Natural Language Processing

Natural Language Processing

Introduction to Chatbots

Instructor: Moushmi Dasgupta

Connect up with me on LinkedIn
www.linkedin.com/in/moushmi1234

Some of the material is from Georgia Institute of Technology, Atlanta, GA, USA.

Session Plans

Date	Day	Topic	Time
2nd Nov 2022	Wednesday	Text Classification	2:00PM – 4:00PM
4th Nov 2022	Friday	Text Classification	2:00PM – 4:00PM
5th Nov 2022	Saturday	IE	9:30AM - 1:30PM
9th Nov 2022	Wednesday	IE	2:00PM – 4:00PM
11th Nov 2022	Friday	Chatbot	2:00PM – 4:00PM
13th Nov 2022	Saturday	Chatbot	9:30AM - 1:30PM

AGENDA

Information Extraction

Module 3

4.1 IE Applications

4.2 IE Tasks

4.3 The General Pipeline for IE

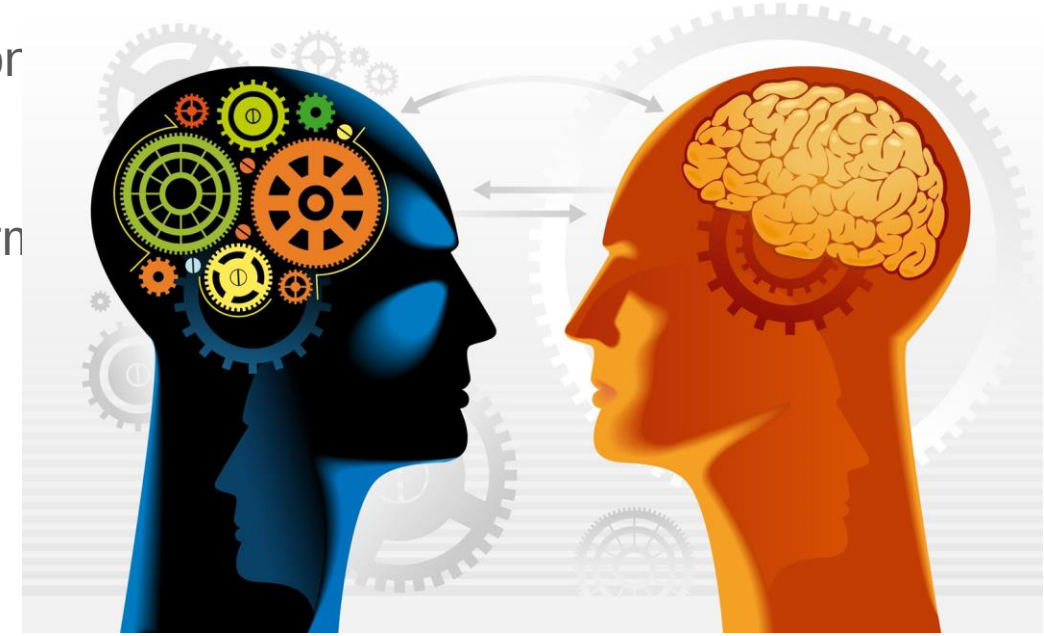
4.4 Case Study on IE

Pre-requisite:

1. Python programming
2. An understanding of Machine Learning
3. Fundamentals of NLP Understanding and hands-on with Python programming on and IDE
4. An Understanding of NLP Pipeline
5. Understanding of Text Classifications
6. Invest in attending classroom sessions (Weekly 1 or 2 classes of 3+ hours duration)
7. Invest in yourself with1 hour of self study everyday

