Lab 1

November 14, 2022

1. Assign the above text to a variable sentence variable the value and print

[1]:
```
sentence = """Natural language processing makes it possible for computers
to understand the human language. In natural language
processing, human language is separated into fragments so that
the grammatical structure of sentences and the meaning of words
can be analysed and understood in context. This helps computers
read and understand spoken or written text in the same way as
humans. I am studying Natural Language Processing at Amrita
University."""
```

2. Check whether the word 'language' belongs to that text.

[3]:
```
"language" in sentence
```

[3]:
```
True
```

3. Find out the index value of the word 'human'.

[4]:
```
sentence.index('human')
```

[4]:
```
78
```

4. Find out the position of the word 'possible'

[6]:
```
sentence.find("possible")
```

[6]:
```
37
```

5. Print the third word of the given text.

[10]:
```
sentence.split()[2]
```

[10]:
```
'processing'
```

6. Find how many

a) lines/sentences are there
b) words are there
c) characters are there (space character excluded)

[16]:
```
print(sentence.count('\n'))
print(len(sentence.split()))
```

```
7
67
```

[20]:
```
words = sentence.split()
count = 0
for i in words:
    count += len(i)
print(count)
```

```
372
```

7. Create a vocabulary (list of unique words) from the text.

[25]:
```
unique = list(set(words))
len(unique)
```

[25]:
```
54
```

8. List the words in the vocabulary along with their frequency (count).

[27]:
```
from collections import Counter
```

[30]:
```
words = sentence.split()
Counter(words)
```

[30]:
```
Counter({'Natural': 2,
         'language': 3,
         'processing': 1,
         'makes': 1,
         'it': 1,
         'possible': 1,
         'for': 1,
         'computers': 2,
         'to': 1,
         'understand': 2,
         'the': 4,
         'human': 2,
         'language.': 1,
         'In': 1,
         'natural': 1,
         'processing,': 1,
         'is': 1,
         'separated': 1,
         'into': 1,
         'fragments': 1,
```

```
'so': 1,
'that': 1,
'grammatical': 1,
'structure': 1,
'of': 2,
'sentences': 1,
'and': 3,
'meaning': 1,
'words': 1,
'can': 1,
'be': 1,
'analysed': 1,
'understood': 1,
'in': 2,
'context.': 1,
'This': 1,
'helps': 1,
'read': 1,
'spoken': 1,
'or': 1,
'written': 1,
'text': 1,
'same': 1,
'way': 1,
'as': 1,
'humans.': 1,
'I': 1,
'am': 1,
'studying': 1,
'Language': 1,
'Processing': 1,
'at': 1,
'Amrita': 1,
'University.': 1})
```

[ ]: