ENCRYPTED FILE SHARING USING KEYWORD SEARCH AND GRAPHICAL PASSWORD

Anjali[1], Shashi Sharma[2]
[1]M.Tech Scholar, [2]Assistant Professor, Computer Science Department, Jaipur Institute of Technology.

Abstract: Data is growing at the enormous rate and all big organization requirements are to manage and access the information from such big data and this information is also required to be access securely. For this in our thesis work we are proposing the concept and algorithm of the multi-keyword search, in which we have categorized the users as data owner, administrator and data user. The Data owner will upload the documents on the cloud server and data user will search for the required documents and administrator will manage the access of documents. Apart from this in order to search the big data, we have used the multi keyword search algorithm and in order to securely accessing the documents we have protected the access using the graphical password with the real time movement of the object and the password is also generated as OTP and the bank and other financial organizations require the data to be encrypted so that the concern persons will able to decrypt the required information.

Keywords: Big Data, Keyword Search, Graphical OTP

I. INTRODUCTION

Headings Big Data security and privacy includes Big Data management and investigation for cyber security. While Big Data has establishes in numerous technologies, database management is at its heart. Consequently in this area we will examine how data management has advanced and will then focus on the Big Data security and privacy issues.[1] Database systems technology has propelled an extraordinary arrangement during the previous four decades from the legacy systems in light of network and various leveled models to relational and question database systems. Database systems can likewise now be gotten to by means of the web and data management administrations have been executed as web administrations. Because of the blast of electronic administrations, unstructured data management and web-based social networking and versatile registering, the measure of data to be taken care of has expanded from terabytes to petabytes and zetabytes in only two decades. Such immeasurable measures of complex data have come to be known as Big Data. Not exclusively does big data need to be overseen effectively, such data additionally must be analyzed to remove helpful chunks to upgrade businesses and enhance society. This has come to be known as Big Data Analytics[1]. Capacity, management and examination of huge amounts of data likewise result in security and privacy violations. Frequently data must be held for different reasons including for administrative consistence. The data held may have touchy information and could disregard client privacy. Moreover, controlling such big data, for example, consolidating sets of various sorts of data could bring about security and privacy violations. For ex-abundant, while the crude data evacuates by and by identifiable information, the determined data may contain private and touchy information. For instance, the crude data about a man might be consolidated with the per-child's address which might be adequate to distinguish the individual. Distinctive communities are taking a shot at the Big Data challenge. For instance, the systems community is developing technologies for enormous stockpiling of big data. The network community is developing solutions for overseeing huge networked data. The data community is developing solutions for effectively man-maturing and examining extensive arrangements of data. Big Data research and development is being completed both in the scholarly world, industry and government research labs. Nonetheless, little consideration has been given to security and privacy contemplations for Big Data. Security cuts over various zones including systems, data and net-works. We require the numerous communities to meet up to develop solutions for Big Data security and privacy.

1.1 Big data analytics for security applications

While the difficulties examined with securing Big Data and guaranteeing the privacy of people, Big Data management, and analytics procedures can be utilized to take care of security problems. For instance, an association can outsource exercises, for example, personality management, email filtering and intrusion detection to the cloud. This is on the grounds that gigantic measures of data are being gathered for such applications and this data must be analyzed. Cloud data management is only one case of big data management. The question is, by what method can the developments in big data management and explanatory systems be utilized to take care of security problems? These problems incorporate malware detection, insider danger detection, intrusion detection, and spam filtering.

1.2 Encrypting and decrypting Data

In network security, cryptography has a long history by gives an approach to store sensitive Information or transmit it crosswise over insecure networks (i.e. the Internet) so it can't be perused by anybody aside from the expected beneficiary, where the cryptosystem is an arrangement of algorithms consolidated with keys to change over the first message (Plain-text) to encrypted message (Cipher-text) and change over it back in the proposed beneficiary side to the first message (Plain-text) [1]. The primary model proposed by Shannon on the cryptosystem is appeared in figure 1 [2].

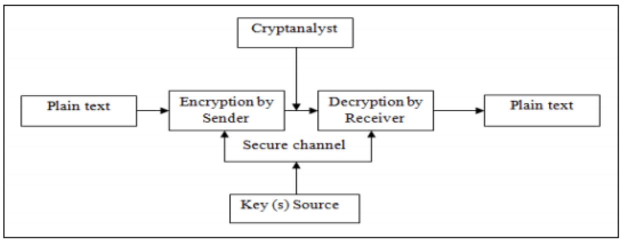
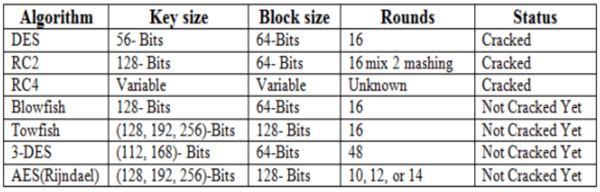


Fig.1 Shannon model of secret communication

1.3 Comparisons of Most Popular Encryption Algorithms

There is a significant number of encryption algorithms utilized for keeping information secured. Their Complexity and capacity to oppose assault fluctuates starting with one algorithm then onto the next. The fundamental component of encryption process is the algorithms that fill essential need in various ways. Popularly utilized algorithms incorporate DES, Triple DES, RC2, RC4, Blowfish, Twofish and Rijndael (AES) as we said in theory. The essential information of the most prominent ciphers is appeared in table 1[5].

Table 1 Comparison of popular encryption algorithms



Algorithm	Key size	Block size	Rounds	Status
DES	56- Bits	64-Bits	16	Cracked
RC2	128- Bits	64- Bits	16 mix 2 mashing	Cracked
RC4	Variable	Variable	Unknown	Cracked
Blowfish	128- Bits	64-Bits	16	Not Cracked Yet
Towfish	(128, 192, 256)-Bits	128- Bits	16	Not Cracked Yet
3-DES	(112, 168)- Bits	64-Bits	48	Not Cracked Yet
AES(Rijndael)	(128, 192, 256)-Bits	128- Bits	10, 12, or 14	Not Cracked Yet

1.4 Advanced Encryption Standard (AES)

In light of the table 1, the National Institute of Standards and Technology (NIST) in 1997, declared authoritatively that Rijndael algorithm would turn into the Advanced Encryption Standard (AES) to supplant the maturing Data Encryption Standard (DES). AES algorithm is a square cipher text the piece size can be 128, 192 or 256 bits. 128(AES - 128), 192(AES - 192) and 256 (AES - 256) bits key lengths [5-7]. The Rijndael algorithm depends on round capacity, and diverse mixes of the algorithm are organized by rehashing these round capacity distinctive circumstances. Each round capacity contains uniform and parallel four stages, byte substitution, push moving, section blending 147 and key addition, the information is gone through Nr rounds (10, 12, and 14), and each progression has its own specific usefulness as appeared in figure 2 [7].

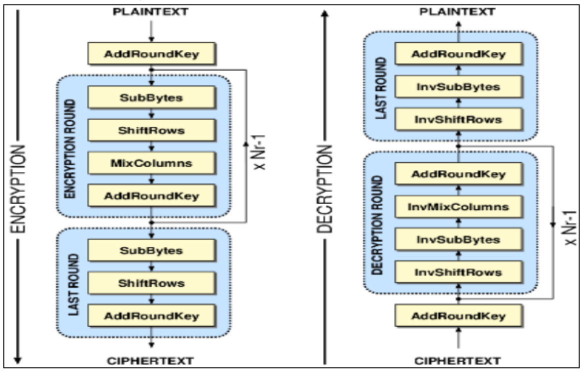


Fig. 2 Advanced encryption standard structure

1.4.1 Public Key Infrastructure (PKI)

PKI gives arrangement of security administrations, for example, authentication, confidentiality, non repudiation, and integrity to the messages being traded [8-10]. In this paper, PKI use in association foundation stage to trade the security esteem between the network terminals i.e. sender, and collector.

1.4.2 Encryption and Decryption of Text using AES Algorithm

In the previous couple of years the security and integrity of information is the fundamental concern. In the present situation every one of the information is exchanged over PC networks because of which it is powerless against different sorts of assaults. To make the information secure from different assaults and for the integrity of information we should scramble the information before it is transmitted or stored. Cryptography is a technique for putting away and transmitting information in a frame that lone those it is expected for can read and process. It is a study of ensuring information by encoding it into a unintelligible configuration. It is a successful method for ensuring sensitive information as it is stored on media or transmitted through network communication ways.

