Int. j. inf. tecnol.
https://doi.org/10.1007/s41870-024-01809-x

ORIGINAL RESEARCH

Developing an agriculture ontology for extracting relationships from texts using Natural Language Processing to enhance semantic understanding

Saurabh Bhattacharya[1] · **Manju Pandey**[1]

Received: 23 November 2023 / Accepted: 24 February 2024

Abstract This paper outlines a methodology for developing an agriculture ontology to extract relationships from texts using web-scraping techniques, Natural Language Processing (NLP) and Artificial Intelligence (AI). The objective of the presented approach is to offer a deeper understanding of the connections among different concepts in the agriculture industry and enhance the decision-making processes. The proposed methodology comprises utilizing web-scraping techniques to gather text data pertaining to agriculture from sources that provide agri-related information. Subsequently, the gathered data is subjected to pre-processing utilizing NLP techniques in order to eliminate any extraneous or insignificant information. Then, a range of Machine-Learning and Deep Learning techniques, specifically Linear SVM, Random Forest, Convolutional Neural Network (CNN), and Long Short-Term Memory (LSTM) network, are employed to derive significant insights from the pre-processed data. The ontology is constructed using Protégé through the identification of concepts and relationships derived from the extracted features using a rule-based methodology. The suggested approach was evaluated using a dataset consisting of articles related to agriculture. The results showcased the effectiveness of the suggested approach in creating an agriculture ontology that accurately identifies connections between concepts. Paper introduces a novel approach to creating an agriculture ontology by employing web-scraping techniques, NLP, and AI to extract semantic relationships from textual data. The proposed approach has the potential to enhance decision-making processes in agriculture by providing insights into the interrelationships among various concepts. This methodology is highly valuable for researchers and professionals in the agriculture sector and it can also be utilized in other fields to derive semantic relationships from textual data.

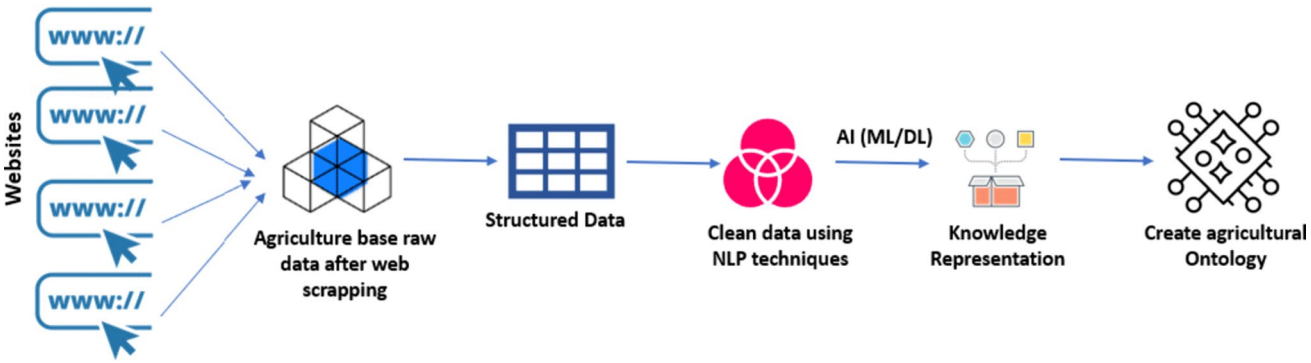
✉ Saurabh Bhattacharya
babu.saurabh@gmail.com

Manju Pandey
mpandey.mca@nitrr.ac.in

[1] Department of Computer Applications, National Institute of Technology, Raipur, CG, India

Published online: 29 March 2024

Graphical abstract



Keywords Ontology · Agriculture · Web-scrapping · Knowledge representation · Semantic extraction

1 Introduction

Ontology is a computer science field that organizes knowledge. Soft computing establishes a precise and widely accepted understanding of a domain or subject. It also represents and organizes ambiguous data.[1]. Artificial intelligence, NLP, and the semantic web use ontologies. They spread knowledge across systems and applications [2, 3]. Ontologies allow machines to reason, infer, and make decisions based on knowledge. They help represent and analyze incomplete, inconsistent, or uncertain knowledge. This is especially useful in AI, NLP, and the semantic web where data is often complex and difficult to express using mathematical models [4]. Ontologies contain concepts, classes, and objects and their relationships. Hierarchically, broader concepts are at the top and more specific concepts are at the bottom. Concept connections can be depicted as properties or attributes to set limits and regulations on their operation [5]. Ontologies help systems and applications share knowledge. A shared ontology lets systems interpret and use information meaningfully. This improves system and application interoperability and collaboration, enabling smarter, more adaptive systems.

The agriculture ontology focuses on the conceptualization and categorization of agricultural knowledge, encompassing the terminology and relationships between various concepts in the field [6–9]. The multidisciplinary approach incorporates knowledge from agriculture, computer science, and information science. It can improve agricultural system understanding and management. Big data on soil nutrient levels and crop growth is analyzed using ML/DL algorithms. These algorithms predict crop yields and nutrient needs [9, 10]. An agricultural ontology can include specialized knowledge about crops, soil, weather, and other important factors in crop production. An ontology provides a systematic representation of domain knowledge to help select and interpret sensors, improving sensor data accuracy.

An ontology serves as a conceptual framework that encompasses the constraints and interrelationships among various entities within a particular area of research. An increasingly popular method in the field of knowledge representation is the development of domain-specific ontology. The field of agriculture generates a substantial amount of data, which can be acquired in the forms of spreadsheets, tables, and textual documents [11]. Unfortunately, the lack of sufficient data processing techniques has led to the underutilization of this data. The Indian agriculture sector is widely recognized as the primary industry of the country and is responsible for employing over 50% of the nation's population [12]. Using cutting-edge technologies and innovative practices can boost sector productivity. However, this requires a well-organized and methodical knowledge representation. With an agricultural ontology, Indian farmers can easily find industry terms and concepts. It can also improve their communication and collaboration with other farmers. This knowledge can also help researchers and stakeholders communicate. An agricultural ontology can help create intelligent systems that give farmers weather and other crop-related data.

Indian farmers face many challenges using agricultural ontologies. One issue is the lack of a common language for terms and concepts. Communication breakdown and ambiguity can hinder their adoption of new agricultural methods. Failure to integrate diverse technologies and systems and lack of current information hinders new practices [13]. This hinders informed decision-making. An agricultural ontology

can address these issues by creating a common lexicon that allows farmers to use terms and concepts and share information and knowledge. It can also help create intelligent systems that make smart decisions. An agricultural ontology is needed to improve Indian agriculture. It improves farmers' communication and collaboration. It can also provide timely crop and other information. This form of knowledge representation must be developed and implemented for industry growth.

Despite the abundance of data, developing countries like India still rely on human experts and government policies for decision-making. Extracting agricultural terms is difficult due to data heterogeneity, format variations, and categorization. Farmers often use past experiences to determine crop cultivation methods, which can be risky. Thus, many farmers have crop-growing questions which can be solved by proposed approach. Rest of the paper is organized: Sect. 2 discusses Ontology literature, while Sect. 3 discusses methods used to achieve the goal. Sect. 4 covered feature extraction methods. AI-based modeling is covered in Sect. 5. Relationship extraction for ontology creation was shown in Sect. 6. Similarity indices are calculated in Sect. 7. Section-8 describes ontology creation. Section-9 discusses outputs and results. Section-10 ends the paper and discusses future plans.

2 Related works

As agricultural knowledge grows and must be stored and managed, ontologies are becoming more important. This review summarizes the literature on ontologies in agriculture and related fields. This text discusses the development and integration of these technologies in knowledge management and soil classification. In recent years, ontology-based approaches have advanced in many fields.

Muñoz et al. proposed an approach based on ontology for efficient soil selection and classification [14]. Their work laid the foundation for understanding how ontologies can contribute to improving agricultural practices. Following this, S. Harish Venu et al. explored unsupervised domain ontology learning, which allowed for greater flexibility in ontology construction [15].

Several studies have furthered the development and understanding of agricultural ontologies. Jebaraj et al. conducted an exploratory study on agriculture ontology from a global perspective [16]. Subsequently, Jonquet et al. demonstrated the power of "unified metadata in ontology" repositories using "AgroPortal" as a case study [17]. Kaladzavi et al. proposed an ontology-based architecture for co-construction of sociocultural knowledge systems [18], while Kaushik et al. explored automatic relationship extraction from "agricultural text" for "ontology construction" [19]. Deb et al. developed a framework for ontology learning from taxonomic data [20].

Agriculture ontology research has expanded to include knowledge graphs pertaining to crop pests and diseases. Xiaoxue et al. conducted a review and trend analysis of these knowledge graphs [21]. In the same year, Chukkapalli et al. discussed about ontologies and AI systems for cooperative smart farming ecosystems were discussed [22]. Ghazal et al. introduced an "intelligent role-based access control model" and framework using semantic business roles in multi-domain environments [23]. Aydin et al. developed an ontology-based data acquisition model for agricultural open data platforms and implemented the "OWL2MVC" tool [24].

Drury et al. conducted a survey of semantic web technology for agriculture, highlighting the importance of ontology-based approaches in the field [25]. Tarus et al. reviewed "ontology-based recommender systems" for e-learning, a domain that shares similarities with agriculture in terms of knowledge dissemination [26]. Rajendran et al. designed an agricultural ontology based on "levy flight distributed optimization" and "Naïve Bayes classifier" [27]. Zaman et al. presented an ontological structure for extracting data from diverse scientific sources [28], while Mughal et al. proposed an ontology-based semantic model for river flow and flood mitigation [29].

N. Kaur et al. [30] propose reformulating queries with a domain-specific ontology to improve semantic information retrieval. To improve information retrieval accuracy and relevance, the authors suggest augmenting search queries with ontology. This work improves semantic search by incorporating domain-specific knowledge structures into query reformulation. A. Thukral et al. [31] used NLP, NER, and biomedical ontologies to improve clinical narrative knowledge graphs. A method for extracting organized knowledge from unorganized clinical data improves healthcare applications. The study addresses the challenges of organizing narrative clinical text data to improve specialized healthcare analytics. A semantic knowledge graph is used to create a subject-based ontology by Ta et al. [32] An innovative ontology creation method using semantic knowledge graphs is proposed. The study develops topic-specific ontologies to improve knowledge representation and organization across fields.

