Abstract: Keyword search is the concept for fetching the related data or information. According to the IR, Information Retrieval, keyword search can be considered as the method or search technique which is used for searching the data according to one or more keyword specified by the user. In Information Retrieval, keyword search is a type of search method that looks for matching documents which contain one or more keywords specified by a user. A keyword search scheme to relational database is an important area of research comprising IR and relational database system. This paper review the various techniques of keyword search and its applications to provide the total information lookout on the same.

1. Introduction

With the progress within the field of the information technology there are variety of effective strategies and efficient techniques of keyword search that are presently in use. Keyword search technique is essentially used for looking unstructured data. With time it's resulted in development of variety of techniques of rating and ranking of query results and to estimate the effectiveness of these techniques. In database community main focus is on huge collection of the structured data that resulted in development of variety of artificial techniques and strategies for process or executing the structured queries on the database. In today's era, the mixture of database techniques and also the info retrieval techniques is incredibly very important. With the large growth of web & increasing users of web demanded requirement of keyword search techniques and to increase thought of keyword search over relational data. Keyword search techniques are much helpful for analyzing both of the structured and also the unstructured data that contains the massive amount of the textual info.

Modeling user browsing behavior is a lively research area with tangible real word application. In internet search applications, queries are submitted to look engines to represent the data wants of users. Typically queries might not precisely represent user’s specific info. It should happen completely different users might want to get info on different aspects once they submit a similar query. Keyword search is that the most famous info discovery technique because the user doesn't got to know either a query language or the underlying structure of the information. Large number of techniques are employed in information Retrieval (IR) system. Keyword search is the technique use for the retrieving information or info. Keyword search will be implement on both structured and semi-structured databases, also is attainable on graph structure which mixes relational, html and XML data. Keyword search use variety of techniques and algorithmic rule for storing and retrieving information, less accuracy, doesn't giving a correct answer, need large time for searching and enormous quantity of storage space for information storage.

In our research paper we'll analyze varied keyword search techniques and that we also will attempt to analyze the areas on that we will work to enhance performance of keyword search algorithms.

2. Relational Keyword Search

Keyword search on semi-structured data (e.g., XML) and relational data differs significantly from ancient IR. A contradiction exists in between the data’s logical concept and physical storage of the data. Relational databases are normalized to eliminate redundancy, and foreign keys identify related info. Search queries often cross these relationships (i.e., a set of search terms is present in one tuple and therefore the remaining terms are found in connected or related tuples), that forces relational keyword search systems to recover a logical concept of the data. The implicit assumption of keyword search is, the search terms are related difficult to the search method as a result of usually there are several possible relationships between 2 search terms. It’s nearly always possible to incorporate occurrence of a search term by adding tuples to an existing result.

I. Schema based approaches:

Including tuples and keys (primary and foreign key). There are some techniques are existed for schema primarily based approaches.

i. DISCOVER:

DISCOVER is that the techniques that multiple info Retrieval approaches follow. DISCOVER permits its user to issue
keyword queries without any information of the databases schema or SQL[2]. DISCOVER returns qualified joining network of tuples, that is set of tuples that are associated as a result of they join on their primary and foreign keys, jointly contain all the keywords of the query. DISCOVER uses static optimization. In future, it applies on dynamic optimization. DIS-COVER returns a monotonic score aggregation operate for ranking a result.

ii. SPARK: With the increasing of the text information hold on in relational databases, there are increase a requirement for RDBMS to support keyword query search on text information. For a similar existing keyword search technique can’t fulfill the need of text information search. This techniques specialize in effectiveness and potency of keyword query search [6]. They propose a brand new ranking formula using existing info retrieval techniques. Major importance of this two technique is works on massive scale real databases (E.g. commercial application that is customer Relationship Management) using 2 widespread RDBMS effectiveness and efficiency. It uses a Top-k join algorithm which incorporates 2 efficient query processing algorithms for ranking function.

II. Graph based Approaches

Graph based approaches assume that the info is modeled as a weighted graph wherever the weight of the edges indicate the importance of relationships. This weighted tree with edges is similar to steiner tree problem [5]. Graph base search techniques is a lot of general than schema primarily based techniques as well as XML, relative databases and web.[1]

A. BANKS: BANKS permits user to extract info in a simple manner without use of any of the database data of schema [2]. A user will get info by writing a couple of keyword, following hyperlinks and interacting with controls on the displayed results. BANKS algorithm is an efficient heuristics algorithm for locating and ranking query results. BANKS is concentrate on browsing and keyword searching. Keyword searching in BANKS is completed using proximity primarily based ranking on foreign key links. Model info could be a graph with the tuple as nodes and cross references between edges.

