An Ontology Based Approach for Information Retrieval in Agriculture

BY

Naba Ranjan Roy
162-15-8135

Abdullah Al Masud
162-15-8051

This Report Presented in Partial Fulfillment of the Requirements for B.Sc in CSE (Computer Science & Engineering)

Supervised By
Mr. Narayan Ranjan Chakraborty
Assistant Professor
Department of CSE, Daffodil International University

Co-Supervised By
Sharmin Akter
Lecturer
Department of CSE, Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY
DHAKA, BANGLADESH
October, 2020

APPROVAL

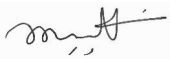
This Project titled **"An Ontology Based Approach for Information Retrieval in Agriculture"**, submitted by Naba Ranjan Roy, ID No: 162-15-8135 and Abdullah Al Masud, ID No: 162-15-8051 to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 7th October 2020.

BOARD OF EXAMINERS

Dr. Syed Akhter Hossain — **Chairman**
Professor and Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Md. Zahid Hasan — **Internal Examiner**
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Most. Hasna Hena — **Internal Examiner**
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University



Dr. Mohammad Shorif Uddin — **External Examiner**
Professor
Department of Computer Science and Engineering
Jahangirnagar University

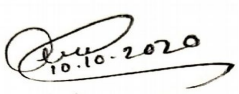
DECLARATION

We hereby declare that, this project has been done by us under the supervision of **Mr. Narayan Ranjan Chakraborty,** Assistant Professor and co-supervision of **Sharmin Akter**, Lecturer**, Department of CSE** Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:

Mr. Narayan Ranjan Chakraborty,
Assistant Professor
Department of CSE
Daffodil International University

Co-Supervised by:



Sharmin Akter
Lecturer
Department of CSE
Daffodil International University

Submitted by:

Naba Ranjan Roy
ID: 162-15-8135
Department of CSE
Daffodil International University

Abdullah Al Masud
ID: 162-15-8051
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First we express our heartiest thanks and gratefulness to almighty God for His divine blessing makes us possible to complete the final year project/internship successfully.

We really grateful and wish our profound our indebtedness to **Supervisor, Mr. Narayan Ranjan Chakraborty, Assistant Professor and Co-Supervisor, Sharmin Akter, Lecturer,** Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "NLP" to carry out this project. His endless patience ,scholarly guidance ,continual encouragement , constant and energetic supervision, constructive criticism , valuable advice ,reading many inferior draft and correcting them at all stage have made it possible to complete this project.

We would like to express our heartiest gratitude to the Almighty Allah and Head**,** Department of CSE, for his kind help to finish our project and also to other faculty member and the staff of CSE department of Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discuss while completing the course work.

Finally, we must acknowledge with due respect the constant support and patients of our parents.

ABSTRACT

Large number of text documents has been found in the web for different subject which contain specific information of a particular concept. These documents are in the form of heterogeneous nature that make difficult task to find information from their that user need. Now it become a challenging task to find relevant information from this type of source. In order to solve this problem to get relevant information according to the user query, we introduce a approach in the information retrieval field which is based on ontology. Here we use ontology concept for knowledge representation in our model which organize information in such way that help improve retrieval task. This technique not only help us to solve heterogeneous problem of information source but also help us to retrieve semantically related information that is the user's expectation. In this approach, we divide the documents into different classes according to their concept matching with a class. Then we find out contextually related feature words from each document which give us metadata for these documents and used for indexing. So when a user asked query first of all find out query's class then matching the query with document's metadata and the highest scoring document show to the user. Our approach has been tested by agriculture domain ontology which give us better result.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

Introduction

1.1 Introduction

Development of computer technology large number of electronic document have been increasing day by day. Where every document contain specific information of a specific topic. For increasing the number of documents they are in heterogeneous form. That means different types of document are store in one place which have no relation among them. So, from this large numbers of document find relevant information that user need become difficult task. Present information retrieval system mostly use keyword based technique. This technique mainly focused on keywords. That means when a user asked a query first of all find out keywords then according to these keywords system find which document carry these keywords and finally show those documents which carry maximum number of keywords. Though the result documents carry the maximum numbers of keywords but they never fulfill the user expectation. Because this system provide many unnecessary documents which are not semantically related to the user query. So solve this problem, we use ontology based knowledge representation technique which help us find semantically related documents. Here in this technique, we categories documents in different classes. In every class contain documents which are must be semantically relevant to the class. And every class are connected each other with contextually. Then we use a machine learning classifier for identify which class query actually this. When we identify the class then searching most relevant document in this class. Since every class organize documents semantically which help retrieve relevant information easily.