Natural Language Processing

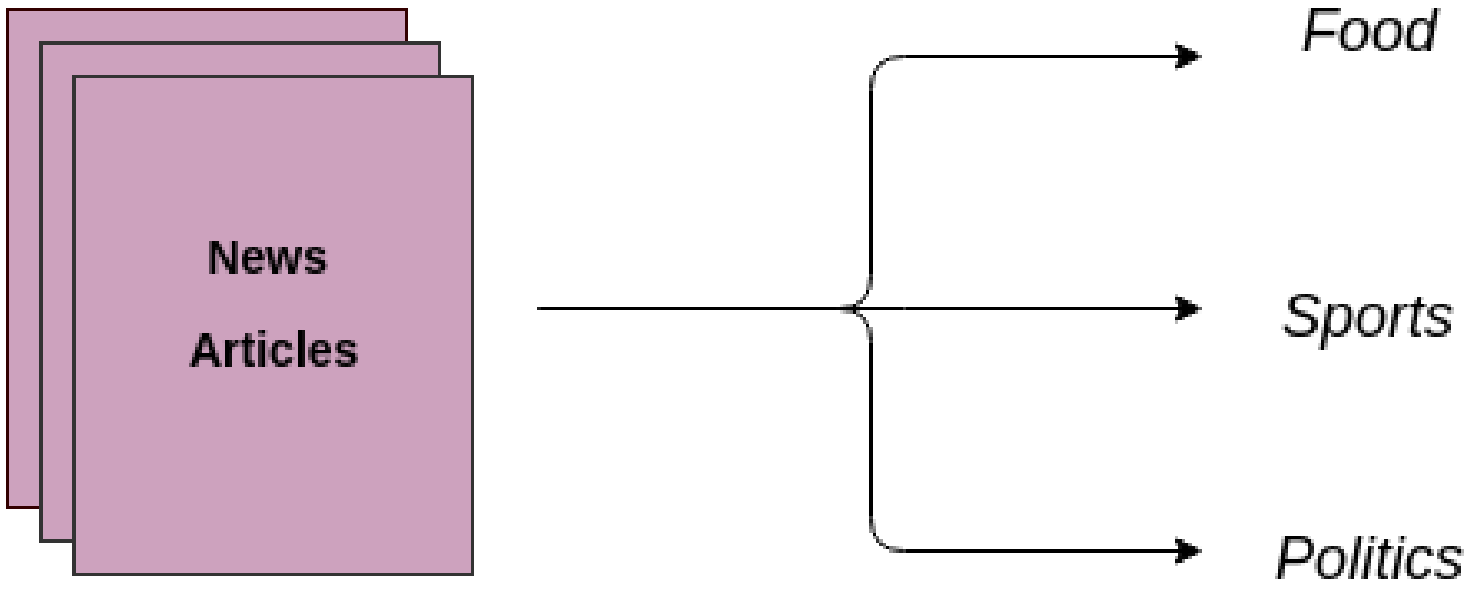
1. Natural Language Processing is a subfield of artificial intelligence concerned with methods of communication between computers and natural languages such as english, hindi, etc.
2. Objective of Natural Language processing is to perform useful tasks involving human languages like
    - Sentiment Analysis
    - Machine Translation
    - Part of Speech Tags
    - Human-Machine communication(chatbots)



NLP Pipeline Steps

Some of the material is from Georgia Institute of Technology, Atlanta, GA, USA.

