

**a. Compare two nouns and find out path similarity between them.**

Code:

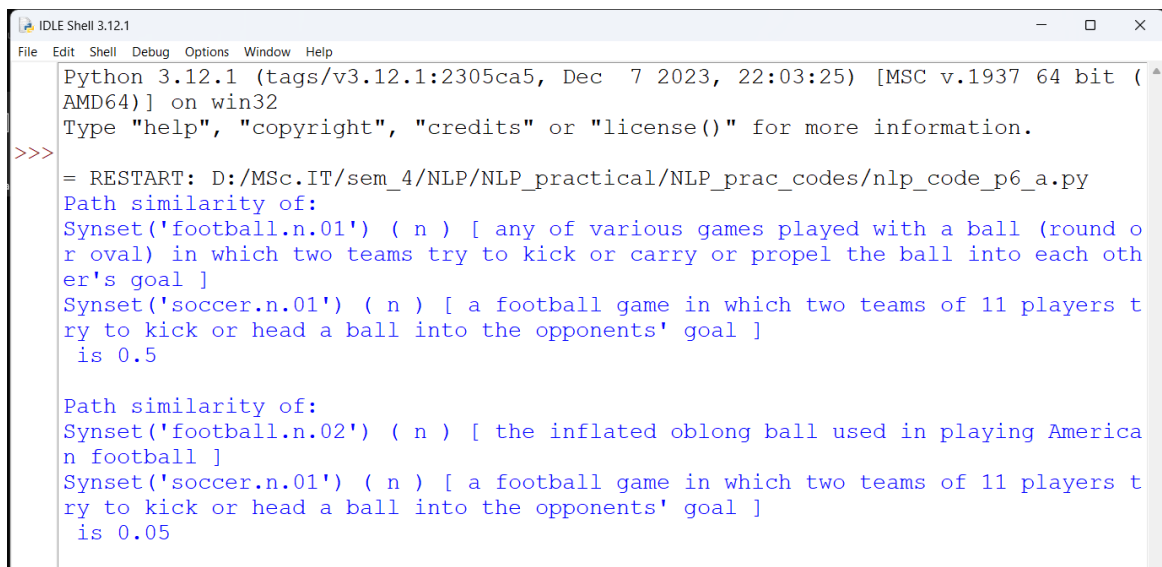
```
import nltk

from nltk.corpus import wordnet

syn1 = wordnet.synsets('football')
syn2 = wordnet.synsets('soccer')

for s1 in syn1:
    for s2 in syn2:
        print("Path similarity of: ")
        print(s1, '(', s1.pos(), ')', '[', s1.definition(), ']')
        print(s2, '(', s2.pos(), ')', '[', s2.definition(), ']')
        print(" is", s1.path_similarity(s2))
        print()
```

Output:



```
IDLE Shell 3.12.1
File Edit Shell Debug Options Window Help
Python 3.12.1 (tags/v3.12.1:2305ca5, Dec 7 2023, 22:03:25) [MSC v.1937 64 bit (AMD64)] on win32
Type "help", "copyright", "credits" or "license()" for more information.
>>>
= RESTART: D:/MSc.IT/sem_4/NLP/NLP_practical/NLP_prac_codes/nlp_code_p6_a.py
Path similarity of:
Synset('football.n.01') ( n ) [ any of various games played with a ball (round or oval) in which two teams try to kick or carry or propel the ball into each other's goal ]
Synset('soccer.n.01') ( n ) [ a football game in which two teams of 11 players try to kick or head a ball into the opponents' goal ]
is 0.5

Path similarity of:
Synset('football.n.02') ( n ) [ the inflated oblong ball used in playing American football ]
Synset('soccer.n.01') ( n ) [ a football game in which two teams of 11 players try to kick or head a ball into the opponents' goal ]
is 0.05
```

**b. Handling Stop words.**

- **Adding or removing stop words in NLTKs Default stop word list.**
- **Using gensim adding and removing stop words in Default gensim stop word list.**
- **Using spacy adding and removing stop words in Default spacy stop word list.**

