

Efficient Information Analysis and Retrieval using Index based EMTSR in E-governance application

Rajan Manro*, Tripat Deep Singh Dua** and A.S. Joshi***

ABSTRACT

The huge amount of information is being accessed to the dynamic operation across the globe as per the requirement of the user in the application of E-governance. The secure information management is complex due to concerns of security while the process of accessing the data from cloud environment. To overcome the issues, the proposed analysis and methodology system is implemented for easy and accurate information access from cloud storage. By secure process of encrypted data in cloud the issues of security are solved, but the retrieval of precise information is difficult. To overcome the issues of secure process and need of information retrieval an Index based Enhanced Multi-keyword Top-k Search and Retrieval (IEMTSR) mechanism has been proposed. Index keyword search service is proposed based on the generating inverted index keywords. By proposing a technique of relevance scoring the search query of the document in cloud storage is evaluated and ranked the documents based on the ranking procedure of the proposed algorithm. By ranking the highest or most relevant data sets will available to users in an efficient manner. The proposed work provides efficient retrieval of data and saves memory to a great extent by adjusting the size of the table according to its capacity. It achieved high speed retrieval of data as well as saving the storage cost of E-governance cloud environment with better performances than the existing in search quality, retrieval of data with high speed and improvement in accuracy.

Keywords: Secure access; Ranking; Index based Search; Searching mechanism; Query Processing;

1. INTRODUCTION

Normally the information delivery and services to the public through internet is mean to be E-government. It is an Information Communication Technologies (ICT) with the development model at sequential time by the scientific institutions, researchers and international organizations. As per the requirement the development model has improved and provides services to the public to access the information as per the growing requirements. The environment of cloud computing appears in this application for the storage and access of government information from anywhere at time via internet. So the resource computing will transfer the services as per the product. The services are Software as a Service (SaaS), Platform as a Service (PaaS) and Infrastructure as a Service (IaaS).

In e-government cloud environment, the solutions for the issues in information access will be useful. But in the existing environment model the state of transition of required access and redevelop is not taken into account. Some issues occur from a random process of transition like security and redundancy. It related to the transaction privacy and ensures security from any issues in the future. It exploits the benefits in ensuring actual resources, professionalism use, cost savings and application management. In this application, it provides high quality services for various resources. It provides ubiquitous network access, location independent resource pooling, usage-based pricing and rapid resource elasticity.

* Research Scholar Punjab Technical University Jalandhar, Punjab, India, *Email: rajanmanro@gmail.com*

** Asst. Prof. Department of Computer Applications Guru Nanak Institute of Management and Technology (GNIMT) Model Town Ludhiana, Punjab, India, *Email: tripatdua@yahoo.com*

*** Former Director Punjab Institute of Management & Technology Mandi Gobindgarh, Punjab, India, *Email: asjoshij@yahoo.com*