1.2 Motivation

People use information in their different task. Information help us know something and done our task easily. Sometime people face difficulties when they have lack of information. On the other hand, their task become easier when they have proper information about this. Today we have large number of information source but from this retrieve relevant information according to user need its become difficult task. That's why people do not get proper information when they need. Present information retrieval systems mainly based on keyword based system. In these system they focused on user query's keywords. According to this keywords system return

those documents which are contain maximum number of these keyword. As a result user get number of documents which are keyword related but these are not semantically related that do not make user satisfy. Sometime user get documents which carry the keywords but these not actually the user want. So most of the time user get irrelevant documents which did not help the user. But always want that they ask query and system will show the user semantic related documents which they want. That is why we try to make a system that will find semantically related document which help users.

1.3 Research Questions

- Is it possible to find semantically related information?
- How we will find relevant information according to user query?
- Can we find find relevant information in limited time?
- What is the lacking of the existing information retrieval system?

1.4 Expected Outcome

People use information retrieval system to find information according to their query. So when user ask query this time user expected that system will show the relevant information.

- That's why we expected that our system find relevant information which are semantically related to the user query.

- We also expected that our system will find relevant information in limited time.

1.5 Layout of the Report

Chapter one have demonstrated an introduction to the project with its motivation, research questions, and expected outcome.

- Chapter two will have “Background” demonstrates related works and challenges.

- Chapter three will have Research Methodology.

- Chapter four will have Experimental Results and Discussion.
- Chapter five will have Conclusion and Future works.

CHAPTER 2

Background Study

2.1 Introduction

In this section, we will discuss related works and challenges about this research. In related works section, we will discuss other research paper and their works, their methods, and accuracy which are related to our work. In research summary section we will give the summary of our related works. In challenges section, we will discuss how we increased the accuracy level

2.2 Related Works

When the ontology concept used into the information retrieval system. Their many researcher try use this concept to retrieve information efficiently. Vijayarajan, Dinakaran, Tejaswin, and Lohani (2016) proposed an ontology based Object-Attribute-Value (O-A-V) generic framework for information retrieval. It provides meaningful insight into the content of the documents, but related terms of those search words are not taken into consideration, thereby results in low precision and recalls of results [3]. Tulasi, Rao, Ankita, and Hgoudar (2017), who proposed a semantic ontology based system with automatic semantic annotation for efficient retrieval of documents in the domain of sports. They are also effort insight into the document compare to less effort into query [4]. Devi and Gandhi (2014), proposed SIRSD algorithm works by using WordNet and domain ontology to enhance the returned results of query [5]. Ma and Tian (2015) developed ontology for mechanical domain in RDF scripts with aid of Protégé 4.2 and adopted query expansion technique for its knowledge retrieval [6]. The work of Alfred et al. (2014) maintains that the retrieval mechanisms of an ontology knowledge based system are drawing research attentions presently. In their research work they focus on ontology based Query Expansion technique [7]. An ontology based information retrieval was presented in the domain of soccer Kara et al. (2012). In their research work, three semantic search issues were taken into consideration. Among which is the retrieval's performance issue [1]. While Ruban, Tendolkar, Rodrigues, and Shetty (2014) argued that the keyword based information retrieval does not satisfy user's query intent in terms of recall and precision.

Therefore, in order to enhance the relevance of results return to user, an information retrieval model based on domesticated plants ontology is developed [8].

2.3 Comparative Analysis and Summary

In the related works, we saw most of the paper try to match ontology class using query expansion technique. For query expansion technique, they use synonyms of the query words for matching ontology class. In this way they create more difficulties because every words have many synonyms words and considering these word expansion query make the query more ambiguous. Using this ambiguous query find ontology class become more difficult task. As a result, the retrieve information saw more irrelevant information.

2.4 Challenges

Finding ontology class according to user query is a big challenge for us. Because we have large number of class set. When our system able to find appropriate user query class that mean half of the retrieval task done by our system. Once find related class means in their must have user query related information. So if we fail find relevant class that means it become difficult task find related information. For this, we used machine learning query classifier technique for matching ontology class. That help us find appropriate class from multiple class. This help us not only make process time fast but also show us relevant information. After finding related class then find which document is more relevant according to user query from this class is another challenge for us. So for this we also use word2vec technique for find similarity among query words and every documents feature word. That help us retrieve relevant information from a class.

CHAPTER 3

Research Methodology

3.1 Data Collection Procedure

Since our research domain is agriculture so we collect agriculture data. For our research we mainly collect two type of data one is query set and another is document set. In the query set we collect five categories query. Collecting query set is become difficult because there is no available query set in the web. So we needed to contact will our friends who are studying in agriculture line. They help us collect many query. Their many online book, article help us to collecting query set. We also get help people collecting query who are good at agriculture. We also discuss with farmer because they involve in this agriculture sector. They are well known about in the many sector of agriculture. But also they know which type of problem they face in the agriculture sector. We collected these problem from them. We also asked them what want to know about their farming because this query help us to make our query set. We also create query which are know about to agriculture. In this way we collect 500 query for our research purposes.

Similarly we need to collect document set. These document are our information source that are retrieve according to user query. So when we collect documents we face many difficulties because their no available source for document set. For collecting documents first of all we decide categories that we want to work. Here we take five categories into account. So we contact with the people who are good in this specific categories. They help us found documents. We also contact with agricultural instituted for finding documents. Online book, article etc about related to the agriculture also help us to find document. We also contact with the friends who are studying in agriculture help us for find documents. In this way we collect total 750 documents which is used for information set of our research that are retrieve according to user query.

3.2 Proposed Methodology

In this section we discuss how we do our research. Here we try to explain all the parts of our system. We also try to show how our system perform.

3.2.1 Building Agriculture ontology

Agriculture ontology refers to the conceptual relationship of agricultural information. For making agriculture ontology, first of all, we divide agricultural information into different classes. Here each class describe specific area of agriculture. Every class may have one or many subclass. Then we identify the relation among these classes. It provides a map among different class in this area which can help computer for search specific information easily. It provides certain semantic help to improve the efficiency of information retrieval. Since every class hold those document which are semantically related of this class. Hence, there no chance found irrelevant information from a specific class. Therefore the semantic relationship concept should be possible to improve based on ontology.

3.2.1.1 Define Classes

Since ontology concept mainly based on classes. So first and most important task for making a better ontology is to find a specific class name which describe certain field of a domain. A class name is a concept which describe of a specific area. Every class have some properties which are explains about which type of information can be found their. So when an information want to become a member of a class first of this information need to fulfill of this class properties. If an information do not fulfill the properties of a class that means this information is not a member of this class. By a class name, it become easily understandable by a computer which type of information are in their. It also helpful for user who maintain the system for update the information. A class name not only provide which type of information are in their but also it help to find out how this class is related with other class. And this help us to make better categories of information so that any class document do not conflict with other class documents. So when we define new class, we need to identify is this type of class previously created into the ontology. We also need to identify as any class is not conflict with other class. If same class name found into the ontology that means same type of information found into this two class which make difficulties when retrieve information. So when we identify class name this time we need to careful so that any class name repeat. It help us to make clean ontology. For our case study of agriculture ontology, we find out top most classes which mainly represent agriculture. Then we try to divide every top classes into many sub-classes. And then every sub-classes into many more sub-class so on. In this way

we get many sub-classes which hold small pieces of information of this ontology. This help computer find information from this sub-classes. So their no need to search into the hole domain instead of this we only search the class where the information will be found. And this help find most relevant information from this domain.

3.2.1.2 Make relationship of classes

When we find out the classes then main task is make relationship among these classes. For this first of all, we need to know the properties of these classes which describe a class. From class properties we are able to know how these classes are and what type of information they actually hold. So making relationship among classes its need to know about the classes which help us make better ontology. Making relationship among classes help us how information is store and how these classes connected with each other. Which help someone able to understand the context of this ontology. This relationship of classes help computer to find out relevant information whenever someone want to know a specific information. These relationships help computer find out conceptual information. Because every class hold relevant information of this class. So when system searching for information into a class this time system must be get relevant information from this class. A good conceptual relationship of the classes help computer to avoid those information which are not relevant to the user query. Figure 3.1 describe our agriculture domain's classes and relationship among those classes. From this figure we can understand how this classes are connected with each other. So when we found new class which properties are not similar with the class properties that we already have classes it become easier for us add this class into our ontology. We also find out is this class a parent class or sub-class of a class. If this is parent class this time we will add this class directly with root. If this is a sub-class this time we will add this class with this parent class. Once connect these class in this way it not only visually make understandable of our system but also help us store information into our system. So from our ontology we can able to know how many class we have and how they are connected with each other. This not only help us retrieve relevant information but also help us normalize a huge heterogeneous information source using divide information into different classes and connect these class perfectly so that no class repeat again. So using this it become easier for us to maintain our system because we divide our information in different part.

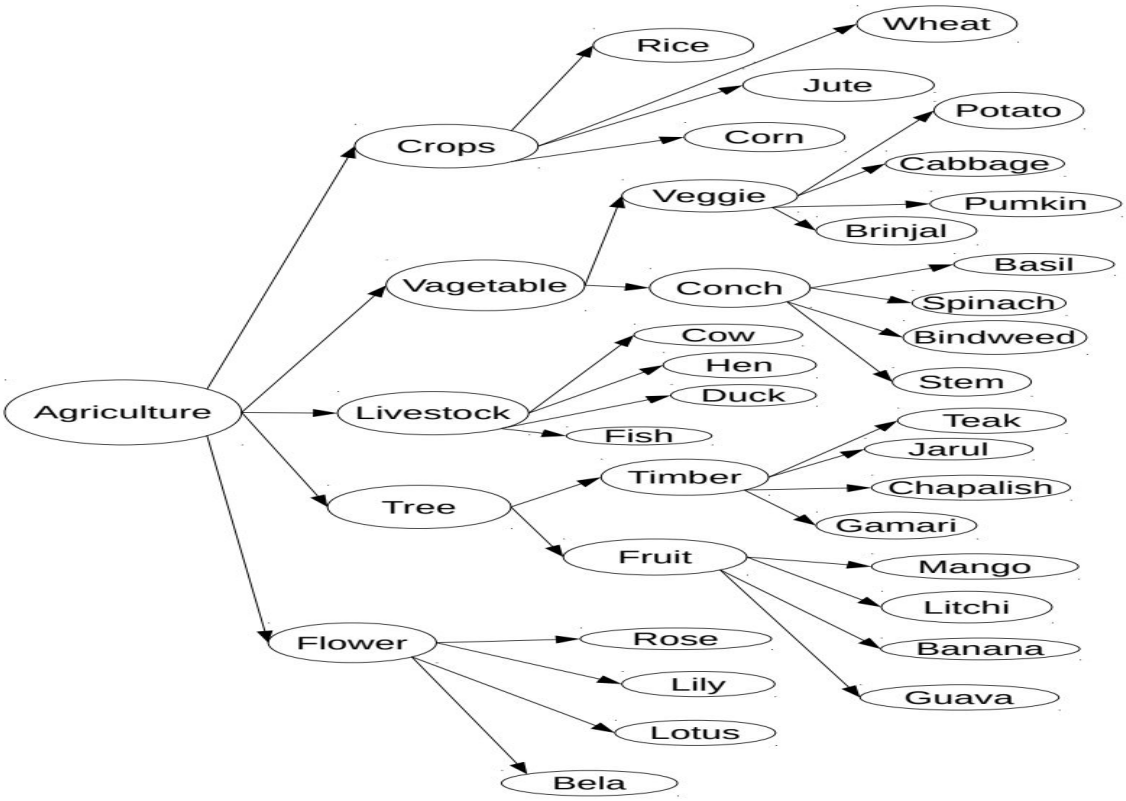


Figure 3.1: Agriculture Ontology

From the above figure we are easily able to represent knowledge into our system. Here every class represent a specific area of agriculture. From our ontology we can see that every class is connected with other classes. This connection help us make clean representation of knowledge for our system. This help store similar information into same class and this help us do better performance because in a class their no irrelevant information will not found. So when someone ask query this time only similar concept information will retrieve. Their no unnecessary information will retrieve which are not relevant to user query.

3.2.2 Retrieve Information from Agriculture Ontology

From agriculture ontology we know how we store information into our system. This also show us different type of information store into different class. So when user asked query what type of task need to be perform this time that is describe in figure3.2. Though their have lots of process need to follow. Here we try to figure out main part of our system.

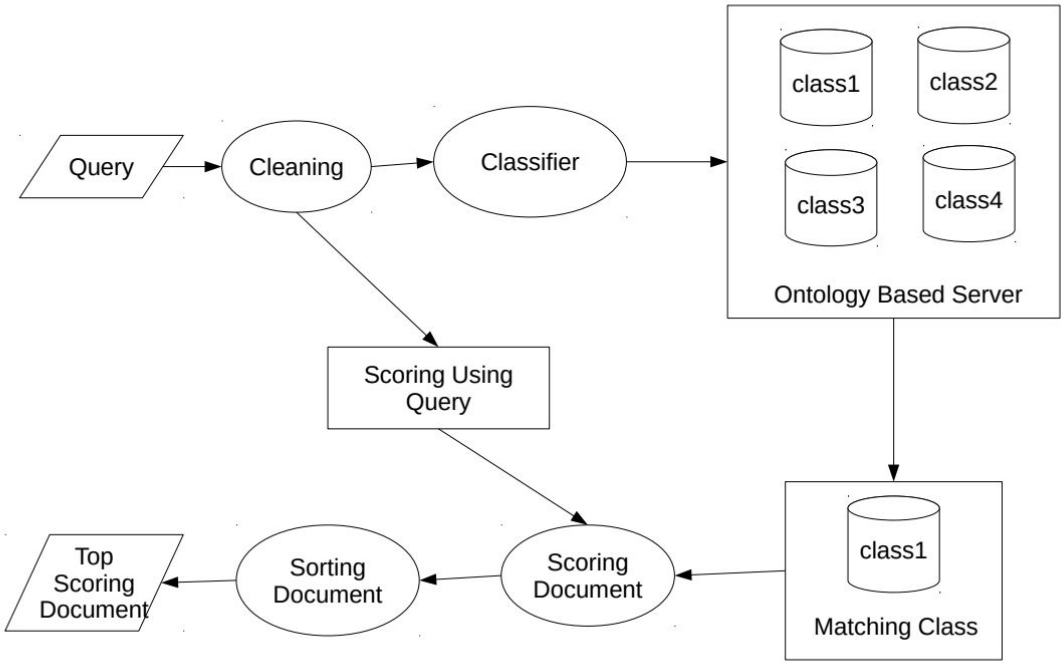


Figure 3.2: Ontology based approach for Information Retrieval

Above figure show how our system work for every query. Now we describe every part of our system in details so that it make clear how our system retrieve information for a specific query.

3.2.2.1 Document processing

For our system first of we need to organize our database with our documents so that when user ask query our system can retrieve information. Here we store these documents into different classes. In every class we apply preprocess technique into these documents for finding metadata from every document. Here metadata is small pieces of data which is explain main context of the document. From this metadata we are able to know which type of information are contain into this document. Instead of searching hole document only searching into the metadata we are able to know is their have our require information into this document. It help us for searching and find out relevant information according to the. user query. So for every information retrieval system document processing is the main task. The system will perform better depending on how better document are processed. Here for our

model, we store document into database in such a way so that we can easily retrieve relevant document from database. Here, create a new column where we store the feature word for each document. Here feature words are those words which represent the hole document. Those words give the main context about the document. Their no need search hole document only the feature word we can make understand which type information into the document. So for finding out feature words from document we use a well known technique which called tf-idf technique. Using this we can identify which words are feature word of a document.

$$TF(t) = \frac{t}{T}$$

Where,

t = Number of times term t appears in a document,

T = Total number of terms in the document

$$IDF(t) = \log\left(\frac{N}{n}\right)$$

Where,

N = Total number of documents,

n = Number of documents with term t in it

$$Tf - idf = TF(t) \times IDF(t)$$

This process take hole document of a class and give feature word for each document. So, when we find out feature words for each document, it is helpful for us to find out relevant document using feature words instead of searching hole document. This is helpful for us reduce retrieval time.

3.2.2.2 Query processing

When user asked a query, first of all the query will processing. Because in query there have many unnecessary word and symbol which are not contain any value in the query. Those word and figure use for human for understand. But for information retrieve those does not

play important role. That's why query need to clean before actual retrieval occur. We need to find out main words from user query. For this task, first of all we are remove punctuation from query. Here punctuation does not perform any role for retrieve information. Then we remove stop-words from query. Stop-words are those words which are not contain so much meaning for retrieve relevant document. That's why we remove stop-words from query. After remove stop-words, we will get few words, these words may be stay in different form. For this reason, we will use stemming technique. Stemming technique will make every word in root form. So using stemming technique, we will get those words which are in root form that will be make easy our retrieval task. This query processing process make an important role for retrieve information. Depending on query processing system will show this type of result. That means if we process query perfectly we will show most relevant information from our system.

Figure 3.3: Query processing

Figure3.3 show how our system clean user query and find out main word from query. So when a user ask any query to our system first of all query will going to in these procedure and find out few words. These words not any other word which are actually want to know. From this word system can able to know about user expectation. This process help system for retrieve relevant information.

3.2.2.3 Train query classifier

Here for our system we use a query classifier. This classifier classify user query and identify which class belong this query. For query classification, we need to make a classifier. Here classifier is a machine learning technique which use machine for classify query. So for making query classifier, first of all we need to train this classifier so that this classify query. For training our classifier, we use supervised machine learning technique. Here we collect

many query from different classes those are must be in our ontology. After collecting query with their class name then we clean those query so that may not happen any problem when we train our classifier. After cleaning query set we find out feature words from those query. Then we divide these query into two different part one is for training and another is for testing. Using training query set we are try to learn our model so that it able to identify query class. Using testing query set we are test validation of our model how this identify of a query class. Then we use multi-class Naive-Bayes classifier algorithm for our query classifier. Here Naive-Bayes classifier algorithm perform well for multiple classes. We train our classifier perfectly which perform well. We see this classifier are able to identify the class of given query.

3.2.2.4 Matching class using query classifier

When we clean query, we will get few words which are in root form. This process word give the actual meaning that the user want to find. Then we will not directly searching into database. If we will do this that will become time consuming. Today's keyword based information retrieval system use this technique. For this reason those retrieval system sometime produce irrelevant and unnecessary information which are not relevant to the user query. For this, we will use a technique which will help us find relevant information easily. Here we use query classifier for finding matching class. In figure3.4 show how our query classifier identity user query class. For this, we trained this classifier by different class query. Whenever user asked a query, the classifier will find related class. So when we find class, their no need to search other class for find information. This technique not only reduce processing time but also give us most relevant information. This help us reduce retrieve unnecessary information which show better performance in information retrieval.

Figure 3.4: Query Classification

3.2.2.5 Scoring retrieve document

After finding contextually related class, then we will find most relevant document according to query. For find relevant of documents, we will not searching hole document instead of this we will matching with the feature words that we was retrieve in section[3.2.1]. This also reduce processing time for our system. For finding similarity between feature word of a document and a query, we use word2vec technique. For matching similarity first of all we train our word2vec model by our dataset. This technique give us cosine similarity between query words and feature word of a documents. So using this technique we calculate the cosine similarity among each document's feature words with query words. After finding score between each document and query, then we try to find out which are most relevant document to the query. For this we sorting scores. Here highest score means more similar with user query. So we show highest score documents as a result.

Chapter 4

Experimental Results and Discussion

4.1 Experimental Setup

In Agriculture ontology there have many classes. For experimental purposes we take some classes. Here we take rice, mango, banana, hen and potato for our experimental class. Since we use classifier for our system that's why need to collect two type of dataset. One is query set that is use for train our classifier. And another is document set which contain information for our information retrieval system.

For query dataset, we take rice, banana, potato, hen and mango classes for experiment. So we need to collect query among those classes. So for this we make various query among those class. Here we did not get any source that we get query. That's why most of the query is made by us. After collecting those query we need to clean those query set so that our classifier give better performance. There are many algorithm which use for classification. But all of these do not give better result. So for our query classifier we use Naive-Bayes classifier which give better result for multi-class classification. Table 4.1 show our dataset that we use for train our query classifier.

Table 4.1: Query that we use training classifier

Query class	No. of query
Rice	100
Banana	100
Potato	100
Hen	100
Mango	100

For document dataset, we take rice, banana, potato, hen and mango classes for experiment. Since we use these classes into query dataset. That's why we use these class for document dataset so that we can measure the performance of our system. Here we collect various document among these classes. After collect these documents we are find out feature words among these document. Here we use tf-idf technique for find out feature words. These word are the main word of the documents. So this help use find similarity among document with query instead of searching hole document. Table 4.2 show our document dataset that we use in our system.

Table 4.2: Document that we use retrieve information

Document class	No. of doc
Rice	170
Banana	160
Potato	140
Hen	120
Mango	160

4.2 Experimental Results and Analysis

In our experiment, we show that our system show better result. Our query classifier give better result when we test it. We take number of query and every query class is known to us. Then we test our classifier using this query. We show most of the time it able to identify the class of query. When we measure the validity of our classifier, we show our query classifier give 89.25% accuracy.

In Information retrieval result, we show the recall ratio of our retrieval system show better result. Our system retrieve maximum number of document which is relevant. The retrieve rate of irrelevant document is vary low. It give most of the time these information which are most relevant to the query.

$$Recall = \frac{n}{N}$$

Where,

n = Total number of documents retrieved that are relevant,

N = Total number of relevant documents in the database.

Using this equation we calculate recall rate of our system. So first of all we take some query and relevant information of these query set for measure performance of our system. Then using selected query we observe retrieve information. We saw most of time our system able to retrieve information that we expected. Figure 4.1 show the recall ratio of our system for different classes.

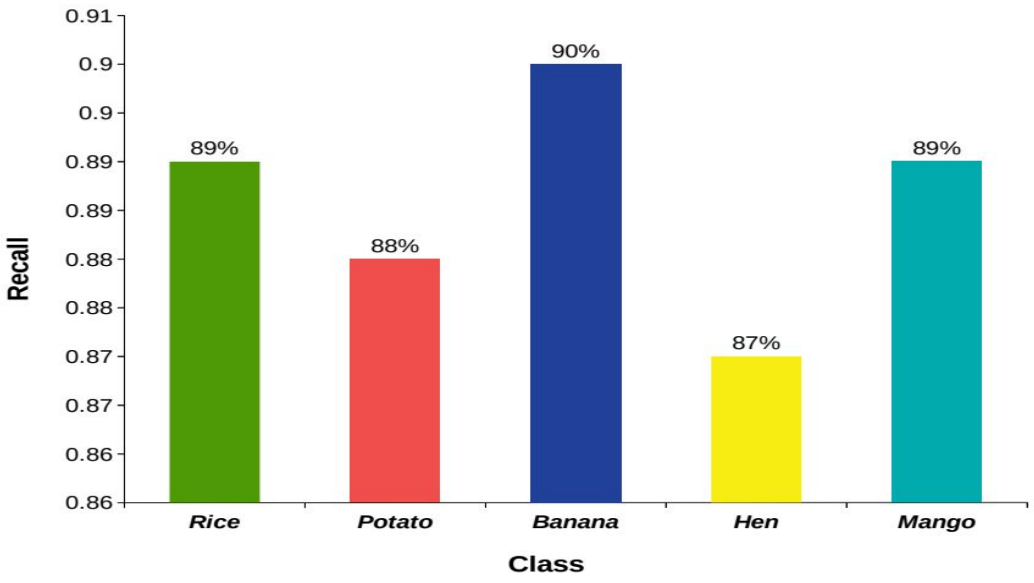


Figure 4.1: The result of recall ratio

The precision ratio of our system show better result. In testing time we show retrieve information according to the query is most similar to the information that we set for the

query. We show most of the time it try to show accurate result for given query. The retrieve of irrelevant information rate of our system is low. So retrieve information of our system is mostly relevant. That's why user do not need to confused which information is relevant and which is not. That help a user find their information.

$$Precision = \frac{n}{x}$$

Where,

n = Total number of documents retrieved that are relevant,

x= Total number of documents that are retrieved.

Using this equation we measure of our precision rate of our system. So we take a query and measure how relevant retrieve information. We saw our system retrieve irrelevant information is low. In this way we calculate all classes precision rate. Figure 4.2 show the precision ratio of our system for different classes.

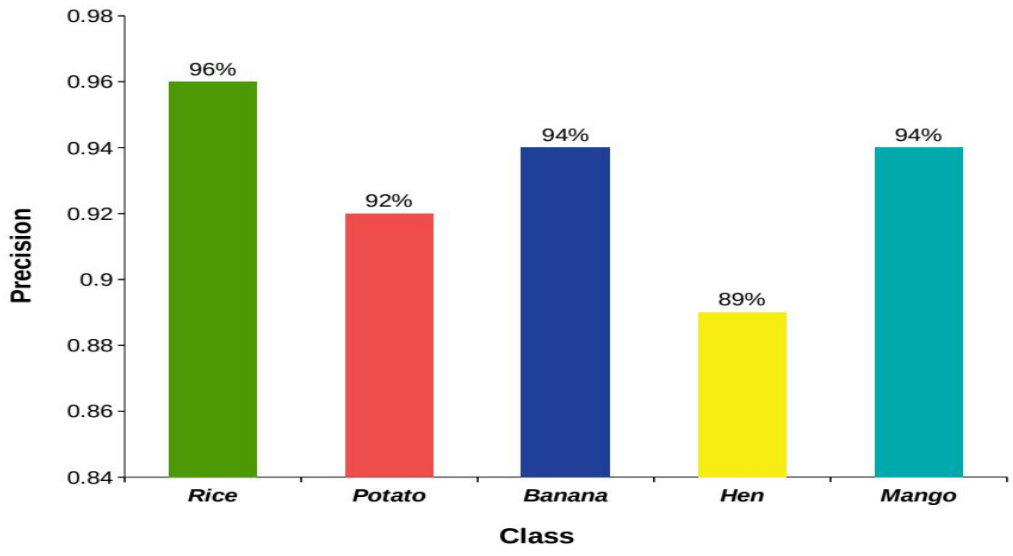


Figure 4.2: The result of precision ratio

From our result we show our system do better result. Our classifier classify query perfectly. If we train our classifier with more query the accuracy rate of our classifier will be increase. But present accuracy rate of our classifier is good, it can able to identify class of query correctly. Information retrieval rate of our system show better result when we experiment. The recall rate and precision rate of our system is good. In the experiment time we saw recall rate and precision rate of some class low but when we increase information source of these class we see the recall rate and precision rate of these class increase. So we can say that recall rate and precision rate of our system will increase if we increase more documents.

Chapter 5

Impact on Society and Sustainability Plan

5.1 Impact on Society

Since our work is on information retrieval and we try to retrieve relevant information according to user query. So it make huge impact on society. When people will get information according to their need it will help them do their task perfectly. There no need to wast time for find out information. They will get relevant information when they need. Since our working domain is agriculture so our farmer and people related to agriculture will get help much more. They will get agriculture related information easily. Our farmer will able to how to cultivate, maintain, store etc from our system which help them farming perfectly. When our farmer will do farming perfectly that means it will help to reduce food and related problem from our society. When our farmer are able to produce more product they can earn more money which help them become self-reliant. They can make their child educate and can buy thing they daily need for different purposes in their family. So when our farmer get relevant information they can doing their farming perfectly which help reduce different type of problem from our society.

5.2 Sustainability Plan

For make our system sustainable, we do lots of plan. We will collect lots of information so that user will get these information when they need. We will make our system online so that people can access our system from anywhere whenever they need. We will make connection with agricultural organization so that they can help us by giving agricultural information. A agricultural organization is much know more about agriculture they will know present problem in agriculture which help add this information into our source since our working domain is agriculture. We will make our system up to date so that people get right information. We will make our system for multiple user so that it can serve multiple user. We will also make our system user friendly so that they fill comfortable when they find information.

Chapter 6

Conclusion and Future works

6.1 Conclusion

In this report, we proposed mechanism which use for retrieve relevant information. For this we divide heterogeneous information into different class so that information store into our system nicely and this help retrieve information perfectly. Then we use machine learning classifier for classify user query. This classifier actually used for identify class from multiple class. Here, for our system we use Naive-Bayes classifier for our machine learning model. This classifier classify query perfectly and accuracy level is better than our expectation. We extract metadata from every document which help us searching information instead of searching hole document. Here we use if-idf technique which help us find out feature words from every document. This not only help us find relevant information but also help us reduce searching time. We also used word2vec model for calculate similarity between query word with document feature words. For this fist all we trained our word2vec model with our documents so that it can find out similarity between document feature words and query words. In this way we can scoring documents and identify which document is relevant for user. And that give a user semantic related document which user always ask for. We use agriculture domain for experiment. So first all we divide agriculture information into different classes. Then we train machine learning classifier with different agriculture related query. Then we evaluate our system and we saw our system retrieve relevant information according to user need. So we can say that our system is better than compare to other.

6.2 Future Works

Now this is our research based project for this others people can not use our system except our team. Keep this in mind in future we want deploy in the web so that anyone can use our system. For this we need to upgrading our system. We also need to train our classifier with huge number of query and also gather large number of document. In addition we want to make a mobile application so that anyone can access our system from anywhere whenever they need

REFERENCES

[1]. Kara, S., Alan, Ö., Sabuncu, O., Akpınar, S., Cicekli, N. K., & Alpaslan, F. N. (2012). An ontology-based retrieval system using semantic indexing. Information Systems, 37(4), 294-305.

[2]. Alfred, R., Chin, K. O., Anthony, P., San, P. W., Im, T. L.,Leong, L. C., & Soon, G. K. (2014). Ontology-based query expansion for supporting information retrieval in agriculture. In The 8th International Conference on Knowledge Management in Organizations (pp. 299-311). Springer.

[3]. Vijayarajan, V., Dinakaran, M., Tejaswin, P., & Lohani, M. (2016). A generic framework for ontology-based information retrieval and image retrieval in web data. Human-centric Computing and Information Sciences, 6(1), 18.

[4]. Tulasi, R. L., Rao, M. S., Ankita, K., & Hgoudar, R. (2017). Ontology-based automatic annotation: an approach for efficient retrieval of semantic results of web documents. In Proceedings of the First International Conference on Computational Intelligence and Informatics (pp. 331-339). Springer.

[5]. Devi, M. U., & Gandhi, G. M. (2014). WordNet and ontology based query expansion for semantic information retrieval in Sports domain. Journal on Computer Science.

[6]. Ma, S., & Tian, L. (2015). Ontology-based semantic retrieval for mechanical design knowledge. International Journal of Computer Integrated Manufacturing, 28(2), 226-238.

[7]. Alfred, R., Chin, K. O., Anthony, P., San, P. W., Im, T. L., Leong, L. C., & Soon, G. K. (2014). Ontology-based query expansion for supporting information retrieval in agriculture. In The 8th International Conference on Knowledge Management in Organizations (pp. 299-311). Springer.

[8]. Ruban, S, Tendolkar. K, Rodrigues, A. P, & Shetty, N, (2014). An ontology-based information retrieval model for domesticated plants. International Journal of Innovative Research in Computer and Communication Engineering, 2(5).

[9]. Akmal, S., Shih, L. H., & Batres, R. (2014). Ontology-based similarity for product information retrieval. Computers in Industry, 65(1), 91-107.

[10] F. Priyatna, D. Chaves-Fraga, A. Alobaid and O. Cor-cho, morph-GraphQL: GraphQL Resolvers Generation from R2RML Mappings., in: SEKE, 2019.

[11]. Uthayan, K. R., & Mala, G. S. A. (2015). Hybrid ontology for semantic information retrieval model using keyword matching indexing system. The Scientific World Journal, 2015.

[12]. Pokharel, S., Sherif, M. A., & Lehmann, J. (2014, August). Ontology based data access and integration for improving the effectiveness of farming in Nepal. In Proceedings of the 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT) (Vol. 2, pp. 319-326). IEEE Computer Society.

[13]. Qi, H., Zhang, L., & Gao, Y. (2010, August). Semantic retrieval system based on corn ontology. In Frontier of Computer Science and Technology (FCST), 2010 Fifth International Conference on (pp. 116-121). IEEE.

[14]. Ranwez, S., Duthil, B., Sy, M. F., Montmain, J., Augereau, P., & Ranwez, V. (2013). How ontology based information retrieval systems may benefit from lexical text analysis. In New Trends of Research in Ontologies and Lexical Resources (pp. 209-231). Springer, Berlin, Heidelberg.

[15]. Thangaraj, M., & Sujatha, G. (2014). An architectural design for effective information retrieval in semantic web. Expert Systems with Applications, 41(18), 8225-8233.

[16]. Xinhua, L., & Xutang, Z. (2012). A domain ontology based Information retrieval approach for technique preparation. Physics Procedia, 25, 1582-1588.

[17]. Zidi, A., & Abed, M. (2013, May). A generalized framework for ontology-based information retrieval: Application to a public-transportation system. In Advanced Logistics and Transport (ICALT), 2013 International Conference on (pp. 165-169). IEEE.

[18]. Bhogal, J., & Macfarlane, A. (2013, October). Ontology based query expansion with a probabilistic retrieval model. In Information Retrieval Facility Conference (pp. 5-16). Springer.

[19]. Fontes, C. A., Cavalcanti, M. C., & Moura, A. M. D. C. (2013, September). An ontology-based reasoning approach for document annotation. In Semantic Computing (ICSC), 2013 IEEE Seventh International Conference on (pp. 160-167). IEEE.

[20]. Jain, V., & Singh, M. (2013). Ontology based information retrieval in semantic web: A survey. International Journal of Information Technology and Computer Science (IJITCS), 5(10), 62.

PLAGIARISM


Final Test
ORIGINALITY REPORT
16%
SIMILARITY INDEX
11%
INTERNET SOURCES
4%
PUBLICATIONS
13%
STUDENT PAPERS