Canito et al. performed a systematic review of time-constrained ontology evolution in predictive maintenance [33]. Mummigatti et al. developed a supervised ontology-oriented deep neural network for predicting soil health [34]. Ngo et al. emphasized the need for knowledge representation in digital agriculture and presented a step towards a standardized model [35]. Zulkipli et al. conducted a systematic literature review of automatic

ontology construction [36]. Yu et al. provided a survey of knowledge-enhanced text generation [37]. Bhuyan et al. systematically reviewed knowledge representation techniques in smart agriculture [38]. Mahmood et al. proposed an ecological and confined domain ontology construction scheme using concept clustering for knowledge management [39]. Lastly, Wilson et al. investigated the identification of quality characteristics for ontology-driven decision support systems, with a focus on the importance of usability in ontology design and implementation [40].

Literature shows ontology-based methods in agriculture and related fields gaining popularity and progress. Researcher have studied ontology construction, acquisition, and implementation to develop smart, contextually-aware solutions. This review highlights the importance of ontology-based methodologies in agriculture, providing valuable insights into current research and future directions. Continuing to study and integrate ontologies in agriculture should improve decision-making, knowledge management, and agricultural practices. The literature review emphasizes ontologies' role in agricultural knowledge management and consolidation. The text describes their use in various fields and discusses their potential to improve knowledge sharing and compatibility. This literature review suggests that ontology-based models could be used to create agricultural decision-making and knowledge management systems. The review highlighted the variety of methods and strategies used to develop and evaluate such systems..

3 Methodology

Agricultural practices are becoming increasingly complex and rely on various factors such as climate, soil, and crop varieties. Therefore, there is a need for a strong knowledge management system. This paper presents a methodology for constructing an agricultural ontology using web-scraped data and sophisticated Natural Language Processing (NLP) techniques. Our goal is to utilize the abundant and ever-changing information found online to build a thorough and organized knowledge repository that empowers individuals involved in the agricultural industry. The methodology consists of various distinct steps: data collection and preprocessing, feature extraction, knowledge representation, implementing DL and ontology generation algorithm 1 express the agricultural ontology development. This section provides an explanation for each stage, emphasizing the specific natural language processing techniques used and their importance in constructing a strong and informative agricultural ontology.

Algorithm 1: Creating agricultural ontology development

	Text	Crop_Name	Class
0	Brinjal or Eggplant is an important crop of su...	Brinjal	1
1	The brinjal is of much importance in the warm ...	Brinjal	1
2	The brinjal is a warm season crop, therefore s...	Brinjal	4
3	A long and warm growing season is desirable fo...	Brinjal	2
4	The brinjal plants can be grown in all types o...	Brinjal	3
...	...	...	...
766	Tobacco is primarily grown during the kharif s...	Tobacco	4
767	Tobacco cultivation requires careful managemen...	Tobacco	1
768	Tobacco is a highly profitable cash crop that ...	Tobacco	1
769	Tobacco farming can have significant environme...	Tobacco	3
770	The Indian government has implemented several ...	Tobacco	1

Fig. 1 Dataset Sample

Input:
  Web-scraped text data related to agriculture
  Domain-specific knowledge - Agriculture
Output:
  Agricultural ontology
Data Preprocessing:
  Clean and filter text ←Remove irrelevant information like whitespace, punctuation, and non-alphabetic characters.
  Lowercase text ←Normalize all text to lowercase for consistency.
  Tokenize text← Split the text into individual words or phrases (tokens).
  Stop word removal← Remove common words that add little meaning
  Stemming or lemmatization: Reduce words to their base forms for better analysis.
Feature Extraction: Apply NLP techniques
  Part-of-speech (POS) tagging← Identify the grammatical roles of words (nouns, verbs, adjectives).
  Named entity recognition (NER)← Identify and classify named entities like crop names, soil types, etc.
Use AI models:
  Train and apply various AI models
  Extract relevant features like crops, diseases, agricultural practices, temporal expressions.
  Identify relationships between features
Ontology Construction:
  Identify concepts, Define hierarchies
  Define properties, Define relationships
  Build the ontology in Protégé using the identified concepts, hierarchies, properties, and relationships.
Evaluation and Refinement:
  Assess ontology completeness and accuracy: Evaluate the constructed ontology for coverage of relevant concepts and relationships.
  Refine and iterate: Review and adjust the ontology based on expert feedback and domain knowledge if needed.

3.1 Dataset

Data are collected from various websites using web-scrapping API "Scrapy". Websites like "www.farmingindia.in" [41], "www.krishijagran.com" [42], "www.agrifarming.in" [43] etc are used to get data based on Indian climate. Data are then converted to ".CSV" file for further processing. Here 22 various crops are taken in consideration namely "Bajra", "Wheat", "Rice", "Tobacco" etc. Dataset are then categories under 4 categories 1= "Crop", 2= "Soil Nutrient", 3= "Soil Type", 4= "Season" for easier analysis and organization of the data for quick identify which category a particular piece of data falls under. Using standardized classification system, it helps to better understand relationship between various categories of data and draw more meaningful insight from dataset (see Fig. 1).

3.2 Data normalization

Normalization requires multiple steps for accurate data extraction. Lowercase all text is one option. Next, remove all word gaps. This reduces data interference. After that, periods and punctuation are removed. Eliminating Unicode characters means removing rare symbols from the language. Substitutions replace contraction short forms with full-forms to improve text consistency and readability. Eliminating meaningless "Stopwords" from the text. Alphabet-free words are excluded. These processes are essential for a precise and meaningful agriculture ontology. Table 1 represents the process of various data normalization methods.

3.3 Stemming and lemmatization

NLP uses "stemming" and "lemmatization" to derive word bases from text. Stemming removes letters to form a base word, while lemmatization reduces words to their dictionary form, the lemma. Lemmatization's main benefit is creating semantically meaningful base words. This process considers word pronunciation and does not remove all letters. The development of an agricultural ontology uses "stemming" and "lemmatization" to extract relevant and significant data about the field's diverse concepts and terminology. This method helps create a complete and accurate picture of the field's complex relationships. Table 2 shows the difference between the processing of both techniques.

4 Feature extraction

4.1 NLP and POS tagging

Text extraction requires POS tagging to extract sentence information. As shown in Table 3, it can classify and identify

Table 1 Various Data normalization techniques

Statement	Process	Output
"Rice cultivated during rabi season is also called as ?summer rice?. It is sown in the months of November to February and harvested during March to June. 9% of total rice crop is grown in this season."[43]	Lowercase	Rice cultivated during rabi season is also called as ?summer rice?. it is sown in the months of November to February and harvested during march to June. 9% of total rice crop is grown in this season
	Removal of Whitespaces	Rice cultivated during rabi season is also called as ?summer rice?. It is sown in the months of November to February and harvested during March to June. 9% of total rice crop is grown in this season
	Removal of Punctuations	Rice cultivated during rabi season is also called as summer rice It is sown in the months of November to February and harvested during March to June 9 of total rice crop is grown in this season
	Removal of Unicode Characters	Rice Farming Practices
	Substitution of Contractions	"Rice cultivated during rabi season is also called as summer rice It is sown in the months of November to February and harvested during March to June 9 of total rice crop is grown in this season"
	Removal of Stopwords	"Rice cultivated rabi season also called summer rice It sown months November February harvested March June 9 total rice crop grown season"
	Discardment of Non-alphabetic Words	"Rice cultivated during rabi season is also called as summer rice It is sown in the months of November to February and harvested during March to June of total rice crop is grown in this season"

Table 2 Stemming and Lemmatization output

	Root word	Stemming	Lemmatization
-	"Rice cultivated during rabi season is also called as ?summer rice?. It is sown in the months of November to February and harvested during March to June. 9% of total rice crop is grown in this season."	"rice cultivate during rabi season is also call as summer rice it is sown in the month of November to February and harvest during march to June 9 of total rice crop is grown in this season"	"rice cultivate during rabi season be also call as ? summer rice ?. it be sow in the month of November to February and harvest during March to June. 9% of total rice crop be grow in this season."

speech parts for naming entity recognition. Language processing relies on POS tagging to identify each word in sentences. It helps machines understand sentence structure and make more accurate predictions.

The weighted terms derived from the regular expressions are based on the assumptions shown in Table 4. Noun is preferred over other words that satisfy similar regular expressions. A verb is preferred over other terms that have similar regular expressions. High frequency words, on the other hand, are significant terms. The weights are increased for words that have multiple patterns when compared to those that only have single patterns.

4.2 Timex feature

NLP's Timex feature extracts temporal expressions from text. This feature can be used to create an agriculture ontology to identify time-related information in farming practices and crop growth documents (Tables 5, 6). Timex can improve agricultural efficiency by creating a comprehensive ontology to analyze and interpret farming data.

4.3 Analysing web-scraped text with N-grams

The three NLP concepts "Uni-gram", "Bi-gram" and "Tri-gram" used to analyze and process text. Uni-gram is one word, bi-gram is two, and tri-gram is three. An agriculture ontology can analyze the context and frequency of terms like crop name, season, soil nutrient, and crop type using these features. The ontology can represent domain relationships and knowledge by categorizing and identifying these terms by frequency and co-occurrence, as shown in Figs. 2, 3. Web scraping can collect a lot of text from agricultural websites. This creates an informative and complete agriculture ontology.

4.4 Integration of the process

Input: "Rice cultivated during rabi season is also called as ?summer rice?. It is sown in the months of November to February and harvested during March to June. 9% of total rice crop is grown in this season."

Output: rice cultivate rabi season also call summer rice month november march june rice crop grown season (see Fig. 4).

4.5 Sample output

See Table 7.

Table 3 POS tagging

Input	"Rice cultivated during rabi season is also called as ?summer rice?. It is sown in the months of November to February and harvested during March to June. 9% of total rice crop is grown in this season."
Tokens	*"['Rice', 'cultivated', 'during', 'rabi', 'season', 'is', 'also', 'called', 'as', 'summer', 'rice', 'It', 'is', 'sown', 'in', 'the', 'months', 'of', 'November', 'to', 'February', 'and', 'harvested', 'during', 'March', 'to', 'June', '9', 'of', 'total', 'rice', 'crop', 'is', 'grown', 'in', 'this', 'season']"*
Tagged Tokens	*"[('Rice', 'NNP'), ('cultivated', 'VBD'), ('during', 'IN'), ('rabi', 'NN'), ('season', 'NN'), ('is', 'VBZ'), ('also', 'RB'), ('called', 'VBN'), ('as', 'IN'), ('summer', 'NN'), ('rice', 'NN'), ('It', 'PRP'), ('is', 'VBZ'), ('sown', 'VBN'), ('in', 'IN'), ('the', 'DT'), ('months', 'NNS'), ('of', 'IN'), ('November', 'NNP'), ('to', 'TO'), ('February', 'NNP'), ('and', 'CC'), ('harvested', 'VBN'), ('during', 'IN'), ('March', 'NNP'), ('to', 'TO'), ('June', 'NNP'), ('9', 'CD'), ('of', 'IN'), ('total', 'JJ'), ('rice', 'NN'), ('crop', 'NN'), ('is', 'VBZ'), ('grown', 'VBN'), ('in', 'IN'), ('this', 'DT'), ('season', 'NN')]"*

Table 4 POS Tagging with NN-VERB count

S. No	Sentence	POS Tagging	NN Count	VERB Count
1	"*wheat main cereal crop mainly rabi (winter) Season crop India*."	('wheat', 'NN'), ('main', 'JJ') ('cereal', 'NN'), ('crop', 'NN') ('mainly', 'RB'), ('rabi', 'NN') ('winter', 'NN'),('season', 'NN') ('crop', 'NN'),('india', 'NN')	8	0
2	*"wheat cultivation suit warm damp climatic areas."*	('wheat', 'NN'), ('cultivation', 'NN'), ('suit', 'NN'), ('warm', 'JJ'), ('damp', 'NN'), ('climatic', 'JJ'), ('areas', 'NNS')	4	0
3	*"heading flowering stages, excessively low high temperatures drought harmful wheat"*	('heading', 'VBG'), ('flowering', 'VBG'), ('stages', 'NNS'), (',', ','), ('excessively', 'RB'), ('low', 'JJ'), ('high', 'JJ'), ('temperatures', 'NNS'), ('drought', 'VBD'), ('harmful', 'JJ'), ('wheat', 'NN')	1	2
4	*"use biofertilizer, rice yield 50% lower n, phosphorus (p) 32% higher chemical fertilizers"*	('use', 'NN'), ('biofertilizer', 'NN'), ('rice', 'NN'), ('yield', 'NN'), ('50', 'CD'), ('%', 'NN'), ('lower', 'JJR')	4	0
5	*"many rice growers claim ideal availability phosphorus rice soil ph less 6.5"*	('many', 'JJ'), ('rice', 'NN'), ('growers', 'NNS'), ('claim', 'VBP'), ('ideal', 'JJ'), ('availability', 'NN'), ('phosphorus', 'NN'), ('rice', 'NN'), ('soil', 'NN'), ('ph', 'NN'), ('less', 'JJR'), ('6.5', 'CD'),	5	1

Table 5 Regular Expression for Timex Feature Extraction

S. No	Timex Features	Regular Expression
1	numbers	"(^a(?=\s)|one|two|three|four|five|six|seven|eight|nine|ten|eleven|twelve|thirteen|fourteen|fifteen|sixteen|seventeen|eighteen|nineteen|twenty|thirty|forty|fifty|sixty|seventy|eighty|ninety|hundred|thousand)"
2	day	"(monday|tuesday|wednesday|thursday|friday|saturday|sunday)"
3	month	"(january|february|march|april|may|june|july|august|september| \ october|november|december)"
4	dmy	"(year|day|week|month)"
5	rel_day	"(today|yesterday|tomorrow|tonight|tonite)"
6	exp1	"(before|after|earlier|later|ago)"
7	exp2	"(this|next|last)"
8	iso	"\d+[/-]\d+[/-]\d+\d+:\d+:\d+\.\d+"
9	year	"((?<=\s)\d{4}|^\d{4})"
10	regxp1	"((\d+|("+numbers+"[-\s]?)+) "+dmy+"s? "+exp1+")"
11	regxp2	"("+exp2+" ("+dmy+"|"+week_day+"|"+month+"))"
12	regxp3_season	"(winter|summer|rainy)"

Table 6 Extracted timex feature from sentences

S. No	Sentence	Extracted Timex Features
1	*"wheat main cereal crop mainly rabi (winter) season crop India"*	Winter
2	*"ideal temperature range ideal germination wheat seed 20c 25c though seeds germinate temperature range 3.5c 35c."*	20c, 25c, 3.5c, 35c
3	*"azatobacter 2.5 kg phosphetica culture 2.5 kg trycoderma powder 2.5 kg mix 100 125 kg farm yard manure (fym) broadcast time last ploughing."*	2.5 kg, 125 kg
4	*"many farmers apply 0.5-tons n-p-k 30–10-10 per hectare day approximately 45–60 days first application, apply 0.2–0.3 tons n-p-k 40–0-0 33–0-0 per hectare"*	0.5, 30-10-10, 0.2,0.3, 40-0-0, 33-0-0
5	*"many rice growers claim ideal availability phosphorus rice soil ph less 6.5."*	6.5

Crop

	Words	Frequency
130	crop	126
275	india	88
240	grow	68
584	use	47
12	also	34
588	variety	31
470	rice	30
220	food	26
198	farmer	23
463	require	20

Soil Nutrient

	Words	Frequency
244	soil	116
223	require	70
197	potassium	63
169	nitrogen	57
191	phosphorus	54
109	growth	53
281	use	37
297	yield	36
145	level	32
86	fertilizer	30

Soil Type

	Words	Frequency
204	soil	212
85	grow	62
235	type	46
124	loam	44
178	range	41
158	ph	34
29	clay	26
241	variety	26
104	include	24
250	welldraine	22

Season

	Words	Frequency
212	season	98
43	crop	78
95	grow	58
264	winter	33
233	summer	33
115	india	29
205	require	28
207	rice	28
238	temperature	27
127	kharif	23

Bigrams Crop

	Bigrams	Frequency
281	crop grow	29
547	grow india	17
284	crop india	14
32	also use	12
145	cash crop	11
1368	use variety	10
301	crop rotation	9
152	cereal crop	9
429	farmer india	8
1358	use make	8

Bigrams Soil Nutrient

	Bigrams	Frequency
464	phosphorus potassium	43
413	nitrogen phosphorus	35
257	growth yield	25
634	soil fertility	25
486	potassium growth	18
339	level nitrogen	18
576	require level	17
284	improve soil	14
653	soil ph	13
647	soil matter	13

Bigrams Soil Type

	Bigrams	Frequency
485	soil type	43
467	soil ph	27
271	loam soil	25
571	variety soil	24
323	ph range	22
548	type include	22
187	grow variety	21
52	clay loam	16
226	include clay	15
374	range soil	14

Bigrams Season

	Bigrams	Frequency
617	season crop	19
853	winter season	18
368	kharif season	18
8	also grow	13
124	crop sow	13
122	crop require	12
721	summer season	12
417	monsoon season	10
244	grow kharif	9
252	grow season	9

Fig. 2 Uni-Gram, Bi-Gram

5 AI Based modelling

5.1 Machine learning algorithms

5.1.1 Logistic regression

Logistic regression is a statistical method that can be used to analyse a dataset in which one or more independent variables determine an outcome. A binary variable is used to measure the experimental outcomes (in which there are only two possible outcomes). It is used to predict the probability of a categorical dependent variable based on one or more predictor variables (independent variables) as shown in Eq 1

$$\hat{p}(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n)}} \tag{1}$$

5.2 Linear SVM

The linear support vector machine (SVM) is a supervised machine learning algorithm applicable to classification and regression problems. It accomplishes this by employing a

Trigrams Crop

	Trigrams	Frequency
298	crop grow india	13
308	crop india grow	5
1163	rice staple food	5
1248	source income farmer	5
666	income farmer india	4
225	consumption animal feed	4
516	food crop grow	4
1131	require management strategy	3
1179	rotation use variety	3
38	also use make	3

Trigrams Soil Nutrient

	Trigrams	Frequency
447	nitrogen phosphorus potassium	34
510	phosphorus potassium growth	18
363	level nitrogen phosphorus	17
549	potassium growth yield	15
647	require level nitrogen	15
205	fertility nutrient content	9
26	also use soil	9
860	use soil fertility	9
727	soil fertility nutrient	9
761	soil ph range	7

Trigrams Soil Type

	Trigrams	Frequency
568	soil type include	22
675	variety soil type	22
210	grow variety soil	21
536	soil ph range	19
644	type include clay	15
252	include clay loam	13
418	range soil type	13
203	grow range soil	11
474	sandy loam soil	9
63	clay loam soil	8

Trigrams Season

Fig. 3 Tri-gram

	Crop_Name	Text	normalized text plus	normalized tokens plus	Class
0	Brinjal	Brinjal or Eggplant is an important crop of su...	brinjal crop subtropic tropic name subcontinen...	[brinjal, crop, subtropic, tropic, name, subco...	1
1	Brinjal	The brinjal is of much importance in the warm ...	brinjal importance area far extensively pakist...	[brinjal, importance, area, far, extensively, ...	1
2	Brinjal	The brinjal is a warm season crop, therefore s...	brinjal warm season crop therefore frost tempe...	[brinjal, warm, season, crop, therefore, frost...	4
3	Brinjal	A long and warm growing season is desirable fo...	long grow season brinjal farming night summer ...	[long, grow, season, brinjal, farming, night, ...	2
4	Brinjal	The brinjal plants can be grown in all types o...	brinjal plant grown soil light clay	[brinjal, plant, grown, soil, light, clay]	3
...	...	...	...	...	...
766	Tobacco	Tobacco is primarily grown during the kharif s...	tobacco primarily grow kharif season start end...	[tobacco, primarily, grow, kharif, season, sta...	4
767	Tobacco	Tobacco cultivation requires careful managemen...	tobacco cultivation require management attenti...	[tobacco, cultivation, require, management, at...	1
768	Tobacco	Tobacco is a highly profitable cash crop that ...	tobacco highly cash crop provide farmer india ...	[tobacco, highly, cash, crop, provide, farmer,...	1
769	Tobacco	Tobacco farming can have significant environme...	tobacco farm impact use fertilizer pesticide w...	[tobacco, farm, impact, use, fertilizer, pesti...	3
770	Tobacco	The Indian government has implemented several ...	government policy program promote crop discour...	[government, policy, program, promote, crop, d...	1

771 rows × 5 columns

Fig. 4 Output after integration of various process

Table 7 Output obtained for creating ontology

S. no	Term-1	Relation	Term-2
1	Wheat	Grows in	Loamy and sandy soil
2	Tomato	Grows in	Well-drained fertile soil
3	Rice	Weather_condition	Hot and humid climate
4	Wheat	Weather_condition	Cool and temperate climate
5	Tomato	Weather_condition	Warm and sunny climate
6	Rice	Affected_by	*Helminthosporium oryzae*
7	Wheat	Affected_by	Fusarium head blight
8	Tomato	Affected_by	Late blight

hyperplane to classify data points. The objective is to choose a hyperplane with the largest possible distance from any point in the training set. This will increase the likelihood of correctly categorizing any newly added data points as shown in Eq. 2

$$minimum \frac{1}{2} ||w| |^2 \tag{2}$$

Subjected to $y_i\left(w^T x_i + b\right) \geq 1, i = 1, 2, 3 \ldots ..n$

5.2.1 Random forest

An ensemble learning method known as Random Forest can be used for classification, regression, and other types of tasks. It accomplishes this by first constructing a large number of decision trees during the training phase, and then outputting the class that corresponds to the mode of the classes (during classification) or the mean prediction (during regression) of the individual trees as shown in Eq. 3.

$$\hat{y}(x) = \frac{1}{B}\sum_{b=1}^{B} T_b(x) \tag{3}$$

where, $T_b(x)$ = prediction of the bth "decision tree in the random forest for input" "x" and "B" is the number of trees present in the forest.

