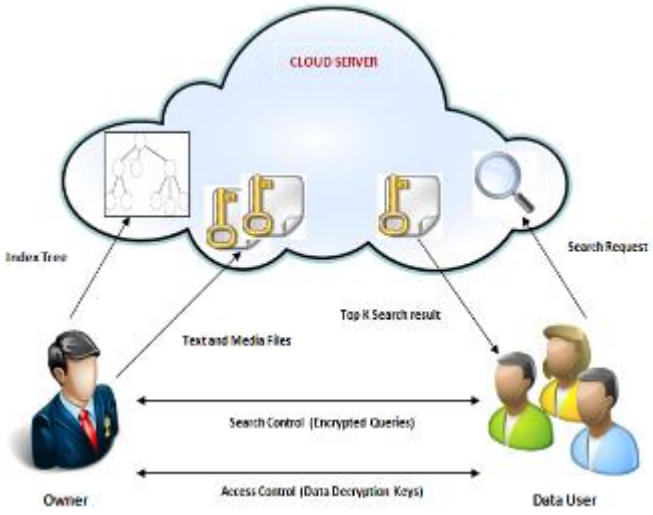

Figure 1: System Diagram

B. Algorithm Used

1) **Stemmer:** used in linguistic morphology and information retrieval to describe the process for reducing inflected (or sometimes derived) words to their word stem, base or root form e.g. word running is converted into run. This algorithm is used for index generation. **Input :** When user upload particular document then its inflected words are input **Steps :** Step 1 : Get rid of plurals and ed or ing suffixes Step 2 : Turns terminal y to i when there is another vowel in the stem Step 3 : Maps double suffixes to single ones ization , -ational etc. Step 4 : Deals with suffixes , -full , -ness etc. Step 5 : Takes off ant , -ence , etc. Step 6 : Removes a final e **Output :** Normalize word forms eg. Destructiveness =¿ destruct.
2) **Stop word:**
In this algorithm stopword like: punctuation marks, words like a,an,the are removed from the given text. **Input :** Stemmed words and sentences **Steps :** Step 1 : Get word list Step 2 : Calculate word count of frequently occurred words Step 3 : Create own stop word list Step 4 : Match these frequent words from stop word list Step 5 : Remove most frequent words **Output :** Filtration of words like the , and , a , to , of , was , it , in , that , he etc.
3) **Frequent word calculation:** Most frequent words above the threshold are calculated in this algorithm **Input :** Raw data with stemmed and without stop word **Steps :** Step 1 : Frequent words are buffered in hash table and word and its count is calculated Step 2 : Rank K is decide by the user and such top K results are shown to user from all calculated word counts **Output :** Dataset with word and its word count
4) **OCR:** For image tagging words are extracted from labeled image. **Input :** Labeled Image
5) **Steps :** Step 1 : Get labeled image Step 2 : Rough Preprocessing of image Step 3 : Search and recognition of the first character Step 4 : If feasible accuracy achieved
6) then goto step 5 else goto step 1 Step 5 : Position Evaluation of next character Step 6 : Again preprocessing of image Step 7 : Search and recognition of first character Step 8 : If feasible accuracy is not achieved goto step 5 else goto step 9 Step 9 : All characters are recognize and stop
Output : Text extracted by labeled images

5) **Keygen: Input :** Selection of key generation option **Steps :** Step 1 : Public key is generated Step 2 : Private key is generated **Output :** Public key for encryption and private keys for sharing is generated
6) **RSA:** RSA is used to encrypt and decrypt data stored on cloud **Input :** Raw text **Steps :** Step 1
: Choose two prime numbers, Prime1 and Prime2 to get the ProductOfPrime1Prime2 variable Step 2

: Find the Totient of ProductOfPrime1Prime2 (ProductOfPrime1Prime2) = (Prime1 -1) * (Prime2 1) Totient = (Prime1 -1) * (Prime2 1)

