Experiment No. 5

Aim: To study Chunking

Theory:

Chunk extraction or partial parsing is a process of meaningful extracting short phrases from the sentence (tagged with Part-of-Speech). Chunks are made up of words and the kinds of words are defined using the part-of-speech tags. One can even define a pattern or words that can't be apart of chuck and such words are known as chinks. A ChunkRule class specifies what words or patterns to include and exclude in a chunk.

Defining Chunk patterns:

Chuck patterns are normal regular expressions which are modified and designed to match the part-of-speech tag designed to match sequences of part-of-speech tags. Angle brackets are used to specify an individual tag for example – to match a noun tag. One can define multiple tags in the same way.

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 in NLP is Changing a perception by moving a "chunk", or a group of bits of information, in the direction of a Deductive or Inductive conclusion through the use of language.

Chunking up or down allows the speaker to use certain language patterns, to utilize the natural internal process through language, to reach for higher meanings or search for more specific bits/portions of missing information. When we "Chunk Up" the language gets more abstract and there are more chances for agreement, and when we "Chunk Down" we tend to be looking for the specific details that may have been missing in the chunk up.

As an example, if you ask the question "for what purpose cars?" you may get the answer "transport", which is a higher chunk and more toward abstract If you asked "what specifically about a car"? you will start to get smaller pieces of information about a car.

Lateral thinking will be the process of chunking up and then looking for other examples: For example, "for what intentions cars?", "transportation", "what are other examples of transportation?" "Buses!"

Code:

Noun Phrase chunking:

```
Import nltk

sentence = [("the", "DT"), ("little", "JJ"), ("yellow", "JJ"), ("dog", "NN"), ("barked",

"VBD"), ("at", "IN"), ("the", "DT"), ("cat", "NN")]

grammar = "NP: {<DT>?<JJ>*<NN>}"

cp = nltk.RegexpParser(grammar)

result = cp.parse(sentence)

print(result)

>>> result.draw()
```

Output:

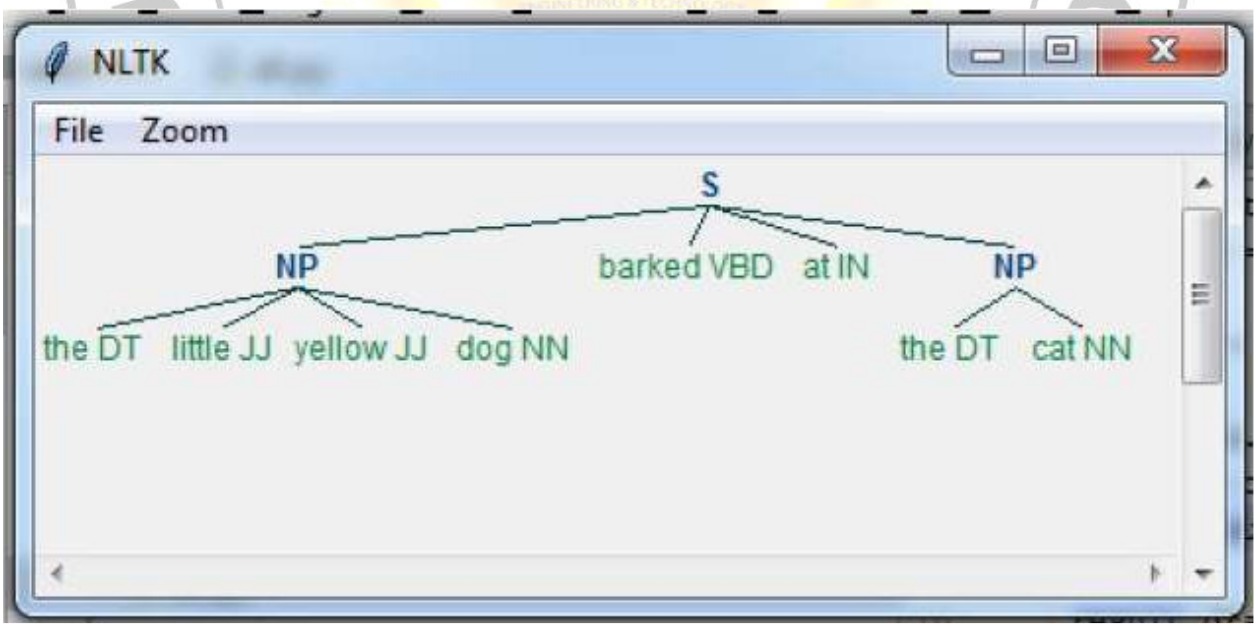
```
(S

(NP the/DT little/JJ yellow/JJ dog/NN)

barked/VBD

at/IN

(NP the/DT cat/NN))
```



Conclusion:

Thus, in the above experiment we have studies regarding chunking and tried to implement the code for same and successfully executed it.