International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 16 (2024) pp. 230–245
© Cerebration Science Publishing

Article

AOQAS: Ontology Based Question Answering System for Agricultural Domain

Krithikha Sanju Saravanan * and Velammal Bhagavathiappan

Department of Computer Science and Engineering, College of Engineering, Guindy, Anna University, Chennai 600025, India; velammalkarthik@gmail.com
* Correspondence author: Krithikhasanju3008@gmail.com

Received date: 4 March 2024; Accepted date: 21 March 2024; Published online: 24 May 2024

Abstract: Agriculture is an indispensable sector for human community that has been transformed by technological innovations. The data handling with information extraction is one of the areas that is benefited by the advancements in information technology. The presented research work aims to develop a question answering system (QAS) for improving the information retrieval from the agricultural text documents. The proposed Agriculture domain Ontology based QAS (AOQAS) processes the given agricultural text documents and constructs it to a knowledge representation called ontology. The domain based ontology is created using the Bidirectional Encoder Representations from Transformers model (BERT model) with Regular Expressions (RE) for withdrawing domain terms and the Bidirectional Long Short Term Memory model (BiLSTM) with RE for relationship extraction between the agricultural terms. From the developed ontology, the answers for the input query are extracted and validated using Natural Language Processing (NLP) techniques and the deep learning model. The proposed AOQAS shows an accuracy and recall of 98.47% and 98.26%. The outcomes of AOQAS shows better performance when it is evaluated against the current systems.

Keywords: natural language processing; ontology; question answering system; agricultural term extraction; relationship extraction; deep learning

1. Introduction

The key purpose of agriculture is to provide food for daily consumption. Its innovations and scope grow persistently as there is a constant demand for food [1,2]. The industrial revolution led to sustainable productivity with increased quality and quantity of agricultural products [3]. The internet has enriched data across different domains and modalities. The amount of text data generated on the internet is growing at dizzying speed which makes it impossible for users to easily read all the required information and comprehend [4]. To resolve the issue, information representations, such as ontologies, knowledge graphs, concept maps etc., are utilized. The information representations are employed to organize and structure the data thus by making it easier to find and understand. Semantic understanding of the text documents is the capability to extract meaning and context from text and other unstructured data [5]. It is a complex task, as it requires the skill to understand the context of words and phrases, as well as the relationships between them [6]. In order to accomplish semantic understanding of the document, it is necessary to apply robust algorithms that are specifically designed for the agriculture domain of the documents. Multiple techniques can be used to mine flat files like text data [7]. One common technique is to employ NLP methodologies for extracting the meaning of the text document. NLP techniques with machine learning [8] and deep learning methods [9] are utilized to retrieve entities, relationships, and other important information in text document to attain more accurate results and efficient performances. However, semantic understanding of the agriculture domain based document is still a formidable and challenging task. It is important to choose the right approach or frame the effective algorithm for converting the text document into information representation with domain based semantic understanding of the data [10]. Ontology is the process of representing information of a particular domain in a graphical way which is both machine readable and easily understandable by humans [11]. Ontologies can be used to represent a wide range of knowledge, including the relationships between different concepts, the properties of those concepts, and the rules that govern how those concepts can be used.

A framework for constructing the ontology from the agricultural text document and retrieving the information from the constructed ontology using the QAS is proposed in this research work. Agriculture domain based ontology for the text documents is constructed using NLP and deep learning methodologies then the proposed QAS receives the agricultural domain based question and pre-processes the question for answer extraction. If the given question is grammatically incorrect or not having enough keywords for extracting the candidate answers, question reformulation is performed which involves grammar error rectification and the addition of relevant keywords. From the reformulated question, candidate answers are extracted by forming queries which are formed using entity and relation identification from the question. The extracted candidate answers are validated with the passage from the input text document that contains the words in the given question using deep learning model and the most relevant answer is selected.

The proposed AOQAS aims at extracting the candidate answers from the agricultural ontology and selects the appropriate answer. The research work comprises of following contributions.

- Creating an effective domain based ontology using NLP and deep learning techniques.
- Solving the problems related to domain-based NLP systems such as lack of semantic understanding, data, and question reformulation.
- The extraction of answers from the ontology by identifying various metadata about the entity, such as relations, concepts and query-related understanding.
- Novel answer validation technique is introduced to handle the domain-based system which includes the semantic understanding of the data.

With the above-mentioned contributions, the ontology is constructed from the corpus and also answers are extracted from the ontology and verified the extracted answers for correctness. In this research article, Section 2 describes about the detailed literature survey that has been done for the proposed AOQAS, Section 3 gives the detailed design and explanation about the overall methodologies, Section 4 discusses the step by step implementation results obtained for the proposed AOQAS, Section 5 is the conclusion and section 6 provides the references for the research work.

2. Literature Survey

The research work AOQAS is proposed by surveying and analyzing the existing works that has been done for developing ontologies and QAS. The literature survey of the research work is classified into two parts namely, ontology formulation and question answering frameworks.

2.1. Ontology Formulation

The key terms and relations for the construction of ontology is extracted from the text using the hierarchy of linguistic filters and NLP techniques [12]. By carrying-out domain analysis on the given text documents, ontology consisting of various relationships with requirements can be developed utilizing Web Ontology Language (OWL) [13]. The sub-domain relationships can be extracted using the knowledge based scheme and it is employed for the ontology creation [14]. Also, by applying the classification algorithms on the feature vector from the dependency syntactic analysis, the relationships for the ontology have been efficiently extracted [15]. Ontology can also be constructed using two complementary approaches namely, HTML Structure-based ontology learner and N-gram-based ontology learner [16]. The relationships between the entities are extracted and matched using the Relational Component Analysis (RCA) [17]. Analyzing the text with the use of concept clustering and taxonomic relation identification with fuzzy based conceptual similarity computing, the relationship between the entities can be retrieved [18]. Semantic ontologies are formed using the web search engine based statistics on identified keywords to find the taxonomic and non-taxonomic relationships [19]. Relationships detected for ontology can be from semantic pattern analysis with domain based keywords [20,21].

The current systems of agricultural term extraction apply NLP techniques and then mostly uses vocabulary for filtering the terms. The tokenization with parts of speech patterns [22], improved keyphrase assignment algorithm (KEA++) with AGROVOC vocabulary [23], Regular expressions using POS with domain patterns and statistical features [24], Regular Expression and NLP based Term extraction scheme [25], customized Named Entity Recognition (NER) model using Spacy model [26] are employed in the existing systems for extracting the agricultural domain terms. The proposed AOQAS work should address the issues in the prevailing systems. So, the compound terms must be extracted effectively and classify the ambiguous terms efficiently.

