CONCEPT OF KEYWORD SEARCH: A REVIEW

Prerna Chopra[1], Vidisha Agrawal[2], Nidhi Sharma[3], Yusera Farooq Khan[2]
[1,4]M.Tech Scholar, [2,3]Asst. Professor
Computer Science and Engineering Department, NIMS University Rajasthan, India.

Abstract: Keyword search is the technique use for the recovering data or information. In Information Retrieval, keyword search is a sort of search strategy that looks for coordinating documents which contain at least one keywords specified by a user. A keyword search scheme to relational database becomes an interesting region of research system inside the IR and relational database system. The assumption and investigation of user search goals can be extremely significant in enhancing search engine importance and user encounter. The user tries to search about any query on the web, Search engine gives many numbers of result identified with that query. These results can be rely on upon metadata or on full ordering, because of this, user need to spend a considerable measure of time in finding the information of his interest. Along these lines, in venture induced user search goals by investigating search engine query logs. System use a structure to discover distinctive user search goals for a query by clustering the propose input sessions.
Index Terms: Search Engine, Keyword Search, Information Retrieval.

I. INTRODUCTION

Demonstrating user browsing conduct is a dynamic research range with unmistakable genuine word application. In web search applications, queries are submitted to search engines to represent the information needs of users. Sometimes queries may not precisely represent user's specific information. It might happen diverse users might need to get information on various aspects when they submit the same query. Keyword search is the most famous information discovery technique because the user does not have to know either a query dialect or the fundamental structure of the data. Huge number of techniques are used in Information Retrieval (IR) system. Keyword search is the technique use for the recovering data or information. Keyword search can be execute on both structured and semi-structured databases, also it possible on diagram structure which combines relational, HTML and XML data. Keyword search use number of techniques and algorithm for storing and recovering data, less exactness, does not giving a right answer, require vast time for searching and substantial measure of storage space for data storage. Data mining or information recovery is the process to recover data from dataset and transform it to user in displaying user browsing conduct is a dynamic research zone with unmistakable genuine word application. In web search applications, queries are submitted to search engines to represent the information needs of users. Sometimes queries may not precisely represent user's specific information. It might happen distinctive users might need to get information on various

aspects when they submit the same query. Keyword search is the most famous information discovery technique because the user does not have to know either a query dialect or the basic structure of the data. Expansive number of techniques are used in Information Retrieval (IR) system. Keyword search is the technique use for the recovering data or information. Keyword search can be actualize on both structured and semi-structured databases, also it possible on chart structure which combines relational, HTML and XML data. Keyword search use number of techniques and algorithm for storing and recovering data, less precision, does not giving a right answer, require substantial time for searching and huge measure of storage space for data storage.

II. KEYWORD SEARCH TECHNIQUES

A. Ranked Keyword Search:

Ranked search significantly enhances system usability by restoring the coordinating files in a ranked request with respect to certain importance criteria (eg. keyword recurrence) thus, making one step closer toward useful organization of security preserving data hosting services with regards to distributed computing. To the best of information it gives a legitimate status surprisingly the issue of viable ranked keyword search over encrypted cloud data. Ranked keyword search strongly provides system usability by restoring the coordinating files in ranked request worried to certain significance criteria, thus moving close towards the down to earth activity of security preserving data presenting services in distributed computing. To accomplish design goals investigate the statistical measure come nearer from Information recovery (IR) and evacuation to insert importance score of each record amid the establishment of searchable file before outsourcing the encrypted document accumulation. An IR system allocates a significance score to every last document and ranks those documents by this score. Significance score is used to manufacture a secure searchable list to legitimately ensure the sensitive information. This technique enables data users to locate the most related information quickly, instead of burdensome sorting through each match in the substance gathering. Ranked search can also carefully dispose of unnecessary network movement by sending back just the most important data. For security assurance, such ranking capacity, be that as it may, should not release any keyword significant information. Another, to enhance search result exactness as well as improve user searching background, it is also essential for such ranking system to support multiple keywords search.

B. Plaintext Fuzzy Keyword Search:

The significance of fuzzy search has gotten consideration with regards to plaintext searching in information recovery group. The issue is addressed in the customary information access worldview by enabling user to search without using attempt and-see approach for finding applicable information based on inexact string coordinating. At the first look, it seems possible for one to specifically apply these string coordinating algorithms to the setting of searchable encryption by registering the tGrapdoors on a character base inside a letter set. This trifling construction suffers from the word reference and statistics attacks and fails to accomplish the search protection.

C. Boolean Keyword Search:

Boolean systems enabled customers to specify their information require using a blend of Boolean operators AND, OR and NOT. Boolean systems have several disadvantages, for instance there are no any features of document ranking, and it is extremely troublesome for a customer to make a decent search request. Thus, the disadvantage of existing system specifies the essential requirement for new techniques that support searching adaptability.

D. Schema based approaches:

Schema based approaches support keyword search over relational databases using execution of SQL commands [1]. These techniques are blend of vertices and edges including tuples and keys (essential and remote key). There are some techniques are existed for schema based approaches

DISCOVER:

DISCOVER is the techniques that multiple Information Retrieval approaches take after. DISCOVER allows its user to issue keyword queries with no learning of the databases schema or SQL[2].DISCOVER returns qualified joining network of tuples, which is set of tuples that are associated because they join on their essential and remote keys, on the whole contain every one of the keywords of the query. DISCOVER uses static advancement. In future, it applies on unique streamlining. DIS-COVER returns a monotonic score conglomeration work for ranking a result. S ii. SPARK: With the increasing of the data stored in there are increase an interest for RDBMS to support keyword query search on data. For the same existing keyword search technique cannot satisfy the necessity of data search. This techniques focus on effectiveness and proficiency of keyword query search [6]. They propose another ranking recipe using existing information recovery techniques. Significant significance of this 2 technique is works on expansive scale genuine databases (Eg. Business application which is Customer Relationship Management) using two well-known RDBMS effectiveness and Efficiency. It uses a Top-k Join algorithm which includes two effective query processing algorithms for ranking capacity.

E. Graph Based Approaches

Graph based approaches assume that the database is demonstrated as a weighted graph where the heaviness of the edges show the significance of relationships. This weighted tree with edges is identified with steiner tree issue [5]. Graph base search techniques is more broad than schema based techniques including XML, relational databases and internet.[1]

[i] BANKS: BANKS enables user to correct information in a simple way with no learning of schema [2]. A user can get information by writing a couple keyword, taking after hyperlinks and interfacing with controls on the displayed results. BANKS algorithm is a productive heuristics algorithm for finding and ranking query results. BANKS is focus on browsing and keyword searching. Keyword searching in BANKS is finished using closeness based ranking on outside key links. Demonstrate database is a graph with the tuple as nodes and cross references between edges.

[ii] BLINKS: In query processing over graph-structured is a top-k keyword search query on a graph finds the top answered by some ranking criteria. Prior to the execution of graph existing system have some drawbacks like poor worst case execution, not taking full advantages of indexes and high memory requirements. To address this issue BLINKS (Bi-level ordering and query processing) scheme for top k-keyword search in graph algorithm will be actualized [4] . To reduce file space BLINKS partition a data graph into blocks. The bi-level file stores summery information at the block level.

III. AREA OF KEYWORD SEARCH

Information retrieval (IR) systems were firstly created to help deal with the colossal measure of information. Numerous universities, corporate, and open Libraries now use IR systems to give access to books, journals, and different documents. Information retrieval is used today in numerous applications [7]. General applications of information retrieval system are as follows:

Digital Library

A digital library is a library in which collections are stored in digital formats and accessible by computers. The digital substance might be stored locally, or accessed remotely through PC networks. A digital library is a kind of information retrieval system [7].

Search Engines

A search engine is a standout amongst the most the handy applications of information retrieval techniques to extensive scale content collections. Web search engines are best-known examples, yet numerous others searches exist, similar to: Desktop search, Enterprise search, Federated search, Mobile search, and Social search [7].

Media Search

A picture retrieval system is a PC system for browsing, searching and recovering images from a vast database of digital images [7].

IV. CONCLUSION

Finally we reason that, information retrieval is a process of searching and recovering the learning based information from gathering of documents. This REVIEW has managed the basics of the information retrieval. In first section we are characterizing the information retrieval system with their basic measurements. After this we concerns with customary IR models and furthermore discuss about the distinctive ordering techniques and searching techniques. This paper also includes the range of IR applications.

REFERENCES

[1] A. Baid, I. Rae, J. Li, A. Doan, and J. Naughton ―Toward Scalable Keyword Search over Relational Data,‖ Proceedings of the VLDB En-dowment, vol. 3, no. 1, pp. 140–149, 2010.

[2] G. Bhalotia, A. Hulgeri, C. Nakhe, S. Chakrabarti, and S. Sudarshan, ―Keyword Searching and Browsing in Databases using BANKS,‖ in Proceedings of the 18th International Conference on Data Engineering, ser. ICDE'02, February 2002, pp. 431–440.

[3] B. B. Dalvi, M. Kshirsagar, and S. Sudarshan ―Keyword Search on External Memory Data Graphs,‖ Proceedings of the VLDB Endowment, vol. 1, no. 1, pp. 1189–1204, 2008.

[4] J. X. Yu, L. Qin, and L. Chang, Keyword d Browsing in Databases using BANKS,‖ in Proceedings of the 18th International Conference on Data Engineering ser. ICDE COMPUSOFT, An international journal of advanced computer technology, 4 (6), June-2015 (Volume-IV, Issue-VI)

[5] J. Coffman and A.C. Weaver, ―A Framework for Evaluating Database Keyword Search Strategies,‖ in Proceedings of the 19th ACM Inter-national Conference on Information and Knowledge Management, ser. CIKM'10, October 2010, pp. Search in Databases, 1st ed.Morgan and Claypool Publishers, 2010.

[6] Y. Luo, X. Lin, W. Wang, and X. Zhou, ―SPARK: Top-k Keyword Query in Relational Databases,‖ in Proceedings of the 2007 ACM SIGMOD International Conference on Management of Data, ser. SIGMOD '07, June 2007, pp.115–126.

[7] S. Cohen, J. Mamou, Y. Kanza, and Y. Sagiv. XSEarch: A semantic search engine for XML. In VLDB, 2011.[8] W.Webber, ―Evaluating the Effectiveness of Keyword Search,‖ IEEE Data Engineering Bulletin, vol. 33, no. 1,pp. 54–59, 2010.

[8] Xiang-Yang Li and Taeho Jung , ―Search Me If You Can: Privacy-preserving Location Query Service‖, National Natural Science Founda-tion of China, arXiv:1208.0107v3 [cs.CR] 11 Apr 2013.

[9] Shiwen Cheng, Arash Termehchy, and Vagelis Hristidis, ―Efficient Pre-diction of Difficult Keyword Queries over Databases‖, IEEE Trans. Knowledge and Data Engineering., June 2014, ISSN :1041-4347.

[10] Reshma Sawant, Akshaya Deshmane, Shweta Sawant,―Personalization of Search Engines for Mobiles‖,International Journal of Advanced Engineering & Innovative Technology, Vol 1, Issue 1, April-2014, 24-29 ISSN: 2348-7208

[11] W. Webber, ―Evaluating the Effectiveness of Keyword Search,‖ IEEE Data Engineering Bulletin, vol. 33, no. 1,pp. 54–59, 2010.

[12] Zheng Lu, Student Member, IEEE, Hongyuan Zha, Xiaokang Yang, Senior Member, IEEE, Weiyao Lin,Member, IEEE, and Zhaohui Zheng A New Algorithm for Inferring User Search Goals with Feedback Sessions VOL. 25, NO. 3, MARCH 2013.