Purpose of cryptography:

1. Authentication: The process of demonstrating one's personality. It is another piece of information security that we experience with ordinary PC utilization. Simply think when you sign into your email, or blog account. The basic sign-in process is a type of authentication that enables you to sign into applications, records, organizers and even a whole PC framework. Once signed in, you have different given benefits until the point when logging out. Some framework will wipe out a session if your machine has been sit still for a specific measure of time, requiring that you demonstrate authentication at the end of the day to re-enter.

The basic sign-on conspire is additionally actualized into solid client authentication systems. Be that as it may, it expects individuals to login utilizing numerous components of authentication. Non-repudiation: In this, the recipient should know whether the sender is not faking. For instance, if assume when one buys something on the web, one ought to make certain that the individual whom one pays is not faking.

2. Integrity: Many a time's information should be refreshed yet this must be finished by validated individuals.

3. Privacy/confidentiality: Ensuring that nobody can read the message with the exception of the planned collector.

Encryption is the process of darkening information to make it ambiguous without unique learning. Encryption has been utilized to secure communications for a considerable length of time, however just associations and individuals with an uncommon requirement for mystery had made utilization of it. In the mid-1970s, in number encryption risen up out of the sole save of shrouded government organizations into the general population space, and is presently utilized as a part of ensuring broadly utilized systems, for example, Web online business, cell phone networks and bank programmed teller machines. Encryption can be utilized to guarantee mystery, however different methods are as yet expected to make communications secure, especially to check the integrity and validness of a message, for instance, a message authentication code (MAC) or advanced marks. Another thought is assurance against activity examination. Interruption can be dealt with by sending a flag to the beneficiary, the one that sends the affirmation motion back to the transmitter, the information will be sent just to that collector. In this way with the utilization of handshaking signals interruption can be dodge.

II. LITERATURE SURVEY

Yue Lu, Chew Lim Tan [1] suggested that a huge amount of document images are accessible in the Internet and digital libraries. They find that, most of them are packed in PDF files and are compressed using CCITT Group 4 standards for saving storage space and speeding up transmission. There is thus significant meaning to develop the methods of directly searching keywords from these documents. In this paper, they present a compressed pattern matching method for searching keywords from the CCITT Group 4 compressed document images, without explicit decompression.

According to the CCITT Group 4 standards, each coded position indicates that the current pixel colour is different from its previous pixel, except for the next coded positions of the pass mode. In their work, they extract these changing elements from the compressed images directly. The changing elements are utilized to segment and bound the word objects, and are used for measuring the similarity of two word images. The associated segments are named based hanging in the balance by-line system as per the relative positions between the changing components of the present coding line and the changing components of the reference line.

Sanket S. Pawar Abhijeet Manepatil Aniket Kadam Prajakta Jagtap[2], This research work is devoted to keyword inquiry and gives two viewpoints of its application in IR and database system. Article show prototype of Machine An and B, where A presents Innovative IR system and B presents Discover approach relational database management system. Article concentrates more on stretching out keyword hunt to database management system as it less tended to subject and all the more difficult. Examination of Machine B demonstrate that execution assessment need to address with successful assessment like inquiry workload memory utilization for adaptable and versatile advanced machine improvement.

Instead of assessment parameters like time delay and so forth mixture versatile report retrieval system is construct and evaluated on memory utilization and inquiry space is decreased significantly with two layer algorithm. Assist extent of system is creating hybridization at machine level and working with pictures as information question.

Qiuxiang Dong, Zhi Guan, ZhongChen[3] In this paper, they grow new techniques that split the computation for the keyword encryption and trapdoor/token era into two stages: an arrangement stage that does by far most of the work to encrypt a keyword or make a token before it knows the keyword or the property list/access control strategy that will be utilized. A moment stage then quickly collects a middle figure content or trapdoor when the specifics get to be distinctly known. The readiness work can be performed while the cell phone is connected to a power source, then it can later quickly perform keyword encryption or token era operations moving without fundamentally depleting the battery. We name our plan Online/Offline ABKS. To the best of our insight, this is the primary work on building productive multi-client searchable encryption conspire for cell phones through moving most of the cost of keyword encryption and token era into a disconnected stage.

