Experiment No.6
Perform POS tagging on the given English and Indian Language Text
Date of Performance:
Date of Submission:

Aim: Perform POS tagging on the given English and Indian Language Text

Objective: To study POS Tagging and tag the part of speech for given input in english and an Indian Language.

Theory:

The primary target of Part-of-Speech (POS) tagging is to identify the grammatical group of a given word. Whether it is a NOUN, PRONOUN, ADJECTIVE, VERB, ADVERBS, etc. based on the context. POS Tagging looks for relationships within the sentence and assigns a corresponding tag to the word.

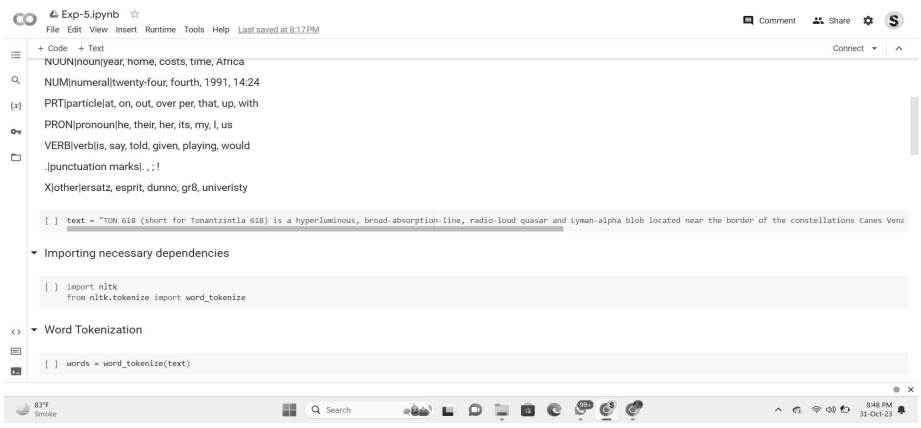
POS Tagging (Parts of Speech Tagging) is a process to mark up the words in text format for a particular part of a speech based on its definition and context. It is responsible for text reading in a language and assigning some specific token (Parts of Speech) to each word. It is also called grammatical tagging.

Steps Involved in the POS tagging example:

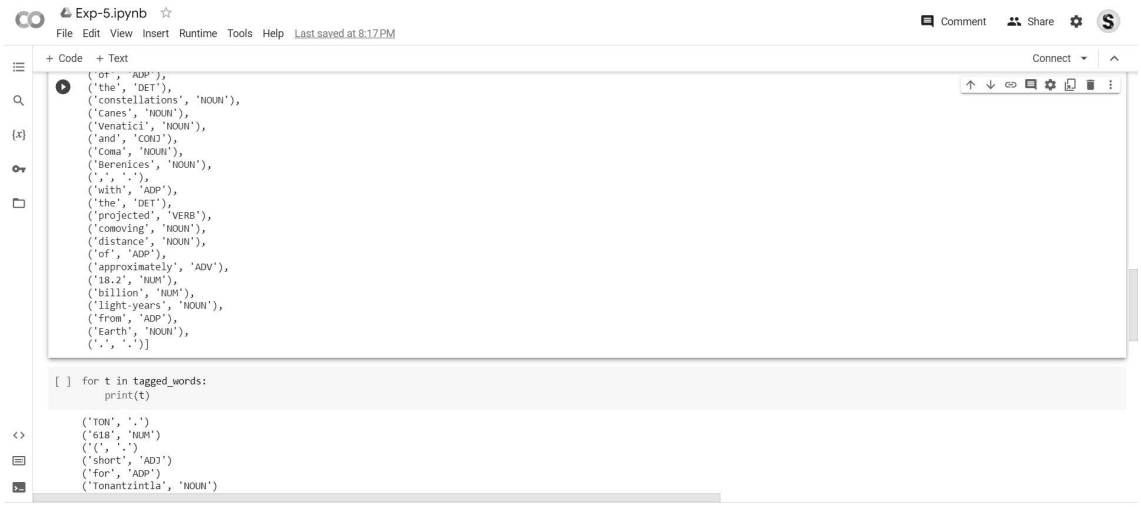
- Tokenize text (word_tokenize)
- apply pos_tag to above step that is nltk.pos_tag(tokenize_text)

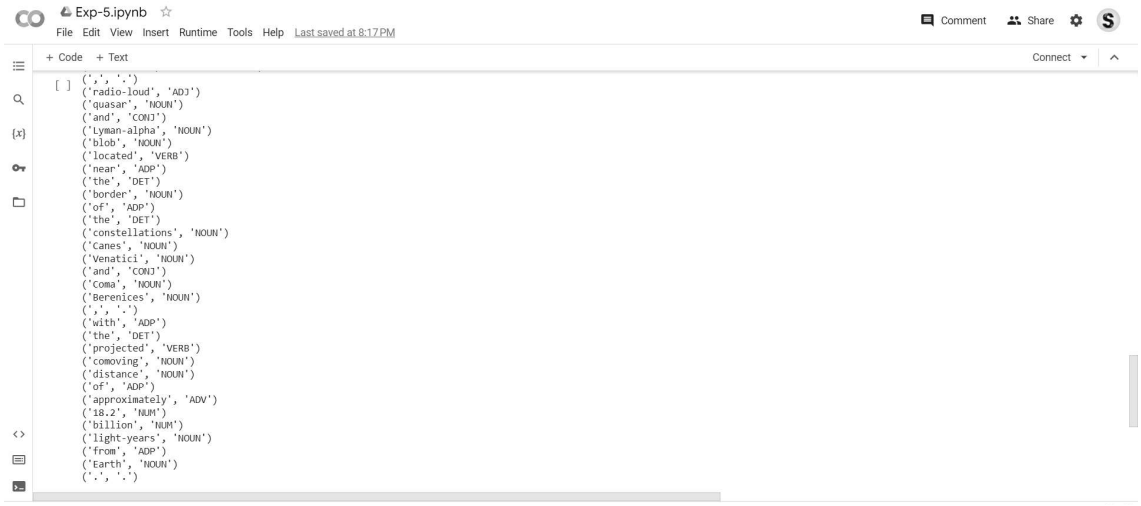
Implementation:


