KrishiBot: A Chatbot for Farmer

A report submitted for the course named Project I (CS-200)

By

Rohit Kesarwani

Roll No. 17010102

Supervised By

Dr. Navanath Saharia

THESIS

SUBMITTED TO

Indian Institute of Information Technology, Manipur

Bachelor of Technology, Computer Science and Engineering

May, 2019

INDIAN INSTITUTE OF INFORMATION TECHNOLOGY, MANIPUR

Certificate

This is to certify that the project report entitled **"KrishiBot: A Chatbot for farmer"**, submitted to the Department of Computer Science and Engineering, Indian Institute of Information Technology, Manipur, in partial fulfillment for the award of the degree of Bachelor of Technology in Computer Science and Engineering, is a record of bona fide work carried out by **Mr. Rohit Kesarwani**, Roll No. 17010102, under my supervision and guidance.

No part of this report has been submitted elsewhere for award of any other degree.

Dr. Navanath Saharia
Assistant Professor

Date:

INDIAN INSTITUTE OF INFORMATION TECHNOLOGY, MANIPUR

Certificate

This is to certify that the project report entitled **"KrishiBot: A Chatbot for farmer"**, submitted to the Department of Computer Science and Engineering, Indian Institute of Information Technology, Manipur, in partial fulfillment for the award of the degree of Bachelor of Technology in Computer Science and Engineering, is a record of bona fide work carried out by **Mr. Rohit Kesarwani**, Roll No.17010102.

No part of this report has been submitted elsewhere for award of any other degree.

Dr. Nongmeikapam Kishorjit Singh
Assistant Professor Head,
Department of CSE
IIIT, Manipur

Date:

INDIAN INSTITUTE OF INFORMATION TECHNOLOGY, MANIPUR

Certificate

This is to certify that the project report entitled **"KrishiBot: A Chatbot for farmer"**, submitted to the Department of Computer Science and Engineering, Indian Institute of Information Technology, Manipur, in partial fulfillment for the award of the degree of Bachelor of Technology in Computer Science and Engineering, is a record of bona fide work carried out by **Mr. Rohit Kesarwani**, Roll No. 17010102.

Examiner's Signature:

Declaration

I declare that this submission represents my idea in my own words and where others' idea or words have been included, I have adequately cited and referenced the original source. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/sources in my submission. I understand that any violation of the above will be a cause for disciplinary action by the institute and can also evoke penal action from the sources which have thus not been properly cited or from proper permission has not been taken when needed.

(Signature)
Rohit Kesarwani
17010102

Date:

Acknowledgement

The success and final outcome of this project required a lot of guidance and I am extremely privileged to have got this all along the completion of my project. All that I have done is only due to such supervision and assistance and I would not forget to thank them. I respect and thank Dr. Navanath Saharia, for providing me an opportunity to do the project work and giving us all support and guidance which made me complete the project duly. I am extremely thankful to him for providing such a nice support and guidance. We pay our respects and love to our parents and all other family members and friends for their love and encouragement through out our career. Last but not the least we express our thanks to our friends for their cooperation and support.

-Rohit Kesarwani

Contents

List of Figures

Abstract

In this work, chatbot is implemented for the farmer purpose. This work is divided into two parts consisting of creating the data from the 'Vikaspedia (A government website for agriculture)' site and normal similarity based function for querying.

In the first part I have fetched some data from the site which I will use for querying and answering in chatbot. In second part I have used document similarity method to match with the given query by the user through chatbot(Krishibot).

Chapter 1

Introduction

“ Verizon Ventures is an active investor in the chatbot market. According to Christie Pitts, Manager – Ventures Development of Verizon Ventures, “Chatbots represent a new trend in how people access information, make decisions, and communicate. We think that chatbots are the beginning of a new form of digital access, which centers on messaging. Messaging has become a huge component of how we interact with our devices, and how we stay connected with the people, businesses and the day-to-day activities of life. Chatbots bring commerce into this part of our lives, and will open up new opportunities.”