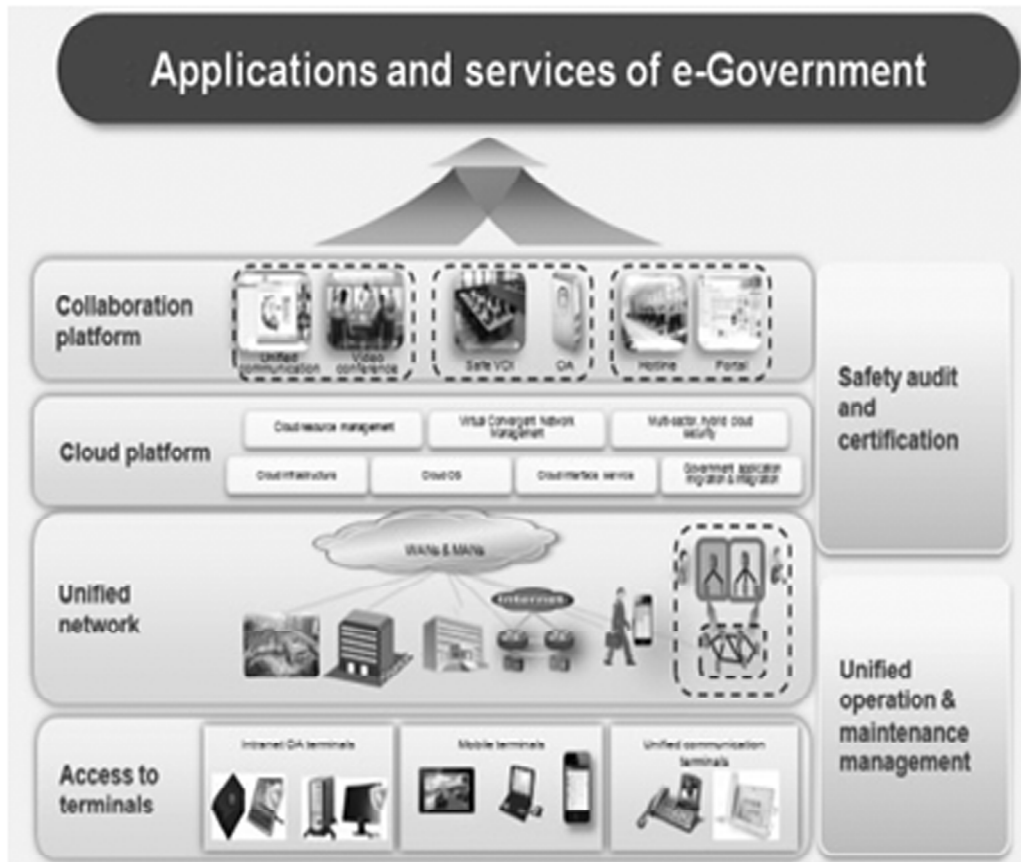


Figure 1: Architecture of E-Government Cloud Service Platform

This application offers the economic savings in outsourcing the data for both enterprises and individuals. The data security is difficult to outsource the confidential and sensitive data in a cloud environment. To ensure secure process various encrypted techniques are developed and implemented. It avoids the unnecessary access and prevent from attacks. Due to encrypting data the access is becoming a challenging task. It converts the plain text data into chipper text.

The retrieval of information from the cloud server is considered to retrieve the relevant data instead of retrieving all chipper data; in order to improve the usability and to save the expensive of the environment usage. The retrieval process is done by ranking the data and the data matching to list the ranked data. From that the appropriate data is retrieved as per the user requirements. In this paper Index based Enhanced Multi-keyword Top-k Search and Retrieval (IEMTSR) mechanism is proposed to improve and solve the issues of the system. This method is used to retrieve and search the highest ranked by the function of similarity scoring method. It reduces the searching time and processing time with cost reduction.

The rest of the paper is organized into section wise; in section II the literature review of the related method of searching mechanism is discussed. In section III the proposed method and its functions are explained with its implementations. In section IV the performance analysis of the proposed methods with the existing methods is presented. Finally, in section V overall proposed system concludes with future work.

2. RELATED WORK

In this section, the review related to the proposed searching mechanism is discussed with various methods. Normally the service of E-government is flexible in on-demand high quality utility with less cost. But when ensuring the security and access control need to be improved for an efficient process. In searching

mechanism the function of retrieving and finding is based on scoring and ranking of the keywords and as well as for preserving privacy. It ensures the semantic find through keyword extension. By TFIDF algorithm the searching process is extending the keyword search with ultimate support on semantic query. It makes for a flexible and fast process of searching mechanism [1].

In many enterprise storage services provides to users with less cost, scalable and trust access from anywhere at any time. The provider provides trust based on process between the cloud users. Encrypting of data is establishing the search to provide security. It explores the techniques based on CRSA and B-Tree in order to enhance the trust level [2].

By ranking score the multiple keywords search is implemented to have data retrieval from cloud storage. In this paper they propose a ranking method to avoid the undifferentiated results. The statistical measurement approach builds the index structure function to protect sensitive information. Based on the data score the keyword search is performed without loss of privacy and leakage of information. Through the analysis encryption scheme is implemented by using the OPM function for multiple keyword search [3].