2.2. Question Answering Framework

The input questions given by the users cannot guarantee for the presence of required number of keywords from the complete set of keywords for the domain [27]. The user may give the questions with very minimal keywords and this makes the answer search process complicated. Ref. [28] proposed a query expansion mechanism to expand the query with the minimal number of keywords using lexical resources and word embedding. Neural networks [29] are used to construct the meaningful query with the extracted and expanded candidate terms from the text dataset. Ref. [30] applied the techniques of semantic and syntactic clustering for identifying the answers for the given query with the objective of reducing the search space. Once the candidate answers are extracted, the correct answer is selected using the likelihood of correctness. Ref. [31] proposed a system that receives the question from the user and important terms are segregated based on the POS tag [32]. The expected answer tag is identified based on the type of question. Then, the answer is extracted from the database using the identified terms and expected entity type. Ref. [33] developed the technique of retrieving the sentence from Twitter dataset, using the text search based on the expected answer type for the natural disaster domain. Ten types of questions are handled in the framework.

Ref. [34] utilizes the context generation method for the given questions where a list of contexts is extracted from the knowledge base, and then the answer is searched and extracted from the semantic web using the semantic search technique. The extracted answers are ranked by the users manually. Ref. [35] proposed an QAS based on the interactive knowledge enhanced attention networks. When the user raises a question and the answer for the corresponding query is selected using BiLSTM [36]. Extracting the keywords from the question and identification of the relevant web pages using semantic analysis is proposed by Ref. [37]. From these identified pages, snippets are extracted and then the appropriate answer is filtered. Ref. [38] constructed a neural network model for both the phases of QAS which are question generation and answer extraction to handle the table-based data. The label of the generated question denotes whether the question is positive or negative through which the collaboration model extracts the answer. Ref. [39] developed a BERT based architecture to solve the problem of answer selection. Candidate answer paragraphs are extracted using inverted indices and fine-tuned BERT model is used to rank those paragraphs and segments them for extracting the answer. Ref. [40] proposed a model to extract candidate answers from a triplet-based knowledge graph and the correct answer is selected using a verification mechanism involving the corpus using three neural networks. Ref. [41] reformulated the given syntactically wrong question and then extract the relevant candidate answers using context-based semantic search method. Ref. [42] addresses the issue of extracting answers for the combined queries by the combined indices concept. For the given question, the relevant paragraphs are extracted using combined indices which include an inverted index and next word list. Then, using information retrieval models [43], the candidate answers are selected and the correct answer is selected using semantic ranking. Extracting the answers from the text can be identified using the semantical search methods but selecting the appropriate algorithm is a crucial process in any QAS. Ref. [44] proposed the validation features for identifying the answers using validation and similarity features. Validation score involves knowledge-based features such as Gazetteers, WordNet Scores and data-driven features which includes Wikipedia and Google Rank. Similarity measures include similarity matrix and string distance metrics.

Now considering the agricultural QAS, Ref. [45] created an agriculture domain QAS based chatbot using multi-layer perceptron and Recurrent Neural Network for assisting the farmers with major inquiries. Based on LSTM with Word2Vec model, the sentence similarity is calculated by [46] on agricultural corpus and developed the rice FAQ based QAS. Ref. [47] constructs a knowledge graph and for the given question the answers are extracted from the knowledge graph by utilizing the NER and multi label text classification techniques. Ref. [48] classified the questions of QAS using machine learning models in two steps of tagging the words in the corpus and classifying the questions based on the tags. Ref. [49] used data in OWL and Resource Description Format (RDF). The QAS is formed using NLP techniques and semantic web technologies for extracting the answers from the OWL and RDF data corpus. Ref. [50] created an agricultural chatbot for the Kisan call center dataset by employing the sentence embedding model. Ref. [51] developed a methodology for agriculture domain QAS using NLP and information retrieval technologies for extracting the answer. Ref. [52] proposed QAS system for agricultural domain that initially generates a Knowledge Graph then generates questions using encoder-decoder neural network and extracts answers using Recurrent Neural Network.

From the above survey, the following issues are identified.

- Lack of effective identification of the agricultural terms which are domain-based terms and that can be a unigram (one word), bigram (two words), trigram (three words) and so on.
- Inefficiency in establishing the relationship between the identified entities.
- Identification of entities from the given question and identifying the relations to extract the

candidate answers.

- Validation of the answers in the agricultural domain is much challenging because there is no benchmark dataset.

3. Materials and Methods

In this section, the proposed AOQAS is explained completely in detail. The AOQAS takes the agricultural question from the user and extracts the exact answer from the ontology which is constructed from the agricultural text. The knowledge for the proposed QAS system is represented using the ontology and the raw text corpus is used for answer validation purpose. The first step of the proposed system AOQAS is to pre-processes the given question after reformulating the question by rectifying the lexical errors. Then keywords are included if the given question does not have sufficing keywords. The question reformation phase enhances the answer retrieval process effectively. From various government websites and text blogs, the input document is prepared and agricultural terms are extracted from the text with relationships needed for the ontology are extracted. Candidate answers are extracted using the entity type, entity sub type, answer type, relationship identification and from the formed query. Corresponding answers are extracted from the constructed ontology. The extracted candidate answers are validated using the corpus data and the exact answer is selected using the deep learning model. The overall architecture of the proposed AOQAS research work is shown in Figure 1.

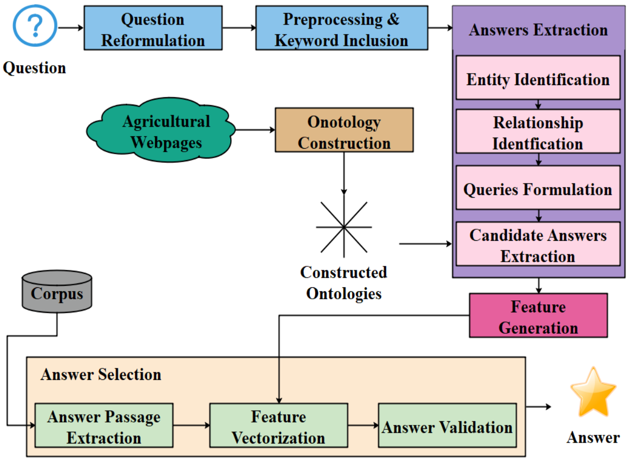


Figure 1. Overall Architecture Diagram of the Proposed AOQAS Work.

3.1. Dataset Description

The agricultural documents are created from government websites and blogs that contain text based information related to the domain. The data collected is collected from multiple sources and they are not uniform which increases the complexity in processing the data. So, the agriculture domain experts are engaged for preparing the contents for the input documents. The documents are collected from Indian Council of Agricultural Research (dogr.icar.gov.in accessed on 10 March 2023), Food and Agriculture Organization of the United States (FAO.org accessed on 10 March 2023), Tamil Nadu Agricultural University Agritech portal (Agritech.tnau.ac.in accessed on 10 March 2023), National Horticulture Research and Development Foundation (nhrdf.org accessed on 10 March 2023), Department of Agriculture & Farmers Welfare (agricoop.nic.in accessed on 10 March 2023), Farmer portal (farmer.gov.in accessed on 10 March 2023) and agricultural blogs during March 2023.

3.2. Question Reformulation

The questions that are given to any QAS are not correct at all occasions. In order to address the linguistic errors and barriers among the users, question reformulation is performed. The primary process involved in the question reformulation is the question correction. The process of question correction in this work involves correcting grammatical errors, spelling errors and basic linguistic corrections for a particular language. The research works prominently focus on the English language.