Code and Output:

```
#Adding or removing stop words in NLTKs Default stop word list.
```

```
import nltk
from nltk.corpus import stopwords
nltk.download('stopwords')
from nltk.tokenize import word_tokenize
print(stopwords.words())
```

```
text = "yashesh likes to play football, however he is not to fond of tennis"
text_tokens = word_tokenize(text)
tokens_without_sw = [word for word in text_tokens if not word in stopwords.words()]
print(tokens_without_sw)
```

```
all_stopwords = stopwords.words('english')
all_stopwords.append('play')
text_tokens = word_tokenize(text)
tokens_without_sw = [word for word in text_tokens if not word in all_stopwords]
print(tokens_without_sw)
```

```
all_stopwords.remove('not')
text_tokens = word_tokenize(text)
tokens_without_sw = [word for word in text_tokens if not word in all_stopwords]
print(tokens_without_sw)
```

```
IDLE Shell 3.12.1
File Edit Shell Debug Options Window Help
Python 3.12.1 (tags/v3.12.1:2305ca5, Dec 7 2023, 22:03:25) [MSC v.1937 64 bit (AMD64)] on win32
Type "help", "copyright", "credits" or "license()" for more information.
>>>
= RESTART: D:/MSc.IT/sem_4/NLP/NLP_practical/NLP_prac_codes/nlp_code_p6_b_1.py
[nltk_data] Downloading package stopwords to
[nltk_data] C:\Users\acer\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!
Squeezed text (1227 lines).
['yashesh', 'likes', 'play', 'football', ',', 'fond', 'tennis']
['yashesh', 'likes', 'football', ',', 'however', 'fond', 'tennis']
['yashesh', 'likes', 'football', ',', 'however', 'not', 'fond', 'tennis']
>>>
```

#Using gensim adding and removing stop words in Default gensim stop word list.

```
import nltk
```

```
nltk.download('punkt_tab')
```

```
import gensim
```

```
from gensim.parsing.preprocessing import remove_stopwords
```

```
text = "yasesh likes to play football, however he is not too fond of tennis"
```

```
filtered_sentence = remove_stopwords(text)
```

```
print(filtered_sentence)
```

```
all_stopwords = gensim.parsing.preprocessing.STOPWORDS
```

```
print(all_stopwords)
```

```
from gensim.parsing.preprocessing import STOPWORDS
```

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

```
all_stopwords_gensim = STOPWORDS.union(set(['likes', 'play']))
```

```
text = "yasesh likes to play football, however he is not too fond of tennis"
```

```
text_tokens = word_tokenize(text)
```

```
tokens_without_sw = [word for word in text_tokens if not word in all_stopwords_gensim]
```

```
print(tokens_without_sw)
```

#the following script removes the word "not" from the set of stop words

```
from gensim.parsing.preprocessing import STOPWORDS
```

```
all_stopwords_gensim = STOPWORDS
```

```
sw_list = {"not"}
```

```
all_stopwords_gensim = STOPWORDS.difference(sw_list)
```

```
text = "yasesh likes to play football, however he is not too fond of tennis"
```

```
text_tokens = word_tokenize(text)
```

```
tokens_without_sw = [word for word in text_tokens if not word in all_stopwords_gensim]
```

```
print(tokens_without_sw)
```

```
> yasesh likes play football, fond tennis
frozenset({'whereby', 'thus', 'three', 'something', 'many', 'through', 'well', 'then', 'anyone', 'sixty',
['yasesh', 'football', ',', 'fond', 'tennis']
['yasesh', 'likes', 'play', 'football', ',', 'not', 'fond', 'tennis']
[nltk_data] Downloading package punkt_tab to /root/nltk_data...
[nltk_data] Package punkt_tab is already up-to-date!
```

```
#Using spacy adding and removing stop words in Default spacy stop word list.
import spacy
import nltk
from nltk.tokenize import word_tokenize
sp = spacy.load("en_core_web_sm")
all_stopwords = sp.Defaults.stop_words
all_stopwords.add("play")
text = "yasesh likes to play football, however he is not too fond of tennis"
text_tokens = word_tokenize(text)
tokens_without_sw = [word for word in text_tokens if not word in all_stopwords]
print(tokens_without_sw)
# remove 'not' from word collection
all_stopwords.remove('not')
tokens_without_sw = [word for word in text_tokens if not word in all_stopwords]
print(tokens_without_sw)
```

```
['yasesh', 'likes', 'football', ',', 'fond', 'tennis']
['yasesh', 'likes', 'football', ',', 'not', 'fond', 'tennis']
```