ISSN: 0974-5823 Vol. 7 No. 2 February, 2022

International Journal of Mechanical Engineering

AN EFFICIENT INTEGRATED APPROACH FOR CLOUD INFORMATION RETRIEVAL USING N-LEVEL VECTOR MODEL

Dr. V. Esther Jyothi, Assistant Professor, Department of Computer Applications, Velagapudi Ramakrishna Siddhartha Engineering College, Vijayawada.

Dr. V. Madhu Latha, Assistant Professor, Department of Business Management, Velagapudi Ramakrishna Siddhartha Engineering College, Vijayawada.

Prof. D.S.B.Bharathi, PG Department of management studies, The Hindu College-MBA, Machilipatnam.

Mrs. A.Madhuri, Assistant Professor, Department of Computer Science and Engineering, PVP Siddhartha Institute of technology, Vijayawada.

ABSTRACT: Nowadays most of the aspects are involved in the techniques of cloud computing by which the information technology is reshaped. Information retrieval technology application and its development is becoming inevitable because of the internet, networks digital information rapid growth. The model of cloud retrieval system developed with current technology by experts according to the mobile search engine and search engine study. Minimum cost and maximize profit are achieved with the help of Cloud computing technology by different possible ways of shared Cloud resource. An efficient integrated approach for cloud information retrieval using N-level vector model (NVM) in Cloud computing environment framework is analyzed in this paper. This cloud information retrieval system includes with many layers as cloud retrieval cluster system, cloud information layer, residual module, the user query analysis and NVM based ranking. Mass information retrieval is having low information search precision when it uses the vector space model because poor similarity values are acquired by the long documents so in the vector space model poorly represented. Results are indicating that, if document size is increased then which leads to increment in the relevancy ranking scheme costing part. Cloud information retrieval system is more efficient with N-level vector model (NVM) and having large precision and recall values.

KEYWORDS: —cloud retrieval system, N-level vector model (NVM), effective ranking, precision, recall ratio.

I. INTRODUCTION

In internet computing resources are delivered by the Cloud computing. Most of the applications are adopted by the Cloud computing network and becomes a more persistent system [1]. The main reasons for
adopting cloud computing are social network data exponential growth, modern data centers, sudden increment in computing data storage and power. Some of them are having low utilization and high costs so these to be reduced by the designing of cost-effective and reliable cloud-based systems [2].

Internet based IT services are modeled and the current consumption supplement by the new way of Cloud computing. Regular virtualization and dynamically scalable are provided as a service [3]. Internet providers are offering many online services and these are benefit for the users when it are used. Consequently the resources are required to be saved in computing and management, data storage [4]. Specific services are provided by the service providers to offering client software and hardware at lower cost. Huge amount of data is also stored on remote servers because computing resources and storage have aggregating trend.

Computing resources (hardware and software) are available in the remote locations also, and these used by the Cloud computing for accessing the entire network. In future expanding techniques are developed for reshaping the information technology processes with cloud computing. Most of the companies are uses this cloud computing technology to share cloud data because of its more advantages as scalability, flexibility and cost-effectiveness [5]. Cloud services are subscribed by the organization in the conventional cloud application in order to share the documents or information between staff in the cloud. Set of keywords are describing every file and staff or authorized users are having the authority to retrieve the file from cloud with certain keywords by querying the cloud [6].

Because of its adaptability and simplification, the retrieval of information is widely uses the Vector space model [7]. Boolean model based Information retrieval system is associated with addressing and without order ranking problem [8]. For the information retrieval the vector space model is used and from this, the search precision of information is low because poor similarities values are acquired by the long documents so poorly represented and in these spontaneous weightings are accepted but in the vector space model not formal. The order of documents in the form of terms in the vector space model is lost.

This problem is efficiently eliminated by the designing of proposed relevancy ranking scheme based on N-level vector model which explores a new concept for the term weighting and the feature term location in the document is also taken into account. Then the document content is properly described and encrypted documents ranking process is investigated by using the proposed scheme [9]. The better retrieving result is achieved by this newly proposed formula and is given for information retrieving. In this method term weighting new formula introduced and the feature term location is also described in the document and its contents are described briefly.

II. CLOUD INFORMATION RETRIEVAL

Cloud information retrieval major contributions are summarized as below:

• The framework of cloud information retrieval and its risks are defined formally. Effective keyword retrieval problem over high users encrypted cloud data retrieval performance [10] and privacy is studied first according to researcher's knowledge.

• Encryption scheme is designed for supporting the information retrieval methods [11]. The integration of cloud information retrieval framework to state-of-the-art retrieval models are illustrated in addition with complexity of communication is also discussed for the retrieval models in CIR.

In the form of septuplet cloud information retrieval (CIR) is mathematically represented as
{S, D, K, C, Q, F, R (Q, ci)}, where:

• Privacy protecting and security policies are denoted by 'S'.
• In the documents, representations are denoted by 'D'.
• In every document all keywords union set, searchable keyword field is denoted by 'K'.
• In the collection D, for the documents searchable encrypted representation set is denoted by 'C'.
• Queries or user information are required to logical representation set as 'Q'.
• Modeling document representations, relationships and quires framework is as 'F'.
• The ranking function is denoted by R (Q, ci) which can calculate among C, the order of ci with regards to the query Q [12]. Privacy and security protecting policy (S) [13] is required by brief information need (Q).

Cloud computing and box Computing are the two popularized search engines. Internet next generation main Momentum is retrieving the cloud information. According to the box computing and cloud computing, the cloud information is retrieved. By the terminal unit only the users are connect in the network. Users required information uses the terminal unit as input terminal. Then the cloud layers can be retrieved automatically by the cloud retrieval system quickly and gives a feedback to the users regarding to their service.

Random mapping function and binary vector are essential for the Bloom filter-based security indexing method [14]. Security can be improved by the Chang et al based on Goh's algorithm and a semi-trusted third party is introduced in multi-server for mediating and coordinating the queries. Inverse hash sequence is generated first by encryption keys in each encryption of their algorithms. After this, bloom filters are takes the encrypted index. Trap-door mechanism is designed for the inverse hash sequence for retrieving and bloom detection. Finally, the returned cipher texts are decrypted for require documents retrieving. Initial searches security is guaranteed by this system. One by one linear faster search operations are conducted in the Bloom filter but if the encrypted information size is too large, this system is insufficient for searching. Inverted index building eliminates the above problem of general information retrieval [15].

By generating number of key sequences, statistical attacks problem is also solved by this method but gives the result as more complexity in the construction then by increasing the data retrieval time. Hence it is clear that, for massive data retrieval process is not to be implemented by this method and it is not practically suitable.

From above stated methods are explains that more relevant documents are not mentioned even the documents are retrieved. Documents with large data cannot be handled properly with these methods. So mass encrypted information retrieval is cannot used directly over multimedia networks. In encrypted documents ranking problem reduction is urgent and prominent. Improving the retrieval precision and retrieval recall demand is increased with information size increment unprecedentedly. Better information retrieval is achieved by the ranking the possible documents.

III. CLOUD INFORMATION RETRIEVAL USING N-LEVEL VECTOR MODEL

The process of retrieving the cloud files may include with many layers such as cloud retrieval cluster system, cloud information layer, user query and its analysis layer, residual module and NVM based ranking. Again the cloud retrieval cluster system is also consisted with many layers as cloud processing, cloud collecting, cloud index, cloud interface, cloud query, cloud data storage layers with cloud management and scheduling system, and cloud retrieval monitoring system. An efficient integrated approach for cloud information retrieval system using N-level vector model (NVM) in Cloud computing environment is analyzed in this paper and its framework is as shown in below Fig. 1.

The information is collected from the Cloud Information Layer by the network robots with Cloud Acquisition Layer. Search performance can be improved by the processes of running set for robot, which takes the parallel processing mode that means collecting the web information work can be completed by the multiple robots on multiple computers distribution. A coordinator is to be designed for coordinating the scheduling task reasonably in order to work together with multiple robots. Batch processing

mode is allotted to the robots, URL is temporarily stored by these robots and which is sent to robot exchange datum and coordinator when certain number is throw up by the storage. The coordinator will get the store URL from robots and gives back to the new URL to the robots in this communication. In the storage clouds the search results are stored by the coordination.

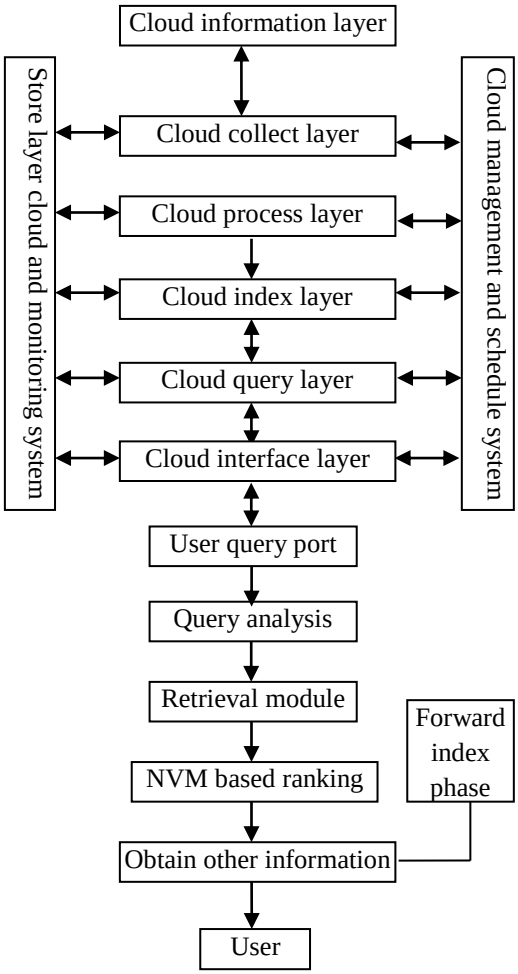

Fig.1: FRAME WORK OF 'NVM' BASED CLOUD INFORMATION RETRIEVAL

3.1 Cloud collection layer:

The information is first collected by the cloud collecting layer work process on the internet and this collected information is processed for the extraction of URL information. This extraction can be used by the URL information to store in the URL original repository which helps to transcode the web documents. Based on the store and design in the database, the final data format is obtained as the web documents. The cloud collecting layer main functions are given as: the web page files encoding conversion, HTTP protocol Simulation achievement, the URL Rank algorithm achievement, robots crawling state preservation, URL algorithms extraction, multi-threaded robot achievement, the original data file storage format definition, the robot program flow control and so on.

3.2 Cloud process layer:

The collected information will be cured by the Cloud processing layer first stage as the network link duplicates, useless information and control codes are must be removed by the filtering. Text classification program is used by the cloud retrieval system in order to achieve the fast information retrieval in the page of search then the retrieval efficiency can be improved in the processing layer. Entry sequence documents are obtained from the word processing and the received raw information is collected by the word program participle which is used in the certain information model.

3.3 Cloud index layer:

Inverted index technology is used by the Cloud index layer in cloud retrieval completion. The index entry operation compiling of original document entry sequence can be obtained from the indexing process by cloud retrieval in the previous step and, is raw data index table generation, file of scanned entry sequence, data format complete conversion for accommodating the input requirements main index. The response speed is must be improved by placing the index in the memory queue. The storage space can be reduced by both operations of index compression and index page sorting. The data amount cannot be expanded directly to the index then in that case multiple or multi-level indexes are necessarily used. Finally in distributed storage clouds all the retrieved information is required to be stored.

3.4 Cloud query layer:

The index tables are reading by the index scanner from the database of main index to the finder and then some weighing operations and set of operations are made by the finder and it sends to the multi-word processor for further weighing and its handling. At finally these are transfers to sorter for sorting which gives to the query front-end.

3.5 Cloud interface layer:

Through the cloud interface layer and queries Terminal, the required information can be passed by users to the cloud interface layer. Through cloud interface layer, the information is can undergo the searching, scanning and contrast treatment, and final results are sending to the users. Simultaneously monitoring systems and storage stratus are monitoring the cloud management systems, all the processes and cloud dispatching system is responsible for handling exceptions. All processes are enabled by the role of operations which were different and used in completion of task efficiently.

3.6 NVM-based information retrieval system:

Three components are present in the information retrieval system and these are, the index module used for the generation of index, search module used for encrypted information search and user interface module. The output interfaces and input interfaces are provided for the process of information retrieval in the user interface module. Search engine, User interface, NVM-based ranking and Index database are the main elements in the information retrieval system. For the submission of queries, user interface enabled by the user and the response is provided to the user queries. The retrieval of information stores documents in the index database after the search engine move slowly, discovers and transforms the data. Index establishment text analysis and text extraction operations are conducted in the index database. Based on ranking scores of NVM model is realized by the system of encrypted information retrieval. Query parsing, language processing, lexical analysis, ranking the retrieval results and searching index are the main steps in the process.

3.7 Index module

The collected information from the collector index was established by the index module and for the purpose of query, the collected information is stored in the index database. From page-to-page correlations, the analysis of information acquisition and tabular form of resulting documents were done. These documents description reorganized and important data items inverted list has build by the index and which are ready for the user's search, and this was the main aim. A relevancy ranking scheme of N-level vector model is adopts the index module.

IV. RESULTS

The cloud information retrieval system based on NVM-model is evaluated by the verification of effectiveness of NVM-based retrieval system. Information retrieval performance from huge document collection can by analyzed by the two main parameters as retrieval precision and retrieval recall.