3.2.1. Question Correction

Question correction involves rectifying the syntactically incorrect question. Syntactic correctness of the question is done by checking the grammar rules with the construction of a parse tree. If the parse tree is successfully constructed for the given question, the given question follows the grammatical rule of linguistics. For checking the spelling errors, two steps are proposed which are performed in a sequential manner. The first step involves identifying the words that are close enough to the given wrong word using Levenshtein distance. Levenshtein distance is a type of distance in which insertion, deletion and substitution operations are allowed to perform while finding the distance between any two strings. Mathematically, the Levenshtein distance between two strings, x and y is calculated using Equation (1).

$$\mathrm{lev}(x, y) = \left\{ \begin{array}{c} |x|, \quad \mathit{if}\,|y| = 0 \\ |y|, \quad \mathit{if}\,|x| = 0 \\ \mathrm{lev}\big(\mathrm{tail}(x),\ \mathrm{tail}(y)\big), \quad \mathit{if}\ x[0] = y[0]) \\ 1 + \min \left\{ \begin{array}{c} \mathrm{lev}\big(\mathrm{tail}(x), (y)\big) \\ \mathrm{lev}\big((x),\ \mathrm{tail}(y)\big) \\ \mathrm{lev}\big(\mathrm{tail}(x),\ \mathrm{tail}(y)\big) \end{array} \right\}, \ \mathit{otherwise} \end{array} \right\} \tag{1}$$

where |x| is the length of the text x and tail(x) is the sequence of text except the first character (x[0]). There are following two cases of spelling mistakes. Case I: This covers the words with a few wrongly typed characters. This kind of word is addressed using the Levenshtein distance since it involves insertion, deletion and substitute operations. For convergence of the algorithm, the maximum error is fixed as square root of x after conducting a series of experiments and this gives an optimal solution. The constraint is mentioned in Equation (2).

$$\mathrm{lev}(x,y) \le (\mathrm{sqrt}(|x|)) \tag{2}$$

Case II: This addresses the words with rearranged characters. Using the basic operations of text (insertion and deletion), a single character-based jumbled text needs two steps to resolve it. For example, the wrongly spelt word “Riec” where “Rice” is the correct word needs two steps to be executed as follows to find the Levenshtein distance:

1. Remove ‘e’ at third position
2. Insert ‘e’ to fourth position

Totally, Levenshtein distance between “Riec” and “Rice” is 2. In order to address this, from the series of experiments, the maximum acceptance is fixed as (2* sqrt(|x|)). From the above analysis, the constraint for the candidate replacing words follows Equation (3).

$$1 \le \mathrm{lev}(x,y) \le (2^{*}\ \mathrm{sqrt}(|x|)) \tag{3}$$

The second process involves the Hidden Markov Model (HMM) to select the correct replacing word. Trained (HMM) model for the English language is used to rank the candidate replacing words. The top-ranked word is replaced with the wrong word. The Algorithm 1 represents the question reformulation process.

3.2.2. Algorithm 1: Question Reformulation by Question Correction

```
Input: Question Q
Output: Corrected Question Qcorrect
if!parse tree(Q) then
//Spell Error Correction
for dic word in Dictionary do
QLev = Lev(Q, dic word)
end
Qcorwords = Word Selection (QLev) //using Equations (2) and (3)
Qwithout Spell errors = HMM Pred (Q, QLev)
Qcorrect = Lexical Suggestion(Q without spell errors)
return Qcorrect
end
return Q
```

3.3. Question Preprocessing

Tokenization and stop word removal are the basic preprocessing in many applications of NLP. In question preprocessing, the reformed question is processed with tokenization followed by stop word removal. In the process of tokenization, the given question is chunked into meaningful tokens which are words in the English vocabulary. Tokenization is often performed with the removal of punctuation that are present in the given sentences. Extracted tokens are fed into the next preprocessing process called stop word removal. Stop word removal involves removing words that convey minimum information and often they are added to match the grammatical rules of the language. The output text from the stop word removal is the output of the question preprocessing and the next step is to check for keyword inclusions.

3.3.1. Keyword Inclusion

The list of keywords from the ontology and the dataset is extracted using Term Frequency-Inverse Document Frequency (TF-IDF) score. Mathematically, TF-IDF of word x in the document d from the set of Document D is calculated using Equation (4) [53].

$$\text{TF-IDF}(x, d, D) = tf\,(x, d).idf\,(t, D) \qquad (4)$$

where,

$$tf\,(x, d) = \log(1 + freq(t, d)) \qquad (5)$$

and

$$idf(t, D) = \log(N/\,(count(d \in D : x \in d)) \qquad (6)$$

The extracted keywords are stored in a buffer and used for question reformulation. The given question is compared with the list of keywords. The Keyword inclusion phase is executed when there is only one keyword present in the question. The given question is semantically compared with the list of keywords and ranked according to the sentence score η and semantic similarity ϕ. The sentence score η for the keyword *k* in the list of keywords and word *x* in the question is identified using the Equation (7).

$$\eta = \{n((k \cup x) \cap S\,) : S \in \text{Sentences in Corpus}\}/$$

$$\{n(S\,) : S \in \text{Sentences in Corpus}\} \qquad (7)$$

The semantic similarity ϕ between the considered keyword *k* and word *x* is calculated using the Synset function from WordNet [54]. Synset function returns a score denoting how semantically similar two words are, based on the shortest path that connects them. The rank is generated using the combination of semantic similarity ϕ and sentence score η. The top 3 matching keywords are attached with the question at the end position. The steps involved in keywords inclusion are shown in Algorithm 2.

3.3.2. Algorithm 2: Keyword Inclusion

```
Input: Preprocessed Question Qp
Output: Question with Keywords Qwithkey
LOC Buffer is populated using TF-IDF of given question on corpus using equation 4.
Qkeywords = Keyword Extraction(Qp, LOC Buffer)
if n(Qkeywords) = 1 then
for keyword ∈ LOC Buffer do
keyword(φ) = Synset(keyword, Qp)
end
 Keywords are selected based on φ
for sentence ∈ Corpus do
Sentence Score η is calculated using Equation (7) for Qp∪ Selected Keyword and Sentence
end
```

```
  The top 3 Keywords are selected using Sentence Score (keyword Selected)
return Qp∪ Keyword Selected
end
return Qwithkey
```

3.4. Ontology Construction

The text documents collected is converted into graphical representations using pretrained BERT model (BERT base uncased model) with RE for domain based term extraction and a BiLSTM model with RE is utilized for retrieving the domain relationships between the terms. From the terms and relationships, agriculture domain based ontologies are created and which is later used for extracting the answers for the given questions.

The input text document is given to the BERT model for tokenization and generates the contextual embeddings. Tokenization of the input document is done with the help of WordPiece tokenizer in the BERT model. Then these tokens are employed in the embedding layer, transformer layers and output layer for producing the contextual embeddings. Next the domain based entity patterns are formed using the RE. The contextual embeddings with RE utilizing rule based approaches are developed for extracting the domain based terms from the text document.

