International Journal of Technical Innovation in Modern Engineering & Science (IJTIMES)
Impact Factor: 5.22 (SJIF-2017), e-ISSN: 2455-2585
Volume 4, Issue 09, September-2018

PRIVACY PRESERVING SMART SEMANTIC RANKED KEYWORD SEARCH RESULT VERIFICATION USING CLOUD COMPUTING TECHNOLOGY

N.Sugapriya[#1], Dr.T.Ramaprabha[#2]

M.Phil Full-Time/Research Scholar[#1], Professor[#2]
PG and Research Department of Computer Science[#1, 2]
Vivekanandha College of Arts and Sciences for Women, (Autonomous)[# 1, 2]
Tiruchengode, Namakkal-DT, TamilNadu, INDIA[#1, 2]

Abstract:
In today's era a brand new generation of technology is remodeling the planet of computing. Advances in web primarily based information storage, process and services. Looking could be a key technology of the online, since it's the first thanks to access content on the online. Current customary looking is actually supported a mix of matter keyword search with an importance ranking of the documents counting on the link structure of the online. There square measure 3 sorts of options to be used. First, what's the linguistics of keyword search; second, what constitutes an honest answer, or, the way to rank the answers; third, the way to perform keyword search with efficiency. We've got projected a brand new theme to subsume protective the privacy in encrypted cloud information exploitation hierarchic keyword looking. currently a day's great amount of information is outsourced on cloud server in order that information privacy is major issue in cloud computing due to cloud information server isn't totally trustworthy there for data is encrypted so keep on cloud. During this paper, we've got an inclination to supply a transient outline of existing such approaches, furthermore as own ones, and sketch some potential future directions of research.

Keywords: Privacy search, Top-k search, Semantic Ranked keyword search.

I. Introduction

Net search could be a key technology of the online that is basically supported a mix of matter keyword search with an importance ranking of the documents counting on the link structure of the online. For this reason, it's several limitations, and there a lot of than way over analysis activities towards more intelligent net search, known as linguistics search on the online, or conjointly linguistics net search, that is presently one in every of the recent analysis topics in each the linguistics net and net search. Keyword search is that the data retrieval mechanism for knowledge on the globe Wide net. It conjointly proves to be an efficient mechanism for querying semi-structured and structured knowledge, due to its easy question interface. Recently, question process over graph-structured knowledge has attracted increasing attention, as myriads of applications are driven by and manufacturing graph structured knowledge. Keyword search permits users to question the databases quickly, with no ought to understand the schema of the various databases. Additionally, will facilitate keyword search discover surprising answers that are usually tough to get via rigid-format SQL queries.[1]

Privacy Requirements for Semantic Ranked Keyword Searching

To build the safety of encrypted data that is distributed by the consumer to cloud server, the elemental objective is to code and decode the data in a very much secured path with less time and fewer value in each the encoding and decoding method. The info might get disclosed or changed by any unauthorized access. A secure storage should be achieved in cloud computing. Thus we have a tendency to adopt cryptologic techniques for the secure storage. The info is encrypted by the info owner before the info is uploaded to the cloud. The foremost feature of a cryptologic storage is that the safety properties that area unit delineated below area unit accomplished.

