Journal of Theoretical and Applied Information Technology
15th April 2024. Vol.102. No 7

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

CROSS CENTROID MERKLE SEARCH WITH LINGUAL-MULTI KEYWORDS OVER ENCRYPTED CLOUD DATA

[1*]DIVYA SURENDRAN AND [2]DR. K. SASIKALA

[1*]Research Scholar, Department of Computer Science,
Vinayaka Missions Kirupananda Variyar Engineering College.
Vinayaka Missions Research Foundation, Salem, India
[2]Professor & Head, Department of Information Technology,
R.P Sarathy Institute of Technology, Salem, India - 636305
Corresponding Author Email ID:[1*]surendrdivya@gmail.com

ABSTRACT

One of the most significant services provided by cloud computing is now cloud storage, from which its consumers have reaped numerous benefits. Users can easily use public-key encoding and keyword searching to search secured data and retrieve desired data from cloud storage. However, there were some problems with speed, accuracy, and security when searching for encrypted keywords. The cross-lingual multi-keyword Centroid Merkle search over encrypted data (CLCMSE) suggested in this study is based on the Open Multilingual Wordnet and is meant to solve this problem. The proposed CLCMSE technique allows data users to query in any language and select the linguistic kind of information provided. First, the Centroid technique is used to cluster data from the cloud and sort the clustered data in this process. Then the Merkle search technique is used to increase the search speed. Finally, the targeted source data is retrieved from the cloud by using the fuzzy information retrieval algorithm. This experimental result proved that our proposed CLCMSE has higher accuracy, security, and speed performance than existing methods. This study compares the multi-keyword rank search over encrypted cloud data, the multi-keyword rank searchable encryption, and the verifiable attribute multi-keyword search.

Keywords: *Cross-Lingual Multi-Keyword, Centroid, Merkle Search, Cloud Storage, Word Net, Encryption, Data Retrieval.*

1. INTRODUCTION

The semantic relationship between words is used by several well-known and trustworthy semantic searching algorithms to broaden their searches on plain text. Precise matching between the keywords in external files is achieved by utilising both query terms and enlarged semantically related words. Reliable semantic searching has three keyword searching schemes: synonym, mutual information model, and idea hierarchy[1]. Cloud platforms have grown in popularity as internet technology has evolved due to their massive storage and processing capabilities. By uploading data to the cloud system, users from various physical places can share resources[2].

Individuals and businesses are increasingly turning to cloud computing to outsource their access to multiple services. Cloud computing services can use individuals and companies to transfer their information to a shared cloud server, which relieves cloud clients of the burden of data storage and administration. Some of the outsourced data is extremely sensitive. In general, searching is hampered when users encrypt data traditionally [3][4]. For secure searches over encrypted cloud data to be conducted while protecting data privacy and confidentiality, searching encryption is essential. Though this is gaining much research attention, barrier security measures in a group-oriented cloud information-sharing environment are often not enabled by existing searchable encryption algorithms [5][6]. Searchable encryption methods were constructed in two-sample settings, including the symmetric essential configuration and the public key configuration, to use the searchable function [7]. The recipient's public key is used by the data administrator to encrypt the data, which is subsequently sent (i.e., ciphertext). The recipient can quickly get the encrypted data by downloading the ciphertexts that go with it from the cloud storage service. Then, they can decrypt

the whole set of encrypted business data and get the goal data locally [8].

Information retrieval and archival have grown in importance as multimedia data processing has advanced to evaluate real-time data. In this instance, searching is a well-known technique used on the Internet. Traditional and general data retrieval is based on keyword searches, which have drawbacks such as a high manpower demand and a reliance on personal information, resulting in poor simulation results. Secure data retrieval-based cloud has lately reached a peak concentration thanks to the implementation of CC. Businesses and consumers adopt the CC for storing and managing sensitive information, such as photographic collections and individual fitness information, because it offers greater convergence and financial savings. Until storing data in the cloud, data encryption is performed to ensure data confidentiality. As a result, traditional encryption makes basic data tasks difficult, such as the IR of encrypted data. In the case of ciphertext it is extremely difficult to reach successful information retrieval when securing customer information [9]. Without revealing the data content or search queries, a searchable interface allows a cloud server to recover keywords from encrypted data securely. The list of searchable encryption techniques is then expanded to include searchable symmetric encryption, public-key encryption with keyword search (PKES), and identity-based encryption with keyword search. However, if the information is exchanged with several data users, the data owner will need to run the encryption algorithm many times in these schemes [10]. This work presents a novel approach to the Open Multilingual Wordnet strategy based on the data searching and retrieval plan: the Cross-Lingual Multi-keyword Centroid Merkle Search over Encrypted Data (CLCMSE). The presented method involves clustering and searching-based keyword searches to retrieve the data of the input keyword. Centroid Merkle search over encrypted data (CLCMSE) scheme is discovered to be a successful cloud data retrieval method since it eliminates similar data with a low retrieving rate. It also offers a fast search and a high level of trustworthiness. The paper's remaining portion is structured as follows: The article's methodology is discussed in Section 3, while Section 2 discusses the literature. Section 4 shows the full system's method. Section 5 discusses exploratory observations. The sixth area concludes the exposition.

1.1 Background of the study

It is possible to configure and change a cloud system into a cloud storage system when it is primarily utilised for data management and storage (rather than computation and processing). Simply said, cloud storage is any form of online storage that users can access from any networkable device at any time, anywhere. Encryption technologies are frequently used to enhance the security of cloud storage because people are becoming more worried regarding data privacy and the fact that cloud servers are vulnerable to menaces from inside and outside of the system. The retrieval of ciphertext data may provide difficulties even if encryption methods can guarantee the privacy of data transferred to the cloud. In recent years, figuring out the ways to empower data searching while maintaining the privacy of sensitive data has been crucial.

1.2 Problem statement

The challenge of secure search over encrypted data has been a research focus for the past few years, is now better understood thanks to the development of searchable encryption technology. Nevertheless, the majority of search algorithms based on searchable encryption now in use only handle specific linguistic queries. The limited number of multilingual searchable encryption systems that have been implemented do not succeed in automatic cross-lingual retrieval, which has a substantial negative influence on the user's search experience. Furthermore, the majority of these systems only support precise matching and are unable to support semantic search. Thus, cross-lingual and semantic search implementation together is still up for discussion in searchable encryption. We are aware of no prior studies that have looked into the issue of cross-lingual ranked search over encrypted cloud data. Our solution, based on the Open Multilingual Wordnet, is a cross-lingual, multi-keyword Centroid Merkle search over encrypted data (CLCMSE). Language barriers can be overcome by our CLCMSE system, which also achieves intelligent and tailored search through adjustable keyword and language preference settings.

2.RELATED WORK

Ge et al. [11] introduced a ciphertext-policy attribute and data sharing (CP AB-KSDS) for encrypted cloud 12 data. Because an opponent with a 450 rating can select a keyword and create

an identical keyword 451 ciphertext, key secrecy captures the fact that the tokens containing keyword 449 cannot be covered. They suggested a concrete scheme and demonstrated that it was safe in the random oracle model against 20 chosen ciphertext and chosen keyword attacks. Finally, the performance and 23 property comparisons showed that the proposed construction was realistic and practical.

Liu et al.[12] presented a quick and accurate seven-searchable encryption (FASE) technique. Furthermore, the two 24 keyword matching degrees and the significance score return reliable search results and improve the search performance. A quick and precise multi-keyword ranking search was achieved by the FASE scheme, according to theoretical research and analytical findings. However, the FASE system still has a lot of issues. Nevertheless, they fail to develop a dynamic, searchable encryption system that can manage changing circumstances.

A secure multi-attribute linguistic keyword search technique was presented by Zhang et al. [13] over encrypted cloud data. This study carried out a sufficient number of experiments to test the proposed structure's efficiency. The presented approach outperformed was superior–linear search performance. The delivered systems are quite feasible, according to extensive analysis and experiment findings. However, it requires less computing power for searches, trapdoor creation, and configuration.

A verified top-k searchable encryption cloud data (VSED) for dynamic updating activities, including adding and removing documents, was presented by Elizabeth et al. [14] the analysis's findings indicated that the structure was effective based on both time and storage complexity. Even though the offered method was insecure and inefficient in the absence of a SCP, the system model that was given contained the Document Owner (DO), CSP, the Secure coprocessor (SCP), and the Data User (DU).

Miao et al. [15] provided a feasible attribute-based keyword search strategy enabling remote access policy in the shared multi-owner setting (ABKS-SM) system. In the general bilinear group model, the disclosed ABKS-SM systems repelled off-line keyword guessing assaults and achieved the required security. However, one drawback of the suggested ABKS-SM frameworks was that as the quantity of framework credits increased, the computational and capacity costs also increased.

Using a fuzzy information technique based on Lattice assumptions, Yang et al. [16] demonstrated a secure multimedia cloud with multi-user capability. The recommended scheme guarantees that a gathering of endorsers can look at scrambled interactive media information without sharing private keys. However, they do not focus on planning grid-based accessible encryption plans with additional adaptable inquiry designs, like boolean, range, and subset inquiry.

Shen et al. [17] suggested P3, a phrase search approach that preserves anonymity for secure encrypted data capture in cloud-based IoT. They used encryption algorithms and a bilinear chart to evaluate a position association of multiple approached keywords over encrypted data. The presented scheme's performance and efficiency were demonstrated in the experimental evaluation results.

A conjunctive multi-keyword ranked secure search approach was presented by Yin et al. [18] for different data owners. Comprehensive findings demonstrated the accuracy and applicability of the suggested approach.

Dai et al. [19] suggested a recently developed privacy-permanent searchable encryption technique based on the Latent Dirichlet Allocation (LDA) topic model. The proposed topic model accelerated the search while decreasing the size of the vectors. The cost of searching is further reduced by using a tree-based index. A precise, privacy-preserving, semantic-linguistic ranked search method over encrypted cloud data has been suggested.

A Multi-keyword ranking cloud data search method that efficiently supports dynamic operations was presented by Guo et al. [20]. To improve the performance of the search, an index tree for the connected details was built using the Bloom filter. Furthermore, their method better supports a document's numerous procedures that involve omissions or settings. The experiments revealed that their approach could efficiently and effectively meet the design goals.

The idea of searchable encryption was first presented by He et al. [21] and uses attribute-based network access to enable hybrid Boolean keyword search on externally encrypted content. More expressive searches, like any necessary logical keyword expression search, can be carried out by authorised individuals. Additionally, the security of the system was demonstrated by the application of their security model.

An effective role-based authorised keyword search strategy for controlling hierarchical access permissions was suggested by Y. Miao et al. [22]. This was expanded to enable effective user management and token creation preprocessing by using it as a building piece. It has been finally established by formal means that the suggested approach resists both CKA and IGKA. Empirical trials have been conducted to confirm the method's effectiveness and viability in real-world situations. Future developments will involve evaluating a real-world prototype of the proposed method and investigating more expressive and effective multi-keyword searches with hierarchical access control. Shen et al. [23] suggested a blockchain-based electronic medical record system with multi-keyword searchable encryption (BMSE). There are two sections to the plan. On the one hand, our method uses symmetric data encryption using the advanced encryption standard (AES) in conjunction with blockchain technology. We also encrypt the search index using attribute-based encryption (ABE). This method aims to solve the problem of centralised CSP's overbearing authority, which might compromise patient privacy. However, to address the issue of the low efficiency of the currently available multi-keyword searchable encryption methods, we employ the K-means algorithm to cluster the documents and use the relevance score of keywords and documents as the search index. Ultimately, we confirm the security of BMSE via safety analysis, and empirical research indicates that BMSE enhances search effectiveness.

By utilising Intel SGX, Liu et al. [24] present a dynamic multi-client fuzzy keyword search technique that may ensure forward privacy at the cost of multiple trapdoor communication. With the help of Intel SGX, the suggested approach can lessen client-side processing and communication overload. Furthermore, we present an enhanced multi-client fuzzy keyword search method that maintains forward privacy while the compromised user is present. Our methods are suitable for real-world applications and can offer the necessary security level, as demonstrated by the efficiency and security evaluation.

For blockchain-assisted cloud-edge storage, Liu et al. [25] presented a reliable and strong multi-keyword search (TRMS) that allows data users to select between a more thorough search based on cloud servers or a faster search based on edge servers. Our approach involves deploying a blockchain-based smart contract to run the search algorithm and update the score-based trust management model, thereby enabling the search for dependable servers and trustworthy search results. This will enable the recording and publication of trust scores and search results on the blockchain. Data consumers can determine whether the papers that are returned are top-k documents by perusing the search results.

An effective confirmed privacy-preserving Boolean range query with suppressed leaking was presented by Q. Tong et al. [26]. First, we use the Bloom filter and Grey code to transform BRQ into a multi-keyword query. Then, by integrating the distributed point function and PRP-based Cuckoo hashing, we accomplish effective oblivious multi-keyword inquiry while protecting the access and search patterns. Additionally, by employing aggregate MAC, keyed-hashing MAC, oblivious query, and XOR-homomorphic pseudorandom function, we provide lightweight and oblivious result verification. It allows query users to utilise a proof whose size is independent of the size of the outsourced dataset to confirm the accuracy of the results. Lastly, our suggested system is efficient and adaptively secure for real-world applications, as shown by thorough experiments and formal security analysis, respectively.

A multi-keyword searchable encryption system was suggested by Wang et al. [27] to increase the efficiency and security of trustworthy exchange of sensitive material. The plan creates a blockchain-based architecture for sharing private data, and the distributed ledgers' tamper-proof functionality guarantees the authenticity of encrypted data and indexes in addition to sharing behaviour monitoring. Based on this, we refined the inverted index structure to achieve effective multi-keyword searchable encryption and prevent keyword-pair result pattern leaking. The results of the simulation demonstrate the effectiveness of the multi-keyword searchable encryption technique, and the recommended fix has withstood a thorough security examination.

A technique known as MKSABE-VaAR (multi-keyword searchable attribute-based encryption with verification and attribute revocation) was introduced by Shen et al. [28]. We first bundle many keywords into a polynomial to enable multi-keyword search, addressing the issues of latency, unnecessary computation during the search process, and only permitting single-keyword search in most attribute-based searchable encryption techniques. This polynomial helps MKSABE-VaAR to increase search efficiency and decrease the number of bilinear pairing operations required for search by decreasing the

amount of search calculation for keyword ciphertext. Simultaneously, we have developed a specific indexing architecture to incorporate user attribute verification in the keyword search procedure, hence enhancing search precision.

2. PROPOSAL METHODOLOGY

Search phrases are used to search the encrypted data is possible using Searchable Encryption (SE). However, most of the existing SE systems are incapable of dealing with shared records with hierarchical topologies. The SE approach, which permits cloud servers to access encrypted data on behalf of data owners without affecting data confidentiality, has significantly improved the security, speed, and relevancy of searching through Encrypted data. This paper suggests a cross-lingual, multi-keyword Centroid Merkle search over encrypted cloud data, called CLCMSE. The Centroid algorithm is used for clustering the cloud data because the cloud consists of a significant amount of data. After the cloud data is divided, the clustered data is sorted using the recommended clustering method. The Merkle search approach is utilised to make the search process go faster.

Additionally, the secure data from the cloud is retrieved using an effective retrieval technique. The fuzzy retrieve technique is used to retrieve absolute data from cloud storage in a secure manner safely. A significant amount of data is efficiently clustered using the machine learning Centroid approach. Also, the speed of the search is being increased by the Merkle search method. The enhanced encrypted keyword searching process retrieves the exact data from the cloud storage, and also the proposed system used Multilingual Wordnet for Multilanguage keyword search.

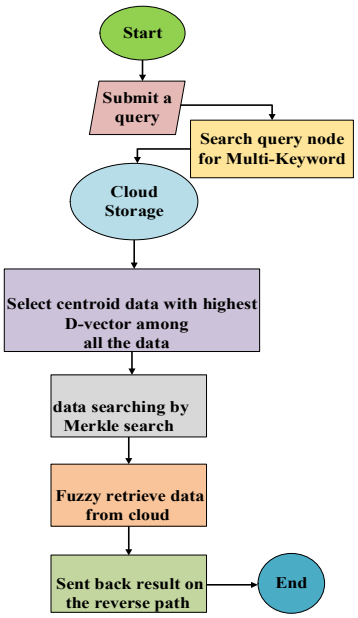

Fig.1.Suggested Model

The general operation of the suggested encrypted keyword search method is depicted in the diagram above. The input used in this procedure is a query word, which is the user's desired search term. Initially, the user queries the server with a request. Encrypted data is the search term, and the system goes into cloud storage. The enhanced centroid clustering algorithm is done in clustering and sorting the data from the cloud storage. The Centroid algorithm clusters a large amount of data based on selecting centroid data with the highest D-vector process. Now the numerous amounts of data have been clustered and sorted to get the exact data. Next, implementing the searching algorithm for efficient searching. The proposed Merkle searching technique explores the targeted data using point-to-point applications, improving searching efficiency and increasing the searching speed.

Implementing the retrieval technique retrieves the data from the cloud based on the membership function. This proposed system employs the fuzzy retrieval technique to retrieve data from cloud storage efficiently. By using the suggested Cross-Lingual Multi-keyword Centroid Merkle Search over Encrypted Data (CLCMSE) scheme, the keyword searching procedure allows the user to query and retrieve data from the cloud. The recommended Cross-lingual multi-keyword Centroid Merkle search over encrypted data (CLCMSE) method is shown in Figure 1. The introduced method increases satisfaction results, and speed and enhances retrieval techniques to

ensure cloud data security. The enhanced keyword search technique gives the absolute result based on the user's search word.

Research Contribution:

Recently, there has been a notable surge in researchers' interest in cross-lingual multi-keyword Centroid Merkle search. Studies in this area focus on developing solutions for accurate results in various languages. In this paper, we make the following contributions.

- We believe we are the first to investigate cross-lingual multi-keyword ranked search over encrypted cloud data. Our system enhances cross-lingual target query capabilities by leveraging Open Multilingual WordNet (OMW) with language conversion and semantic extension, aiming to overcome language barriers in searchable encryption.
- The flexible keyword and language preference settings, along with automated scoring based on semantics, enable intelligent and personalised sorting search, enhancing accuracy across all languages.
- We assess our scheme's performance based on accuracy, security, and speed through extensive experiments.

3.1. Clustering

Data clustering is a complicated method in machine learning, pattern recognition, image processing, and information extraction domains. Using similarity metrics (such as Euclidean distance) to arrange related data into a single cluster is known as clustering. Even though several data clustering strategies have been presented. Due to the inefficiency of the similarity metrics used in traditional clustering approaches, they typically perform poorly on high-dimensional data. Furthermore, on large-scale datasets, these approaches have a significant computation time [29]. In this paper, the Centroid clustering technique is proposed. The data from the cloud is clustered using the Centroid clustering technique, and the data is accurately clustered using the suggested clustering algorithm.

3.2. Searching

The data-searching mechanism is required to search and receive the shared data by authorised devices. There are currently few solutions available to handle the difficulties of safe data sharing and cloud search. Because the basic goal of cloud storage is to reuse data in the future, data consumers must be able to rapidly and precisely identify a specific group of data [30]. In this paper, the enhanced Merkle searching technique is used for the data searching process. The proposed searching technique works efficiently with the keyword searching process. It also searches the data from the cloud very fast.

3.3. Data Retrieval

Many real-world approaches depend heavily on information retrieval, such as online databases, expert findings, internet browsing, and so on. Information Retrieval is the process of extracting information resources from enormous collections relevant to a specific information demand. Because there may be various relevant resources, the retrieved results are usually sorted by some criterion of significance [31]. The process of extracting text, images, or multimedia content tailored to a query from online resources is known as data retrieval (IR). Diagonal searches, in general, retrieve information with substantially less precision than topical or perpendicular searches. Consequently, scientists are continuously working to improve information retrieval methods to improve accuracy [32]. This research uses the fuzzy retrieval technique to enhance data retrieval to retrieve secure data from cloud storage. This process involves searching the clustered cloud data, after which the data is securely retrieved from the cloud by the retrieval system.

4. PROPOSED CENTROID MERKLE SEARCH RETRIEVAL METHOD

The technique of grouping a set of objects into non-overlapping groups is known as clustering analysis. Every subset consists of a group of items that are similar to each other, not those in other groups (intra-similarity), (inter-dissimilarity). There are numerous methods for clustering. Based on partitioning, two of the most popular clustering methods are the K-means and K-medoids algorithms. The K-medoids clustering algorithm is more precise and resistant to noise and variations than the K-means technique, considering its slower processing speed. As a result, the K-medoids approach is also popular.

Consider that the data set D = $\{c_1\ c_{2,\ldots.}\ c_n\}$(1) consists of n objects, each of which has m Characteristics $c_i \in P^m$. The term dist $(c_i\ ,c_j)$ refers to the distance between the objectsc_i and $c_j$in D.The fast K clustering technique selects K medoids

b(i),i=1,….K (2)

whereB=$\{b_1\ , b_2\ , \ldots\ldots b_k\} \subseteq D$, (3)

then, based on the distance dist (·,·), allocates each object in D to the nearest medoids, then K clusters L_1 L_2 , ... L_k are designed to lower the total cluster cost F,

$$F=\sum_{K=1}^{K}\sum_{C\subseteq K^L} di(st_k, b)^2 \quad (4)$$

Formulation 1: The object distance dist (·,·) is defined as the Euclidean distance.

$$\text{dist}(c_i, c_j) = \sqrt{\sum_{t=1}^{m}(c_i^t - c_j^t)^2} \quad (5)$$

where m is the object's attribute number.

Formulation 2: The variation of data set D is expressed as:

$$\sigma = \sqrt{\frac{1}{n-1}}\sum_{i=1}^{n}(dist_i, \overline{C}^2) \quad (6)$$

Where,

$$\bar{C} = \sum_{i=1}^{n} c_i/n \quad (7)$$

is the object mean. The $c_i$variance object is defined as follows:

$$\sigma_i = \sqrt{\frac{1}{n-1}}(\text{dist}_i, \ c_j)^2 \quad (8)$$

The variance represents the data degree of divergence. In Eq. (3), the difference between all items and their average is calculated, whereas in Eq. (4), the difference between c_i i and all other objects are calculated. Since the outliers have high variance because they are generally located distant from the central region, the larger an object's variance, the less probable it is to be a medoid. The proposed Centroid clustering algorithm clusters the cloud data accurately. But this technique is not to do the other keyword search process like searching and retrieving data from the cloud. So, the improved search algorithm is implemented in this paper.

The Merkle tree, also known as a Hash tree, was proposed by Ralph Merkle. Block size hash values are in the leaf nodes, and child hash values are in the parent nodes. The ternary tree is a three-node data structure with three child nodes: left, middle, and right. The proposed methodology uses it as a hash tree. In the ternary hash tree, each parent node has exactly three child nodes. The data integrity of the massive tree data blocks is validated using the generated root hash value. Merkle Search Trees produce an ordered set of elements in a space S that are ultimately sorted. To create maps, elements can be paired with tags from a set U (the values, in which case S becomes the set of keys). U presents a default element ⊥('bottom'), which signals a key missing from the map

$$f: S \rightarrow U. \quad (9)$$

If U is a CLCMSE with a merge operation $\sqcup_U$such that

$$\forall y, \perp \sqcup_U y = y \sqcup_U \perp = y \quad (10)$$

Merkle Search Trees then apply CLCMSE on S→ U as specified by a point-to-point application of

$$\sqcup_U: (f \sqcup h)(y) = f(y) \sqcup_U h(x). \quad (11)$$

The sequence of states a CLCMSE takes is meant to be monotonic concerning⊔, a transition from a state y to a state y'must be of the form

$$y' = y \sqcup o. \quad (12)$$

In Merkle Search Trees to those that produce monotonic sequences for each key, this means limiting the operations that can be employed. As a result, instead of using the operations put and delete directly, updates of the type update (f; s; v) = put should be used.

$$(f; s; \text{get}(f; s) \sqcup u). \quad (13)$$

By choosing 𝕌properly, this technique can get different types of data. For instance if

$$\mathbb{U} = \{\perp, \top\} \quad (14)$$

To get an improved set defined on S, a Boolean indicates if an item is available. To get a crucial store with last-writer-wins reconciliation, if U is one last register with the latest version.

As the value type 𝕌, any current CLCMSE type can be used, resulting in a map CLCMSE construction that efficiently identifies distinct things. The Merkle searching algorithm completes the search process, and ultimately, this system implements a data retrieval approach to recover data from the cloud, in this case, the fuzzy retrieval system. The fuzzy method works well with data retrieval systems.TheMerkle searching algorithm is done through the searching process. Finally, this system implements the data retrieval technique for retrieving the data from the cloud; here, the fuzzy retrieval system retrieves the data from the server.

The fuzzy technique functions well for the system that retrieves data. The linguistic word 'LV_i' describes all variables (input or output), according to the idea of fuzzy sets. Linguistic values represent the value y_i^j that fits into the conversation environment LV_i'. The discourse's universe elements are partially "belonging to" the linguistic value [0,1], which is supplied by a membership function $\mu(LV_i)$, utilising FRS_{iv} as the inputs (fuzzy retrieval system). The clear values are different.

$$\mu_{y_i^j}(LV_i) = \text{FRS}_{\text{iv}} \rightarrow [0,1] \quad (15)$$

The whole set value is represented by the value 1. The fuzzification procedure must change the crisp values instead of a fuzzy value, and different techniques are used for singleton fuzzification. The law of If Else inference governs how i/p is assigned to o/p. To extract assumptions from both

the rule base and i/p, an inferential phase is necessary.

IF

$$LV_i \rightarrow y_i^1 LV_i \rightarrow y_i^2 LV_i \rightarrow y_i^3 LV_i \cdots\cdots \rightarrow y_i^n \quad (16)$$

Then,

$$a_i = G_i(.) y_i^j \quad (17)$$

The fundamental argument is given by "." in the preceding steps. A method containing the term data is generated using the Takagi-Sugeno inference. The defuzzification approach is required at the end of the process to obtain crisp output data.

$$Z = \frac{\sum_{i=1}^{R} a_i \mu_i}{\sum_{i=1}^{R} \mu_i} \quad (18)$$

The process of fuzzification is the transformation of given input data into knowledge representations. Next, FL has created functions primarily based on "categories" that are easier for physicians to recognise and information clustering with similar qualities for decision-making operations. Finally, the fuzzy retrieve system retrieves the data from the cloud efficiently.

3. RESULT

In this section, to run many tests to see how well the CLCMSE-suggested search method works. To test this strategy, use a PC with a 3.40GHz Intel®Core TM i7-6700 processor and 16.0GB RAM. The Eclipse integrated development environment and the Java programming language are used to implement each algorithm and protocol in CLCMSE. Using the Natural Language Tool-Kit wordnet interface (NLTK), which also retrieves the open multilingual wordnet corpus, the query extension is created. To access the precision of the presented search scheme is compared to other techniques in terms of performance and efficiency. The proposed scheme's satisfaction result determines whether the suggested results match the needs of the data users. Your search results will be more accurate depending on the type of extended language you choose because the OMW (Open Multilingual Wordnet) supports multiple languages. We tested our approach on one hundred native speakers of different popular languages to conduct this study. The target language in this experiment is the participants' native tongues, while the query language is English.

Table1. Accuracy of the major languages spoken throughout the world

	Satisfaction	BS	DS
French	65%	30%	0%
Spanish	97%	5%	3%
Russian	70%	10%	10%
Portuguese	86%	23%	5%
Japanese	98%	11%	2%
German	72%	4%	11%
Indonesian	99%	2%	0%
Chinese	63%	24%	9%

Relevant data were categorised using the satisfaction, basic satisfaction, and dissatisfaction categories. To counteract the negative effects of dataset selection, we generate datasets in many languages using almost identical keyword collections on the result. We recorded the input from the participants in Table 1.

Table 2. Index tuple storage overhead with varying document collection sizes

No.Of. Documents	2000	4000	6000	8000	10000
MRSE (MB)	9.75	19.5	29.25	39	48.75
CLRSE (MB)	10.7	21.44	32.16	42.88	53.6
CLCMSE (MB)	12.1	23.65	35.1	43.96	61.49

Table 2 also contrasts the index tuple storage overheads between the two systems, taking into account the varying document collection sizes and keyword counts. Total 4000 documents are used in the experiment which contains two keywords in each document.

Table 3. Storage overhead of index tuples in the document containing varying numbers of keywords

No.Of.Keyword in Doc	2	4	6	8	10
MRSE(MB)	11.79	19.49	27.28	35.09	42.88
CLRSE(MB)	13.64	21.44	29.24	37.03	44.83
MRSE(MB)	15.85	24.37	30.92	41.65	45.76

The index tuple storage overheads in the two systems with varying numbers of keywords per document are contrasted in Table 3. The findings show that compared to the MRSE method, the suggested CLCMSE method has a considerably higher index storage overhead.

Table 4.Execution time of proposed searching scheme with the different bit length of S.

S(N)	512	768	1024	1280	1536	2048
CP	2.096	7.253	15.685	38.63	85.84	301
CSP	0.425	1.145	1.958	3.632	10.25	36.426
Total	2.924	8.215	20.116	43.26	96.25	295.45

The proposed technique also considered the processing speed. The enhanced searching algorithm increasing the processing speed also reduced the time. The proposed searching scheme's execution time with varied bit lengths of S (N) is The query response period is recorded in this experiment to determine query efficiency, which relates to the time between issuing the query request and receiving the search results. Set up a multilingual dataset with 2048 encrypted documents and safe correlated indices in Chinese, English, and other languages to test response times. The original query has four keywords. T is the threshold value of 0.5. In comparison to the MRSE schemes, Figure 3 shows the overhead of our basic and modified CLCMSE approaches vs the quantity of returned results. In comparison to the MRSE system, our scheme necessitates an additional query extension phase. This technique has a very low estimate using cp, csp, and the overall process, as shown in the table. 4

Fig.2.Graphical representation of Execution time of the proposed searching scheme time overhead.

The proposed approach achieves reasonable satisfaction and works Fig.2.Graphical representation for the Execution time of proposed searching scheme (S) with a different bit length efficiently.

Table.5.Execution time of proposed searching scheme (S) with a different bit length

D	128	512	1024	2048	4096
CS	10.012	12.253	14.863	16.15	18.6
CSP	0.936	1.248	0.28	1.469	1.84
Total	10.216	15.635	15.985	18.06	20.2

Furthermore, our improved CLCMSE approach builds the initial heap and modifies the heap shape dynamically. Merkle searching does not require running the whole K-rounds sorting algorithm with this technique. The greater the amount of N, the more efficient the query will be overall. Table 5 indicates all protocols that operate on the CP and CSP can be run concurrently, resulting in much-enhanced method execution performance.

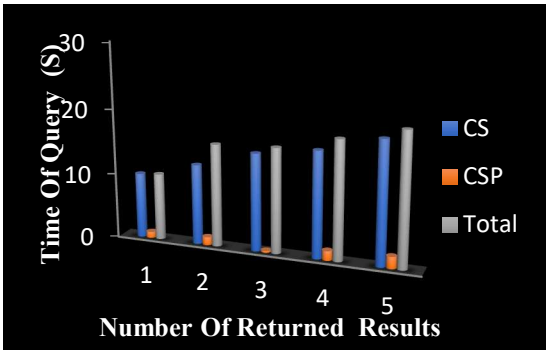

Figure 3. Execution time of proposed searching scheme (S) with different bit lengths.

The larger the value N, the higher overall query efficiency can be achieved. Fig 3 indicates the efficiency of all protocols that make use of the CP and CSP can be improved by running them all simultaneously.

5.1 Discussion

This research effectively resolves the problem of Cross-lingual multi-keyword rank search with a semantic extension over encrypted data. Through lingual-multi keywords and language preference settings, our CLCMSE scheme also achieves intelligent and personalized search. We evaluate the performance of our scheme in terms of index tuple storage overheads, variable number of keywords per document, execution time with varied bit length of S(N) using CP, CSP.The Merkle search technique increases the search speed when compared with other methods such as cross-lingual multi-keyword rank search, Multi-keyword Rank Searchable Encryption, Verifiable Attribute multi-keyword search over encrypted cloud data. The proposed method proved the accuracy result in all languages; specifically, Indonesian (99%) and Japanese(98%) are achieved. Practically, expressive search queries should be supported because single keyword searches may yield many irrelevant results and decrease user search experience.

6. CONCLUSION

Traditional and broad data retrieval relies on keyword searches, which have limitations like a high manpower need and a dependence on private data, leading to poor simulation outcomes.To overcome the above issue, we have suggested the CLCMSE technique. This technique allows data users to query in any language and choose the linguistic kind of information.The problem of cross-lingual multi-keyword rank search over

encrypted material with a semantic extension was solved by this work. A cross-lingual multi-keyword rank search system (CLCMSE) based on OMW is provided as a solution to this problem. The presented system created a significant advance in searchable encryption by removing linguistic restrictions in searchable encryption. This system also features keyword and language request parameters that may be customised and automated preference score computations. Extensive experiments are used to assess one's performance of our plans, which are then compared to existing schemes. The enhanced centroid algorithm clustered the cloud data inefficiently. Also, the Merkle search technique is increasing the search speed. Finally, the enhanced fuzzy retrieval system securely retrieves the data. We also compare our method with other state-of-the-art methods such as Cross-lingual multi-keyword rank search, Multi-keyword Rank Searchable Encryption, and Verifiable Attribute multi-keyword search over encrypted cloud data. The suggested approach demonstrated accuracy in all languages; in particular, 99% and 98% of the results were obtained in Indonesian and Japanese. As a part of future work, we will continue to improve the security and further improve the response speed of this work.

REFERENCE

[1] W. Yang and Y. Zhu, "A Verifiable Semantic Searching Scheme by Optimal Matching Over Encrypted Data in Public Cloud," in *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 100-115, 2021, doi: 10.1109/TIFS.2020.3001728.

[2] . Guo, Z. Li, W. -C. Yau and S. -Y. Tan, "A Decryptable Attribute-Based Keyword Search Scheme on eHealth Cloud in Internet of Things Platforms," in *IEEE Access*, vol. 8, pp. 26107-26118, 2020, doi: 10.1109/ACCESS.2020.2971088.

[3] Guo, Ziqing, et al. "Secure multi-keyword ranked search over encrypted cloud data for multiple data owners." Journal of Systems and Software 137 (2018): 380-395, Volume 137, March 2018, Pages 380-395.

[4] Y. Zhang, C. Xu, J. Ni, H. Li and X. S. Shen, "Blockchain-Assisted Public-Key Encryption with Keyword Search Against Keyword Guessing Attacks for Cloud Storage," in *IEEE Transactions on Cloud Computing*, vol. 9, no. 4, pp. 1335-1348, 1 Oct.-Dec. 2021, doi: 10.1109/TCC.2019.2923222.

[5] Y. Cui, F. Gao, Y. Shi, W. Yin, E. Panaousis and K. Liang, "An Efficient Attribute-Based Multi-Keyword Search Scheme in Encrypted Keyword Generation," in *IEEE Access*, vol. 8, pp. 99024-99036, 2020, doi: 10.1109/ACCESS.2020.2996940.

[6] Y. Miao, R. H. Deng, K. -K. R. Choo, X. Liu and H. Li, "Threshold Multi-Keyword Search for Cloud-Based Group Data Sharing," in *IEEE Transactions on Cloud Computing*, vol. 10, no. 3, pp. 2146-2162, 1 July-Sept. 2022, doi: 10.1109/TCC.2020.2999775.

[7] Y. Zhang, R. H. Deng, J. Shu, K. Yang and D. Zheng, "TKSE: Trustworthy Keyword Search Over Encrypted Data With Two-Side Verifiability via Blockchain," in *IEEE Access*, vol. 6, pp. 31077-31087, 2018, doi: 10.1109/ACCESS.2018.2844400.

[8] X. Zhang, C. Xu, H. Wang, Y. Zhang and S. Wang, "FS-PEKS: Lattice-Based Forward Secure Public-Key Encryption with Keyword Search for Cloud-Assisted Industrial Internet of Things," in *IEEE Transactions on Dependable and Secure Computing*, vol. 18, no. 3, pp. 1019-1032, 1 May-June 2021, doi: 10.1109/TDSC.2019.2914117

[9] A. F. S. Devaraj *et al.*, "An Efficient Framework for Secure Image Archival and Retrieval System Using Multiple Secret Share Creation Scheme," in *IEEE Access*, vol. 8, pp. 144310-144320, 2020, doi:10.1109/ACCESS.2020.3014346.

[10] C. Lou *et al.*, "A Secure Key-Aggregate Keyword Retrieval Scheme Over Encrypted Data in Cloud Computing," in *IEEE Access*, pp. 1-1, 2020, doi: 10.1109/ACCESS.2020.2980886.

[11] C. Ge, W. Susilo, Z. Liu, J. Xia, P. Szalachowski and L. Fang, "Secure Keyword Search and Data Sharing Mechanism for Cloud Computing," in *IEEE Transactions on Dependable and Secure Computing*, vol. 18, no. 6, pp. 2787-2800, 1 Nov.-Dec. 2021, doi: 10.1109/TDSC.2020.2963978.

[12] G. Liu, G. Yang, S. Bai, H. Wang and Y. Xiang, "FASE: A Fast and Accurate Privacy-Preserving Multi-Keyword Top-k Retrieval Scheme Over Encrypted Cloud Data," in *IEEE Transactions on Services Computing*, vol. 15, no. 4, pp. 1855-1867, 1 July-Aug. 2022, doi: 10.1109/TSC.2020.3023393.

[13] L. Zhang, Y. Zhang and H. Ma, "Privacy-Preserving and Dynamic Multi-Attribute Conjunctive Keyword Search Over Encrypted Cloud Data," in *IEEE Access*, vol.

6, pp. 34214-34225, 2018, doi: 10.1109/ACCESS.2018.2823718.

[14] Elizabeth, B. Lydia, and A. John Prakash. "Verifiable top-k searchable encryption for cloud data, 45, (2020). https://doi.org/10.1007/s12046-019-1227-5

[15] Y. Miao *et al*., "Privacy-Preserving Attribute-Based Keyword Search in Shared Multi-owner Setting," in *IEEE Transactions on Dependable and Secure Computing*, vol. 18, no. 3, pp. 1080-1094, 1 May-June 2021, doi: 10.1109/TDSC.2019.2897675.

