

OntoFuzz: An Information Retrieval Model in a Multi-Tenant Cloud Environment using Neuro-Fuzzy and Ontological-based Approach

Dimpy Jindal^{1*}, Manju Kaushik¹ & Barkha Behl²

¹Amity University, Rajasthan, Jaipur 303 002, India

²Trinity Institute of Professional Studies, Delhi 110 075, India

Received 23 April 2023; revised 18 April 2024; accepted 05 July 2024

There is a tremendous amount of data present on the web and accessing useful/relevant information from a cluster of random documents is a tedious and time-consuming task. Traditional information retrieval techniques and information management systems are not that intelligent to extract relevant information from pre-defined datasets or documents. This necessitates the researchers to create and enhance a sophisticated information retrieval system. Also, the similarity between information is equipped with uncertainties due to its computing measures. Keeping these issues in mind, a neuro-fuzzy and ontological-based model in a multi-tenant cloud environment is proposed in this research study. The model comprises modules like query expansion, the weighting of terms and queries, and hashing function to ease the retrieval process followed by validation of the dataset using a neuro-fuzzy network to retrieve relevant information from the cloud service provider. The simulation results prove the validation of the proposed model in terms of higher accuracy and better retrieval performance as compared to traditional models (support vector machines and deep neural networks) as well as existing recent works.

Keywords: Cloud computing, Information retrieval, Multi-tenancy, Neuro-fuzzy, Ontology

Introduction

Information extraction is a critical step in cloud computing because it permits the storage and recovery of useful information from the local host to the cloud server and vice versa. The system for managing information experiences a significant transformation in a short time due to the availability of resources and technological innovation. With its cutting-edge web-based information management solutions, technology has also begun removing the last vestiges of conventional information management techniques. Similarly to this, the demand for information management systems is continually growing as a result of modern advanced search engines, internet-based services, network mediums, and electronic libraries. With its cutting-edge web-based information management solutions, technology has also begun removing the last vestiges of conventional information management techniques. Similarly to this, the demand for information management systems is continually growing as a result of modern advanced search engines, internet-based services, network mediums, and electronic libraries. This necessitates

the researchers to create and enhance a sophisticated information retrieval system.

Problem Statement and Significance of the Study

The significance of information retrieval is demonstrated by the way it extracts the material from databases that are most appropriate for a given query. But the question is: How can the relevant data for the query to be retrieved? The precision and recall will increase when information is extracted based on semantics instead of traditional text-based keywords. According to the query, a big group of documents is separated into smaller sets for this reason to extract data mining's ultimate objective to gather data based on patterns that can be recovered to gain crucial knowledge from the massive amounts of data that have been stored. The best example of information management nowadays is cloud computing, which is based on pay per usage policy and is extensively used by both large and small enterprises. The cloud provides services based on needs, reducing substantial expenses for businesses because they do not have to pay for hardware or environment setup. Because cloud computing is widely used, users can easily expand or decrease storage capacity as needed, which further lowers organizational costs. Hence, this

*Author for Correspondence
E-mail: dimpy.jindal23@gmail.com

research work attempts to provide an efficient and reliable means of retrieving relevant information in the context of user queries using the fuzzy and ontological-based approach in a multi-tenant cloud environment.

Related Works

This section summarises relevant research on information retrieval that has already been done to identify potential problems with retrieval system implementation. Jiaying *et al.*¹ analyze the data mining and information retrieval system. The survey provides in-depth information regarding applications of information retrieval systems in the cloud environment. Yongjun *et al.*² identified problems with the bibliographic information retrieval system's natural language interface since database management systems have trouble organizing data in natural language, it is difficult to perform information retrieval using natural language. In their research, Andrei *et al.*³ describe a ranking-based information retrieval technique. When calculating model ranks based on document description and phrase frequency, the user request is taken into account. The research employs a modified genetic algorithm and offers helpful data immediately. A genetically based map criterion and ranking models are used in the proposed method to lessen the structural complexity of conventional information retrieval. However, the fact that the recommended system is unable to react to questions posed in natural language is regarded as a model restriction. Issues with the multimodal retrieval method for getting document images are reported by Datta *et al.*⁴ Graphs are used to represent the words in a text, which makes it easier to assess how similar the words are. Synonymy and a semantic index are added to the similarity measure to improve performance by obtaining the data required for the user query. The studies by Tajeddine *et al.*⁵ and Yang *et al.*⁶ employing distributed databases for information retrieval discovered similar models which increase retrieval performance and data security. A cluster-based information retrieval paradigm was put forth by Djenouri *et al.*⁷. Based on the user question, the suggested approach determines the frequency of the information and returns the most common things as results. The recommended approach has advantages because it returns to the user the most well-liked terms that are frequently used by others. This model's weakness is the incorporation of unnecessary data, though. Angelini *et al.*⁸ proposed

multimedia information retrieval exchange based on a case-based dataset. Information processing problems in a multi-cloud context were reported by Jennifer *et al.*⁹. Due to its interface, service delivery, and technology, data upkeep in a multi-cloud context is a difficult procedure. Joby *et al.*¹⁰ performed an information retrieval process with traditional deep learning models like Support Vector Machines (SVM) and Convolutional Neural Networks (CNN). It produces repetitive results due to a lack of filtering and expansion of queries entered by the user. Chandrika *et al.*¹¹ perform accessing of data in a multi-tenant environment using an Elliptic Curve Cryptography (ECC) algorithm but its limitation is that the data is not refined with relevant results. It does not make use of any deep learning and ontological-based methodology to refine the retrieval process. Krishnaraj *et al.*¹² devised a model based on semantic indexing for large-scale image retrieval in a cloud environment. It does not involve the expansion of terms and text-based keywords which may lead to higher precision and retrieval efficiency of entered data. In addition to this, a concise view of some extra literature review studies is presented in Table 1.

Ontology in Cloud

A hierarchical representation of data in the form of classes, properties, and instances to make it more meaningful and representable is termed Ontology. Ontology is treated as the formal, explicit specification of a shared conceptualization.¹⁶ Besides its formal nature, philosophical aspects, and handling of real-world scenarios, it also acts as a medium for linking humans and machines.¹⁷

Proposed Model

It begins with the expansion of the input query entered by the user. Query expansion involves pre-processing data sources, term extraction, term weighing and ranking, term selection, and query rephrasing. It is described as follows.

Pre-processing of Sources of Extraction of Terms and Data

The main objective of this stage is to retrieve from the data source a list of phrases that significantly broadens the user's initial query. It includes extrapolation of text from the data source (full-text extraction from the particular data source utilized for query expansion), tokenization (technique of word-by-word splitting of a text stream), and stop words elimination (removing terms that are used frequently, such as articles, adjectives, prepositions, etc.).

Table 1 — Detailed insights of a few selected papers

| Author | Year | Description | Shortfalls indicated in referred articles | Algorithm/ Technique used |
|--|------|--|---|---|
| Jiaying <i>et al.</i> ¹ | 2019 | Detailed information about information retrieval processes in a cloud environment. | Not suited for domain-specific applications and training datasets. | Not applicable |
| Yongjun <i>et al.</i> ² | 2017 | Identified issues with existing information retrieval systems. | Cannot validate dataset on any simulation platform | Natural language processing |
| Andrei <i>et al.</i> ³ | 2017 | Calculates model ranks based on document description and phrase frequency. | The model is unable to react to queries in natural language | Genetic algorithm |
| Datta <i>et al.</i> ⁴ | 2017 | Semantic indexing of queries is done to improve the retrieval process | No algorithm and training dataset validation | Learning graphs |
| Tajeddine <i>et al.</i> ⁵ | 2018 | Uses distributed databases for information retrieval | Although it provides high security and higher efficiency, failed to provide validation of results. | Not applicable |
| Yang <i>et al.</i> ⁶ | 2018 | Uses distributed databases for information retrieval | Although it provides high security and higher efficiency, failed to provide validation of results. | Not applicable |
| Djenouri <i>et al.</i> ⁷ | 2019 | Performs retrieval based on the frequency of terms entered by the user. | Does not consider the user's previous preferences | Cluster-based approach |
| Angelini <i>et al.</i> ⁸ | 2018 | Works on the assumed dataset by computing similarity measure | Low accuracy | Visual analytics approach |
| Jennifer <i>et al.</i> ⁹ | 2019 | Data is accessed through cryptographic and hash functions in a cloud environment | Irrelevant search results | Computational intelligence |
| Joby <i>et al.</i> ¹⁰ | 2020 | Identified security issues in multi-tenant cloud systems by computing similarity measure | No concept of semantic indexing and ontology is being used | Probabilistic and vector space model |
| Chandrika <i>et al.</i> ¹¹ | 2021 | Self-adaptive system | It does not make use of any deep learning and ontological-based methodology to refine the retrieval process | Elliptic curve cryptography algorithm (ECC) |
| Krishnaraj <i>et al.</i> ¹² | 2021 | Image-based retrieval only | It does not involve the expansion of terms and text-based keywords | Semantic indexing |
| Babu <i>et al.</i> ¹³ | 2022 | Secure deduplication in Multi-tenant cloud | No concept of semantic integration and ontology is used | Hashing technique |
| Ubaidullah <i>et al.</i> ¹⁴ | 2023 | Virtual machine trusted cloud computing | Does not perform validation in testing phase | Elliptic cryptography |
| Vajpayee & Hossain ¹⁵ | 2024 | Risk prediction through multi-tenant cloud | No official dataset is taken into consideration | Homomorphic encryption |

