching mechanisms, including fuzzy search, wildcard keyword search, and multi-keyword dictionary with boolean search, is addressed in this phase 2 project work. This is done to protect sensitive content from unauthorized users by using cloud data. However, cloud customers find it challenging to efficiently search for their data due to encryption. [3]

Massive amounts of data are stored on cloud servers; the speed at which a user can retrieve his data depends on the effectiveness of the searching strategies used. Not only should keyword search methods be quick, but the security of such private information cannot be jeopardized. It has been attempted in this project work to analyze different keyword searching strategies and ascertain how they make sense while looking over encrypted data. [4]

Thesis has been organized as follows Chapter 1 describes the objective and problem statement of the project. Chapter 2 describes the notable research literature relevant to the study. Chapter 3 contains detailed design with architecture and flow

chart. It contains list of modules and algorithms used for implementing each module. Chapter 4 contains input and output for each module and results obtained for each module with screenshots are shown. Final results and time analysis for each module is presented. Chapter 5 concludes the thesis and explains the future works. [5]

II.LITERATURE SURVEY

Netflix is the leading provider of on-demand Internet video streaming in the US and Canada, accounting for 29.7% of the peak downstream traffic in US. Understanding the Netflix architecture and its performance can shed light on how to best optimize its design as well as on the design of similar on-demand streaming services. In this paper, we perform a measurement study of Netflix to uncover its architecture and service strategy. We find that Netflix employs a blend of data centers and Content Delivery Networks (CDNs) for content distribution. We also perform active measurements of the three CDNs employed by Netflix to quantify the video delivery bandwidth available to users across the US. Finally, as improvements to Netflix's current CDN assignment strategy, we propose a measurement-based adaptive CDN selection strategy and a multiple-CDN-based video delivery strategy, and demonstrate their potentials in significantly increasing user's average bandwidth. [6].

ClosestNode.com is an accurate, scalable, and backwards-compatible service for mapping clients to a nearby server. It provides a DNS interface by which unmodied clients can look up a service name, and get the IP address of the closest server. A shared system for performing such a mapping amortizes the administration and implementation costs of proximity-based server selection. It is aimed at minimizing the amount of effort required for system developers to make new and existing infrastructure services proximity-aware. [7]

The shared nature of the network in today's multi-tenant datacenters implies that network performance for tenants can vary significantly. This applies to both production datacenters and cloud environments. To this effect, the key contribution of this paper is the design of virtual network abstractions that capture the trade-off between the performance guarantees offered to tenants, their costs and the provider revenue. To illustrate the feasibility of virtual networks, we develop Oktopus, a system that implements the proposed abstractions. Using realistic, large-scale simulations and an Oktopus deployment on a 25-node two-tier testbed, we demonstrate that the use of virtual networks yields significantly better and more predictable tenant performance. Further, using a simple pricing model, we find that the our abstractions can reduce tenant costs by up to 74% while maintaining provider revenue neutrality. [8]

Large datacenter operators with sites at multiple locations dimension their key resources according to the peak demand of the geographic area that each site covers. The demand of specific areas follows strong diurnal patterns with high peak to valley ratios that result in poor average utilization across a day. In this paper, we show how to rescue unutilized bandwidth across multiple datacenters and backbone networks and use it for non-real-time applications, such as backups, propagation of bulky updates, and migration of data. Achieving the above is non-trivial since leftover bandwidth appears at different times, for different durations, and at different places in the world. To this end, we have designed, implemented, and validated Net Stitcher, a system that employs a network of storage nodes to stitch together unutilized bandwidth, whenever and wherever it exists. It gathers information about leftover resources, uses a store-and-forward algorithm to schedule data transfers, and adapts to resource fluctuations. We have compared NetStitcher with other bulk transfer mechanisms using both a testbed and a live deployment on a real CDN. [9].