[16] Yang, Yang, et al. "Lattice assumption based fuzzy information retrieval scheme support multi-user for secure multimedia cloud." Multimedia Tools and Applications 77,9927-9941, 2018, doi.org/10.1007/s11042-017-4560-x.

[17] .M. Shen, B. Ma, L. Zhu, X. Du and K. Xu, "Secure Phrase Search for Intelligent Processing of Encrypted Data in Cloud-Based IoT," in *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 1998-2008, April 2019, doi: 10.1109/JIOT.2018.2871607.

[18] Yin, Hui, et al. "Secure conjunctive multi-keyword ranked search over encrypted cloud data for multiple data owners." Future Generation Computer Systems 100 (2019): 689-700, https://doi.org/10.1016/j.future.2019.05.001.

[19] Dai, Hua, et al. "Semantic-aware multi-keyword ranked search scheme over encrypted cloud data." Journal of Network and Computer Applications 147 (2019): 102442, https://doi.org/10.1016/j.jnca.2019.102442.

[20] C. Guo, R. Zhuang, C.Chang and Q. Yuan, "Dynamic Multi-Keyword Ranked Search Based on Bloom Filter Over Encrypted Cloud Data," in *IEEE Access*, vol. 7, pp. 35826-35837, 2019, doi: 10.1109/ACCESS.2019.2904763.

[21] K. He, J. Guo, J. Weng, J. Weng, J. K. Liu and X. Yi, "Attribute-Based Hybrid Boolean Keyword Search over Outsourced Encrypted Data," in *IEEE Transactions on Dependable and Secure Computing*, vol. 17, no. 6, pp. 1207-1217, 1 Nov.-Dec. 2020, doi: 10.1109/TDSC.2018.2864186.

[22] 22. Y. Miao *et al*., "REKS: Role-Based Encrypted Keyword Search With Enhanced Access Control for Outsourced Cloud Data," in *IEEE Transactions on Dependable and Secure Computing*, pp.1-15, doi: 10.1109/TDSC.2023.3324640.

[23] Shen, Fanfan, Lin Shi, Jun Zhang, Chao Xu, Yong Chen, and Yanxiang He. "BMSE: Blockchain-based multi-keyword searchable encryption for electronic medical records." *Computer Standards & Interfaces* 89 (2024): 103824,https://doi.org/10.1016/j.csi.2023.103824

[24] Liu, Shuqin, Xialin Liu, Wanxuan Huang, and Kai Du. "Efficient dynamic multi-client searchable encryption supporting fuzzy search." *Computer Standards & Interfaces* 88 (2024): 103772, https://doi.org/10.1016/j.csi.2023.103772

[25] Liu, Xingchen, Shaohui Zhang, Haiping Huang, and Reza Malekian. "A trustworthy and reliable multi-keyword search in blockchain-assisted cloud-edge storage." *Peer-to-Peer Networking and Applications* (2024): 1-16, https://doi.org/10.1007/s12083-024-01635-9.

[26] Q. Tong, X. Li, Y. Miao, Y. Wang, X. Liu and R. H. Deng, "Beyond Result Verification: Efficient Privacy-Preserving Spatial Keyword Query With Suppressed Leakage," in *IEEE Transactions on Information Forensics and Security*, vol. 19, pp. 2746-2760, 2024, doi: 10.1109/TIFS.2024.3354414.

[27] Wang, Miaomiao, Lanlan Rui, Siya Xu, Zhipeng Gao, Huiyong Liu, and Shaoyong Guo. "A multi-keyword searchable encryption sensitive data trusted sharing scheme in multi-user scenario." *Computer Networks* 237 (2023): 110045, https://doi.org/10.1016/j.comnet.2023.110045.

[28] H. Shen, J. Zhou, G. Wu and M. Zhang, "Multi-Keywords Searchable Attribute-Based Encryption With Verification and Attribute Revocation Over Cloud Data," in *IEEE Access*, vol. 11, pp. 139715-139727, 2023, doi: 10.1109/ACCESS.2023.3334733.

[29] Y. Yang, X. Liu and R. H. Deng,(2020). "Multi-User Multi-Keyword Rank Search Over Encrypted Data in Arbitrary Language," in *IEEE Transactions on Dependable and Secure Computing*, vol. 17, no. 2, pp. 320-334, doi: 10.1109/TDSC.2017.2787588.

[30] Bond, Francis, and Ryan Foster (2013). "Linking and extending an open multilingual wordnet." In *Proceedings of the 51st Annual Meeting of the Association for*

Computational Linguistics (Volume 1: Long Papers), pp. 1352-1362.

[31] Min, Erxue, et al. "A survey of clustering with deep learning: From the perspective of network architecture." IEEE Access 6 (2018): 39501-39514, https://doi.org/10.1109/ACCESS.2018.2855437.

[32] J. -S. Fu, Y. Liu, H. -C. Chao, B. K. Bhargava and Z. -J. Zhang, "Secure Data Storage and Searching for Industrial IoT by Integrating Fog Computing and Cloud Computing," in *IEEE Transactions on Industrial Informatics*, vol. 14, no. 10, pp. 4519-4528, Oct. 2018, doi: 10.1109/TII.2018.2793350.

[33] Guo, Jiafeng, et al. "A deep look into neural ranking models for information retrieval." Information Processing & Management 57.6 (2020): 102067, https://doi.org/10.1016/j.ipm.2019.102067.

[34] .Bhopale, Amol P., and Ashish Tiwari. "Swarm optimised cluster based framework for information retrieval." Expert Systems with Applications 154 (2020): 113441, DOI:10.1016/j.eswa.2020.113441.