5.3 Deep learning algorithms

Deep learning models are a class of machine learning models that use multiple layers of artificial neural networks to learn complex patterns in data. Deep learning has been successful in a variety of applications such as computer vision, natural language processing, and speech recognition.

5.3.1 LSTM

Long Short-Term Memory (LSTM) is an RNN architecture designed to remember long-term dependencies in sequential data. LSTM consist of memory cells that can forget or remember information selectively. This allows LSTM to learn patterns in sequences of data such as natural language sentences or time series data Eqs. 4–10 represent input gate, cell state, output gate and predicted probability.

The forget gate and input gate at time t:

$$f_t = \sigma\left(W_f\left[h_{t-1}, x_t\right]\right) + b_f) \tag{4}$$

$$i_t = \sigma\left(W_i\left[h_{t-1}, x_t\right]\right) + b_i) \tag{5}$$

The candidate cell state and memory cell update:

$$\widetilde{C}_t = tanh\left(W_c\left[h_{t-1}, x_t\right]\right) + b_c) \tag{6}$$

$$C_t = f_t * C_{t-1} + i_t * \widetilde{C}_t \tag{7}$$

The output gate and current hidden state:

$$o_t = \sigma(W_o[h_{t-1}, x_i] + b_o) \tag{8}$$

$$h_t = o_t * \tanh(C_t) \tag{9}$$

The predicted probability distribution:

$$y_t = softmax(W_y h_t + b_y) \tag{10}$$

The model is trained by minimizing a loss function such as cross-entropy between the predicted probability distribution and the true label of the crop type, using backpropagation through time (BPTT) algorithm.

5.3.2 GRU

Similar to LSTM, but with fewer parameters, Gated Recurrent Units (GRU) are a type of RNN architecture. GRU also use gating mechanisms to forget or remember information selectively in memory cells. GRU are easier and quicker to train than LSTM making them useful for NLP.

Let $X = x_1, x_2, \ldots ..x_n$ be the input sequence of soil nutrient measurements, and $Y = y_1, y_2, \ldots ..y_n$ be the corresponding crop types. The GRU model consists of a single GRU network that takes the input sequence X as input and generates a sequence of hidden states $H = h_1, h_2, \ldots ..h_n$; as shown in Eqs. 11–15.

The update gate and reset gate at time t:

$$z_{t=\sigma}\left(W_z\left[h_{t-1}, x_t\right] + b_z\right) \tag{11}$$

$$r_{t=\sigma}\left(W_r\left[h_{t-1}, x_t\right] + b_r\right) \tag{12}$$

The candidate hidden state at time t:

$$\widetilde{h}_t = \tanh(W_h\left[r_t * h_{t-1}, x_t\right] + b_h) \tag{13}$$

The current hidden state at time t:

$$\tilde{h}_t = (1 - z_t) + h_{t-1} + z_t * \tilde{h}_t \tag{14}$$

The predicted probability distribution:

$$y_t = softmax(W_y h_t + b_y) \tag{15}$$

where σ = "sigmoid function", tanh = "hyperbolic tangent function", * → "element-wise multiplication", W= "Weight" and b="bias matrices/vectors".

5.3.3 Bi-GRU

Bidirectional GRUs (Bi-GRU) are a variation of GRUs that process sequential data in both directions, similar to Bi-LSTM. Bi-GRUs have two GRU layers, one which processes data in the forward direction and the other in the reverse direction. Let $X = x_1, x_2, \ldots ..x_n$ be the input sequence of soil nutrient measurements, and $Y = y_1, y_2, \ldots ..y_n$ be the corresponding crop types. The Bi-GRU model consists of two GRU networks, one processing the input sequence in

the reverse direction and the other in the forward direction. The outputs of both networks are concatenated and fed into a dense layer to make the final prediction as shown in Eqs. 16–20.

The forward GRU network takes the input sequence X as input and generates a sequence of hidden states $H_f = h_{f1}, h_{f2}, \ldots ..h_{fn}$;

$$h_{fi} = GRUf(x_i, h_{f,i-1}) \tag{16}$$

where GRUf = forward GRU function.

The reverse GRU network takes the input sequence X in reverse order and generates a sequence of hidden states

$$Hb = h_{b1},\ h_{b2}, \ldots ..h_{bn} \tag{17}$$

$$h_{bi} = GRUb\left(x_{n-i+1}, h_{b,i+1}\right) \tag{18}$$

where GRUb = backward GRU function.

The final prediction is obtained by concatenating the forward and backward hidden states at each time step and feeding them into a dense layer:

$$h_i = \left[h_{fi};h_{bi}\right] \tag{19}$$

$$y_i = softmax\left(W_h h_i + b\right) \tag{20}$$

where [;] denotes the concatenation operation, W_h = *"Weight matrix"*, b = *"bias vector"*.

5.3.4 Bi-LSTM

Bidirectional LSTM (Bi-LSTM) is a modified version of LSTM that analyzes sequential data in both forward and backward directions. Bi-LSTM consists of two LSTM layers, one for processing data in the forward direction and another for processing data in the backward direction. Bi-LSTM have the ability to capture information from both the past and the future in a sequence, which makes them valuable in NLP.

Let $X = x_1, x_2, \ldots ..x_n$ be the input sequence of soil nutrient measurements, and $Y = y_1, y_2, \ldots ..y_n$ be the corresponding crop types. Bi-LSTM uses two LSTM networks, one forward and one reverse, to process the input sequence. Combining both networks' results into a dense layer yields the final prediction. $H_f = h_{f1}, h_{f2}, \ldots \ldots h_{fn}$ as depicted in Eqs. 21–24.

$$h_{fi} = LSTMf\left(x_i, h_{f,i-1}\right) \tag{21}$$

where *LSTMf* is the forward LSTM function.

The reverse LSTM network takes the input sequence X in reverse order and generates a sequence of hidden states $H_b = h_{b1}, h_{b2}, \ldots .h_{bn}$:

$$h_{bi} = LSTMb\left(x_{n-i+1}, h_{b,i+1}\right) \tag{22}$$

where *LSTMb* is the backward LSTM function.

The final prediction is obtained by concatenating the forward and backward hidden states at each time step and feeding them into a dense layer:

$$h_i = \left[h_{fi};h_{bi}\right] \tag{23}$$

$$y_i = softmax\left(W_h h_i + b\right) \tag{24}$$

where [;] denotes the concatenation operation, W_h = εw eightmatrix", b= "bias vector".

6 Relationship extraction

Relationship extraction is essential to NLP workflows. It helps identify and represent textual concept relationships. An agricultural ontology shows how fertilizers and crops are related. Table 8 shows how NLP help organize and represent agricultural industry knowledge by identifying relationships between entities. Relationship extraction can help identify the correlation between ideal crop growth conditions and fertilizer quantity.

Table 8 Textual patterns and corresponding regular expressions for term extraction

S. No	Extracted Word	Regular Expression
1	season	r'\b(?i)' + 'season' + r'\b'
2	cultivation	r'\b(?i)' + 'cultivation' + r'\b'
3	use of	r'\b(?i)' + 'use of' + r'\b'
4	systems	r'\b(?i)' + 'systems' + r'\b'
5	consumption of	r'\b(?i)' + 'consumption of' + r'\b'
6	such as	r'\b(?i)' + 'such as' + r'\b'
7	production of	r'\b(?i)' + 'production of' + r'\b'
8	hybrid	r'\b(?i)' + 'hybrid' + r'\b'
9	growth in	r'\b(?i)' + 'growth in' + r'\b'
10	cultivation of	r'\b(?i)' + 'cultivation of' + r'\b'
11	production	r'\b(?i)' + 'production' + r'\b'
12	revolution	r'\b(?i)' + 'revolution' + r'\b'
13	sector	r'\b(?i)' + 'sector' + r'\b'
14	including	r'\b(?i)' + 'including' + r'\b'
15	growth of	r'\b(?i)' + 'growth of' + r'\b'
16	millions of	r'\b(?i)' + 'millions of' + r'\b'
17	include	r'\b(?i)' + 'include' + r'\b'
18	consumption	r'\b(?i)' + 'consumption' + r'\b'
19	productivity	r'\b(?i)' + 'productivity' + r'\b'
20	sensors	r'\b(?i)' + 'sensors' + r'\b'

7 Similarity index calculation

NLP relies on the similarity index to find similarities between textual pieces. An agricultural ontology, a collection of related concepts and terms, is built using it. NLP uses word embedding and word similarity to measure text similarity. Word embedding vectorizes words. These methods can be used in agriculture to compare concepts and terms. The agricultural term relevance is determined by a high word similarity index score. Table 9 shows how "Euclidean similarity" and "Cosine similarity" are used to gather data to improve agricultural knowledge.

7.1 Euclidean similarity

Euclidean similarity is a quantitative measure that mathematically assesses the degree of similarity between two vectors. Euclidean similarity is employed in the agricultural domain to quantify the resemblance between various agricultural elements, including crop varieties, soil compositions, and fertilizers. The Euclidean similarity between two vectors can be computed using the following formula, as depicted in Eq. 25:

$$Euclidean\ Similarity\,(A,B) = \sqrt{\sum_{i-1}^{n}(A_i - B_i)^2} \tag{25}$$

where A and B = "vectors (statements) being compared", n= "no. of dimensions in the vectors".

7.2 Cosine similarity

Cosine similarity is employed metric for quantifying the similarity between two vectors. It is especially advantageous in NLP tasks for measuring the similarity between statements and categorizing text. Cosine similarity quantifies the cosine of the angle formed by two vectors, which represent the documents or texts. The cosine similarity ranges from – 1 to 1. A value of 1 signifies that the two vectors are identical, 0 indicates no similarity, and – 1 indicates complete dissimilarity between the two vectors. Cosine similarity is a useful tool in relationship extraction as it allows for the identification of the similarity between two sentences. If the cosine similarity between two sentences is high, it indicates that they are likely to be conveying a similar relationship, as depicted in Eq. 26.

$$Cosine\ similarity(s_{1,}s_2) = \frac{s_1.s_2}{|s_1| \times |s_2|} \tag{26}$$

where $s_1.s_2$ = "dot product of vectors $s_1 and s_2$", $\|s_1\|$ and $\|s_2\|$ "norms of vectors $s_1 and s_2$ respectively". Sample statements are show in statements 1 to 4.

