PHRASE DETECTION

Given a query related reminders, detect the custom message/phrase that user wants to set the reminder for (if present).

Eg:

Query: Please remind me to go to gym.

Reminder phrase: “go to gym”

Solution :-

This is problem statement of Natural Language Processing (NLP).

Here I will implement this using Python in Jupyter Notebook.

First, I am importing libraries like **nltk , csv ,pandas , numpy.**

Nltk is a NLP Toolkit.

Then I am reading the data using read_csv command and giving the file name ‘training_data.tsv’

Then I am checking the length of data using len(data).

I am using Textblob because **TextBlob** is a library for processing textual data which I am going to give in this problem.

Now I am Tokenizing sentence into individual words and every word is assigned a tag.

I am defining Nouns and Verbs as important words so that they can be differentiated in test sentence easily.

For statistical analysis of text or speech content to find n (a number) of some sort of item in the text , we are using N-Grams

I am using function to extract the phrase from given input sentence. But for this I am defining a grammar because the phrase will be given as output accordance with the grammar defined. I am creating a parser from the grammar defined so that I can know the sentence given as input. I am removing the the keywords which are not used in training data, So I am not using those in test set also. I am generating the phrase using the grammer and the training set and generating phrase for every sentence.

Now , I am reading the text file which needs to be evaluated and passing it through the function and sending the the evaluation to the new file which I create using **open('result.csv', mode='w', newline='').**

Accuracy :- For accuracy , I am checking the accuracy using

Accuracy :- count/len(data))*100

Count :- The count is initiated with 0 and incremented when data in the ‘result.csv’ is matching with the predicted value by the algorithm in the ‘result.csv.

Len(data) :- It stores the total number present in the dataset.