Text Classification



3.2 Using Existing Text Classification APIs

Open-Source Libraries for Text Classification

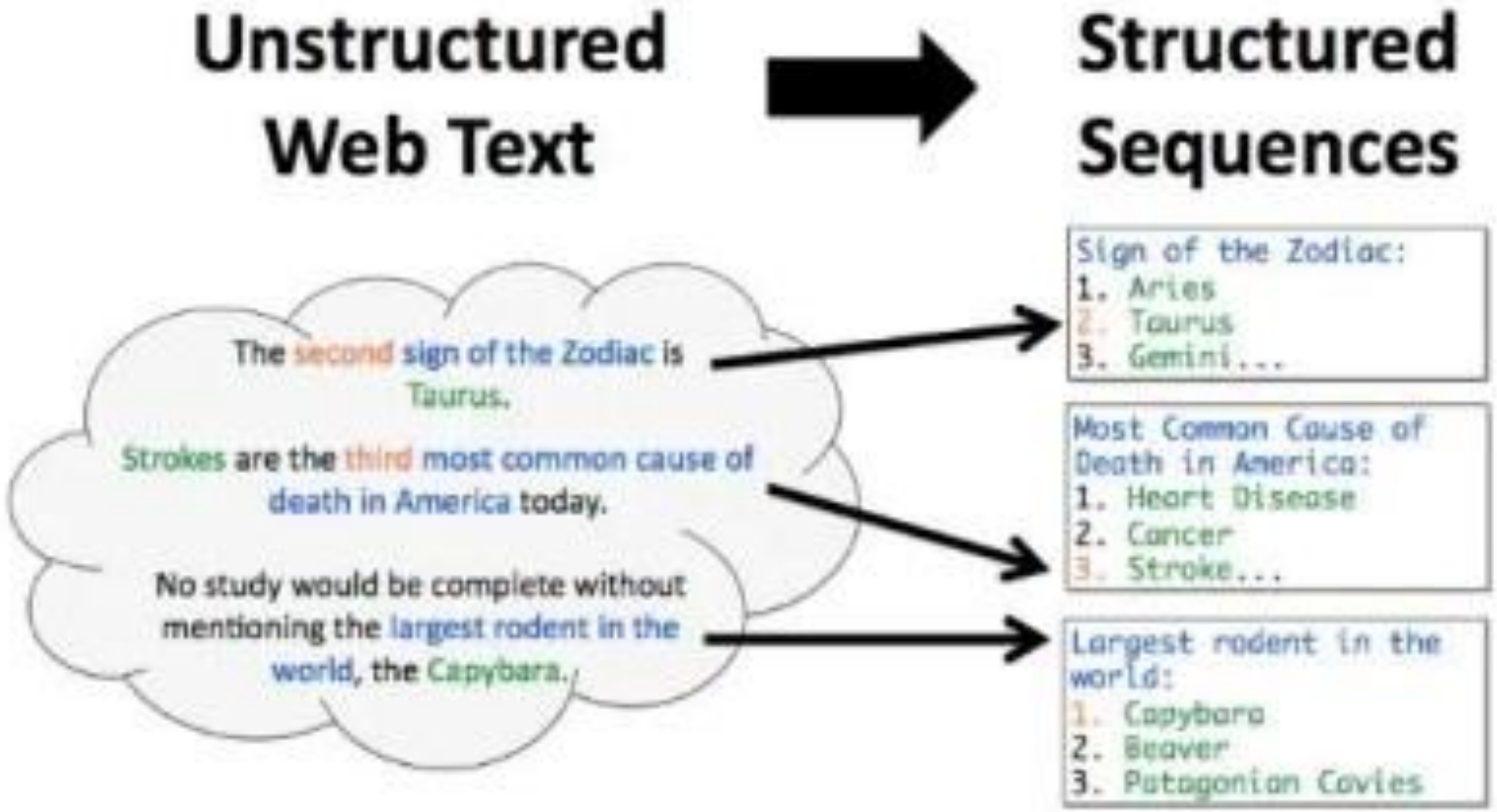
1.Scikit-learn
2.Natural Language Toolkit (NLTK)
3. SpaCy

(Other Open source Libraries of interest)
1.TensorFlow
2.PyTorch
3.Keras

Paid API's

Google Cloud NLP
IBM Watson
Lexaytics
Amazon Comprehend
Aylien

What is IE?


Unstructured
Web Text
Structured
Sequences
The second sign of the Zodiac is Taurus.
Strokes are the third most common cause of death in America today.
No study would be complete without mentioning the largest rodent in the world, the Capybara.
Sign of the Zodiac:
1. Aries
2. Taurus
3. Gemini...
Most Common Cause of Death in America:
1. Heart Disease
2. Cancer
3. Stroke...
Largest rodent in the world:
1. Capybara
2. Beaver
3. Patagonian Cavies

Get Familiar with these concepts

NAME	DESCRIPTION
Tokenization	Segmenting text into words, punctuations marks etc.
Part-of-speech (POS) Tagging	Assigning word types to tokens, like verb or noun.
Dependency Parsing	Assigning syntactic dependency labels, describing the relations between individual tokens, like subject or object.
Lemmatization	Assigning the base forms of words. For example, the lemma of “was” is “be”, and the lemma of “texts” is “text”.
Sentence Boundary Detection (SBD)	Finding and segmenting individual sentences.

Get Familiar with these concepts

NAME	DESCRIPTION
Named Entity Recognition (NER)	Labelling named “real-world” objects, like persons, companies or locations.
Entity Linking (EL)	Disambiguating textual entities to unique identifiers in a knowledge base.
Similarity	Comparing words, text spans and documents and how similar they are to each other.
Text Classification	Assigning categories or labels to a whole document, or parts of a document.
Rule-based Matching	Finding sequences of tokens based on their texts and linguistic annotations, similar to regular expressions.
Training	Updating and improving a statistical model’s predictions.
Serialization	Saving objects to files or byte strings.