– Christie Pitts

It all began in 1950, when Alan Turing, an English computer scientist, published a paper entitled "Computer Machinery and Intelligence.". In this paper Alan Turing he wrote "'Can machines think?'" In his article, he outlined the Turing Test, a way to measure whether one was speaking to a human or to a chatbot. In many ways, this was the beginning of AI, a test to discover the answer to his question.

In 1966, ELIZA was created by **"Joseph Weizenbaum"**, was one of the first chatbots. Although she was able to fool some users into thinking that they were actually talking to a human, she failed the Turing Test. Despite that, the principles used in ELIZA laid a foundation for the structures of chatbots, such as keywords, specific phrases, and preprogrammed responses.

In 1995, a popular online bot was A.L.I.C.E. was made which was basesd on language-processing. Although she was unable to pass the Turing Test, she did receive many other rewards for being the most advanced bot of her time.

In 2001, that is, until smarterchild came out. In many ways, it was the precursor to Apple's Siri and Samsung's S Voice.

In 2010-2015, over the next decade or so, bots became very popular among big tech companies, starting with in Siri (2010), Google Now (2012), Alexa (2015), and Cortana in (2015).

1.1 Chatbot

A chatbot is a computer program or an artificial agent which conducts a conversation via auditory or textual methods Now a days, Chatbots are used in the various practical purposes including cutomer services, college sevices, e-commerce sites and applications, etc.

1.2 Why chatbots are important?

Chatbot applications streamline interactions between people and services, enhancing customer experience. At the same time, they offer companies new opportunities to improve the customers engagement process and operational efficiency by reducing the typical cost of customer service.

1.3 Gantt chart

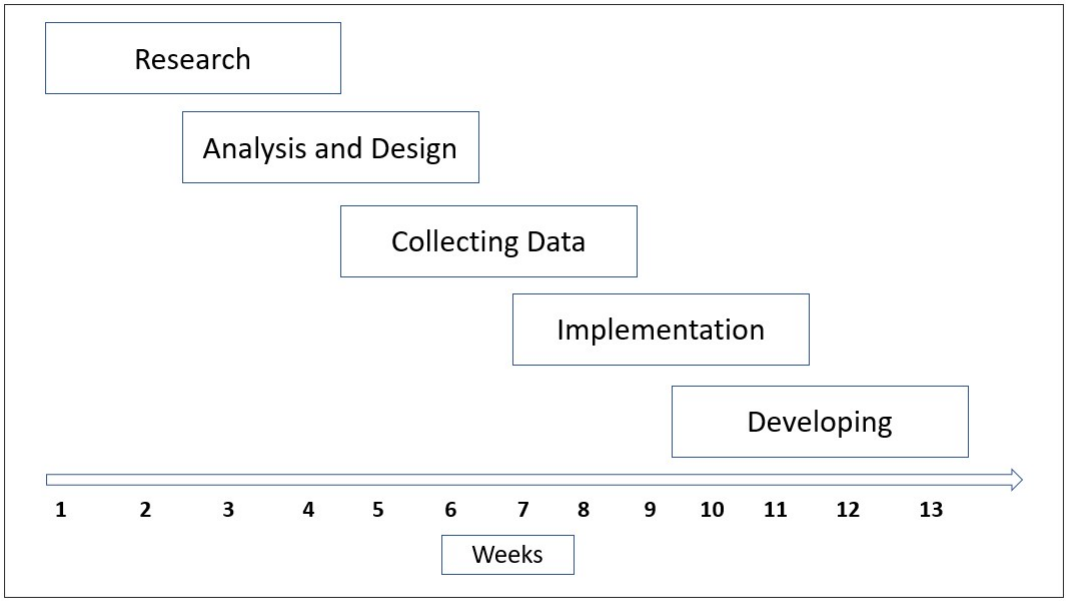

Figure 1.1: Gantt Chart

The Figure 1.1 shows the steps that I have taken to implement the project. The area which is shown in the brackets represent the work has been done in the part wise. The whole system implementation was divided into the five steps as shown in the Figure 1.1. It can be seen that some part of each steps was done parallely. The unit on week axis move by 1 week in Figure 1.1

Chapter 2

Existing System Study

2.1 Introduction

Many more chatbots are implemented in most of the domain but only few work has been done in agriculture area. So I have choosed to do on the agriculture domain because many farmers they don't have much knowledege about the how to grow better crops in the better way.
Few commercial chatbots are listed below which are implemented in the several domains.