Ranking and Weighing the Terms used in Query Expansion

The inputs for this step include the user's query and the texts that were obtained from the data sources in the preceding stage. The relevance of the phrases in the expanded query is shown by the assigned weights, which are subsequently applied to the documents that have been retrieved to score them. Romano *et al.*¹⁸ divide the strategies into three groups based on how the expansion features and query terms interact.

(a) *One-to-One relationship*: -Connect each enlarged term to at least one query phrase. The use of a glossary is necessary. One of the most well-known thesauruses is WordNet.¹⁹ WordNet is used to map each query phrase to its synonyms and a related set of terms in the expanded query. For instance, the term "java" has three synsets in WordNet, each with a

distinct sense: place. According to the common understanding of word similarity, two terms are semantically similar if they exist in the same text. Like this, texts are judged comparable if they include similar language. There are various methods for figuring out word similarity. The degree of phrase similarity between WordNet synsets is assessed using path length-based metrics.

Let t_1 and t_2 be the specified terms, and let ' len_s ' stand for the length of the shortest path in WordNet between t_1 and t_2 . Then, the definition of the "Shortest Path Similarity Score"²⁰ is:

$$SimPath(t_1, t_2) = 1 \div len_s \quad \dots (1)$$

The similarity between the terms t_1 and t_2 is also computed by the Cosine similarity measure defined as:

$$Sim_{cosine} = C_{t1, t2} \sqrt{\sum_{d \in j} w \times w(t1, j) \cdot \sum_{d \in j} w \times w(t2, j)}$$

... (2)

where $C_{t1, t2}$ denotes similarity between terms $t1$ and $t2$; $w(t1, j)$ and $w(t2, j)$ are the weight of terms $t1, t2$ in j^{th} document

(b) One-to-many relationship:-In this, a candidate term is included in the enlarged query if it is linked to numerous terms from the original query.

(c) Query Language Modelling:- It involves selecting the most likely growth terms and developing a quantitative approach for the inquiry. This method is also known as the model-based technique. The likelihood that a document will be among the top results is as follows:

$$log_p(Tr \downarrow \theta q) = \sum_{\substack{D \in t1 \\ D \in t2}} C(t1, t2, D) log(\mu(t \downarrow C) + (1 - \mu)(t, \theta q))$$

... (3)

where, Tr is the set of top-ranked documents; θq is the query model; $C(t1, t2, D)$ is the number of occurrences of terms in Document D ; μ is the weighing parameter having a value between 0 and 1

Selecting the Query Expansion Terms

The selection of the expansion terms based on the ranks assigned to each term has been recommended using a variety of techniques that utilize the additional information. A system that chooses the most popular terms for each query while utilizing a variety of term rating methods was proposed by Carpineto *et al.*²¹ To do this, corpus papers must be reexamined, and new questions must be created that relate to the original one. Fang *et al.*²² suggested an impact-sorted indexing method that makes use of a specific index data model and improves information retrieval scoring systems.

