ISSN: 2321-7782 (Online)
e-ISJN: A4372-3114
Impact Factor: 7.327

Volume 6, Issue 1, January 2018

International Journal of Advance Research in Computer Science and Management Studies

Research Article / Survey Paper / Case Study

Available online at: www.ijarcsms.com

An Approach for Synonym Based Fuzzy Multi Keyword Ranked Search over Encrypted Cloud Data

Syed Shabbeer Ahmad
Dept.of CSE, MJ College of Engineering & Technology
Hyderabad – India

Abstract: In the proposed system, secure and efficient multi-keyword ranked search scheme over encrypted data.

Construct an index based on vector space model and (term frequency)TF X IDF (inverse document frequency) .

Construct a special index structure and propose a "Greedy Depth first Search" Algorithm to provide efficient multi-keyword ranked search. The secure kNN method by using Euclidean distance is utilized to ensure accurate relevance score calculation between encrypted index and query vectors. Fuzzy search method is used to search for relevant records, the system also tries to find those records that include words similar to the keywords in the query, even if they do not match exactly.

Keywords: component; formatting; style; styling; insert.

I. INTRODUCTION

Cloud computing has been considered as a new model of enterprise IT infrastructure, which can organize huge resource of computing, storage and applications, and enable users to enjoy ubiquitous, convenient and on-demand network access to a shared pool of configurable computing resources with great efficiency and minimal economic overhead. With the prevalence of cloud services, more and more sensitive information are being centralized into the cloud servers, such as emails, personal health records private videos and photos, company finance data, government documents, etc.

In this paper, we present a secure and efficient multi-keyword ranked search scheme over encrypted data, which additionally supports dynamic update operations like deletion and insertion of documents. Specifically, we construct an index based on vector space model to provide multi-keyword search, which meanwhile supports flexible update operations. Besides, cosine similarity measure is utilized to support accurate ranking for search result. To improve search efficiency, we further propose a search algorithm based on "Greedy Depth-first Traverse Strategy". Moreover, to protect the search privacy, we propose a secure scheme to meet various privacy requirements in the known cipher text threat model.

II. EXISTING SYSTEM

Wherever Times is specified, Times Roman or Times New Roman may be used. If neither is available on your word processor, please use the font closest in appearance to Times. Avoid using bit-mapped fonts if possible. True-Type 1 or Open Type fonts are preferred. Please embed symbol fonts, as well, for math, etc.

Disadvantages of Existing system: The CSPs that keep the data for users may access users sensitive information without authorization. Without encrypt the data, users directly upload the files into the cloud means cannot applied encryption directly on data. Existing system methods are not practical due to their high computational overhead for both the cloud server and user.

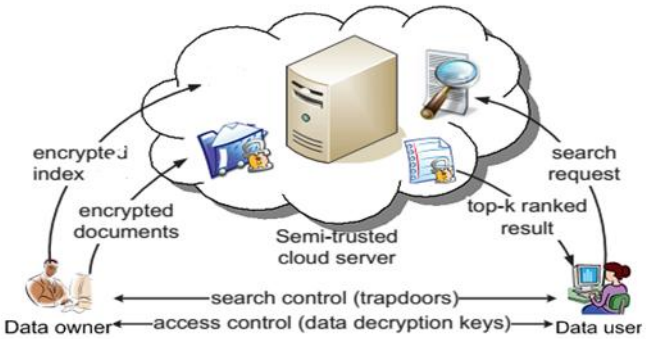
III. PROPOSED SYSTEM

The CSPs that keep the data for users may access users sensitive information without authorization. Without encrypt the data, users directly upload the files into the cloud means cannot applied encryption directly on data. Existing system methods are not practical due to their high computational overhead for both the cloud server and user. The proposed system is a secure based search scheme, which supports multi keyword ranked search and dynamic operation on the document collection. Construct a special index structure. Propose a GDFS algorithm to provide efficient multi-keyword ranked search. The secure kNN method by using Euclidean distance is utilized to ensure accurate relevance score calculation between encrypted index and query vectors. Fuzzy search method is used to search for relevant records, the system also tries to find those records that include words similar to the keywords in the query, even if they do not match exactly.

Advantages of Existing system:

Searchable encryption scheme supports both the accurate multi keyword ranked search and flexible dynamic operation on document collection. The proposed scheme can achieve higher search efficiency. Search can flexibly performed to further reduce the search process.

IV. SYSTEM ARCHITECTURE



V. MODULES

Data owner: Data owner has a collection of documents F=(f1, f2,....., fn).Data owner builds a secure searchable tree index I from F and generates an encrypted document collection C for F.

Data users: Users are authorized ones to access the documents of data owner. With t query keywords and generate trapdoor TD to fetch k encrypted documents from server. Then data user can decrypt the documents with the shared secret key.

Cloud server: The cloud server stores the C and tree index I for data owner upon receiving the TD from user, the cloud server Executes search over index. Returns the corresponding top- k ranked encrypted documents.

VI. ALGORITHMS

1) BUILD INDEX

- Create an ArrayList to store document unique words Return 0 if there is no common term.
- Calculate the IDF values of documents, to specify whether any common or rare terms appear across all documents.IDF value can be calculated by

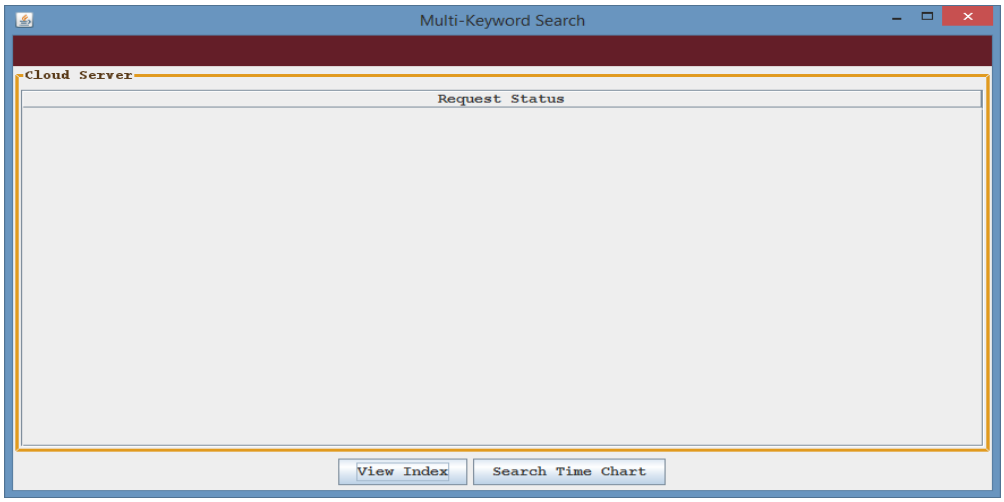
- $idf_{term} = \log_{10}(^{\text{total number of documents}}/_{\text{Number of documents the term appears in}})$
- Calculate TF values for all the documents in C i.e no of times the term occur in a document.
- Calculate TF×IDF and store in matrix
- For the the given query calculate the tf-idf vectors and store in array
- Calculate the length of document and query by
- Length (d,q)=sqrt(TF×IDF)^2
- Calculate the similarity score between two vectors i.e documents and query by inner product multiplication using eucledean distance
- Sscore=∑TF×IDF where ∑TF.IDF=D.Q/Lemgth(d,q)
- According to the similarity values the index store decreasing order of Sscore in HashSet

2) GDFS SEARCH

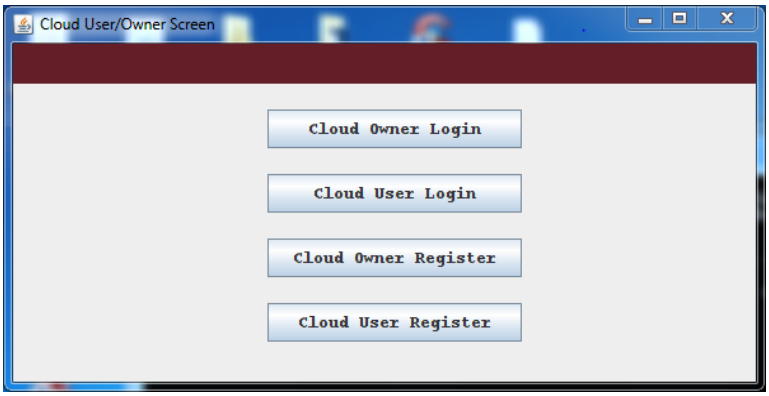
- User enter the keyword to be searched
- Data Owner generate the trapdoor file consists of filename and index
- if Sscore(D,Q)> 0 then return file.
- Cloud server check into index if the keyword present into the documents then return the documents with similarity score by using HashMap
- For authorized user provide the decrypting keys.
- Download the document
- Else return no documents found

