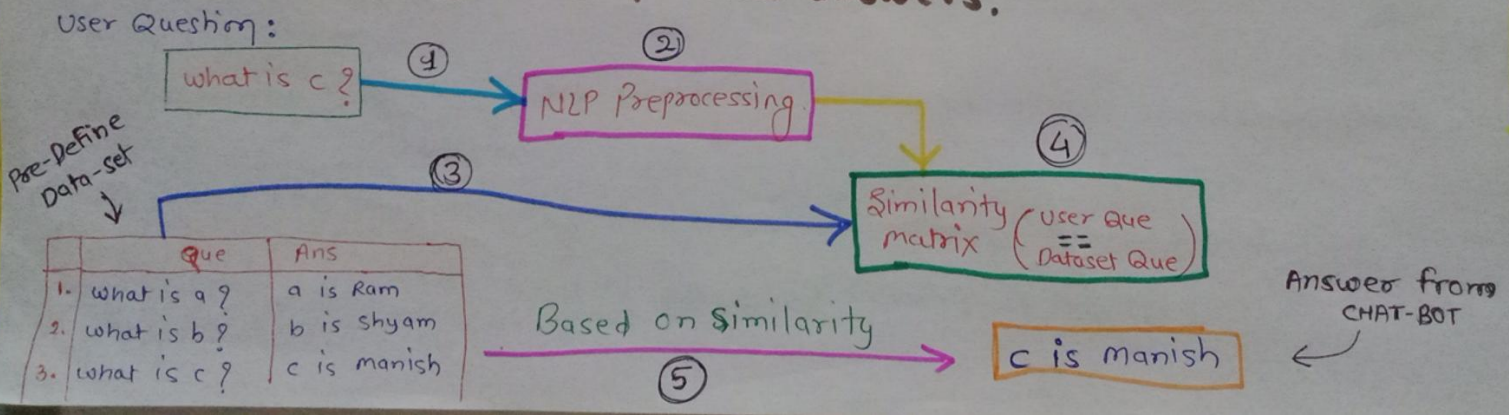
CHAT-BOT

* **CHAT-BOT** :- chatbot are soft. application that use artificial intelligence and NLP technique, in which you ask a question and based on some rules, this system provide you an answer.

* **Types of CHAT-BOT** :-

→ **Retrieval Based (Rule Based)** :- with this type of chat-bot, communication work by having a pre-set of question/ Answer rules. Based on similarity in input que. and data set que., model provide answers.



→ Generative model Based (Self-learning) :- This chat-bot work by generating new response using ML/DL on a lot of historical data and previous conversation, whitout needing a pre-define dataset. 2)

3)

* Basic NLP Pre-Processing steps *

* Tokenization * Stemming * Lemmatization * Vectorization
* Word Embedding * NER * Text classification

↳ Tokenization : Tokenization is the task of breaking or seperating the whole sentence into words / appostopy or punctuation.

Exp :- This is a dog → [This] [is] [a] [dog]

split the string (text) using space (" ")

Exp: We're friends → [we] ['re] [friends]

Appostopy Problem

Exp: What is your name ? → [What] [is] [your] [name] [?]

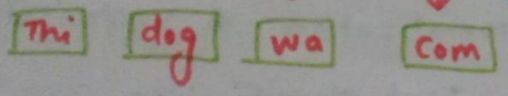
Punctuation Problem seperate token consider

↳ Stemming : → This take input as tokenized output words in form of array / list.

→ Stemming remove suffix / Pre-fix from word and improve the quality of word which is in list.

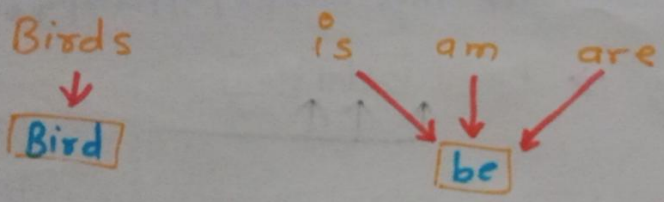
Manish Bendale

4) Ex: This dog was comming



↳ Lemmatization :- Lemmatization find the root of the words, not just remove the affix from the words.

Ex



↳ Named Entity Recognition :- A task of NLP, whose purpose is to locate and classify named entities in string into pre-defined categories.

Exp: Precident Abdul Kalam (Person) of India (Loc) on his first (ordinal) state visit to the united state (Location) on tuesday (Date) night. (Time)

Manish Bendale

5)

2> Vectorization : vectorization is a process which is used to turning a text document into a numerical vector.

→ Basic approch are:

- Bag of words (countvectorizer)
- TF-IDF (Term frequency inverse term frequency)
- word2vec

ex: "Bag of word" —

1. first define fixed length vocabulary.

["I", "am", "you", "are", "manish"]

2. map each word to an index in voc.

[I=1, am=2, you=3, are=4, manish=5]

3. Based on this index, construct a vector in which the word index is a 1 if the word in document else 0.

exp: "I am manish" → [1, 1, 0, 0, 1]

I am you are manish

Manish Bendale