Query Rephrasing

The larger query is modified in the final stage of the expansion process to deliver better results when utilized to fetch pertinent documents. Query reweighting is the process of reformulating the query based on the weights given to each distinct term in the extended query. It can be formulated as:

$$RW_{t,q} = (1 - \mu) \cdot W_{t1} + \mu \cdot W_{t2}$$

... (4)

where, $RW_{t,q}$ is the reweighting of term t of the query q ; W_{t1} = weight assigned to term $t1$; W_{t2} = weight assigned to term $t2$; μ is the weighing parameter having a value between 0 and 1

After query expansion, it is fed into the Cloud Service Provider (CSP). The model is proposed in a distributed multi-tenant cloud environment comprising a pre-defined domain-specific dataset. The query is inputted by the user related to the dataset. Using the hash function, a hash table is created to maintain a database for easy retrieval of information. Hamming distance is calculated to map the queries of the same class with similar codes. Once it is identified, the data is set to train and validate using a neuro-fuzzy network based on fuzzy logic splicing and parallelization. As soon as a query is entered by a user, the relevant information from the trained dataset is mapped from a Cloud Service Provider (CSP) and provided to the user in form of ontologies (hierarchical structure of relevant results). With ontology, a user can easily identify classes, properties, and instances related to the entered query. The process is illustrated in algorithms as shown below:

Algorithm 1: Query expansion

Input: Query entered by the user

Output: Reformulated query

Step 1. Pre-processing of query

1.1 Extrapolation of text

1.2. Word-by-word splitting of text

1.3. Stop words elimination

1.4 Stemming of words

Step 2. The weighting of terms. Let terms be $t1, t2$, and weights associated with them be $w(t1)$ and $w(t2)$

2.1. Computing similarity between terms $t1$ and $t2$ using cosine similarity measure using Eq. 2:

2.2 Selecting the most likely occurring terms in the query using Eq. 3

Step 3. Ranking of selected terms

Step 4. Reformulation of query based on weights associated with each term using Eq. 4

Algorithm 2: Hashing function

Input: Expanded query features

Output: Binary codes, hamming distance

Step 1. After the expansion of the query, it is mapped by hash codes, and a mediator function is introduced to achieve better similarity between pairs of terms.

Step 2. Let $q1, q2$, and $q3$ be the pairs for the given query (Q)

Step 3. Let hash codes associated with pairs are $h1, h2, h3$ having range of n bit binary codes $\{+1, -1, +1\}^n$

Step 4. for $n \leq$ length of terms do

map $\{(h1, q1), (h2, q2) \text{ and } (h3, q3)\}$

compute mediator function M_f as

$M_f(h1, h2, h3, Q) = \frac{1}{2} (1-q) d_h(h1, h2, h3)$ where d_h is hamming distance computed to obtain similar pairs of terms with 0 code and dissimilar pairs with 1

Algorithm 3: Fuzzy logic splicing

Step 1. The neuron-activated function in the neuro-fuzzy model is used to set the threshold based on M_f computed in the above step. It works on the same codes for similar terms and binary codes for dissimilar terms.

Step 2. The similar terms are spliced using fuzzy rules and are computed as:

$$F_1(n_w \wedge o) = (I \div M_f) \sum_{n=1} w(t1) + (I \div Ln_w) \sum_{n=1} w(t2) - (2 \div Ln_w) n_w o \quad \dots (5)$$

where, F_1 = Fuzzy logic function, n_w = neuron weight, o = output function, Ln_w = neuron length

3. This neuro-fuzzy function will process either similar or dissimilar results at one time thus blocking the irrelevant results to the user.