2.2 FarmChat: A Convertational Agent to Answer Farmer Queries

In this the authors acknowledge two sources of knowledge that informed the development of FarmChat: 1- Farmer's information inquiries from the Kisan Call Center (KCC) 2- Findings from a formative study with local farmers and agriculture experts.

The Government of India has made all logs of calls to the KCC that has asked by the farmers from January 2015 to September 2017 which is available publicly. In total, this corpus contains data for around 8,012,856 calls. Each call log has 11 fields, including the date and time of the call, location, crop (one of the 306 crop types), query, and the answer provided by the KCC agriculture experts.

The paper authors conducted semi-structured interviews with 14 farmers (9 male, 5 female) and 2 male agriculture-experts, in September 2017. They worked closely with a local agriculture NGO, where the two agriculture experts were employed. They helped recruit the farmers and obtain their consent for participation, following their own internal ethics policies.

The farmers and agriculture experts provided the researchers with similar questions as the ones they found in the KCC dataset. Based on both the sources, they identified that the four major areas requiring information support:

1- Plant Protection: In the KCC dataset, 60.6% of the potato farming calls were related to remedies for protecting.

2- Pests and diseases: Agriculture experts stated that a majority of farmers seek suggestions on which medicine to spray for a particular crop disease. None of the farmers aware about any disease when the researchers interviewed. Usually, farmers describe crop diseases by their visible symptoms to the agri-expert with a few back-and-forth questions, the agriculture expert hypothesizes the issue and recommends medicine with dosage information that will be provided to the crops.

3- Weather: In the KCC dataset, 39.4% of the overall calls were about weather-related questions; 13.5% of potato farming questions were about weather. Farmers eagerly wanted to know weather information, as rains can wash away expensive sprayed pesticides and weather conditions determine the best time to harvest the crops.

4- Best Practices: Information related to best practices can help increase yield in terms of the quantity or quality of potatoes. Common questions were asked by the farmers: "Till what height should I put water?" "After how many days, should I harvest?" These best practices questions comprise of 6.6% of the potato farming calls in the KCC dataset. Agriculture experts also stated that farmers consistently asked them tips to increase yield and consequently income.

Unbiased Recommendations on Products: Farmers wanted recommendations from agri-experts on products they should purchase. Questions such as "Which fertilizer to put and how many times?" and "Which seeds are the best for red potatoes?" were commonly asked. They prefer to ask these questions to agri-experts instead of local shopkeepers, believing that agri-experts would provide unbiased and trustworthy response; they feared that shopkeepers may be motivated by the profit margin of products.

2.3 College Enquiry Chatbot Using A.L.I.C.E

The author {Balbir Singh Bani, Ajay Pratap Singh} is carried on to explain the design of a chat bot specifically tailored as an application which is going to help new students to solve all the problems that they face and the questions which arises in their mind during and after the admission . In particular, the proposal investigates the implementation of ALICE chat bot system as an application named as college enquiry chat bot. A keywords-based human-computer dialog system makes it possible that the user could chat with the computer using a natural language, i.e. in English.

2.4 Online Shopping Management System Chatbot with Customer Query handling Chatbot

In the e-commerce sites, there are deal with many kind of products throughout the world. The author proposed shopping system contains different services to make user feasible in e-shopping time. When user want to buy anything from these sites, he needs guideline about product and other things in this system just like make shopping in a store. To provide this kind of things in online, they integrate an artificial chatting system with e-commerce site which gives unlimited chatting services. When user first get into the

e- commerce site, he can ask queries to know in the system. E-commerce system sends customer query to the AIML Knowledge Base System to get answer by applying pattern matching algorithm. Then this answer return back to the system and then back to the user.

2.5 Chatbot for Laundry and Dry Cleaning

In this they present a chatbot for laundry and dry cleaning service. At first, they introduced a Facebook Messenger chatbot. This chatbot was used to acquaint interactions with users and to bring new customers. Based on the experience with the firstprototype that they have experienced earlier.

A customer just needs to follow these steps if he wants to interact with the chatbot:

1. make an order in the application, on the website or by phone,
2. hand clothes to Jeff at the preferred place and time,
3. take the clothes from Jeff at the arranged time and place.

Chapter 3

System Design

3.1 Architectural Design

Architectural design is a concept that focuses on components or elements of a structure and determinig the requirements for the system.The input or requirements to the analysis activity can come from any number of stakeholders and include items such as:

- What the system will do when operational (the functional requirements)

• How well the system will perform runtime non-functional requirements such as reliability, operability, performance efficiency, security, compatibility defined in ISO/IEC 25010:2011 standard.

• Development-time non-functional requirements such as maintainability and transferability defined in ISO 25010:2011 standard.

• Business requirements and environmental contexts of a system that may change over time, such as legal, social, financial, competitive, and technology concerns.

The outputs of the analysis activity are those requirements that have a measurable impact on a software system's architecture, called architecturally significant requirements.

3.1.1 Context Diagram

In the context diagram output and input to the system is defined. In this the input is the agriculture data that I have taken from 'Vikaspedia' {http://vikaspedia.in/} website. The output of the system is through the chatbot based on the keyword matching.

The Figure 3.1 represent the input and output of the system.

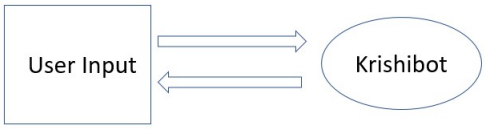

Figure 3.1: Input and Output of the system

3.1.2 1st level DFD

The Figure 3.2 shows the detail design of the system. The whole system is divided into 4 process local data file is used.

User Query:- One form of input from the user's environment as a query will come, and it will be a expression of the user's query that what type of query they want to ask to the information retreival system. The user has ordinarily thought of as a human being and the response will be generated by the machine indirectly to the human being.

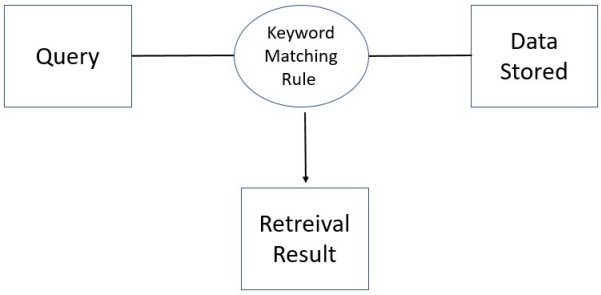

Figure 3.2: Data Flow Diagram

Keyword Matchig Rule:- TF-IDF stands for Term Frequency and Inverse Document Frequency which is a simple method to calculate weight of word in a document. The weight is a statistical measure which is used to evaluate how the word is important] in a document.

The importance of a word increases proportionally to the number of times a word appears in the document.Tf-idf is simple and can be successfully used for stop-words filtering in various subject fields including text summarization.

TF:- Term Frequency which measures how frequently a term or a word appears in a document. Since we know that every document is different in length wise, it is possible that a term would appear much more times in long documents than shorter ones. Thus, the term frequency is often divided by the document length (the total number of terms in the document) as a way of normalization:

TF(t) = (Number of times term t appears in a document) / (Total number of terms present in the document)

IDF:- Inverse Document Frequency which measures how important a term is in the document. While computing term frequency, all terms are considered equally important. However it is known that certain terms, such as "is", "of", and "that" (called as stop words), may appear a lot of times but have little importance. Thus we need to weight down the frequent terms while scale up the rare ones, by computing the following Inverse Document Frequency for the term:

IDF(t) = log(Total number of documents / Number of documents with term t in it)

Cosine Similarity:- Cosine similarity gives a useful measure of how two documents are similar likely to be in terms of their subject matter. Cosine similarity measures the similarity between two document using an inner product space that measures the cosine of the angle between the two documents. As we know that, cosine of 0° is 1, and it is less than 1 for any angle in the interval (0,pi] radians.

