Importing Libraries and Dataset

The first step is to import the libraries required to execute the scripts, along with the dataset. The following code imports the required libraries:

In [1]:

```
from __future__ import print_function
import os
import nltk
import nltk.corpus
nltk.download('popular')
print(os.listdir(nltk.data.find("corpora")))
```

```
[nltk_data] Downloading collection 'popular'
[nltk_data]    |
[nltk_data]    | Downloading package cmudict to
[nltk_data]    |     C:\Users\User\AppData\Roaming\nltk_data...
[nltk_data]    |   Package cmudict is already up-to-date!
[nltk_data]    | Downloading package gazetteers to
[nltk_data]    |     C:\Users\User\AppData\Roaming\nltk_data...
[nltk_data]    |   Package gazetteers is already up-to-date!
[nltk_data]    | Downloading package genesis to
[nltk_data]    |     C:\Users\User\AppData\Roaming\nltk_data...
[nltk_data]    |   Package genesis is already up-to-date!
[nltk_data]    | Downloading package gutenberg to
[nltk_data]    |     C:\Users\User\AppData\Roaming\nltk_data...
[nltk_data]    |   Package gutenberg is already up-to-date!
[nltk_data]    | Downloading package inaugural to
[nltk_data]    |     C:\Users\User\AppData\Roaming\nltk_data...
[nltk_data]    |   Package inaugural is already up-to-date!
[nltk_data]    | Downloading package movie_reviews to
[nltk_data]    |     C:\Users\User\AppData\Roaming\nltk_data...
[nltk_data]    |   Package movie_reviews is already up-to-date!
[nltk_data]    | Downloading package names to
[nltk_data]    |     C:\Users\User\AppData\Roaming\nltk_data...
[nltk_data]    |   Package names is already up-to-date!
[nltk_data]    | Downloading package shakespeare to
[nltk_data]    |     C:\Users\User\AppData\Roaming\nltk_data...
[nltk_data]    |   Package shakespeare is already up-to-date!
[nltk_data]    | Downloading package stopwords to
[nltk_data]    |     C:\Users\User\AppData\Roaming\nltk_data...
[nltk_data]    |   Package stopwords is already up-to-date!
[nltk_data]    | Downloading package treebank to
[nltk_data]    |     C:\Users\User\AppData\Roaming\nltk_data...
[nltk_data]    |   Package treebank is already up-to-date!
[nltk_data]    | Downloading package twitter_samples to
[nltk_data]    |     C:\Users\User\AppData\Roaming\nltk_data...
[nltk_data]    |   Package twitter_samples is already up-to-date!
[nltk_data]    | Downloading package omw to
[nltk_data]    |     C:\Users\User\AppData\Roaming\nltk_data...
[nltk_data]    |   Package omw is already up-to-date!
[nltk_data]    | Downloading package wordnet to
[nltk_data]    |     C:\Users\User\AppData\Roaming\nltk_data...
[nltk_data]    |   Package wordnet is already up-to-date!
[nltk_data]    | Downloading package wordnet_ic to
[nltk_data]    |     C:\Users\User\AppData\Roaming\nltk_data...
[nltk_data]    |   Package wordnet_ic is already up-to-date!
[nltk_data]    | Downloading package words to
[nltk_data]    |     C:\Users\User\AppData\Roaming\nltk_data...
[nltk_data]    |   Package words is already up-to-date!
[nltk_data]    | Downloading package maxent_ne_chunker to
[nltk_data]    |     C:\Users\User\AppData\Roaming\nltk_data...
[nltk_data]    |   Package maxent_ne_chunker is already up-to-date!
[nltk_data]    | Downloading package punkt to
[nltk_data]    |     C:\Users\User\AppData\Roaming\nltk_data...
```

```
['cmudict', 'cmudict.zip', 'gazetteers', 'gazetteers.zip', 'genesis', 'genesis.zip', 'gutenberg', 'gu
tenberg.zip', 'inaugural', 'inaugural.zip', 'movie_reviews', 'movie_reviews.zip', 'names', 'names.zi
p', 'omw', 'omw.zip', 'shakespeare', 'shakespeare.zip', 'stopwords', 'stopwords.zip', 'treebank', 'tr
eebank.zip', 'twitter_samples', 'twitter_samples.zip', 'wordnet', 'wordnet.zip', 'wordnet_ic', 'wordn
et_ic.zip', 'words', 'words.zip']
```

```
[nltk_data]    |   Package punkt is already up-to-date!
[nltk_data]    | Downloading package snowball_data to
[nltk_data]    |     C:\Users\User\AppData\Roaming\nltk_data...
[nltk_data]    |   Package snowball_data is already up-to-date!
[nltk_data]    | Downloading package averaged_perceptron_tagger to
[nltk_data]    |     C:\Users\User\AppData\Roaming\nltk_data...
[nltk_data]    |   Package averaged_perceptron_tagger is already up-
[nltk_data]    |       to-date!
[nltk_data]    |
[nltk_data]  Done downloading collection popular
```

The next step is to download the dataset. We will use Python's NLTK library to download the dataset. We will be using the Gutenberg Dataset, which contains 3036 English books written by 142 authors, including the "Hamlet" by Shakespeare.

The following script downloads the Gutenberg dataset and prints the names of all the files in the dataset.

In [2]:

```
nltk.download('gutenberg')
from nltk.corpus import gutenberg
nltk.corpus.gutenberg.fileids()
```

```
[nltk_data] Downloading package gutenberg to
[nltk_data]     C:\Users\User\AppData\Roaming\nltk_data...
[nltk_data]   Package gutenberg is already up-to-date!
```

Out[2]:

```
['austen-emma.txt',
 'austen-persuasion.txt',
 'austen-sense.txt',
 'bible-kjv.txt',
 'blake-poems.txt',
 'bryant-stories.txt',
 'burgess-busterbrown.txt',
 'carroll-alice.txt',
 'chesterton-ball.txt',
 'chesterton-brown.txt',
 'chesterton-thursday.txt',
 'edgeworth-parents.txt',
 'melville-moby_dick.txt',
 'milton-paradise.txt',
 'shakespeare-caesar.txt',
 'shakespeare-hamlet.txt',
 'shakespeare-macbeth.txt',
 'whitman-leaves.txt']
```

The file shakespeare-hamlet.txt contains raw text for the novel "Hamlet". To read the text from this file, the raw method from the gutenberg class can be used:

Let's print the first 5000 characters from out dataset:

In [3]:

```
hamlet_text = nltk.corpus.gutenberg.raw('shakespeare-hamlet.txt')
print(hamlet_text[:5000])
```

```
[The Tragedie of Hamlet by William Shakespeare 1599]


Actus Primus. Scoena Prima.

Enter Barnardo and Francisco two Centinels.

  Barnardo. Who's there?
  Fran. Nay answer me: Stand & vnfold
your selfe

   Bar. Long liue the King

   Fran. Barnardo?
  Bar. He

   Fran. You come most carefully vpon your houre

   Bar. 'Tis now strook twelue, get thee to bed Francisco

   Fran. For this releefe much thankes: 'Tis bitter cold,
And I am sicke at heart

   Barn. Haue you had quiet Guard?
  Fran. Not a Mouse stirring

   Barn. Well, goodnight. If you do meet Horatio and
Marcellus, the Riuals of my Watch, bid them make hast.
Enter Horatio and Marcellus.

  Fran. I thinke I heare them. Stand: who's there?
  Hor. Friends to this ground

   Mar. And Leige-men to the Dane

   Fran. Giue you good night

   Mar. O farwel honest Soldier, who hath relieu'd you?
  Fra. Barnardo ha's my place: giue you goodnight.

Exit Fran.

  Mar. Holla Barnardo

   Bar. Say, what is Horatio there?
  Hor. A peece of him

   Bar. Welcome Horatio, welcome good Marcellus

   Mar. What, ha's this thing appear'd againe to night

   Bar. I haue seene nothing

   Mar. Horatio saies, 'tis but our Fantasie,
And will not let beleefe take hold of him
Touching this dreaded sight, twice seene of vs,
Therefore I haue intreated him along
With vs, to watch the minutes of this Night,
That if againe this Apparition come,
He may approue our eyes, and speake to it

   Hor. Tush, tush, 'twill not appeare

   Bar. Sit downe a-while,
And let vs once againe assaile your eares,
That are so fortified against our Story,
What we two Nights haue seene

   Hor. Well, sit we downe,
And let vs heare Barnardo speake of this

   Barn. Last night of all,
When yond same Starre that's Westward from the Pole
Had made his course t' illume that part of Heauen
Where now it burnes, Marcellus and my selfe,
The Bell then beating one

   Mar. Peace, breake thee of:
Enter the Ghost.

Looke where it comes againe

   Barn. In the same figure, like the King that's dead

   Mar. Thou art a Scholler; speake to it Horatio

   Barn. Lookes it not like the King? Marke it Horatio

   Hora. Most like: It harrowes me with fear & wonder
  Barn. It would be spoke too

   Mar. Question it Horatio

   Hor. What art thou that vsurp'st this time of night,
Together with that Faire and Warlike forme
In which the Maiesty of buried Denmarke
Did sometimes march: By Heauen I charge thee speake

   Mar. It is offended
```

```
   Barn. See, it stalkes away

   Hor. Stay: speake; speake: I Charge thee, speake.

Exit the Ghost.

  Mar. 'Tis gone, and will not answer

   Barn. How now Horatio? You tremble & look pale:
Is not this something more then Fantasie?
What thinke you on't?
  Hor. Before my God, I might not this beleeue
Without the sensible and true auouch
Of mine owne eyes

   Mar. Is it not like the King?
  Hor. As thou art to thy selfe,
Such was the very Armour he had on,
When th' Ambitious Norwey combatted:
So frown'd he once, when in an angry parle
He smot the sledded Pollax on the Ice.
'Tis strange

   Mar. Thus twice before, and iust at this dead houre,
With Martiall stalke, hath he gone by our Watch

   Hor. In what particular thought to work, I know not:
But in the grosse and scope of my Opinion,
This boades some strange erruption to our State

   Mar. Good now sit downe, & tell me he that knowes
Why this same strict and most obseruant Watch,
So nightly toyles the subiect of the Land,
And why such dayly Cast of Brazon Cannon
And Forraigne Mart for Implements of warre:
Why such impresse of Ship-wrights, whose sore Taske
Do's not diuide the Sunday from the weeke,
What might be toward, that this sweaty hast
Doth make the Night ioynt-Labourer with the day:
Who is't that can informe me?
  Hor. That can I,
At least the whisper goes so: Our last King,
Whose Image euen but now appear'd to vs,
Was (as you know) by Fortinbras of Norway,
(Thereto prick'd on by a most emulate Pride)
Dar'd to the Combate. In which, our Valiant Hamlet,
(For so this side of our knowne world esteem'd him)
Did slay this Fortinbras: who by a Seal'd Compact,
Well ratified by Law, and Heraldrie,
Did forfeite (with his life) all those his Lands
Which he stood seiz'd on, to the Conqueror:
Against the which, a Moity competent
Was gaged by our King: which had return'd
To the Inheritance of Fortinbras,
Had he bin Vanquisher, as by the same Cou'nant
And carriage of the Article designe,
His fell to Hamlet. Now sir, young Fortinbras,
Of vnimproued Mettle, hot and full,
Hath in the skirts of Norway, heere and there,
Shark'd vp a List of Landlesse Resolutes,
For Foode and Diet, to some Enterprize
That hath a stomacke in't: which is no other
(And it doth well appeare vnto our State)
But to recouer of vs by strong hand
And termes Compulsatiue, those foresaid Lands
So by his Father lost: and this (I take it)
Is the maine Motiue of our Preparations,
The Sourse of this our Watch, and the cheefe head
Of this post-hast, and Romage in the Land.
Enter Ghos
```

Tokenizing words

In [4]:
```
AI = """In computer science, artificial intelligence (AI), sometimes called machine intelligence, is in
telligence demonstrated by machines, unlike the natural intelligence displayed by humans and animals. L
eading AI textbooks define the field as the study of "intelligent agents": any device that perceives it
s environment and takes actions that maximize its chance of successfully achieving its goals."""
```

In [5]:
```
type(AI)
```

Out[5]:
```
str
```

In [6]:
```
from nltk.tokenize import word_tokenize
```

In [7]:
```
AI_tokens = word_tokenize(AI)
AI_tokens
```

Out[7]:
```
['In',
 'computer',
 'science',
 ',',
 'artificial',
 'intelligence',
 '(',
 'AI',
 ')',
 ',',
 'sometimes',
 'called',
 'machine',
 'intelligence',
 ',',
 'is',
 'intelligence',
 'demonstrated',
 'by',
 'machines',
 ',',
 'unlike',
 'the',
 'natural',
 'intelligence',
 'displayed',
 'by',
 'humans',
 'and',
 'animals',
 '.',
 'Leading',
 'AI',
 'textbooks',
 'define',
 'the',
 'field',
 'as',
 'the',
 'study',
 'of',
 '``',
 'intelligent',
 'agents',
 "''",
 ':',
 'any',
 'device',
 'that',
 'perceives',
 'its',
 'environment',
 'and',
 'takes',
 'actions',
 'that',
 'maximize',
 'its',
 'chance',
 'of',
 'successfully',
 'achieving',
 'its',
 'goals',
 '.']
```

In [8]:
```
len(AI_tokens)
```

Out[8]:
```
65
```

Frequency of each word

In [9]:
```
from nltk.probability import FreqDist
fdist = FreqDist()
```

In [10]:
```
for word in AI_tokens:
    fdist[word.lower()]+=1
fdist
```

Out[10]:
```
FreqDist({',': 4, 'intelligence': 4, 'the': 3, 'its': 3, 'ai': 2, 'by': 2, 'and': 2, '.': 2, 'of': 2,
'that': 2, ...})
```

In [11]:
```
fdist_top10 = fdist.most_common(10) ## frequency of first 10 words
fdist_top10
```

Out[11]:
```
[(',', 4),
 ('intelligence', 4),
 ('the', 3),
 ('its', 3),
 ('ai', 2),
 ('by', 2),
 ('and', 2),
 ('.', 2),
 ('of', 2),
 ('that', 2)]
```

In [12]:
```
from nltk.tokenize import blankline_tokenize
AI_blank = blankline_tokenize(AI)
len(AI_blank)
```

Out[12]:
```
1
```

1 indicates how many paragraphs we have separated by a new line.