VII. IMPLEMENTATION

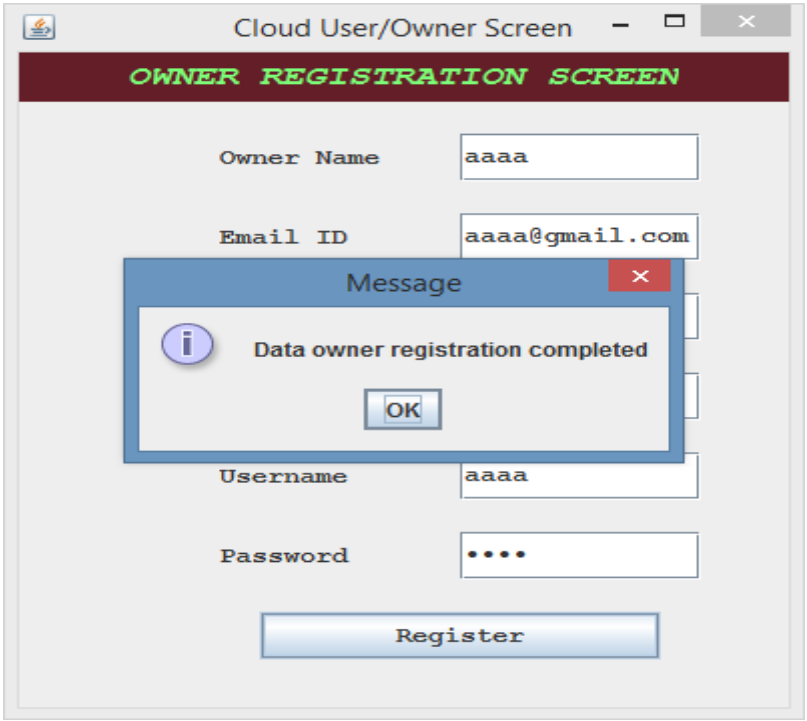
1. Server Gui



2. CLIENT GUI:WHERE OWNER AND USER REGISTER :



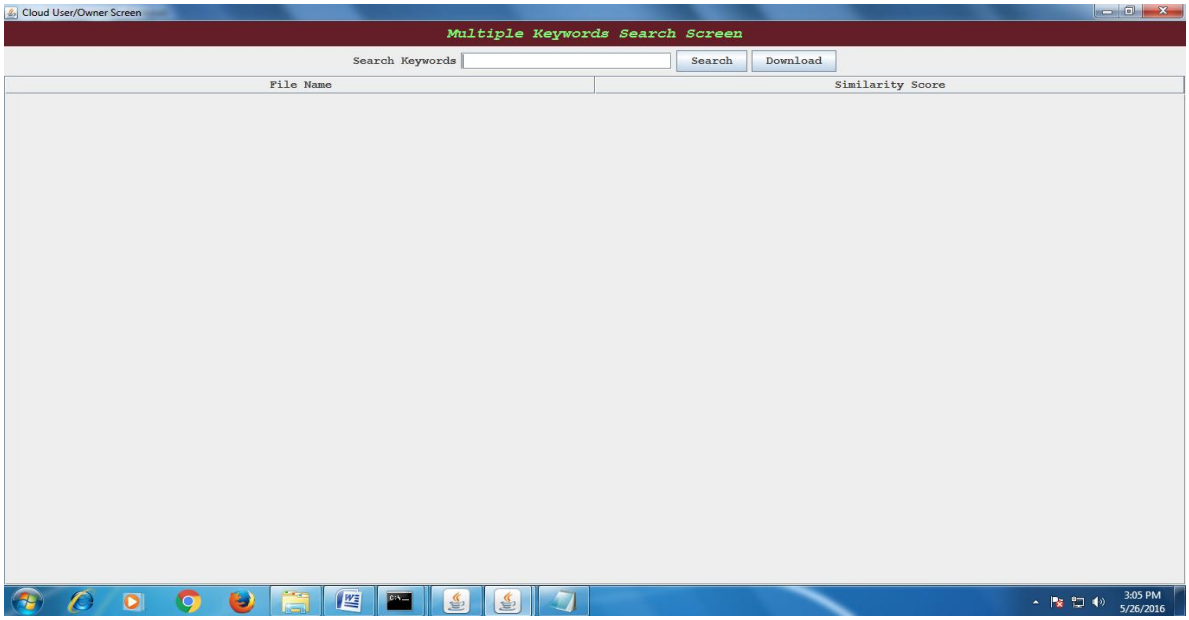
3. OWNER REGISTRATION :



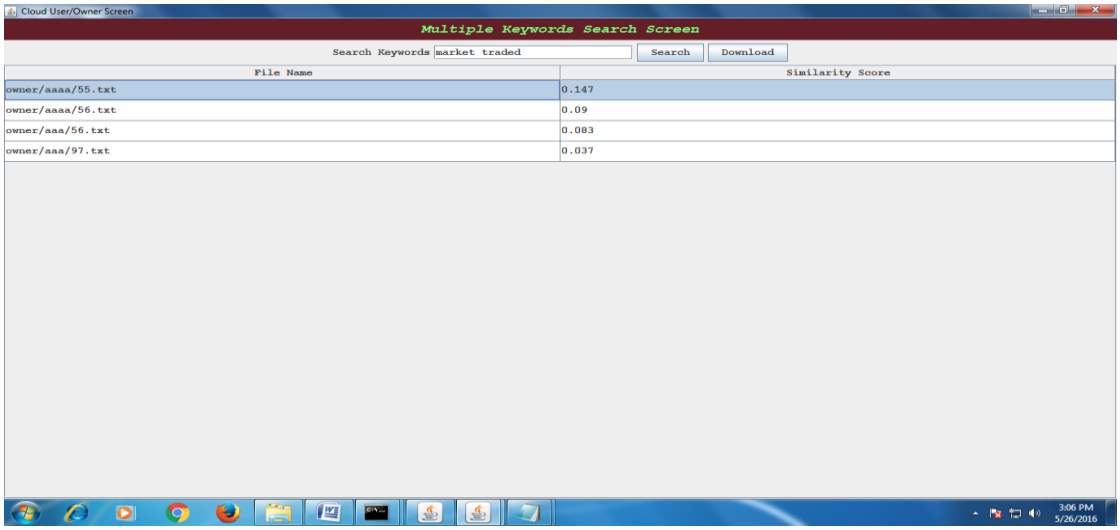
4. AFTER REGISTRATION OF OWNER, OWNER LOGIN AND OUTSOURCE THE FILES.