- Recall = Number of retrieved relevant documents / Number of all relevant documents

Given that in the network environment, information is a dynamic process (e.g. updating, deleting, adding, etc), so the retrieve recall measuring is not easy. The search engine relative retrieval recall is estimated by satisfactory practice as enhancement to the recall. Maneuverable is subjected by the information relative recall.

- Precision = Number of retrieved relevant documents / Number of all retrieved documents

Here from the above expansions of precision and recall, these two are having inverse relation. After reaching the threshold point of recall must be reduce the precision value, and conversely high precision value which gives the result as low recall.

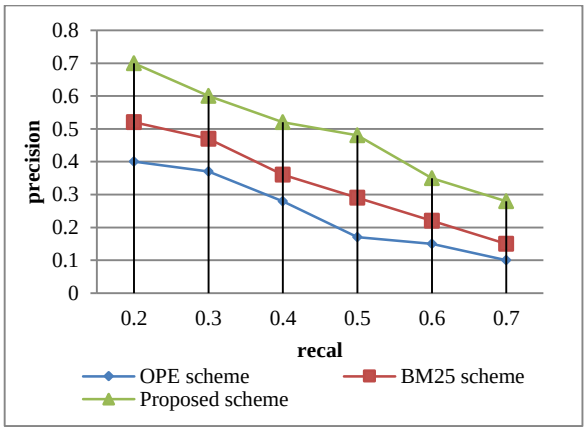

Fig. 2: COMPARISION BETWEEN RECALL AND PRECISION IN VARIOUS SCHEMES

From above Fig. 2, states that precision and recall (P-R) curves of order preserving encryption-based (OPE) is having very low precision and recall ratio when compared to both BM25 scheme and as well as proposed scheme therefore the P-R curve is lower than proposed and BM25 schemes. The OPE scheme is having less effectiveness because of only the consideration of term frequency usage in the encryption scheme. Among these three schemes, the proposed relevancy ranking scheme of N-Level vector model is having the great retrieval recall and retrieval precision ratio than both BM25 scheme and OPE scheme. So the encrypted

documents are transmitted to entire network effectively by using relevancy ranking scheme based on N-Level vector model.

Effect of document size

The performance of information retrieval systems are evaluated by one of the technical parameters as response time. This evaluation process is includes with the two aspects such as, entering into the search engine waiting time and waiting time for getting query results. An important role is played by the response time in performance evaluation of the information retrieval system. The encrypted information retrieval system based on NVM model effectiveness is verified by measuring the index building time in experiment. The index building test results are depicted in Table 1, including the document size, index building time and file type.

Table 1: TEST RESULTS OF INDEX BUILDING

File type	Document size (Byte)	Index building time {ms)
Text	15124	942
Text	161278	1542
Doc	135546	1256
Doc	126745	1259
Pdf	961872	1089

Percentage of Returned Files: from 0 to 4 ranks are classified in the users as Rank-0, Rank-1, Rank-2, Rank-3, and Rank-4 and they are retrieving file percentages as 100%, 75%, 50%, 25% and 0% respectively. Differential query services are provided by the EIRQ system with matched files 100%, 75%, 50%, 25%, 0% retrieved by Rank-0, Rank-1, Rank-2, Rank-3, and Rank-4, and this is explained with the help of bar graph in the Fig. 3.

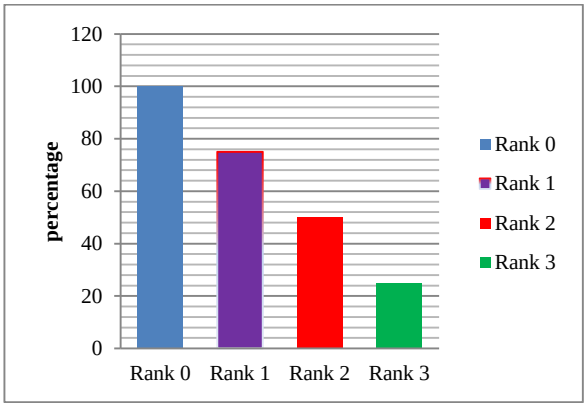

Fig. 3: RETURNED FILE PERCENTAGE

V. CONCLUSION

An efficient integrated approach for cloud information retrieval using N-level vector model (NVM) in Cloud computing environment framework is analyzed in this paper. Encrypted information can be retrieved by using relevancy ranking scheme based on proposed N-level vector model in the network. The process of retrieving the cloud files include with many layers such as cloud retrieval cluster system, cloud information layer, user query and its analysis layer, residual module and NVM based ranking. From the experimental results, it is clear that if document size is increased then the index building time is also increased. The proposed information retrieval system requires the long index building for which number of operations is required to establishment of index. Cloud information retrieval system based on NVM model is having effective performance cloud documents transmission ranking with high recall and precision values. Different rank users are retrieving the files with different percentages by using the proposed scheme and match it to the quires which make the cloud services more flexible and scalable.

VI. REFERENCES

[1] Xuemin Shen, Chuntao Ding, Shangguang Wang, Xiulong Liu, Ning Zhang, Jiannong Cao, Ao Zhou, "A cloud-Guided Feature Extraction Approach for Image Retrieval in Mobile Edge computing", IEEE Trans. on Mobile comp., Vol.: 20, Iss.: 2, Year: 2021

[2] Vibhutikumar G. Joshi, Hiteshkumar J. Lad, Hiral M Joshi, "cloud-based Portable and Cost-effective Particulate Matters Concentration Estimation system", 2020 4th International Conf. on Electronics, Comm. and Aerospace Tech. (ICECA), Year: 2020

[3] Yuko Ikkatai, Eiri Ono, "Internet-based services to obtain information on science and technology according to the degree of interest", 2020 9th International Congress on Advanced Applied Infor. (IIAI-AAI), Year: 2020

[4] Wenchun He, Yufeng Chen, Ming Yang, Yongjun Xu, Zhiqiang Zhang, Xiaolong Xu, "An Efficient storage Service Method for Multidimensional Meteorological data in Cloud Environment", 2019 International Conf. on Iot (iThings) and IEEE Green computing and Comm. (GreenCom) and IEEE Cyber, Physical and Social ompcu. (CPSCom) and IEEE Smart data (SmartData), Year: 2019

[5] Jantima Polpinij, Sirinun Karnbunjong, Khanista Namee, "The Integration of File Server Function and Task Management Function to Replace Web Application on Cloud Platform for Cost Reduction", 2019 IEEE Asia Pacific Conf. on Circ. and Sys. (APCCAS), Year: 2019
[6] Drishti Gugal, Manish Kumar Yadav, Sanket Waghmare, Shivani Matkar, "Encrypted keyword Search in Cloud Computing using Fuzzy Logic", 2019 1st International Conf. on Innovations in Information and Comm. Tech. (ICIICT), Year: 2019
[7] Ajit Ku. Nayak, Mamata Nayak, Upasana Parida, " Ranking of Odia Text Document Relevant to User Query UsingVector space model", 2019 International Conf. on Applied ML (ICAML), Year: 2019
[8] Anestis Koutsoudis, George Ioannakis, Christodoulos Chamzas, Ioannis Pratikakis, "RETRIEVAL —An Online Performance Evaluation Tool for Information-retrieval method", IEEE Trans. on Multimedia, Vol.: 20, Iss.: 1, Year: 2018
[9] Xingming Sun, Xinhui Wang, Qian Wang, Zhihua Xia, "A Secure and Dynamic Multi-Keyword Ranked Search Scheme over Encrypted Cloud Data", IEEE Trans. on Parallel and Distri. Sys., Vol.: 27, Iss.: 2,Year: 2016
[10] Abhijeet Manepatil, Sanket S. Pawar, Prajakta Jagtap, Aniket Kadam, "Keyword search in information retrieval and relational database system: Two class view", 2016 International Conf. on Electrical, Electronics, and Optimi. Techni. (ICEEOT), Year: 2016
[11] Zhao-Li Wang, Jing-Li Han, Jian Liu, "Searchable Encryption-Scheme on the Cloud via Fully Homomorphic Encryotion", 2016 Sixth International Conf. on Instrumentation & Measurement, Computer, Comm. and Control (IMCCC), Year: 2016
[12] A. K. Saxena, Ashish Saini, Yogesh Gupta, "A comparative analysis of fuzzy based ranking functions for information retrieval", 2016 3rd International Conf. on Compu. for Sustainable Global Devel. (INDIACom), Year: 2016
[13] Hai Zhuge, Xiaoping Sun, Li Zhang, "Density Based Collective Spatial Keyword Query", 2012 Eighth International Conf. on Semantics, Knowl. and Grids, Year: 2012
[14] Yang Ying-Hui, Li Jian-Hua, Xiong Jin-Shi, "Method of Determine Index Weight in Security Risk Evaluation based on Information Entropy", 2012 Fourth International Conf. on Multimedia Information Networking and Sec., Year: 2012
[15] Xuechang Chen, Chi Ma, Qiuying Bai, "A new index model based on inverted-index ", 2012 IEEE International Conf. on Computer Sci. and Automation Engg. Year: 2012