International Journal For Technological Research In Engineering
Volume 5, Issue 9, May-2018 *ISSN (Online): 2347 - 4718*

IN-DEPTH REVIEW ON KEYWORD SEARCH AND SECURITY CONCEPTS

Deepshikha Bhati[1], Khushbu Soni[2], Sonal Sharma[3]
[1,2]Mtech. Scholar, [3]Assistant Professor
[1,2,3]Computer Science, Rajasthan College Of Engineering For Women, Jaipur Rajasthan.

Abstract: Keyword Search is very important for the relevant document retrieval and the documents numbers is rising every day it is required the most accurate and efficient method of keyword retrieval should be used , this paper review regarding the various method which are available in the keyword search and also extend a review related to the security techniques for securing the online documents.
Keyword: Keyword Search, Information Retrieval, Database, Security.

I. INTRODUCTION

Keyword search techniques are outstandingly useful for separating both the structured and likewise the unstructured data which contains the broad measure of the literary information. In our research paper we will explore distinctive keyword search frameworks and we will in like manner endeavor to separate the domains on which we can work to upgrade execution of keyword search algorithms.

II. CLASSIFICATION OF DATA

Structured and Unstructured Data:

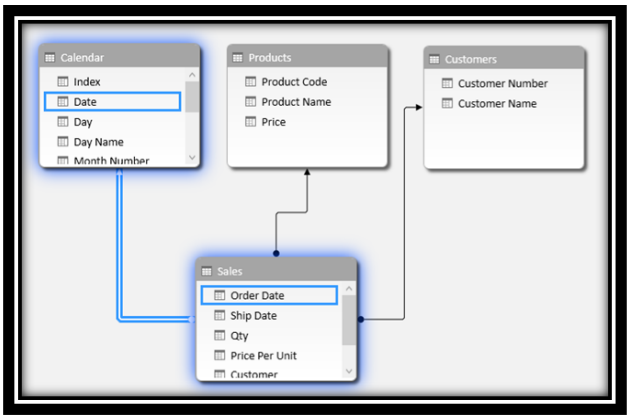

Fig. 1 Relationship among tables via Structured Data.

Structured Data is one in which data is dealt with to the extent structures i.e. relations or tables and that structure will take after a strict database mapping Like in SQL. These tables furthermore form the data to the extent lines and areas, where lines imply tuples or records and sections suggests characteristics and all tables are restricted together with some cardinality or associations e.g. one to many, various to various and so on as show in fig. 1.

Unstructured data is completely backwards to structured data. It contains data that don't dealt with in any predefined Schema. It can be in any shape like Audio, Videos, JPEG Files, Pdfs, Text Files et cetera and it is normally implies information that doesn't live in a conventional line section database.

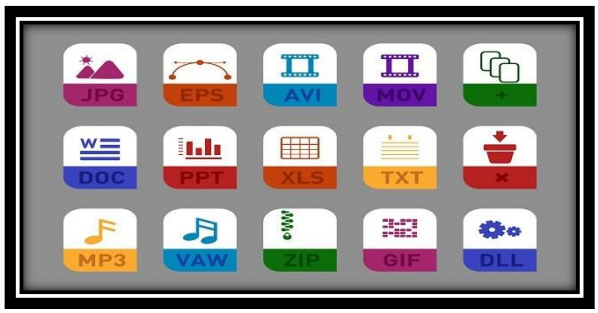

Fig. 2 Unstructured Data

III. KEYWORD SEARCH

Information recuperation is the path toward get-together information by using keywords from the appropriate record and that report can be unstructured or structured data. It hides its diserse quality from customer by giving calculated view. As customer don't have any information about example and some other inquiry dealing with tongue, he can search through unique interface by putting keywords. By using Keyword Search customer can submit keyword to search engines (Internet Search) or structured data and subsequently it reestablishes an once-over of records to customer according to situating. Situating of reports are given in perspective of the keywords match and occasion of keyword facilitate particularly record. Situating is given in plunging solicitation of occasion of keyword facilitate and the document with most extraordinary occasion get higher need.