The cosine of two non-zero vectors can be derived by using the Euclidean dot product formula shown in equation

$$\mathbf{A} \cdot \mathbf{B} = \|\mathbf{A}\| \, \|\mathbf{B}\| \cos\theta$$

where :

$Ais\ First\ Vector/document$
$BisSecondVector/document$
$\theta istheanglebetweenAandB$

Given two vectors of attributes, A and B, the cosine similarity, cos( ), is represented using a dot product and magnitude as:

$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}},$$

Retrieved Result:- A retrieved information is logically a subset of the representations as partitioned off by the outcome of the matching rule applied to the formal query to the chatbot.

3.2 Physical Requirements

1- Input Requirement:- I have used the "Parsehub"(https://www.parsehub.com/) an open source software to take out the data from the website. The input is required only a user text and it is based on the data that I have stored in a json(format) file.

2- Output Requirement:- The Program will be run from command line and the output will be shown in the terminal. The output is shown in Appendix A.

3- Storage Requirement I have local file to store the data.

4- Data Requirement:- I have taken the data from the Vikaspedia(A government website for agriculture purpose) website and stored it locally.

Vikaspedia Agriculture Website –[1] **http://vikaspedia.in/agriculture**

[1]Vikaspedia Agriculture Website –

Chapter 4

Coding and Implementation

There are some challenges faced by the development team while implementing the software. Some of them are mentioned below: Code-reuse Programming interfaces of present-day languages are very sophisticated and are equipped huge library functions. Still, to bring the cost down of end product, the organization management prefers to re-use the code, which was created earlier for some other software. There are huge issues faced by programmers for compatibility checks and deciding how much code to re-use.

4.1 Code to generate text to speech

```
engine = pyttsx3.init()
engine.setProperty('rate', 150)
```

```
engine.setProperty('volume', 0.9)
print("Hey user I am KrishiBot How can I help you!!")
engine.say("Hey user I am Krishibot How can I help you")
engine.runAndWait()


engine.say(krishiBotReply)
engine.runAndWait()
```

4.2 Code to calculate cosine similarity

```
"compute cosine similarity of v1 to v2: (v1 dot v2)/||v1||*||v2||)"

def cos_similarity(v1, v2):
cosine = cosine_similarity(v1, v2)
return cosine
```

4.3 Code to load query from json file

```
def dataFromJsonFile(jsonDataFile, document):
i = 0
with open(jsonDataFile, "r") as sentences_file:
reader = json.load(sentences_file)
for row in reader:
document.append(row["question"])
i += 1
return document
```

4.4 Code to generate response

```
def generateResponse(jsonDataFile, minScore, cosine, max):
responseIndex = 0
if (max > minScore):
new_max = max - 0.01 list = np.where(cosine > new_max)
responseIndex = random.choice(list[0])
else:
return ”Chat with Krishibot...”, 0
j = 0
with open(jsonDataFile, ”r”) as sentences_file:
reader = json.load(sentences_file)
for row in reader:
j += 1
if j == responseIndex:
return row[”answer”], max
break
```

4.5 Code to reply through chatbot

```
while True:
query = input(”Krishibot : ”)
krishiBotReply = krishiChats(query)
print(krishiBotReply + ””)
```

Chapter 5

Conclusion

The keyword matching model that I have used in my project is very simple and easy to understand. Currently this project is not working for the ambiguous word, for eg: we can see in Figure 5.1 that if we provide the word ‘seed’ or ‘farming’ this will not respond. So that later on I can resolve the ambiguity that is coming in my project.

5.1 Future direction

Currently this project I have implemented is terminal based. Further more work I can do like it can be implemented for the web based or mobile application based. The data I have used from website are lesser, later on I can increase a much more data so that it can better response. And In future this project can be made multilingual because in India may be only few farmers know english so it can be implemented in another language like Hindi, Manipuri, Assamese also.

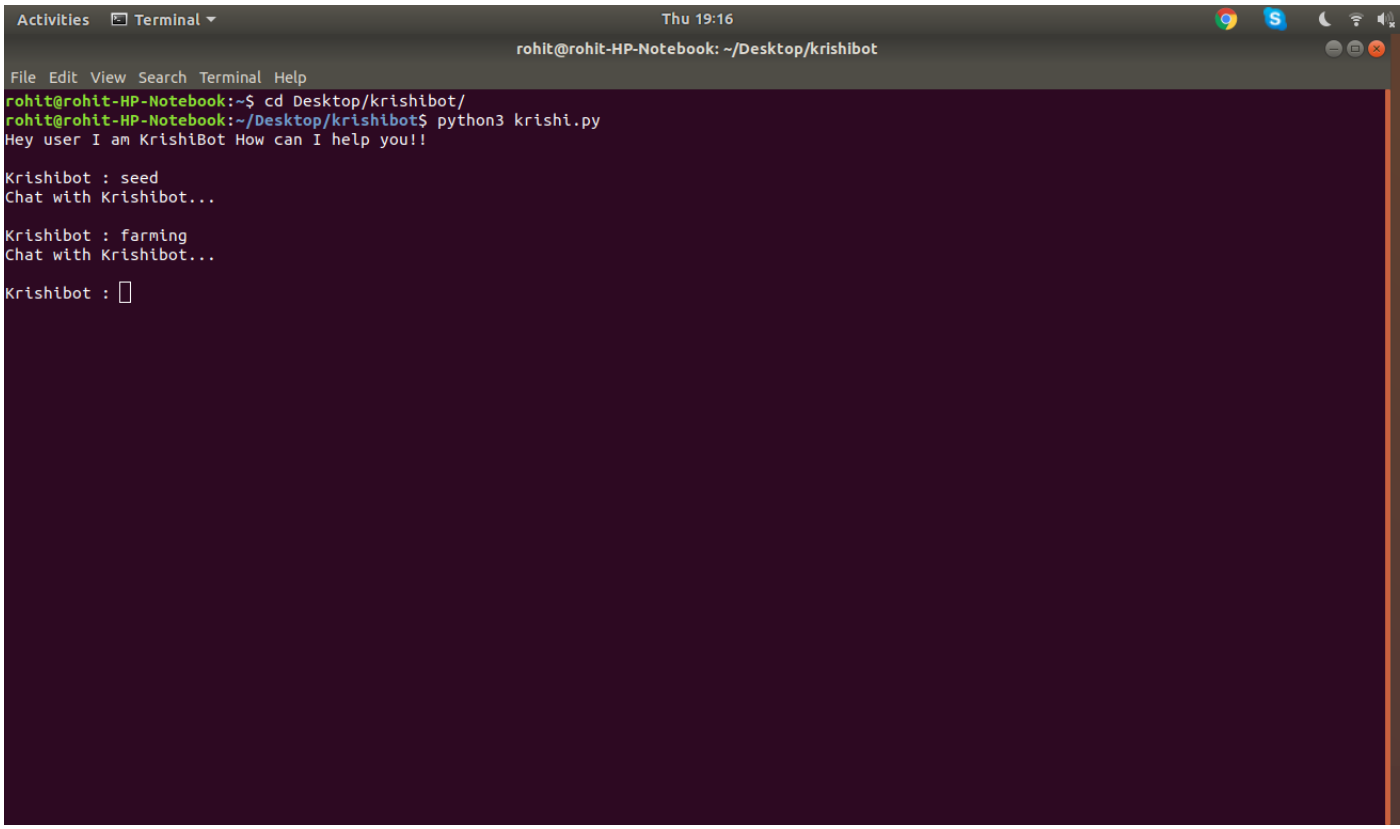

Figure 5.1: Ambiguity coming in Project

Appendix A

Screenshot and Description of the Implemented System