An empty directed graph is created with relationship patterns using RE. After this, created a list of unique entities and a mapping from entities to indices. The adjacency matrix is then constructed, where each entry represents the connection between two entities. The adjacency matrix is then converted to edge indices, and a Graph Data object is created. Now, the BiLSTM model is defined. It has three layers namely an LSTM layer, a linear layer, and a dropout layer. The LSTM layer takes the input features and produces a sequence of hidden states. The linear layer then takes the hidden states and produces a sequence of output features. The dropout layer is used to prevent overfitting. The dimensions for input, hidden, and output features are then set. The BiLSTM model is then created with the specified dimensions (input_dim = number of entities, hidden_dim = 16, output_dim = input_dim) and the model uses Gelu activation function. The loss function and optimizer are then defined. The loss function is Mean Squared Error (MSE) and the optimizer is Adam. Next, the input features are generated. The input features are one-hot encoded which means that each entity is represented by a vector of length of entities number, where the only non-zero entry is at the index of the entity. The created BiLSTM model is trained and the training loop iterates for a fixed number of 100 epochs. Learned node embeddings are received from the output feature of the model with MSE loss for each 10 epochs. From these node embeddings and relationship patterns using RE, the relationships between the entities are extracted for the input text document. For the graphical representation, the nodes are represented by the entities and the edges are represented by the relationships. The network library with torch_geometric module is utilized for visualizing the created ontology graph. The ontologies are stored as the database and it can be employed for searching, extracting the answers for the given questions.

3.5. Candidate Answers Extraction

For the given question, candidate answers are extracted for relevant relationships. The entities are extracted from the question using the NLP techniques of tokenization with POS tagging and stopword removal. The relevant relationships for the given question are identified using a semantic matrix.

3.5.1. Entities Identification

Entity represents agricultural terms that are used in the agricultural domain documents. Identifying the entities that are available in the question is useful to map the main context of the question. Often the entity identification from the given question results in identifying the crop to which the question is related. For example, let the given question be “When is the cultivation period of paddy?” and entities identified from the question are {“cultivation”, “period”, “paddy”}. Here, Paddy is the crop with which the question is dealing about. The entities in the question are detected using tokenization with POS tagging followed by stopword removal. This simple NLP techniques are used because more enhanced methods extract the entities with some information loss which are important for the question. For example, “what are the two types of rice” is the question and in the question “two”, “types” are important for answer search and that may not be properly filtered by the more enhanced entity extraction methods.

3.5.2. Relationship Identification

Relationship identification module identifies the relationships to which the given question is related.

The proposed AOQAS work mainly focuses on eight types of agriculture domain based relationships namely is a, is also/ are also, type of/ types of, cultivated in/ cultivated during, disease in/ diseases in, fertilizer for/ fertilizers for, intercrop of/ intercrops of, production of, for the construction of the ontology. Semantic similarity matrix S is identified for reformed question and list of relationships using Synsets. The semantic matrix maps the question with the relationship through a semantic similarity score. The entry in the semantic similarity matrix S is represented using σ ij where i represents the words from the preprocessed and reformulated question and j represents relationships in the ontology. The final relationship is identified using the maximum of the summation of the similarity value in the column. The top ranked relations are selected and then processed for query formulation.

3.5.3. Queries Formulation

The given question is tagged with the NER tags and then processed for the query formulation. The relationships identified in the previous sections are analyzed with necessary entities and queries are identified. Employing the type and sub-type of the given input question, the expected named entity type for the given question is found. The type of the input question is identified using multinomial logistic regression. The input question is vectorized using GLOVE framework [55]. Then, the vectorized question is fed into multinomial logistic regression with lbfgs solver. The same statistical model is used with different training sets for identifying the subtype of the question. From the entity type and subtype, answer tag type is estimated.

The relationships along with entities are analyzed and the relations are rejected if they do not contain the expected answer tag. Otherwise, relations are taken into consideration and the entity are accepted if it contains the expected answer tag. After processing the relationships, the required queries are extracted. If all the relations that are considered, are rejected, then the algorithm stops. For all the formulated queries, the entities are identified using the traversal technique from the ontology. The extracted entities should satisfy the expected named entity tag from the type and sub-type of the question.

3.6. Answer Selection

The most relevant answer is selected in the answer selection phase using answer validation. For the given question, the relevant passage is extracted from the text using TF-IDF. The proposed AOQAS considers not only the semantic relevancy between the question and candidate answers but also considers the relevant answer with the text passage. Features are extracted from the question, candidate answers with extracted passage and are unified. The unified feature is used to rank the answer through the neural network where the validation of the answer occurs. Then, the final relevant answer is selected.

3.6.1. Answer Passage Extraction

For the given question, the relevant passage is extracted using TF-IDF measure. The concept underlying the answer passage extraction is double verification. From the extracted passage, the relative sentence is extracted using the semantic relativity between the sentence and the input question. The extracted sentence contains the answer to the input question but it is the extracted sentence from the considered corpus.

3.6.2. Feature Vectorization

Feature Vectorization converts the text into numerical vectors using the generic framework. The Candidate Answers extraction process extracts different answers with varying identified relationships. For validating the answers, the feature vector should be generated. A tuple in the feature consists of input questions, extracted relationships, extracted answers and extracted passages from the corpus which are converted into a vectorized format using GLOVE network.

3.6.3. Answer Validation

Answer validation confirms the extracted answer using statistical machines which involves the features from the question, candidate answer and extracted sentence from the passage. The structure of Convolutional Neural Network (CNN) used for answer validation in the proposed system AOQAS is shown in Figure 2 which acts as a binary classifier for selecting the appropriate answer. The extracted features are fed into the CNN and the relevant answer selected is validated. The steps involved in proposed CNN model layers are,

- The first layer is the Input layer. This layer takes in the input data, which in this case is a paragraph and a question. The input data is a sequence of words, and each word is represented by a number. The number represents the index of the word in the vocabulary.

- Embedding layer: This layer converts each word in the paragraph and question into a vector representation. The vector representation is a fixed-length vector that captures the meaning of the word.
- Convolutional layer: This layer applies a convolution operation to the word vectors. The convolution operation extract features from the word vectors that are relevant to the question.
- Pooling layer: This layer reduces the dimensionality of the feature map produced by the convolutional layer. This is done to reduce the number of parameters in the model and to improve its performance.
- Dense layer: This layer applies a fully connected layer to the output of the pooling layer. The fully connected layer produces a probability distribution over the possible answers to the question.
- The last layer is the output layer. This layer outputs the classification result.