Statement-1: "*Wheat grows well in well-drained loamy soil with good organic matter content*".

Statement-2: "*Black soil is ideal for wheat cultivation as it is rich in nutrients and retains moisture for a long time*".

Statement-3: "*Wheat requires a cool climate with moderate rainfall*".

Statement-4: "*Wheat is sown in October to November and harvested in March to April*".

8 Ontology creation

The software tool "Protégé" is utilized for the creation and administration of intricate agricultural knowledge bases. Its purpose is to generate and oversee knowledge bases that adhere to standardized properties, classes, and relationships. This facilitates comprehension and implementation

Table 9 Cosine and Euclidean Similarity

Statement 1	Statement 2	Cosine Similarity	Euclidean Similarity
"Wheat grows well in well-drained loamy soil with good organic matter content."	*"Black soil is ideal for wheat cultivation as it is rich in nutrients and retains moisture for a long time."*	0.447	1.291
"Wheat grows well in well-drained loamy soil with good organic matter content."	*"Wheat requires a cool climate with moderate rainfall."*	0.5	1.224
"Wheat grows well in well-drained loamy soil with good organic matter content."	*"Wheat is sown in October to November and harvested in March to April."*	0.96	2.828
"Black soil is ideal for wheat cultivation as it is rich in nutrients and retains moisture for a long time."	*"Wheat requires a cool climate with moderate rainfall."*	0.97	2.449
"Black soil is ideal for wheat cultivation as it is rich in nutrients and retains moisture for a long time."	*"Wheat is sown in October to November and harvested in March to April."*	0.71	2.236
"Wheat requires a cool climate with moderate rainfall."	*"Wheat is sown in October to November and harvested in March to April."*	0.7	3.162

of knowledge within their respective domain. Protégé is utilized for the purpose of collaborating and reutilizing the knowledge base, rendering it an optimal tool for the development and management of intricate agricultural ontology systems. To create an ontology for agriculture, it is important to follow a systematic process, as depicted in Fig. 5, to ensure that the ontology is informative and well-structured.

8.1 Creating various class hierarchy and Data property for agriculture ontology

The "Protégé" software is employed to establish a class hierarchy and data property hierarchy for the agriculture ontology, which was generated through diverse techniques including web scraping, NLP and AI. The class hierarchy is a graphical depiction of the fundamental concepts that are frequently employed in the field of agriculture. Users are able to arrange and classify them within the ontology. The data property hierarchy is employed to depict the diverse attributes or traits of a class within an ontology. These may encompass the title, explanation, or attributes of the class as depicted in Figs. 6, 7a. It aids in delineating the attributes and qualities of the concepts within the ontology. The Protégé software enables users to create and oversee a diverse range of classes and data properties using a graphical interface. It allows to define relationships between these properties and classes. This process can help them develop effective decision-making systems and improve the efficiency of their operations. The development of the data property and class hierarchies in Protégé is very important in order to create an agricultural ontology. These components provide a framework for categorizing and organizing the various concepts within the framework. The resulting ontology will be used to support the decisions made in the field of agriculture.

9 Results and outputs

This section presents the results of agricultural ontology for the extraction of semantic connections from texts. It

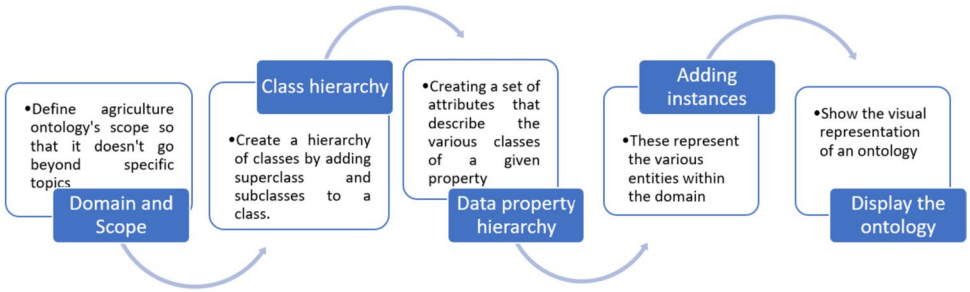


Fig. 5 Steps to create ontology in agriculture domain

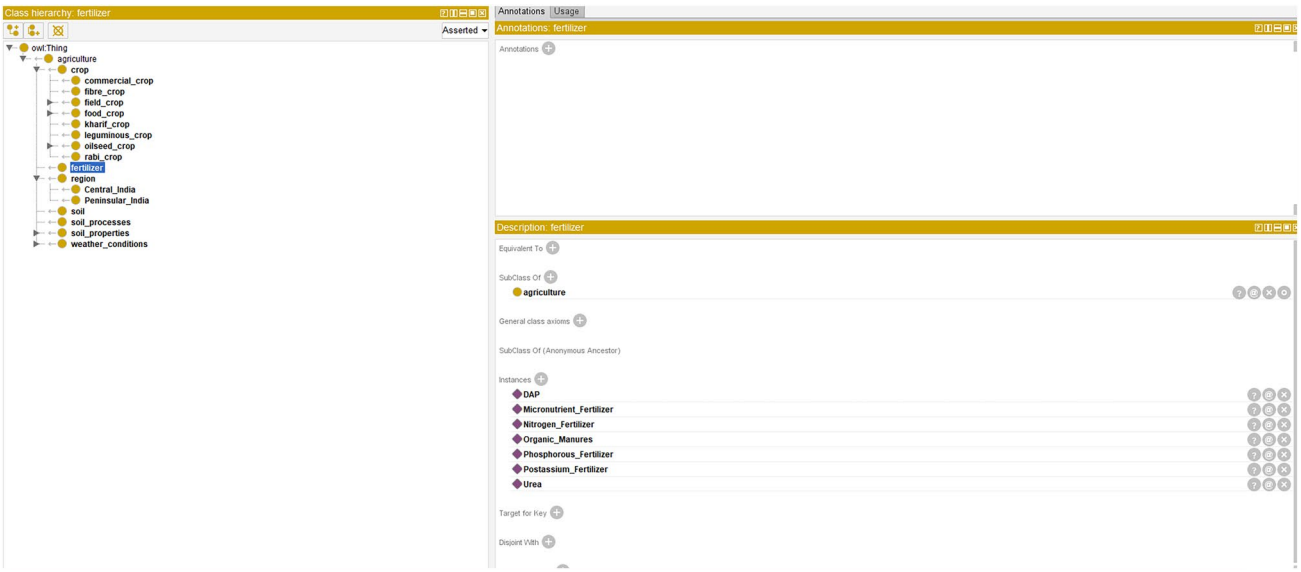


Fig. 6 Class hierarchy

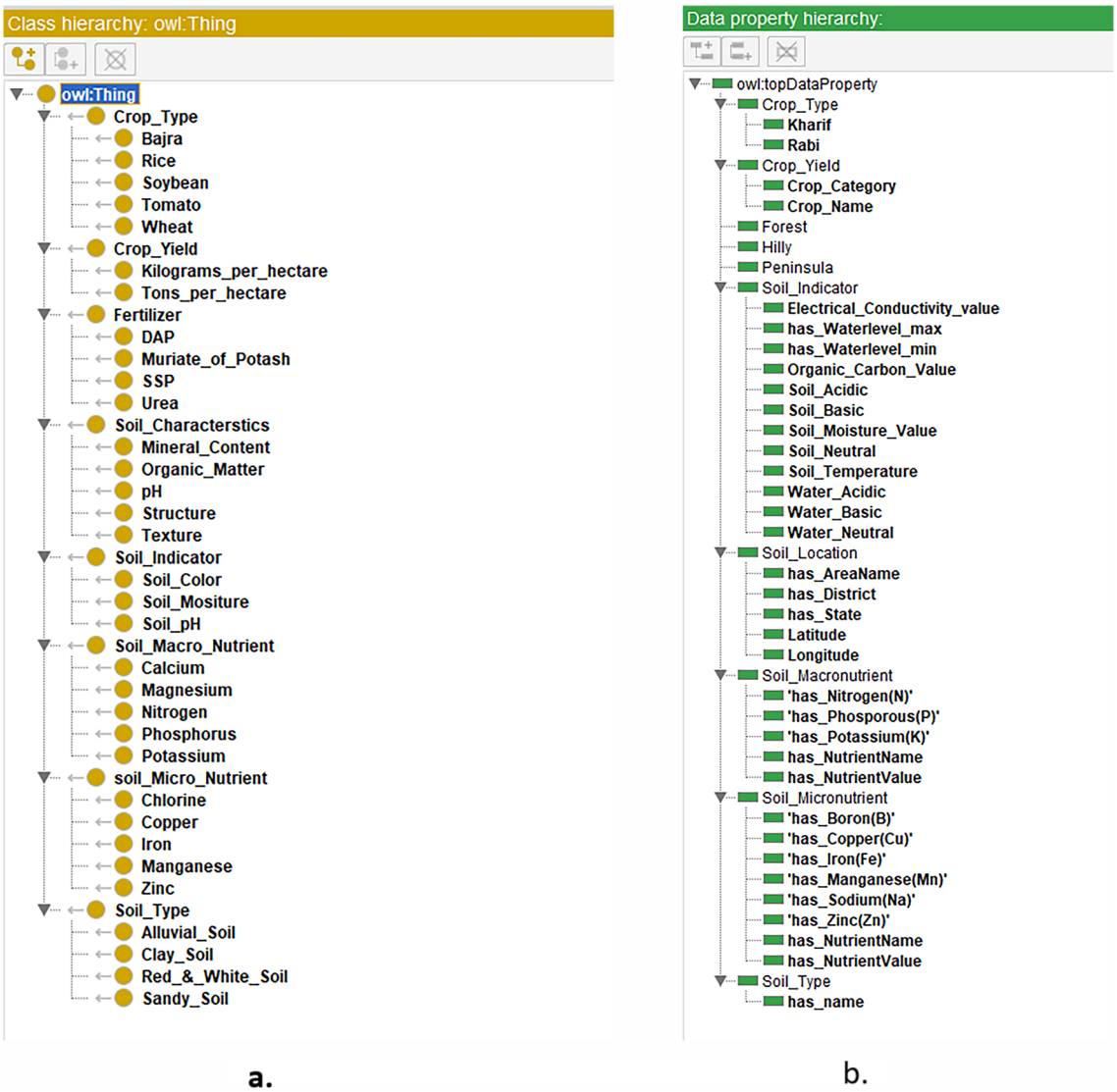


Fig. 7 A, b Detail class and data property

focuses on the use of various models, including ML/DL algorithms, as well as evaluation metrics, confusion matrix and accuracy loss graphs.