Tokenization Types:

1. Bigrams: Tokens of two consecutive written words
2. Trigrams Tokens of three consecutive written words
3. Ngrams: Tokens of any no. of consecutive written words

In [13]:
```
from nltk.util import bigrams,trigrams,ngrams
```

In [14]:
```
string  = "Instead of worrying about what you cannot control, shift your energy to what you can creat
e."
quotes_tokens = nltk.word_tokenize(string)
quotes_tokens
```

Out[14]:
```
['Instead',
 'of',
 'worrying',
 'about',
 'what',
 'you',
 'can',
 'not',
 'control',
 ',',
 'shift',
 'your',
 'energy',
 'to',
 'what',
 'you',
 'can',
 'create',
```

```
 '.']
```

In [15]:
```
quotes_bigrams = list (nltk.bigrams(quotes_tokens)) ##tokens in paired form
quotes_bigrams
```

Out[15]:
```
[('Instead', 'of'),
 ('of', 'worrying'),
 ('worrying', 'about'),
 ('about', 'what'),
 ('what', 'you'),
 ('you', 'can'),
 ('can', 'not'),
 ('not', 'control'),
 ('control', ','),
 (',', 'shift'),
 ('shift', 'your'),
 ('your', 'energy'),
 ('energy', 'to'),
 ('to', 'what'),
 ('what', 'you'),
 ('you', 'can'),
 ('can', 'create'),
 ('create', '.')]
```

In [16]:
```
quotes_ngrams = list (nltk.ngrams(quotes_tokens,5)) ##tokens in paired form
quotes_ngrams
```

Out[16]:
```
[('Instead', 'of', 'worrying', 'about', 'what'),
 ('of', 'worrying', 'about', 'what', 'you'),
 ('worrying', 'about', 'what', 'you', 'can'),
 ('about', 'what', 'you', 'can', 'not'),
 ('what', 'you', 'can', 'not', 'control'),
 ('you', 'can', 'not', 'control', ','),
 ('can', 'not', 'control', ',', 'shift'),
 ('not', 'control', ',', 'shift', 'your'),
 ('control', ',', 'shift', 'your', 'energy'),
 (',', 'shift', 'your', 'energy', 'to'),
 ('shift', 'your', 'energy', 'to', 'what'),
 ('your', 'energy', 'to', 'what', 'you'),
 ('energy', 'to', 'what', 'you', 'can'),
 ('to', 'what', 'you', 'can', 'create'),
 ('what', 'you', 'can', 'create', '.')]
```

Stemming

Normalize words into its base form or root form

In [17]:
```
from nltk.stem import PorterStemmer
pst = PorterStemmer()
pst.stem("Having")
```

Out[17]:
```
'have'
```

In [18]:
```
words_to_stem=['give','giving','given','gave']
for words in words_to_stem:
    print(words+":"+pst.stem(words))
```

```
give:give
giving:give
given:given
gave:gave
```

In [19]:
```
from nltk.stem import LancasterStemmer
lst=LancasterStemmer()
for words in words_to_stem:
    print(words+":"+lst.stem(words))
```

```
give:giv
giving:giv
given:giv
gave:gav
```

Lemmatization

- Groups together different inflected forms of a word called Lemma
- Somehow similar to stemming, as it maps several words into one common root
- Output of Lemmatization is a proper word

In [20]:
```
from nltk.stem import wordnet
from nltk.stem import WordNetLemmatizer
word_len=WordNetLemmatizer()
word_len.lemmatize('corpora')
```

Out[20]:
```
'corpus'
```

In [21]:
```
words_to_lem=['gone', 'Going']
for words in words_to_lem:
    print(words+":"+word_len.lemmatize(words))
```

```
gone:gone
Going:Going
```

Stopwords

In [22]:
```
from nltk.corpus import stopwords
stopwords.words('english')
```

Out[22]:
```
['i',
 'me',
 'my',
 'myself',
 'we',
 'our',
 'ours',
 'ourselves',
 'you',
 "you're",
 "you've",
 "you'll",
 "you'd",
 'your',
 'yours',
 'yourself',
 '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',
 'an',
 'the',
 'and',
 'but',
 'if',
 'or',
 'because',
 'as',
 'until',
 'while',
 'of',
 'at',
 'by',
 'for',
 'with',
 'about',
 'against',
 'between',
 'into',
 'through',
 'during',
 'before',
 'after',
 'above',
 'below',
 'to',
 'from',
 'up',
 'down',
 'in',
 'out',
 'on',
 'off',
 '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',
 'aren',
 "aren't",
 'couldn',
 "couldn't",
 'didn',
 "didn't",
 'doesn',
 "doesn't",
 'hadn',
 "hadn't",
 'hasn',
 "hasn't",
 'haven',
 "haven't",
 'isn',
 "isn't",
 'ma',
 'mightn',
 "mightn't",
 'mustn',
 "mustn't",
 'needn',
 "needn't",
 'shan',
 "shan't",
 'shouldn',
 "shouldn't",
 'wasn',
 "wasn't",
 'weren',
 "weren't",
 'won',
 "won't",
 'wouldn',
 "wouldn't"]
```

In [23]:

```
len(stopwords.words('english'))
```

Out[23]:

```
179
```

In [24]:

```
fdist_top10
```

Out[24]:

```
[(',', 4),
 ('intelligence', 4),
 ('the', 3),
 ('its', 3),
 ('ai', 2),
 ('by', 2),
 ('and', 2),
 ('.', 2),
 ('of', 2),
 ('that', 2)]
```

In [25]:

```
import re
punctuation=re.compile(r'[-.?!,:;()\'|0-9]')
#compile function from re module to form a list with any digit or special charater
```

In [26]:

```
post_punctuation=[]
for words in AI_tokens:
    word=punctuation.sub("/",words)
    if len(word)>0:
        post_punctuation.append(word)
```

```
In [27]: post_punctuation
```

```
Out[27]: ['In',
 'computer',
 'science',
 '/',
 'artificial',
 'intelligence',
 '/',
 'AI',
 '/',
 '/',
 'sometimes',
 'called',
 'machine',
 'intelligence',
 '/',
 'is',
 'intelligence',
 'demonstrated',
 'by',
 'machines',
 '/',
 'unlike',
 'the',
 'natural',
 'intelligence',
 'displayed',
 'by',
 'humans',
 'and',
 'animals',
 '/',
 'Leading',
 'AI',
 'textbooks',
 'define',
 'the',
 'field',
 'as',
 'the',
 'study',
 'of',
 '``',
 'intelligent',
 'agents',
 '//',
 '/',
 'any',
 'device',
 'that',
 'perceives',
 'its',
 'environment',
 'and',
 'takes',
 'actions',
 'that',
 'maximize',
 'its',
 'chance',
 'of',
 'successfully',
 'achieving',
 'its',
 'goals',
 '/']
```

```
In [28]: len(post_punctuation)
```

```
Out[28]: 65
```

POS: Parts of speech

- Generally speaking gramatical types of words like noun,verb,adverb, ajectives
- A word can have more then one parts of speech based on the context it is used
- Ex. "Google something on the internet", Here google is used as a verb although its a noun.
- These are some sort of ambiquities or difficulties which makes the NLU much more harder as compared to NLG
- Because once u understand the language Generation is quite easy

```
In [29]: sent='krishna is a natural when it comes to singing'
         sent_tokens=word_tokenize(sent)
```

```
In [30]: for token in sent_tokens:
             print(nltk.pos_tag([token]))
```

```
[('krishna', 'NN')]
[('is', 'VBZ')]
[('a', 'DT')]
[('natural', 'JJ')]
[('when', 'WRB')]
[('it', 'PRP')]
[('comes', 'VBZ')]
[('to', 'TO')]
[('singing', 'VBG')]
```

Named Entity Recognition

Naming such as--

- movie
- monetary value
- organization
- location
- quantities
- person from a text

Three phases of NAE

1. Noun Pharase identification:-Extract all the noun phrases from a text using dependency passing and parts of speech tagging
2. Phase classification:- in this step all the extracted nouns and phrases are classified into respective categories such as location names and much more
3. Validation layer to evalute if something goes wrong using knowledge graphs</br>

- Popular knowledge craft:- Google knowledge graph, IBM Watson, Wkipedia
- Google's CEO Sundar Pichai introduced the new pixel at Minnesota Roi Centre Event
- Google-Organization
- Sundar Pichai- person
- Minnesota-Location
- Roi centre event-location

```
In [31]: import nltk
         from nltk import ne_chunk
         NE_sent="The Indian Politicians shouts in the Parliament House"
```

```
In [32]: NE_tokens=word_tokenize(NE_sent)
         NE_tags=nltk.pos_tag(NE_tokens)
```

```
In [33]: NE_NER=ne_chunk(NE_tags)
         print(NE_NER)
```

```
(S
  The/DT
  (GPE Indian/JJ)
  Politicians/NNPS
  shouts/NNS
  in/IN
  the/DT
  (ORGANIZATION Parliament/NNP House/NNP))
```

Chunking

picking up individual pieces of information and grouping them into bigger pieces

```
In [34]: new = "The cat sat on a mat and ate the rat"
         new_Tokens = nltk.pos_tag(word_tokenize(new))
         new_Tokens
```

```
Out[34]: [('The', 'DT'),
 ('cat', 'NN'),
 ('sat', 'VBD'),
 ('on', 'IN'),
 ('a', 'DT'),
 ('mat', 'NN'),
 ('and', 'CC'),
 ('ate', 'VB'),
 ('the', 'DT'),
 ('rat', 'NN')]
```

```
In [35]: grammar_np = r"NP:{<DT>?<JJ>*<NN>}"
         chunk_parser = nltk.RegexpParser(grammar_np)
```

```
In [36]: chunk_result = chunk_parser.parse(new_Tokens)
         chunk_result
```

```
The Ghostscript executable isn't found.
See http://web.mit.edu/ghostscript/www/Install.htm
If you're using a Mac, you can try installing
https://docs.brew.sh/Installation then `brew install ghostscript`
```

```
---------------------------------------------------------------------------
LookupError                               Traceback (most recent call last)
~\anaconda3\lib\site-packages\nltk\tree.py in _repr_png_(self)
    805                                 env_vars=['PATH'],
--> 806                                 verbose=False,
    807                             )

~\anaconda3\lib\site-packages\nltk\internals.py in find_binary(name, path_to_bin, env_vars, searchpat
h, binary_names, url, verbose)
    696         find_binary_iter(
--> 697             name, path_to_bin, env_vars, searchpath, binary_names, url, verbose
    698         )

~\anaconda3\lib\site-packages\nltk\internals.py in find_binary_iter(name, path_to_bin, env_vars, sear
chpath, binary_names, url, verbose)
    680     for file in find_file_iter(
--> 681         path_to_bin or name, env_vars, searchpath, binary_names, url, verbose
    682     ):

~\anaconda3\lib\site-packages\nltk\internals.py in find_file_iter(filename, env_vars, searchpath, fil
e_names, url, verbose, finding_dir)
    638         div = '=' * 75
--> 639         raise LookupError('\n\n%s\n%s\n%s' % (div, msg, div))
    640 

LookupError: 

===========================================================================
NLTK was unable to find the gs file!
Use software specific configuration paramaters or set the PATH environment variable.
===========================================================================

During handling of the above exception, another exception occurred:

LookupError                               Traceback (most recent call last)
~\anaconda3\lib\site-packages\IPython\core\formatters.py in __call__(self, obj)
    343             method = get_real_method(obj, self.print_method)
    344             if method is not None:
--> 345                 return method()
    346             return None
    347         else:

~\anaconda3\lib\site-packages\nltk\tree.py in _repr_png_(self)
    817                         "https://docs.brew.sh/Installation then `brew install
 ghostscript`")
    818                 print(pre_error_message, file=sys.stderr)
--> 819                 raise LookupError
```

```
    820
    821             with open(out_path, 'rb') as sr:

LookupError: 
```

Out[36]:

```
Tree('S', [Tree('NP', [('The', 'DT'), ('cat', 'NN')]), ('sat', 'VBD'), ('on', 'IN'), Tree('NP',
[('a', 'DT'), ('mat', 'NN')]), ('and', 'CC'), ('ate', 'VB'), Tree('NP', [('the', 'DT'), ('rat', 'N
N')])])
```

ML Classifier - Movie Review from the nltk corpora

In [37]:

```python
import pandas as pd
import numpy as np
```

In [38]:

```python
from sklearn.feature_extraction.text import CountVectorizer
print (os.listdir(nltk.data.find("corpora")))
```

```
['cmudict', 'cmudict.zip', 'gazetteers', 'gazetteers.zip', 'genesis', 'genesis.zip', 'gutenberg', 'gu
tenberg.zip', 'inaugural', 'inaugural.zip', 'movie_reviews', 'movie_reviews.zip', 'names', 'names.zi
p', 'omw', 'omw.zip', 'shakespeare', 'shakespeare.zip', 'stopwords', 'stopwords.zip', 'treebank', 'tr
eebank.zip', 'twitter_samples', 'twitter_samples.zip', 'wordnet', 'wordnet.zip', 'wordnet_ic', 'wordn
et_ic.zip', 'words', 'words.zip']
```

In [39]:

```python
from nltk.corpus import movie_reviews
print(movie_reviews.categories())
```

```
['neg', 'pos']
```

