**EXPERIMENT 7 – Chunking**

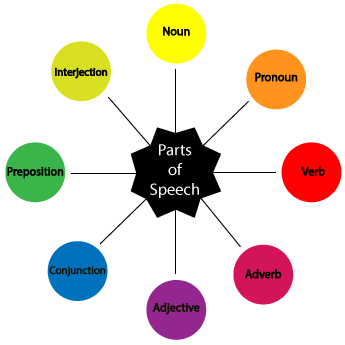
**Problem Statement:**  To understand and implement **Chunking**

**Theory:**

Chunking is a process of extracting phrases from unstructured text. Instead of just simple tokens which may not represent the actual meaning of the text, its advisable to use phrases such as “**South Africa**” as a single word instead of **‘South’** and **‘Africa’** separate words.

Chunking works on top of POS tagging, it uses pos-tags as input and provides chunks as output. Similar to POS tags, there are a standard set of Chunk tags like Noun Phrase(NP), Verb Phrase (VP), etc. Chunking is very important when you want to extract information from text such as Locations, Person Names etc. In NLP called Named Entity Extraction.

The **part of speech** explains how a word is used in a sentence. There are eight main parts of speech - **nouns**, **pronouns**, **adjectives**, **verbs**, **adverbs**, **prepositions**, **conjunctions** and **interjections**.



Noun (N)- Daniel, London, table, dog, teacher, pen, city, happiness, hope

Verb (V)- go, speak, run, eat, play, live, walk, have, like, are, is

Adjective(ADJ)- big, happy, green, young, fun, crazy, three

Adverb(ADV)- slowly, quietly, very, always, never, too, well, tomorrow

Preposition (P)- at, on, in, from, with, near, between, about, under

Conjunction (CON)- and, or, but, because, so, yet, unless, since, if

Pronoun(PRO)- I, you, we, they, he, she, it, me, us, them, him, her, this

Interjection (INT)- Ouch! Wow! Great! Help! Oh! Hey! Hi!

**Code:**

import nltk

sentence = "the little yellow dog barked at the cat"

#Define your grammar using regular expressions

grammar = ('''

NP: {<DT>?<JJ>\*<NN>} # NP

''')

chunkParser = nltk.RegexpParser(grammar)

tagged = nltk.pos\_tag(nltk.word\_tokenize(sentence))

print(tagged)

**Output:**

[('the', 'DT'),

('little', 'JJ'),

('yellow', 'JJ'),

('dog', 'NN'),

('barked', 'VBD'),

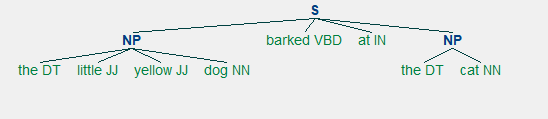
('at', 'IN'),

('the', 'DT'),

('cat', 'NN')]

tree = chunkParser.parse(tagged)

tree.draw()

****

**Conclusion:**

**Post Lab:**

**1]** What is Chunking?

**2]** Apply Chunking to string “The Book has many chapters”. Explain with neatly labelled trees.