

EXPERIMENT 4

AIM: Implement Stemming, Lemmatization in Python

SOURCE CODE:

```
from nltk.tokenize import word_tokenize
from nltk.corpus import stopwords
from nltk.stem import PorterStemmer
from nltk.stem import WordNetLemmatizer

f = open("tanay.txt", "r")

#wordtokenizer
for line in f:
    print("Tokenization without NLTK \n")
    print(line.split())
    print("\n")
    print("Length of Tokenization without NLTK:")
    print(len(line.split()))
    print("\n")

    print("Tokenization with NLTK \n")
    print(word_tokenize(line))
    print("\n")
    print("Length of Tokenization with NLTK:")
    print(len(word_tokenize(line)))
    print("\n")

#filteration
    print("Filteration \n")
```

```

bad_chars = [',',';',':','!',',','*','#','<','>','?', '@','.',']
words = word_tokenize(line)
print(list(filter(lambda i: i not in bad_chars,words)))
print("\n")
print("Length of Filtration:")
print(len(list(filter(lambda i: i not in bad_chars,words))))
print("\n")

```

#stop word removal

```

print("Stop word removal \n")
stop_words = set(stopwords.words("English"))
without_stop_words = [word for word in words if not word in stop_words]
with_stopwords = list(filter(lambda i: i not in bad_chars, without_stop_words))
print(with_stopwords)
print("\n")
print("Length of Stopword Removal:")
print(len(with_stopwords))
print("\n")

```

#Stemming and Lemmatization

```

ps = PorterStemmer()
l = WordNetLemmatizer()

```

#Tablular form to compare

```

print("Word \t \tStemming \tLemmatization")
for w in with_stopwords:
    print ("{:<15} {:<15} {:<15}".format( w, ps.stem(w), l.lemmatize(w)))
print("\n")

```

OUTPUT:

```
Python 3.7.9 Shell
File Edit Shell Debug Options Window Help
===== RESTART: C:\Users\Dell\Desktop\SEM8\NLP\Exp3.py =====
Tokenization without NLTK

['Manchester', 'United', 'Football', 'Club', 'is', 'a', 'professional', 'football', 'club', 'based', 'in', 'Old', 'Trafford', 'Greater', 'Manchester', 'England', 'that', 'competes', 'in', 'the', 'Premier', 'League', 'the', 'top', 'flight', 'of', 'English', 'football.', 'It', 'is', 'nicknamed', 'as', 'the', 'Red', 'Devils.', 'Tanay', 'runs', 'a', 'Manchester', 'United', 'fanbase.']]

Length of Tokenization without NLTK:
41

Tokenization with NLTK

['Manchester', 'United', 'Football', 'Club', 'is', 'a', 'professional', 'football', 'club', 'based', 'in', 'Old', 'Trafford', ' ', 'Greater', 'Manchester', ' ', 'England', ' ', 'that', 'competes', 'in', 'the', 'Premier', 'League', ' ', 'the', ' ', 'top', 'flight', 'of', 'English', 'football', '.', 'It', 'is', 'nicknamed', 'as', ' ', 'the', 'Red', 'Devils', '"', '.', 'Tanay', 'runs', 'a', 'Manchester', 'United', 'fanbase', '.']]

Length of Tokenization with NLTK:
50

Filtration

['Manchester', 'United', 'Football', 'Club', 'is', 'a', 'professional', 'football', 'club', 'based', 'in', 'Old', 'Trafford', 'Greater', 'Manchester', 'England', 'that', 'competes', 'in', 'the', 'Premier', 'League', 'the', 'top', 'flight', 'of', 'English', 'football', 'It', 'is', 'nicknamed', 'as', ' ', 'the', 'Red', 'Devils', '"', 'Tanay', 'runs', 'a', 'Manchester', 'United', 'fanbase']]

Length of Filtration:
43

Stop word removal

['Manchester', 'United', 'Football', 'Club', 'professional', 'football', 'club', 'based', 'Old', 'Trafford', 'Greater', 'Manchester', 'England', 'competes', 'Premier', 'League', 'top', 'flight', 'English', 'football', 'It', 'nicknamed', ' ', 'Red', 'Devils', '"', 'Tanay', 'runs', 'Manchester', 'United', 'fanbase']]

Length of Stopword Removal:
31
```

Word	Stemming	Lemmatization
Manchester	manchest	Manchester
United	unit	United
Football	footbal	Football
Club	club	Club
professional	profession	professional
football	footbal	football
club	club	club
based	base	based
Old	old	Old
Trafford	trafford	Trafford
Greater	greater	Greater
Manchester	manchest	Manchester
England	england	England
competes	compet	competes
Premier	premier	Premier
League	leagu	League
top	top	top
flight	flight	flight
English	english	English
football	footbal	football
It	it	It
nicknamed	nicknam	nicknamed
Red	red	Red
Devils	devil	Devils
Tanay	tanay	Tanay
runs	run	run
Machester	machest	Machester
United	unit	United
fanbase	fanbas	fanbase