In [40]:

```python
print(len(movie_reviews.fileids('pos')))
print(' ')
print(movie_reviews.fileids('pos'))
```

```
1000

['pos/cv000_29590.txt', 'pos/cv001_18431.txt', 'pos/cv002_15918.txt', 'pos/cv003_11664.txt', 'pos/cv0
04_11636.txt', 'pos/cv005_29443.txt', 'pos/cv006_15448.txt', 'pos/cv007_4968.txt', 'pos/cv008_29435.t
xt', 'pos/cv009_29592.txt', 'pos/cv010_29198.txt', 'pos/cv011_12166.txt', 'pos/cv012_29576.txt', 'po
s/cv013_10159.txt', 'pos/cv014_13924.txt', 'pos/cv015_29439.txt', 'pos/cv016_4659.txt', 'pos/cv017_22
464.txt', 'pos/cv018_20137.txt', 'pos/cv019_14482.txt', 'pos/cv020_8825.txt', 'pos/cv021_15838.txt',
'pos/cv022_12864.txt', 'pos/cv023_12672.txt', 'pos/cv024_6778.txt', 'pos/cv025_3108.txt', 'pos/cv026_
29325.txt', 'pos/cv027_25219.txt', 'pos/cv028_26746.txt', 'pos/cv029_18643.txt', 'pos/cv030_21593.tx
t', 'pos/cv031_18452.txt', 'pos/cv032_22550.txt', 'pos/cv033_24444.txt', 'pos/cv034_29647.txt', 'pos/
cv035_3954.txt', 'pos/cv036_16831.txt', 'pos/cv037_18510.txt', 'pos/cv038_9749.txt', 'pos/cv039_6170.
txt', 'pos/cv040_8276.txt', 'pos/cv041_21113.txt', 'pos/cv042_10982.txt', 'pos/cv043_15013.txt', 'po
s/cv044_16969.txt', 'pos/cv045_23923.txt', 'pos/cv046_10188.txt', 'pos/cv047_1754.txt', 'pos/cv048_16
828.txt', 'pos/cv049_20471.txt', 'pos/cv050_11175.txt', 'pos/cv051_10306.txt', 'pos/cv052_29378.txt',
'pos/cv053_21822.txt', 'pos/cv054_4230.txt', 'pos/cv055_8338.txt', 'pos/cv056_13133.txt', 'pos/cv057_
7453.txt', 'pos/cv058_8025.txt', 'pos/cv059_28885.txt', 'pos/cv060_10844.txt', 'pos/cv061_8837.txt',
'pos/cv062_23115.txt', 'pos/cv063_28997.txt', 'pos/cv064_24576.txt', 'pos/cv065_15248.txt', 'pos/cv06
6_10821.txt', 'pos/cv067_19774.txt', 'pos/cv068_13400.txt', 'pos/cv069_10801.txt', 'pos/cv070_12289.t
xt', 'pos/cv071_12095.txt', 'pos/cv072_6169.txt', 'pos/cv073_21785.txt', 'pos/cv074_6875.txt', 'pos/c
v075_6500.txt', 'pos/cv076_24945.txt', 'pos/cv077_22138.txt', 'pos/cv078_14730.txt', 'pos/cv079_1193
3.txt', 'pos/cv080_13465.txt', 'pos/cv081_16582.txt', 'pos/cv082_11080.txt', 'pos/cv083_24234.txt',
'pos/cv084_13566.txt', 'pos/cv085_1381.txt', 'pos/cv086_18371.txt', 'pos/cv087_1989.txt', 'pos/cv088_
24113.txt', 'pos/cv089_11418.txt', 'pos/cv090_0042.txt', 'pos/cv091_7400.txt', 'pos/cv092_28017.txt',
'pos/cv093_13951.txt', 'pos/cv094_27889.txt', 'pos/cv095_28892.txt', 'pos/cv096_11474.txt', 'pos/cv09
7_24970.txt', 'pos/cv098_15435.txt', 'pos/cv099_10534.txt', 'pos/cv100_11528.txt', 'pos/cv101_10175.t
xt', 'pos/cv102_7846.txt', 'pos/cv103_11021.txt', 'pos/cv104_18134.txt', 'pos/cv105_17990.txt', 'pos/
cv106_16807.txt', 'pos/cv107_24319.txt', 'pos/cv108_15571.txt', 'pos/cv109_21172.txt', 'pos/cv110_277
88.txt', 'pos/cv111_11473.txt', 'pos/cv112_11193.txt', 'pos/cv113_23102.txt', 'pos/cv114_18398.txt',
'pos/cv115_25396.txt', 'pos/cv116_28942.txt', 'pos/cv117_24295.txt', 'pos/cv118_28980.txt', 'pos/cv11
9_9867.txt', 'pos/cv120_4111.txt', 'pos/cv121_17302.txt', 'pos/cv122_7392.txt', 'pos/cv123_11182.tx
t', 'pos/cv124_4122.txt', 'pos/cv125_9391.txt', 'pos/cv126_28971.txt', 'pos/cv127_14711.txt', 'pos/cv
128_29627.txt', 'pos/cv129_16741.txt', 'pos/cv130_17083.txt', 'pos/cv131_10713.txt', 'pos/cv132_5618.
txt', 'pos/cv133_16336.txt', 'pos/cv134_22246.txt', 'pos/cv135_11603.txt', 'pos/cv136_11505.txt', 'po
s/cv137_15422.txt', 'pos/cv138_12721.txt', 'pos/cv139_12873.txt', 'pos/cv140_7479.txt', 'pos/cv141_15
686.txt', 'pos/cv142_22516.txt', 'pos/cv143_19666.txt', 'pos/cv144_5007.txt', 'pos/cv145_11472.txt',
'pos/cv146_18458.txt', 'pos/cv147_21193.txt', 'pos/cv148_16345.txt', 'pos/cv149_15670.txt', 'pos/cv15
0_12916.txt', 'pos/cv151_15771.txt', 'pos/cv152_8736.txt', 'pos/cv153_10779.txt', 'pos/cv154_9328.tx
t', 'pos/cv155_7308.txt', 'pos/cv156_10481.txt', 'pos/cv157_29372.txt', 'pos/cv158_10390.txt', 'pos/c
v159_29505.txt', 'pos/cv160_10362.txt', 'pos/cv161_11425.txt', 'pos/cv162_10424.txt', 'pos/cv163_1005
2.txt', 'pos/cv164_22447.txt', 'pos/cv165_22619.txt', 'pos/cv166_11052.txt', 'pos/cv167_16376.txt',
'pos/cv168_7050.txt', 'pos/cv169_23778.txt', 'pos/cv170_3006.txt', 'pos/cv171_13537.txt', 'pos/cv172_
11131.txt', 'pos/cv173_4471.txt', 'pos/cv174_9659.txt', 'pos/cv175_6964.txt', 'pos/cv176_12857.txt',
'pos/cv177_10367.txt', 'pos/cv178_12972.txt', 'pos/cv179_9228.txt', 'pos/cv180_16113.txt', 'pos/cv181
_14401.txt', 'pos/cv182_7281.txt', 'pos/cv183_18612.txt', 'pos/cv184_2673.txt', 'pos/cv185_28654.tx
t', 'pos/cv186_2269.txt', 'pos/cv187_12829.txt', 'pos/cv188_19226.txt', 'pos/cv189_22934.txt', 'pos/c
v190_27052.txt', 'pos/cv191_29719.txt', 'pos/cv192_14395.txt', 'pos/cv193_5416.txt', 'pos/cv194_1207
9.txt', 'pos/cv195_14528.txt', 'pos/cv196_29027.txt', 'pos/cv197_29328.txt', 'pos/cv198_18180.txt',
'pos/cv199_9629.txt', 'pos/cv200_2915.txt', 'pos/cv201_6997.txt', 'pos/cv202_10654.txt', 'pos/cv203_1
7986.txt', 'pos/cv204_8451.txt', 'pos/cv205_9457.txt', 'pos/cv206_14293.txt', 'pos/cv207_29284.txt',
'pos/cv208_9020.txt', 'pos/cv209_29118.txt', 'pos/cv210_9312.txt', 'pos/cv211_9953.txt', 'pos/cv212_1
0027.txt', 'pos/cv213_18934.txt', 'pos/cv214_12294.txt', 'pos/cv215_22240.txt', 'pos/cv216_18738.tx
t', 'pos/cv217_28842.txt', 'pos/cv218_24352.txt', 'pos/cv219_18626.txt', 'pos/cv220_29059.txt', 'pos/
cv221_2695.txt', 'pos/cv222_17395.txt', 'pos/cv223_29066.txt', 'pos/cv224_17661.txt', 'pos/cv225_2922
4.txt', 'pos/cv226_2618.txt', 'pos/cv227_24215.txt', 'pos/cv228_5806.txt', 'pos/cv229_13611.txt', 'po
s/cv230_7428.txt', 'pos/cv231_10425.txt', 'pos/cv232_14991.txt', 'pos/cv233_15964.txt', 'pos/cv234_20
643.txt', 'pos/cv235_10217.txt', 'pos/cv236_11565.txt', 'pos/cv237_19221.txt', 'pos/cv238_12931.txt',
'pos/cv239_3385.txt', 'pos/cv240_14336.txt', 'pos/cv241_23130.txt', 'pos/cv242_10638.txt', 'pos/cv243
_20728.txt', 'pos/cv244_21649.txt', 'pos/cv245_8569.txt', 'pos/cv246_28807.txt', 'pos/cv247_13142.tx
t', 'pos/cv248_13987.txt', 'pos/cv249_11640.txt', 'pos/cv250_25616.txt', 'pos/cv251_22636.txt', 'pos/
cv252_23779.txt', 'pos/cv253_10077.txt', 'pos/cv254_6027.txt', 'pos/cv255_13683.txt', 'pos/cv256_1474
0.txt', 'pos/cv257_10975.txt', 'pos/cv258_5792.txt', 'pos/cv259_10934.txt', 'pos/cv260_13959.txt', 'p
os/cv261_10954.txt', 'pos/cv262_12649.txt', 'pos/cv263_19259.txt', 'pos/cv264_12801.txt', 'pos/cv265_
10814.txt', 'pos/cv266_25779.txt', 'pos/cv267_14952.txt', 'pos/cv268_18834.txt', 'pos/cv269_21732.tx
t', 'pos/cv270_6079.txt', 'pos/cv271_13837.txt', 'pos/cv272_18974.txt', 'pos/cv273_29112.txt', 'pos/c
v274_25253.txt', 'pos/cv275_28887.txt', 'pos/cv276_15684.txt', 'pos/cv277_19091.txt', 'pos/cv278_1304
1.txt', 'pos/cv279_18329.txt', 'pos/cv280_8267.txt', 'pos/cv281_23253.txt', 'pos/cv282_6653.txt', 'po
s/cv283_11055.txt', 'pos/cv284_19119.txt', 'pos/cv285_16494.txt', 'pos/cv286_25050.txt', 'pos/cv287_1
5900.txt', 'pos/cv288_18791.txt', 'pos/cv289_6463.txt', 'pos/cv290_11084.txt', 'pos/cv291_26635.txt',
'pos/cv292_7282.txt', 'pos/cv293_29856.txt', 'pos/cv294_11684.txt', 'pos/cv295_15570.txt', 'pos/cv296
_12251.txt', 'pos/cv297_10047.txt', 'pos/cv298_23111.txt', 'pos/cv299_16214.txt', 'pos/cv300_22284.tx
t', 'pos/cv301_12146.txt', 'pos/cv302_25649.txt', 'pos/cv303_27520.txt', 'pos/cv304_28706.txt', 'pos/
cv305_9946.txt', 'pos/cv306_10364.txt', 'pos/cv307_25270.txt', 'pos/cv308_5016.txt', 'pos/cv309_2257
1.txt', 'pos/cv310_13091.txt', 'pos/cv311_16002.txt', 'pos/cv312_29377.txt', 'pos/cv313_18198.txt',
'pos/cv314_14422.txt', 'pos/cv315_11629.txt', 'pos/cv316_6370.txt', 'pos/cv317_24049.txt', 'pos/cv318
_10493.txt', 'pos/cv319_14727.txt', 'pos/cv320_9530.txt', 'pos/cv321_12843.txt', 'pos/cv322_20318.tx
t', 'pos/cv323_29805.txt', 'pos/cv324_7082.txt', 'pos/cv325_16629.txt', 'pos/cv326_13295.txt', 'pos/c
v327_20292.txt', 'pos/cv328_10373.txt', 'pos/cv329_29370.txt', 'pos/cv330_29809.txt', 'pos/cv331_827
3.txt', 'pos/cv332_16307.txt', 'pos/cv333_8916.txt', 'pos/cv334_10001.txt', 'pos/cv335_14665.txt', 'p
os/cv336_10143.txt', 'pos/cv337_29181.txt', 'pos/cv338_8821.txt', 'pos/cv339_21119.txt', 'pos/cv340_1
3287.txt', 'pos/cv341_24430.txt', 'pos/cv342_19456.txt', 'pos/cv343_10368.txt', 'pos/cv344_5312.txt',
'pos/cv345_9954.txt', 'pos/cv346_18168.txt', 'pos/cv347_13194.txt', 'pos/cv348_18176.txt', 'pos/cv349
_13507.txt', 'pos/cv350_20670.txt', 'pos/cv351_15458.txt', 'pos/cv352_5524.txt', 'pos/cv353_18159.tx
t', 'pos/cv354_8132.txt', 'pos/cv355_16413.txt', 'pos/cv356_25163.txt', 'pos/cv357_13156.txt', 'pos/c
v358_10691.txt', 'pos/cv359_6647.txt', 'pos/cv360_8398.txt', 'pos/cv361_28944.txt', 'pos/cv362_15341.
txt', 'pos/cv363_29332.txt', 'pos/cv364_12901.txt', 'pos/cv365_11576.txt', 'pos/cv366_10221.txt', 'po
s/cv367_22792.txt', 'pos/cv368_10466.txt', 'pos/cv369_12886.txt', 'pos/cv370_5221.txt', 'pos/cv371_76
30.txt', 'pos/cv372_6552.txt', 'pos/cv373_20404.txt', 'pos/cv374_25436.txt', 'pos/cv375_9929.txt', 'p
os/cv376_19435.txt', 'pos/cv377_7946.txt', 'pos/cv378_20629.txt', 'pos/cv379_21963.txt', 'pos/cv380_7
574.txt', 'pos/cv381_20172.txt', 'pos/cv382_7897.txt', 'pos/cv383_13116.txt', 'pos/cv384_17140.txt',
'pos/cv385_29741.txt', 'pos/cv386_10080.txt', 'pos/cv387_11507.txt', 'pos/cv388_12009.txt', 'pos/cv38
9_9369.txt', 'pos/cv390_11345.txt', 'pos/cv391_10802.txt', 'pos/cv392_11458.txt', 'pos/cv393_29327.tx
t', 'pos/cv394_5137.txt', 'pos/cv395_10849.txt', 'pos/cv396_17989.txt', 'pos/cv397_29023.txt', 'pos/c
v398_15537.txt', 'pos/cv399_2877.txt', 'pos/cv400_19220.txt', 'pos/cv401_12605.txt', 'pos/cv402_1442
5.txt', 'pos/cv403_6621.txt', 'pos/cv404_20315.txt', 'pos/cv405_20399.txt', 'pos/cv406_21020.txt', 'p
os/cv407_22637.txt', 'pos/cv408_5297.txt', 'pos/cv409_29786.txt', 'pos/cv410_24266.txt', 'pos/cv411_1
5007.txt', 'pos/cv412_24095.txt', 'pos/cv413_7398.txt', 'pos/cv414_10518.txt', 'pos/cv415_22517.txt',
'pos/cv416_11136.txt', 'pos/cv417_13115.txt', 'pos/cv418_14774.txt', 'pos/cv419_13394.txt', 'pos/cv42
0_28795.txt', 'pos/cv421_9709.txt', 'pos/cv422_9381.txt', 'pos/cv423_11155.txt', 'pos/cv424_8831.tx
t', 'pos/cv425_8250.txt', 'pos/cv426_10421.txt', 'pos/cv427_10825.txt', 'pos/cv428_11347.txt', 'pos/c
v429_7439.txt', 'pos/cv430_17351.txt', 'pos/cv431_7085.txt', 'pos/cv432_14224.txt', 'pos/cv433_10144.
txt', 'pos/cv434_5793.txt', 'pos/cv435_23110.txt', 'pos/cv436_19179.txt', 'pos/cv437_22849.txt', 'po
s/cv438_8043.txt', 'pos/cv439_15970.txt', 'pos/cv440_15243.txt', 'pos/cv441_13711.txt', 'pos/cv442_13
846.txt', 'pos/cv443_21118.txt', 'pos/cv444_9974.txt', 'pos/cv445_25882.txt', 'pos/cv446_11353.txt',
'pos/cv447_27332.txt', 'pos/cv448_14695.txt', 'pos/cv449_8785.txt', 'pos/cv450_7890.txt', 'pos/cv451_
10690.txt', 'pos/cv452_5088.txt', 'pos/cv453_10379.txt', 'pos/cv454_2053.txt', 'pos/cv455_29000.txt',
'pos/cv456_18985.txt', 'pos/cv457_18453.txt', 'pos/cv458_8604.txt', 'pos/cv459_20319.txt', 'pos/cv460
_10842.txt', 'pos/cv461_19600.txt', 'pos/cv462_19350.txt', 'pos/cv463_10343.txt', 'pos/cv464_15650.tx
t', 'pos/cv465_22431.txt', 'pos/cv466_18722.txt', 'pos/cv467_25773.txt', 'pos/cv468_15228.txt', 'pos/
cv469_20630.txt', 'pos/cv470_15952.txt', 'pos/cv471_16858.txt', 'pos/cv472_29280.txt', 'pos/cv473_736
7.txt', 'pos/cv474_10209.txt', 'pos/cv475_21692.txt', 'pos/cv476_16856.txt', 'pos/cv477_22479.txt',
'pos/cv478_14309.txt', 'pos/cv479_5649.txt', 'pos/cv480_19817.txt', 'pos/cv481_7436.txt', 'pos/cv482_
10580.txt', 'pos/cv483_16378.txt', 'pos/cv484_25054.txt', 'pos/cv485_26649.txt', 'pos/cv486_9799.tx
t', 'pos/cv487_10446.txt', 'pos/cv488_19856.txt', 'pos/cv489_17906.txt', 'pos/cv490_17872.txt', 'pos/
cv491_12145.txt', 'pos/cv492_18271.txt', 'pos/cv493_12839.txt', 'pos/cv494_17389.txt', 'pos/cv495_145
18.txt', 'pos/cv496_10530.txt', 'pos/cv497_26980.txt', 'pos/cv498_8832.txt', 'pos/cv499_10658.txt',
'pos/cv500_10251.txt', 'pos/cv501_11657.txt', 'pos/cv502_10406.txt', 'pos/cv503_10558.txt', 'pos/cv50
4_29243.txt', 'pos/cv505_12090.txt', 'pos/cv506_15956.txt', 'pos/cv507_9220.txt', 'pos/cv508_16006.tx
t', 'pos/cv509_15888.txt', 'pos/cv510_23360.txt', 'pos/cv511_10132.txt', 'pos/cv512_15965.txt', 'pos/
cv513_6923.txt', 'pos/cv514_11187.txt', 'pos/cv515_17069.txt', 'pos/cv516_11172.txt', 'pos/cv517_1921
9.txt', 'pos/cv518_13331.txt', 'pos/cv519_14661.txt', 'pos/cv520_12295.txt', 'pos/cv521_15828.txt',
'pos/cv522_5583.txt', 'pos/cv523_16615.txt', 'pos/cv524_23627.txt', 'pos/cv525_16122.txt', 'pos/cv526
_12083.txt', 'pos/cv527_10123.txt', 'pos/cv528_10822.txt', 'pos/cv529_10420.txt', 'pos/cv530_16212.tx
t', 'pos/cv531_26486.txt', 'pos/cv532_6522.txt', 'pos/cv533_9821.txt', 'pos/cv534_14083.txt', 'pos/cv
535_19728.txt', 'pos/cv536_27134.txt', 'pos/cv537_12370.txt', 'pos/cv538_28667.txt', 'pos/cv539_2034
7.txt', 'pos/cv540_3421.txt', 'pos/cv541_28835.txt', 'pos/cv542_18980.txt', 'pos/cv543_5045.txt', 'po
s/cv544_5108.txt', 'pos/cv545_12014.txt', 'pos/cv546_11767.txt', 'pos/cv547_16324.txt', 'pos/cv548_17
731.txt', 'pos/cv549_21443.txt', 'pos/cv550_22211.txt', 'pos/cv551_10565.txt', 'pos/cv552_10016.txt',
'pos/cv553_26915.txt', 'pos/cv554_13151.txt', 'pos/cv555_23922.txt', 'pos/cv556_14808.txt', 'pos/cv55
7_11449.txt', 'pos/cv558_29507.txt', 'pos/cv559_0050.txt', 'pos/cv560_17175.txt', 'pos/cv561_9201.tx
t', 'pos/cv562_10359.txt', 'pos/cv563_17257.txt', 'pos/cv564_11110.txt', 'pos/cv565_29572.txt', 'pos/
cv566_8581.txt', 'pos/cv567_29611.txt', 'pos/cv568_15638.txt', 'pos/cv569_26381.txt', 'pos/cv570_2908
2.txt', 'pos/cv571_29366.txt', 'pos/cv572_18657.txt', 'pos/cv573_29525.txt', 'pos/cv574_22156.txt',
'pos/cv575_21150.txt', 'pos/cv576_14094.txt', 'pos/cv577_28549.txt', 'pos/cv578_15094.txt', 'pos/cv57
9_11605.txt', 'pos/cv580_14064.txt', 'pos/cv581_19381.txt', 'pos/cv582_6559.txt', 'pos/cv583_29692.tx
t', 'pos/cv584_29722.txt', 'pos/cv585_22496.txt', 'pos/cv586_7543.txt', 'pos/cv587_19162.txt', 'pos/c
v588_13008.txt', 'pos/cv589_12064.txt', 'pos/cv590_19290.txt', 'pos/cv591_23640.txt', 'pos/cv592_2231
5.txt', 'pos/cv593_10987.txt', 'pos/cv594_11039.txt', 'pos/cv595_25335.txt', 'pos/cv596_28311.txt',
'pos/cv597_26360.txt', 'pos/cv598_16452.txt', 'pos/cv599_20988.txt', 'pos/cv600_23878.txt', 'pos/cv60
1_23453.txt', 'pos/cv602_8300.txt', 'pos/cv603_17694.txt', 'pos/cv604_2230.txt', 'pos/cv605_11800.tx
t', 'pos/cv606_15985.txt', 'pos/cv607_7717.txt', 'pos/cv608_23231.txt', 'pos/cv609_23877.txt', 'pos/c
v610_2287.txt', 'pos/cv611_21120.txt', 'pos/cv612_5461.txt', 'pos/cv613_21796.txt', 'pos/cv614_10626.
txt', 'pos/cv615_14182.txt', 'pos/cv616_29319.txt', 'pos/cv617_9322.txt', 'pos/cv618_8974.txt', 'pos/
cv619_12462.txt', 'pos/cv620_24265.txt', 'pos/cv621_14368.txt', 'pos/cv622_8147.txt', 'pos/cv623_1535
6.txt', 'pos/cv624_10744.txt', 'pos/cv625_12440.txt', 'pos/cv626_7410.txt', 'pos/cv627_11620.txt', 'p
os/cv628_19325.txt', 'pos/cv629_14909.txt', 'pos/cv630_10057.txt', 'pos/cv631_4967.txt', 'pos/cv632_9
610.txt', 'pos/cv633_29837.txt', 'pos/cv634_11101.txt', 'pos/cv635_10022.txt', 'pos/cv636_15279.txt',
'pos/cv637_1250.txt', 'pos/cv638_2953.txt', 'pos/cv639_10308.txt', 'pos/cv640_5378.txt', 'pos/cv641_1
2349.txt', 'pos/cv642_29867.txt', 'pos/cv643_29349.txt', 'pos/cv644_17154.txt', 'pos/cv645_15668.tx
t', 'pos/cv646_15065.txt', 'pos/cv647_13691.txt', 'pos/cv648_15792.txt', 'pos/cv649_12735.txt', 'pos/
cv650_14340.txt', 'pos/cv651_10492.txt', 'pos/cv652_13972.txt', 'pos/cv653_19583.txt', 'pos/cv654_182
46.txt', 'pos/cv655_11154.txt', 'pos/cv656_24201.txt', 'pos/cv657_24513.txt', 'pos/cv658_10532.txt',
'pos/cv659_19944.txt', 'pos/cv660_21893.txt', 'pos/cv661_2450.txt', 'pos/cv662_13320.txt', 'pos/cv663
_13019.txt', 'pos/cv664_4389.txt', 'pos/cv665_29538.txt', 'pos/cv666_18963.txt', 'pos/cv667_18467.tx
t', 'pos/cv668_17604.txt', 'pos/cv669_22995.txt', 'pos/cv670_25826.txt', 'pos/cv671_5054.txt', 'pos/c
v672_28083.txt', 'pos/cv673_24714.txt', 'pos/cv674_10732.txt', 'pos/cv675_21588.txt', 'pos/cv676_2109
0.txt', 'pos/cv677_17715.txt', 'pos/cv678_13419.txt', 'pos/cv679_28559.txt', 'pos/cv680_10160.txt',
'pos/cv681_9692.txt', 'pos/cv682_16139.txt', 'pos/cv683_12167.txt', 'pos/cv684_11798.txt', 'pos/cv685
_5947.txt', 'pos/cv686_13900.txt', 'pos/cv687_21100.txt', 'pos/cv688_7368.txt', 'pos/cv689_12587.tx
t', 'pos/cv690_5619.txt', 'pos/cv691_5043.txt', 'pos/cv692_15451.txt', 'pos/cv693_18063.txt', 'pos/cv
694_4876.txt', 'pos/cv695_21108.txt', 'pos/cv696_29740.txt', 'pos/cv697_11162.txt', 'pos/cv698_15253.
txt', 'pos/cv699_7223.txt', 'pos/cv700_21947.txt', 'pos/cv701_14252.txt', 'pos/cv702_11500.txt', 'po
s/cv703_16143.txt', 'pos/cv704_15969.txt', 'pos/cv705_11059.txt', 'pos/cv706_24716.txt', 'pos/cv707_1
0678.txt', 'pos/cv708_28729.txt', 'pos/cv709_10529.txt', 'pos/cv710_22577.txt', 'pos/cv711_11665.tx
t', 'pos/cv712_22920.txt', 'pos/cv713_29155.txt', 'pos/cv714_18502.txt', 'pos/cv715_18179.txt', 'pos/
cv716_10514.txt', 'pos/cv717_15953.txt', 'pos/cv718_11434.txt', 'pos/cv719_5713.txt', 'pos/cv720_538
9.txt', 'pos/cv721_29121.txt', 'pos/cv722_7110.txt', 'pos/cv723_8648.txt', 'pos/cv724_13681.txt', 'po
s/cv725_10103.txt', 'pos/cv726_4719.txt', 'pos/cv727_4978.txt', 'pos/cv728_16133.txt', 'pos/cv729_101
54.txt', 'pos/cv730_10279.txt', 'pos/cv731_4136.txt', 'pos/cv732_12245.txt', 'pos/cv733_9839.txt', 'p
os/cv734_21568.txt', 'pos/cv735_18801.txt', 'pos/cv736_23670.txt', 'pos/cv737_28907.txt', 'pos/cv738_
10116.txt', 'pos/cv739_11209.txt', 'pos/cv740_12445.txt', 'pos/cv741_11890.txt', 'pos/cv742_7751.tx
t', 'pos/cv743_15449.txt', 'pos/cv744_10038.txt', 'pos/cv745_12773.txt', 'pos/cv746_10147.txt', 'pos/
cv747_16556.txt', 'pos/cv748_12786.txt', 'pos/cv749_17765.txt', 'pos/cv750_10180.txt', 'pos/cv751_157
19.txt', 'pos/cv752_24155.txt', 'pos/cv753_10875.txt', 'pos/cv754_7216.txt', 'pos/cv755_23616.txt',
'pos/cv756_22540.txt', 'pos/cv757_10189.txt', 'pos/cv758_9671.txt', 'pos/cv759_13522.txt', 'pos/cv760
_8597.txt', 'pos/cv761_12620.txt', 'pos/cv762_13927.txt', 'pos/cv763_14729.txt', 'pos/cv764_11739.tx
t', 'pos/cv765_19037.txt', 'pos/cv766_7540.txt', 'pos/cv767_14062.txt', 'pos/cv768_11751.txt', 'pos/c
v769_8123.txt', 'pos/cv770_10451.txt', 'pos/cv771_28665.txt', 'pos/cv772_12119.txt', 'pos/cv773_1881
7.txt', 'pos/cv774_13845.txt', 'pos/cv775_16237.txt', 'pos/cv776_20529.txt', 'pos/cv777_10094.txt',
'pos/cv778_17330.txt', 'pos/cv779_17881.txt', 'pos/cv780_7984.txt', 'pos/cv781_5262.txt', 'pos/cv782_
19526.txt', 'pos/cv783_13227.txt', 'pos/cv784_14394.txt', 'pos/cv785_22600.txt', 'pos/cv786_22497.tx
t', 'pos/cv787_13743.txt', 'pos/cv788_25272.txt', 'pos/cv789_12136.txt', 'pos/cv790_14600.txt', 'pos/
cv791_16302.txt', 'pos/cv792_3832.txt', 'pos/cv793_13650.txt', 'pos/cv794_15868.txt', 'pos/cv795_1012
2.txt', 'pos/cv796_15782.txt', 'pos/cv797_6957.txt', 'pos/cv798_23531.txt', 'pos/cv799_18543.txt', 'p
```

```
os/cv800_12368.txt', 'pos/cv801_25228.txt', 'pos/cv802_28664.txt', 'pos/cv803_8207.txt', 'pos/cv804_1
0862.txt', 'pos/cv805_19601.txt', 'pos/cv806_8842.txt', 'pos/cv807_21740.txt', 'pos/cv808_12635.txt',
'pos/cv809_5009.txt', 'pos/cv810_12458.txt', 'pos/cv811_21386.txt', 'pos/cv812_17924.txt', 'pos/cv813
_6534.txt', 'pos/cv814_18975.txt', 'pos/cv815_22456.txt', 'pos/cv816_13655.txt', 'pos/cv817_4041.tx
t', 'pos/cv818_10211.txt', 'pos/cv819_9364.txt', 'pos/cv820_22892.txt', 'pos/cv821_29364.txt', 'pos/c
v822_20049.txt', 'pos/cv823_15569.txt', 'pos/cv824_8838.txt', 'pos/cv825_5063.txt', 'pos/cv826_11834.
txt', 'pos/cv827_18331.txt', 'pos/cv828_19831.txt', 'pos/cv829_20289.txt', 'pos/cv830_6014.txt', 'po
s/cv831_14689.txt', 'pos/cv832_23275.txt', 'pos/cv833_11053.txt', 'pos/cv834_22195.txt', 'pos/cv835_1
9159.txt', 'pos/cv836_12968.txt', 'pos/cv837_27325.txt', 'pos/cv838_24728.txt', 'pos/cv839_21467.tx
t', 'pos/cv840_16321.txt', 'pos/cv841_3967.txt', 'pos/cv842_5866.txt', 'pos/cv843_15544.txt', 'pos/cv
844_12690.txt', 'pos/cv845_14290.txt', 'pos/cv846_29497.txt', 'pos/cv847_1941.txt', 'pos/cv848_10036.
txt', 'pos/cv849_15729.txt', 'pos/cv850_16466.txt', 'pos/cv851_20469.txt', 'pos/cv852_27523.txt', 'po
s/cv853_29233.txt', 'pos/cv854_17740.txt', 'pos/cv855_20661.txt', 'pos/cv856_29013.txt', 'pos/cv857_1
5958.txt', 'pos/cv858_18819.txt', 'pos/cv859_14107.txt', 'pos/cv860_13853.txt', 'pos/cv861_1198.txt',
'pos/cv862_14324.txt', 'pos/cv863_7424.txt', 'pos/cv864_3416.txt', 'pos/cv865_2895.txt', 'pos/cv866_2
9691.txt', 'pos/cv867_16661.txt', 'pos/cv868_11948.txt', 'pos/cv869_23611.txt', 'pos/cv870_16348.tx
t', 'pos/cv871_24888.txt', 'pos/cv872_12591.txt', 'pos/cv873_18636.txt', 'pos/cv874_11236.txt', 'pos/
cv875_5754.txt', 'pos/cv876_9390.txt', 'pos/cv877_29274.txt', 'pos/cv878_15694.txt', 'pos/cv879_1490
3.txt', 'pos/cv880_29800.txt', 'pos/cv881_13254.txt', 'pos/cv882_10026.txt', 'pos/cv883_27751.txt',
'pos/cv884_13632.txt', 'pos/cv885_12318.txt', 'pos/cv886_18177.txt', 'pos/cv887_5126.txt', 'pos/cv888
_24435.txt', 'pos/cv889_21430.txt', 'pos/cv890_3977.txt', 'pos/cv891_6385.txt', 'pos/cv892_17576.tx
t', 'pos/cv893_26269.txt', 'pos/cv894_2068.txt', 'pos/cv895_21022.txt', 'pos/cv896_16071.txt', 'pos/c
v897_10837.txt', 'pos/cv898_14187.txt', 'pos/cv899_16014.txt', 'pos/cv900_10331.txt', 'pos/cv901_1101
7.txt', 'pos/cv902_12256.txt', 'pos/cv903_17822.txt', 'pos/cv904_24353.txt', 'pos/cv905_29114.txt',
'pos/cv906_11491.txt', 'pos/cv907_3541.txt', 'pos/cv908_16009.txt', 'pos/cv909_9960.txt', 'pos/cv910_
20488.txt', 'pos/cv911_20260.txt', 'pos/cv912_5674.txt', 'pos/cv913_29252.txt', 'pos/cv914_28742.tx
t', 'pos/cv915_8841.txt', 'pos/cv916_15467.txt', 'pos/cv917_29715.txt', 'pos/cv918_2693.txt', 'pos/cv
919_16380.txt', 'pos/cv920_29622.txt', 'pos/cv921_12747.txt', 'pos/cv922_10073.txt', 'pos/cv923_1105
1.txt', 'pos/cv924_29540.txt', 'pos/cv925_8969.txt', 'pos/cv926_17059.txt', 'pos/cv927_10681.txt', 'p
os/cv928_9168.txt', 'pos/cv929_16908.txt', 'pos/cv930_13475.txt', 'pos/cv931_17563.txt', 'pos/cv932_1
3401.txt', 'pos/cv933_23776.txt', 'pos/cv934_19027.txt', 'pos/cv935_23841.txt', 'pos/cv936_15954.tx
t', 'pos/cv937_9811.txt', 'pos/cv938_10220.txt', 'pos/cv939_10583.txt', 'pos/cv940_17705.txt', 'pos/c
v941_10246.txt', 'pos/cv942_17082.txt', 'pos/cv943_22488.txt', 'pos/cv944_13521.txt', 'pos/cv945_1216
0.txt', 'pos/cv946_18658.txt', 'pos/cv947_10601.txt', 'pos/cv948_24606.txt', 'pos/cv949_20112.txt',
'pos/cv950_12350.txt', 'pos/cv951_10926.txt', 'pos/cv952_25240.txt', 'pos/cv953_6836.txt', 'pos/cv954
_18628.txt', 'pos/cv955_25001.txt', 'pos/cv956_11609.txt', 'pos/cv957_8737.txt', 'pos/cv958_12162.tx
t', 'pos/cv959_14611.txt', 'pos/cv960_29007.txt', 'pos/cv961_5682.txt', 'pos/cv962_9803.txt', 'pos/cv
963_6895.txt', 'pos/cv964_6021.txt', 'pos/cv965_26071.txt', 'pos/cv966_28832.txt', 'pos/cv967_5788.tx
t', 'pos/cv968_24218.txt', 'pos/cv969_13250.txt', 'pos/cv970_18450.txt', 'pos/cv971_10874.txt', 'pos/
cv972_26417.txt', 'pos/cv973_10066.txt', 'pos/cv974_22941.txt', 'pos/cv975_10981.txt', 'pos/cv976_102
67.txt', 'pos/cv977_4938.txt', 'pos/cv978_20929.txt', 'pos/cv979_18921.txt', 'pos/cv980_10953.txt',
'pos/cv981_14989.txt', 'pos/cv982_21103.txt', 'pos/cv983_22928.txt', 'pos/cv984_12767.txt', 'pos/cv98
5_6359.txt', 'pos/cv986_13527.txt', 'pos/cv987_6965.txt', 'pos/cv988_18740.txt', 'pos/cv989_15824.tx
t', 'pos/cv990_11591.txt', 'pos/cv991_18645.txt', 'pos/cv992_11962.txt', 'pos/cv993_29737.txt', 'pos/
cv994_12270.txt', 'pos/cv995_21821.txt', 'pos/cv996_11592.txt', 'pos/cv997_5046.txt', 'pos/cv998_1411
1.txt', 'pos/cv999_13106.txt']
```

In [41]:

```python
print(len(movie_reviews.fileids('neg')))
print(' ')
print(movie_reviews.fileids('neg'))
```

```
1000

['neg/cv000_29416.txt', 'neg/cv001_19502.txt', 'neg/cv002_17424.txt', 'neg/cv003_12683.txt', 'neg/cv0
04_12641.txt', 'neg/cv005_29357.txt', 'neg/cv006_17022.txt', 'neg/cv007_4992.txt', 'neg/cv008_29326.t
xt', 'neg/cv009_29417.txt', 'neg/cv010_29063.txt', 'neg/cv011_13044.txt', 'neg/cv012_29411.txt', 'ne
g/cv013_10494.txt', 'neg/cv014_15600.txt', 'neg/cv015_29356.txt', 'neg/cv016_4348.txt', 'neg/cv017_23
487.txt', 'neg/cv018_21672.txt', 'neg/cv019_16117.txt', 'neg/cv020_9234.txt', 'neg/cv021_17313.txt',
'neg/cv022_14227.txt', 'neg/cv023_13847.txt', 'neg/cv024_7033.txt', 'neg/cv025_29825.txt', 'neg/cv026
_29229.txt', 'neg/cv027_26270.txt', 'neg/cv028_26964.txt', 'neg/cv029_19943.txt', 'neg/cv030_22893.tx
t', 'neg/cv031_19540.txt', 'neg/cv032_23718.txt', 'neg/cv033_25680.txt', 'neg/cv034_29446.txt', 'neg/
cv035_3343.txt', 'neg/cv036_18385.txt', 'neg/cv037_19798.txt', 'neg/cv038_9781.txt', 'neg/cv039_5963.
txt', 'neg/cv040_8829.txt', 'neg/cv041_22364.txt', 'neg/cv042_11927.txt', 'neg/cv043_16808.txt', 'ne
g/cv044_18429.txt', 'neg/cv045_25077.txt', 'neg/cv046_10613.txt', 'neg/cv047_18725.txt', 'neg/cv048_1
8380.txt', 'neg/cv049_21917.txt', 'neg/cv050_12128.txt', 'neg/cv051_10751.txt', 'neg/cv052_29318.tx
t', 'neg/cv053_23117.txt', 'neg/cv054_4101.txt', 'neg/cv055_8926.txt', 'neg/cv056_14663.txt', 'neg/cv
057_7962.txt', 'neg/cv058_8469.txt', 'neg/cv059_28723.txt', 'neg/cv060_11754.txt', 'neg/cv061_9321.tx
t', 'neg/cv062_24556.txt', 'neg/cv063_28852.txt', 'neg/cv064_25842.txt', 'neg/cv065_16909.txt', 'neg/
cv066_11668.txt', 'neg/cv067_21192.txt', 'neg/cv068_14810.txt', 'neg/cv069_11613.txt', 'neg/cv070_132
49.txt', 'neg/cv071_12969.txt', 'neg/cv072_5928.txt', 'neg/cv073_23039.txt', 'neg/cv074_7188.txt', 'n
eg/cv075_6250.txt', 'neg/cv076_26009.txt', 'neg/cv077_23172.txt', 'neg/cv078_16506.txt', 'neg/cv079_1
2766.txt', 'neg/cv080_14899.txt', 'neg/cv081_18241.txt', 'neg/cv082_11979.txt', 'neg/cv083_25491.tx
t', 'neg/cv084_15183.txt', 'neg/cv085_15286.txt', 'neg/cv086_19488.txt', 'neg/cv087_2145.txt', 'neg/c
v088_25274.txt', 'neg/cv089_12222.txt', 'neg/cv090_0049.txt', 'neg/cv091_7899.txt', 'neg/cv092_27987.
txt', 'neg/cv093_15606.txt', 'neg/cv094_27868.txt', 'neg/cv095_28730.txt', 'neg/cv096_12262.txt', 'ne
g/cv097_26081.txt', 'neg/cv098_17021.txt', 'neg/cv099_11189.txt', 'neg/cv100_12406.txt', 'neg/cv101_1
0537.txt', 'neg/cv102_8306.txt', 'neg/cv103_11943.txt', 'neg/cv104_19176.txt', 'neg/cv105_19135.txt',
'neg/cv106_18379.txt', 'neg/cv107_25639.txt', 'neg/cv108_17064.txt', 'neg/cv109_22599.txt', 'neg/cv11
0_27832.txt', 'neg/cv111_12253.txt', 'neg/cv112_12178.txt', 'neg/cv113_24354.txt', 'neg/cv114_19501.t
xt', 'neg/cv115_26443.txt', 'neg/cv116_28734.txt', 'neg/cv117_25625.txt', 'neg/cv118_28837.txt', 'ne
g/cv119_9909.txt', 'neg/cv120_3793.txt', 'neg/cv121_18621.txt', 'neg/cv122_7891.txt', 'neg/cv123_1216
5.txt', 'neg/cv124_3903.txt', 'neg/cv125_9636.txt', 'neg/cv126_28821.txt', 'neg/cv127_16451.txt', 'ne
g/cv128_29444.txt', 'neg/cv129_18373.txt', 'neg/cv130_18521.txt', 'neg/cv131_11568.txt', 'neg/cv132_5
423.txt', 'neg/cv133_18065.txt', 'neg/cv134_23300.txt', 'neg/cv135_12506.txt', 'neg/cv136_12384.txt',
'neg/cv137_17020.txt', 'neg/cv138_13903.txt', 'neg/cv139_14236.txt', 'neg/cv140_7963.txt', 'neg/cv141
_17179.txt', 'neg/cv142_23657.txt', 'neg/cv143_21158.txt', 'neg/cv144_5010.txt', 'neg/cv145_12239.tx
t', 'neg/cv146_19587.txt', 'neg/cv147_22625.txt', 'neg/cv148_18084.txt', 'neg/cv149_17084.txt', 'neg/
cv150_14279.txt', 'neg/cv151_17231.txt', 'neg/cv152_9052.txt', 'neg/cv153_11607.txt', 'neg/cv154_956
2.txt', 'neg/cv155_7845.txt', 'neg/cv156_11119.txt', 'neg/cv157_29302.txt', 'neg/cv158_10914.txt', 'n
eg/cv159_29374.txt', 'neg/cv160_10848.txt', 'neg/cv161_12224.txt', 'neg/cv162_10977.txt', 'neg/cv163_
10110.txt', 'neg/cv164_23451.txt', 'neg/cv165_2389.txt', 'neg/cv166_11959.txt', 'neg/cv167_18094.tx
t', 'neg/cv168_7435.txt', 'neg/cv169_24973.txt', 'neg/cv170_29808.txt', 'neg/cv171_15164.txt', 'neg/c
v172_12037.txt', 'neg/cv173_4295.txt', 'neg/cv174_9735.txt', 'neg/cv175_7375.txt', 'neg/cv176_14196.t
xt', 'neg/cv177_10904.txt', 'neg/cv178_14380.txt', 'neg/cv179_9533.txt', 'neg/cv180_17823.txt', 'neg/
cv181_16083.txt', 'neg/cv182_7791.txt', 'neg/cv183_19826.txt', 'neg/cv184_26935.txt', 'neg/cv185_2837
2.txt', 'neg/cv186_2396.txt', 'neg/cv187_14112.txt', 'neg/cv188_20687.txt', 'neg/cv189_24248.txt', 'n
eg/cv190_27176.txt', 'neg/cv191_29539.txt', 'neg/cv192_16079.txt', 'neg/cv193_5393.txt', 'neg/cv194_1
2855.txt', 'neg/cv195_16146.txt', 'neg/cv196_28898.txt', 'neg/cv197_29271.txt', 'neg/cv198_19313.tx
t', 'neg/cv199_9721.txt', 'neg/cv200_29006.txt', 'neg/cv201_7421.txt', 'neg/cv202_11382.txt', 'neg/cv
203_19052.txt', 'neg/cv204_8930.txt', 'neg/cv205_9676.txt', 'neg/cv206_15893.txt', 'neg/cv207_29141.t
xt', 'neg/cv208_9475.txt', 'neg/cv209_28973.txt', 'neg/cv210_9557.txt', 'neg/cv211_9955.txt', 'neg/cv
212_10054.txt', 'neg/cv213_20300.txt', 'neg/cv214_13285.txt', 'neg/cv215_23246.txt', 'neg/cv216_2016
5.txt', 'neg/cv217_28707.txt', 'neg/cv218_25651.txt', 'neg/cv219_19874.txt', 'neg/cv220_28906.txt',
'neg/cv221_27081.txt', 'neg/cv222_18720.txt', 'neg/cv223_28923.txt', 'neg/cv224_18875.txt', 'neg/cv22
5_29083.txt', 'neg/cv226_26692.txt', 'neg/cv227_25406.txt', 'neg/cv228_5644.txt', 'neg/cv229_15200.tx
t', 'neg/cv230_7913.txt', 'neg/cv231_11028.txt', 'neg/cv232_16768.txt', 'neg/cv233_17614.txt', 'neg/c
v234_22123.txt', 'neg/cv235_10704.txt', 'neg/cv236_12427.txt', 'neg/cv237_20635.txt', 'neg/cv238_1428
5.txt', 'neg/cv239_29828.txt', 'neg/cv240_15948.txt', 'neg/cv241_24602.txt', 'neg/cv242_11354.txt',
'neg/cv243_22164.txt', 'neg/cv244_22935.txt', 'neg/cv245_8938.txt', 'neg/cv246_28668.txt', 'neg/cv247
_14668.txt', 'neg/cv248_15672.txt', 'neg/cv249_12674.txt', 'neg/cv250_26462.txt', 'neg/cv251_23901.tx
t', 'neg/cv252_24974.txt', 'neg/cv253_10190.txt', 'neg/cv254_5870.txt', 'neg/cv255_15267.txt', 'neg/c
v256_16529.txt', 'neg/cv257_11856.txt', 'neg/cv258_5627.txt', 'neg/cv259_11827.txt', 'neg/cv260_1565
2.txt', 'neg/cv261_11855.txt', 'neg/cv262_13812.txt', 'neg/cv263_20693.txt', 'neg/cv264_14108.txt',
'neg/cv265_11625.txt', 'neg/cv266_26644.txt', 'neg/cv267_16618.txt', 'neg/cv268_20288.txt', 'neg/cv26
9_23018.txt', 'neg/cv270_5873.txt', 'neg/cv271_15364.txt', 'neg/cv272_20313.txt', 'neg/cv273_28961.tx
t', 'neg/cv274_26379.txt', 'neg/cv275_28725.txt', 'neg/cv276_17126.txt', 'neg/cv277_20467.txt', 'neg/
cv278_14533.txt', 'neg/cv279_19452.txt', 'neg/cv280_8651.txt', 'neg/cv281_24711.txt', 'neg/cv282_683
3.txt', 'neg/cv283_11963.txt', 'neg/cv284_20530.txt', 'neg/cv285_18186.txt', 'neg/cv286_26156.txt',
'neg/cv287_17410.txt', 'neg/cv288_20212.txt', 'neg/cv289_6239.txt', 'neg/cv290_11981.txt', 'neg/cv291
_26844.txt', 'neg/cv292_7804.txt', 'neg/cv293_29731.txt', 'neg/cv294_12695.txt', 'neg/cv295_17060.tx
t', 'neg/cv296_13146.txt', 'neg/cv297_10104.txt', 'neg/cv298_24487.txt', 'neg/cv299_17950.txt', 'neg/
cv300_23302.txt', 'neg/cv301_13010.txt', 'neg/cv302_26481.txt', 'neg/cv303_27366.txt', 'neg/cv304_284
89.txt', 'neg/cv305_9937.txt', 'neg/cv306_10859.txt', 'neg/cv307_26382.txt', 'neg/cv308_5079.txt', 'n
eg/cv309_23737.txt', 'neg/cv310_14568.txt', 'neg/cv311_17708.txt', 'neg/cv312_29308.txt', 'neg/cv313_
19337.txt', 'neg/cv314_16095.txt', 'neg/cv315_12638.txt', 'neg/cv316_5972.txt', 'neg/cv317_25111.tx
t', 'neg/cv318_11146.txt', 'neg/cv319_16459.txt', 'neg/cv320_9693.txt', 'neg/cv321_14191.txt', 'neg/c
v322_21820.txt', 'neg/cv323_29633.txt', 'neg/cv324_7502.txt', 'neg/cv325_18330.txt', 'neg/cv326_1477
7.txt', 'neg/cv327_21743.txt', 'neg/cv328_10908.txt', 'neg/cv329_29293.txt', 'neg/cv330_29675.txt',
'neg/cv331_8656.txt', 'neg/cv332_17997.txt', 'neg/cv333_9443.txt', 'neg/cv334_0074.txt', 'neg/cv335_1
6299.txt', 'neg/cv336_10363.txt', 'neg/cv337_29061.txt', 'neg/cv338_9183.txt', 'neg/cv339_22452.txt',
'neg/cv340_14776.txt', 'neg/cv341_25667.txt', 'neg/cv342_20917.txt', 'neg/cv343_10906.txt', 'neg/cv34
4_5376.txt', 'neg/cv345_9966.txt', 'neg/cv346_19198.txt', 'neg/cv347_14722.txt', 'neg/cv348_19207.tx
t', 'neg/cv349_15032.txt', 'neg/cv350_22139.txt', 'neg/cv351_17029.txt', 'neg/cv352_5414.txt', 'neg/c
v353_19197.txt', 'neg/cv354_8573.txt', 'neg/cv355_18174.txt', 'neg/cv356_26170.txt', 'neg/cv357_1471
0.txt', 'neg/cv358_11557.txt', 'neg/cv359_6751.txt', 'neg/cv360_8927.txt', 'neg/cv361_28738.txt', 'ne
g/cv362_16985.txt', 'neg/cv363_29273.txt', 'neg/cv364_14254.txt', 'neg/cv365_12442.txt', 'neg/cv366_1
0709.txt', 'neg/cv367_24065.txt', 'neg/cv368_11090.txt', 'neg/cv369_14245.txt', 'neg/cv370_5338.txt',
'neg/cv371_8197.txt', 'neg/cv372_6654.txt', 'neg/cv373_21872.txt', 'neg/cv374_26455.txt', 'neg/cv375_
9932.txt', 'neg/cv376_20883.txt', 'neg/cv377_8440.txt', 'neg/cv378_21982.txt', 'neg/cv379_23167.txt',
'neg/cv380_8164.txt', 'neg/cv381_21673.txt', 'neg/cv382_8393.txt', 'neg/cv383_14662.txt', 'neg/cv384_
18536.txt', 'neg/cv385_29621.txt', 'neg/cv386_10229.txt', 'neg/cv387_12391.txt', 'neg/cv388_12810.tx
t', 'neg/cv389_9611.txt', 'neg/cv390_12187.txt', 'neg/cv391_11615.txt', 'neg/cv392_12238.txt', 'neg/c
v393_29234.txt', 'neg/cv394_5311.txt', 'neg/cv395_11761.txt', 'neg/cv396_19127.txt', 'neg/cv397_2889
0.txt', 'neg/cv398_17047.txt', 'neg/cv399_28593.txt', 'neg/cv400_20631.txt', 'neg/cv401_13758.txt',
'neg/cv402_16097.txt', 'neg/cv403_6721.txt', 'neg/cv404_21805.txt', 'neg/cv405_21868.txt', 'neg/cv406
_22199.txt', 'neg/cv407_23928.txt', 'neg/cv408_5367.txt', 'neg/cv409_29625.txt', 'neg/cv410_25624.tx
t', 'neg/cv411_16799.txt', 'neg/cv412_25254.txt', 'neg/cv413_7893.txt', 'neg/cv414_11161.txt', 'neg/c
v415_23674.txt', 'neg/cv416_12048.txt', 'neg/cv417_14653.txt', 'neg/cv418_16562.txt', 'neg/cv419_1479
9.txt', 'neg/cv420_28631.txt', 'neg/cv421_9752.txt', 'neg/cv422_9632.txt', 'neg/cv423_12089.txt', 'ne
g/cv424_9268.txt', 'neg/cv425_8603.txt', 'neg/cv426_10976.txt', 'neg/cv427_11693.txt', 'neg/cv428_122
02.txt', 'neg/cv429_7937.txt', 'neg/cv430_18662.txt', 'neg/cv431_7538.txt', 'neg/cv432_15873.txt', 'n
eg/cv433_10443.txt', 'neg/cv434_5641.txt', 'neg/cv435_24355.txt', 'neg/cv436_20564.txt', 'neg/cv437_2
4070.txt', 'neg/cv438_8500.txt', 'neg/cv439_17633.txt', 'neg/cv440_16891.txt', 'neg/cv441_15276.txt',
'neg/cv442_15499.txt', 'neg/cv443_22367.txt', 'neg/cv444_9975.txt', 'neg/cv445_26683.txt', 'neg/cv446
_12209.txt', 'neg/cv447_27334.txt', 'neg/cv448_16409.txt', 'neg/cv449_9126.txt', 'neg/cv450_8319.tx
t', 'neg/cv451_11502.txt', 'neg/cv452_5179.txt', 'neg/cv453_10911.txt', 'neg/cv454_21961.txt', 'neg/c
v455_28866.txt', 'neg/cv456_20370.txt', 'neg/cv457_19546.txt', 'neg/cv458_9000.txt', 'neg/cv459_2183
4.txt', 'neg/cv460_11723.txt', 'neg/cv461_21124.txt', 'neg/cv462_20788.txt', 'neg/cv463_10846.txt',
'neg/cv464_17076.txt', 'neg/cv465_23401.txt', 'neg/cv466_20092.txt', 'neg/cv467_26610.txt', 'neg/cv46
8_16844.txt', 'neg/cv469_21998.txt', 'neg/cv470_17444.txt', 'neg/cv471_18405.txt', 'neg/cv472_29140.t
xt', 'neg/cv473_7869.txt', 'neg/cv474_10682.txt', 'neg/cv475_22978.txt', 'neg/cv476_18402.txt', 'neg/
cv477_23530.txt', 'neg/cv478_15921.txt', 'neg/cv479_5450.txt', 'neg/cv480_21195.txt', 'neg/cv481_793
0.txt', 'neg/cv482_11233.txt', 'neg/cv483_18103.txt', 'neg/cv484_26169.txt', 'neg/cv485_26879.txt',
'neg/cv486_9788.txt', 'neg/cv487_11058.txt', 'neg/cv488_21453.txt', 'neg/cv489_19046.txt', 'neg/cv490
_18986.txt', 'neg/cv491_12992.txt', 'neg/cv492_19370.txt', 'neg/cv493_14135.txt', 'neg/cv494_18689.tx
t', 'neg/cv495_16121.txt', 'neg/cv496_11185.txt', 'neg/cv497_27086.txt', 'neg/cv498_9288.txt', 'neg/c
v499_11407.txt', 'neg/cv500_10722.txt', 'neg/cv501_12675.txt', 'neg/cv502_10970.txt', 'neg/cv503_1119
6.txt', 'neg/cv504_29120.txt', 'neg/cv505_12926.txt', 'neg/cv506_17521.txt', 'neg/cv507_9509.txt', 'n
eg/cv508_17742.txt', 'neg/cv509_17354.txt', 'neg/cv510_24758.txt', 'neg/cv511_10360.txt', 'neg/cv512_
17618.txt', 'neg/cv513_7236.txt', 'neg/cv514_12173.txt', 'neg/cv515_18484.txt', 'neg/cv516_12117.tx
t', 'neg/cv517_20616.txt', 'neg/cv518_14798.txt', 'neg/cv519_16239.txt', 'neg/cv520_13297.txt', 'neg/
cv521_1730.txt', 'neg/cv522_5418.txt', 'neg/cv523_18285.txt', 'neg/cv524_24885.txt', 'neg/cv525_1793
0.txt', 'neg/cv526_12868.txt', 'neg/cv527_10338.txt', 'neg/cv528_11669.txt', 'neg/cv529_10972.txt',
'neg/cv530_17949.txt', 'neg/cv531_26838.txt', 'neg/cv532_6495.txt', 'neg/cv533_9843.txt', 'neg/cv534_
15683.txt', 'neg/cv535_21183.txt', 'neg/cv536_27221.txt', 'neg/cv537_13516.txt', 'neg/cv538_28485.tx
t', 'neg/cv539_21865.txt', 'neg/cv540_3092.txt', 'neg/cv541_28683.txt', 'neg/cv542_20359.txt', 'neg/c
v543_5107.txt', 'neg/cv544_5301.txt', 'neg/cv545_12848.txt', 'neg/cv546_12723.txt', 'neg/cv547_18043.
txt', 'neg/cv548_18944.txt', 'neg/cv549_22771.txt', 'neg/cv550_23226.txt', 'neg/cv551_11214.txt', 'ne
g/cv552_0150.txt', 'neg/cv553_26965.txt', 'neg/cv554_14678.txt', 'neg/cv555_25047.txt', 'neg/cv556_16
563.txt', 'neg/cv557_12237.txt', 'neg/cv558_29376.txt', 'neg/cv559_0057.txt', 'neg/cv560_18608.txt',
'neg/cv561_9484.txt', 'neg/cv562_10847.txt', 'neg/cv563_18610.txt', 'neg/cv564_12011.txt', 'neg/cv565
_29403.txt', 'neg/cv566_8967.txt', 'neg/cv567_29420.txt', 'neg/cv568_17065.txt', 'neg/cv569_26750.tx
t', 'neg/cv570_28960.txt', 'neg/cv571_29292.txt', 'neg/cv572_20053.txt', 'neg/cv573_29384.txt', 'neg/
cv574_23191.txt', 'neg/cv575_22598.txt', 'neg/cv576_15688.txt', 'neg/cv577_28220.txt', 'neg/cv578_168
25.txt', 'neg/cv579_12542.txt', 'neg/cv580_15681.txt', 'neg/cv581_20790.txt', 'neg/cv582_6678.txt',
'neg/cv583_29465.txt', 'neg/cv584_29549.txt', 'neg/cv585_23576.txt', 'neg/cv586_8048.txt', 'neg/cv587
_20532.txt', 'neg/cv588_14467.txt', 'neg/cv589_12853.txt', 'neg/cv590_20712.txt', 'neg/cv591_24887.tx
t', 'neg/cv592_23391.txt', 'neg/cv593_11931.txt', 'neg/cv594_11945.txt', 'neg/cv595_26420.txt', 'neg/
cv596_4367.txt', 'neg/cv597_26744.txt', 'neg/cv598_18184.txt', 'neg/cv599_22197.txt', 'neg/cv600_2504
3.txt', 'neg/cv601_24759.txt', 'neg/cv602_8830.txt', 'neg/cv603_18885.txt', 'neg/cv604_23339.txt', 'n
eg/cv605_12730.txt', 'neg/cv606_17672.txt', 'neg/cv607_8235.txt', 'neg/cv608_24647.txt', 'neg/cv609_2
5038.txt', 'neg/cv610_24153.txt', 'neg/cv611_2253.txt', 'neg/cv612_5396.txt', 'neg/cv613_23104.txt',
'neg/cv614_11320.txt', 'neg/cv615_15734.txt', 'neg/cv616_29187.txt', 'neg/cv617_9561.txt', 'neg/cv618
_9469.txt', 'neg/cv619_13677.txt', 'neg/cv620_2556.txt', 'neg/cv621_15984.txt', 'neg/cv622_8583.txt',
'neg/cv623_16988.txt', 'neg/cv624_11601.txt', 'neg/cv625_13518.txt', 'neg/cv626_7907.txt', 'neg/cv627
_12603.txt', 'neg/cv628_20758.txt', 'neg/cv629_16604.txt', 'neg/cv630_10152.txt', 'neg/cv631_4782.tx
t', 'neg/cv632_9704.txt', 'neg/cv633_29730.txt', 'neg/cv634_11989.txt', 'neg/cv635_0984.txt', 'neg/cv
636_16954.txt', 'neg/cv637_13682.txt', 'neg/cv638_29394.txt', 'neg/cv639_10797.txt', 'neg/cv640_5380.
```

```
txt', 'neg/cv641_13412.txt', 'neg/cv642_29788.txt', 'neg/cv643_29282.txt', 'neg/cv644_18551.txt', 'ne
g/cv645_17078.txt', 'neg/cv646_16817.txt', 'neg/cv647_15275.txt', 'neg/cv648_17277.txt', 'neg/cv649_1
3947.txt', 'neg/cv650_15974.txt', 'neg/cv651_11120.txt', 'neg/cv652_15653.txt', 'neg/cv653_2107.txt',
'neg/cv654_19345.txt', 'neg/cv655_12055.txt', 'neg/cv656_25395.txt', 'neg/cv657_25835.txt', 'neg/cv65
8_11186.txt', 'neg/cv659_21483.txt', 'neg/cv660_23140.txt', 'neg/cv661_25780.txt', 'neg/cv662_14791.t
xt', 'neg/cv663_14484.txt', 'neg/cv664_4264.txt', 'neg/cv665_29386.txt', 'neg/cv666_20301.txt', 'neg/
cv667_19672.txt', 'neg/cv668_18848.txt', 'neg/cv669_24318.txt', 'neg/cv670_2666.txt', 'neg/cv671_516
4.txt', 'neg/cv672_27988.txt', 'neg/cv673_25874.txt', 'neg/cv674_11593.txt', 'neg/cv675_22871.txt',
'neg/cv676_22202.txt', 'neg/cv677_18938.txt', 'neg/cv678_14887.txt', 'neg/cv679_28221.txt', 'neg/cv68
0_10533.txt', 'neg/cv681_9744.txt', 'neg/cv682_17947.txt', 'neg/cv683_13047.txt', 'neg/cv684_12727.tx
t', 'neg/cv685_5710.txt', 'neg/cv686_15553.txt', 'neg/cv687_22207.txt', 'neg/cv688_7884.txt', 'neg/cv
689_13701.txt', 'neg/cv690_5425.txt', 'neg/cv691_5090.txt', 'neg/cv692_17026.txt', 'neg/cv693_19147.t
xt', 'neg/cv694_4526.txt', 'neg/cv695_22268.txt', 'neg/cv696_29619.txt', 'neg/cv697_12106.txt', 'neg/
cv698_16930.txt', 'neg/cv699_7773.txt', 'neg/cv700_23163.txt', 'neg/cv701_15880.txt', 'neg/cv702_1237
1.txt', 'neg/cv703_17948.txt', 'neg/cv704_17622.txt', 'neg/cv705_11973.txt', 'neg/cv706_25883.txt',
'neg/cv707_11421.txt', 'neg/cv708_28539.txt', 'neg/cv709_11173.txt', 'neg/cv710_23745.txt', 'neg/cv71
1_12687.txt', 'neg/cv712_24217.txt', 'neg/cv713_29002.txt', 'neg/cv714_19704.txt', 'neg/cv715_19246.t
xt', 'neg/cv716_11153.txt', 'neg/cv717_17472.txt', 'neg/cv718_12227.txt', 'neg/cv719_5581.txt', 'neg/
cv720_5383.txt', 'neg/cv721_28993.txt', 'neg/cv722_7571.txt', 'neg/cv723_9002.txt', 'neg/cv724_15265.
txt', 'neg/cv725_10266.txt', 'neg/cv726_4365.txt', 'neg/cv727_5006.txt', 'neg/cv728_17931.txt', 'neg/
cv729_10475.txt', 'neg/cv730_10729.txt', 'neg/cv731_3968.txt', 'neg/cv732_13092.txt', 'neg/cv733_989
1.txt', 'neg/cv734_22821.txt', 'neg/cv735_20218.txt', 'neg/cv736_24947.txt', 'neg/cv737_28733.txt',
'neg/cv738_10287.txt', 'neg/cv739_12179.txt', 'neg/cv740_13643.txt', 'neg/cv741_12765.txt', 'neg/cv74
2_8279.txt', 'neg/cv743_17023.txt', 'neg/cv744_10091.txt', 'neg/cv745_14009.txt', 'neg/cv746_10471.tx
t', 'neg/cv747_18189.txt', 'neg/cv748_14044.txt', 'neg/cv749_18960.txt', 'neg/cv750_10606.txt', 'neg/
cv751_17208.txt', 'neg/cv752_25330.txt', 'neg/cv753_11812.txt', 'neg/cv754_7709.txt', 'neg/cv755_2488
1.txt', 'neg/cv756_23676.txt', 'neg/cv757_10668.txt', 'neg/cv758_9740.txt', 'neg/cv759_15091.txt', 'n
eg/cv760_8977.txt', 'neg/cv761_13769.txt', 'neg/cv762_15604.txt', 'neg/cv763_16486.txt', 'neg/cv764_1
2701.txt', 'neg/cv765_20429.txt', 'neg/cv766_7983.txt', 'neg/cv767_15673.txt', 'neg/cv768_12709.txt',
'neg/cv769_8565.txt', 'neg/cv770_11061.txt', 'neg/cv771_28466.txt', 'neg/cv772_12971.txt', 'neg/cv773
_20264.txt', 'neg/cv774_15488.txt', 'neg/cv775_17966.txt', 'neg/cv776_21934.txt', 'neg/cv777_10247.tx
t', 'neg/cv778_18629.txt', 'neg/cv779_18989.txt', 'neg/cv780_8467.txt', 'neg/cv781_5358.txt', 'neg/cv
782_21078.txt', 'neg/cv783_14724.txt', 'neg/cv784_16077.txt', 'neg/cv785_23748.txt', 'neg/cv786_2360
8.txt', 'neg/cv787_15277.txt', 'neg/cv788_26409.txt', 'neg/cv789_12991.txt', 'neg/cv790_16202.txt',
'neg/cv791_17995.txt', 'neg/cv792_3257.txt', 'neg/cv793_15235.txt', 'neg/cv794_17353.txt', 'neg/cv795
_10291.txt', 'neg/cv796_17243.txt', 'neg/cv797_7245.txt', 'neg/cv798_24779.txt', 'neg/cv799_19812.tx
t', 'neg/cv800_13494.txt', 'neg/cv801_26335.txt', 'neg/cv802_28381.txt', 'neg/cv803_8584.txt', 'neg/c
v804_11763.txt', 'neg/cv805_21128.txt', 'neg/cv806_9405.txt', 'neg/cv807_23024.txt', 'neg/cv808_1377
3.txt', 'neg/cv809_5012.txt', 'neg/cv810_13660.txt', 'neg/cv811_22646.txt', 'neg/cv812_19051.txt', 'n
eg/cv813_6649.txt', 'neg/cv814_20316.txt', 'neg/cv815_23466.txt', 'neg/cv816_15257.txt', 'neg/cv817_3
675.txt', 'neg/cv818_10698.txt', 'neg/cv819_9567.txt', 'neg/cv820_24157.txt', 'neg/cv821_29283.txt',
'neg/cv822_21545.txt', 'neg/cv823_17055.txt', 'neg/cv824_9335.txt', 'neg/cv825_5168.txt', 'neg/cv826_
12761.txt', 'neg/cv827_19479.txt', 'neg/cv828_21392.txt', 'neg/cv829_21725.txt', 'neg/cv830_5778.tx
t', 'neg/cv831_16325.txt', 'neg/cv832_24713.txt', 'neg/cv833_11961.txt', 'neg/cv834_23192.txt', 'neg/
cv835_20531.txt', 'neg/cv836_14311.txt', 'neg/cv837_27232.txt', 'neg/cv838_25886.txt', 'neg/cv839_228
07.txt', 'neg/cv840_18033.txt', 'neg/cv841_3367.txt', 'neg/cv842_5702.txt', 'neg/cv843_17054.txt', 'n
eg/cv844_13890.txt', 'neg/cv845_15886.txt', 'neg/cv846_29359.txt', 'neg/cv847_20855.txt', 'neg/cv848_
10061.txt', 'neg/cv849_17215.txt', 'neg/cv850_18185.txt', 'neg/cv851_21895.txt', 'neg/cv852_27512.tx
t', 'neg/cv853_29119.txt', 'neg/cv854_18955.txt', 'neg/cv855_22134.txt', 'neg/cv856_28882.txt', 'neg/
cv857_17527.txt', 'neg/cv858_20266.txt', 'neg/cv859_15689.txt', 'neg/cv860_15520.txt', 'neg/cv861_128
09.txt', 'neg/cv862_15924.txt', 'neg/cv863_7912.txt', 'neg/cv864_3087.txt', 'neg/cv865_28796.txt', 'n
eg/cv866_29447.txt', 'neg/cv867_18362.txt', 'neg/cv868_12799.txt', 'neg/cv869_24782.txt', 'neg/cv870_
18090.txt', 'neg/cv871_25971.txt', 'neg/cv872_13710.txt', 'neg/cv873_19937.txt', 'neg/cv874_12182.tx
t', 'neg/cv875_5622.txt', 'neg/cv876_9633.txt', 'neg/cv877_29132.txt', 'neg/cv878_17204.txt', 'neg/cv
879_16585.txt', 'neg/cv880_29629.txt', 'neg/cv881_14767.txt', 'neg/cv882_10042.txt', 'neg/cv883_2762
1.txt', 'neg/cv884_15230.txt', 'neg/cv885_13390.txt', 'neg/cv886_19210.txt', 'neg/cv887_5306.txt', 'n
eg/cv888_25678.txt', 'neg/cv889_22670.txt', 'neg/cv890_3515.txt', 'neg/cv891_6035.txt', 'neg/cv892_18
788.txt', 'neg/cv893_26731.txt', 'neg/cv894_22140.txt', 'neg/cv895_22200.txt', 'neg/cv896_17819.txt',
'neg/cv897_11703.txt', 'neg/cv898_1576.txt', 'neg/cv899_17812.txt', 'neg/cv900_10800.txt', 'neg/cv901
_11934.txt', 'neg/cv902_13217.txt', 'neg/cv903_18981.txt', 'neg/cv904_25663.txt', 'neg/cv905_28965.tx
t', 'neg/cv906_12332.txt', 'neg/cv907_3193.txt', 'neg/cv908_17779.txt', 'neg/cv909_9973.txt', 'neg/cv
910_21930.txt', 'neg/cv911_21695.txt', 'neg/cv912_5562.txt', 'neg/cv913_29127.txt', 'neg/cv914_2856.t
xt', 'neg/cv915_9342.txt', 'neg/cv916_17034.txt', 'neg/cv917_29484.txt', 'neg/cv918_27080.txt', 'neg/
cv919_18155.txt', 'neg/cv920_29423.txt', 'neg/cv921_13988.txt', 'neg/cv922_10185.txt', 'neg/cv923_119
51.txt', 'neg/cv924_29397.txt', 'neg/cv925_9459.txt', 'neg/cv926_18471.txt', 'neg/cv927_11471.txt',
'neg/cv928_9478.txt', 'neg/cv929_1841.txt', 'neg/cv930_14949.txt', 'neg/cv931_18783.txt', 'neg/cv932_
14854.txt', 'neg/cv933_24953.txt', 'neg/cv934_20426.txt', 'neg/cv935_24977.txt', 'neg/cv936_17473.tx
t', 'neg/cv937_9816.txt', 'neg/cv938_10706.txt', 'neg/cv939_11247.txt', 'neg/cv940_18935.txt', 'neg/c
v941_10718.txt', 'neg/cv942_18509.txt', 'neg/cv943_23547.txt', 'neg/cv944_15042.txt', 'neg/cv945_1301
2.txt', 'neg/cv946_20084.txt', 'neg/cv947_11316.txt', 'neg/cv948_25870.txt', 'neg/cv949_21565.txt',
'neg/cv950_13478.txt', 'neg/cv951_11816.txt', 'neg/cv952_26375.txt', 'neg/cv953_7078.txt', 'neg/cv954
_19932.txt', 'neg/cv955_26154.txt', 'neg/cv956_12547.txt', 'neg/cv957_9059.txt', 'neg/cv958_13020.tx
t', 'neg/cv959_16218.txt', 'neg/cv960_28877.txt', 'neg/cv961_5578.txt', 'neg/cv962_9813.txt', 'neg/cv
963_7208.txt', 'neg/cv964_5794.txt', 'neg/cv965_26688.txt', 'neg/cv966_28671.txt', 'neg/cv967_5626.tx
t', 'neg/cv968_25413.txt', 'neg/cv969_14760.txt', 'neg/cv970_19532.txt', 'neg/cv971_11790.txt', 'neg/
cv972_26837.txt', 'neg/cv973_10171.txt', 'neg/cv974_24303.txt', 'neg/cv975_11920.txt', 'neg/cv976_107
24.txt', 'neg/cv977_4776.txt', 'neg/cv978_22192.txt', 'neg/cv979_2029.txt', 'neg/cv980_11851.txt', 'n
eg/cv981_16679.txt', 'neg/cv982_22209.txt', 'neg/cv983_24219.txt', 'neg/cv984_14006.txt', 'neg/cv985_
5964.txt', 'neg/cv986_15092.txt', 'neg/cv987_7394.txt', 'neg/cv988_20168.txt', 'neg/cv989_17297.txt',
'neg/cv990_12443.txt', 'neg/cv991_19973.txt', 'neg/cv992_12806.txt', 'neg/cv993_29565.txt', 'neg/cv99
4_13229.txt', 'neg/cv995_23113.txt', 'neg/cv996_12447.txt', 'neg/cv997_5152.txt', 'neg/cv998_15691.tx
t', 'neg/cv999_14636.txt']
```

In [42]:

```
neg_rev=movie_reviews.fileids('neg')
rev = nltk.corpus.movie_reviews.words('pos/cv000_29590.txt')
rev
```

Out[42]:

```
['films', 'adapted', 'from', 'comic', 'books', 'have', ...]
```

In [43]:

```
rev_list=[]
for rev in neg_rev:
    rev_text_neg=rev=nltk.corpus.movie_reviews.words(rev)
    review_one_string= " ".join(rev_text_neg)
    review_one_string=review_one_string.replace(' ,',',') # Remove space that comes with comma
    review_one_string=review_one_string.replace(' .','.') # replace space that comes with dot
    rev_list.append(review_one_string)
```

In [44]:

```
len(rev_list)
```

Out[44]:

```
1000
```

In [46]:

```
pos_rev = movie_reviews.fileids('pos')
for rev_pos in pos_rev:
    rev_text_pos=rev=nltk.corpus.movie_reviews.words(rev_pos)
    review_one_string= " ".join(rev_text_neg)
    review_one_string=review_one_string.replace(' ,',',') # Remove space that comes with comma
    review_one_string=review_one_string.replace(' .','.') # replace space that comes with dot
    rev_list.append(review_one_string)
```

In [47]:

```
len(rev_list)
```

Out[47]:

```
3000
```

In [48]:

```
neg_target=np.zeros((1000,),dtype=np.int)
pos_target=np.ones((1000,),dtype=np.int)
```

In [49]:

```
target_list=[]
for neg_tar in neg_target:
 target_list.append(neg_tar)
for pos_tar in pos_target:
 target_list.append(pos_tar)
```

In [50]:

```
len(target_list)
```

Out[50]:

```
2000
```

In [51]:

```
y=pd.Series(target_list)
```

In [52]:

```
type(y)
```

Out[52]:

```
pandas.core.series.Series
```

In [53]:

```
y.head(20)
```

Out[53]:

```
0     0
1     0
2     0
3     0
4     0
5     0
6     0
7     0
8     0
9     0
10    0
11    0
12    0
13    0
14    0
15    0
16    0
17    0
18    0
19    0
dtype: int64
```

In [54]:

```
from sklearn.feature_extraction.text import CountVectorizer
```

In [55]:

```
count_vect=CountVectorizer(lowercase=True,stop_words='english',min_df=2)
x_count_vect=count_vect.fit_transform(rev_list)
x_count_vect.shape # will print (2000,16228)
```

Out[55]:

```
(3000, 16228)
```

```
In [56]: x_names=count_vect.get_feature_names()
         x_names
```

```
Out[56]: ['00',
          '000',
          '007',
          '05',
          '10',
          '100',
          '1000',
          '101',
          '102',
          '103',
          '105',
          '106',
          '107',
          '108',
          '10th',
          '11',
          '113',
          '115',
          '11th',
          '12',
          '126',
          '129',
          '13',
          '132',
          '137',
          '13th',
          '14',
          '15',
          '15th',
          '16',
          '1600',
          '16x9',
          '17',
          '17th',
          '18',
          '180',
          '1800s',
          '1869',
          '1871',
          '18th',
          '19',
          '1900',
          '1914',
          '1925',
          '1928',
          '1933',
          '1939',
          '1940s',
          '1947',
          '1948',
          '1949',
          '1950',
          '1950s',
          '1954',
          '1958',
          '1960',
          '1960s',
          '1961',
          '1962',
          '1963',
          '1964',
          '1966',
          '1967',
          '1968',
          '1969',
          '1970',
          '1970s',
          '1971',
          '1972',
          '1973',
          '1974',
          '1976',
          '1977',
          '1978',
          '1979',
          '1980',
          '1980s',
          '1981',
          '1982',
          '1983',
          '1984',
          '1985',
          '1986',
          '1987',
          '1988',
          '1989',
          '1990',
          '1990s',
          '1991',
          '1992',
          '1993',
          '1994',
          '1995',
          '1996',
          '1997',
          '1998',
          '1999',
          '19th',
          '20',
          '200',
          '2000',
          '2001',
          '2017',
          '2020',
          '2029',
          '20s',
          '20th',
          '21',
          '2176',
          '22',
          '23',
          '24',
          '25',
          '25th',
          '26',
          '26th',
          '27',
          '28',
          '29',
          '2d',
          '30',
          '300',
          '3000',
          '30th',
          '31',
          '32',
          '33',
          '35',
          '36',
          '37',
          '39',
          '40',
          '400',
          '43',
          '45',
          '47',
          '48',
          '4th',
          '50',
          '500',
          '5000',
          '50s',
          '51',
          '53',
          '54',
          '57',
          '60',
          '600',
          '6000',
          '607',
          '60s',
          '64',
          '66',
          '666',
          '70',
          '700',
          '70s',
          '75',
          '76',
          '78',
          '80',
          '800',
          '80s',
          '83',
          '84',
          '85',
          '87',
          '89',
          '8mm',
          '8th',
          '90',
          '90210',
          '90s',
          '92',
          '93',
          '94',
          '95',
          '96',
          '97',
          '98',
          '99',
          '999',
          '_21_jump_street_',
          '_____',
          '_a_night_at_the_roxbury_',
          '_and_',
          '_babe_',
          '_clueless_',
          '_does_',
          '_don',
          '_have_',
          '_in',
          '_is_',
          '_jerry_maguire_',
          '_last_',
          '_must_',
          '_not_',
          '_real_',
          '_really_',
          '_roxbury_',
          '_saturday_night_live_',
          '_saving',
          '_scream_',
          '_snl_',
          '_the',
          '_the_',
          'aa',
          'aaa',
          'aaliyah',
          'aaron',
          'aback',
          'abandon',
          'abandoned',
          'abandonment',
          'abandons',
          'abdomen',
          'abduction',
          'abe',
          'abel',
          'abetting',
          'abhorrent',
          'abilities',
          'ability',
          'ablaze',
          'able',
          'ably',
          'abo',
          'aboard',
          'abode',
```

```
'abominable',
'abomination',
'aborted',
'abortion',
'abound',
'abraham',
'abrahams',
'abrams',
'abrasive',
'abrupt',
'abruptly',
'abs',
'absence',
'absences',
'absent',
'absentee',
'absolute',
'absolutely',
'absorb',
'absorbed',
'absorbing',
'abstract',
'absurd',
'absurdity',
'absurdly',
'abundance',
'abundant',
'abundantly',
'abuse',
'abused',
'abuses',
'abusing',
'abusive',
'abysmal',
'abysmally',
'abyss',
'academic',
'academy',
'accent',
'accented',
'accents',
'accept',
'acceptable',
'acceptance',
'accepted',
'accepting',
'accepts',
'access',
'accident',
'accidental',
'accidentally',
'accidently',
'acclaim',
'acclaimed',
'accommodating',
'accompanied',
'accompaniment',
'accompany',
'accompanying',
'accomplice',
'accomplices',
'accomplish',
'accomplished',
'accomplishes',
'accomplishment',
'accomplishments',
'accordance',
'according',
'accordingly',
'account',
'accountant',
'accounted',
'accounts',
'accumulation',
'accuracy',
'accurate',
'accurately',
'accusation',
'accusations',
'accused',
'accuses',
'accusing',
'accustomed',
'ace',
'acerbic',
'aces',
'achieve',
'achieved',
'achievement',
'achievements',
'achieves',
'achieving',
'acid',
'acknowledge',
'acknowledged',
'acknowledges',
'acknowledgment',
'acquaintances',
'acquired',
'acquit',
'acquittal',
'acrobatic',
'acrobatics',
'act',
'acted',
'acting',
'action',
'actioner',
'actioners',
'actions',
'active',
'actively',
'activist',
'activities',
'activity',
'actor',
'actors',
'actress',
'actresses',
'acts',
'actual',
'actuality',
'actually',
'acute',
'ad',
'adage',
'adam',
'adams',
'adaptation',
'adaptations',
'adapted',
'adapting',
'adaption',
'add',
'addams',
'added',
'addict',
'addicted',
'addiction',
'adding',
'addition',
'additional',
'additionally',
'additions',
'addled',
'address',
'addressed',
'adds',
'adept',
'adequate',
'adequately',
'adherence',
'adheres',
'adjacent',
'adjective',
'adjectives',
'administration',
'admirable',
'admirably',
'admiration',
'admire',
'admired',
'admission',
'admit',
'admits',
'admitted',
'admittedly',
'ado',
'adolescence',
'adolescent',
'adolescents',
'adopt',
'adopted',
'adopting',
'adoptive',
'adopts',
'adorable',
'adore',
'adores',
'adorned',
'adrenalin',
'adrenaline',
'adrian',
'adrien',
'ads',
'adult',
'adulthood',
'adults',
'advance',
'advanced',
'advancement',
'advances',
'advancing',
'advantage',
'adventure',
'adventures',
'adventurous',
'advertised',
'advertisement',
'advertisements',
'advertising',
'advice',
'advise',
'advised',
'adviser',
'advises',
'advisor',
'advocate',
'advocating',
'aerosmith',
'aesthetic',
'afar',
'affable',
'affair',
'affairs',
'affect',
'affected',
'affecting',
'affection',
'affectionate',
'affections',
'affects',
'affleck',
'afflicted',
'affluent',
'afford',
'afforded',
'aficionado',
'afield',
'afloat',
'afoot',
'afore',
'aforementioned',
```

```
    'afoul',
    'afraid',
    'africa',
    'african',
    'afterlife',
    'aftermath',
    'afternoon',
    'afternoons',
    'afterthought',
    'afterward',
    'age',
    'aged',
    'agency',
    'agenda',
    'agent',
    'agents',
    'ages',
    'aggravating',
    'aggression',
    'aggressive',
    'aggressively',
    'agile',
    'agility',
    'aging',
    'ago',
    'agonizingly',
    'agree',
    'agreeable',
    'agreed',
    'agreeing',
    'agreement',
    'agrees',
    'ah',
    'ahabs',
    'ahead',
    'ahem',
    'ahern',
    'ahh',
    'ahmed',
    'aid',
    'aidan',
    'aide',
    'aided',
    'aiding',
    'aids',
    'aiello',
    'ailing',
    'aim',
    'aimed',
    'aiming',
    'aimless',
    'aimlessly',
    'aims',
    'ain',
    'air',
    'aircraft',
    'aired',
    'aires',
    'airline',
    'airlock',
    'airplane',
    'airplanes',
    'airport',
    'airwolf',
    'airy',
    'aisles',
    'aka',
    'akin',
    'akiva',
    'akroyd',
    'al',
    'ala',
    'alan',
    'alarm',
    'alarmed',
    'alarming',
    'alas',
    'alaska',
    'albeit',
    'albert',
    'albino',
    'albinos',
    'album',
    'alcohol',
    'alcoholic',
    'alcoholics',
    'alcoholism',
    'alcott',
    'aldys',
    'alec',
    'alert',
    'alerted',
    'alessandro',
    'alex',
    'alexander',
    'alexandre',
    'alfred',
    'ali',
    'alias',
    'alice',
    'alicia',
    'alien',
    'alienate',
    'alienated',
    'alienation',
    'aliens',
    'align',
    'alignment',
    'alike',
    'alison',
    'alive',
    'allan',
    'alleged',
    'allegedly',
    'allegiance',
    'allegory',
    'allen',
    'allergy',
    'alley',
    'alliance',
    'alliances',
    'alligators',
    'allison',
    'allow',
    'allowed',
    'allowing',
    'allows',
    'alluded',
    'allure',
    'alluring',
    'allusions',
    'ally',
    'alongside',
    'aloof',
    'alot',
    'alright',
    'altar',
    'alter',
    'altered',
    'altering',
    'alternate',
    'alternately',
    'alternates',
    'alternating',
    'alternative',
    'altman',
    'altogether',
    'alumnus',
    'alyson',
    'alzheimer',
    'amalgamation',
    'amanda',
    'amarillo',
    'amateur',
    'amateurish',
    'amateurishly',
    'amaze',
    'amazed',
    'amazing',
    'amazingly',
    'amazon',
    'ambassador',
    'ambience',
    'ambiguity',
    'ambiguous',
    'ambition',
    'ambitions',
    'ambitious',
    'ambitiously',
    'ambrose',
    'ambush',
    'america',
    'american',
    'americans',
    'ames',
    'amiable',
    'amicable',
    'amid',
    'amidala',
    'amidst',
    'amis',
    'amish',
    'amiss',
    'amistad',
    'ammo',
    'ammunition',
    'amnesia',
    'amok',
    'amoral',
    'amounts',
    'ample',
    'amply',
    'amuck',
    'amuse',
    'amused',
    'amusement',
    'amusements',
    'amusing',
    'amusingly',
    'amy',
    'anaconda',
    'anacondas',
    'anakin',
    'anal',
    'analogy',
    'analysis',
    'analyze',
    'analyzed',
    'anarchic',
    'anastasia',
    'anatomy',
    'anchor',
    'ancient',
    'anders',
    'anderson',
    'andie',
    'andre',
    'andrea',
    'andrew',
    'andrews',
    'android',
    'andrzej',
    'andy',
    'anemic',
    'anew',
    'angel',
    'angela',
    'angeles',
    'angelic',
    'angelina',
    'angelo',
    'angels',
    'anger',
```

```
            'angered',
            'angle',
            'angles',
            'angrily',
            'angry',
            'angst',
            'angsty',
            'anguish',
            'anguished',
            'angus',
            'animal',
            'animals',
            'animated',
            'animation',
            'animator',
            'animators',
            'animatronic',
            'animatronics',
            'anime',
            'aniston',
            'anita',
            'anjelica',
            'ann',
            'anna',
            'annabella',
            'annabeth',
            'anne',
            'annette',
            'annie',
            'annihilate',
            'annihilation',
            'anniversary',
            'announce',
            'announced',
            'announcement',
            'announcer',
            'announces',
            'announcing',
            'annoy',
            'annoyance',
            'annoyed',
            'annoying',
            'annoyingly',
            'annoys',
            'anonymous',
            'answer',
            'answered',
            'answering',
            'answers',
            'ant',
            'antagonism',
            'antagonist',
            'antagonists',
            'antagonizes',
            'anthony',
            'anthropologist',
            'anthropology',
            'anti',
            'antichrist',
            'anticipate',
            'anticipated',
            'anticipating',
            'anticipation',
            'anticlimactic',
            'antics',
            'antidote',
            'antoine',
            'antonio',
            'antonioni',
            'ants',
            'antsy',
            'antwerp',
            'antz',
            'anxiety',
            'anybody',
            'anymore',
            'anytime',
            'anyways',
            'apart',
            'apartment',
            'ape',
            'apes',
            'apex',
            'aplenty',
            'aplomb',
            'apocalypse',
            'apocalyptic',
            'apollo',
            'apologize',
            'apologizes',
            'apology',
            'appalled',
            'appalling',
            'appallingly',
            'apparantly',
            'apparent',
            'apparently',
            'apparitions',
            'appeal',
            'appealing',
            'appeals',
            'appear',
            'appearance',
            'appearances',
            'appeared',
            'appearing',
            'appears',
            'applaud',
            'applauding',
            'applause',
            'apple',
            'applegate',
            'applicable',
            'applied',
            'applies',
            'apply',
            'appointed',
            'appointment',
            'appreciable',
            'appreciate',
            'appreciated',
            'appreciation',
            'apprehend',
            'apprentice',
            'approach',
            'approachable',
            'approached',
            'approaches',
            'approaching',
            'appropriate',
            'appropriately',
            'approved',
            'approximately',
            'april',
            'apt',
            'aptitude',
            'aptly',
            'arab',
            'araki',
            'aramis',
            'arbitrary',
            'arbogast',
            'arc',
            'arcade',
            'arch',
            'archeologist',
            'archer',
            'archetype',
 'archetypes',
 'architect',
 'arduous',
 'area',
 'areas',
 'aren',
 'arena',
 'aretha',
 'argonautica',
 'arguable',
 'arguably',
 'argue',
 'argued',
 'argues',
 'arguing',
 'argument',
 'arguments',
 'ari',
 'arise',
 'arisen',
 'arises',
 'aristocrat',
 'arizona',
 'ark',
 'arkin',
 'arlington',
 'arliss',
 'arm',
 'armageddon',
 'armand',
 'armed',
 'armies',
 'armin',
 'armor',
 'armored',
 'arms',
 'army',
 'arnie',
 'arnold',
 'arouses',
 'arquette',
 'arranged',
 'arrangement',
 'arrangements',
 'array',
 'arrest',
 'arrested',
 'arresting',
 'arrival',
 'arrive',
 'arrived',
 'arrives',
 'arriving',
 'arrogance',
 'arrogant',
 'arrow',
 'art',
 'artagnan',
 'artemus',
 'arthur',
 'article',
 'articles',
 'articulate',
 'artie',
 'artifacts',
 'artificial',
 'artificiality',
 'artificially',
 'artist',
 'artistic',
 'artistically',
 'artistry',
 'artists',
 'arts',
 'artsy',
 'artwork',
 'arty',
 'ash',
 'ashamed',
 'ashes',
 'ashley',
 'ashton',
 'asia',
 'asian',
 'aside',
 'asides',
 'asinine',
 'ask',
 'asked',
 'askew',
 'asking',
 'asks',
 'asleep',
 'asner',
 'aspect',
 'aspects',
 'aspirations',
 'aspire',
 'aspiring',
 'ass',
 'assante',
 'assasin',
 'assasination',
 'assassin',
 'assassinate',
 'assassinated',
 'assassination',
 'assassins',
 'assault',
 'assaulted',
 'assaulting',
 'assaults',
 'assed',
 'assemble',
 'assembled',
 'assembles',
 'assembly',
 'assert',
 'assertive',
 'asses',
 'assess',
 'assessment',
 'asset',
 'assets',
 'asshole',
 'assigned',
 'assigning',
 'assignment',
 'assigns',
 'assimilating',
 'assist',
 'assistance',
 'assistant',
 'assisted',
 'associate',
 'associated',
 'associates',
 'association',
 'assorted',
 'assortment',
 'assume',
 'assumed',
 'assumes',
 'assuming',
 'assumption',
 'assumptions',
 'assure',
 'assured',
 'assures',
 'assuring',
 'asswhole',
 'asteroid',
 'asteroids',
 'asthma',
 'astonished',
 'astonishing',
 'astonishingly',
 ...]
```

```
In [57]: x_count_vect.shape
```

```
Out[57]: (3000, 16228)
```