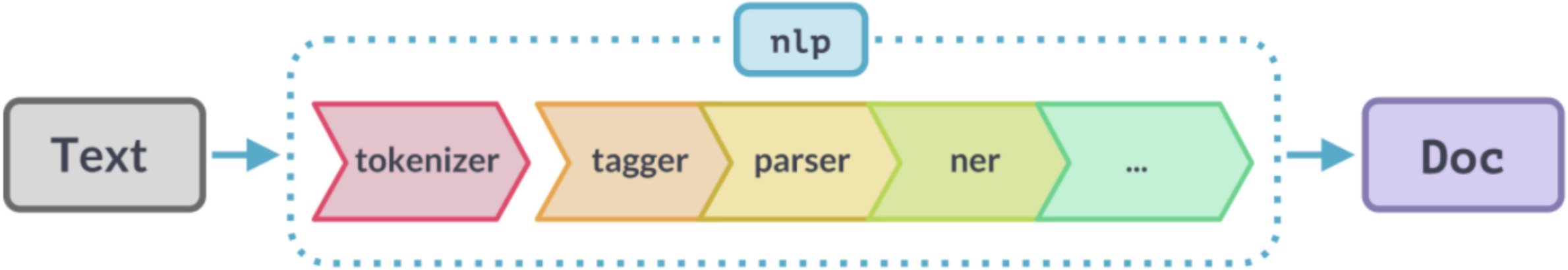
4.1 IE Applications

Identify specific pieces of information (data) in a unstructured or semi-structured textual document.
Transform unstructured information in a corpus of documents or web pages into a structured database.

Applied to different types of text:

- Newspaper articles
- Web pages
- Scientific articles
- Newsgroup messages
- Classified ads
- Medical notes

4.3 The General Pipeline for IE



Evaluating IE Accuracy

Always evaluate performance on independent, manually-annotated test data not used during system development.

Measure for each test document:

Total number of correct extractions in the solution template: *N*

Total number of slot/value pairs extracted by the system: *E*

Number of extracted slot/value pairs that are correct (i.e. in the solution template): *C*

Compute average value of metrics adapted from IR:

Recall = *C*/*N*

Precision = *C*/*E*

F-Measure = Harmonic mean of recall and precision

Demo Application

Introduction to Chatbot

A bot and a Chatbot

Although the terms chatbot and bot are used interchangeably, there's a significant difference between them.

A **chatbot** is a computer program designed to communicate with users. It analyzes users' questions to provide matching answers. Businesses use chatbots to support customers and help them accomplish simple tasks without the help of a human agent.

A **bot** is an algorithm that interacts with web content. Bots help businesses and users perform helpful, mundane, or complex tasks faster online. Below are some different types of bots.:

A Bot

A bot is an algorithm that interacts with web content. Bots help businesses and users perform helpful, mundane, or complex tasks faster online. Below are some different types of bots.:

•Search engine bots called crawlers are used by Google and Yahoo to index the web content and scale web cataloging. This helps users easily find information related to their search intent.

•Feed bots look for new information on the web to add to news sites.

•Copyright bots look for content that violates copyright laws. They help companies and authors check whether their proprietary content has been used without approval.

What is a chatbot?

Chatbots or virtual assistants, or conversational AI assistants are some of the popular buzzwords today and every organization/industry is aping towards adapting them.

- A chatter robot (chatbot) is a computer program designed to simulate an intelligent conversation with human users in natural language via voice or textual methods.
- Its key task is to help users by providing answers to their questions.

- Chat Bots can chat with multiple users simultaneously and provide information within seconds.

- Because of that, they are now used on a wide scale to help businesses and consumers communicate with each other on websites and mobile messaging apps.

Types of Chatbots

Types of Chatbots

There are many types of chatbots available. A few of them can be majorly classified as follows:

- ***Text-based chatbot***: In a text-based chatbot, a bot answers the user's questions via a text interface.

- ***Voice-based chatbot***: In a voice or speech-based chatbot, a bot answers the user's questions via a human voice interface.

Types of Chatbots based on their design

There are mainly two approaches used to design the chatbots, described as follows:

•**Rule-based** Chatbots – This type of chatbot answers questions based on some rules on which it is trained on. The rules defined can be very simple to very complex. The bots can handle simple queries but fail to manage complex ones.

•**AI Based Chatbots -** are the ones that use some Machine Learning-based approaches and are definitely more efficient than rule-based bots. These bots can be further classified into two types: Retrieval Based or Generative.

The brief history of chatbots

1950 Alan Turing — the man that started it all

In 1950 Alan Turing, a computer pioneer, wrote a scientific paper titled “Computing Machinery and Intelligence.” In the paper, the scientist implied that a computer program can think and talk like a human. Turing proposed an experiment called the Imitation Game, which is known as the Turing Test, to prove the point. In the Turing experiment, the person designated as a judge was chatting over a computer with a human and a machine who could not be seen.