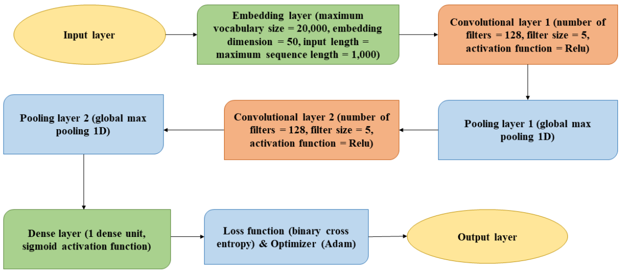
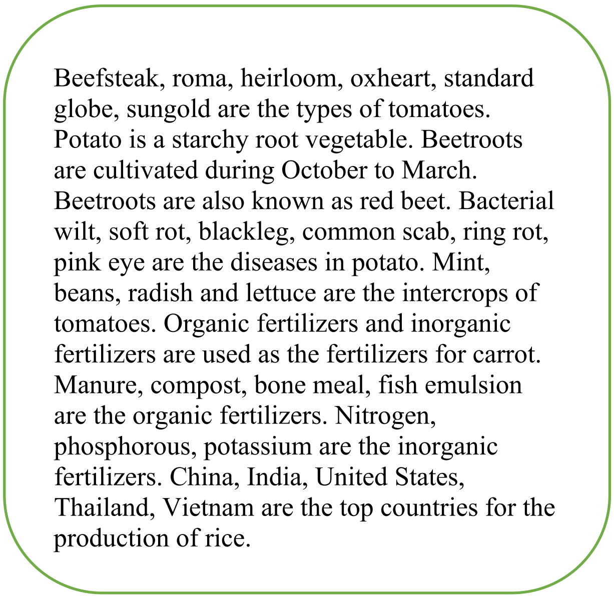

Input layer
Embedding layer (maximum vocabulary size = 20,000, embedding dimension = 50, input length = maximum sequence length = 1,000)
Convolutional layer 1 (number of filters = 128, filter size = 5, activation function = Relu)
Pooling layer 2 (global max pooling 1D)
Convolutional layer 2 (number of filters = 128, filter size = 5, activation function = Relu)
Pooling layer 1 (global max pooling 1D)
Dense layer (1 dense unit, sigmoid activation function)
Loss function (binary cross entropy) & Optimizer (Adam)
Output layer

Figure 2. Proposed CNN model layers for answer validation.

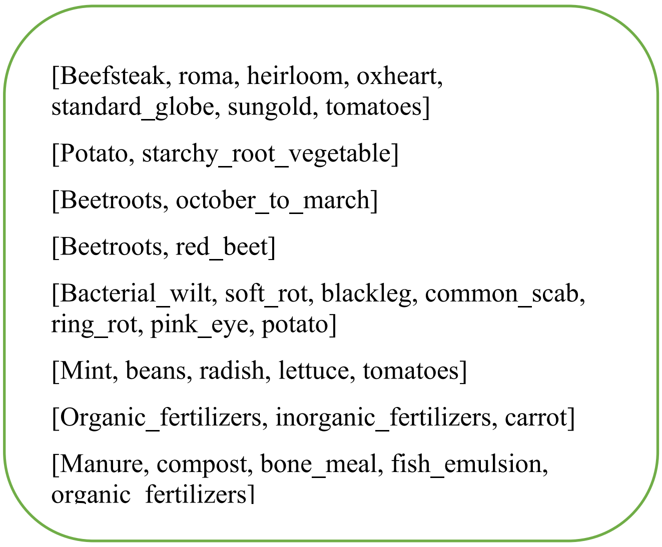
4. Results and Discussion

In this section, the step by step implementation results have been discussed. The first step of the proposed AOQAS work is to construct the ontology for the input text document. For creating the domain based ontology, agriculture term extraction and relationships between the terms are retrieved using pretrained Bert model with RE and BiLSTM model with RE. This section provides a visual explanation of the steps involved in developing the ontology. Figure 3 which represents the sample of the input text document, Figure 4 represents the domain terms extracted for the sample data, Table 1 represents the relationships between the agriculture terms that are extracted, Figure 5 represents the ontology created for the sample data.



Beefsteak, roma, heirloom, oxheart, standard globe, sungold are the types of tomatoes. Potato is a starchy root vegetable. Beetroots are cultivated during October to March. Beetroots are also known as red beet. Bacterial wilt, soft rot, blackleg, common scab, ring rot, pink eye are the diseases in potato. Mint, beans, radish and lettuce are the intercrops of tomatoes. Organic fertilizers and inorganic fertilizers are used as the fertilizers for carrot. Manure, compost, bone meal, fish emulsion are the organic fertilizers. Nitrogen, phosphorous, potassium are the inorganic fertilizers. China, India, United States, Thailand, Vietnam are the top countries for the production of rice.

Figure 3. Represents the sample of the text document.



[Beefsteak, roma, heirloom, oxheart, standard_globe, sungold, tomatoes]

[Potato, starchy_root_vegetable]

[Beetroots, october_to_march]

[Beetroots, red_beet]

[Bacterial_wilt, soft_rot, blackleg, common_scab, ring_rot, pink_eye, potato]

[Mint, beans, radish, lettuce, tomatoes]

[Organic_fertilizers, inorganic_fertilizers, carrot]

[Manure, compost, bone_meal, fish_emulsion, organic_fertilizers]

Figure 4. Represents the extracted domain terms from the sample data.

Table 1. Extracted relationships between the terms for the sample data.

Term 1	**Relationship**	**Term 2**
Beefsteak	Type of	Tomatoes
Roma	Type of	Tomatoes
Heirloom	Type of	Tomatoes
Oxheart	Type of	Tomatoes
Standard_globe	Type of	Tomatoes
sungold	Type of	Tomatoes
Potato	Is a	Starchy_root_vegetable
Beetroots	Cultivated during	October_to_march
Beetroots	Is also	Red_beet
Bacterial_wilt	Disease in	Potato
Soft_rot	Disease in	Potato
Blackleg	Disease in	Potato
Common_scab	Disease in	Potato
Ring_rot	Disease in	Potato
Pink_eye	Disease in	Potato
Mint	Intercrop of	Tomatoes
Beans	Intercrop of	Tomatoes
Radish	Intercrop of	Tomatoes

Lettuce	Intercrop of	Tomatoes
Organic_fertilizer	Fertilizer for	Carrot
Inorganic_fertilizer	Fertilizer for	Carrot
Manure	Is a	Organic_fertilizer
Compost	Is a	Organic_fertilizer
Bone_meal	Is a	Organic_fertilizer
Fish_emulsion	Is a	Organic_fertilizer
Nitrogen	Is a	Inorganic_fertilizer
Phosphorous	Is a	Inorganic_fertilizer
Potassium	Is a	Inorganic_fertilizer
China	Production of	Rice
India	Production of	Rice
United_states	Production of	Rice
Thailand	Production of	Rice
Vietnam	Production of	Rice

4.1. Evaluation for Term Extraction

The evaluation of the agriculture term extraction from the text document have been done by calculating the True Positive for Term Extraction (TPTE), True Negative for Term Extraction (TNTE), False Positive for Term Extraction (FPTE), False Negative for Term Extraction (FNTE). TPTE is the total count of both the actual and predicted terms are agriculture domain based terms. TNTE is the total count of both the actual and predicted terms are not agriculture terms. FPTE denotes the count of the predicted terms as the domain terms but in actual case they are not the domain terms. FNTE denotes the count of the terms which are actually domain terms but they are not predicted as the domain terms. Sensitivity is calculated by dividing the number of TPTE by the summation of number of TPTE and FNTE. Mathematically, it is represented as TPTE/(TPTE+FNTE). Specificity is estimated by dividing the number of TNTE by the summation of number of FPTE and TNTE. Mathematically, it is represented as TNTE/(FPTE+TNTE). Precision is the number of TPTE divided by the summation of number of TPTE and FPTE. Mathematically, precision is denoted as TPTE/(TPTE+FPTE). Negative Predictive Value is the TNTE divided by the summation of number of TNTE and FNTE. Mathematically, it is denoted as TNTE/(TNTE+FNTE). False Positive Rate is the number of FPTE divided by the summation of number of FPTE and TNTE. Mathematically, it is denoted as FPTE/(FPTE+TNTE). False Discovery Rate is the number of FPTE divided by the summation of number of FPTE and TPTE. Mathematically, it is denoted as FPTE/(FPTE+TPTE). False Negative Rate is the number of FNATE divided by the summation of number of FNTE and TPTE. Mathematically, it is denoted as FNTE/(FNTE+TPTE). Accuracy is defined as the measure that finds how close the actual and predicted terms are similar. Mathematically, it is represented as (TPTE+TNTE)/ (TPTE+TNTE+FNTE+FPTE). F1-Score is calculated using precision and recall for evaluating the terms extracted. Mathematically, it is represented as (2*TPTE)/(2*TPTE+FPTE+FNTE). Matthews Correlation Coefficient measures the quality of the domain term extraction technique. It is mathematically expressed as, (TPTE*TNTE-FPTE*FNTE)/[sqrt[(TPTE+FPTE)*(TPTE+FNTE)*(TNTE+FPTE)*(TNTE+FNTE)]]. The pretrained BERT model with RE methods used in this research work is assessed these evaluation metrics and the results are shown in Table 2.

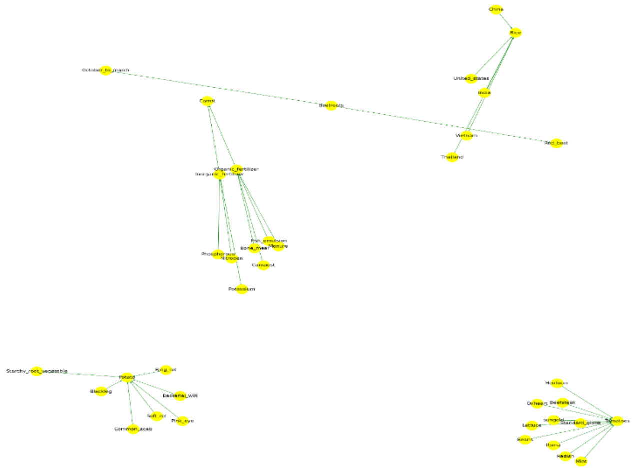


Figure 5. Ontology created for the sample data.

4.2. Evaluation for Relationship Extraction and Ontology

The evaluation of the extracted relationships and constructed ontologies have been done by estimating the True Positive for extracted relationships and constructed ontologies (TPERCO), True Negative for extracted relationships and constructed ontologies (TNERCO), False for extracted relationships and constructed ontologies (FPERCO), False Negative for extracted relationships and constructed ontologies (FNERCO) have been calculated with respect to the input text document and extracted domain terms. TPERCO is the total count of having both the actual and predicted relationships as same and also correctly available in the document between the terms. TNERCO is the total count of absence of the domain relationships between the terms both in cases. FPERCO denotes the count of the predicted relationships that are incorrect or no relation is actually available between entities. FNERCO denotes the count of the relationships which are actually there between terms but it is not predicted. Sensitivity is calculated by dividing the number of TPERCO by the summation of number of TPERCO and FNERCO. Mathematically, sensitivity is represented as TPERCO/(TPERCO + FNERCO). Specificity is estimated by dividing the number of TNERCO by the summation of number of FPERCO and TNERCO. Mathematically, specificity is represented as TNERCO/(FPERCO+ TNERCO). Precision is defined as the number of TPERCO divided by the summation of number of TPERCO and FPERCO. Mathematically, precision is denoted as TPERCO/(TPERCO+FPERCO). Negative Predictive Value is the TNERCO divided by the summation of number of TNERCO and FNERCO. Mathematically, it is denoted as TNERCO/(TNERCO+FNERCO). False Positive RERCO is the number of FPERCO divided by the summation of number of FPERCO and TNERCO. Mathematically, it is denoted as FPERCO/(FPERCO+TNERCO). False Discovery RERCO is the number of FPERCO divided by the summation of number of FPERCO and TPERCO. Mathematically, it is denoted as FPERCO/(FPERCO+TPERCO). False Negative RERCO is the number of FNERCO divided by the summation of number of FNERCO and TPERCO. Mathematically, it is denoted as FNERCO/(FNERCO+TPERCO). Accuracy is defined as the measure that finds how close the actual and predicted relationships are similar. Mathematically, it is represented as (TPERCO+TNERCO)/(TPERCO+TNERCO+FNERCO+FPERCO). F1-Score is calculated using precision and recall for evaluating the relationships extracted. Mathematically, it is represented as (2*TPERCO)/(2*TPERCO+FPERCO+FNERCO). Matthews Correlation Coefficient measures the quality of the domain relationships extraction methodology. It is mathematically expressed as, (TPERCO*TNERCO-FPERCO*FNERCO)/ [sqrt[(TPERCO+ FPERCO)*(TPERCO+FNERCO)*(TNERCO+FPERCO)*(TNERCO+FNERCO)]]. After assessing the proposed relationship extraction method and ontology constructed using these evaluation metrics the results are shown in Table 2.

Table 2. Evaluation for Agriculture term extraction and relationship extraction with ontology construction.

Evaluation Metric	Term Extraction Using Pretrained BERT Model + RE	Relationship Extraction and Ontology Construction Using BiLSTM Model + RE
Sensitivity or Recall	0.9740	0.9672
Specificity	0.9773	0.9600
Precision	0.9868	0.9833
Negative Predictive Value	0.9556	0.9231
False Positive Rate	0.0227	0.0400
False Discovery Rate	0.0132	0.0167
False Negative rate	0.0260	0.0328
Accuracy	0.9752	0.9651
F1 Score	0.9804	0.9752
Matthews Correlation Coefficient	0.9468	0.9168

After developing the domain based ontologies, the proposed AOQAS is framed. The Table 3 provides the output of the AOQAS for the sample data.

Table 3. Output of the AOQAS with respect to the sample data.

