

# Enhanced Neuro-Fuzzy-Based Crop Ontology for Effective Information Retrieval

K. Ezhilarasi<sup>1,\*</sup> and G. Maria Kalavathy<sup>2</sup>

<sup>1</sup>Computer sceince and Engineering, Anna university, Chennai, 600025, India
<sup>2</sup>Computer sceince and Engineering, St. Joseph's College of Engineering, Chennai, 600119, India
\*Corresponding Author: K. Ezhilarasi. Email: kezhilarasi2021@yahoo.com
Received: 18 May 2021; Accepted: 26 July 2021

Abstract: Ontology is the progression of interpreting the conceptions of the information domain for an assembly of handlers. Familiarizing ontology as information retrieval (IR) aids in augmenting the searching effects of user-required relevant information. The crux of conventional keyword matching-related IR utilizes advanced algorithms for recovering facts from the Internet, mapping the connection between keywords and information, and categorizing the retrieval outcomes. The prevailing procedures for IR consume considerable time, and they could not recover information proficiently. In this study, through applying a modified neuro-fuzzy algorithm (MNFA), the IR time is mitigated, and the retrieval accuracy is enhanced for trouncing the above-stated downsides. The proposed method encompasses three phases: i) development of a crop ontology, ii) implementation of the IR system, and iii) processing of user query. In the initial phase, a crop ontology is developed and evaluated by gathering crop information. In the next phase, a hash tree is constructed using closed frequent patterns (CFPs), and MNFA is used to train the database. In the last phase, for a specified user query, CFP is calculated, and similarity assessment results are retrieved using the database. The performance of the proposed system is measured and compared with that of existing techniques. Experimental results demonstrate that the proposed MNFA has an accuracy of 92.77% for simple queries and 91.45% for complex queries.

**Keywords:** Ontology; crop ontology; information retrieval (IR); *k*-medoids algorithm; neuro-fuzzy algorithm (NFA); modified NFA (MNFA)

# **1** Introduction

The information on the webspace is augmenting at a vast speed with the progressive improvement in information technology. With dynamic research being conducted for over 30 years, only information retrieval (IR) has become omnipresent with the World Wide Web initiation. The retrieval of information proficiently and precisely has become highly imperative [1]. Although most IR systems depend on ontologies, they recurrently practice one of two subsequent extreme methodologies [2]; either they utilize the maximum of the ontology semantic expressiveness and, hence, necessitate intricate query languages,



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

which are inapplicable for nonspecialists; or they proffer a simple query language, which almost mitigates the ontology to a wordlist of synonyms utilized in Boolean retrieval replicas [3]. IR is introductory in ontology for indexing and recovery. The information formulation phase intends to derive sequential data and their connected happenings from a script [4]. An ontology comprises various components, such as classes, individuals, relations, attributes, functions, axioms, and restrictions. Many languages are available to build ontologies [5]. An ontology epitomizes a general comprehension of a domain, where data semantics is machine understandable. Here, ontology functions on metadata and accelerates the semantic matching functionality to the search engine with transaction and incorporation of knowledge [6,7]. In contemporary days, domain ontology acts as the pillar of the Semantic Web through proffering vocabularies, and formal conceptualization is provided for assisting the allocation and altercation of domain information [8,9]. It can heighten the web function by refining the exactitude of web searches [10]. Regarding the transfiguration of words to a connotation, the crucial concern is to ascertain the apt conceptions that elucidate and recognize documents and the language applied in request by the user. The ontology usage to mitigate the precincts of keyword-related search has been heralded as one of the stimuli of the Semantic Web from its occurrence in the late 1990s. Web ontology language (OWL) was premeditated for highly intricate class structures and properties [11]. Enormous assistances occurred in the preceding years; however, most accomplishments were either fractional usage of the complete expressive power in ontology-related knowledge representation or Boolean retrieval-based models, thereby lacking an apt ranking model for measuring massive information sources [12].

Farmers require information on seasonal weather, seeds, fertilizers, and best cultivars, as well as erudition on pests and diseases, controlling techniques, harvesting and postharvesting techniques, precise market costs, and present supply and demand, to make learned judgments at innumerable stages of the agricultural cycle [13]. Farmers articulate their demands in an accepted language that is generally responded by human specialists. In such a condition, a gap might exist between the astuteness of farmers and agriculture specialists, which is perplexing for farmers in selecting the apt knowledge amid several adoptions. This problem must be addressed for capturing knowledge in a system, which comprehends a query and proffers the pre-eminent solution for farmers' concern. Therefore, an idea evolves for developing a system that can fill such a gap and establishing an exclusive system to address farmers' problems [14].