The Enhanced Multi keyword Top-k Search and Retrieval (EMTR) scheme is proposed in this paper to have high efficiency. In this scheme based on inverted indexing the document efficiency is estimated. By using ranking and scoring method the data is indexed in multi keyword search [4].

In E-Government application, the addressing of information flow is re-observed somewhat inadequate by the classical MAC not by the classical DAC [14]. In constructing the resource sharing is the main target in this paper with the integration and sharing of information. Integrated tactics are proposed based on the planning catalog system [12]. Also, the management mode is it reform in this application. The impact of services provided to the users or the public is pointed out and developed for efficient process and access of government information which is directly reaching the people [13].

Searchable symmetric encryption (SSE) is proposed for encrypting the data for secure retrieve process. It addresses the issues of data privacy and proposed a log generating module based on the two-round searchable encryption (TRSE) scheme. It provides secure and efficient process with the leakage elimination [5].

In order to have an efficient security process without leakage trust management domain is implemented. The retrieval of cipher data is based on the ranking making, privacy and secure keyword search. The multi-keyword query scheme is proposed to address query results during the expansion of keyword dictionary. Based on matching the result set is available from the database [6].

The secure data access is utilized by the proposed secure ranked keyword search by matching the set of ordered data based on ranking. It enhances the matching criteria and deployed the data hosting services with privacy-preserving without losing privacy of keywords [7].

The privacy-preserving based multi-keyword text search (MTS) scheme is implemented to have a similarity based addressing the issues. It builds the index based similarity measure in terms of vector space and frequency model. The tree-based index structure is performed by the multi-dimensional (MD) algorithm in a linear manner with an efficient process. Secure index schemes are performed with the requirement of privacy [8]. For economic savings and flexibility, multi keyword search based on ranking is proposed to improve the information retrieval of encrypted cloud data [9]. Unique storage access is addressed to the path or location for quick access of XML files. The storage and retrieval of XML files indexed the data for utilization of query search [10].

In AHP application, the decision policy for evaluating information is performed based on the information security and its components. The process of decision making is done by analysis the methodology. As per the security and management the e-government information systems provides an availability of secure information to the users or public [11].

3. PROPOSED WORK

In this section, the proposed methods explanation of searching methodology and analysis is presented with its implementation. In order to provide an efficient methodology and analysis of information retrieval as per the requirement of user a new secure searching mechanism is proposed. In that after the data are stored in the storage server with the indexing and ranking. The information searching is performed as per the data ranking and scoring. For ensuring data security the encrypted data is decrypted with the data is retrieved in the receiving side by using improved RSA double encryption and decryption technique – IRDEDT.

The basic function of dynamic operation is performed by indexing the keywords by using indexing method, as well as ranked the data according to the score and index search. The search methodology is performed with the similarity score, similarity ranking and index based keyword search of top-k multi keyword. In this proposed algorithm the cipher data is decrypted by using the IRDERT. According to the block diagram the proposed system is performed and accessed in E-Government application. An efficient data analysis for search service and information retrieval is achieved by the proposed system.

The main aim is to retrieve the data in a secure manner without any privacy leakage. The proposed searching methodology is based on Index search with Fibonacci hash function in order to have an accurate search of keywords from the server. As well as based on the index hash value the data are stored in the cloud server and it is retrieved with the similarity score and ranking of the data value. So, the user can access the information as per their requirements. The hash value function is computed by using Multiplication Method.

In everyday life accessing of data service has been increased, so the searching mechanism efficiency needs to be improved for quick access by using indexing and ranking. In security domain the process of search engine is defined in a secure manner with accuracy and fast process. The rate of document will help to access the data in speed process. In this paper, an index based enhanced multi-keyword top-k search and retrieval scheme is proposed for friendly access of data query processing with access rate of data. Fig [2] shows the implementation of the proposed system in E-government cloud server.

In a document several keywords are consists and indexed for outsourced. The documents are encrypted by the IRDEDT in order to provide security in cloud server. The retrieval of data is based on the query process on multiple keyword searches with its weight. If the user wants to access any information from e-government server the search process is based on the keywords.