Dr Kehinde K. Agbele, Eniafe F. Ayetiran, Kehinde D. Aruleba and Daniel O. Ekong [4] proposed this article to create algorithms that improve the positioning of records recovered from IRS as per client seek setting. Specifically, the positioning assignment that drove the client to take part in information-chasing conduct amid inquiry errands. This article examines and portrays a Document Ranking Optimization (DROPT) algorithm for IR in an Internet-based or assigned databases environment. On the other hand, as the volume of information accessible on the web and in assigned databases is developing persistently, positioning algorithms can assume a noteworthy part with regards to list items. In this article, a DROPT technique for archives recovered from a corpus is produced as for report list keywords and the question vectors. This depends on figuring the weight ( ) of keywords in the report list vector, ascertained as a component of the frequency of a keyword over a record. The motivation behind the DROPT technique is to reflect how human clients can judge the setting changes in IR result rankings as per information significance. This article demonstrates that it is workable for the DROPT technique to beat a portion of the limitations of existing conventional (×) algorithms by means of adjustment. The observational assessment utilizing measurements measures on the DROPT technique helped out through human client cooperation indicates change over the conventional importance input technique to show enhancing IR viability.

XiaoliLian, Mona Rahimi,Remo Ferrari and Michael Smith [4] ,In this paper they first investigate the exertion expected to physically fabricate an abnormal state space demonstrate catching the utilitarian segments. They then present MaRK (Mining Requirements Knowledge), which recognizes and recovers the records containing depictions of practical parts in the area demonstrate. Area investigators can utilize this information to indicate prerequisites. They present and assess an algorithm which positions space archives as indicated by their significance to a part and after that highlights segments of content which are probably going to contain prerequisites related information. They portray prepare inside the setting of the Positive Train Control (PTC) area with a vault of 523 archives, speaking to 852MB of information. They experimentally assess the MaRK significance algorithm and its capacity to recover important prerequisites information for necessities identified with PTC's On-Board Unit.

SanjayAgrawal, SurajitChaudhuri,Gautam Das[5] ,In this paper, they examine DBXplorer, a system that empowers keyword-based pursuit in relational databases. DBXplorer has been executed utilizing a business relational database and web server and permits clients to cooperate through a program front-end. They layout the difficulties and talk about the usage of our system including aftereffects of broad trial assessment.

III. PROBLEM DESCRIPTION

Base Paper Approach (Zhihua Xia, Member, IEEE, Xinhui Wang, Xingming Sun, and Qian Wang ,A Secure and Dynamic Multi-keyword Ranked Search Scheme over Encrypted Cloud Data) :Multi-keyword Boolean search permits the clients to input multiple query keywords to ask for reasonable records. Among these works, conjunctive keyword search plots just give back the records that contain the greater part of the query keywords. Disjunctive keyword search plans give back the greater part of the records that contain a subset of the query keywords.

Positioned search can empower fast search of the most significant information. Sending back just the top-k most significant archives can adequately diminish organize activity.

System demonstrate

The system demonstrates in this paper includes three unique elements: information proprietor, information client and cloud server.

Gaps in the Base Paper:

Initially, every one of the clients as a rule keep the same secure key for trapdoor era in a symmetric SE conspire. For this situation, the denial of the customer is gigantic test. If it is relied upon to deny a customer in this arrangement, we need to alter the list and scatter the new secure keys to all the approved clients.

Also, symmetric SE conspires usually assume that all the data users are dependable. It is not practical and an exploitative data client will lead to many secure issues. For example, a deceptive data client may search the reports and convey the unscrambled archives to the unauthorized ones. Considerably more, an unscrupulous data client may disseminate his/her secure keys to the unauthorized ones. Later on works, we will attempt to enhance the SE plan to handle these challenge issues.

Additional Solution:

1. Fast Multi-keyword search algorithm is proposed to actualize the search using the Associative Mapping so will take lesser time as compare to the normal search.
2. For the secure key we have contrived the novel approach of the password generation or the key generation of the access of the records shared,

We take a matrix of the 6x6 in which we will place the 6 pictures at any of the random locations.

Proposed Work and Implementation

In our proposed approach we have make use of two algorithms,

Algorithm 1: For Keyword Search

Step1: Capture the Keyword String user entered for Searching

Step 2: Split the multi-keyword string into an array. Now each element of array is the keyword to be searched.

Step 3: In the keyword search, we will maintain the following data structures,

Structure 1 :
Filename
Uploaded By
Keyword matched
Line Number

By making this structure we will get access the lines of the file containing the keyword.