A.1 Data from 'Vikaspedia' website

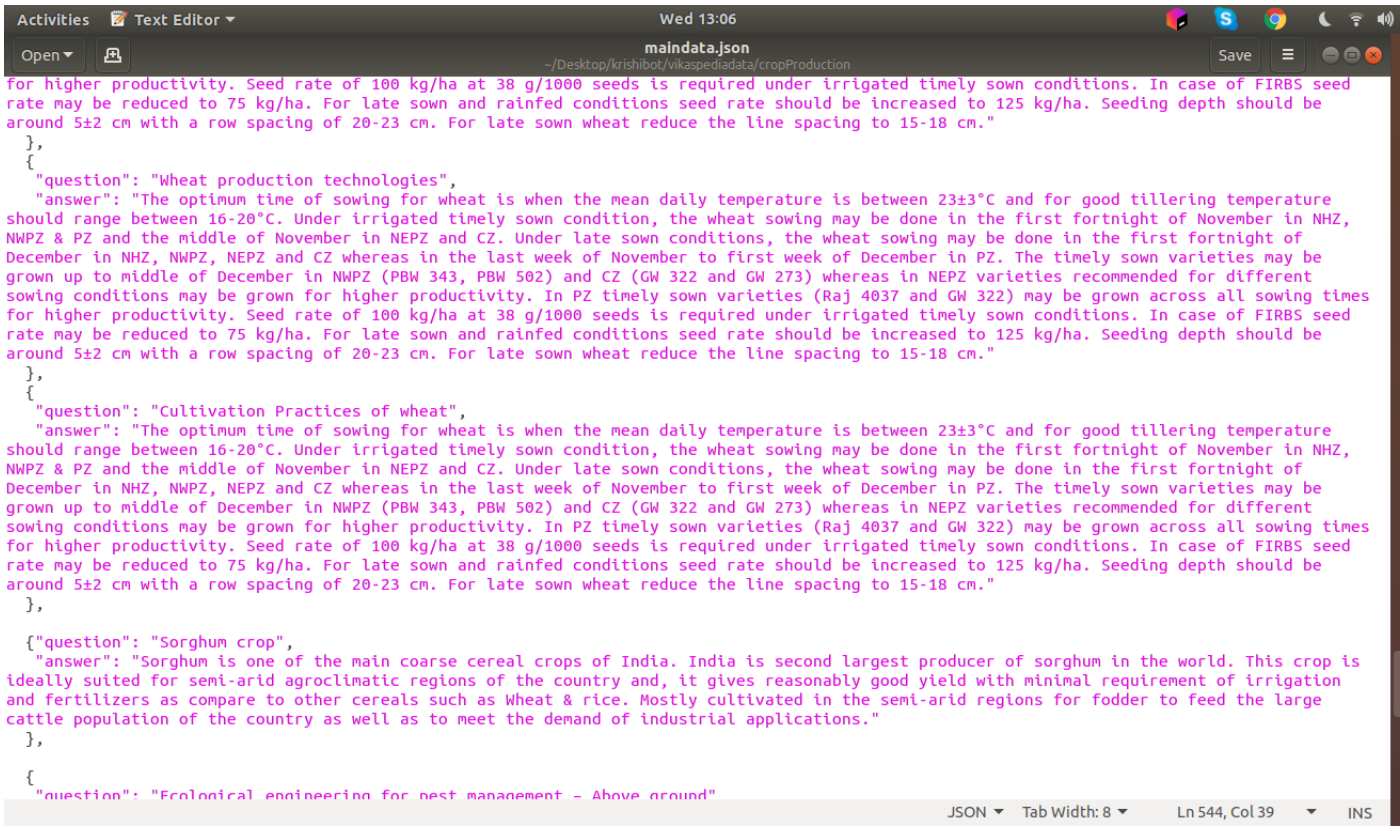

Figure A.1: Data Stored

In this Figure A.1 we can see that data are in the json format.

