NAME: LAKSHITHA S K

REG NO: 20BCE1649

CSE4022 Natural Language Processing
Digital Assignment -1

1. Utilize Python NLTK (Natural Language Tool Kit) Platform and do the following.
Install relevant Packages and Libraries

```python
In [10]: import nltk.corpus
         print(str(nltk.corpus.brown).replace('\\\\','/'))
         nltk.corpus.brown.fileids()
```

```
<CategorizedTaggedCorpusReader in 'C:/Users/laksh/AppData/Roaming/nltk_data/corpora/brown'>
```

```
Out[10]: ['ca01',
          'ca02',
          'ca03',
          'ca04',
          'ca05',
          'ca06',
          'ca07',
          'ca08',
          'ca09',
          'ca10',
          'ca11',
          'ca12',
          'ca13',
          'ca14',
          'ca15',
          'ca16',
          'ca17',
          'ca18'
```

```python
from nltk.corpus import brown

print(brown.categories())
print(type(brown.words()))
print(brown.words())
print(brown.sents())
print(brown.paras())
print(len(brown.words()))
print(len(brown.sents()))
print(len(brown.paras()))
print(len(brown.words(categories='government')))
vocab=nltk.FreqDist(brown.words())
print(vocab.most_common)
```

```
['adventure', 'belles_lettres', 'editorial', 'fiction', 'government', 'hobbies', 'humor', 'learned', 'lore', 'mystery', 'news',
'religion', 'reviews', 'romance', 'science_fiction']
<class 'nltk.corpus.reader.util.ConcatenatedCorpusView'>
['The', 'Fulton', 'County', 'Grand', 'Jury', 'said', ...]
[['The', 'Fulton', 'County', 'Grand', 'Jury', 'said', 'Friday', 'an', 'investigation', 'of', "Atlanta's", 'recent', 'primary',
'election', 'produced', '``', 'no', 'evidence', "''", 'that', 'any', 'irregularities', 'took', 'place', '.'], ['The', 'jury',
'further', 'said', 'in', 'term-end', 'presentments', 'that', 'the', 'City', 'Executive', 'Committee', ',', 'which', 'had', 'ove
r-all', 'charge', 'of', 'the', 'election', ',', '``', 'deserves', 'the', 'praise', 'and', 'thanks', 'of', 'the', 'City', 'of',
'Atlanta', "''", 'for', 'the', 'manner', 'in', 'which', 'the', 'election', 'was', 'conducted', '.'], ...]
[[['The', 'Fulton', 'County', 'Grand', 'Jury', 'said', 'Friday', 'an', 'investigation', 'of', "Atlanta's", 'recent', 'primary',
'election', 'produced', '``', 'no', 'evidence', "''", 'that', 'any', 'irregularities', 'took', 'place', '.']], [['The', 'jury',
'further', 'said', 'in', 'term-end', 'presentments', 'that', 'the', 'City', 'Executive', 'Committee', ',', 'which', 'had', 'ove
r-all', 'charge', 'of', 'the', 'election', ',', '``', 'deserves', 'the', 'praise', 'and', 'thanks', 'of', 'the', 'City', 'of',
'Atlanta', "''", 'for', 'the', 'manner', 'in', 'which', 'the', 'election', 'was', 'conducted', '.']], ...]
1161192
57340
53
70117
<bound method Counter.most_common of FreqDist({'the': 62713, ',': 58334, '.': 49346, 'of': 36080, 'and': 27915, 'to': 25732,
'a': 21881, 'in': 19536, 'that': 10237, 'is': 10011, ...})>
```

2. Explore the corpora available in NLTK (any two)

```python
from nltk.corpus import brown
from nltk.corpus import ieer
from nltk.corpus import indian

indian.fileids()
ieer.fileids()
print(brown.raw('ca07'))
print(brown.tagged_words())
docs = ieer.parsed_docs('APW_19980429')
print(docs[0])
print(docs[0].docno)
print(docs[0].doctype)
print(docs[0].date_time)
print(docs[0].headline)
print(docs[0].text)
nltk.corpus.indian.words('hindi.pos')
```

```
        Resentment/nn welled/vbd up/rp yesterday/nr among/in Democratic/jj-tl district/nn leaders/nns and/cc some/dti county/
nn leaders/nns at/in reports/nns that/cs Mayor/nn-tl Wagner/np had/hvd decided/vbn to/to seek/vb a/at third/od term/nn with/i
n Paul/np R./np Screvane/np and/cc Abraham/np D./np Beame/np as/cs running/vbg mates/nns ./.

        At/in the/at same/ap time/nn reaction/nn among/in anti-organization/jj Democratic/jj-tl leaders/nns and/cc in/in the/
at Liberal/jj-tl party/nn to/in the/at Mayor's/nn$-tl reported/vbn plan/nn was/bedz generally/rb favorable/jj ./.

        Some/dti anti-organization/jj Democrats/nps saw/vbd in/in the/at program/nn an/at opportunity/nn to/to end/vb the/at
bitter/jj internal/jj fight/nn within/in the/at Democratic/jj-tl party/nn that/wps has/hvz been/ben going/vbg on/rp for/in th
e/at last/ap three/cd years/nns ./.

        The/at resentment/nn among/in Democratic/jj-tl organization/nn leaders/nns to/in the/at reported/vbn Wagner/np plan/n
n was/bedz directed/vbn particularly/rb at/in the/at Mayor's/nn$-tl efforts/nns to/to name/vb his/pp$ own/jj running/vbg mate
s/nns without/in consulting/vbg the/at leaders/nns ./.
Some/dti viewed/vbd this/dt attempt/nn as/cs evidence/nn that/cs Mr./np Wagner/np regarded/vbd himself/ppl as/in bigger/jjr t
han/in the/at party/nn ./.
```

```
Mayor/nn-tl-hl visits/vbz-hl Buckley/np-hl
As/cs usual/jj ,/, he/pps made/vbd no/at attempt/nn to/to get/vb in/in touch/nn with/in Carmine/np G./np De/np Sapio/np ,/, t
he/at Manhattan/np leader/nn ./.
He/pps is/bez publicly/rb on/in record/nn as/cs believing/vbg Mr./np De/np Sapio/np should/md be/be replaced/vbn for/in the/a
t good/nn of/in the/at party/nn ./.

        Last/ap night/nn the/at Mayor/nn-tl visited/vbd Mr./np Buckley/np at/in the/at Bronx/np leader's/nn$ home/nn for/in
a/at discussion/nn of/in the/at situation/nn ./.
Apparently/rb he/pps believes/vbz Mr./np Buckley/np holds/vbz the/at key/nn to/in the/at Democratic/jj-tl organization's/nn$
acceptance/nn of/in his/pp$ choices/nns for/in running/vbg mates/nns without/in a/at struggle/nn ./.

        In/in talks/nns with/in Mr./np Buckley/np last/ap week/nn in/in Washington/np ,/, the/at Mayor/nn-tl apparently/rb re
ceived/vbd the/at Bronx/np leader's/nn$ assent/nn to/in dropping/vbg Controller/nn-tl Lawrence/np E./np Gerosa/np ,/, who/wps
lives/vbz in/in the/at Bronx/np ,/, from/in this/dt year's/nn$ ticket/nn ./.
But/cc Mr./np Buckley/np seems/vbz to/to have/hv assumed/vbn he/pps would/md be/be given/vbn the/at right/nn to/to pick/vb M
r./np Gerosa's/np$ successor/nn ./.
```

```
  [('The', 'AT'), ('Fulton', 'NP-TL'), ...]
  <IEERDocument APW19980429.1258: 'Tickets for 1999 championship to go on sale Friday'>
  APW19980429.1258
  NEWS STORY
  04/29/1998 15:10:00
  (DOCUMENT
    Tickets
    for
    (DATE 1999)
    championship
    to
    go

Out[16]:  ['पूर्ण', 'प्रतिबंध', 'हटाओ', ':', 'इराक', 'संयुक्त', ...]
```

3. Create a text corpus with a minimum of 200 words (unique content). Implement the following text processing

```
In [1]: import os, os.path
        path = os.path.expanduser('~/nltk_data')
        import nltk.data
        path in nltk.data.path
```

```
Out[1]: True
```

```
In [9]:
        newfile=nltk.data.load('corpora/your_corpus/file.txt')
        print(newfile)
```

When Americans think of cheesecake now, it's most often associated with a product that has a cream cheese base. Cream cheese was invented in 1872 by American dairyman William Lawrence of Chester, New York, who accidentally stumbled on a method of producing cream cheese while trying to reproduce a French cheese called Neufchâtel.

In 1880, Lawrence began distributing his cream cheese in foil wrappers under the auspices of the Empire Cheese Company of South Edmeston, New York, where he manufactured the product. However, you might know it better by the more famous name Lawrence came up with for his "not Neufchâtel"—Philadelphia Brand Cream Cheese.

In 1903, the Phoenix Cheese Company bought Lawrence's business—and with it, the Philadelphia trademark. In 1928, the brand was bought by the Kraft Cheese Company. James L. Kraft invented pasteurized cheese in 1912, which led to the development of pasteurized Philadelphia Brand cream cheese, currently the most popular cheese used for cheesecake making. Kraft Foods still owns and produces Philadelphia Cream Cheese today.

Fast Facts: Cheesecake Favorites
Traditional Greek Cheesecake—Most "traditional" Greek cheesecake is made using ricotta cheese, however, for the real deal, try to find authentic unsalted anthotyros or myzirtha cheeses which are made with either goat's or sheep's milk. Greek cheesecake is usually sweetened with honey. Some recipes incorporate flour directly into the cheese/honey mixture prior to baking, while others employ a crust.
Cream Cheese Cheesecake—The cheesecake most Americans grew up with is one or another version of a cream cheese cheesecake. At the bottom of such cheesecakes, you'll usually find a crust made of crushed Graham crackers or other cookies (Oreos are a top choice for chocolate cheesecakes) that have been blended with butter and tamped into the bottom of a pan or mold. Cheesecakes that rely on a custard base must be baked. (The original New York Cheesecake that hails from Junior's on Flatbush Avenue in Brooklyn is a baked cheesecake.) However, there are scads of recipes that use a blend of other rich ingredients—such as sour cream, Greek yogurt, or heavy cream—that firm up in the refrigerator to create a "no-bake cheesecake."

• Word segmentation

```
from nltk.tokenize import word_tokenize

print(word_tokenize(newfile))
```

```
['When', 'Americans', 'think', 'of', 'cheesecake', 'now', ',', 'it', "'s", 'most', 'often', 'associated', 'with', 'a', 'product', 'that', 'has', 'a', 'cream', 'cheese', 'base', '.', 'Cream', 'cheese', 'was', 'invented', 'in', '1872', 'by', 'American', 'dairyman', 'William', 'Lawrence', 'of', 'Chester', ',', 'New', 'York', ',', 'who', 'accidentally', 'stumbled', 'on', 'a', 'method', 'of', 'producing', 'cream', 'cheese', 'while', 'trying', 'to', 'reproduce', 'a', 'French', 'cheese', 'called', 'Neufchâtel', '.', 'In', '1880', ',', 'Lawrence', 'began', 'distributing', 'his', 'cream', 'cheese', 'in', 'foil', 'wrappers', 'under', 'the', 'auspices', 'of', 'the', 'Empire', 'Cheese', 'Company', 'of', 'South', 'Edmeston', ',', 'New', 'York', ',', 'where', 'he', 'manufactured', 'the', 'product', '.', 'However', ',', 'you', 'might', 'know', 'it', 'better', 'by', 'the', 'more', 'famous', 'name', 'Lawrence', 'came', 'up', 'with', 'for', 'his', '``', 'not', 'Neufchâtel', "''", '—Philadelphia', 'Brand', 'Cream', 'Cheese', '.', 'In', '1903', ',', 'the', 'Phoenix', 'Cheese', 'Company', 'bought', 'Lawrence', "'s", 'business—and', 'with', 'it', ',', 'the', 'Philadelphia', 'trademark', '.', 'In', '1928', ',', 'the', 'brand', 'was', 'bought', 'by', 'the', 'Kraft', 'Cheese', 'Company', '.', 'James', 'L.', 'Kraft', 'invented', 'pasteurized', 'cheese', 'in', '1912', ',', 'which', 'led', 'to', 'the', 'development', 'of', 'pasteurized', 'Philadelphia', 'Brand', 'cream', 'cheese', ',', 'currently', 'the', 'most', 'popular', 'cheese', 'used', 'for', 'cheesecake', 'making', '.', 'Kraft', 'Foods', 'still', 'owns', 'and', 'produces', 'Philadelphia', 'Cream', 'Cheese', 'today', '.', 'Fast', 'Facts', ':', 'Cheesecake', 'Favorites', 'Traditional', 'Greek', 'Cheesecake—Most', '"', 'traditional', '"', 'Greek', 'cheesecake', 'is', 'made', 'using', 'ricotta', 'cheese', ',', 'however', ',', 'for', 'the', 'real', 'deal', ',', 'try', 'to', 'find', 'authentic', 'unsalted', 'anthotyros', 'or', 'myzirtha', 'cheeses', 'which', 'are', 'made', 'with', 'either', 'goat', "'", 's', 'or', 'sheep', "'", 's', 'milk', '.', 'Greek', 'cheesecake', 'is', 'usually', 'sweetened', 'with', 'honey', '.', 'Some', 'recipes', 'incorporate', 'flour', 'directly', 'into', 'the', 'cheese/honey', 'mixture', 'prior', 'to', 'baking', ',', 'while', 'others', 'employ', 'a', 'crust', '.', 'Cream', 'Cheese', 'Cheesecake—The', 'cheesecake', 'most', 'Americans', 'grew', 'up', 'with', 'is', 'one', 'or', 'another', 'version', 'of', 'a', 'cream', 'cheese', 'cheesecake', '.', 'At', 'the', 'bottom', 'of', 'such', 'cheesecakes', ',', 'you', "'", 'll', 'usually', 'find', 'a', 'crust', 'made', 'of', 'crushed', 'Graham', 'crackers', 'or', 'other', 'cookies', '(', 'Oreos', 'are', 'a', 'top', 'choice', 'for', 'chocolate', 'cheesecakes', ')', 'that', 'have', 'been', 'blended', 'with', 'butter', 'and', 'tamped', 'into', 'the', 'bottom', 'of', 'a', 'pan', 'or', 'mold', '.', 'Cheesecakes', 'that', 'rely', 'on', 'a', 'custard', 'base', 'must', 'be', 'baked', '.', '(', 'The', 'original', 'New', 'York', 'Cheesecake', 'that', 'hails', 'from', 'Junior', "'s", 'on', 'Flatbush', 'Avenue', 'in', 'Brooklyn', 'is', 'a', 'baked', 'cheesecake', '.', ')', 'However', ',', 'there', 'are', 'scads', 'of', 'recipes', 'that', 'use', 'a', 'blend', 'of', 'other', 'rich', 'ingredients—such', 'as', 'sour', 'cream', ',', 'Greek', 'yogurt', ',', 'or', 'heavy', 'cream—that', 'firm', 'up', 'in', 'the', 'refrigerator', 'to', 'create', 'a', '"', 'no-bake', 'cheesecake', '.', '"']
```