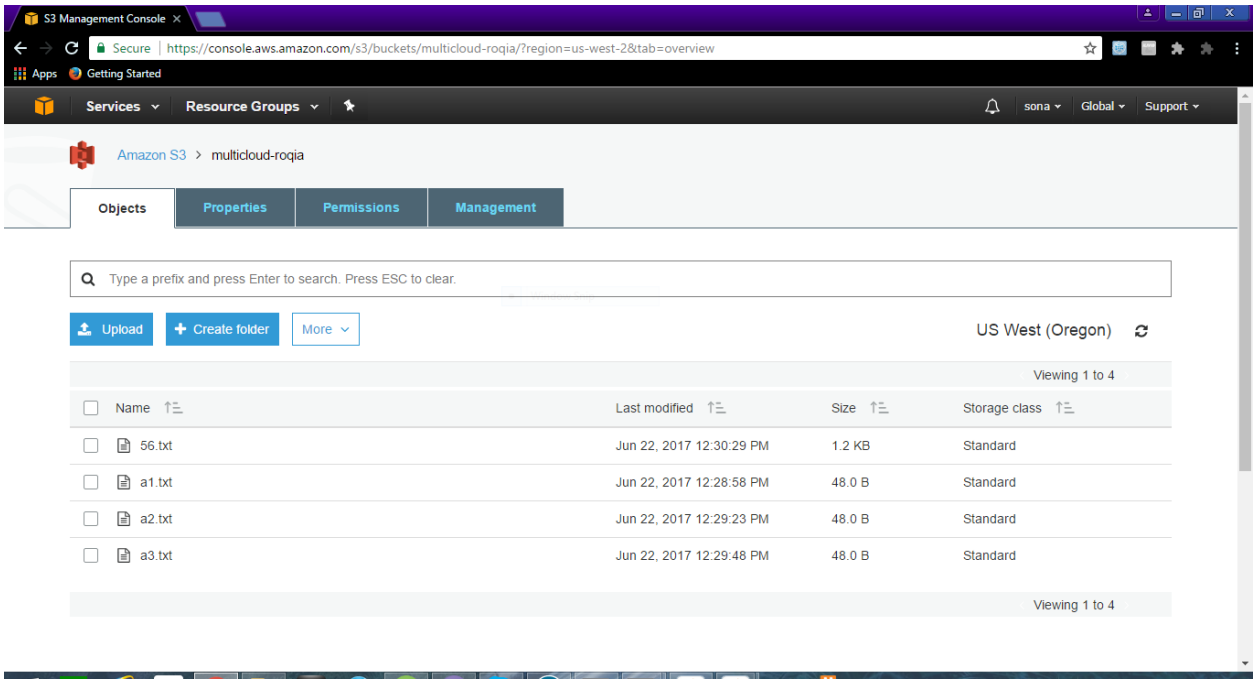
5. FILE OUTSOURCING

6. USER SEARCHING FILES: CLIENT GUI



7. SEARCHED RESULTS

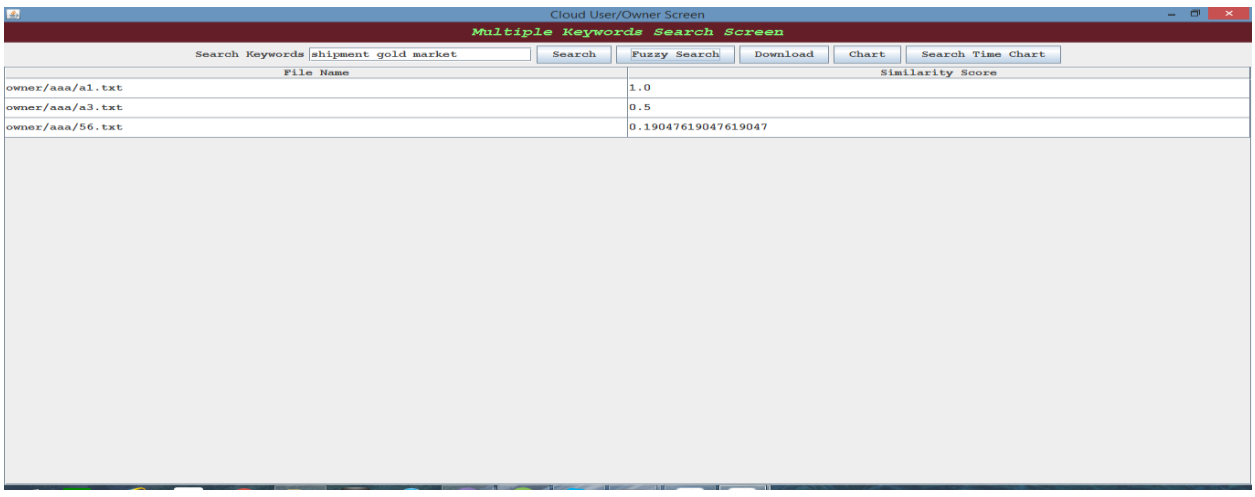




I. RESULT ANALYSIS

1. In existing system the search operation give the result only if exact keywords match.To solve this problem we are proposing fuzzy search.
2. Fuzzy search means when searching for relevant records, the system also tries to find those records that include words similar to the keywords in the query, even if they do not match exactly.
3. If the user's searching input exactly matches the pre-defined set of keywords, the server will return the files containing the keyword;
4. If there exist some error in spelling or some format inconsistencies in the searching input, the server will return the closest possible results based on pre-specified similarity semantics.
5. We used an advanced technique(i.e.,wild card-based technique) to construct the storage efficient fuzzy keyword sets by using edit distance technique. With the help of symbol base trisearch scheme we enhance searching efficiency.ed to further reduce the search process.