B. BLINKS: In query processing over graph-structured is a top-k keyword search query on a graph finds the top k answered in line with some ranking criteria. Before Schema primarily based approaches support keyword search over relational databases using execution of SQL commands [1]. These techniques are combination of vertices and edges the implementation of graph existing system have some drawbacks like poor worst case performance, not taking full benefits of indexes and high memory needs. To deal with this drawback BLINKS (Bi-level categorization and query processing) scheme for top k-keyword search in graph algorithm are enforced [4]. To cut back index space BLINKS partition a data graph into blocks. The bi-level index stores summery info at the block level.

3. Privacy and Keyword Search

To eliminate unnecessarily network traffic by not causing back the immaterial data, ranked keyword search is employed. This method is extremely desirable within the “pay-as-you-use” cloud paradigm. For privacy protection, such ranking operation shouldn't leak any keyword related data. to enhance the search result accuracy likewise on enhance the user searching experience, it's necessary for such ranking system to support multi-keyword search, as single keyword search typically yields far too coarse results (5). The data is retrieved from the matching files to calculate the relevance scores of given request.

If ranking system supports multiple keyword search then, it's possible to enhance the search result accuracy in addition as user searching experience will be increased. Altogether internet search engines, users give a collection of keywords rather than just one keyword to indicate that they're curious about a specific area. Every keywo rd within the user question is used to slim down the search process.

4. Multi-Keyword Algorithms

Ning Cao et al (6) established a set of strict privacy requirements for solving Privacy Preserving Multi-Keyword Ranked Search over Encrypted data in Cloud Computing (MRSE). They proposed a basic idea for the MRSE based on secure inner product computation. In their proposed technique “Coordinate Matching” is used as similarity measure to capture relevant data for a search query. The “Inner Product Similarity” scheme is used to quantitatively evaluate the similarity measure. The algorithm is practical, flexible and has low overhead
The author proposed a lightweight search multi-keyword search is proposed by Yanzhi et al. documents are present. So, a support for the efficient search or become inefficient when a large amount of data search that either focus on single keyword or multiple keywords. There are various approaches on encrypted cloud service provider (CSP) and the third party user. Thus, by hiding the user's identity, the cloud service provider as well as the third party user. A unique ID. This user ID is kept hidden from the cloud user a unique ID. This user ID is kept hidden from the cloud service provider as well as the third party user. Thus, by hiding the user's identity, the confidentiality of user’s data is maintained.

There are various approaches on encrypted cloud data search that either focus on single keyword search or become inefficient when a large amount of documents are present. So, a support for the efficient multi-keyword search is proposed by Yanzhi et al. The author proposed a light weight search that supports efficient multi-keyword ranked search in cloud computing system. First a basic scheme using polynomial function is used to hide the encrypted keyword and search patterns for efficient multi-keyword ranked search. To enhance the search privacy, the authors proposed a privacy preserving scheme which utilizes the secure inner product method for protecting the privacy of the searched multi-keywords. In this work, the authors encrypted the keywords and then constructed a polynomial function to hide them in search trapdoor generation. The detailed scheme to achieve the ranked multi-keyword search over encrypted data are performed using the polynomial functions which provides guaranteed privacy over two threat models. Through the proposed scheme, the authors conducted extensive experiments based on the real-world dataset. The experimental results demonstrated that the scheme proposed by Yanzhi et al. (11) enables the encrypted multi-keyword ranked search service is highly efficient in cloud computing. C. Wang et al. (3) proposed a definition for Ranked searchable encryption scheme to hide the keywords in the search query using polynomial functions which provides guaranteed privacy over two threat models.

Shieba et al. (8) solves the challenging problem of privacy preserving MRSE over encrypted cloud data based on secure inner product computation and efficient similarity measure of coordinate matching, i.e., as many matches as possible in order to capture the relevancy of data documents to the search query. The authors proposed significantly improved MRSE scheme to achieve various privacy requirements in two different threat models. An algorithm for anonymous sharing of private data among N parties is developed. This technique is used iteratively to assign these nodes ID numbers ranging from 1 to N. This assignment is anonymous in that the identities received are unknown to the members of the group. In the proposed system, stringent privacy is provided by assigning the cloud user a unique ID. This user ID is kept hidden from the cloud service provider as well as the third party user in order to protect the user’s data on cloud from the cloud service provider (CSP) and the third party user. Thus, by hiding the user’s identity, the confidentiality of user’s data is maintained.

5. Conclusion

Keyword search algorithm is the vast field and has its implications in various related field. In this paper we have tried to cover an overall concept of keyword search and its sub-types in future scope we will try to use the concept of the keyword search and will try to refine and design a new algorithm for the usage of searching in the unstructured documents.

6. References