Eliza – the first chat bot made by *Joseph Weizenbaum.*

Eliza Effect

tendency of humans to attach associations to terms from prior experience.

Working of Eliza is based on

Knowledge Representation
Pattern Recognition
Substitution of key words into known phrases.

1966 ELIZA — the first chatbot

In 1966, an MIT professor, Joseph Weizenbaum, developed a computer program called Eliza. It's considered to be the first chat robot in history.

Eliza was a simple keyword-based conversational interaface that mimicked a human psychiatrist. The program communicated by matching user questions with scripted responses entered into its database.

When a patient would say, “My mother loves flowers,” Eliza would reply, “Tell me more about your mother.”

1971 Parry

In 1971, Kenneth Colby, a Stanford Artificial Intelligence Laboratory psychiatrist, wondered whether computers could contribute to understanding brain function.

He believed that the computer could help in treating patients with mental diseases.

These thoughts led Colby to develop Parry, a computer program that simulated a person with schizophrenia.

Colby believed that Parry could help educate medical students before they started treating patients. Parry was considered the first chat robot to pass the Turing Test. Back then, its creation initiated a serious debate about the possibilities of artificial intelligence.

Fun fact:

In 1978, Colby developed the first intelligent speech prosthesis. It was a computer program that helped people with communication disorders to speak.

1988 Jabberwacky

In 1988, a self-taught programmer called Rollo Carpenter created Jabberwacky.

It was a program designed to simulate human conversation entertainingly.

Jabberwacky learned from past experiences and developed over time.

It reflected users' personalities and behaviors.

1992 Dr.Sbaitso

In 1992, Creative Labs, a technology company based in Singapore, developed Dr. Sbaitso.

It was an AI speech synthesis program that imitated a psychologist.

The program was distributed with sound cards sold by the company.

They wanted to show the digitized voices their cards were able to produce.

1995 A.L.I.C.E

Developed in 1995 by Richard Wallace, Alice was an NLP application that simulated a chat with a woman.

Wallace Alice was inspired by Eliza and designed to have a natural conversation with users.

Its code was released as open-source, which means it can be reused by other developers to power their conversational interfaces.

Fun fact:

Alice was an inspiration for an American science-fiction romantic drama Her. It's a film about a man, Theodore Twombly, who falls in love with a chatbot.

2001 SmarterChild

SmarterChild was an intelligent chat interface built on AOL Instant Messenger in 2001 by ActiveBuddy, the brand creating conversational interfaces.

SmarterChild was designed to have a natural conversation with users.

It's considered to be a precursor to Apple's Siri.

2010 Virtual Assistants

Since 2010, when Apple launched Siri, virtual assistants have been on the rise.

Siri was the first personal assistant available worldwide.

Google followed in Apple's footsteps by releasing Google Now in 2012.

Microsoft's Cortana and Amazon's Alexa were both released in 2014.

2016 Chatbot platforms

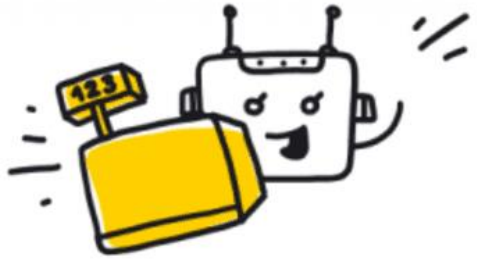
In 2016, Facebook opened its Messenger platform for chatbots. This helped fuel the development of automated communication platforms.

In 2018, LiveChat released ChatBot, a framework that lets users build chatbots without coding.

So far, there have been over 300,000 active bots on Messenger.

Applications of Chatbots

Marketing



Sales

Support

And many more!

Fun fact:

If you block bots on your website, they won't be able to index your content in search engines. That means that users won't be able to find your website.

How do chatbots work?

Chatbots are pre-programmed responses, artificial intelligence, or both.

Based on the applied mechanism, they process a user's question to deliver a matching answer.