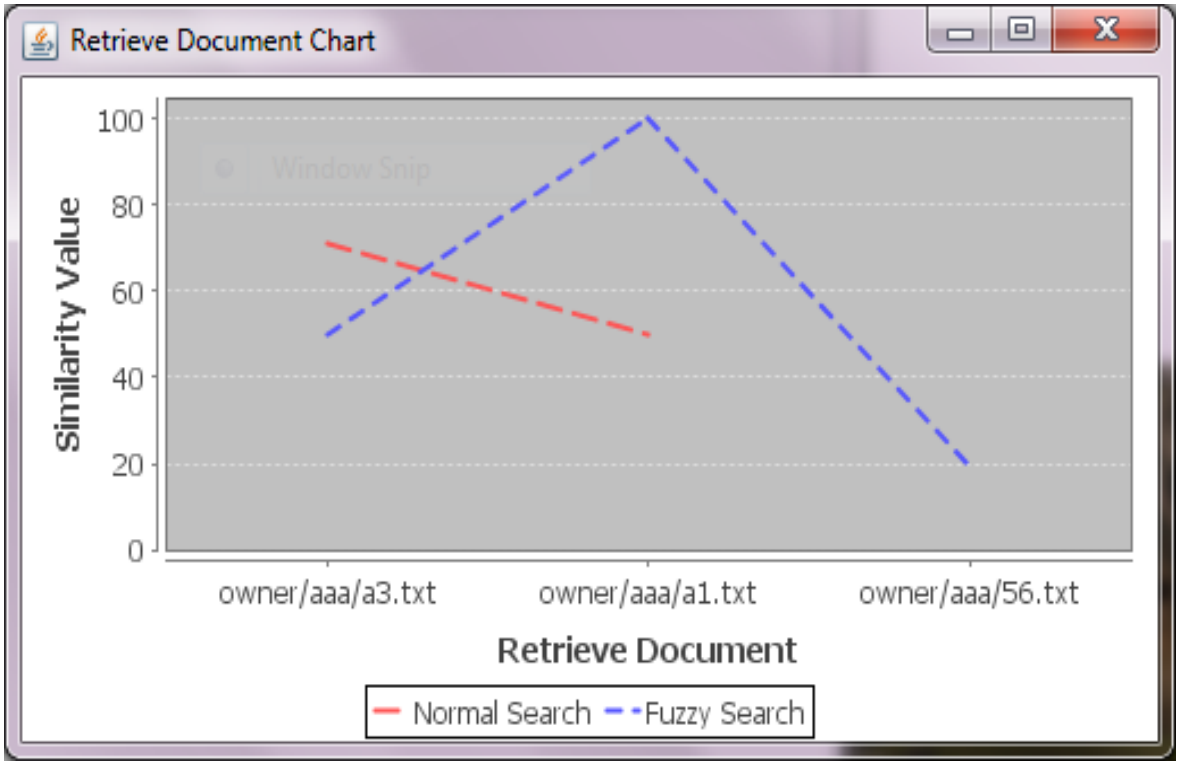
Fuzzy keyword Searched Results



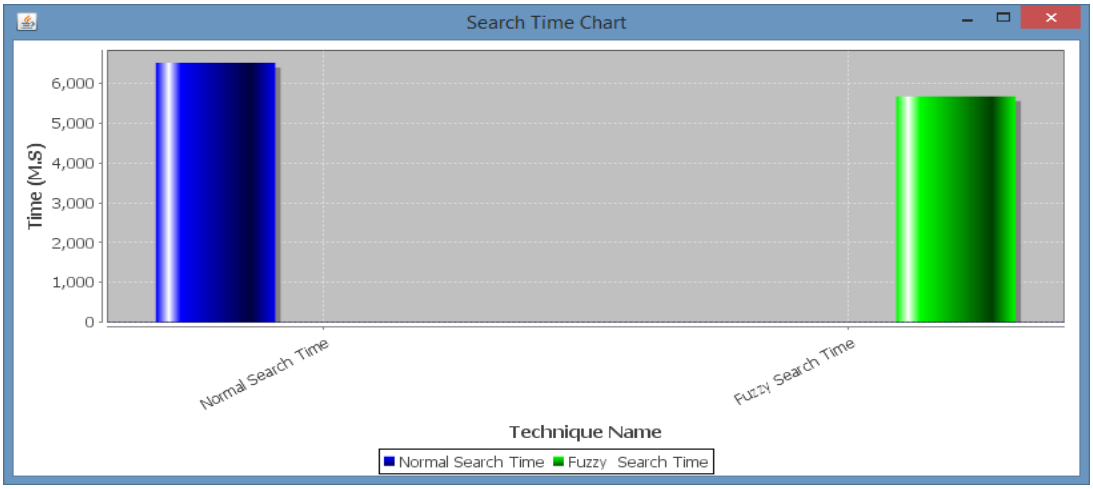
The table below gives the file name and similarity score for normal search and fuzzy search.

No. of documents	Normal search Sscore	Similarity(%)	Fuzzy search Sscore	Similarity(%)
a1.txt	0.79	79	0.1	100
a2txt	0	0	0	0
a3.txt	0.74	74	0.50	50
57.txt	0	0	0.19	19

Fuzzy keyword Searched Results shows the similarity value comparison of normal and fuzzy search.



Fuzzy keyword Searched Results give the result of the no. of documents and normal search time and fuzzy search time. The search time varies as the more no. of documents uploaded by the owner.



II. Conclusion and Future Scope

1. A secure, efficient and dynamic search scheme is proposed, which supports not only the accurate multi-keyword ranked search but also the dynamic deletion and insertion of documents. We construct a special keyword balanced index, and propose a "Greedy Depth-first Search" algorithm to obtain better efficiency than linear search. In addition, the parallel search process can be carried out to further reduce the time cost.

2. Fuzzy Search is proposed to searching for relevant records, the system also tries to find those records that include words similar to the keywords in the query, even if they do not match exactly.