In further we will modify the concept of uploading the document on the category basis.

i.e. Structure for File Details
Filename
Uploaded By
Date Time

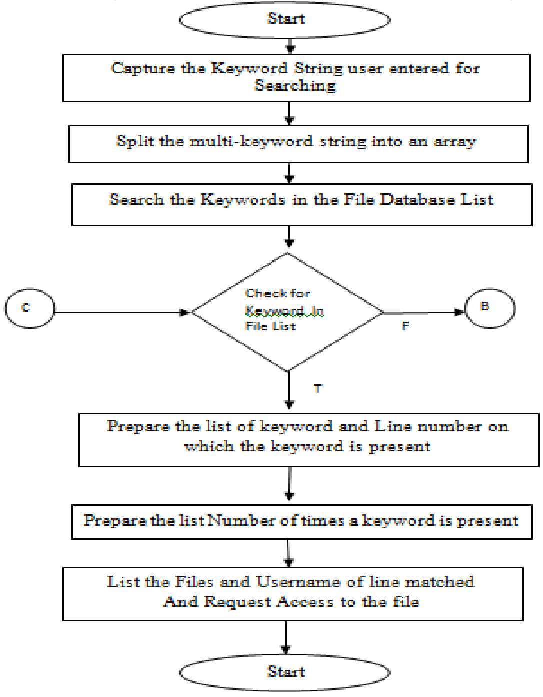
Structure for Keywords
Category Id
Category Name
Keywords

When the user uploads the file then on the basis of the category a detailed record is stored in the following table structure
File Name
Keyword Matched
Line Number

This structure can contain multiple entries for the same keyword as the same keyword can appear in the various lines.

In order to speed up the search we can use an associative memory structure
Filename
Keyword
Match Times
Uploaded By

In order to just get the matched document will the keyword.


Start
Capture the Keyword String user entered for Searching
Split the multi-keyword string into an array
Search the Keywords in the File Database List
C
Check for Keyword In File List
F
B
T
Prepare the list of keyword and Line number on which the keyword is present
Prepare the list Number of times a keyword is present
List the Files and Username of line matched And Request Access to the file
Start

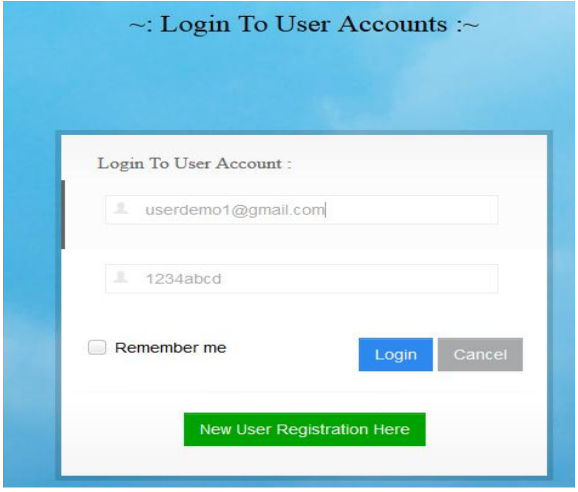
Algorithm 2: Secure Graphical OTP pin generation

Step 1: Place the Images in the Grid first by clicking on the image and then on the position in the grid where we want to place the image.
Step2 : After all the images are arranged in the grid the code will scan the grid starting from the first row and then processing to the last row and scanning each column in the row.
Step 3: If the column contains an image then it will participate in creating the pin and the concept involved First letter of the image following by row and the column number and this process is repeated for all the images in grid.

Step4 : Then mail the generated pin to the user and user then reenter the pin using the same process as mentioned in the step 1.

We have created the implementation on line on the website, using the PHP and MYSQL.

4.1 Login Form



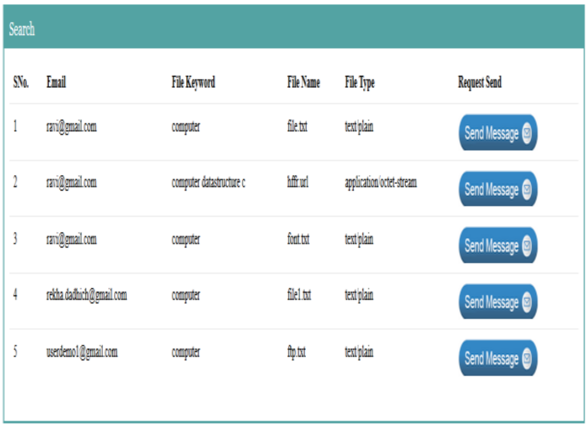
~: Login To User Accounts :~

Login To User Account :

userdemo1@gmail.com

1234abcd

Fig 3 Login Form



Search

SNo.	Email	File Keyword	File Name	File Type	Request Send
1	ravi@gmail.com	computer	file.txt	text/plain	Send Message
2	ravi@gmail.com	computer datastructure c	hfff.url	application/octet-stream	Send Message
3	ravi@gmail.com	computer	font.txt	text/plain	Send Message
4	rekha.dadhich@gmail.com	computer	file1.txt	text/plain	Send Message
5	userdemo1@gmail.com	computer	ftp.txt	text/plain	Send Message

Fig 4 Result of Keyword Search

4.2 Encryption

In term of encryption handle, the algorithm comprises of blend of open key framework for cross breed framework and RC6 algorithm for disarray and dispersion operations as appeared. The proposed encryption algorithm comprises of the accompanying procedures as appeared in figure 5

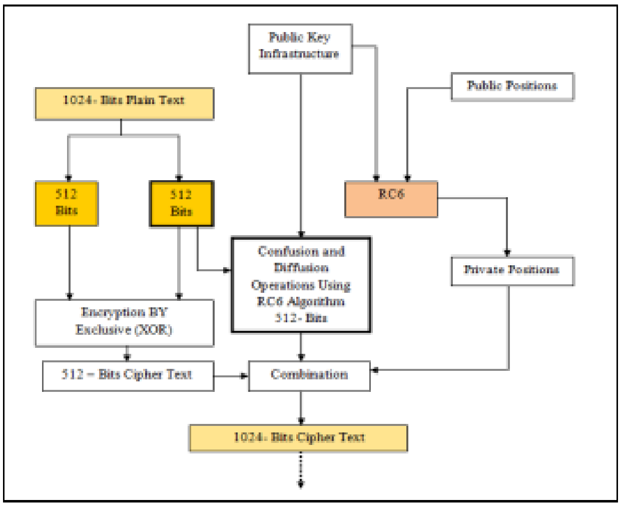
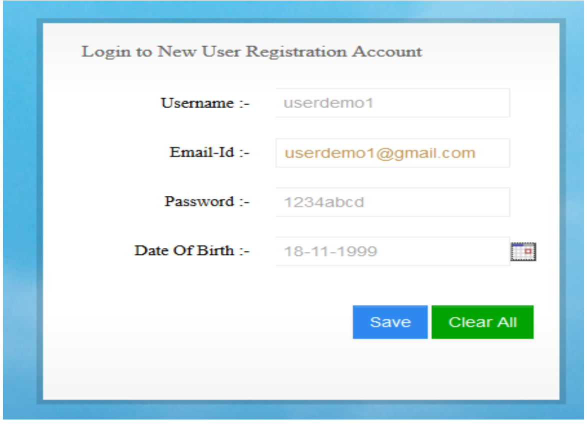


Fig 5 Proposed encryption structures

IV. IMPLEMENTATION

The We have crated the implementation on line on the website, using the PHP and MYSQL.

5.1 Client Registration Page



Login to New User Registration Account

Username :- userdemo1

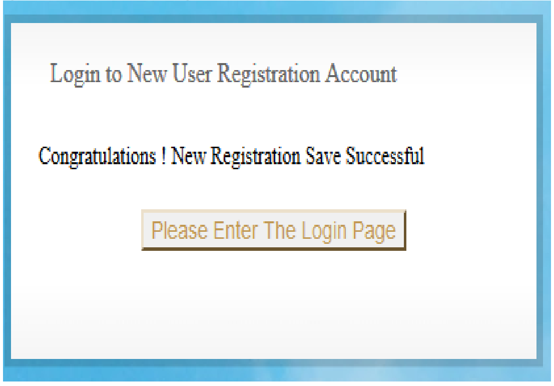
Email-Id :- userdemo1@gmail.com

Password :- 1234abcd

Date Of Birth :- 18-11-1999

Fig 6 New User Registration Form

This is the client registration page and note that the username and email id are unique for the registration and the registered client can act as the data user or the data owner for the sharing or downloading the file.



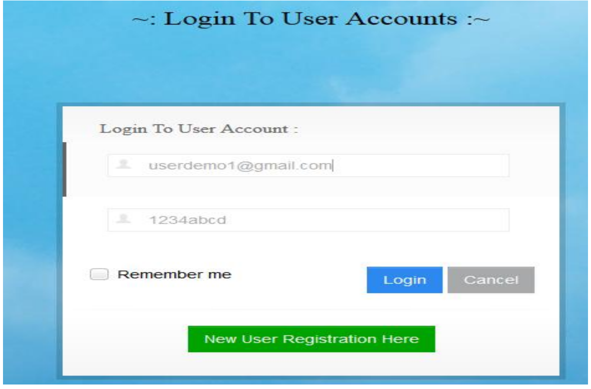
Login to New User Registration Account

Congratulations ! New Registration Save Successful

Fig 7 Registration Confirmation Message

The acknowledgement is received after the registration of the user is completed.

5.2 Login Page



~: Login To User Accounts :~

Login To User Account :

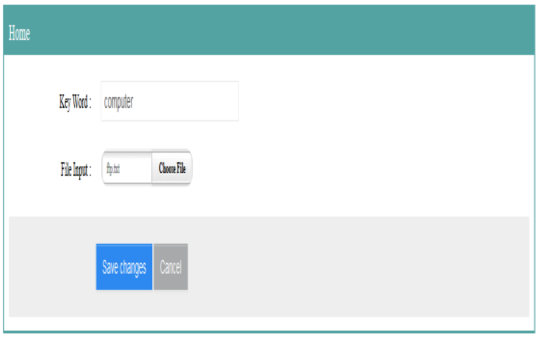
userdemo1@gmail.com

1234abcd

Fig 8: Login Form

This is the user account login page and it will be used by the user for getting access the services for the secure file sharing.

5.3 File Upload Form



Home

Key Word : computer

File Input : ftp.txt Choose File

Fig. 9: File Upload Form

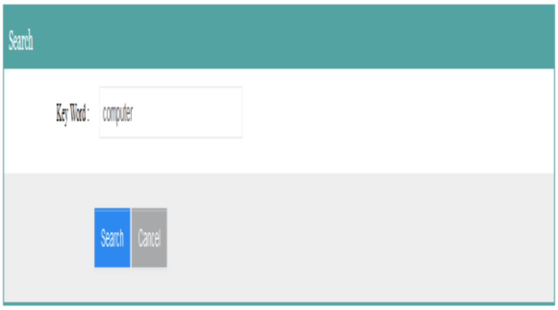
This form is used for uploading the file and while uploading the file we will also specify the keywords which are then matched into the file for specifying the lines which are matched with the file.

Home

Record Save Successful

Fig 10 File Upload Success Message

5.4 Keyword Search

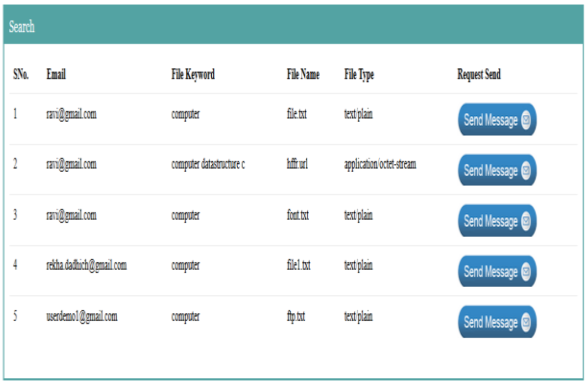


Search

Key Word : computer

Fig 11 Keyword Search Form

This is the keyword search form and we will specify the keyword we want to search in the file. (This module is partially completed, we are working on refining the algorithm for the keyword search) Then the files which are matched for the keyword are listed and we cannot directly download the file we have to request the access for downloading the file.

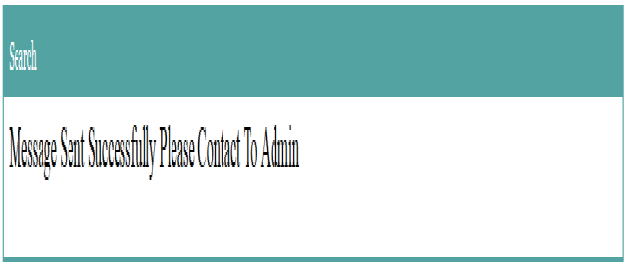


Search

SNo.	Email	File Keyword	File Name	File Type	Request Send
1	ravi@gmail.com	computer	file.txt	text/plain	Send Message
2	ravi@gmail.com	computer datastructure c	hfff.url	application/octet-stream	Send Message
3	ravi@gmail.com	computer	font.txt	text/plain	Send Message
4	rekha.dadhich@gmail.com	computer	file1.txt	text/plain	Send Message
5	userdemo1@gmail.com	computer	ftp.txt	text/plain	Send Message

Fig 12 Result of Keyword Search

When we click on the send message then the request for accessing the file is generated and then the role of admin will come into play that will generate the secure graphical pin for accessing the file. (Here the userdemo2 has requested for the access)\



Search

Message Sent Successfully Please Contact To Admin

Fig 13 Message for Successful Request

V. CONCLUSION AND FUTURE SCOPE

As of now information security is crucial to all organization to ensure their information and behaviors their business. Information security is characterized as the assurance of information and the framework, and equipment that utilization store and transmit that information. Information security performs four vital for an organization which is ensure the organization's capacity to work, empower the sheltered operation of utilizations actualized on the organization's IT systems, secure the information the organization gather and uses, and in conclusion is shields the technology resources being used at the organization. There are likewise difficulties and hazard includes in actualized information security in organization. In an organization, information is vital business resources and basic for the business and along these lines require fitting ensured. This is particularly essential in a business domain progressively interconnected, in which information is presently presented to a developing number and a more extensive assortment of dangers and vulnerabilities. Cause harm, for example, noxious code, PC hacking, and disavowal of administration assaults have turned out to be more typical, more goal-oriented, and more complex. Along these lines, by executed the information security in an organization, it can ensure the technology resources being used at the organization. In term of ensuring the functionality of an organization, both general management and IT management are in charge of executing information security that secures the organization capacity to work. Information is the most critical component in organization to work together. Other than that an organization is kept their clients information, so it is crucial for them to ensure the information. Without information, the business can't be run. By secure the information store; it can empower the organization to run business also. That is the reason the information security is critical in organizations.

REFERENCES

[1] Gary Pan, Seow Poh Sun, Calvin Chan and Lim Chu Yeong,"Analytics and Cybersecurity: The shape of things to come",CPA ,2015

[2] Erol Gelenbe and Omer H. Abdelrahman,"Search in the Universe of Big Networks and Data",IEEE ,2014

[3] Shengli Wu,Chunlan Huang,Jieyu Li,"Combining Retrieval Results for Balanced Effectiveness and Efficiency in the Big Data Search Environment",IEEE International Conference on Computer and Information Technology,2014

[4] Ajeet Lakhani,Ashish Gupta,K. Chandrasekaran,"IntelliSearch: A Search Engine based on Big Data Analytics integrated with Crowdsourcing and category-based search",International Conference on Circuit, Power and Computing Technologies ,2015

[5] Zhihua Xia, Member, Xinhui Wang, Xingming Sun, and Qian Wang,"A Secure and Dynamic Multi-keyword Ranked Search Scheme over Encrypted Cloud Data",IEEE,2015

[6] Bing Wang, Wei Song, Wenjing Lou Y. ,Thomas Hou,"Inverted Index Based Multi-Keyword Public-key Searchable Encryption with Strong Privacy Guarantee",IEEE Conference on Computer Communications (INFOCOM),2015

[7] N. L. Sarda and A. Jain. Mragyati: A system for keyword-based searching in databases. http://arxiv.org/abs/cs.DB/0110052.

[8] Marcos D. Assuncao, Rodrigo N. Calheiros, Silvia Bianchi, Marco A.S. Netto, Rajkumar Buyya "Big Data computing and clouds: Trends and future directions",J. Parallel Distrib. Comput. 79–80 (2015) 3–15.

[9] Venkata Narasimha Inukollu, Sailaja Arsi and Srinivasa Rao Ravuri "SECURITY ISSUES ASSOCIATED WITH BIG DATA IN CLOUD COMPUTING",International Journal of Network Security & Its Applications (IJNSA), Vol.6, No.3, May 2014.

[10] Puneet Singh Duggal, Sanchita Paul "Big Data Analysis: Challenges and Solutions" International Conference on Cloud, Big Data and Trust 2013, Nov 13-15, RGPV.

[11] Zan Mo, Yanfei Li "Research of Big Data Based on the Views of Technology and Application" American Journal of Industrial and Business Management, 2015, 5, 192-197.