EXPERIMENT 13: Implementation of Natural Language Processing

**Name :** D V Ravi Sridhar

**Reg No. :** RA1811003010732

**Section :** CSE – D2

# Date : 26/04/2021

**Problem Description :- Lexicon Normalization**

**Performing stemming and Lemmatization**

A type of textual noise is about the multiple representations exhibited by single word. For example – “play”, “player”, “played”, “plays” and “playing” are the different variations of the word – “play”, Though they mean different but contextually all are similar. The step converts all the disparities of a word into their normalized form (also known as lemma). Normalization is a pivotal step for feature engineering with text as it converts the high dimensional features (N different features) to the low dimensional space (1 feature), which is an ideal ask for any ML model.The most common lexicon normalization practices are :

**Stemming:** Stemming is a process of linguistic normalization, which reduces words to their word root word or chops off the derivational affixes.

**Lemmatization:** Lemmatization, on the other hand, is an organized & step by step procedure of obtaining the root form of the word, it makes use of vocabulary (dictionary importance of words) and morphological analysis (word structure and grammar relations). Lemmatization reduces words to their base word, which is linguistically correct lemmas. It transforms root word with the use of vocabulary and morphological analysis. Lemmatization is usually more sophisticated than stemming. Stemmer works on an individual word without knowledge of the context. For example, The word "better" has "good" as its lemma. This thing will miss by stemming because it requires a dictionary look-up.

# Code :-

#Lexicon Normalization

#performing stemming and Lemmatization

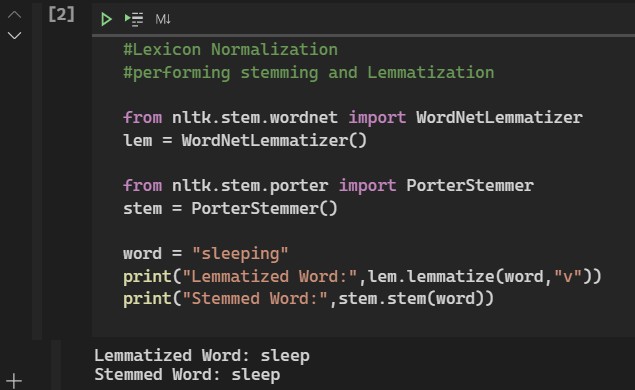
from nltk.stem.wordnet import WordNetLemmatizer lem = WordNetLemmatizer()

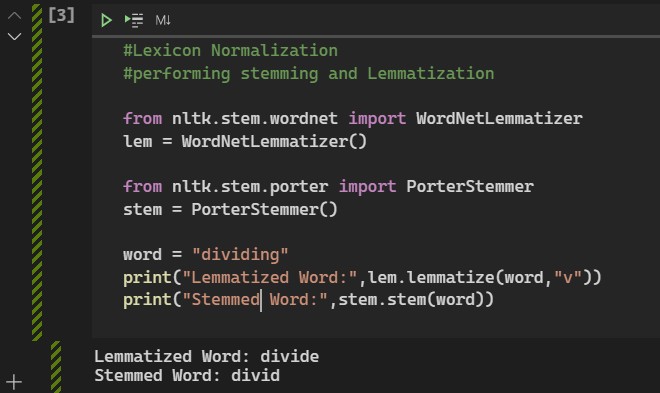
from nltk.stem.porter import PorterStemmer stem = PorterStemmer()

word = "dividing"

print("Lemmatized Word:",lem.lemmatize(word,"v")) print("Stemmed Word:",stem.stem(word))

**Screenshot from Output :-**





**Result**

The experiment was successfully implemented and executed.