9.1 Confusion matrix

The classification capabilities of the proposed models were presented in an easy-to-follow and comprehensive manner through the utilization of the confusion matrix as shown in Fig. 8. The number of classifications displayed in the table indicated the strengths and weaknesses of each model. The matrix helped us identify the weaknesses and strengths of each proposed model. This allowed us to select the most suitable one for semantic extraction and to understand the areas where we need to improve.

9.2 Agriculture ontology visualization

The creation of an agricultural ontology is accomplished through a hierarchical structure, which included attributes, entities, and relationships as shown in Fig. 9. This

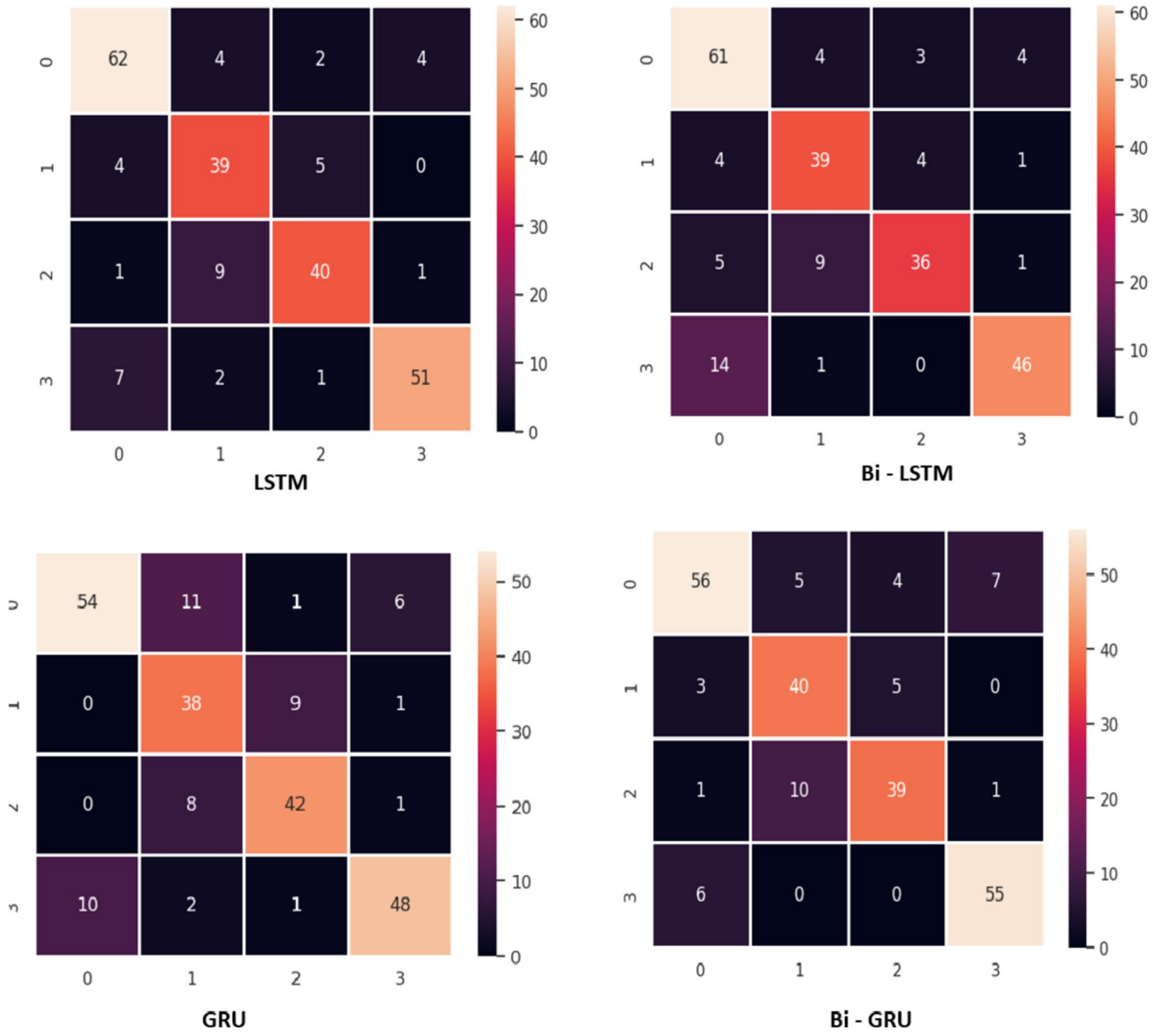


Fig. 8 Confusion Matrix for DL algorithms

represented the knowledge of the domain and allowed extraction of semantic connections within the text, which improved the way researchers can retrieve and understand agricultural data.

9.3 Evaluation metrices

The ontology specification generated offered a thorough and precise depiction of the relationships and concepts within the agricultural field. The construction process follows a rule-based approach, which includes identifying the concepts and relationships of the extracted features.

Tables 10, 11 and Figs. 10, 11 depict the various results obtained.

9.3.1 Machine Learning algorithm with N-gram

9.3.2 Deep learning algorithm with Cohen-kappa and Mathew's correlation

9.3.3 Generation of agricultural ontology

The ontology specification generated offered a thorough and precise depiction of the relationships and concepts within

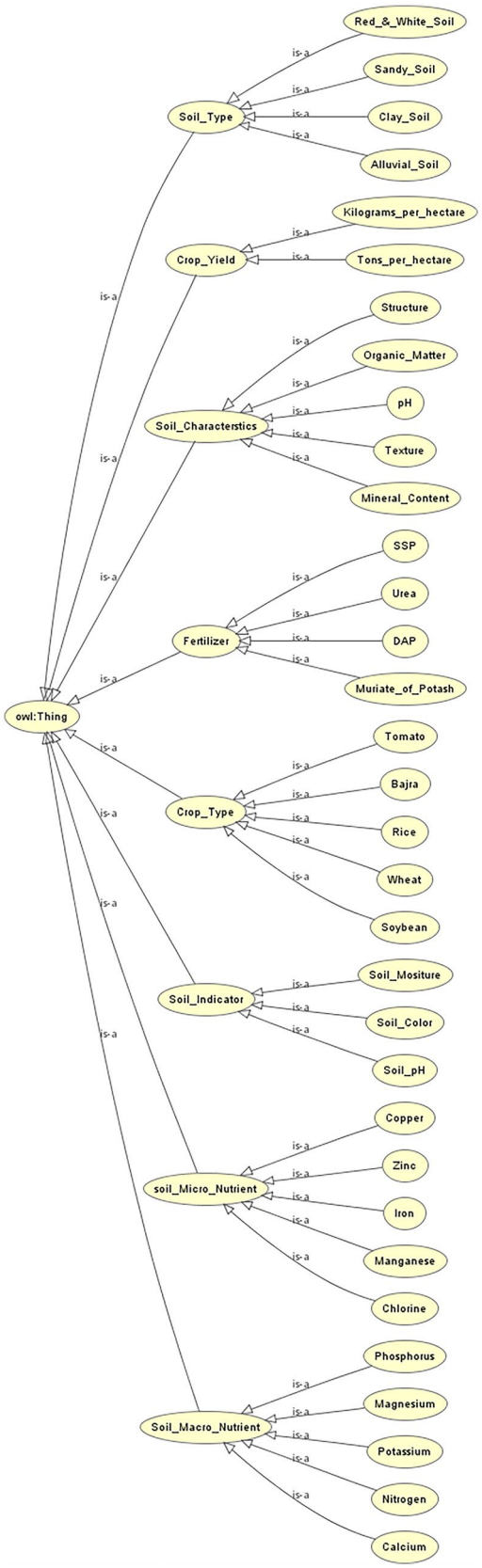


Fig. 9 Agriculture ontology

Table 10 Machine Learning Algorithm Evaluation Metrices—Accuracy

ML Algorithms	All Feature	Uni-Gram	Bi-Gram	(Uni + Bi) -Gram
		Selected Feature (25%)		
Logistic Regression	81.43	80.42	75.38	82.62
Linear SVM	82.34	80.26	73.18	81.79
Random Forest	79.56	78.29	61.32	79.61

Table 11 Evaluation metrices

DL Algorithms	Accuracy	Cohen-Kappa	Mathews- correlation coefficient
LSTM	82.7	76.82	76.89
Bi-LSTM	78.45	70.9	71.18
GRU	78.44	71.19	71.38
Bi-GRU	83.67	77.89	78.13

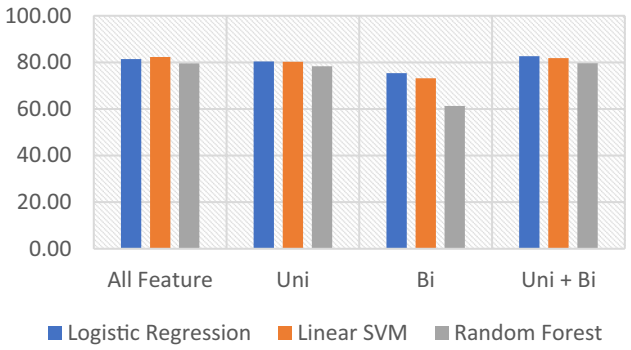


Fig. 10 Graphical Representation of ML Algorithms

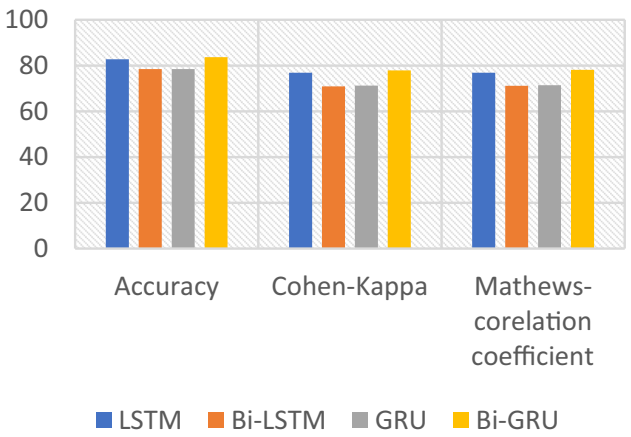


Fig. 11 Graphical Representation of Deep Learning Algorithms

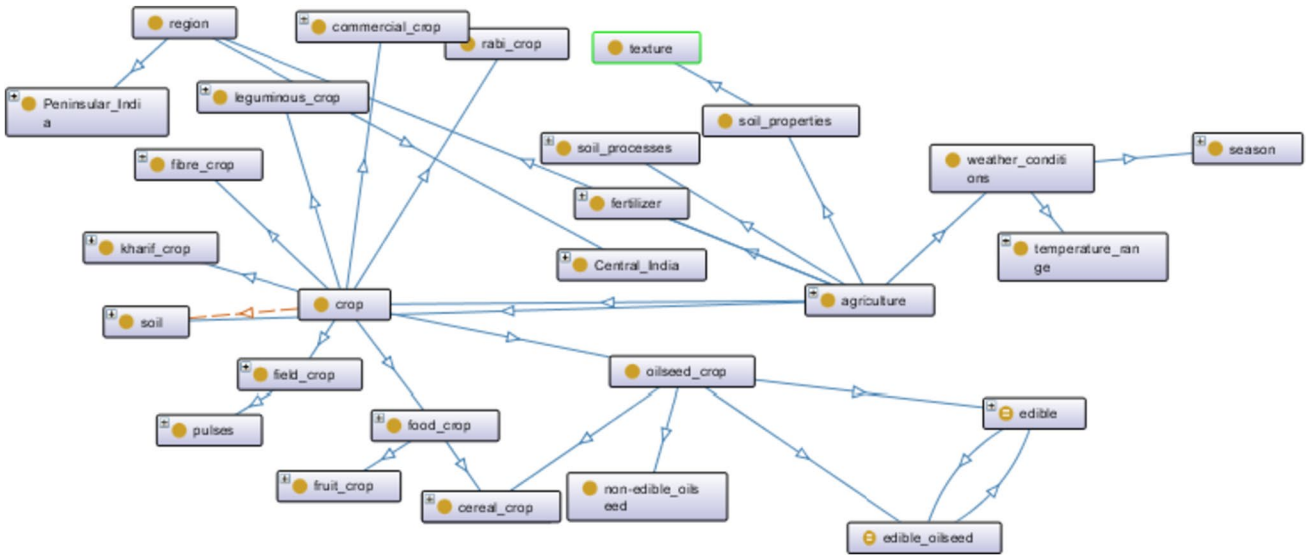


Fig. 12 Agriculture Ontology

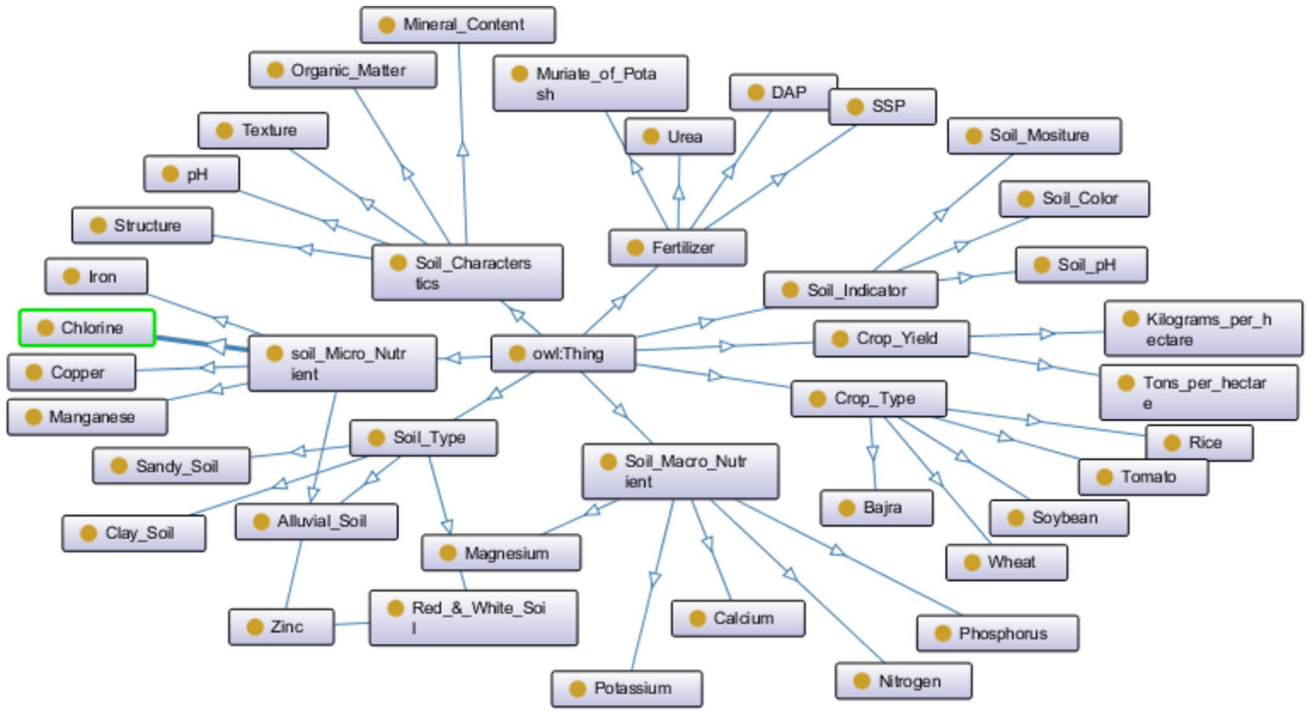


Fig. 13 Detailed View of Agriculture Ontology

the agricultural field. The construction process follows a rule-based approach, which entails identifying the concepts and relationships of the extracted features. Figures 12, 13, and 14 depict the agriculture ontology generated.

The proposed method has several novel features. The data source uses web-scraped text to collect diverse and extensive information beyond traditional agricultural sources. Natural Language Processing (NLP) techniques make it stand out for pre-processing, ensuring a more sophisticated and nuanced agricultural text analysis. The methodology uses AI models like SVM, Random Forest, CNN, and LSTM for feature extraction, making it more robust and advanced than existed ontology-based methods [9, 27]. Protégé's rule-based ontology construction distinguishes it from other models. The proposed

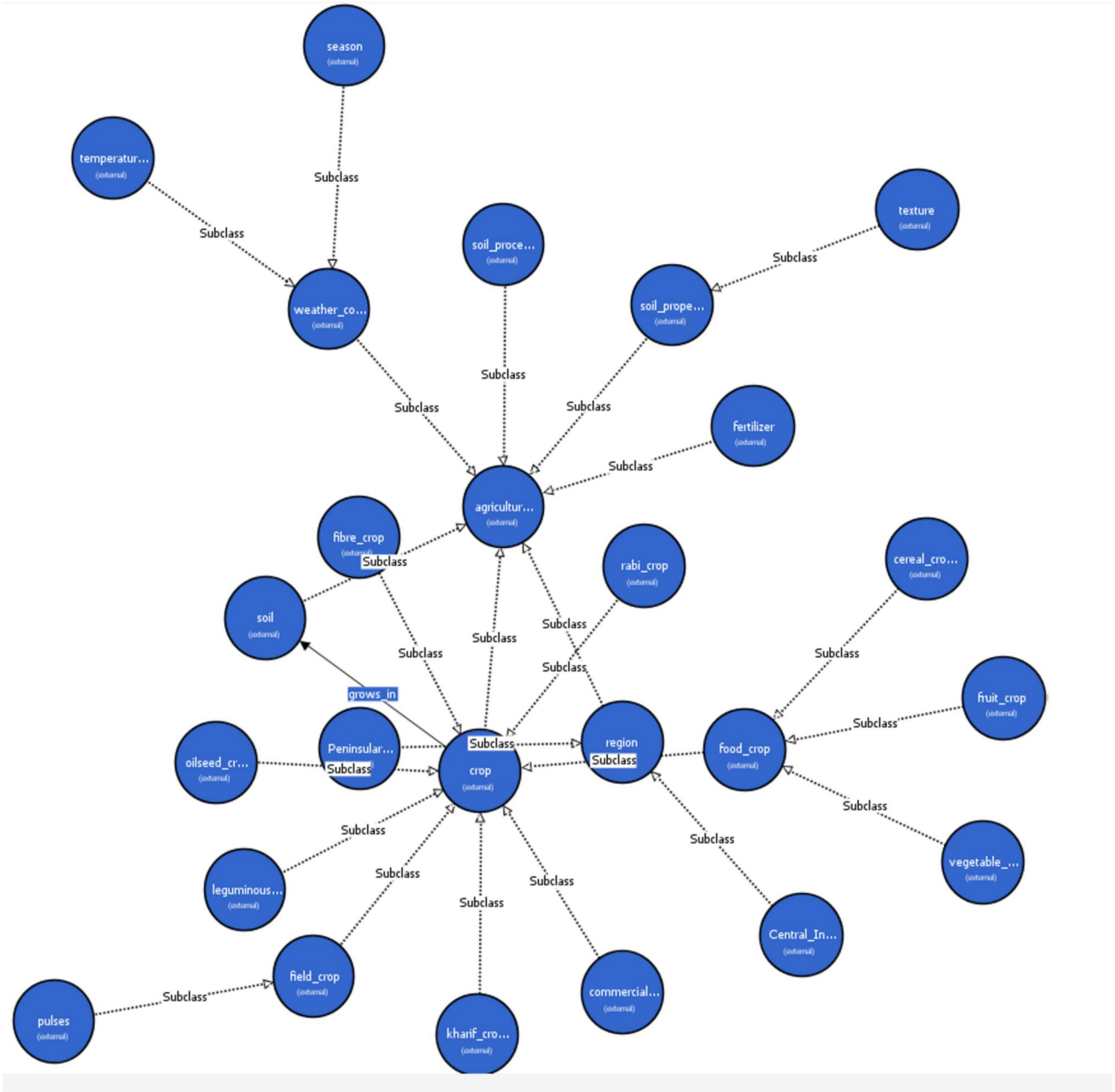


Fig. 14 Agriculture Ontology developed using Protégé

work focuses on semantic relationship extraction and insights into connections, offering a unique perspective on agricultural information. The novel approach combines web-scraping, NLP, and AI for relationship extraction with rule-based ontology construction. This method increases applicability beyond agriculture to text data with semantic relationships.

10 Conclusion and future scope

The paper proposes a novel ontology construction method for agricultural text relationships. DL, NLP and web-based methods are used. The method successfully extracts important information across concepts. The resulting ontology can benefit agricultural professionals and researchers. The proposed method is adaptable to finance, healthcare, and law, providing many applications. It can also be improved

with diverse external sources and decision support systems. Integrating new avenues improves the suggested method. It will generate a more accurate and relevant ontology. In NLP and agriculture, the proposed ontology method for extracting relationships from agricultural texts is a breakthrough. It helps researchers and decision-makers make informed sector decisions. This approach could transform multi-domain decision-making. It can provide more accurate and comprehensive knowledge to help users make informed decisions. In multilingual countries, expanding the proposed approach to multiple languages would make it more relevant. Developing language models for the NLP pipeline and translating are ways to do this. To understand agriculture's many aspects, current decision support systems can use the suggested methodology. The ontology can be used to create prognostic instruments for agricultural production decision support systems to provide timely industry insights.