The layout of the proposed model is presented in Fig. 1

Implementation Process

Hardware requirements: i3/i5/i7 processor with 2.2 GHz with 4 GB RAM or higher

Operating system: Windows 64 bit

IDE: NetBeans 8.2, Java Development Kit (JDK)

Simulator: CloudSim 3.03 version and MATLAB R2023a

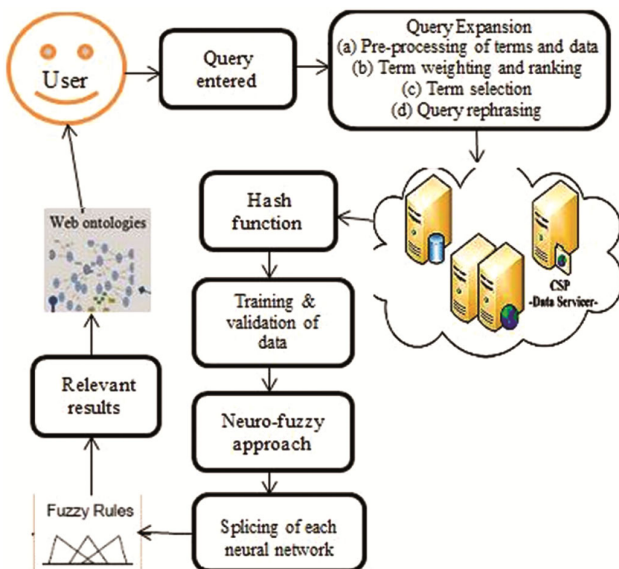


Fig. 1 — OntoFuzz- Proposed information retrieval model

Experimental Dataset

The proposed OntoFuzz algorithm for information retrieval in a multi-tenant cloud environment is experimentally verified in CloudSim²³ and MATLAB²⁴ and validated on the annual rainfall dataset taken from the Kaggle database.²⁵ The efficiency and retrieval performance of the proposed model is estimated using Precision, Recall, and f-measure. Conventional support vector machine (SVM) information retrieval systems and deep neural network models are compared with the proposed OntoFuzz system.

Precision (P) is calculated as $t_p \div (t_p \pm f_p) \quad \dots (6)$

Recall (R) is calculated as $t_p \div (t_p \pm f_n) \quad \dots (7)$

F-measure is calculated as $\frac{2 \times P \times R}{P + R} \quad \dots (8)$

The dataset has many training samples in the form of periods, a few of them are depicted in Table 2. In the dataset, S denotes seeded, U denotes unseeded, NC denotes rainfall in the north-central area, SC denotes rainfall in the south-central area, NWC denotes rainfall in the northwest-central area and TE denotes east target areas.

Simulation of Model in its Training and Testing Phase

The dataset used in the study initially has 10,000 training samples, 100 numerical attributes and 500 testing samples. Suppose a user enters a query ‘Plot the rainfall intensity wave graph in north-central area and south-central area during AUTUMN season’. The simulation of our proposed model is done in MATLAB and CloudSim and the results are shown in Fig. 2 and Fig. 3.

Results and Discussion

In Table 3, the proposed model results are collated with the conventional models^{26,27} and recent studies.¹²⁻¹⁵

It can be seen that the proposed OntoFuzz model outperforms the conventional models as well as recent studies in terms of precision, recall and f-measure. Also, the retrieval efficiency of the individual proposed OntoFuzz model is depicted in Fig. 4. It implies that performance increases even as the number of queries rises in the system.

Better accuracy is seen in the proposed OntoFuzz model as compared to traditional information systems as depicted in Fig. 5. The accuracy of the SVM system based on the number of queries is around 67–76%, and that of the Deep neural network is around 56–83%.

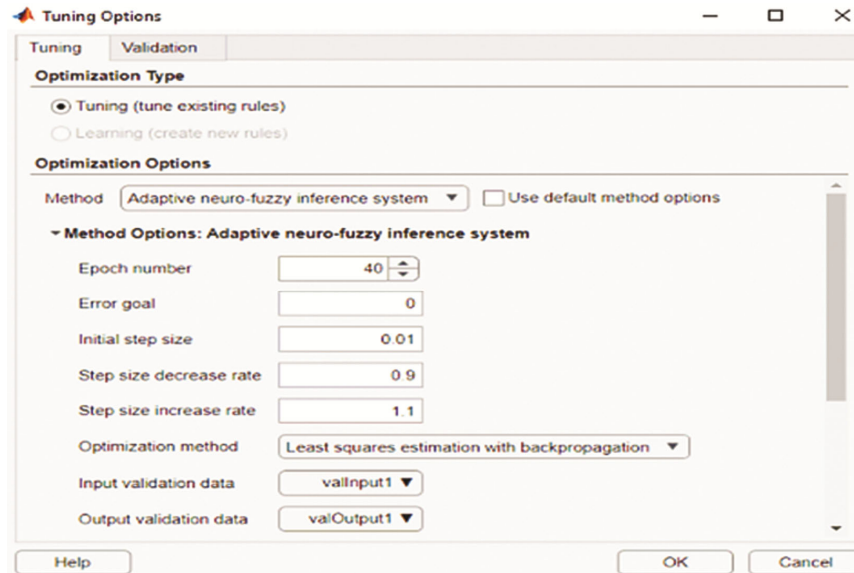


Fig. 2 — Tuning training and testing parameters using Neuro-fuzzy approach

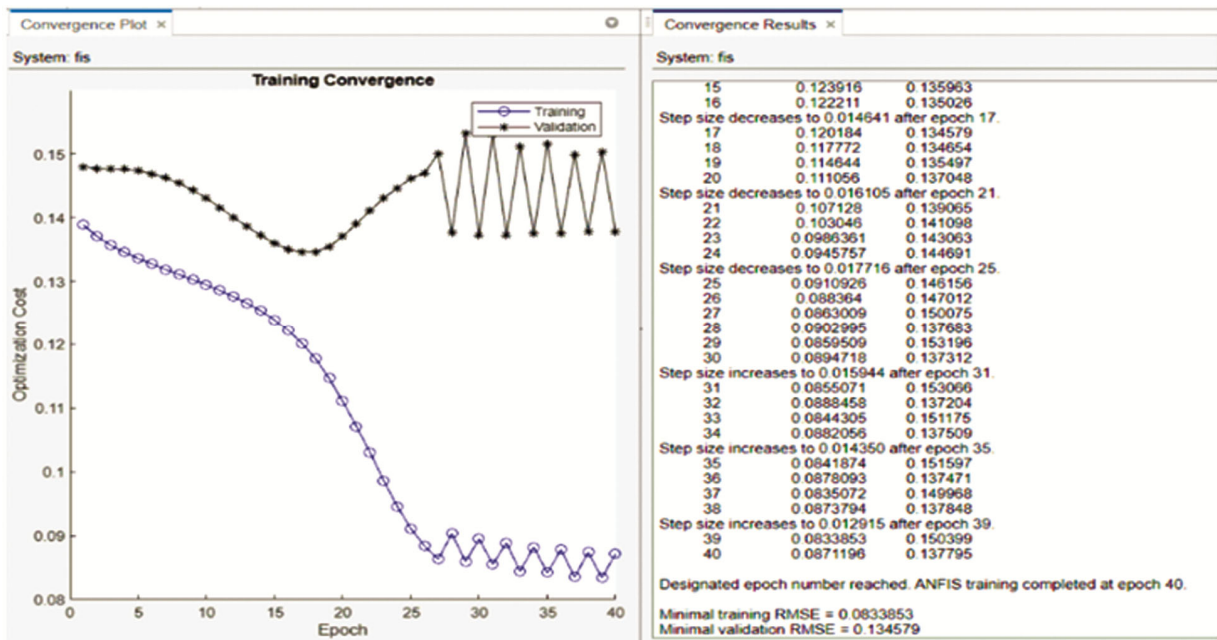


Fig. 3 — Splicing or convergence of neural network and generation of fuzzy rules

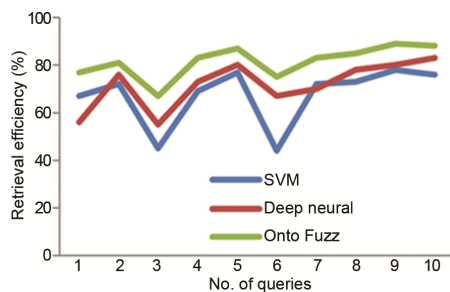


Fig. 4 — Information retrieval efficiency of the combined systems

Table 3 — Comparative analysis of the proposed model with conventional models and recent studies

| Approaches/Models | Precision (%) | Recall (%) | f-measure (%) |
|--|---------------|------------|---------------|
| Krishnaraj <i>et al.</i> ¹² | 65.67 | 64.33 | 66.78 |
| Babu <i>et al.</i> ¹³ | 68.23 | 69.55 | 69.76 |
| Ubaidullah <i>et al.</i> ¹⁴ | 72.44 | 73.55 | 74.67 |
| Vajpayee & Hossain ¹⁵ | 78.48 | 79.12 | 78.66 |
| SVM-based system ²⁶ | 76.45 | 75.34 | 77.65 |
| Deep neural network ²⁷ | 82.34 | 84.23 | 84.61 |
| Proposed OntoFuzz system | 85.67 | 86.54 | 85.49 |

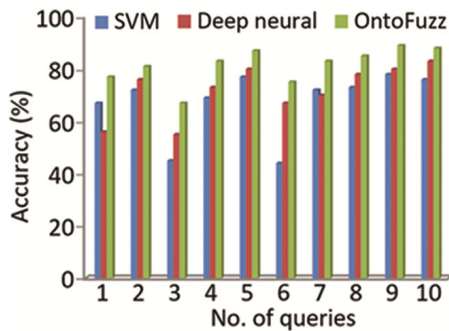


Fig. 5 — Accuracy of combined systems based on the number of queries

OntoFuzz system is having an accuracy of around 77–89% thereby retrieving more relevant results.

Conclusions

Validated on CloudSim using an annual rainfall dataset from the Kaggle repository. It is compared with traditional Support Vector Machine (SVM) information retrieval systems and deep neural network systems based on evaluation metrics including precision, recall, f-measure, and accuracy. The results found to be better. However, Optimization of hyper-parameters involved in neuro-fuzzy approach may involve higher processing time which lead to overlapping of fuzzy rules and distorted results. Sometimes, the integration of ontology with the cloud server provider requires higher application cost for the users. As a future scope, ontology can be designed using ontology editors like PROTÉGÉ, HOZO, etc. and the results from the designed ontology are required to produce inferences based on semantic web rule language and ontology web language. In this way, the results from enhanced ontology are added to the Software-as-a-Service (SaaS) layer of the cloud.

References

- Liu J, Kong X, Zhou X, Wang L, Zhang D, Lee I, Xu B & Xia F, Data mining and information retrieval in the 21st century: A bibliographic review, *Comput Sci Rev*, **34** (2019) 1–15, <https://doi.org/10.1016/j.cosrev.2019.100193>.
- Zhu Y, Yan E & Song Y, A natural language interface to a graph-based bibliographic information retrieval system, *Data Knowl Eng*, **3** (2017) 73–89, <https://doi.org/10.1016/j.datak.2017.06.006>.
- Kulunchakov A S & Strijov V, Generation of simple structured information retrieval functions by a genetic algorithm without stagnation, *Expert Syst with Appl*, **85** (2017) 221–230, <https://doi.org/10.1016/j.eswa.2017.05.019>.
- Datta D, Varma S, Chowdary R & Singh S K, Multimodal retrieval using mutual information based textual query reformulation, *Expert Syst Appl*, **68** (2017) 81–92, <https://doi.org/10.1016/j.eswa.2016.09.039>.
- Tajeddine R, Gnilke O W & Rouayheb S E, Private information retrieval from MDS coded data in distributed storage systems, *IEEE Trans Inf Theory*, **64(11)** (2018) 7081–7093, doi: 10.1109/ISIT.2016.7541531.
- Yang H, Shin W & Lee J, Private information retrieval for secure distributed storage systems, *IEEE Trans Info Forensics Security*, **13(12)** (2018) 2953–2964, doi: 10.1109/TIFS.2018.2833050.
- Djenouri Y, Belhadi A, Viger P F & Lin J C W, Fast and effective cluster-based information retrieval using frequent closed itemsets, *Info Sci*, **453** (2018) 154–167, <https://doi.org/10.1016/j.ins.2018.04.008>.
- Angelini M, Fazzini V, Ferro N, Santucci G & Silvello G, CLAIRE: A combinatorial visual analytics system for information retrieval evaluation, *Info Proces Mgmt*, **54(6)** (2018) 1077–1100, <https://doi.org/10.1016/j.ipm.2018.04.006>.
- Raj & Jennifer S, Efficient information maintenance using computational intelligence in the multi-cloud architecture, *J Soft Comp Paradigm*, **1(2)** (2019) 113–124, doi: 10.36548/jscp.2019.2.006.
- Joby D P P, Expedient information retrieval system for web pages using the natural language modeling, *J Artif Intell*, **2(2)** (2020) 100–110, doi: 10.36548/jaicn.2020.2.003.
- Chandrika S U & Perumal T M, Modified ECC for Secure data transfer in multi-tenant cloud computing, *Int J Comput Net Info Sec*, **6** (2022) 76–88, doi: 10.5815/ijcnis.2022.06.06.
- Krishnaraj N, Elhoseny M, Lydia E L, Shankar K & Dabbas O A L, An efficient radix trie-based semantic visual indexing model for large-scale image retrieval in a cloud environment, *Soft: Pract Exp*, **51(3)** (2021) 489–502, <https://doi.org/10.1002/spe.2834>.
- Babu B S & Jothi K R, Secure deduplication with dynamic updates in multi-tenant cloud environment, *Proc Int Conf Adv Comput Tech Appl* (IEEE) 2022, 1–4, doi: 10.1109/ICACTA54488.2022.9752987.
- Kashif U L, Memon Z A, Ghanghro S A, Channa W A & Soomro A, Centralized accessibility of vm for distributed trusted cloud computing, *Proc Int Conf Comput, Math Eng* (IEEE) 2023, 1–6, doi: 10.1109/iCoMET57998.2023.10099351.
- Vajpayee P & Hossain G, Multi-tenant cloud security- risk prediction through cyber-value-at-risk (CVaR), *Proc Int Conf on Digital Forensics and Security* (IEEE) 2024, 1–7, doi: 10.1109/ISDFS60797.2024.10527323.
- Singh G & Jain V, Information retrieval (IR) through Semantic Web (SW): An overview, *Proc Int Conf on Confluence* (IEEE) 2012, 23–27.
- Yadav U, Narula G S, Duhan N & Jain V, Ontology engineering, and development aspects: A survey, *Int J Edu Manag Eng*, **6 (3)** (2016) 9–19, <https://doi.org/10.5815/ijeme.2016.03.02>.
- Carpineto C & Romano G, A survey of automatic query expansion in information retrieval, *ACM Comput Surv*, **44(1)** (2012) 1–50, <https://doi.org/10.1145/2071389.2071390>.

- 19 Pal D, Mitra M & Datta K, Improving query expansion using wordnet, *J Assoc Inf Sci Technol*, **65(12)** (2014) 2469–2478, <https://doi.org/10.1002/asi.23143>.
- 20 Zingla M A, Chiraz L & Slimani Y, Short query expansion for microblog retrieval, *Procedia Comput Sci*, **96** (2016) 225–234, <https://doi.org/10.1016/j.procs.2016.08.135>.
- 21 Carpineto C, Romano G & Giannini V, Improving retrieval feedback with multiple term-ranking functions combination, *ACM Trans Inf Syst*, **20(3)** (2002) 259–290, <https://doi.org/10.1145/568727.568728>.
- 22 Chen F, Fang H & Wang M, Exploiting entity relationship for query expansion in enterprise search, *Information retrieval*, **17(3)** (2014) 265–294, <https://doi.org/10.1007/s10791-013-9237-0>.
- 23 Sundas A & Panda S N, An introduction of cloudsims simulation tool for modelling and scheduling, *Proc Int Conf on Emerg Smart Comput and Informat* (IEEE) 2020, 263–268, doi: 10.1109/ESCI48226.2020.9167549.
- 24 <https://matlab.mathworks.com/> (20 May 2024)
- 25 <https://www.kaggle.com/datasets/> (20 May 2024)
- 26 Shukla A K & Sharma A, Cloud base intrusion detection system using convolutional and supervised machine learning, *Proc Int Conf on Info Syst and Compu Netw* (IEEE) 2023, 1–5, doi: 10.1109/ISCON57294.2023.10112007.
- 27 Yan H, Yu P & Long D, Study on deep unsupervised learning optimization algorithm based on cloud computing, *Proc Int Conf on Intelligent Transportation, Big Data & Smart City* (IEEE) 2019, 679–681, doi: 10.1109/ICITBS.2019.00168.