Exp-5.ipynb
File Edit View Insert Runtime Tools Help Last saved at 8:17 PM
Comment
Share
+ Code + Text
Connect
Experiment 05 ~ NLP DLOC ~ Sahil Gujral ~ Roll No. 15 ~ Dept. of CSE (DS) ~ VCET
Parts of Speech
Tag|Meaning|English Examples
ADJ|adjective|new, good, high, special, big, local
ADP|adposition|on, of, at, with, by, into, under
ADV|adverb|really, already, still, early, now
CONJ|conjunction|and, or, but, if, while, although
DET|determiner, article|the, a, some, most, every, no, which
NOUN|noun|year, home, costs, time, Africa
NUM|numeral|twenty-four, fourth, 1991, 14:24
PRT|particle|at, on, out, over per, that, up, with
PRON|pronoun|he, their, her, its, my, I, us
VERB|verb|is, say, told, given, playing, would


Exp-5.ipynb
File Edit View Insert Runtime Tools Help Last saved at 8:17 PM
Comment
Share
+ Code + Text
Connect
NOUN|noun|year, home, costs, time, Africa
NUM|numeral|twenty-four, fourth, 1991, 14:24
PRT|particle|at, on, out, over per, that, up, with
PRON|pronoun|he, their, her, its, my, I, us
VERB|verb|is, say, told, given, playing, would
.|punctuation marks|. , ; !
X|other|ersatz, esprit, dunno, gr8, univeristy
[ ] text = "TON 618 (short for Tonantzintla 618) is a hyperluminous, broad-absorption-line, radio-loud quasar and Lyman-alpha blob located near the border of the constellations Canes Venz
Importing necessary dependencies
[ ] import nltk
from nltk.tokenize import word_tokenize
Word Tokenization
[ ] words = word_tokenize(text)
83°F
Smoke
Search
8:48 PM
31-Oct-23


Exp-5.ipynb
File Edit View Insert Runtime Tools Help Last saved at 8:17 PM
Comment
Share
+ Code + Text
Connect
Parts of Speech Tagging
[ ] tagged_words = nltk.pos_tag(words, tagset = 'universal')
tagged_words
[('TON', '.'),
('618', 'NUM'),
('(', '.'),
('short', 'ADJ'),
('for', 'ADP'),
('Tonantzintla', 'NOUN'),
('618', 'NUM'),
(')', '.'),
('is', 'VERB'),
('a', 'DET'),
('hyperluminous', 'ADJ'),
(',', '.'),
('broad-absorption-line', 'ADJ'),
(',', '.'),
('radio-loud', 'ADJ'),
('quasar', 'NOUN'),
('and', 'CONJ'),
('Lyman-alpha', 'NOUN'),
('blob', 'NOUN'),
('located', 'VERB'),
('near', 'ADP'),
('the', 'DET'),
('border', 'NOUN'),


Exp-5.ipynb
File Edit View Insert Runtime Tools Help Last saved at 8:17 PM
Comment
Share
+ Code + Text
Connect
('of', 'ADP'),
('the', 'DET'),
('constellations', 'NOUN'),
('Canes', 'NOUN'),
('Venatici', 'NOUN'),
('and', 'CONJ'),
('Coma', 'NOUN'),
('Berenices', 'NOUN'),
(',', '.'),
('with', 'ADP'),
('the', 'DET'),
('projected', 'VERB'),
('comoving', 'NOUN'),
('distance', 'NOUN'),
('of', 'ADP'),
('approximately', 'ADV'),
('18.2', 'NUM'),
('billion', 'NUM'),
('light-years', 'NOUN'),
('from', 'ADP'),
('Earth', 'NOUN'),
('.', '.')]
[ ] for t in tagged_words:
print(t)
('TON', '.')
('618', 'NUM')
('(', '.')
('short', 'ADJ')
('for', 'ADP')
('Tonantzintla', 'NOUN')


Exp-5.ipynb
File Edit View Insert Runtime Tools Help Last saved at 8:17 PM
Comment
Share
+ Code + Text
Connect
[ ] (',', '.')
('radio-loud', 'ADJ')
('quasar', 'NOUN')
('and', 'CONJ')
('Lyman-alpha', 'NOUN')
('blob', 'NOUN')
('located', 'VERB')
('near', 'ADP')
('the', 'DET')
('border', 'NOUN')
('of', 'ADP')
('the', 'DET')
('constellations', 'NOUN')
('Canes', 'NOUN')
('Venatici', 'NOUN')
('and', 'CONJ')
('Coma', 'NOUN')
('Berenices', 'NOUN')
(',', '.')
('with', 'ADP')
('the', 'DET')
('projected', 'VERB')
('comoving', 'NOUN')
('distance', 'NOUN')
('of', 'ADP')
('approximately', 'ADV')
('18.2', 'NUM')
('billion', 'NUM')
('light-years', 'NOUN')
('from', 'ADP')
('Earth', 'NOUN')
('.', '.')

Conclusion: