Experiment No - 04

Aim:- Study the different POS taggers and Perform POS tagging on the given text

Theory:-

Tagging is a kind of classification that may be defined as the automatic assignment of description to the tokens. Here the descriptor is called tag, which may represent one part-of-speech, semantic information and so on.

Now, if we talk about Part-of-Speech (PoS) tagging, then it may be defined as the process of assigning one of the parts of speech to the given word. It is generally called POS tagging. In simple words, we can say that POS tagging is a task of labeling each word in a sentence with its appropriate part of speech. We already know that parts of speech include nouns, verbs, adverbs, adjectives, pronouns, conjunction and their sub-categories.

Most of the POS tagging falls under Rule Base POS tagging, Stochastic POS tagging and Transformation based tagging.

Rule-based POS Tagging

One of the oldest techniques of tagging is rule-based POS tagging. Rule-based taggers use a dictionary or lexicon for getting possible tags for tagging each word. If the word has more than one possible tag, then rule-based taggers use hand-written rules to identify the correct tag. Disambiguation can also be performed in rule-based tagging by analyzing the linguistic features of a word along with its preceding as well as following words. For example, suppose if the preceding word of a word is an article

then the word must be a noun.

As the name suggests, all such information in rule-based POS tagging is coded in the form of rules. These rules may be either –

- Context-pattern rules
- Or, as Regular expression compiled into finite-state automata, intersected with lexically ambiguous sentence representation.

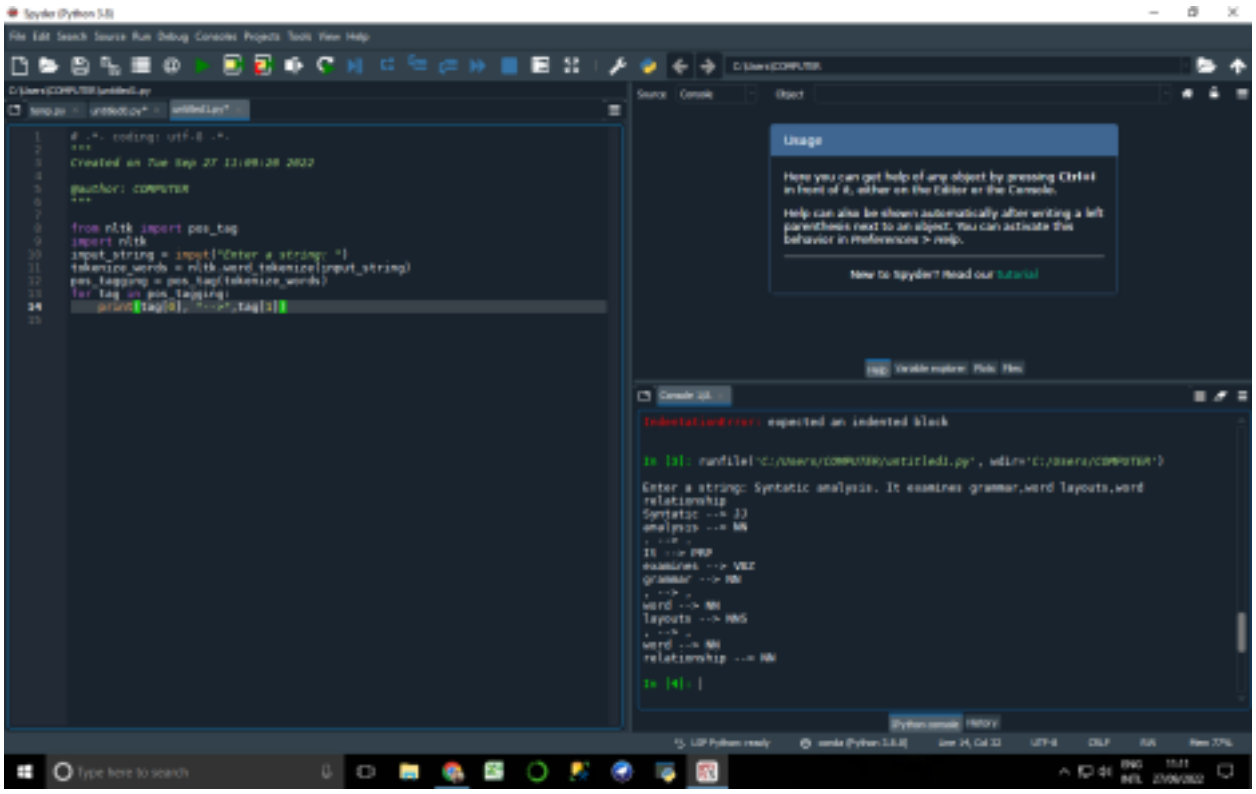
We can also understand Rule-based POS tagging by its two-stage architecture –

- First stage – In the first stage, it uses a dictionary to assign each word a list of potential parts-of-speech.
- Second stage – In the second stage, it uses large lists of hand-written disambiguation rules to sort down the list to a single part-of-speech for each word.

Code:-

```
from nltk import pos_tag
import nltk
input_string = input("Enter a string: ")
tokenize_words = nltk.word_tokenize(input_string)
pos_tagging = pos_tag(tokenize_words)
for tag in pos_tagging:
    print(tag[0], "-->",tag[1])
```

Output:-



Conclusion :- Thus we have successfully implemented different POS taggers and Perform POS tagging on the given text.