A.2 Output of the Krishibot

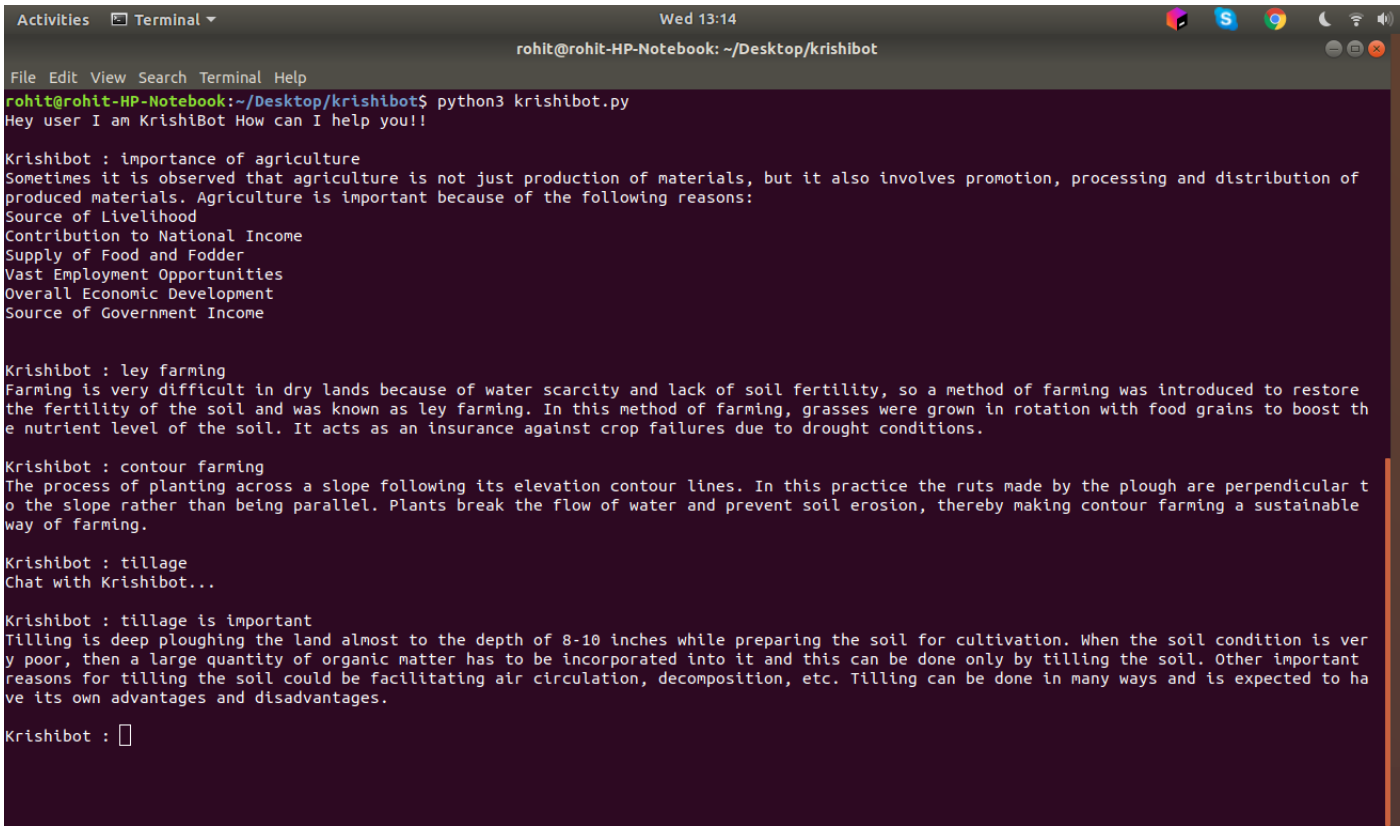

Figure A.2: Output

In this Figure A.2 we can see that how the krishibot is responding based on the keyword matching.

Appendix B

User manual

B.1 Step to install my implemented system

sudo apt-get update
sudo apt-get upgrade
pip install numpy
pip install random
pip install os
pip install json
pip install timeit
pip install pyttsx3
pip install pickle
pip install cosine_similarity
pip install TfidfVectorizer

Bibliography

[1] Abraham. Silberschatz, H. F. Korth, and S. Sudarshan, *Database System Concepts*, ee ed. (2013).

[2] Manning, C.D., Raghavan, P., Schütze, H.: *I*ntroduction to Information Retrieval. Cambridge University Press, New York (2008)

[3] Yan, Z., et al.: DocChat: an information retrieval approach for chatbot engines*u*sing unstructured documents. In: Proceedings of the 54th*A*nnual Meeting of the Association for Computational Linguistics*(*Volume 1: Long Papers), vol. 1, pp. 516–525 (2016)