The data centers used to create cloud services represent a signifi- cant investment in capital outlay and ongoing costs. Accordingly, we first examine the costs of cloud service data centers today. The cost breakdown reveals the importance of optimizing work completed per dollar invested. Unfortunately, the resources inside the data centers often operate at low utilization due to resource stranding and fragmentation. users and increases reliability in the presence of an outage taking out an entire site. However, without appropriate design and management, these geo-diverse data center networks can raise the cost of providing service. Moreover, leveraging geo-diversity requires services be designed to benefit from it. To attack this problem, we propose (1) joint optimization of network and data center resources, and (2) new systems and mechanisms for geo-distributing state. In this paper we present two constructions of Fuzzy IBE schemes.[10].

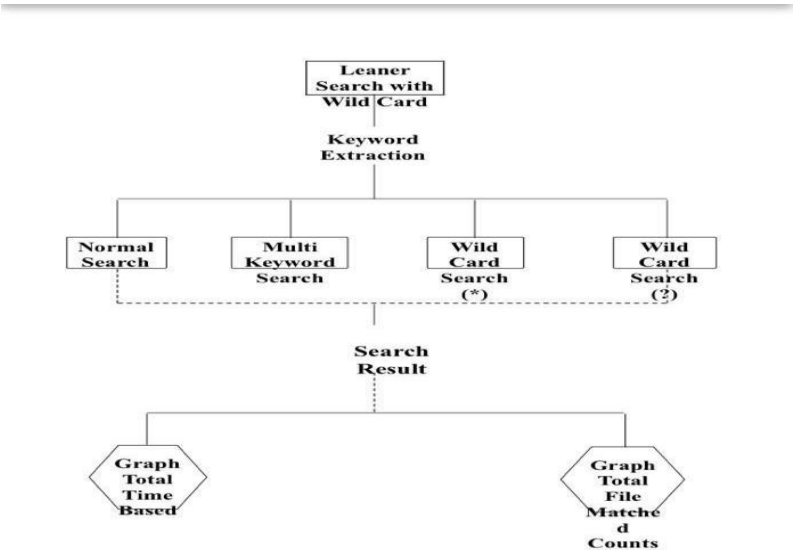
Proposed System:

The Wildcard is an advance search technique that can be used to maximize your search results in library databases. Wildcard are used in search terms to represent one or more other characters. A question mark (?) may be used to represent a single character, anywhere in the world. It is most useful when there are variable spellings for a word, and you want to search for all variants at once.

For example: Searching for Java would return java.

The results show that semantic knowledge is indispensable for short text understanding, and in this knowledge-intensive approaches are both effective and efficient in discovering semantics of short texts Out performs existing state-of-the-art approaches in the field of short text understanding. Then the query will be applied for XML and appropriate clause, document which was sit for by the users. This gives the result related to the query. Then the XML document is modified and the process is repeated until the goal is reached. The modifications are stored in the same XML document which was already formed by the values of support and confidence combine all this document parsed is evaluate as again when required.

Architecture Diagram



III.MODULES

Admin: An administrator typically has the highest level of access and control over a system or platform. They manage user accounts, set permissions, configure settings, and ensure the smooth operation of the system. Admins often have the authority to make critical decisions regarding the system's configuration and security.

Data Owner: The data owner is responsible for the overall management and control of specific sets of data within an organization. They determine who has access to the data, establish data usage policies, and ensure compliance with regulations such as data privacy laws. The data owner has the ultimate responsibility for the security, integrity, and accuracy of the data.

User: A user is an individual who interacts with a system or platform to perform tasks or access resources. Users have limited access compared to administrators and typically have permissions tailored to their specific role or job function. Users may include employees, customers, or any other individuals authorized to access the system.

IV.RESULT AND DISCUSSION

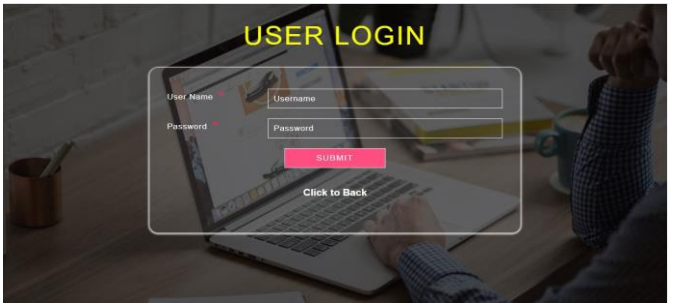


Fig 1.1:login page.

Login: The admin interface starts with a login page where administrators can enter their credentials (username and password) to access the system. This step ensures that only authorized individuals can access the administrative features and perform administrative tasks.

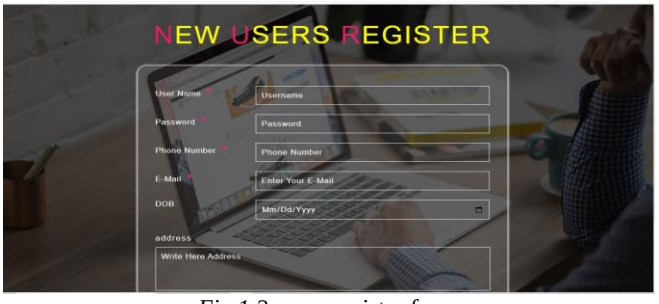


Fig 1.2: user register form.

The user registration form in Dictionary Lexical Analysis collects essential details such as name, email, and password.

This information enables users to create accounts, access personalized details to admin.

Fig 1.3: Admin login Page.

An admin login page is a web page specifically designed for administrators or privileged users to access the backend or administrative features of a website, application, or system. It typically requires a username and password for authentication and grants access to administrative functions such as managing users, content, settings, and other system configurations.

V.CONCLUSION

The proposed model of linear search algorithm will give the most frequent feeds of all the files. After that the feed search will be performed by the wildcard based Search on XQuery which will give us the filtered result. The obtained results are better than the existing methods. It will help the user to find his resources completely. The comparative analysis shows that time taken to fetch the searched keyword in XML document from the proposed linear search algorithm is lesser than the time taken to fetch the searched keyword in XML document from the existing CMT algorithm and also total file matched count in proposed linear search algorithm is better than the total file matched count in existing CMT algorithm. In terms of using wild card search (pattern like ??, ** and Multi keyword) time taken to fetch the searched keyword is lesser than existing method and also total file matched count is better than the existing method. In future, the proposed model of linear search algorithm has used to retrieve image files, audio files and video files. Similarly, the proposed wildcard search is an advance search technique which has used to search the file name as the best result from the library databases

REFERENCES

1. *Ferguson N et al, "A simple algebraic representation of Rinjindal," in Selected Areas in Cryptography. Springer, 2001, pp.103-111.*
2. *Hongwei Li et al, "Multi keyword search supporting classified sub dictionaries over encrypted cloud data," DOI 10.1109/TDSC.2015.2406704, IEEE Transactions on Dependable and Secure Computing.*
3. *Jiadi Yu et al, "Multi keyword Top-K retrieval over encrypted cloud data," IEEE Transactions on Dependable and Secure Computing, vol. 10. No. pp. 239-250, 2013.*
4. *LI D Liu, Y.Dai et al, "Multi keyword ranked query over encrypted cloud data," Future Generation Computer Systems. Vol. 30. Pp. 179-190, 2014.*
5. *Liu H. Li. D et al, "Multi keyword search over encrypted cloud data through blind storage," IEEE Transactions on Emerging Topics in Computing, 2014, DO110.1109/TETC.2014.2371239.*
6. *Ren W. K., C.Wang et al, "Security challenge for the encrypted public cloud," IEEE Internet Computing, vol. 16, no. 1, pp, 69-73, 2012.*
7. *Wenhai Sun et al, "Multi keyword text search in the cloud for similarity based ranking," IEEE Transactions on Parallel and Distributed Systems, vol. DOI: 10.1109/TPDS.2013.282. 2013.*
8. *Wong W. K et al, "KNN computation on encrypted databases," in Proceedings of SIGMOD Internal Conference on Extending Database technology. Advances in database technology. ACM, 2008, pp. 287-298.*
9. *Yang Yang et al, "Flexible Wildcard Searchable Encryption System", in Proceeding of IEEE transaction on cloud computing DOI 10.1109/TSC.2017.27146.*
10. *W. K. Wong et al, "Multi keyword Supporting Gram Based Fuzzy Search", in conference on Information Communication and Embedded System (ICICES 2016).*