The foundational agricultural crop ontology comprises information that is general to every sort of crop. This information comprises production practices, postproduction practices, environmental data, varieties, cropping systems, botanical description, and origin [15]. Such information is considerably valuable, particularly to farmers, for developing their production in consideration of fluctuating conditions and circumstances [16]. This agriculture information is promulgated on the Internet in divergent formats of relational databases, XML, RSS, webpages, and others.

Here, an ontology-related IR system specifically for the agriculture domain is proposed. The objectives are to request a database with an IR system, to assure exactness and reliability in outcomes, and to present an ontology-centered, rapid, and effectual IR system. Ontologies refer to the modeling techniques that possess the power of formally representing domain knowledge. They have a significant role in annotating and organizing considerable experimental, clinical, and real-world data, and their everyday usage is well recognized in the scientific community [17]. Thus, such techniques should be improved to enrich information and increase the accuracy performance to ensure that users will obtain the relevant information when searching for provided information. The system can succor users by mitigating agricultural inputs and abridge the response time while providing an acceptable solution for their search queries.

The rest of this paper is organized as follows. Section 2 evaluates the related work regarding the proposed technique. Section 3 presents a concise confab about the proposed work. Section 4 assesses the investigational outcomes. Lastly, Section 5 concludes the paper.

## 2 Related Work

A paper for publication should be divided into multiple sections, including a title and full names of all authors. Ming et al. [18] recommended a methodology for hastening semantic object search and vegetable trading information detection by utilizing a Steiner tree (ST). Sequences of ontology construction methodologies were established through exploration in accordance with domain ontology for vegetable transaction facts. Jena2 proffered a rule-related reasoning engine. The results indicated that the recall and precision rate of an ontology-related IR system were much superior to those of a keyword-related IR system and presented few concrete values. However, the IR's efficiency was extremely low.

Rajendran et al. [19] introduced a multilevel object relational similarity (MORS)-related image retrieval algorithm. Manifolds were educated through extricating the objects and supporting the labeling with the administered learning procedure. Concerning the MORS value, a solo class was acknowledged [20]. Regarding the acknowledged semantic class, the upshot was raised. The presented method augmented the image-mining performance; nonetheless, it rendered a high false ratio in the retrieval procedure [21].

Sayed et al. [22] proposed an Ontological Search Engine termed Ibri College of Applied Sciences Engine Ontology (IBRI-CASONTO) for the Academies of Applied Sciences, Oman. This engine assisted Arabic and English languages [23]. It engaged two sorts of exploration: a keyword-related search and a semantic-related search. IBRI-CASONTO was regarded a divergent technology of Resource Description Framework data along with ontological graph. Thus, it rendered some unrelated data in a big dataset [24].

Selvalakshmi et al. [25] suggested a semantic IR system that employed feature selection and classification for ameliorating the relevancy score (RS). An intelligent fuzzy rough set-centered feature selection algorithm and an intelligent ontology and Latent Dirichlet Allocation-centered semantic IR algorithm were employed for IR [26]. Outcomes exhibited that the system augmented the RS to 98%. Nevertheless, the main downside of this system was that it was unsuitable in a big-data environment.

## 3 Modified Fuzzy Algorithm and Crop Ontology-Based IR

For achieving rapid and efficacious IR, a reformed neuro-fuzzy-centered crop ontology system is recommended. Three strides are encompassed by the proposed work: i) development of a crop ontology, ii) implementation of the IR system, and iii) processing user query. The first progression of the proposed method is dataset creation, in which the appropriate information concerning farmers and their crops is accumulated from various resources. Next, these data are reprocessed through restructuring and repositioning them into a more understandable manner in the Apache Jena Fuseki database. Then, a crop ontology evaluation. The derived files are again saved in the Apache Jena Fuseki database. Subsequently, IR system implementation is executed by performing the following operations on the dataset: establishing closed frequent patterns (CFPs), hash code generation, and applying a modified neuro-fuzzy algorithm (MNFA) for IR. Afterward, CFPs are established for the data values. Thereafter, hash values are created for all CFPs by using the Secure Hash Algorithm (SHA) 512 algorithm. Lastly, a hash tree is generated concerning the hash values for CFPs. For the information recovery step, MNFA is applied, in which *k*-medoids clustering comprises neuro-fuzzy stratification. The architecture of the proposed technique is presented in Fig. 1.

# 3.1 Development of a Crop Ontology

This phase comprises three main steps. They are knowledge acquisition, ontology development, and ontology evaluation. First, the information regarding each crop and their diseases and the solution for the diseases are accumulated from the website. This sort of data collection for crops is known as knowledge acquisition. Once knowledge acquisition is done, the crop ontology is advanced through creating an OWL file, which comprises massive facts about the accumulated crops. Centered on the similarity of semantics between the indexed data and the user query, the ontology-centered IR system retrieves data. Consequently, only the pertinent data are retrieved during the process of information recovery, and the recovery period is decreased. After the OWL file creation, the file is envisaged. For file creation and visualization, the proposed system utilizes Protégé that is inbuilt with the Eclipse IDE platform. Protégé is a free, open-source ontology editor and a knowledge management system. It proffers a graphic user interface for delineating ontologies. Similar to Eclipse, Protégé is a structure in which different projects advocate plug-ins. Such an application is inscribed in Java and utilizes Swingheavily for creating the user interface. The generated crop ontology in the recommended system by utilizing the Protégé tool is demonstrated in Fig. 2.



Figure 1: Proposed architecture for effective IR

Then, the evaluation of the ontology phase is accomplished by abstracting the values from the OWL file formed in the foregoing stage and utilizing the reasoning in the Protégé tool. The reasoning is the task of deriving implicit facts from a set of proffered explicit facts. The derived details are stored in the Apache Jena Fuseki dataset.

### 3.1.1 Dataset Creation

The initial phase of the proposed technique is data collection. Here, specifics of crops and farmers are collected from divergent resources and kept as a dataset. The dataset comprises two sorts of information. They are the details of farmers and crops. The farmers' details comprise farmer's name, address, contact number, and other information; the crops' details comprise the locality of the paddy field, yield, the duration of beginning and ending, paddy type, and the length of paddy growth.

#### 3.1.2 Preprocessing

Preprocessing of the proposed method encompasses reorganizing and the prearrangement of the farmers' and crops' details. During this phase, after performing the analysis on farmers' data, the following data are removed from the dataset:

- a) Ambiguous information
- b) Incomplete data
- c) Irrelevant data
- d) Noisy data

The final filtered dataset is kept in the Apache Jena Fuseki dataset.



Figure 2: Crop ontology by utilizing the Protégé tool

## 3.2 Execution of the IR System

The execution of the IR system consists of three processes: creating CFPs, hash code generation, and MNFA application. First, CFPs are established from derived values (crops' and farmers' details). Next, the hash code for all CFPs is created by utilizing the SHA 512 algorithm, and a hash tree is created concerning the hash values for every CFP. A hash value is principally utilized in the hash tree for indexing. Every leaf node has the CFPs indexed and a matching hash value. Lastly, MNFA is utilized for powerful IR. Each of the progressions is elucidated in the following sections.

#### 3.2.1 Constructing CFPs

To detect the CFPs in the dataset, frequent patterns (FPs) are established. The patterns indicate the number n of data in the dataset. From the FPs, the CFPs are found. "n" number of forms in the dataset are expressed as

$$D_s = \{P_1, P_2, \dots, P_n\}$$

$$\tag{1}$$

where  $D_s$  signifies the dataset, and  $P_n$  denotes the number of forms in the dataset.

Afterward, the FPs from the dataset are computed. The FPs of the dataset refer to the addition of occurrences of a precise form on the dataset. They are also acknowledged as the count of patterns or the frequency of patterns.

 $F_1 = P_1, P_2$  $F_2 = P_3, P_4$  $\dots$ 

. . . . . . . . . . . .

$$F_n = P_{n-1}, P_n \tag{2}$$

where  $F_n$  indicates the number of FPs.

$$C(\Delta P) = C(P_1), C(P_2), C(P_3), \dots, C(P_n)$$
(3)

where  $C(P_n)$  symbolizes the CFPs of the dataset.

## 3.2.2 Hash Code Generation

After CFP discovery, the hash value for every CFP is created by utilizing the SHA 512 algorithm. The SHA 512 algorithm is a hash algorithm that utilizes a one-way hash function. This algorithm is an advanced version of prevailing hash algorithms called SHA 0, SHA 1, SHA 256, and SHA 384 algorithms. The SHA 512 hash function collects the input data of any size and creates a message digest of 512-bit size and 1024-bit block length. First, message bits are amplified with extra bits for forming a multiple of 1024 bits. Next, this block is split into smaller parts of 1024 bits. The chief block is integrated with the initializing vector, and the hash code is created. The consequent blocks are integrated with the formerly generated hash codes. Afterward, one hash tree is initiated, and the hash code values obtained are traversed to the hash tree's leaf node. Whether the leaf node is full is ascertained. If not, then the hash values are inserted to the hash tree is built using the formed values of hash linked with the CFPs. Hash values are utilized to index in the hash tree. Each leaf node signifies the CFPs indexed with an associated hash value.

#### 3.2.3 MNFA Application

All CFPs from leaf nodes are provided as the input to MNFA. The recommended MNFA is an amalgamation of two practices. They are KMA and NFA. KMA is combined with NFA to enhance the significance of IR in the proposed work. Thus, the proposed system of IR is termed as MNFA. In MNFA, initially, the disarranged CFPs are clustered *via* KMA. Next, the clustered CFPs are provided as the input to NFA. In NFA, rules are created, and hash values with specific CFPs are tested concerning the generated rules. Here, NFA is first reformed through a clustering method by utilizing KMA. Therefore, the recommended NFA is termed MNFA. The phases in MNFA are explicated as follows.

KMA functions in two phases: build and swap. First, k centrally located objects are sequentially chosen and regarded as the first medoid. Next, KMA examines the ensuing condition. The targeted function is mitigated by substituting (swapping) a certain medoid with a nonchosen object, then the swap is implemented. This process is repeated until the targeted function is not mitigated. The KMA steps are elucidated below.

Step 1: k number of random points, such as the medoids from the itemized n CFPs of the dataset, are chosen.

**Step 2:** Each data point of the closest medoid utilizing any of the distance metrics is associated. The distance  $(dist_i)$  between every pair of all CFPs centered on the selected dissimilarity measure is computed.

$$dist_i = \sqrt{\sum_{i=1}^{n} CFP_{1i} - CFP_{2i}}; i = 1, 2, ..., n$$
(4)

where  $CFP_{1i}$  and  $CFP_{2i}$  denote the "two" data points of  $CFP_i$ . The cluster centroids  $s_i$  are calculated as

$$s_i = \sum_{i=1}^n \frac{CFP_i}{\sum_{c=1}^n CFP_{ic}}$$
(5)

where  $CFP_{ic}$  is the distance between data points *i* and *c*. *J* CFPs having the first *j* smallest values are selected as initial medoids. The initial cluster outcome is acquired by assigning every CFP value to the nearest medoid.

Step 3: The cost of the configuration decreases.

For each selected medoid  $S_m$  and nonselected object  $NS_m$ ,

 $S_m$  and  $NS_m$  are swapped, each data point of the closest medoid is related, and the total swapping cost  $TS_c$  (total of distances of points to their medoids) is recalculated.

If  $TS_c < 0$ ,  $S_m$  is replaced with  $NS_m$ .

If  $TS_c$  of the configuration is augmented in Step 3.1, then the swap is undone.

Step 4: The present medoid in every cluster is updated *via* replacement with a new medoid. Every object is allocated to the closest medoid, and the result of the cluster is obtained.

Step 5: The distance from every CPF to their medoid is summed. The algorithm is stopped if the sum is equivalent to the former one. Otherwise, Step 2 is repeated. Lastly, the attained K number of clusters is expressed as

 $FC_i | i = 1, 2, ...K$ 

The pseudo code of the *k*-medoid algorithm is presented in Fig. 3.

```
Input: The number of closed frequent patterns CF<sub>n</sub>
Output: K - clusters \{FC_i | i = 1, 2...K\}
Begin
         Randomly select K data points from CF_n as mediods Sm_1, Sm_2, \dots, Sm_n
        Repeat
           Assign remaining non-selected medoids NS<sub>m</sub> to its closest medoid
           Compute total distance of FC_i, TS<sub>c</sub> between S<sub>m</sub> NS<sub>mi</sub>
            for each mediod S_m do
                Select non-medoid NS_{m_i} for which TS_c is minimal
                Compute TS_c between S_m NS_{m_i}
                if TS_c Ns_m \rightarrow S_m < \text{current } TS_c
                  Swap S_m and NS_{m_i}
                End if
           End for
   Until no S_m changes;
End
```

Figure 3: Pseudo code for *k*-medoid algorithm

(6)

The hash code values are grouped, and the KMA outcomes are integrated into the neuro-fuzzy system in IR by following the above given steps of KMA. A neuro-fuzzy system is basically a fuzzy system that utilizes the learning algorithm motivated by the theory of neural network for determining its parameters (fuzzy sets and rules) through data sample processing. A neuro-fuzzy system is always elaborated as a system of fuzzy rules. It likely generates the system through training data from scratch, given that it is feasible to commence it by using former knowledge in the practice of fuzzy rules. Such systems are generally characterized as distinct multilayer feed-forward neural networks. The NFA structure contains five layers. Here, the hash value-clustered data for a specific CFP resulted from the former step are the inputs to the first layer of NFA. In "five" layers, the first and fourth layers have adaptive nodes, whereas the other layers comprise fixed nodes. The information concerning a farmer or crop is precisely retrieved by utilizing NFA. The two elementary rules of NFA are stated in the following equations.

**Rule 1:** If  $H_0$  is  $A_i$  and  $H_1$  is  $T_i$ , then

$$R_i = x_i H_i + y_i H_{i+1} + z_i (8)$$

**Rule 2:** If  $H_0$  is  $A_{i+1}$  and  $H_1$  is  $T_{i+1}$ , then

$$R_{i+1} = x_{i+1}H_i + y_{i+1}H_{i+1} + z_{i+1}$$
(9)

where  $A_i$ ,  $T_i$ ,  $A_{i+1}$ , and  $T_{i+1}$  denote the fuzzy sets.  $H_0$  and  $H_1$  present the divergent clustered hash values derived from KMA.  $x_i$ ,  $y_i$ ,  $z_i$ ,  $x_{i+1}$ ,  $y_{i+1}$ , and  $z_{i+1}$  values are the parameter set. The layers in NFA are elucidated as follows:

**Layer 1:** This layer is termed the fuzzification layer. Every node in this layer is an adaptive node with a function.

$$L_{1,i} = \mu_{A_i}(H_i)$$
(10)

where  $H_i$  is the input to node *i*. Every node acclimatizes to a function parameter. The output of each node is a grade of membership value that is given by the membership function (MF) input. The MF utilized in NFA is the bell MF, as indicated in the succeeding equation.

$$\mu_{A_i}(H_i) = \frac{1}{1 + \left(H_0 - \frac{z_i}{x_i}\right)^{2y_i}} \tag{11}$$

where  $x_i$ ,  $y_i$ , and  $z_i$  are the MF parameters, which may deduce the figure of the MF. The parameters are indicated as the premise parameters.

Layer 2: Each node in this layer is an immobile node marked as X, whose output is the total of all incoming signals.

$$L_{2,i} = B_i = \mu_{A_i}(H_i) \times \mu_{A_i}(H_{i+1})$$
(12)

The throughput of this layer  $L_{2,i}$  indicates the firing strength of a rule.

**Layer 3:** Every node in this layer is a firm node marked as K. All these nodes have a purpose, that is, computing the ratio of the  $i^{th}$  rule's FS to the total of all the rules' FSs. The outcome is marked as the normalizing FS. The mathematical delineation of the formulations is elucidated as follows:

$$L_{3,i} = \bar{B}_i = \frac{B_i}{B_i}, i = 1, 2....6$$
(13)

For accessibility, the throughput of this layer is termed normalized firing strength.

Layer 4: All nodes in this layer are adaptive nodes having functions.

$$L_{4,i} = \bar{B}_i \cdot R_i \tag{14}$$

where  $\overline{B}_i$  suggests the standardized FS based on the prior layer, and *Rules<sub>i</sub>* indicates the rule of the system. The employed parameters are labeled succeeding parameters.

**Layer 5:** The solo node of this layer is a firm node called NFA, which totals the complete output as the sum of all incoming signals. In this layer, the circle node is termed

$$\sum L_{5,i} = \sum_{i} \bar{B}_{i} R_{i} = \frac{\sum_{i} B_{i} R_{i}}{\sum_{i} B_{i}}$$
(15)

### 3.3 Processing of User Query

After the execution of the IR system, user query processing is performed by utilizing MNFA. The query of the user by using the Semantic Web search engine is provided as an input. The testing procedure of the proposed method is similar to the training process. Here, the user input query is preprocessed, and the CFPs for the input query are established. Next, the SHA 512 algorithm is accomplished on CFPs for deriving hash values, and a hash tree is created concerning the hash values that are created. Then, the CFP hash values of the input query are equated with the trained database, and the result is retrieved by utilizing MNFA. An assessment of work is accomplished by utilizing the PageRank (PR) algorithm, given that the work is completed at the center of the Semantic Web search engine.

## **4** Result and Discussion

The recommended IR methodology utilizing MFNA is applied in the Java working platform.

#### 4.1 Performance Assessment

In this subsection, the performance of the proposed MNFA is equated with that of the prevailing techniques of IBRI-CASONTO and ST regarding precision, recall, F-score, accuracy, returned *vs.* effective information, retrieved results, and query retrieval time. Such measures are matched for divergent kinds of input queries, such as simple and complex queries. The acquired outcomes of the recommended MNFA and prevalent methods are presented in Tab. 1.

| Performance<br>metrics (%) | Simple queries   |       |                  | Complex queries  |        |                  |
|----------------------------|------------------|-------|------------------|------------------|--------|------------------|
|                            | IBRI-<br>CASONTO | ST    | Proposed<br>MNFA | IBRI-<br>CASONTO | ST     | Proposed<br>MNFA |
| Precision                  | 52.49            | 95.2  | 96.56            | 50.12            | 93.562 | 94.45            |
| Recall                     | 75               | 80.8  | 82.77            | 73.56            | 78.34  | 81.45            |
| F-score                    | 68.54            | 74.32 | 84.45            | 66.34            | 72.34  | 83.44            |
| Accuracy                   | 45               | 67    | 92.77            | 45               | 67     | 91.45            |

Table 1: Performance outline of the suggested MNFA and prevailing practices for simple and complex queries

Tab. 1 presents the comparison outcomes of the suggested MNFA and prevailing IBRI-CASONTO and ST for uncomplicated and intricate queries regarding precision, recall, F-score, and accuracy. For simple and intricate queries, the precision values of the recommended MNFA are 96.56 and 94.45, respectively. By contrast, IBRI-CASONTO and ST provide 52.49 and 95.2 for simple queries and 50.12 and 93.56 for

intricate queries, respectively. The accuracy of MNFA is 92.77 for uncomplicated queries and 91.45 for intricate queries. On the contrary, IBRI-CASONTO and ST derive the accuracy values of 45 and 67 for simple and intricate queries, respectively. For the continuous measures, namely, recall and F-score, MNFA attains the highest value compared with IBRI-CASONTO and ST. In all measures, IBRI-CASONTO achieves extremely poor performance. ST proffers good performance comparable with MNFA. Both prevailing techniques' performance is low. In sum, the proposed MNFA proffers superior performance for uncomplicated and intricate queries.

# 4.2 Performance Analysis for Simple Queries

The demonstration of the proposed MNFA and prevailing methods regarding precision, recall, F-score, and accuracy for simple queries is designed in Fig. 4.



**Figure 4:** Performance assessment of the suggested MNFA and prevailing procedures for simple queries. (a) Recall (b) Precision (c) F-Score and (d) Accuracy

Fig. 4 presents the recovery presentation of the recommended MNFA scheme and prevailing IBRI-CASONTO and ST schemes regarding (a) precision, (b) recall, (c) F-score, and (d) accuracy. Here, MNFA proffers the uppermost values for precision and accuracy compared with the others. For recall and F-measure, the suggested MNFA proffers 82.77 and 84.45, respectively, whereas the prevailing IBRI-CASONTO and ST obtain 75 and 80.8 for recall and 68.54 and 74.32 for F-score, respectively. Overall, MNFA acquires the most exceptional outcomes for the IR system. The presentation of MNFA and existing methods regarding retrieved outcome and query retrieval duration is plotted in Fig. 5.

Fig. 5 demonstrates the performance of the suggested MNFA and prevailing IBRI-CASONTO and ST concerning the measures of (a) returned *versus* effective information, (b) retrieved outcomes, and (c) query retrieval time. From 30 records, the recommended MNFA returns 28 records, of which 26 records are dynamic. By contrast, IBRI-CASONTO and ST return 23 and 27 records, of which 17 and 23 records are dynamic, respectively. With MNFA, the system recovers 92% of information about crops, whereas the

recovered information of IBRI-CASONTO and ST are 73 and 85, respectively. Thus, MNFA derives the utmost retrieval performance. For query retrieval, MNFA takes 9887 ms, whereas IBRI-CASONTO and ST take 12450 and 11670 ms, respectively. IBRI-CASONTO consumes more time in retrieving the query. MNFA consumes minimal time in retrieving the input query compared with the prevailing IBRI-CASONTO and ST. Hence, the overall performance of MNFA is greater.





### 4.3 Performance Investigation for Intricate Queries

A comparison of the proposed MNFA and prevailing IBRI-CASONTO and ST on intricate queries based on precision, recall, F-score, and accuracy is demonstrated in Fig. 6.

