TASK P – 5.2 P

LUIS and QnA Maker

Summary of Week 3
Learning

Saurabh Dharmadhikari
Saurabh.dharma01@gmail.com

# Contents

NLP (Natural language processing):

It is a field in Computer science and AI. NLP refers to the ability of computers to understand languages used by humans. It may also be used to generate appropriate responses to questions asked by humans.

Structured and unstructured data:

Speech or text are unstructured. It does not fit in rows and columns. Most data in actual world are unstructured. Thus, we need to structure this data in such a way that a machine can process and understand it. In NLP we deal with a lot of unstructured data and various techniques are used to structure this data that we have discussed latter in this document.
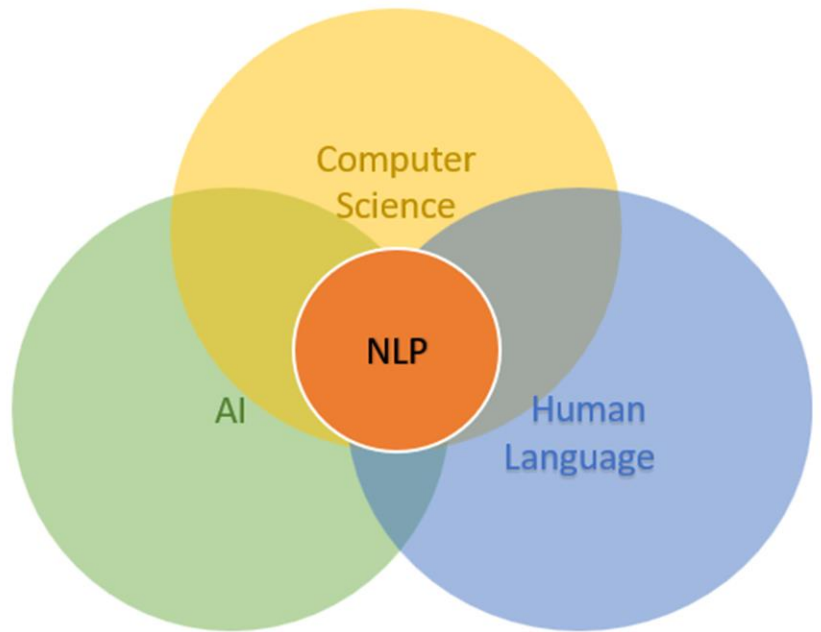
Structured data is data that is in rows and columns and thus it has specific features and entries.

Impact of NLP:

Language translator, auto correct and auto complete are some of the impacts that NLP has brought in our day-to-day work and made our lives easier. Also, some popular virtual assistant such as Siri, Alexa and Google use NLP to create human like responses and assist today humans in their daily tasks.

With such uses and applications of NLP, it is penetrating our lives and in future it will develop more. It is hard to tell already where human is assisting in a task and where is a machine responding. Thus, it is very important for creator of these chatbots to follow certain good principles.

Application of NLP:

a. Sentiment analysis

b. Speech recognition

c. Text classification

d. Text summarization

e. Building chatbots

NLP techniques:

a. Bag of words (BOW): It is a collection of words with their frequency and it does not represent their order or grammar.

b. TF-IDF: This technique helps identify the most relevant words in a sentence or document. It is calculated by multiplying frequency of a word with inverse frequency of word in the corpus.

c. Tokenization: Tokens are created by breaking down the sentence or document. Each token represents a word or sequence of word that have similar meaning.

d. Stemming: Slicing of words. It is reducing the word to its base form. For example, "running" or "ran" is "run" and "went" or "going" is "go"

e. Lemmatization: Is reducing the word to its base form but with the context it is in the text or speech. For example, "Red Bull" may mean a literal red bull or an energy drink called "Red bull". There are many other nuances in natural human language which are not easy to understand with their contexts. Lemma is the reduced word with its context.

Cloud based NLP:

There are many cloud-based NLP systems offered by various companies. Cloud based NLP is technology that is hosted and delivered by internet through cloud computing platforms. Using this a user does not need to build his or her own infrastructure.

Advantages of cloud-based NLP:

a. Scalability: It is able to scale up or down according to traffic it observes.
b. Cost effective: User has to pay only for the resources that he uses.
c. Flexibility: These can easily be integrated with other platforms to create better solution.
d. Accessibility: User can access models developed remotely and it helps in collaboration.