• Sentence segmentation

```python
from nltk.tokenize import sent_tokenize

print(sent_tokenize(newfile))
```

["When Americans think of cheesecake now, it's most often associated with a product that has a cream cheese base.", 'Cream chee
se was invented in 1872 by American dairyman William Lawrence of Chester, New York, who accidentally stumbled on a method of pr
oducing cream cheese while trying to reproduce a French cheese called Neufchâtel.', 'In 1880, Lawrence began distributing his c
ream cheese in foil wrappers under the auspices of the Empire Cheese Company of South Edmeston, New York, where he manufactured
the product.', 'However, you might know it better by the more famous name Lawrence came up with for his "not Neufchâtel"—Philad
elphia Brand Cream Cheese.', 'In 1903, the Phoenix Cheese Company bought Lawrence's business—and with it, the Philadelphia trad
emark.", 'In 1928, the brand was bought by the Kraft Cheese Company.', 'James L. Kraft invented pasteurized cheese in 1912, whi
ch led to the development of pasteurized Philadelphia Brand cream cheese, currently the most popular cheese used for cheesecake
making.', 'Kraft Foods still owns and produces Philadelphia Cream Cheese today.', 'Fast Facts: Cheesecake Favorites\nTraditiona
l Greek Cheesecake—Most "traditional" Greek cheesecake is made using ricotta cheese, however, for the real deal, try to find au
thentic unsalted anthotyros or myzirtha cheeses which are made with either goat's or sheep's milk.', 'Greek cheesecake is usual
ly sweetened with honey.', 'Some recipes incorporate flour directly into the cheese/honey mixture prior to baking, while others
employ a crust.', 'Cream Cheese Cheesecake—The cheesecake most Americans grew up with is one or another version of a cream chee
se cheesecake.', 'At the bottom of such cheesecakes, you'll usually find a crust made of crushed Graham crackers or other cooki
es (Oreos are a top choice for chocolate cheesecakes) that have been blended with butter and tamped into the bottom of a pan or
mold.', 'Cheesecakes that rely on a custard base must be baked.', "(The original New York Cheesecake that hails from Junior's o
n Flatbush Avenue in Brooklyn is a baked cheesecake.)", 'However, there are scads of recipes that use a blend of other rich ing
redients—such as sour cream, Greek yogurt, or heavy cream—that firm up in the refrigerator to create a "no-bake cheesecake."']

• Convert to Lowercase

```python
lower = newfile.lower()
print(lower)
```

when americans think of cheesecake now, it's most often associated with a product that has a cream cheese base. cream cheese wa
s invented in 1872 by american dairyman william lawrence of chester, new york, who accidentally stumbled on a method of produci
ng cream cheese while trying to reproduce a french cheese called neufchâtel.

in 1880, lawrence began distributing his cream cheese in foil wrappers under the auspices of the empire cheese company of south
edmeston, new york, where he manufactured the product. however, you might know it better by the more famous name lawrence came
up with for his "not neufchâtel"—philadelphia brand cream cheese.

in 1903, the phoenix cheese company bought lawrence's business—and with it, the philadelphia trademark. in 1928, the brand was
bought by the kraft cheese company. james l. kraft invented pasteurized cheese in 1912, which led to the development of pasteur
ized philadelphia brand cream cheese, currently the most popular cheese used for cheesecake making. kraft foods still owns and
produces philadelphia cream cheese today.

fast facts: cheesecake favorites
traditional greek cheesecake—most "traditional" greek cheesecake is made using ricotta cheese, however, for the real deal, try
to find authentic unsalted anthotyros or myzirtha cheeses which are made with either goat's or sheep's milk. greek cheesecake i
s usually sweetened with honey. some recipes incorporate flour directly into the cheese/honey mixture prior to baking, while ot
hers employ a crust.
cream cheese cheesecake—the cheesecake most americans grew up with is one or another version of a cream cheese cheesecake. at t
he bottom of such cheesecakes, you'll usually find a crust made of crushed graham crackers or other cookies (oreos are a top ch
oice for chocolate cheesecakes) that have been blended with butter and tamped into the bottom of a pan or mold. cheesecakes tha
t rely on a custard base must be baked. (the original new york cheesecake that hails from junior's on flatbush avenue in brookl
yn is a baked cheesecake.) however, there are scads of recipes that use a blend of other rich ingredients—such as sour cream, g
reek yogurt, or heavy cream—that firm up in the refrigerator to create a "no-bake cheesecake."

• Stop words removal

```python
import nltk
from nltk.corpus import stopwords
sw_nltk = stopwords.words('english')
print(sw_nltk)
words = [word for word in newfile.split() if word.lower() not in sw_nltk]
new_text = " ".join(words)
print(new_text)
print("Old length: ", len(newfile))
print("New length: ", len(new_text))
```

['i', 'me', 'my', 'myself', 'we', 'our', 'ours', 'ourselves', 'you', "you're", "you've", "you'll", "you'd", 'your', 'yours', 'y
ourself', 'yourselves', 'he', 'him', 'his', 'himself', 'she', "she's", 'her', 'hers', 'herself', 'it', "it's", 'its', 'itself',
'they', 'them', 'their', 'theirs', 'themselves', 'what', 'which', 'who', 'whom', 'this', 'that', "that'll", 'these', 'those',
'am', 'is', 'are', 'was', 'were', 'be', 'been', 'being', 'have', 'has', 'had', 'having', 'do', 'does', 'did', 'doing', 'a', 'a
n', 'the', 'and', 'but', 'if', 'or', 'because', 'as', 'until', 'while', 'of', 'at', 'by', 'for', 'with', 'about', 'against', 'b
etween', 'into', 'through', 'during', 'before', 'after', 'above', 'below', 'to', 'from', 'up', 'down', 'in', 'out', 'on', 'of
f', 'over', 'under', 'again', 'further', 'then', 'once', 'here', 'there', 'when', 'where', 'why', 'how', 'all', 'any', 'both',
'each', 'few', 'more', 'most', 'other', 'some', 'such', 'no', 'nor', 'not', 'only', 'own', 'same', 'so', 'than', 'too', 'very',
's', 't', 'can', 'will', 'just', 'don', "don't", 'should', "should've", 'now', 'd', 'll', 'm', 'o', 're', 've', 'y', 'ain', 'ar
en', "aren't", 'couldn', "couldn't", 'didn', "didn't", 'doesn', "doesn't", 'hadn', "hadn't", 'hasn', "hasn't", 'haven', "have
n't", 'isn', "isn't", 'ma', 'mightn', "mightn't", 'mustn', "mustn't", 'needn', "needn't", 'shan', "shan't", 'shouldn', "should
n't", 'wasn', "wasn't", 'weren', "weren't", 'won', "won't", 'wouldn', "wouldn't"]
Americans think cheesecake now, often associated product cream cheese base. Cream cheese invented 1872 American dairyman Willia
m Lawrence Chester, New York, accidentally stumbled method producing cream cheese trying reproduce French cheese called Neufchâ
tel. 1880, Lawrence began distributing cream cheese foil wrappers auspices Empire Cheese Company South Edmeston, New York, manu
factured product. However, might know better famous name Lawrence came "not Neufchâtel"—Philadelphia Brand Cream Cheese. 1903,
Phoenix Cheese Company bought Lawrence's business—and Philadelphia trademark. 1928, brand bought Kraft Cheese Company. Jame
s L. Kraft invented pasteurized cheese 1912, led development pasteurized Philadelphia Brand cream cheese, currently popular che
ese used cheesecake making. Kraft Foods still owns produces Philadelphia Cream Cheese today. Fast Facts: Cheesecake Favorites T
raditional Greek Cheesecake—Most "traditional" Greek cheesecake made using ricotta cheese, however, real deal, try find authent
ic unsalted anthotyros myzirtha cheeses made either goat's sheep's milk. Greek cheesecake usually sweetened honey. recipes inco
rporate flour directly cheese/honey mixture prior baking, others employ crust. Cream Cheese Cheesecake—The cheesecake Americans
grew one another version cream cheese cheesecake. bottom cheesecakes, you'll usually find crust made crushed Graham crackers co
okies (Oreos top choice chocolate cheesecakes) blended butter tamped bottom pan mold. Cheesecakes rely custard base must baked.
(The original New York Cheesecake hails Junior's Flatbush Avenue Brooklyn baked cheesecake.) However, scads recipes use blend r
ich ingredients—such sour cream, Greek yogurt, heavy cream—that firm refrigerator create "no-bake cheesecake."
Old length: 2242
New length: 1763

• Stemming

```
#Below is the implementation of stemming words using NLTK:
from nltk.stem import PorterStemmer
from nltk.tokenize import word_tokenize

ps = PorterStemmer()

# choose some words to be stemmed
words = word_tokenize(newfile)

for w in words:
    print(w, " : ", ps.stem(w))
```

```
When  :  when
Americans  :  american
think  :  think
of  :  of
cheesecake  :  cheesecak
now  :  now
,  :  ,
it  :  it
's  :  's
most  :  most
often  :  often
associated  :  associ
with  :  with
a  :  a
product  :  product
that  :  that
has  :  ha
a  :  a
cream  :  cream
```

```
#Stemming words from sentences
from nltk.stem import PorterStemmer
from nltk.tokenize import word_tokenize

ps = PorterStemmer()

words = sent_tokenize(newfile)

for w in words:
    print(w, " : ", ps.stem(w))
```

When Americans think of cheesecake now, it's most often associated with a product that has a cream cheese base. : when americ
ans think of cheesecake now, it's most often associated with a product that has a cream cheese base.
Cream cheese was invented in 1872 by American dairyman William Lawrence of Chester, New York, who accidentally stumbled on a me
thod of producing cream cheese while trying to reproduce a French cheese called Neufchâtel. : cream cheese was invented in 18
72 by american dairyman william lawrence of chester, new york, who accidentally stumbled on a method of producing cream cheese
while trying to reproduce a french cheese called neufchâtel.
In 1880, Lawrence began distributing his cream cheese in foil wrappers under the auspices of the Empire Cheese Company of South
Edmeston, New York, where he manufactured the product. : in 1880, lawrence began distributing his cream cheese in foil wrappe
rs under the auspices of the empire cheese company of south edmeston, new york, where he manufactured the product.
However, you might know it better by the more famous name Lawrence came up with for his "not Neufchâtel"—Philadelphia Brand Cre
am Cheese. : however, you might know it better by the more famous name lawrence came up with for his "not neufchâtel"—philade
lphia brand cream cheese.
In 1903, the Phoenix Cheese Company bought Lawrence's business—and with it, the Philadelphia trademark. : in 1903, the phoeni
x cheese company bought lawrence's business—and with it, the philadelphia trademark.
In 1928, the brand was bought by the Kraft Cheese Company. : in 1928, the brand was bought by the kraft cheese company.
James L. Kraft invented pasteurized cheese in 1912, which led to the development of pasteurized Philadelphia Brand cream chees
e, currently the most popular cheese used for cheesecake making. : james l. kraft invented pasteurized cheese in 1912, which
led to the development of pasteurized philadelphia brand cream cheese, currently the most popular cheese used for cheesecake ma
king.
Kraft Foods still owns and produces Philadelphia Cream Cheese today. : kraft foods still owns and produces philadelphia cream
cheese today.
Fast Facts: Cheesecake Favorites
Traditional Greek Cheesecake—Most "traditional" Greek cheesecake is made using ricotta cheese, however, for the real deal, try
to find authentic unsalted anthotyros or myzirtha cheeses which are made with either goat's or sheep's milk. : fast facts: ch
eesecake favorites

• Lemmatization

```python
from nltk.stem import WordNetLemmatizer

lemmatizer = WordNetLemmatizer()
b=newfile

print(lemmatizer.lemmatize(b))


# a denotes adjective in "pos"
print(lemmatizer.lemmatize(b, pos ="a"))
```

When Americans think of cheesecake now, it's most often associated with a product that has a cream cheese base. Cream cheese was invented in 1872 by American dairyman William Lawrence of Chester, New York, who accidentally stumbled on a method of producing cream cheese while trying to reproduce a French cheese called Neufchâtel.

In 1880, Lawrence began distributing his cream cheese in foil wrappers under the auspices of the Empire Cheese Company of South Edmeston, New York, where he manufactured the product. However, you might know it better by the more famous name Lawrence came up with for his "not Neufchâtel"—Philadelphia Brand Cream Cheese.

In 1903, the Phoenix Cheese Company bought Lawrence's business—and with it, the Philadelphia trademark. In 1928, the brand was bought by the Kraft Cheese Company. James L. Kraft invented pasteurized cheese in 1912, which led to the development of pasteurized Philadelphia Brand cream cheese, currently the most popular cheese used for cheesecake making. Kraft Foods still owns and produces Philadelphia Cream Cheese today.

Fast Facts: Cheesecake Favorites
Traditional Greek Cheesecake—Most "traditional" Greek cheesecake is made using ricotta cheese, however, for the real deal, try to find authentic unsalted anthotyros or myzirtha cheeses which are made with either goat's or sheep's milk. Greek cheesecake is usually sweetened with honey. Some recipes incorporate flour directly into the cheese/honey mixture prior to baking, while others employ a crust.
Cream Cheese Cheesecake—The cheesecake most Americans grew up with is one or another version of a cream cheese cheesecake. At the bottom of such cheesecakes, you'll usually find a crust made of crushed Graham crackers or other cookies (Oreos are a top choice for chocolate cheesecakes) that have been blended with butter and tamped into the bottom of a pan or mold. Cheesecakes that rely on a custard base must be baked. (The original New York Cheesecake that hails from Junior's on Flatbush Avenue in Brooklyn is a baked cheesecake.) However, there are scads of recipes that use a blend of other rich ingredients—such as sour cream, Greek yogurt, or heavy cream—that firm up in the refrigerator to create a "no-bake cheesecake."

• Part of speech tagger

```python
In [24]: import nltk
         from nltk.corpus import stopwords
         from nltk.tokenize import word_tokenize, sent_tokenize
         stop_words = set(stopwords.words('english'))
```

```python
In [28]: tokenized = sent_tokenize(newfile)
         for i in tokenized:
             wordsList = nltk.word_tokenize(i)
             wordsList = [w for w in wordsList if not w in stop_words]
             tagged = nltk.pos_tag(wordsList)
             print(tagged)
```

```
[('When', 'WRB'), ('Americans', 'NNPS'), ('think', 'VBP'), ('cheesecake', 'VB'), (',', ','), ("'s", 'POS'), ('often', 'RB'),
('associated', 'JJ'), ('product', 'NN'), ('cream', 'NN'), ('cheese', 'JJ'), ('base', 'NN'), ('.', '.')]
[('Cream', 'NNP'), ('cheese', 'NN'), ('invented', 'VBD'), ('1872', 'CD'), ('American', 'NNP'), ('dairyman', 'NN'), ('William',
'NNP'), ('Lawrence', 'NNP'), ('Chester', 'NNP'), (',', ','), ('New', 'NNP'), ('York', 'NNP'), (',', ','), ('accidentally', 'R
B'), ('stumbled', 'VBD'), ('method', 'NN'), ('producing', 'VBG'), ('cream', 'NN'), ('cheese', 'NN'), ('trying', 'VBG'), ('repro
duce', 'VB'), ('French', 'JJ'), ('cheese', 'NN'), ('called', 'VBN'), ('Neufchâtel', 'NNP'), ('.', '.')]
[('In', 'IN'), ('1880', 'CD'), (',', ','), ('Lawrence', 'NNP'), ('began', 'VBD'), ('distributing', 'VBG'), ('cream', 'NN'), ('c
heese', 'JJ'), ('foil', 'NN'), ('wrappers', 'NNS'), ('auspices', 'NNS'), ('Empire', 'NNP'), ('Cheese', 'NNP'), ('Company', 'NN
P'), ('South', 'NNP'), ('Edmeston', 'NNP'), (',', ','), ('New', 'NNP'), ('York', 'NNP'), (',', ','), ('manufactured', 'VBD'),
('product', 'NN'), ('.', '.')]
[('However', 'RB'), (',', ','), ('might', 'MD'), ('know', 'VB'), ('better', 'RB'), ('famous', 'JJ'), ('name', 'NN'), ('Lawrenc
e', 'NNP'), ('came', 'VBD'), ('``', '``'), ('Neufchâtel', 'NNP'), ("''", "''"), ('—Philadelphia', 'NN'), ('Brand', 'NNP'), ('Cr
eam', 'NNP'), ('Cheese', 'NNP'), ('.', '.')]
[('In', 'IN'), ('1903', 'CD'), (',', ','), ('Phoenix', 'NNP'), ('Cheese', 'NNP'), ('Company', 'NNP'), ('bought', 'VBD'), ('Lawr
ence', 'NNP'), ("'s", 'POS'), ('business—and', 'NN'), (',', ','), ('Philadelphia', 'NNP'), ('trademark', 'NN'), ('.', '.')]
[('In', 'IN'), ('1928', 'CD'), (',', ','), ('brand', 'NN'), ('bought', 'VBD'), ('Kraft', 'NNP'), ('Cheese', 'NNP'), ('Company',
'NNP'), ('.', '.')]
[('James', 'NNP'), ('L.', 'NNP'), ('Kraft', 'NNP'), ('invented', 'VBD'), ('pasteurized', 'VBN'), ('cheese', 'JJ'), ('1912', 'C
D'), (',', ','), ('led', 'VBD'), ('development', 'NN'), ('pasteurized', 'VBN'), ('Philadelphia', 'NNP'), ('Brand', 'NNP'), ('cr
eam', 'NN'), ('cheese', 'NN'), (',', ','), ('currently', 'RB'), ('popular', 'JJ'), ('cheese', 'NNS'), ('used', 'VBN'), ('cheese
cake', 'NN'), ('making', 'NN'), ('.', '.')]
[('Kraft', 'NNP'), ('Foods', 'NNP'), ('still', 'RB'), ('owns', 'VBZ'), ('produces', 'NNS'), ('Philadelphia', 'NNP'), ('Cream',
'NNP'), ('Cheese', 'NNP'), ('today', 'NN'), ('.', '.')]
[('Fast', 'JJ'), ('Facts', 'NNS'), (':', ':'), ('Cheesecake', 'VB'), ('Favorites', 'NNP'), ('Traditional', 'NNP'), ('Greek', 'N
NP'), ('Cheesecake—Most', 'NNP'), ('“', 'NNP'), ('traditional', 'JJ'), ('”', 'NNP'), ('Greek', 'NNP'), ('cheesecake', 'NN'),
('made', 'VBD'), ('using', 'VBG'), ('ricotta', 'NN'), ('cheese', 'NN'), (',', ','), ('however', 'RB'), (',', ','), ('real', 'J
J'), ('deal', 'NN'), (',', ','), ('try', 'VB'), ('find', 'VBP'), ('authentic', 'JJ'), ('unsalted', 'JJ'), ('anthotyros', 'NN
S'), ('myzirtha', 'NN'), ('cheeses', 'NNS'), ('made', 'VBD'), ('either', 'CC'), ('goat', 'JJ'), ('’', 'NN'), ('sheep', 'JJ'),
('’', 'NNP'), ('milk', 'NN'), ('.', '.')]
[('Greek', 'JJ'), ('cheesecake', 'NN'), ('usually', 'RB'), ('sweetened', 'VBD'), ('honey', 'NN'), ('.', '.')]
[('Some', 'DT'), ('recipes', 'NNS'), ('incorporate', 'VBP'), ('flour', 'NNS'), ('directly', 'RB'), ('cheese/honey', 'VBP'), ('m
ixture', 'NN'), ('prior', 'RB'), ('baking', 'NN'), (',', ','), ('others', 'NNS'), ('employ', 'VBP'), ('crust', 'NN'), ('.',
'.')]

[('Cream', 'NNP'), ('Cheese', 'NNP'), ('Cheesecake—The', 'NNP'), ('cheesecake', 'NN'), ('Americans', 'NNPS'), ('grew', 'VBD'),
('one', 'CD'), ('another', 'DT'), ('version', 'NN'), ('cream', 'NN'), ('cheese', 'JJ'), ('cheesecake', 'NN'), ('.', '.')]
[('At', 'IN'), ('bottom', 'JJ'), ('cheesecakes', 'NNS'), (',', ','), ('’', 'NNP'), ('usually', 'RB'), ('find', 'VBP'), ('crus
t', 'JJ'), ('made', 'VBN'), ('crushed', 'JJ'), ('Graham', 'NNP'), ('crackers', 'NNS'), ('cookies', 'NNS'), ('(', '('), ('Oreo
s', 'NNP'), ('top', 'JJ'), ('choice', 'NN'), ('chocolate', 'NN'), ('cheesecakes', 'NNS'), (')', ')'), ('blended', 'VBD'), ('but
ter', 'NN'), ('tamped', 'VBD'), ('bottom', 'JJ'), ('pan', 'NN'), ('mold', 'NN'), ('.', '.')]
[('Cheesecakes', 'NNP'), ('rely', 'RB'), ('custard', 'JJ'), ('base', 'NN'), ('must', 'MD'), ('baked', 'VB'), ('.', '.')]
[('(', '('), ('The', 'DT'), ('original', 'JJ'), ('New', 'NNP'), ('York', 'NNP'), ('Cheesecake', 'NNP'), ('hails', 'VBZ'), ('Jun
ior', 'NNP'), ("'s", 'POS'), ('Flatbush', 'NNP'), ('Avenue', 'NNP'), ('Brooklyn', 'NNP'), ('baked', 'VBD'), ('cheesecake', 'N
N'), ('.', '.'), (')', ')')]
[('However', 'RB'), (',', ','), ('scads', 'NNS'), ('recipes', 'NNS'), ('use', 'VBP'), ('blend', 'NN'), ('rich', 'JJ'), ('ingred
ients—such', 'JJ'), ('sour', 'JJ'), ('cream', 'NN'), (',', ','), ('Greek', 'NNP'), ('yogurt', 'NN'), (',', ','), ('heavy', 'J
J'), ('cream—that', 'NN'), ('firm', 'NN'), ('refrigerator', 'NN'), ('create', 'VBP'), ('“', 'JJ'), ('no-bake', 'JJ'), ('cheesec
ake', 'NN'), ('.', '.'), ('”', 'NN')]
```