Stemming and Lemmatization:

Stemming: playing-play-played-plays → play

Lemmatization: is-am-are → be + Stemming

Lemmatization considers (takes into account) the meaning, for instance: if meeting is verb → meet, if noun → stays meeting.

Stemming&Lemmatization-Pipline:

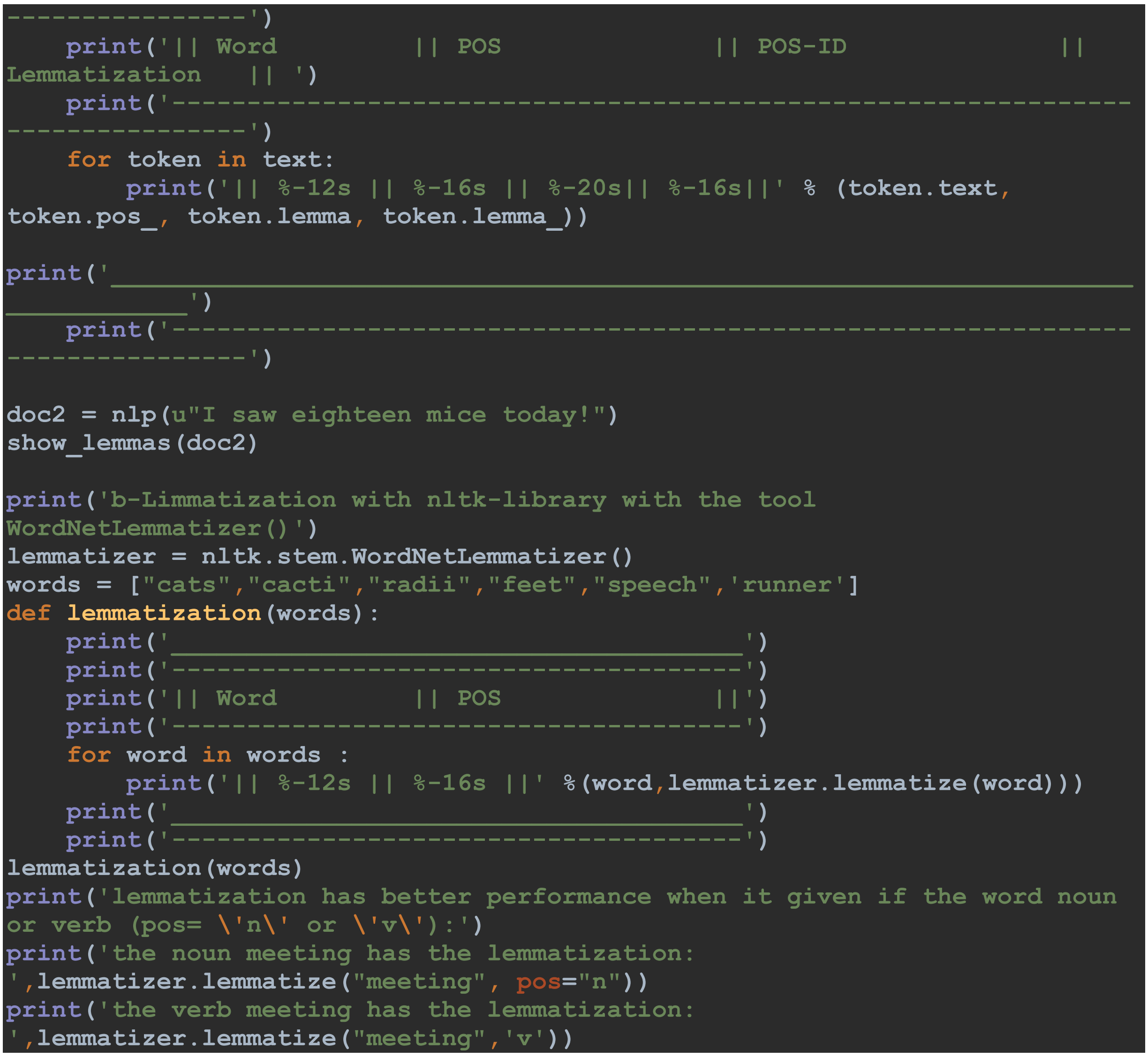
Normalization(U.S.A→USA)→case folding (Gold-goLd-GOLD→gold but USA stays USA)→ lemmatization(is am are→ be(stem)).(cars → car(stem and s is affixes).

The Lemmatization is done with the Porter-Algorithm.

The Code:

```
print('Stemming and Lemmatization:')
print('--------------------------------------------------------------')
print('1- Stemming: we start with Stemming:')
print('the Stemming is only available in nltk, because spacy-library doesn\'t support it.')
print('--------------------------------------------------------------')
print('''Comparison between the variants of Stemming's tool in nltk-library:
PorterStemmer(p_stemmer), SnowballStemmer(s_stemmer) and
LancasterStemmer(l_stemmer: ''')
words = ['run','runner','running','ran','runs','easily','fairly',"is","was","be","been","are","were"]
import nltk
p_stemmer=nltk.stem.porter.PorterStemmer()
s_stemmer= nltk.stem.snowball.SnowballStemmer(language='english')
l_stemmer = nltk.stem.LancasterStemmer()
print('_________________________________________________________________________')
print('-------------------------------------------------------------------------')
print('|| Word         || PorterStemmer    || SnowballStemmer|| LancasterStemmer||  ')
print('---------------------------------------------------------------------------------')
for word in words:
    print('|| %-12s || %-16s || %-15s|| %-16s||'%(word,p_stemmer.stem(word),s_stemmer.stem(word),l_stemmer.stem(word)))
print('_________________________________________________________________________')
print('-------------------------------------------------------------------------')
print(' ')
print('---------------------------------------------------------------')
print('2-Lemmatization:')
print('a- we start with spacy-library: ')
import spacy
nlp = spacy.load('en_core_web_sm')
def show_lemmas(text):

print('________________________________________________________________________________')
    print('-----------------------------------------------------------------
```



```
-----------------')
    print('|| Word          || POS                || POS-ID              ||
Lemmatization    || ')
    print('----------------------------------------------------------------------------
-----------------')
    for token in text:
        print('|| %-12s || %-16s || %-20s|| %-16s||' % (token.text,
token.pos_, token.lemma, token.lemma_))

print('_____________________________________________________________________
____________')
    print('----------------------------------------------------------------------------
-----------------')

doc2 = nlp(u"I saw eighteen mice today!")
show_lemmas(doc2)

print('b-Limmatization with nltk-library with the tool
WordNetLemmatizer()')
lemmatizer = nltk.stem.WordNetLemmatizer()
words = ["cats","cacti","radii","feet","speech",'runner']
def lemmatization(words):
    print('_______________________________________')
    print('---------------------------------------')
    print('|| Word          || POS                ||')
    print('---------------------------------------')
    for word in words :
        print('|| %-12s || %-16s ||' %(word,lemmatizer.lemmatize(word)))
    print('_______________________________________')
    print('---------------------------------------')
lemmatization(words)
print('lemmatization has better performance when it given if the word noun
or verb (pos= \'n\' or \'v\'):')
print('the noun meeting has the lemmatization:
',lemmatizer.lemmatize("meeting", pos="n"))
print('the verb meeting has the lemmatization:
',lemmatizer.lemmatize("meeting",'v'))
```

Output:

Stemming and Lemmatization:

1- Stemming: we start with Stemming:

the Stemming is only available in nltk, because spacy-library doesn't support it.

Comparison between the variants of Stemming's tool in nltk-library:

PorterStemmer(p_stemmer), SnowballStemmer(s_stemmer) and LancasterStemmer(l_stemmer:

____

--

|| Word || PorterStemmer || SnowballStemmer|| LancasterStemmer||

|| run || run || run || run ||

```
|| runner     || runner        || runner       || run          ||
|| running    || run           || run          || run          ||
|| ran        || ran           || ran          || ran          ||
|| runs       || run           || run          || run          ||
|| easily     || easili        || easili       || easy         ||
|| fairly     || fairli        || fair         || fair         ||
|| is         || is            || is           || is           ||
|| was        || wa            || was          || was          ||
|| be         || be            || be           || be           ||
|| been       || been          || been         || been         ||
|| are        || are           || are          || ar           ||
|| were       || were          || were         || wer          ||
________________________________________________________________________________
-----------------------------------------------------------------------


-------------------------------------------------------------
```

2-Lemmatization:

a- we start with spacy-library:

```
________________________________________________________________________________________
------------------------------------------------------------------------------
|| Word       || POS          || POS-ID          || Lemmatization   ||
------------------------------------------------------------------------------
|| I          || PRON          || 4690420944186131903 || I            ||
|| saw        || VERB          || 11925638236994514241|| see          ||
|| eighteen   || NUM           || 9609336664675087640 || eighteen      ||
|| mice       || NOUN          || 1384165645700560590 || mouse        ||
|| today      || NOUN          || 11042482332948150395|| today        ||
|| !          || PUNCT         || 17494803046312582752|| !            ||
________________________________________________________________________________________
------------------------------------------------------------------------------
```

b-Limmatization with nltk-library with the tool WordNetLemmatizer()

```
__________________________________________
```

```
-------------------------------------
|| Word       || POS           ||
-------------------------------------
|| cats       || cat           ||
|| cacti      || cactus        ||
|| radii      || radius        ||
|| feet       || foot          ||
|| speech     || speech        ||
|| runner     || runner        ||
_____________________________________________
-------------------------------------
lemmatization is more accurate when it given if the word noun or verb (pos= 'n' or 'v')
the noun meeting has the lemmatization:  meeting
the verb meeting has the lemmatization:  meet
```