Fig. 6 presents the retrieval flow of the recommended MNFA scheme and prevailing IBRI-CASONTO and ST schemes regarding (a) precision, (b) recall, (c) F-score, and (d) accuracy. MNFA proffers the highest values for precision and accuracy compared with others. For recall and F-measure, the proposed MNFA proffers 81.45 and 83.44, respectively, whereas the prevailing IBRI-CASONTO and ST derive 73.56 and 78.34 for recall and 66.34 and 72.34 for F-score, respectively. In all measures, the entire flow of IBRI-CASONTO is excessively slow. ST and MNFA present comparatively good performance. However, the recommended MNFA attains the highest outcomes for every metric. The performance of the proposed MNFA and prevailing techniques regarding returned *vs.* effective information, retrieved outcomes, and query retrieval duration is elucidated in Fig. 7.

Fig. 7 presents that from 30 records, the proposed MNFA returns 26 records, where 24 records are effective. By contrast, the prevailing IBRI-CASONTO and ST return 24 and 28 records, of which only 16 and 22 records are efficient, respectively. The retrieved information level of the suggested MNFA is

92%, whereas those of the prevailing IBRI-CASONTO and ST are only 66% and 78%, respectively. Thus, MNFA provides the greatest retrieval outcomes. For the query retrieval duration MNFA takes 13964 ms, whereas the prevailing IBRI-CASONTO and ST take 18984 and 16687 ms, respectively. MNFA consumes minimal time in retrieving the input query compared with IBRI-CASONTO and ST. In sum, from the graph mentioned above, MNFA provides the maximum percentage of retrieved outcomes and takes minimal time to retrieve the query.



Figure 6: CASONTO and ST schemes (a) Precision (b) Recall (c) F-score (d) Accuracy



**Figure 7:** Assessment graph of MNFA and prevailing IBRI-CASONTO and ST. (a) Returned Vs Effective Information and (b) Retrieved Results Percentage

#### **5** Conclusions

In this paper, an ontology-related IR system is proposed using MNFA. Three methodologies are included in this system: training, testing, and assessment. After the three phases are performed, the results of the proposed MNFA are compared with those of the prevailing technologies of IBRI-CASONTO and ST. Here, the performance valuation of the proposed and existing methods is accomplished for two sorts of queries: simple and complex queries. For both sorts of queries, the valuation is made in consideration of precision, recall, F-score, accuracy, returned *vs.* effective information, retrieved outcomes, and query retrieval duration. The proposed MNFA proffers an accuracy of 92.77% for simple queries and 91.45% for complex queries. It retrieves 92.77% and 92% of the information for simple and intricate queries, respectively. The time needed by the proposed MNFA for retrieving the query is 9887 ms for simple queries and 13964 ms for intricate queries. Hence, the recommended MNFA attains superior performance for all aforesaid performance metrics and is comparable with prevailing techniques for simple and intricate queries is comparable with prevailing techniques for simple and intricate queries is comparable with prevailing techniques for simple and intricate queries for all aforesaid performance metrics and is comparable with prevailing techniques for simple and intricate queries. In the future, the proposed work can be extended by incorporating natural language processing to lessen the retrieval time.

Funding Statement: The authors received no specific funding for this study.

**Conflicts of Interest:** The authors declare that they have no conflicts of interest to report regarding the present study.

#### References

- U. Shah, T. Finin, A. Joshi, R. S. Cost and J. Matfield, "Information retrieval on the semantic web," in *Proc. ICIKM*, Chongqing, China, pp. 461–468, 2002.
- [2] S. Kumar, R. K. Rana and P. Singh, "Ontology based semantic indexing approach for information retrieval system," *Int. Journal of Computer Applications*, vol. 49, no. 12, pp. 14–18, 2012.
- [3] F. Mohameth, S. Ranwez, J. Montmain, A. Regnault, M. Crampes *et al.*, "User centered and ontology based information retrieval system for life sciences," *BMC Bioinformatics*, vol. 13, no. 1, pp. 1–12, 2012.
- [4] P. Basile, A. Caputo, G. Semeraro and L. Siciliani, Time event extraction to boost an information retrieval system. *Book: Information Filtering and Retrieval*. Studies in Computational Intelligence, vol. 668. Springer, Cham, pp. 1–12, 2017.
- [5] K. Navjot and H. Aggarwal, "Query reformulation approach using domain specific ontology for semantic information retrieval," *Int. Journal of Information Technology*, 2020.
- [6] R. Bansal and S. Chawla, "Design and development of semantic web-based system for computer science domain-specific information retrieval," *Perspectives in Science*, vol. 8, no. 4, pp. 330–333, 2016.
- [7] D. H. Deepenti and S. D. Deshpande, "A review of ontology based information retrieval," *Int. Journal of Advance Research in Computer Science and Management Studies*, vol. 1, no. 7, pp. 263–265, 2013.
- [8] S. Ruban, K. Tendolkar, A. P. Rodrigues and S. Niriksha, "An ontology-based information retrieval model for domesticated plants," *Int. Journal of Innovative Research in Computer and Communication Engineering*, vol. 2, no. 5, pp. 207–213, 2014.
- [9] F. Beirade, H. Azzoune and D. E. Zegour, "Semantic query for Quranic ontology," Journal of King Saud University-Computer and Information Sciences, vol. 33, no. 6, pp. 753–760, 2021.
- [10] R. Suganyakal and R. R. Rajalaxmi, "Movie related information retrieval using ontology based semantic search," in *Proc. ICICES*, chennai, India, pp. 421–424, 2013.
- [11] M. Schiessl and M. Bräscher, "Ontology lexicalization: Relationship between content and meaning in the context of information retrieval," *Transinformação*, vol. 29, no. 1, pp. 57–72, 2017.
- [12] T. Mya Mya Swe, "Intelligent information retrieval within digital library using domain ontology," in *Proc. ICACS*, Yangon Myanmar, 2011.

- [13] A. I. Walisadeera, G. N. Wikramanayake and A. Ginige, "An ontological approach to meet information needs of farmers," in *Proc. ICCSA*, Ho Chi Minh City, Vietnam, pp. 228–240, 2013.
- [14] M. D. Titiya and V. A. Shah, "Ontology building and reasoning process for resource description framework data," *Int. Journal of Engineering Sciences & Research Technology*, vol. 6, no. 12, pp. 43–56, 2017.
- [15] M. Sini, V. Yadav, J. Singh, V. Awasthi and T. V. Prabhakar, "Knowledge models in agropediaindica," FAO, Rome (Italy). Knowledge Exchange and Capacity Building Div. eng, 2009.
- [16] R. Alfred, K. O. Chin, P. Anthony, P. W. San, L. M. Tan et al., "Ontology-based query expansion for supporting information retrieval in agriculture," in Proc. ICKMO, Taiwan, pp. 299–311, 2014.
- [17] I. Harrow, R. Balakrishnan, E. Jimenez Ruiz, S. Jupp, J. Lomax et al., "Ontology mapping for semantically enabled applications," *Drug Discovery Today*, vol. 24, no. 10, pp. 2068–2075, 2019.
- [18] M. Zhao, T. Tengyang and X. Duanmu, "Fast semantic object search and detection for vegetable trading information using Steiner tree," *Artificial Intelligence Review*, vol. 41, no. 3, pp. 415–427, 2014.
- [19] T. Rajendran and T. Gnanasekaran, "Multi level object relational similarity based image mining for improved image search using semantic ontology," *Cluster Computing*, vol. 22, pp. 1–8, 2018.
- [20] M. Abramovici, P. Gebus, J. Christian Göbel and H. B. Dang, "A semantic information retrieval framework within the scope of IPS2-PLM," in *Proc. CIRP*, Linkoping, Sweden, vol. 47, pp. 294–299, 2016.
- [21] J. Ingram and P. Gaskell, "Searching for meaning: Co-constructing ontologies with stakeholders for smarter search engines in agriculture," NJAS - Wageningen Journal of Life Sciences, vol. 90-91, no. 3, pp. 100300, 2019.
- [22] A. Sayed and A. I. Muqrishi, "IBRI-CASONTO: Ontology-based semantic search engine," *Egyptian Informatics Journal*, vol. 18, no. 3, pp. 181–192, 2017.
- [23] R. Bansal and S. Chawla, "Design and development of semantic web-based system for computer science domain-specific information retrieval," *Perspectives in Science*, vol. 8, no. 4, pp. 330–333, 2016.
- [24] J. Lee, J. K. Min and C. W. Chung, "Effective ranking and search techniques for Web resources considering semantic relationships," *Information Processing & Management*, vol. 50, no. 1, pp. 132–155, 2014.
- [25] C. S. Kumar and R. Santhosh, "Effective information retrieval and feature minimization technique for semantic web data," *Computers & Electrical Engineering*, vol. 81, pp. 106518, 2020.
- [26] B. Selvalakshmi and M. Subramaniam, "Intelligent ontology based semantic information retrieval using feature selection and classification," *Cluster Computing*, vol. 22, no. S5, pp. 12871–12881, 2019.