Mainly two types of Chatbots exist based on how they are designed:

Rule Based Chatbot
AI Based Chatbot

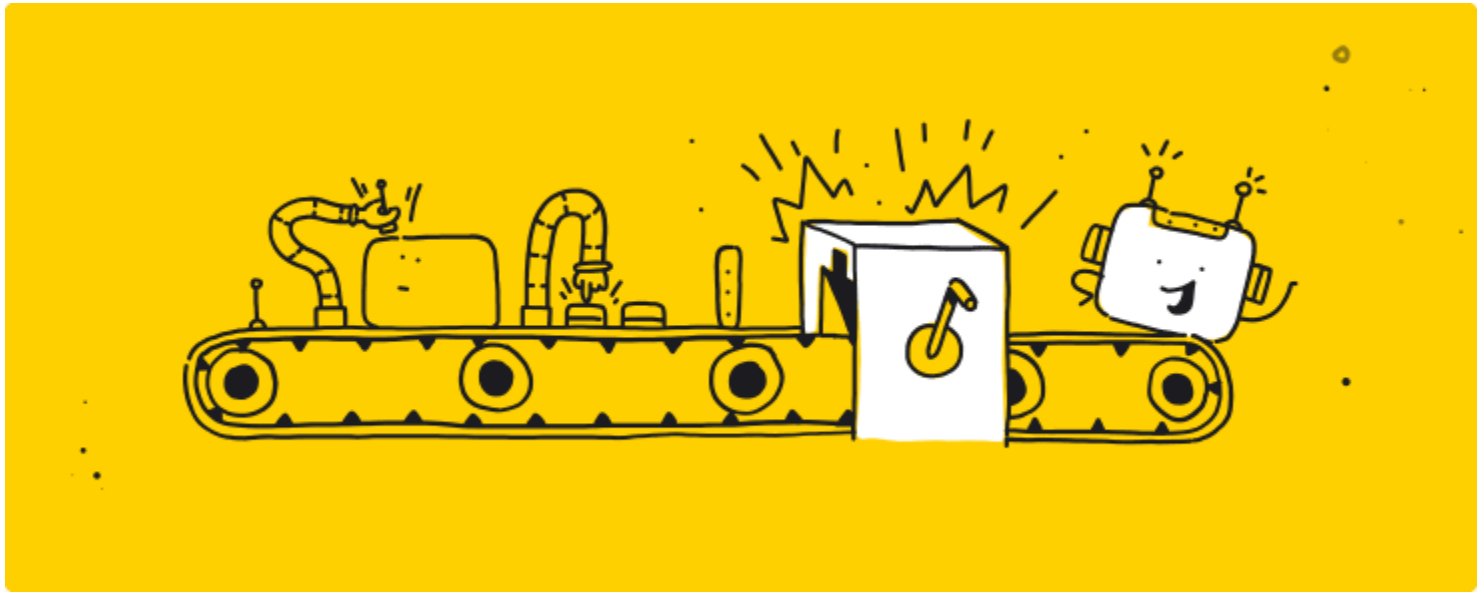
Difference between a Chatbot and AI Bot

A chatbot is a conversational interface that communicates with users via text or voice.

On the other hand,

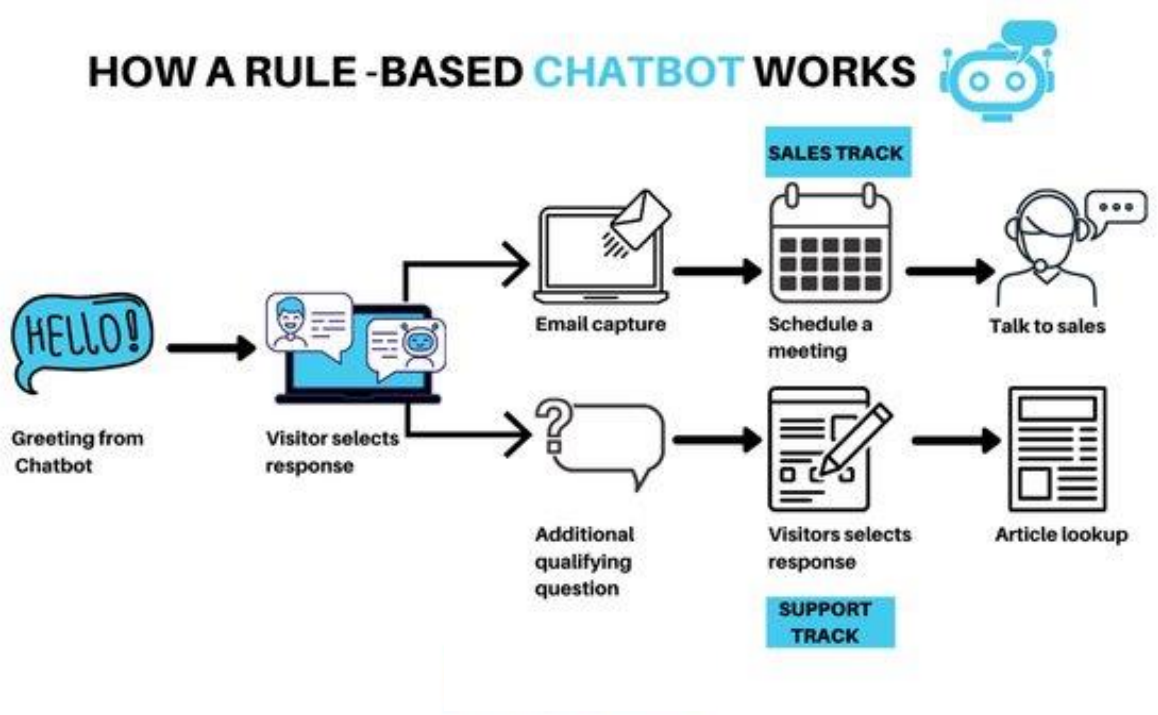
the bot is a type of software that automates repetitive tasks like searching for information and adding it to sites, but it doesn’t chat with users.

The difference between a Rule based Chatbot and an AI Chatbot



Rule Based Chatbots (Scripted Chatbots)

Rule-based AI agents work based on definite rules and provide pre-defined replies. They can only answer correctly the questions they were programmed to answer. Rule-based bots can’t improve independently.



Rule Based Chatbot

Rule-based (also command-based, keyword, or transactional) chatbots communicate using predefined answers.

They can be playfully compared to movie actors because, just like them, they always stick to the script. Rule-based bots provide answers based on a set of if/then rules that can vary in complexity.

These rules are defined and implemented by a chatbot designer.

Rule-based chatbots don't understand the context of the conversation.

They provide matching answers only when users use a keyword or a command they were programmed to answer.

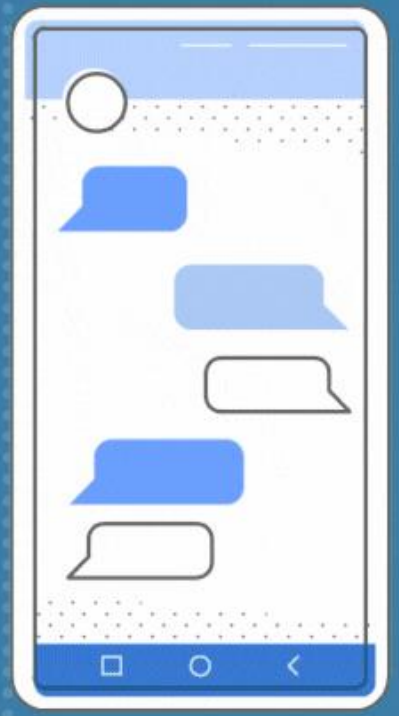
Rule Based Chatbot

When a rule-based bot is asked a question like, “Book a flight ticket to London?” it first looks for familiar keywords in the sentence.

In this example, ‘flight,’ , “book” and ‘London' are the keywords. Then, it matches these keywords with responses available in its database to provide the answer.

However, if anything outside the AI agent's scope is presented, like a different spelling or dialect, it might fail to match that question with an answer.

Because of this, rule-based bots often ask a user to rephrase their question. Bots can also transfer a customers to a human agent when needed.