3. In the proposed scheme, the data owner is responsible for generating updating information and sending them to cloud server. For further research Future work is to design a dynamic searchable encryption scheme whose updating operation can be completed by cloud server only.

4. All users usually keep the same secure key for trapdoor generation in SE schemes.

References

1. R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky, "Searchable symmetric encryption: Improved definitions and efficient con- structions," in Proc. 13th ACM Conf. Comput. Commun. Secur., 2006, pp. 79–88.
2. J. Li, Q. Wang, C. Wang, N. Cao, K. Ren, and W. Lou, "Fuzzy key- word search over encrypted data in cloud computing," in IEEE Proc. INFOCOM, 2010, pp. 1–5.
3. M. Kuzu, M. S. Islam, and M. Kantarcioglu, "Efficient similarity search over encrypted data," in Proc. IEEE 28th Int. Conf. Data Eng., 2012, pp. 1156–1167.
4. C. Wang, N. Cao, K. Ren, and W. Lou, "Enabling secure and effi- cient ranked keyword search over outsourced cloud data," IEEE Trans. Parallel Distrib. Syst., vol. 23, no. 8, pp. 1467–1479, Aug.2012
5. C. Orencik, M. Kantarcioglu, and E. Savas, "A practical and secure multi-keyword search method over encrypted cloud data," in Proc. IEEE 6th Int. Conf. Cloud Comput., 2013, pp. 390–397.
6. Ning Cao, Cong Wang, Ming Li, Kui Ren, and Wenjing Lou, **"**Privacy-Preserving Multi-Keyword Ranked Search over Encrypted Cloud Data" IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, VOL. 25, NO. 1, JANUARY 2014.