As per the data query search in server, the query process is evaluated by calculating the similarity rank and score of the most accessed keywords (Top-K documents). According to the results the data is accessed by the user after the decryption of content to view the original data. After selecting the wanted data the decoding process is done by using IRDEDT. By evaluating the keywords weight, encode and decode, ranking, scoring and index based mechanism are performed in the proposed system for efficient access. The index keywords are identified in the server by the vector space model.

The implementation process of the query search is done with the inverted index and evaluation of weight for each keyword for search. The basic scheme to evaluate weight is TF-IDF weighting scheme which is used to find the weight and equalize the document vector length. The accuracy has been improved by the standardization of query.

$$Tf-Idf_{k,d} = (1 + \log Tf_{k,d} * \log N/Idf_{k,d}) \quad (1)$$

$$Tf_{k,d} = 0.5 + \frac{0.5 * f(k,d)}{\max\{f(k,d) : k \in d\}} \quad (2)$$

$$Idf_{k,d} = \log \frac{N}{|\{d \in D : k \in d\}|} \quad (3)$$

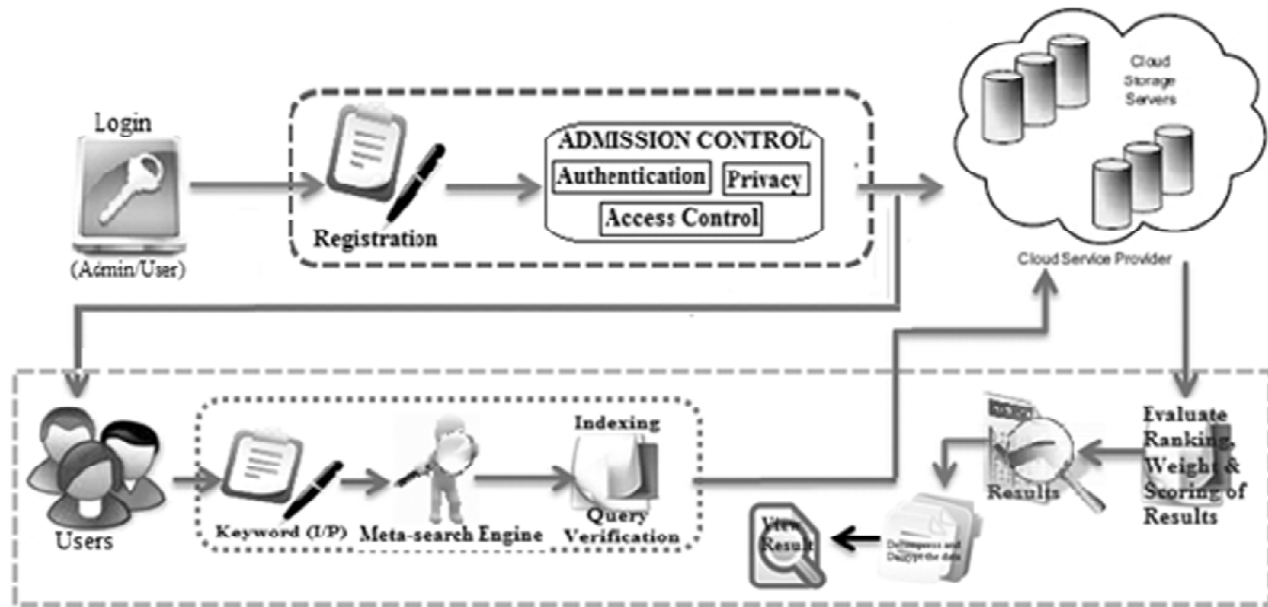


Figure 2: Block Diagram of Proposed System

$$DS(d) = \frac{1}{\sqrt{T f_{k,d}}} \quad (4)$$

$$QS(q) = \frac{1}{\sqrt{\sum (T f_{k,d} * Idf_{k,d})^2}} \quad (5)$$

Tf-IDF (Term frequency-Inverse document frequency) weighting scheme is used to measure the weight of the keywords which is often used for mining text and information retrieval. Based on the frequency of word the keywords are found and indexed. According to frequency of document rank and score similarity the search process is performed in the proposed system. By sophisticated ranking functions the model variants are functioned with the classification and summarization of the filtering the keywords.

In Term frequency (Tf), the appearances of repeated words in the documents are measured and divided with the total word's appearances. In Inverse document frequency (Idf), the amount of logarithm documents divided by the number of specific term appearances in the documents. After selecting the required documents from the server, the document needs to be decrypted for original document access. For decoding the document IRDEDT is used to compute the original document. The document is retrieved by decoding with the same key which is used for encryption and it support 256 bit or 128 bit key for decryption.

$$D = (((c^\lambda \bmod m^2 - 1)/m) * \mu \bmod m)^d \bmod n \quad (6)$$

In the proposed system, the similarity scoring is measured by using enhanced similarity scoring method (ESSM) for efficient ranking search process. As per the requirements of user query the documents are evaluated by measuring similarity score between query keywords and documents. As per the procedure given below the process of measuring the similarity score is performed. The document and query standard is included with the proposed indexed scheme for evaluating score.

ESSM - Score (q)

1. Float score[N] = 0; Float length [N];
2. For each query keyword K
3. Evaluate Tf-Idf_{k,d} by using equation [1] and fetch postings list for K

4. For each pair($d, Tf_{k,d}$) in postings list
5. do score [d] += $Tf-Idf_{k,d} \times QS(q)$; Using equation [5]
6. Read the array length
7. For each d
8. do score[d] = scores [d]/Length [d];
9. $score[q, d] = \sum_{k \in q} TF - Idf_{k,d}$;
10. return Top k components of score.

Algorithm-ESSM using multi keyword search

Procedure Begin

for every keyword $k \in Q$

if (keyword constraint satisfies with k)

 perform keyword processing;

end

for each keyword $k \in Q$

 Calculate query Standardization ($QS(q)$), keyword frequency ($Tf_{k,d}$), inverse document frequency ($Idf_{k,d}$), weight for each keyword ($Tf-Idf_{k,d}$) by using equation [1 to 5]

end

for each index keyword $I_k \in I$ and query keyword $Q_k \in Q$

if(index keyword matches query keyword)

 display list($DS(d), Tf-Idf_{k,d}$);

end

end

for each matched document

 Calculate co-equalization ($Ce = M_k / Q_k$);

 Calculate the ESSM score measure for each

 document and query by ESSM - Score (q);

 display score list;

end

end procedure

In proposing approach various parameters are included like co-equalization factor, term frequency, document length Standardization, query standardization and inverse document frequency. It improves to search most relevant data with quality which means it provide with a high score. Based on the strong matching and similarity of query and document the score is defined for the data. The document score is increased by the more query keywords and it measured based on the searching algorithm.

The proposed scheme provides the high score relevant document based on the top-k selection method from cloud server. It performs with a partial listing of retrieved documents in a linear time from the database.

Algorithm-Proposed Top-k Selection Scheme

Procedure Begin

```

Read ESSM - Score list
if(EMSSM – Score-List! = empty)
    perform Partial sorting in linear time;
    generate ranked list;
end
if(size[ESSM – Score-List]==k)
    for i {1,...M}
        display top-k list
    end else
if(size[ESSM – Score-List]>k)
    initialize s;
    for(s=0; s<=k; s++)
        display top-k list;
    end
end procedure

```

4. PERFORMANCES ANALYSIS

In this section, the proposed scheme performances analysis is evaluated and illustrated the results. The data searching is performed in a real world data set. For analysis of performances 500 documents are used by the implementation of the proposed algorithm. From the set of keywords the stop words are reduced by the indexed inverted method. The results are evaluated based on the Top-k searching accuracy of retrieval, time for retrieval and similarity measure of ranking and scoring with time for execution.

Fig [3] shows the comparison of various algorithm similarity scores of the document. The analysis of Top-k selection Vs Precision for various algorithms is shown in Fig [4]. Fig [5] shows the measurement of

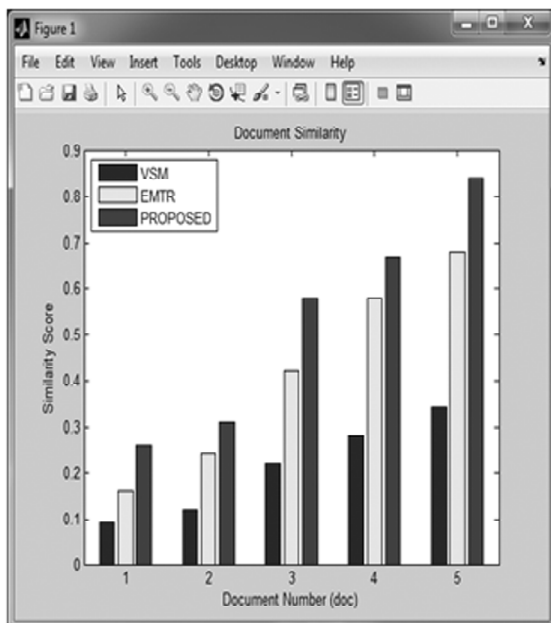


Figure 3: Comparison of Document Similarity

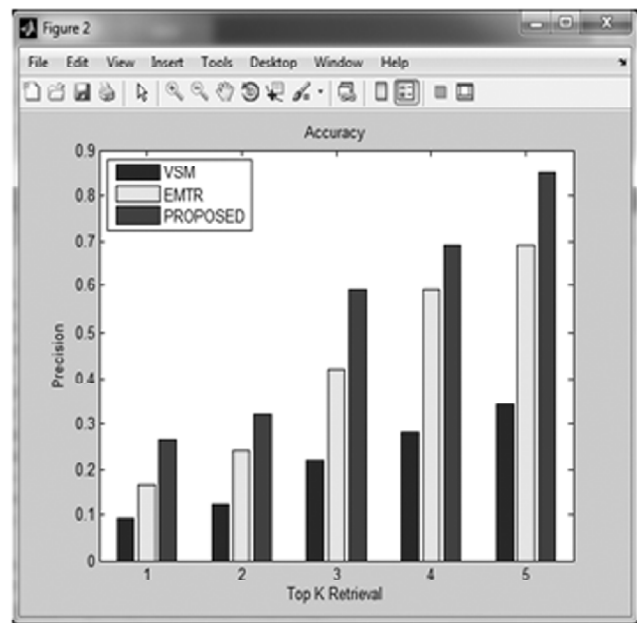


Figure 4: Analysis of Proposed approach Accuracy

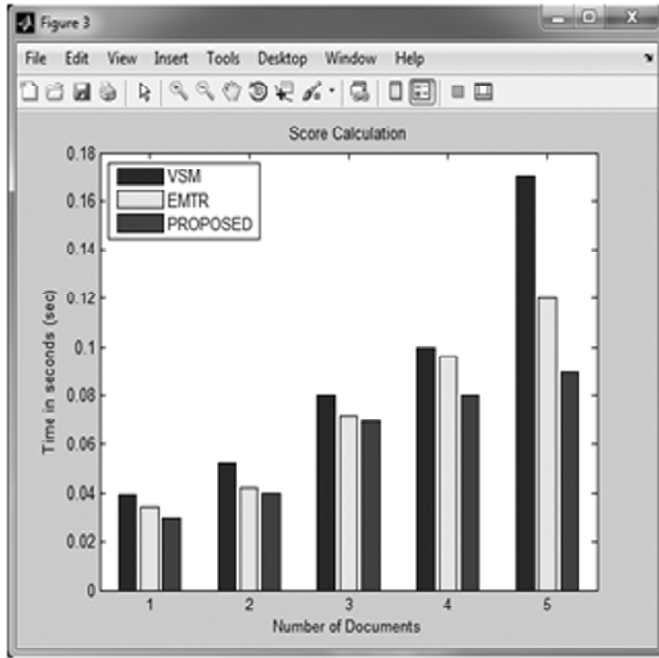


Figure 5: Analysis of Proposed approach score calculation

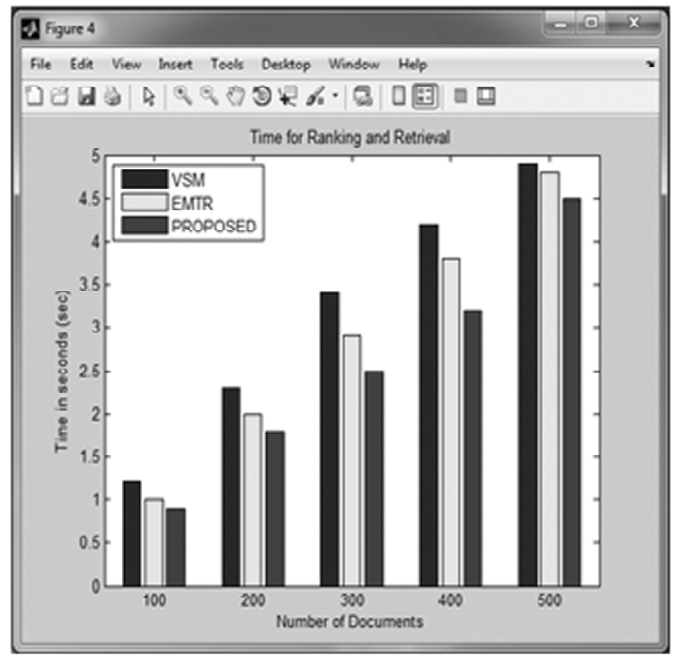


Figure 6: Analysis of Time requirement for Ranking and Retrieval

Table 1
Comparison of Keyword Search Algorithm

Number of documents	Single keyword search		Multi keyword search		IEMTSR	
	Retrieved documents	Execution time in Seconds	Retrieved documents	Execution time in Seconds	Retrieved documents	Execution time in Seconds
50	40	4	20	2	16	1.25
100	93	6	35	3.5	27	2.78
150	142	7	41	5.2	33	4.22
200	180	10	48	6	39	5.20
250	233	12	55	7.2	46	6.48
300	277	14	62	9.1	50	8.30

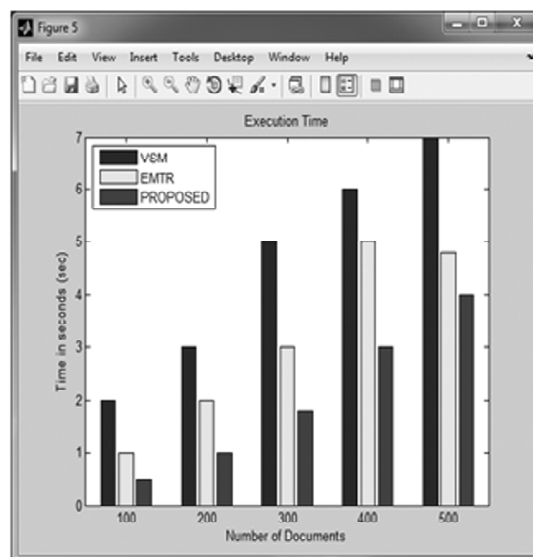


Figure 7: Analysis of Execution time

similarity score for various algorithms and Fig [6] shows the analysis and comparison of time required for ranking and retrieval.

Table [1] shows the illustrated results of retrieval and execution time for various algorithms; also, shows the better performances of proposed work. Fig [7] shows the analysis of execution time for various algorithms.

5. CONCLUSION AND FUTURE WORK

In this paper, an Index based enhanced multi keyword Top-K search and retrieval (IEMTSR) is proposed for the analysis and methodology of searching mechanism for quick access of from the server. The measurement of score for query relevant document is done by using ESSM from the outsourced data from server. The proposed approach provides improvement in searching and retrieval process with efficient and less time. Through the analysis of the approach scheme secure data access and efficient process are shown with high speed. In future, this approach may extend with elimination of duplicate and integration of retrieval of documents in an efficient way by proposing a new machine learning approach.

The contribution in this paper is the improvement of information retrieval in a secure manner and quick access without any burden from anywhere at any time in cloud environment.

REFERENCES

- [1] Dipika Chavan, Dinesh Yadav, "Achieving Efficiency of Encrypted Cloud Data with Synonym Based Search and Multi-Keyword Ranked Search", *International Journal of Science and Research (IJSR)*, Volume 4 Issue 1, January 2015.
- [2] Prasanna B. T., C. B. Akki, "Dynamic Multi-Keyword Ranked Searchable Security Algorithm Using CRSA and B-Tree", (*IJCSIT*) *International Journal of Computer Science and Information Technologies*, Vol. 6 (1), 2015, 826-832.
- [3] Amol D. Sawant and Prof. M.D. Ingle, "Indexing and Advanced Relevance Ranking Score Preserving for Multi-Keyword Search over Encrypted Cloud Data", (*IJCSIT*) *International Journal of Computer Science and Information Technologies*, Vol. 5 (3), 2014, 3165–3169.
- [4] Geethalakshmi, S.; Umamaheswari, S., "An efficient technique for Multi keyword based Search and Retrieval of cloud data," *Recent Trends in Information Technology (ICRTIT)*, 2014 International Conference on , vol., no., pp.1, 6, 10-12 April 2014, doi: 10.1109/ ICRTIT.2014.6996147
- [5] P. Shanmuga Priya, Preethi. D, Priya. J, shanthini. B, "Retrieval of Encrypted Data Using Multi Keyword Top–K Algorithm", *International Journal of Scientific and Research Publications*, Volume 4, Issue 4, April 2014, ISSN 2250-3153.
- [6] Ruixuan Li, Zhiyong Xu, Wanshang Kang, Kin Choong Yow, Cheng-Zhong Xu, "Efficient multi-keyword ranked query over encrypted data in cloud computing", *Future Generation Computer Systems* 30 (2014) 179–190.
- [7] R. Aravind, "An Efficient Search in Cloud Computing Using Ranked Keyword Search Algorithm", *International Journal of Science and Research (IJSR)*, Volume 3 Issue 5, May 2014.
- [8] Wenhai Sun, Bing Wang, Ning Cao, "Privacy-preserving Multi-keyword Text Search in the Cloud Supporting Similarity-based Ranking", *ASIA CCS'13*, May 8–10, 2013.
- [9] Ankatha Samuyelu Raja, Vasanthi A, "Secured Multi-keyword Ranked Search over Encrypted Cloud Data", *International Journal of Advanced Research in Computer Science and Software Engineering*, Volume 2, Issue 10, October 2012.
- [10] Anagha Vaidya and Dr. Arpita Gopal, "XQUERY : Efficient searching and indexing Algorithm Based on Multi-tries", *International Journal of Database Management Systems (IJDMS)*, Vol. 3, No. 3, August 2011.
- [11] Gaigole, M.; Khere, N., "Notice of Violation of IEEE Publication Principles Application of AHP to support information security decision making in case of Indian e-government systems," *Electronics Computer Technology (ICECT)*, 2011 3rd International Conference on , vol. 3, no., pp.150,153, 8-10 April 2011doi: 10.1109/ ICECTECH. 2011. 5941820.
- [12] Tai-tian, Mao; Chun-lin, Deng; Han-ye, Xia, "Research of planning and integration about E-government information resource based on catalog system," *E-Business and E-Government (ICEE)*, 2011 International Conference on , vol., no., pp.1,4, 6-8 May 2011, doi: 10.1109/ICEBEG.2011.5882634
- [13] Zhu-lan Bai; Bing Cai, "Effect of E-Government on the Efficiency of Information Service," *E-Business and Information System Security*, 2009. *EBISS '09*. International Conference on , vol., no., pp.1,4, 23-24 May 2009 doi: 10.1109/ EBISS.2009.5137970.
- [14] Lin, T.Y., "Information flows in e-government," *Computer Software and Applications Conference*, 2003. 27th Annual International, vol., no., pp.616, 621,3-6,Nov.2003,doi: 10.1109/ CMPSAC. 2003. 1245405