Exp.5 Manas Parab-40 CSE(DS) DLOC-NLP

```
text = "TON 618 (short for Tonantzintla 618) is a hyperluminous, broad-absorption-line, radio-loud quasar and Lyman-alpha blob located ne
```

```
import nltk
from nltk.tokenize import word_tokenize
```

```
nltk.download('punkt')
words = word_tokenize(text)
```

```
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]   Unzipping tokenizers/punkt.zip.
```

```
nltk.download('averaged_perceptron_tagger')
nltk.download('universal_tagset')
tagged_words = nltk.pos_tag(words, tagset = 'universal')
```

```
[nltk_data] Downloading package averaged_perceptron_tagger to
[nltk_data]     /root/nltk_data...
[nltk_data]   Package averaged_perceptron_tagger is already up-to-
[nltk_data]       date!
[nltk_data] Downloading package universal_tagset to /root/nltk_data...
[nltk_data]   Unzipping taggers/universal_tagset.zip.
```

```
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'),
 ('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')
('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')
('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')
('.', '.')
```