3.1 Keyword Search Techniques Classification:

Keyword search techniques are classified into two main groups:

3.1.1 Schema Based Keyword Search
3.1.2 Graph based Keyword Search

3.1.1 Schema Based Keyword Search:

Schema based methodologies bolster keyword search over social database (like SQL) using execution of SQL summons [1]. These strategies are mix of vertices and edges including tuples and keys (fundamental and remote key). Each tuple in database uses as vertex and edges portray interdependency among tuples.

By virtue of RDBMS, keyword search using the Schema Based Approach is performed through making use SQL. Mapping Based approach working is divided into the two essential advances:

(i) Determine how to make and create SQL addresses in order to find the structures among tuples.

(ii) Determine how to evaluate the inquiries which are created in step (I) adequately

Discover:
Discover is procedure empowering their client to search into database by means of keywords with any question dialect Knowledge. As indicated by searching keywords Discover First Create Candidate organize chart of tuples and relations at that point diagram yield most limited arrangement first.
It plays out every single searching operation in two noteworthy strides as.
(i) Candidate Network Generator: It helps in creating all candidate networks of relations, which are known as join expressions one that produce the joining networks of tuples.
(ii) Plan Generator: It builds plans for the efficient and the proper evaluation of the set of candidate networks, by making use of the opportunities to reuse common sub expressions of the candidate networks [3].
DISCOVER make utilization of the eager calculation. One primary part of the DISCOVER is that, it performs keyword search without utilizing the prerequisite of the client to know outline of the database.
For positioning the outcome, DISCOVER restores a monotonic score aggregation function [5]. The principle disadvantage of this calculation is that the cost of producing CNs set is high [4].

Spark:
The interest for RDBMs to help keyword search on content data is expanding as there is wide increment in the content data put away in the relational databases. The current keyword search techniques are not achievable for the content data search. The fundamental point of these systems is to centre around viability and productivity of the keyword question search [7].Spark concept devise another positioning equation by making utilization of the current information recovery procedures. The principle utilization of the Spark is that is takes a shot at vast scale genuine databases (Eg. Client Relationship Management by taking in consideration both the RDBMS adequacy and proficiency).
It influences utilization of the Top-k to join calculation which incorporates two effective inquiry handling algorithms for positioning function.
(a) Dealing with Non-monotonic scoring function.
A non-monotonic function is a function that is expanding and diminishing on various interims of its area.
For instance, consider our underlying case f(x) measures up to x2. We saw that this function is expanding on the interim x is more prominent than 0, and diminishing on the interim x is under 0. Since the function is expanding and diminishing on various interims of its area, the function is a non-monotonic function. Fundamentally, if a function isn't expanding on its whole space or diminishing on its whole area, at that point the function isn't monotonic, and we say that it is non-monotonic.
(b) Skyline Sweeping Algorithm.
The Skyline Point Algorithm includes:
(I) Block Nested Loop.
(ii) Divide and Conquer.
(iii) Plane-Sweep.
(iv) Nearest Neighbour Search.
(v) Branch and Bound Algorithms..

(i) Outline of Block Nested Loop
1. Look over a rundown of point and test each point for predominance criteria.
2. Rundown of potential horizon focuses seen so far are kept up by fulfilling a solitary dimension, each went by point is contrasted and all components in the rundown. The rundown is appropriately refreshed.
3. This calculation completes a ton of repetitive work. It has no provision for early termination. Add up to work done relies upon the request in which focuses were experienced.
(ii) Divide-and-Conquer
1. The algorithm recursively isolates extensive datasets into littler parcels. The algorithm proceeds till each littler parcel of the dataset fits in the essential memory.
2. We figure the midway skyline for each segment using any in-memory approach and later unite these fragmentary skyline focuses to shape the last skyline inquiry.

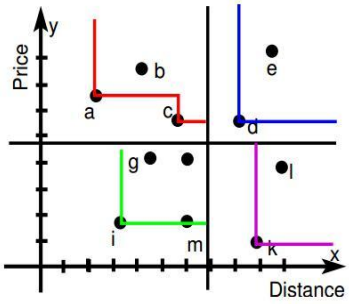

Fig. 3 Skyline Sweeping Algorithm.

(iii) Nearest Neighbour Search
1. Expect that a spatial record structure on the data focuses is open for use.
2. Recognizes skyline focuses by reiterated use of a nearest neighbor search method on the data focuses, using a sensibly portrayed L1 isolate standard.
(iv) Branch and Bound Algorithm
A R-tree depends on the data focuses. Build a need line that coordinates dissents in a MinDist asking for as for the root.
Variations of the Skyline Point Query
1. Situated skyline questions: another slant work is used instead of the base foundation.
2. Compelled skyline questions: The skyline inquiry returns skyline focuses just from the data-space described by the requirement.
3. Recognizing questions: For each skyline point in the dataset, find the amount of focuses in the dataset administered by it.
4. K-Dominating inquiries recuperate the ☆points that lead the greatest number of focuses in the dataset.

3.1.2 Graph Based Keyword Search:
A data graph DG is the sensible portrayal of Relational Database. In this graph algorithm is connected with tackle the keyword search questions. In this portrayal we have, DG (V, E), where DG is the planned graph, v is the arrangement of the vertices address data or record and E is the arrangement of edges which likewise describe relationship substance. In this graph, weights are dole out to the edges in order to address the region of the relating tuples, for

example, we have two vertices u and v then the proximity of u and v is addressed by a weight connoted by We {(u,v)}.

Types of Graph Based Search:
(a) BANKS:-
BANKS stands for Browsing and Keyword Searching. BANKS framework speaks to relational model into data graph and as per coordinating keyword it enacts graph hubs.
Structuring of BANKS algorithm:
In BANKS system, the database is addressed using an organized graph and the record or tuple is addressed as a center point in the planned graph. Outside Key or Primary keys are displayed using the edge, which relates to the association between the comparing tuples. The result handling of BANKS system will restores a sub-graph which is addressed in sort of associating centers, one which arranges the inquiry keyword. This sub-graph can likewise also refined with a particular true objective to get the more exact or more appropriate reaction for the keyword we have searched.
(b) Data Spot
Data Spot is a database passing on instrument and it lets the end client to research the considerable database without making use of any demand vernacular. DataSpot make use of arrangement less semi-made graph which is known as hyperbase. As appeared by the likelihood of the DATASPOT, the Search Server performs searches for inside the hyper base and subsequently as necessities be it returns either HTML pages or question API[9]. The Data Spot utilized as a bit of electronic stock, business list, depicted advancements, help work areas and back.
(c) Proximity search
Proximity search handles general associations among objects each together inquiry replies, these techniques obliging for the savvy ask for sessions. In content Processing Proximity Search searches for Documents or substance records where no under two term events as indicated by sort out are inside chosen detachments. Where seclude is number of broadly engaging words. Web Movie Site Database (IMDB) webpage impacts use of the vicinity to search recollecting the genuine target to answer its database ask. IMDB goals incorporate 140,000 movies and information about more than 500,000 film industry specialists. The idea driving is that the database can be seen as set of associated articles, where objects address films, performers, manager and so on. And the division work in light of joins segregating things [10].

IV. AUTHENTICATION METHODS

Current authentication methods can be divided into three main areas:

- Token based authentication
- Biometric based authentication
- Knowledge based authentication

Token based techniques, such as key cards, bank cards and keen cards are generally utilized. Numerous token-based authentication frameworks likewise utilize knowledge based techniques to upgrade security. For instance, ATM cards are for the most part utilized together with a PIN number.
Biometric based authentication techniques, for example, fingerprints, iris output, or facial recognition, are not yet generally received. The significant downside of this approach is that such frameworks can be costly, and the ID procedure can be moderate and frequently untrustworthy. Be that as it may, this kind of technique gives the most elevated amount of security.
Knowledge based techniques are the most broadly utilized authentication techniques and incorporate both content based and picture-based passwords. The photo based techniques can be additionally divided into two classes: recognition - based and review based graphical techniques. Utilizing recognition-based techniques, a client is given an arrangement of pictures and the client passes the authentication by perceiving and distinguishing the pictures he or she chose amid the enrollment organize. Utilizing review based techniques, a client is requested to duplicate something that he or she made or chose before amid the enlistment arrange.

As we probably am aware graphical pictures are all the more effortlessly reviewed then content. In this segment, graphical secret word framework based on recognition and review based are talked about as beneath:-
Recognition-Based Technique: In this kind of technique, clients will choose pictures, logos or any images from prestored picture. For authentication process client need to perceive the picture, which he pick as a secret word.
Review Based Technique: Again review based secret word authentication are arrange in two sections [2] : (I) Pure Recall Based Technique (ii) Cued Recall Based Technique
Recognition based technique require the client to distinguish and perceive the mystery, or part of it, that the client chose previously. By and large amid secret key creation the clients are required to remember a progression of pictures, and then should perceive their pictures from among fakes to sign in. Phishing assaults are to some degree more troublesome with recognition-based frameworks as a right arrangement of pictures must be displayed to the client before secret key passage. Shoulder-surfing is by all accounts of specific worry in recognition-based frameworks when an assailant is standing behind the client and sees or watches the pictures chose by clients amid login [3][4]. Different recognition based secret word schema are clarified underneath:
(a) Passfaces: The recognition-based framework considered most widely to date is Passfaces. For the most part amid setting a watchword the client chooses an arrangement of human appearances. A board of candidate faces is displayed amid his/her login. Among the given arrangement of distractions the client must choose the faces he/she chose amid setting the watchword. Pass faces basically works by having the client select a subgroup of x faces from a gathering of k faces. For authentication, the framework indicates p appearances and one of the faces has a place with the subgroup q. The client needs to do the determination ordinarily to finish the authentication procedure [5].
(b) Story: The Story conspire, which requires the determination of pictures of articles (individuals, autos, nourishments, planes, touring, and so on.) to shape a story line.

Signaled review based secret key history is for the most part commanded by passpoints. In passpoints the client needs to tap on the five distinct positions or zones of a similar picture. Henceforth it is clicked based graphical secret word. The snap is mouse based and client must recollect the right grouping or arrangement of snap focuses on that foreordained picture for the following fruitful login. It is a tick based plan where clients select a single tick point on every one of 5 pictures in grouping, each one in turn; this gives one-to - one prompting. Amid the following login the client must recall that specific snap point on the offered picture to open the following right picture, if the snap isn't right the following opened picture will be a phony one and not from the picked arrangement of pictures. This will stop current client authentication [7].

V. CONCLUSION

This paper reviews the concepts of the Keyword Search and provides the slight reviews regarding the security models.

REFERENCES

[1] Lu, Yue & Tan, Chew Lim.," Keyword searching in compressed document images". DCC,2003..

[2] S. S. Pawar, A. Manepatil, A. Kadam and P. Jagtap, "Keyword search in information retrieval and relational database system: Two class view," International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT), 2016.

[3] Q. Dong, Z. Guan and Z. Chen, "Attribute-Based Keyword Search Efficiency Enhancement via an Online/Offline Approach," IEEE 21st International Conference on Parallel and Distributed Systems (ICPADS), 2015.

[4] Kehinde K. Agbele, Kehinde Daniel Aruleba, Eniafe F. Ayetiran,"Efficient schema based keyword search in relational databases." University of Computer Studies, Mandalay, Myanmar, International Journal of Computer Science, Engineering and Information Technology (IJCSEIT) 2.6 (2012).

[5] Sanjay Agrawal, Surajit Chaudhuri, Gautam Das, "DBXplorer: enabling keyword search over relational databases",SIGMOD,2002.

[6] Karapakula, M. Puramchand and G. M. Rafi, "Coordinate matching for effective capturing the similarity between query keywords and outsourced documents," IET Chennai 3rd International on Sustainable Energy and Intelligent Systems (SEISCON 2012), Tiruchengode, 2012.

[7] W. Tang, L. Yan, Z. Yang and Q. H. Wu, "Improved document ranking in ontology-based document search engine using evidential reasoning," in IET Software, vol. 8, no. 1, pp. 33-41, February 2014.

[8] Shengli Wu, Jieyu Li, "Merging Results from Overlapping Databases in Distributed Information Retrieval",PDP,2013.

[9] Lakhani, A. Gupta and K. Chandrasekaran, "IntelliSearch: A search engine based on Big Data analytics integrated with crowdsourcing and category-based search", International Conference on Circuits, Power and Computing Technologies , 2015.

[10] Roy Goldman, Narayanan Shivakumar, Suresh Venkatasubramanian, Hector Gercia Molina "Proximity Search In Database" In Proceedings of the 24th VLDB Conference, New York, USA, 1998.

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

