



DIGITAL ASSIGNMENT-4

STEMMING IN NLTK AND LEMMATIZATION IN SPACY

Practice Questions

December 19, 2024

SAI VARUN AITHA

saivarun.aitha18103@gmail.com

```
In [ ]: !pip install nltk
```

```
In [2]: import nltk
nltk.download('punkt')
```

```
[nltk_data] Downloading package punkt to
[nltk_data] C:\Users\hello\AppData\Roaming\nltk_data...
[nltk_data] Unzipping tokenizers\punkt.zip.
```

```
Out[2]: True
```

```
In [3]: from nltk.stem import PorterStemmer, LancasterStemmer, SnowballStemmer
```

```
In [4]: # Initialize stemmers
porter = PorterStemmer()
lancaster = LancasterStemmer()
snowball = SnowballStemmer(language='english')
```

```
In [5]: words = ["running", "runner", "ran", "easily", "easier", "flies", "studies", "st
```

```
In [6]: porter_stemmed = [porter.stem(word) for word in words]
lancaster_stemmed = [lancaster.stem(word) for word in words]
snowball_stemmed = [snowball.stem(word) for word in words]
```

```
In [7]: print("Original Words:")
print(words)

print("\nPorter Stemmer Results:")
print(porter_stemmed)

print("\nLancaster Stemmer Results:")
print(lancaster_stemmed)

print("\nSnowball Stemmer Results:")
print(snowball_stemmed)
```

Original Words:

```
['running', 'runner', 'ran', 'easily', 'easier', 'flies', 'studies', 'studying']
```

Porter Stemmer Results:

```
['run', 'runner', 'ran', 'easili', 'easier', 'fli', 'studi', 'studi']
```

Lancaster Stemmer Results:

```
['run', 'run', 'ran', 'easy', 'easy', 'fli', 'study', 'study']
```

Snowball Stemmer Results:

```
['run', 'runner', 'ran', 'easili', 'easier', 'fli', 'studi', 'studi']
```

```
In [ ]: !pip install spacy
```

```
In [ ]: !python -m spacy download en_core_web_sm
```

```
In [10]: import spacy
nlp = spacy.load("en_core_web_sm")
```

```
In [11]: text = "The striped bats are hanging on their feet for best"
```

```
In [12]: doc = nlp(text)
```

```
In [13]: print("Original Text:", text)
print("Lemmatized Words:")
for token in doc:
    print(f"{token.text} -> {token.lemma_}")
```

Original Text: The striped bats are hanging on their feet for best

Lemmatized Words:

The -> the

striped -> striped

bats -> bat

are -> be

hanging -> hang

on -> on

their -> their

feet -> foot

for -> for

best -> good

```
In [ ]:
```