Step 3 : Get a list of possible integers that result in 1 mod Totient EncryptPrime * DecryptPrime = 1 mod Totient (Totient * AnyInteger) + 1 = 1 mod Totient Step 4 : Choose a 1 mod Totient value with exactly two prime factors: EncryptPrime and DecryptPrime
Step 5 : Actual Encryption CipherText = Plain-
^
TextEncryptPrime mod ProductOfPrime1Prime2 Step 6 : Actual Decryption
^
PlainText = CipherTextDecryptPrime mod ProductOfPrime1Prime2
Output : For encryption we get cipher text and for decryption we get plain text

7) **BST :** Binary search tree. To find appropriate docu-ment from I : searchable index tree
Input : Keyword for search **Steps :** Here k is the key that is searched for and x is the start node. BST-Search(x, k) Step 1 : y x Step 2: while y != nil do Step 3: if key[y] = k then return y Step 4: else if key[y] ¡ k then y right[y] Step 5: else y left[y] Step

6: return (NOT FOUND)
Output : Expected Search result

C. Mathematical Model

S = { I, O, P, U } U = { DO, DU }
Where, DO = Data owner DU = Data User

I={ DC, UAD, SK, PK,W }
where, DC = Uploaded Document UAD= User Authentication SK = Secret key W = set of n keyword to search

O={ EDC, DDC, In, TPK }
where, EDC = Encrypted Document DDC = Decrypted Document In = Index Tree TPK = Top K document

F={ UA, KG, ENC, DEC, GI ,SE, GQ, SS }
where, UA = User Authentication KG = Key Generation ENC = Encryption of Document using RSA DEC = Decryption of document usng RSA
GI = Generation of index tree

SE = search GQ = Query generation for W keywords SS = Synonym search

D. Set Theory

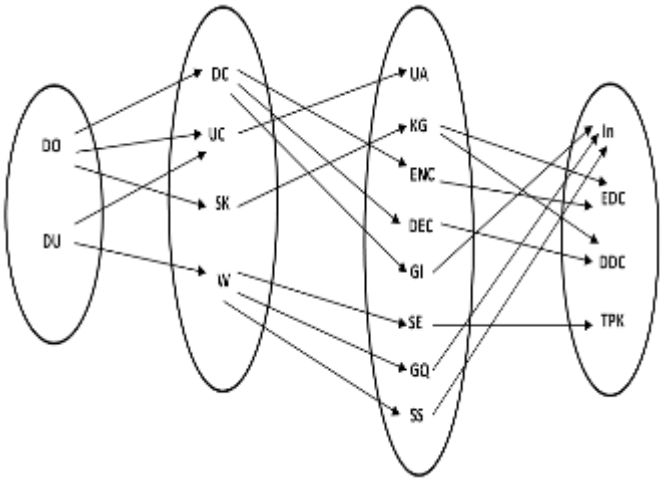

Figure 2: Set Diagram

E. Experimental Setup

For the betterment of search result over encrypted data on cloud , we have worked on synonym query technique. For this we have used wordnet API that provides us nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms (synsets), each expressing a distinct concept. Synsets are interlinked by means of conceptual-semantic and lexical relations. Also we have purchased cloud for 2 months to get real world system experience. We have used dataset Reuters News stories. From this we put 1000 training document and 400 test document on live cloud. For accuracy of result , dictionary of 2000 words is created and placed on cloud which is used for searching purpose.

F. Results

To verify some basic ideas we observe output for search result time and index generation time.

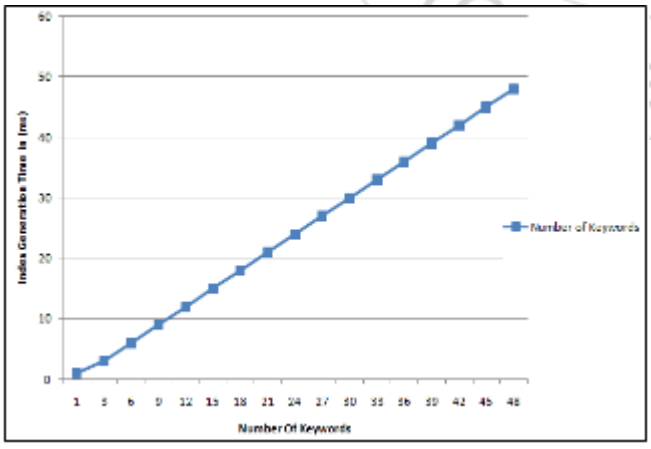

Figure A: Number of keyword Vs Index Generation Time

From fig A it is clear that for same size of dataset , time required to generate index is directly proportional to the number of keywords as expected. Time is shown in milliseconds.

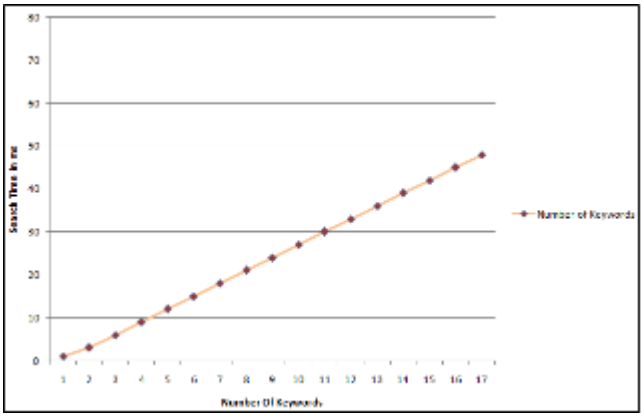

Figure B: Number of keyword Vs Document Search Time (in ms)

From fig B it is also clear that as keywords increases , then for same size of dataset time required is directly proportional to the number of search keywords given by the user. While searching number of keywords are mapped and for betterment ranking of search result is done. Hence it is obvious that time should increased with number of keywords.

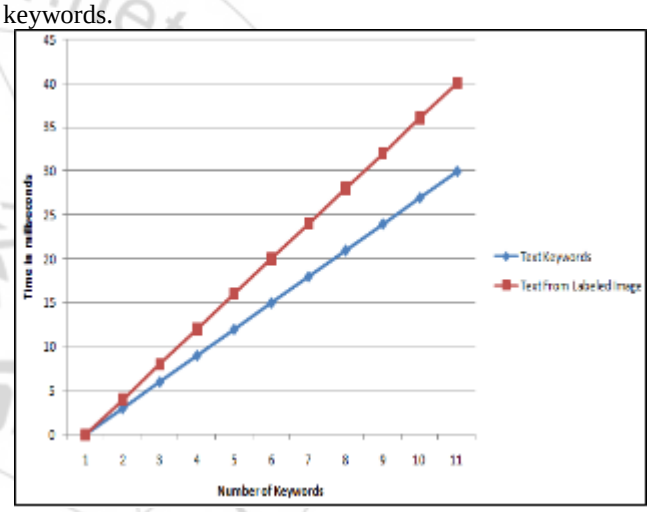

Figure C: Number of keyword Vs Document Search Time

(in ms) for text as input to search and labeled image as input to search

In our system we can give labeled image as input. Labeled are extracted through OCR algorithm hence extraction time is added in search time. Hence in comparison with normal text as input , time required is greater for labeled image as input.

It is prime requirement that labeled image should contain labels present in built dictionary so mapping can be done.

4. Conclusion

Proposed system is decentralized system in which distributed nodes work together for data security on cloud by implementing encryption facility , also these nodes manages multi user tasks like sharing , writing data , reading data etc. Due to this decentralized approach keys are managed at different node hence cloud is not having keys for decryption hence data security is assured. Also KDC is not having data hence only encryption keys are not useful to it. Three types of user like owner , writer and reader has respective access

control to the data. Hence this system is also manages hierarchical scenarios as far as users role is concern.

References

[1] Zhangjie Fu,Xingming Sun Nigel Linge, Lu Zhou,"Achieving Effective Cloud Search Services:Multi-keyword Ranked Search over Encrypted Cloud Data Supporting Synonym Query ", IEEE Transactions on Consumer Electronics, Vol. 60, No. 1, February 2014.
[2] Georgia Koutrika,Saint Petersburg, Russia "Data Clouds: Summarizing Keyword Search Results over Structured Data", EDBT 2009, March 2426,ACM, pp. 391-402,2009.
[3] A multimedia search for the cloud architecture is suggested by Wei-Ying Ma
[4] Byron Y-L. Kuo (2007), "Tag Clouds for Summarizing Web Search Results" ,WWW 2007, May 812, 2007, Banff, Alberta,Canada.pp.1203.
[5] Daniel E. Rose (2012), "CloudSearch and the Democratization of Information Retrieval Wei-Ying Ma (2009),Rethinking Multimedia Search in the Clients + Cloud Era, LS-MMRM09, October 23, 2009,Beijing, China. Pp. 1-1.
[6] J. Li, Q. Wang, C. Wang, N. Cao, K. Ren, and W. Lou, "Fuzzy keyword search over encrypted data in cloud computing", Proceedings of IEEE INFOCOM10 Mini-Conference, San Diego, CA, USA, pp. 1-5, Mar.2010.
[7] Cengiz Orencik (2012), Efficient and Secure Ranked Multi-Keyword Search on Encrypted Cloud Data , PAIS 2012, March 30, 2012, Berlin,Germany. ACM, p 186-195.
[8] Mathew J. Wilson (2012), "Keyword Clouds: Having Very Little Effect on Sensemaking in Web Search Engines" , CHI 2012, May 510, 2012, Austin,Texas, USA,ACM,p2069-2074.
[9] S. Kamara, and K. Lauter, "Cryptographic cloud storage", ,FC 2010 Workshops, LNCS 6054, PP. 136-149, Jan. 2010.
[10] I. H. Witten, A. Moffat, and T. C. Bell, "Managing gigabytes: Compressing and indexing documents and images", Morgan Kaufmann Publishing:San Francisco, May 1999, PP. 36-56
[11] S. Grzonkowski, and P. M. Corcoran, "Sharing cloud services: user authentication for social enhancement of home networking", IEEE Trans.Consumer Electron., vol. 57, no. 3, pp. 1424-1432, 2011.
[12] Ayatullah Faruk Mollah , Nabamita Majumder , Subhadip Basu and Mita Nasipuri , "Design of an Optical Character Recognition System for Camerabased Handheld Devices", IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 4, No 1, July 2011.
[13] W. Sun, B. Wang, N. Cao, M. Li, W. Lou, and Y. T. Hou, "Privacy preserving multi-keyword text search in the cloud supporting similarity based ranking", ASIACCS 2013, Hangzhou, China, May 2013, pp. 71-82,2013.
[14] Mehmet Kuzu, Mohammad Saiful Islam, "Efficient Similarity Search over Encrypted Data Department of Computer Science", The University of Texas at Dallas Richardson, TX 75080, USA.
[15] D. A. Grossman, and O. Frieder," Information retrieval: algorithms and heuristics", 2nd ed., Springer Publisher: Berlin, 2004, pp. 18-20.

References

Manish M. Pardeshi completed B.E.(Computer) from Gokhale Education society's college of Engineering, Nashik and Pursuing Master degree (Computer Science and Engineering) from UoP, Amrutvahini College Engineering, (AVCOE), Sangamner.

Prof. Rahul L. Paikrao is a associate professor and head of computer engineering department at Amrutvahini College of Engineering (AVCOE), Sangamner (Computer Science and Engineering), Pursuing PhD in Cloud computing security from UoP, He has 10 years of teaching experience.