Examples of cloud-based NLP services:

a. Dialogflow.com (API.AI): It is a service owned by Google and provides a platform that enables user to create chatbots and personal assistants on smart devices.

b. Wit.AI: It is owned by Facebook and provides services to understand and generate human like responses

c. LUIS.AI: This is provided by Microsoft from Azure cognitive services.

d. Pandorabots: Pandorabots is an open-source platform that uses AIML (Artificial Intelligence Markup Language). It can be integrated to other platforms such as Facebook messenger.

Chatbot:

Chatbots are an artificial program that uses NLP to simulate human like conversation. They can be rule based which have predefined rules or can be governed by machine learning algorithms that enable them to learn over time.

These have an automated response to the questions asked by the end user. These are deployed in various industries now a days like healthcare, finance, education and more.

Uses of intelligent chatbot:

a. Customer support: It can provide custom services to customer for a business-like answering question.

b. Information retrieval: It can perform search in knowledge base to find relevant information to the questions asked.

c. Appointment scheduling: Chatbots can prepare schedules for the user to fix meetings and manage appointments in calendar.

d. Training and education: Increasingly chatbots are being used by many institutes to provide interactive learning experiences.

Limitations of chatbot:

a. Lack human like understanding: Chatbots are not yet capable of understanding complex ambiguous statements and may misinterpret intent and entity that are not true.

b. Limited domain knowledge: They can not handle queries outside their knowledge base.

c. Lack emotional intelligence: Chatbots are fail to recognise an appropriate response for sensitive emotional subjects.

d. Need continuous improvement: Regular analysis, training and updates are necessary to improve the performance of the machine.

Principles of responsible AI:

Chatbot have still been able to come very close to simulate human like conversations and are improving day by day. This has brought a debate on how these services should be used- ethical and non-ethical practices of these machines. It has become increasingly important to define set of rules that are meant to discourage and avoid this technology to be misused.

Certain set of principles have been advised to developers around the world for the same purpose as below:

a. Be transparent about what the bot can (and can't) do
b. Make it clear that the user is communicating with a bot
c. Enable the bot to seamlessly hand-off to a human if necessary
d. Ensure the bot respects cultural norms
e. Ensure the bot is reliable
f. Respect user privacy
g. Handle data securely
h. Ensure the bot meets accessibility standards
i. Assume accountability for the bot's actions

Input components of chatbot:

Input can be in form of speech or text. There are three major components of input:

a. Utterance: It is the query by the user.
b. Intent: It defines the purpose of the query.
c. Entity: It is the specific information provided in the utterance that could define the parameters and variables.

Examples of utterance, intent and entity:

If there is an input from the user as a query- "What's the weather like in Raipur tomorrow?"
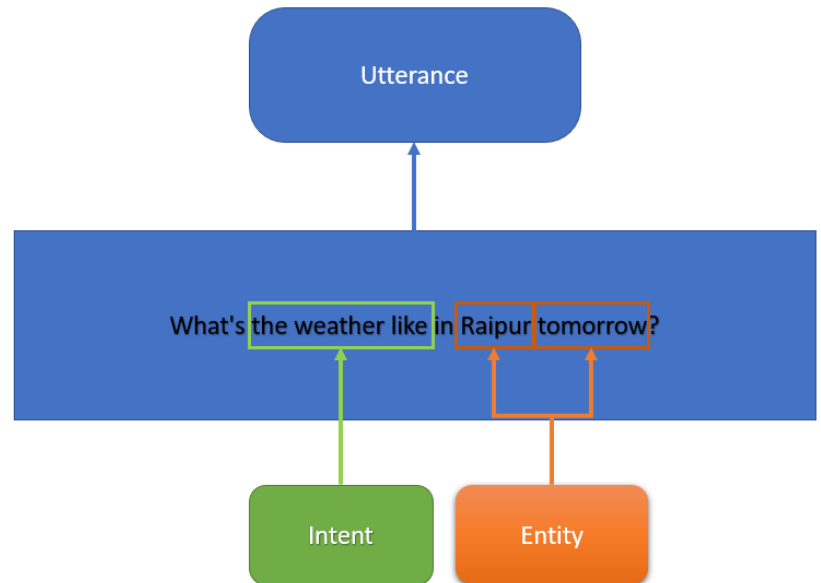
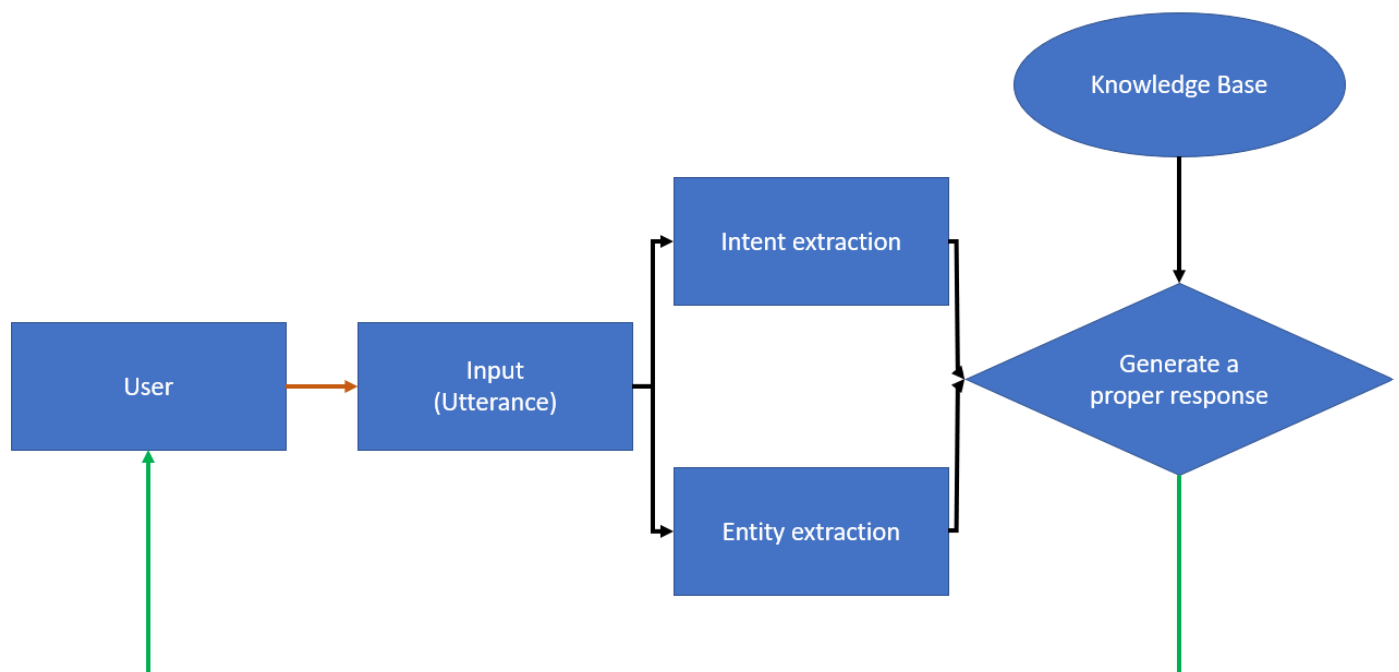Utterance: What's the weather like in Raipur tomorrow?

Intent: WeatherForecast

Entity:

      Location: Raipur

      Time: Tomorrow

How chatbot works:

Above image is an oversimplified way to describe how a chatbot works.

Once user puts in a query to the machine. The machine extracts intent and entity from the utterance using NLP techniques. ML then looks in the knowledgebase for the correct information that suits the query. Chatbot then generates a human like response to the query using NLP.

How to create LUIS:

To create a Language Understanding (LUIS) application, you can follow these general steps:

a. Identify domain: Determine the specific domain or area of application for your LUIS model. This involves understanding the types of user queries or intents you want to support.

b. Create intents and entities: Intents represent the intentions or actions behind user queries, while entities represent specific pieces of information within those queries. Define the intents and entities relevant to your application. For example, in a weather application, you may have intents like "GetWeather" and entities like "Location" and "Date."

c. Add utterances: Utterances are example user queries that help train the LUIS model. Provide a set of representative utterances for each intent, including variations and possible entity values. This allows LUIS to learn patterns and identify user intents accurately.

d. Train and test: Once you have added sufficient utterances, train the LUIS model. Training involves processing the provided data to build a language understanding model. After training, test the model by providing sample queries and verifying if the predicted intents and entities match the expected results.

e. Publish and deploy: After training and testing, publish the LUIS model to make it available for consumption. This involves deploying it as an API endpoint that can be accessed by applications or chatbots. You can integrate the LUIS endpoint into your application to receive predictions for user queries in real-time.