Author contributions Saurabh Bhattacharya: Conceptualization, Data Collection, Methodology, Formal analysis, Paper drafting. Dr. Manju Pandey: Supervision, Conceptualization, Review and Editing.

Funding Not Applicable.

Data availability The dataset used in this study is currently undergoing further refinement and analysis. Once this process is complete, the dataset will be made publicly available on Kaggle for broader access and utilization by the research community.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Noy NF, McGuinness DL (2001) A guide to creating your first ontology. Biomed Inform Res, pp. 7–25, [Online]. Available: http://bmir.stanford.edu/file_asset/index.php/108/BMIR-2001-0880.pdf. Accessed 28 June 2023
2. Femi Aminu E, Oyefolahan IO, Bashir Abdullahi M, Salaudeen MT (2020) A Review on Ontology Development Methodologies for Developing Ontological Knowledge Representation Systems for various Domains. Int. J. Inf. Eng. Electron. Bus. 12(2):28–39. https://doi.org/10.5815/ijieeb.2020.02.05
3. Konys A (2018) Knowledge Systematization for ontology learning methods. Procedia Comput Sci 126:2194–2207. https://doi.org/10.1016/j.procS.2018.07.229
4. Al-Zoghby AM, Elshiwi A, Atwan A (2018) Semantic relations extraction and ontology learning from Arabic texts—a survey. Stud Comput Intell 740:199–225. https://doi.org/10.1007/978-3-319-67056-0_11
5. Aman SS, Agbo DDA, N'guessan BG, Kone T (2023) Design of a data storage and retrieval ontology for the efficient integration of information in artificial intelligence systems. Int J Inf Technol. https://doi.org/10.1007/s41870-023-01583-2
6. Prabhu SM (2021) Transforming India's Agricultural Sector using Ontology-based Tantra Framework. arXiv Prepr. arXiv2102.04206, 2021, [Online]. Available: https://arxiv.org/abs/2102.04206
7. Al-Aswadi FN, Chan HY, Gan KH (2020) Automatic ontology construction from text: a review from shallow to deep learning trend. Artif Intell Rev 53(6):3901–3928. https://doi.org/10.1007/s10462-019-09782-9
8. Nismi Mol EA, Santosh Kumar MB (2022), *Review on knowledge extraction from text and scope in agriculture domain*, no. 0123456789. Springer Netherlands
9. Murali E, Anouncia SM (2022) An ontology-based knowledge mining model for effective exploitation of agro information. IETE J Res. https://doi.org/10.1080/03772063.2022.2058629
10. Khadir AC, Aliane H, Guessoum A (2021) Ontology learning: Grand tour and challenges. Comput Sci Rev 39:100339. https://doi.org/10.1016/j.cosrev.2020.100339
11. Lopes AG, Carbonera JL, Schimidt D, Abel M (2022) Predicting the top-level ontological concepts of domain entities using word embeddings, informal definitions, and deep learning[Formula presented]. Expert Syst Appl 203(October 2021):117291. https://doi.org/10.1016/j.eswa.2022.117291
12. Deepa R, Vigneshwari S (2022) An effective automated ontology construction based on the agriculture domain. ETRI J 44(4):573–587. https://doi.org/10.4218/etrij.2020-0439
13. Sharma A, Vora D, Shaw K, Patil S (2023) Sentiment analysis-based recommendation system for agricultural products. Int J Inf Technol. https://doi.org/10.1007/s41870-023-01617-9
14. Muñoz A, Soriano-Disla JM, Janik LJ (2017) An ontology-based approach for an efficient selection and classification of soils. Intell Environ. https://doi.org/10.3233/978-1-61499-796-2-69
15. Venu SH, Mohan V, Urkalan K (2017) Unsupervised domain ontology learning from text. 30(April): 13–23. https://doi.org/10.1007/978-3-319-58130-9.
16. Jebaraj J, Sathiaseelan JG (2017) An exploratory study on agriculture ontology: a global perspective. Int J Adv Res Comput Sci Softw Eng 7(6):202–206. https://doi.org/10.23956/ijarcsse/v7i6/0148
17. Jonquet C et al (2018) AgroPortal: A vocabulary and ontology repository for agronomy. Comput Electron Agric 144(Ocotber 2017):126–143. https://doi.org/10.1016/j.compag.2017.10.012
18. Kaladzavi G et al (2018) Ontologies-based architecture for socio-cultural knowledge co-construction systems. Online J Appl Knowl Manag 6(1):226–239. https://doi.org/10.36965/ojakm.2018.6(1)226-239
19. Kaushik N, Chatterjee N (2018) Automatic relationship extraction from agricultural text for ontology construction. Inf Process Agric 5(1):60–73. https://doi.org/10.1016/j.inpa.2017.11.003
20. Deb CK, Marwaha S, Arora A, Das M (2018) A framework for ontology learning from taxonomic data. Adv Intell Syst Comput 654:29–37. https://doi.org/10.1007/978-981-10-6620-7_4
21. Xiaoxue L, Xuesong B, Longhe W, Bingyuan R, Shuhan L, Lin L (2019) Review and trend analysis of knowledge graphs for crop pest and diseases. IEEE Access 7:62251–62264. https://doi.org/10.1109/ACCESS.2019.2915987
22. Chukkapalli SSL et al (2020) Ontologies and artificial intelligence systems for the cooperative smart farming ecosystem. IEEE Access 8:164045–164064. https://doi.org/10.1109/ACCESS.2020.3022763
23. Ghazal R, Malik AK, Qadeer N, Raza B, Shahid AR, Alquhayz H (2020) Intelligent Role-based access control model and framework using semantic business roles in multi-domain environments. IEEE Access 8:12253–12267. https://doi.org/10.1109/ACCESS.2020.2965333
24. Aydin S, Aydin MN (2020) Ontology-based data acquisition model development for agricultural open data platforms and

implementation of OWL2MVC tool. Comput Electron Agric 175(June):105589. https://doi.org/10.1016/j.compag.2020.105589
25. Drury B, Fernandes R, Moura MF, de Andrade Lopes A (2019) A survey of semantic web technology for agriculture. Inf Process Agric 6(4):487–501. https://doi.org/10.1016/j.inpa.2019.02.001
26. Tarus JK, Niu Z, Mustafa G (2018) Knowledge-based recommendation: a review of ontology-based recommender systems for e-learning. Artif Intell Rev 50(1):21–48. https://doi.org/10.1007/s10462-017-9539-5
27. Rajendran D, Vigneshwari S (2021) Design of agricultural ontology based on levy flight distributed optimization and Naïve Bayes classifier. Sadhana Acad Proc Eng Sci. https://doi.org/10.1007/s12046-021-01652-x
28. Zaman G, Mahdin H, Hussain K, Atta-Ur-Rahman S, Abawajy J, Mostafa SA (2021) An ontological framework for information extraction from diverse scientific sources. IEEE Access 9(Ml):42111–42124. https://doi.org/10.1109/ACCESS.2021.3063181
29. Mughal MH, Shaikh ZA, Wagan AI, Khand ZH, Hassan S (2021) ORFFM: an ontology-based semantic model of river flow and flood mitigation. IEEE Access 9:44003–44031. https://doi.org/10.1109/ACCESS.2021.3066255
30. Kaur N, Aggarwal H (2021) Query reformulation approach using domain specific ontology for semantic information retrieval. Int J Inf Technol 13(5):1745–1753. https://doi.org/10.1007/s41870-020-00464-2
31. Thukral A, Dhiman S, Meher R, Bedi P (2023) Knowledge graph enrichment from clinical narratives using NLP, NER, and biomedical ontologies for healthcare applications. Int J Inf Technol 15(1):53–65. https://doi.org/10.1007/s41870-022-01145-y
32. Ta CDC, Tran TK (2023) Constructing a subject-based ontology through the utilization of a semantic knowledge graph. Int J Inf Technol. https://doi.org/10.1007/s41870-023-01575-2
33. Canito A, Corchado J, Marreiros G (2022) A systematic review on time-constrained ontology evolution in predictive maintenance. Artif Intell Rev 55(4):3183–3211. https://doi.org/10.1007/s10462-021-10079-z
34. Mummigatti KVK, Chandramouli SM (2022) Supervised ontology oriented deep neural network to predict soil health. Rev d'Intelligence Artif 36(2):341–346. https://doi.org/10.18280/ria.360220
35. Ngo QH, Kechadi T, Le-Khac NA (2022) Knowledge representation in digital agriculture: A step towards standardised model. Comput Electron Agric 199(May):107127. https://doi.org/10.1016/j.compag.2022.107127
36. Zulkipli ZZ, Maskat R, Teo NHI (2022) A systematic literature review of automatic ontology construction. Indones J Electr Eng Comput Sci 28(2):878–889. https://doi.org/10.11591/ijeecs.v28.i2.pp878-889
37. Yu W et al (2022) A survey of knowledge-enhanced text generation. ACM Comput Surv 54(11s):1–38. https://doi.org/10.1145/3512467
38. Bhuyan BP, Tomar R, Cherif AR (2022) A systematic review of knowledge representation techniques in smart agriculture (Urban). Sustain 14(22):1–36. https://doi.org/10.3390/su142215249
39. Mahmood K, Mokhtar R, Raza MA, Noraziah A, Alkazemi B (2023) Ecological and confined domain ontology construction scheme using concept clustering for knowledge management. Appl Sci. https://doi.org/10.3390/app13010032
40. Wilson SI, Goonetillake JS, Ginige A, Walisadeera AI (2022) Towards a usable ontology: the identification of quality characteristics for an ontology-driven decision support system. IEEE Access 10:12889–12912. https://doi.org/10.1109/ACCESS.2022.3146331
41. Farming_India (2023) "Farming India_ India's No." [Online]. Available: https://www.farmatma.in/blog/. Accessed 28 June 2023
42. Krishijagran (2023) Agriculture News, latest news updates on Agriculture, Farming, Food Processing, Farm Tools & Machinery. https://krishijagran.com, [Online]. Available: https://krishijagran.com/agriculture-world. Accessed 28 June 2023
43. AgriFarming (2023) Agri Farming - Agriculture _ Livestock _ Gardening _ Aquaculture _ Horticulture _ Farming." https://www.agrifarming.in, [Online]. Available: https://www.agrifarming.in/category/agriculture-farming. Accessed 28 June 2023