Lets create a Rule based Chatbot in Python Over to Google Colab

AI based Chatbots

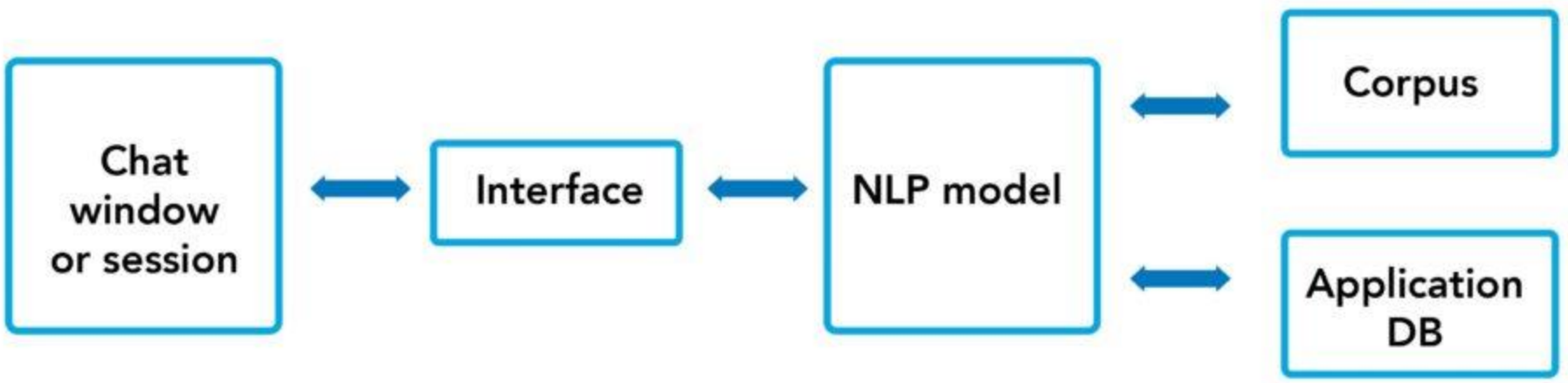
An AI chatbot is **an online program that uses NLP (natural language processing) to simulate human conversations on websites and mobile apps and is powered by artificial intelligence**.

Architecture of an AI Chatbot

The Architecture of chatbots

Typical chatbot architecture should consist of the following:

•Chat window/session/front end application interface
•The deep learning model for Natural Language Processing [NLP]
•Corpus or training data for training the NLP model
•Application Database for processing actions to be performed by the chatbot



Natural Language Understanding (NLU)

Natural language understanding is a branch of artificial_intelligence that uses computer software to understand input in the form of sentences using text or speech.

NLU

NLU enables human-computer interaction.

It is the comprehension of human language such as English, Spanish and French, for example, that allows computers to understand commands without the formalized syntax of computer languages.

NLU also enables computers to communicate back to humans in their own languages.

The main purpose of NLU is to create chat- and voice-enabled bots that can interact with the public without supervision.

Many major IT companies, such as Amazon, Apple, Google and Microsoft, and startups have NLU projects.

Natural language understanding applications

NLU is the basis of speech recognition software -- such as Siri on iOS -- that works toward achieving human-computer understanding.

IVR and message routing. Interactive Voice Response (IVR) is used for self-service and call routing.

Early iterations were strictly touchtone and did not involve AI. However, as IVR technology advanced, features such as NLP and NLU have broadened its capabilities and users can interact with the phone system via voice.

The system processes the user's voice, converts the words to text, and then parses the grammatical structure of the sentence to determine the probable intent of the caller.

Customer support and service through intelligent personal assistants.
NLU is the technology behind AI chatbots which is a computer program that converses with a human in natural language via text or voice.

Natural language understanding applications

Machine translation.

Machine learning (ML) is a branch of AI that enables computers to learn and change behavior based on training data.

Machine learning algorithms are also used to generate natural language text from scratch. In the case of translation, a machine learning algorithm analyzes millions of pages of text -- say, contracts or financial documents -- to learn how to translate them into another language.

The more documents it analyzes, the more accurate the translation. For example, if a user is translating data with an automatic language tool such as a dictionary, it will perform a word-for-word substitution. However, when using machine translation, it will look up the words in context, which helps return a more accurate translation.

Natural language understanding applications

Data capture.

Data capture is the process of gathering and recording information about an object, person or event. For example, if an e-commerce company used NLU, it could ask customers to enter their shipping and billing information verbally. The software would understand what the customer meant and enter the information automatically.

Conversational interfaces.

Many voice-activated devices -- including Amazon Alexa and Google Home -- allow users to speak naturally. By using NLU, conversational interfaces can understand and respond to human language by segmenting words and sentences, recognizing grammar, and using semantic knowledge to infer intent.

Natural Language Generation (NLG)

NLG enables computers to automatically generate natural language text, mimicking the way humans naturally communicate -- a departure from traditional computer-generated text.

Generally, computer-generated content lacks the fluidity, emotion and personality that makes human-generated content interesting and engaging.

NLG can use NLP so that computers can produce humanlike text in a way that emulates a human writer.

This is done by identifying the main topic of a document, and then using NLP to determine the most appropriate way to write the document in the user's native language.

Text is generated based on this decision.

Lets move on to Google Colab for an application on AI based Chatbot

Thankyou and it was a pleasure interacting with you.
Stay in touch!

Moushmi