Question	Formulated Question	Entity and Relationship Extraction	Entity Type, Sub type	Answer Tag Type	Answer	CNN Response
What is the types of toomtoes?	What are the types of tomatoes?	Tomatoes and Type of	What What	subject	Beefsteak, roma, heirloom,Oxheart, standard globe, sungold	Yes
What are the cultivation period of beetroot?	What is the cultivation period of beetroot?	Beetroot And Cultivation period	WhatWhen	Date	October to March	Yes
When does pady is cultivated?	When does paddy is cultivated?	Paddy And Cultivated	When When	Date	Answer not available	No

Now, the proposed AOQAS work is assessed with the basic evaluation metrics (precision, recall, F1-score and accuracy). The precision of AOQAS is the ratio of number of questions with correct answer to the number of questions with answer. The recall of AOQAS is the ratio of number of questions with correct answer to the total number of questions with correct answer and incorrect answer. F1 measure is the harmonic mean of precision and recall, which is mathematically represent as [2*[(precision*recall)/(precision+recall)]]. Accuracy is the ratio of number of questions with correct answer to the total number of questions. The AOQAS has precision of 98.69%, recall of 98.26%, f1-

measure of 98.47% and an accuracy of 97.91%. The proposed QAS is then compared with the current systems. Ref. [45] has scored an accuracy of 97.83% with RNN and 96.97% with MLP. Ref. [46] employed LSTM with Word2Vec and had achieved an accuracy of 93.1%. Ref. [47] utilizing knowledge graph with NER and multilabel text classification have produced f1-score of 88%. Ref. [48] used Support Vector Machine and K- Nearest Neighbor shows an f1-measure of 88%. The ADANS model [49] has gained the precision and recall of 86.7%. The Agribot [50] for Kisan Call Center dataset has increased the accuracy from 86% to 91% by varying the dimensions of the word vectors. The Agri-QAS [51] has scored a very low accuracy of 69%.Ref. [52] have generated question-answer pairs from documents and achieved an accuracy of 87.3%. Thus, by comparing to the existing systems, the AOQAS shows remarkable results.

5. Conclusions

The proposed AOQAS framework constructs ontology from the given input document by employing the BERT model with RE for term extraction and BiLSTM model with RE for relationship extraction. Next, for the given question, the answer is extracted from the ontology using question reformulation, question preprocessing with keyword inclusion, entity identification, relationship identification, query formation. After extracting the answer for the input question, the answer is validated with the help of the proposed CNN model. Finally, the proposed AOQAS model is assessed by utilizing the evaluation metrics and the AOQAS shows the promising results when it is compared with the current systems, with an accuracy of 97.91%. In future, the work can be improved by giving more complex questions and the system must be able to handle the telegraphic questions. Also, more relationships should be considered in the future for the creation of ontology.

Author Contributions

K.S.S.: Conceptualization, Methodology, Resources, Data curation, Writing—original draft preparation, Writing—review and editing, Investigation, Validation. V.B.: Methodology, Resources, Supervision and Validation. All authors have read and confirmed for this published work.

Funding

No funding has been claimed for this research work.

Conflict of Interest Statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability Statement

The text documents for the proposed work is collected from Indian Council of Agricultural Research (dogr.icar.gov.in accessed on 10 March 2023), Food and Agriculture Organization of the United States (FAO.org accessed on 10 March 2023), Tamil Nadu Agricultural University Agritech portal (Agritech.tnau.ac.in accessed on 10 March 2023), National Horticulture Research and Development Foundation (nhrdf.org accessed on 10 March 2023), Department of Agriculture & Farmers Welfare (agricoop.nic.in accessed on 10 March 2023), Farmer portal (farmer.gov.in accessed on 10 March 2023) and agricultural blogs during March 2023.

References

1. Sanju, S. K. and Velammal, B. L.: An automated detection and classification of plant diseases from the leaves using image processing and machine learning techniques: A state-of-the-art review. Annals of the Romanian Society for Cell Biology, 15933-15950 (2021).
2. Sanju, S. K. and Velammal, B. L.: A Novel Deep Learning Model from Modified VGG16 with Resnet-50 Algorithm for Predicting the Diseases in Green Paddy Plants. Journal of Green Engineering, 11, 1696-1718 (2021).
3. Saravanan, K. S. and Bhagavathiappan, V.: A comprehensive approach on predicting the crop yield using hybrid machine learning algorithms. Journal of Agrometeorology, 24(2), 179-185 (2022).
4. Choudhary, K., DeCost, B., Chen, C., Jain, A., Tavazza, F., Cohn, R., et al.: Recent advances and applications of deep learning methods in materials science. npj Computational Materials, 8(1), 59 (2022).
5. Taneja, K., Vashishtha, J. and Ratnoo, S.: Efficient Deep Pre-trained Sentence Embedding Model for Similarity Search. International Journal of Computer Information Systems & Industrial Management Applications, 15 (2023).
6. Rathi, R. N. and Mustafi, A.: The importance of Term Weighting in semantic understanding of text: A review of techniques. Multimedia Tools and Applications, 82(7), 9761-9783 (2023).
7. César, I., Pereira, I., Madureira, A., Coelho, D., Rebelo, M.Â. and A de Oliveira, D.: Customer Success Analysis and Modeling in Digital Marketing. International Journal of Computer Information Systems & Industrial Management Applications, 15 (2023).

8. Sharma, A. and Kumar, S.: Machine learning and ontology-based novel semantic document indexing for information retrieval. Computers & Industrial Engineering,176, p.108940 (2023).
9. Bharadiya, J.: A comprehensive survey of deep learning techniques natural language processing. European Journal of Technology, 7(1), pp.58-66 (2023).
10. Onan, A.: SRL-ACO: A text augmentation framework based on semantic role labeling and ant colony optimization. Journal of King Saud University-Computer and Information Sciences, 35(7), p.101611 (2023).
11. Okoye, K.: Linked Open Data: State-of-the-Art Mechanisms and Conceptual Framework. In: Linked Open Data-Applications, Trends and Future Developments. IntechOpen (2020).
12. Hu, J., Huang, Z., Ge, X., Shen, Y., Xu, Y., Zhang, Z., et al.: Development and application of Chinese medical ontology for diabetes mellitus. BMC Medical Informatics and Decision Making, 24(1), 18 (2024).
13. Patel, A. and Debnath, N. C.: A Comprehensive Overview of Ontology: Fundamental and Research Directions. Current Materials Science: Formerly: Recent Patents on Materials Science, 17(1), 2-20 (2024).
14. Chatterjee, N., Kaushik, N., and Bansal, B.: Inter-subdomain relation extraction for agriculture domain. IETE Technical Review, 36(2), 157-163 (2019).
15. Fernando, S. and Stevenson, M.: A semantic similarity approach to paraphrase detection. In: Proceedings of the 11th annual research colloquium of the UK special interest group for computational linguistics, pp. 45-52 (2008).
16. Hazman, M., El-Beltagy, S. R., and Rafea, A.: Ontology learning from domain specific web documents. International Journal of Metadata, Semantics and Ontologies, 4(1-2), 24-33 (2009).
17. Bendaoud, R., Hacene, A. M. R., Delecroix, B., and Napoli, A.: Text-based ontology construction using relational concept analysis. In: International Workshop on Ontology Dynamics-IWOD 2007 (2007).
18. Lee, C. S., Kao, Y. F., Kuo, Y. H., and Wang, M. H.: Automated ontology construction for unstructured text documents. Data & Knowledge Engineering, 60(3), 547-566 (2007).
19. Sánchez, D. and Moreno, A.: Learning non-taxonomic relationships from web documents for domain ontology construction. Data & Knowledge Engineering, 64(3), 600-623 (2008).
20. Dahab, M. Y., Hassan, H. A., and Rafea, A.: TextOntoEx: Automatic ontology construction from natural English text. Expert Systems with Applications, 34(2), 1474-1480 (2008).
21. Sharma, A. and Kumar, S.: Ontology-based semantic retrieval of documents using Word2vec model. Data & knowledge Engineering, 144, p.102110 (2023).
22. Deepa, R. and Vigneshwari, S.: An effective automated ontology construction based on the agriculture domain. ETRI Journal 44(4), 573-587 (2022).
23. Medelyan, O. and Witten, I.H.: Thesaurus-based index term extraction for agricultural documents. In: Proceedings of the 6th International Conference on Computational Linguistics and Intelligent Text Processing, pp. 309-320. Springer (2005).
24. Kaushik, N. and Chatterjee, N.: A practical approach for term and relationship extraction for automatic ontology creation from agricultural text. In: 2016 International Conference on Information Technology (ICIT), pp. 241-247. IEEE (2016).
25. Chatterjee, N. and Kaushik, N.: RENT: Regular expression and NLP-based term extraction scheme for agricultural domain. In: Proceedings of the International Conference on Data Engineering and Communication Technology: ICDECT 2016, Volume 1, pp. 511-522. Springer Singapore (2017).
26. Panoutsopoulos, H., Brewster, C., and Espejo-Garcia, B.: Developing a Model for the Automated Identification and Extraction of Agricultural Terms from Unstructured Text. Chemistry Proceedings 10(1), 94 (2022).
27. Amur, Z.H., Hooi, Y.K., Soomro, G.M., Bhanbhro, H., Karyem, S. and Sohu, N.: Unlocking the Potential of Keyword Extraction: The Need for Access to High-Quality Datasets.Applied Sciences, 13(12), p.7228 (2023).
28. Esposito, M., Damiano, E., Minutolo, A., De Pietro, G., and Fujita, H.: Hybrid query expansion using lexical resources and word embeddings for sentence retrieval in question answering. Information Sciences 514, 88-105 (2020).
29. Cakir, A. and Gurkan, M.: Modified query expansion through generative adversarial networks for information extraction in e-commerce. Machine Learning with Applications, 14, p.100509 (2023).
30. Karpagam, K. and Saradha, A.: A framework for intelligent question answering system using semantic context-specific document clustering and Wordnet. Sādhanā 44(3), 62 (2019).
31. Banerjee, P.S., Chakraborty, B., Tripathi, D., Gupta, H., and Kumar, S.S.: A information retrieval based on question and answering and NER for unstructured information without using SQL. Wireless Personal Communications 108, 1909-1931 (2019).
32. Machado, M. and Ruiz, E.: Evaluating large language models for the tasks of PoS tagging within the Universal Dependency framework. In Proceedings of the 16th International Conference on Computational Processing of Portuguese, pp. 454-460 (2024).
33. Kemavuthanon, K. and Uchida, O.: Integrated question-answering system for natural disaster domains based on social media messages posted at the time of disaster. Information 11(9), 456 (2020).
34. Ali, I., Yadav, D., and Sharma, A.K.: SWFQA Semantic Web Based Framework for Question Answering. International Journal of Information Retrieval Research (IJIRR) 9(1), 88-106 (2019).
35. Huang, W., Qu, Q., and Yang, M.: Interactive knowledge-enhanced attention network for answer selection. Neural Computing and Applications 32, 11343-11359 (2020).
36. Yang, T., Mei, Y., Xu, L., Yu, H. and Chen, Y.: Application of question answering systems for intelligent agriculture production and sustainable management: A review. Resources, Conservation and Recycling,204, p.107497 (2024).
37. Kavitha, G. and Khanna, V.: Scheme for question answering system by using optimized knowledge graphs. PalArch's Journal of Archaeology of Egypt/Egyptology 17(7), 5386-5393 (2020).

38. Sun, Y., Tang, D., Duan, N., Qin, T., Liu, S., Yan, Z., Zhou, M., and Liu, T.: Joint learning of question answering and question generation. IEEE Transactions on Knowledge and Data Engineering 32(5), 971-982 (2019).
39. Yang, W., Xie, Y., Lin, A., Li, X., Tan, L., Xiong, K., Li, M., and Lin, J.: End-to-end open-domain question answering with bertserini. arXiv preprint arXiv:1902.01718 (2019).
40. Sawant, U., Garg, S., Chakrabarti, S., and Ramakrishnan, G.: Neural architecture for question answering using a knowledge graph and web corpus. Information Retrieval Journal 22, 324-349 (2019).
41. Ali, I. and Yadav, D.: Question reformulation based question answering environment model. International Journal of Information Technology 13, 59-67 (2021).
42. Khushhal, S., Majid, A., Abbas, S.A., Nadeem, M.S.A., and Shah, S.A.: Question retrieval using combined queries in community question answering. Journal of Intelligent Information Systems 55, 307-327 (2020).
43. Hambarde, K.A. and Proenca, H.: Information retrieval: recent advances and beyond. IEEE Access (2023).
44. Ko, J., Si, L., and Nyberg, E.: A probabilistic framework for answer selection in question answering. In: Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Proceedings of the Main Conference, pp. 524-531 (2007).
45. Rose Mary, C.A., Raji Sukumar, A., and Hemalatha, N.: Text based smart answering system in agriculture using RNN. AgriRxiv, 20210310498 (2021).
46. Liang, J., Cui, B., Jiang, H., Shen, Y., and Xie, Y.: Sentence similarity computing based on word2vec and LSTM and its application in rice FAQ question-answering system. Journal of Nanjing Agricultural University 41(5), 946-953 (2018).
47. Zhu, P., Yuan, Y., Chen, L., and Wu, H.: Question Answering on Agricultural Knowledge Graph Based on Multi-label Text Classification. In: International Conference on Cognitive Systems and Signal Processing, pp. 195-208. Singapore: Springer Nature Singapore (2022).
48. Oo, C.Z., Thu, Y.K., Nwe, H.M., and Thant, H.A.: Question Classification for Automatic Question-Answering in Agriculture Domain. Journal of Intelligent Informatics and Smart Technology 6 (2021).
49. Devi, M. and Dua, M.: ADANS: An agriculture domain question answering system using ontologies. In: 2017 International Conference on Computing, Communication and Automation (ICCCA), pp. 122-127. IEEE (2017).
50. Jain, N., Jain, P., Kayal, P., Sahit, J., Pachpande, S., and Choudhari, J.: AgriBot: agriculture-specific question answer system (2019).
51. Gaikwad, S., Asodekar, R., Gadia, S., and Attar, V.Z.: AGRI-QAS question-answering system for agriculture domain. In: 2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI), pp. 1474-1478. IEEE (2015).
52. Sanju, S.K. and Velammal, B.L.: An Agricultural Domain based Question Answering System Using Natural Language processing and Deep Learning Methodologies. In: 23rd International Conference on Hybrid Intelligent Systems (2024).
53. Aizawa, A.: An information-theoretic perspective of tf–idf measures. Information Processing & Management 39(1), 45-65 (2003). 800021-3).
54. Wang, M. and Wang, Y.: A synset relation-enhanced framework with a try-again mechanism for word sense disambiguation. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 6229-6240 (2020).
55. Pennington, J., Socher, R., and Manning, C.D.: Glove: Global vectors for word representation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1532-1543 (2014).