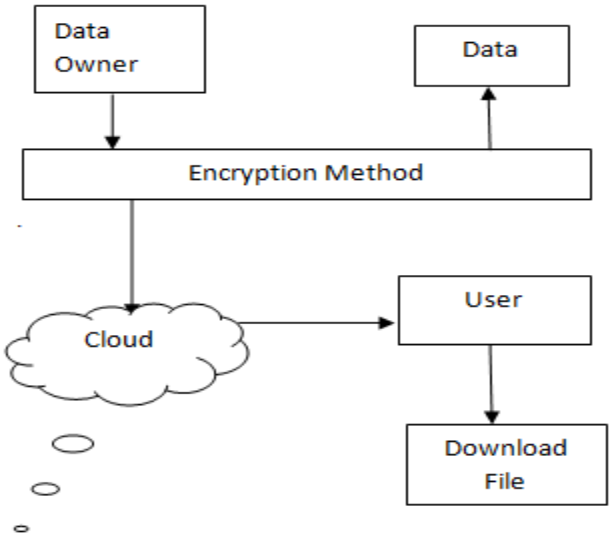

Figure 1: Semantic Ranked Keyword Search

The on top of diagram represents encrypted cloud storage. The owner of the information applies cryptanalytic ways to the sensitive data to safeguard the knowledge from unauthorized access. Information the information owner uploads the encrypted data to the cloud setting. The approved user will decode the info and transfer the specified file. The Strength of cryptanalytic Cloud Storage is usually looking on 2 factors they're Confidentiality and Integrity. The information's were encrypted with the advanced cryptanalytic techniques and so the privacy is maintained. Cloud Storage provides Integrity to the info and so it prevents any unauthorized individuals to change the info.

Design goals

The design of our system model supports linguistics keyword search over outsourced encrypted information in cloud with the subsequent security and performance paradigms.

1) Information privacy: which means we should always shield documents from the cloud server whereas keeping accessible for users. And ancient biradial cryptography may be a smart choice. We are able to encode the documents by biradial key cryptography before outsourcing.

2) Index privacy: we have a tendency to should guarantee that the cloud server can't predict the connection between the documents and keywords through the index.

3) Keyword privacy: during this paper, we have a tendency to read each a part of CG as a keyword and also the keyword is expounded with one another. Therefore it's necessary to preserve users' question CG (sentence) and that we ought to generate secure trapdoors to avoid the escape of data regarding question.

4) Efficiency: The on top of functionalities is achieved with low storage, low network traffic and with low computation and search time.

II Existing Methodology

Introduce abstract graphs **(**CGs) as a data illustration tool. Then, gift the 2 schemes (PRSCG and PRSCG-TF) supported metric system in keeping with completely different eventualities. So as to conduct numerical calculation, we have a tendency to transfer original metric system into their linear kind with some modification and map them to numerical vectors. Second, use the technology of multi-keyword stratified search over encrypted cloud information because the basis against 2 threat models and lift PRSCG and PRSCG-TF to resolve the matter of privacy-preserving sensible linguistics search supported metric system[14].

Advantages:

- Practical and process schemes to resolve the difficult drawback
- Improvement of security definition
- Fuzzy retrieval on abstract graphs in linguistics level

Disadvantages:

- Difficult to analyze the data
- Keyword is not used to verify the documents

III Proposed Methodology

The proposed system is for probing the data from the encrypted data. The data gets encrypted by data owner with the keyword and store in cloud. The user search for the data, the system will search for the outcome from the encrypted data. The significance scoring and ranking methods are used for only if the truthful top k results. Data encryption protects records safety to some degree, but at the charge of compromised inefficiency. Searchable encryption idea allows recovery of encrypted data over cloud. In this paper, we focus on secure keyword search using the ranked and top-k method.

Advantages:

- The ranked keyword search result verification problem where multiple data owners are involved and the cloud server would probably behave dishonestly.
- The aim of propose a novel secure and efficient deterrent based verification scheme for secure ranked keyword search.
- The project is proposed to optimize the value of data used in the construction of verification data buffer.

This section gives the overview of the ranking and top- k scheme. An efficient keyword ranked Search scheme using synonyms is designed as follows:

Architecture of proposed work:

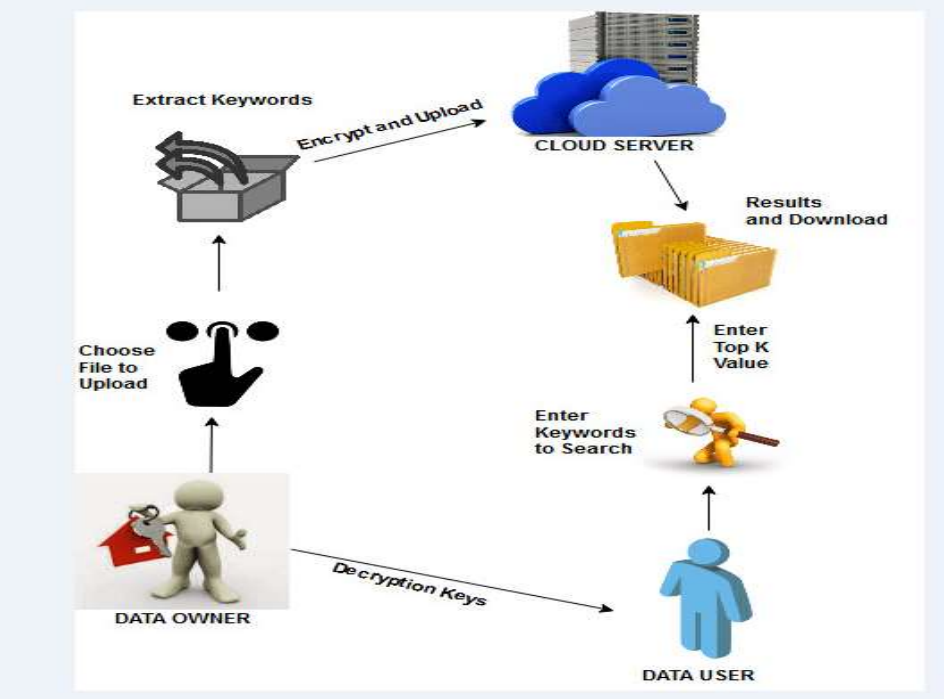

Figure 2: Architecture of search method

Search services involve whole 3 completely different entities:

1. The data owner
2. The data user
3. The Cloud server

In this design represent at first, the information Owner needs to transfer their go in cloud server victimization coding, decoding keys. Then the encrypted File to be uploaded in cloud server. Then user needs to any info suggests that to induce permission the search permission to the information owner. The information owner provides the key keys. Then enter the keyword of the go in cloud server. It shows the top-k leads to relevant file. Next the user to be accessed and transfer the file. During this methodology 2 styles of rule to be used. These are ranking and Top-k retrieval ideas to be enforced the linguistics keyword search design.

Data owner

A cloud computer system processing system ADPS system hosting data service, within which 3 completely different entities are involved: Cloud server, information owner and information user. Since information could contain sensitive info, the cloud servers can't be absolutely entrusted in protective information. For this reason, outsourced files should be encrypted. Any quite info discharge that may have an effect on information privacy is thought to be unacceptable.

Encryption

To improve the machine burden on user aspect, computing work ought to be at the server aspect, want associate coding theme to ensure the operability and security at an equivalent time on server aspect. Homomorphic coding permits specific styles of computations to be allotted on the corresponding cipher text. The results the cipher text of the results of an equivalent operations performed on the plaintext.

Cloud Server:

Cloud server receives the shop request from the information owner and execute the operation of storing the encrypted documents and searchable indexes. Once the information users send the trapdoor to the cloud server, the cloud server makes a computation of relevancy scores and returns top-k connected documents to the information users. The cloud server is additionally to blame for death penalty the command of change documents and searchable indexes.

Figure 3: Entities of Cloud Architecture

Data owner

A cloud computing system hosting data service, in which three different entities are involved: Cloud server, Data owner and Data user. Because data may enclose susceptible in order, the cloud servers cannot be fully entrusted in protecting data. For this reason, outsourced files must be encrypted. Some kind of in sequence outflow that would affect data privacy is regarded as improper.

Encryption

To improve the computational burden on user side, computing work should be at the server side, need an encryption scheme to guarantee the operability and security at the same time on server side. The result is the cipher text of the result of the same operations performed on the plaintext. That is, homomorphic encryption allows computation of cipher text without knowing anything about the plaintext to get the correct encrypted result[2].

Cloud Server:

Cloud server receives the store request from the data owner and execute the operation of storing the encrypted documents and searchable indexes. When the data users send the trapdoor to the cloud server, the cloud server makes a computation of relevance scores and returns top-k related documents to the data users. The cloud server is also responsible for executing the command of updating documents and searchable indexes.

IV Algorithm

Ranking and Top K query processing algorithm

Top k (Q, k)
Step: 1 node Heap← ∅;
Result Heap ← ∅;
Step: 2 θ ← 0;
Step: 3 compute the bloom filter indexes {hi} associated with all keywords in Q;
Step: 4 using PIR retrieve from the root node the paillier cipher texts corresponding to all hi's;
Step: 5 compute the aggregate score for every child of the root node, and insert that node into node Heap;
Step: 6 while (node Heap not empty) do
Step: 7 Remove the top entry from node Heap and store it into e;
Step: 8 if (e.score≤ θ) then

Step: 9 break;
Step: 10 end if
Step: 11 using PIR, retrieve from node e the paillier ciper texts corresponding to all hi's;
Step: 12 compute the aggregate score for every child of e;
Step: 13 if (e is a leaf node) than
Step: 14 Retrieve the metadata files for all document with score > θ;
Step: 15 compute the actual scores of these documents and insert them into result Heap;
Step: 16 16θ ←score of k-th
Document in result Heap;
Step: 17 else /* is an internal node*/
Step: 18 insert every child of e into node Heap;
Step: 19 end if
Step: 20 end while
Step: 21 return results Heap;

V Experimental Analysis

Result and Discussion:

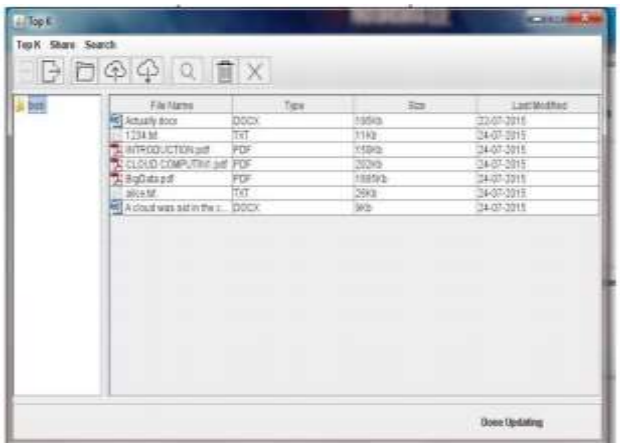

Figure5.1: Document Upload

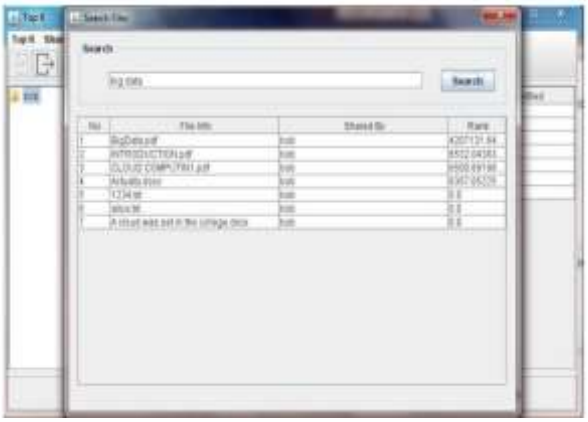

Figure5.2: Ranking data based on keyword

Graphical Representation of Ranking Model

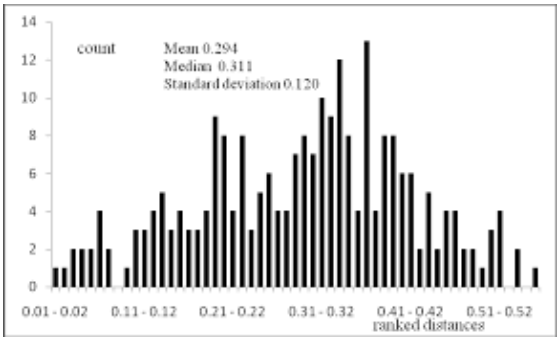

Chart1: Graphical representation of ranking method

VI. Conclusions and future directions

In this paper, we have a tendency to propose a unique top-k learning to rank framework, together with labeling, ranking and analysis, which might be effectively adopted for real search systems. Firstly, a top-k labeling strategy is planned to get reliable relevancy. The planned technique i.e. keyword search and Top-k search is useful in large databases within the cloud. This paper presents the way to increase the search theme for the user intention. Finally, to research the performance of the theme all right by experimenting on real-world dataset. In our future work, we will enhance the security to sustain resourceful active data operations and ranked keyword search in excess of the encrypted big data in cloud.

VII Reference

[1] W. Ogata and K. Kurosawa, "Oblivious keyword search," in Journal of Complexity, Vol.20, 2004, pp. 356–371.

[1] Deepali D. Rane and Dr.V.R.Ghorpade " Multi-User Multi-Keyword Privacy Preserving Ranked Based Search Over Encrypted Cloud Data" International Conference on Pervasive Computing (ICPC), 2015. [2] C. Wang, N. Cao, J. Li, K. Ren, and W. Lou, "Secure ranked keyword search over encrypted cloud data," in Proc. of ICDCS'10, 2010.

[4] D. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," in Proc. of S&P, 2000

[5] Mikhail Strizhov and Indrajit Ray "Multi-keyword Similarity Search Over Encrypted Cloud Data" International Conference on Pervasive Computing (ICPC), 2012.

[6] J. Li, Q. Wang, C. Wang, N. Cao, K. Ren, and W. Lou, "Fuzzy keyword search over encrypted data in cloud computing," in Proc. of IEEE INFOCOM'10 Mini-Conference, San Diego, CA, USA, March 2010.

[7] A. Swaminathan, Y. Mao, G.-M. Su, H. Gou, A.L. Varna, S. He, M. Wu, and D.W. Oard, "Confidentiality-Preserving RankOrdered Search", Proc. Workshop Storage Security and Survivability, 2007.

[8] P. Naresh K. Pavan kumar D. K. Shareef "Implementation Of Secure Ranked Keyword Search By Using RSSE" International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181 Vol. 2 Issue 3, March – 2013.

[9] Q. Tian and S. Chen, "Cross-heterogeneous-database age estimation through correlation representation learning," Neurocomputing, vol. 238, pp. 286–295, May 2017

[10] Z. Wu, B. Liang, and L. You, "High dimension space projection-based biometric encryption for fingerprint with fuzzy minutia," Soft Comput., vol. 20, no. 12, pp. 4907–4918, 2016.

[11] J.F.Sowa, Conceptual structure: Information processing in mind andmachine, Reeading, MA, USA: Addision-Wesely, 1983.

[12] D. X. Song, D. Wagner, and A. Perrig, ―Practical techniques forsearches on encrypted data, in Proc. IEEE Symp. Secur. Privacy, May2000,pp. 44–55.

[13] S.Miranda-Jimnez, A.Gelbukh, and G.Sidorov, "Summarizing conceptual graphs for automatic summarization task," Conceptual Structures for STEM Research and Education,Springer Berlin Heidelberg,pp. 245-253,2013.

[14] Zhangjie Fu and Fengxiao Huang, "Privacy-Preserving Smart Semantic Search Based on Conceptual Graphs Over Encrypted Outsourced Data, "Member, IEEE.Pp-1874-1884,2017.