

I081_Aniruddh_Kulkarni_NLP_EXP3

May 28, 2023

1 Name: Aniruddh Kulkarni

2 Roll no: I081

3 Stream: CS (AI)

4 Division: I

5 Semester: 5th Semester

6 Batch: I-3

7 Subject: NLP

8 Assignment-3

9 Analyzing Gutenberg and Brown corpus with python

- (a) Working with corpus file lists
- (b) Working with file contents
- (c) Visualization

10 NLTK includes a small selection of texts from the Project Gutenberg electronic text

archive, which contains some 25,000 free electronic books, hosted at <http://www.gutenberg.org/>.

```
[1]: import nltk
nltk.download('all')
```

```
[nltk_data] Downloading collection 'all'
[nltk_data] |
[nltk_data] | Downloading package abc to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package abc is already up-to-date!
[nltk_data] | Downloading package alpino to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
```

```

[nltk_data] | Package alpino is already up-to-date!
[nltk_data] | Downloading package averaged_perceptron_tagger to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package averaged_perceptron_tagger is already up-
[nltk_data] | to-date!
[nltk_data] | Downloading package averaged_perceptron_tagger_ru to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package averaged_perceptron_tagger_ru is already
[nltk_data] | up-to-date!
[nltk_data] | Downloading package basque_grammars to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package basque_grammars is already up-to-date!
[nltk_data] | Downloading package bcp47 to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package bcp47 is already up-to-date!
[nltk_data] | Downloading package biocreative_ppi to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package biocreative_ppi is already up-to-date!
[nltk_data] | Downloading package bllip_wsj_no_aux to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package bllip_wsj_no_aux is already up-to-date!
[nltk_data] | Downloading package book_grammars to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package book_grammars is already up-to-date!
[nltk_data] | Downloading package brown to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package brown is already up-to-date!
[nltk_data] | Downloading package brown_tei to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package brown_tei is already up-to-date!
[nltk_data] | Downloading package cess_cat to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package cess_cat is already up-to-date!
[nltk_data] | Downloading package cess_esp to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package cess_esp is already up-to-date!
[nltk_data] | Downloading package chat80 to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package chat80 is already up-to-date!
[nltk_data] | Downloading package city_database to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package city_database is already up-to-date!
[nltk_data] | Downloading package cmudict to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package cmudict is already up-to-date!
[nltk_data] | Downloading package comparative_sentences to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package comparative_sentences is already up-to-

```

```

[nltk_data] | date!
[nltk_data] | Downloading package comtrans to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package comtrans is already up-to-date!
[nltk_data] | Downloading package conll2000 to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package conll2000 is already up-to-date!
[nltk_data] | Downloading package conll2002 to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package conll2002 is already up-to-date!
[nltk_data] | Downloading package conll2007 to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package conll2007 is already up-to-date!
[nltk_data] | Downloading package crubadan to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package crubadan is already up-to-date!
[nltk_data] | Downloading package dependency_treebank to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package dependency_treebank is already up-to-date!
[nltk_data] | Downloading package dolch to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package dolch is already up-to-date!
[nltk_data] | Downloading package europarl_raw to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package europarl_raw is already up-to-date!
[nltk_data] | Downloading package extended_omw to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package extended_omw is already up-to-date!
[nltk_data] | Downloading package floresta to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package floresta is already up-to-date!
[nltk_data] | Downloading package framenet_v15 to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package framenet_v15 is already up-to-date!
[nltk_data] | Downloading package framenet_v17 to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package framenet_v17 is already up-to-date!
[nltk_data] | Downloading package gazetteers to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package gazetteers is already up-to-date!
[nltk_data] | Downloading package genesis to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package genesis is already up-to-date!
[nltk_data] | Downloading package gutenber to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package gutenber is already up-to-date!
[nltk_data] | Downloading package ieer to
[nltk_data] | /Users/pushpakulkarni/nltk_data...

```

```

[nltk_data] | Package ieer is already up-to-date!
[nltk_data] | Downloading package inaugural to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package inaugural is already up-to-date!
[nltk_data] | Downloading package indian to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package indian is already up-to-date!
[nltk_data] | Downloading package jeita to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package jeita is already up-to-date!
[nltk_data] | Downloading package kimmo to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package kimmo is already up-to-date!
[nltk_data] | Downloading package knbc to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package knbc is already up-to-date!
[nltk_data] | Downloading package large_grammars to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package large_grammars is already up-to-date!
[nltk_data] | Downloading package lin_thesaurus to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package lin_thesaurus is already up-to-date!
[nltk_data] | Downloading package mac_morpho to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package mac_morpho is already up-to-date!
[nltk_data] | Downloading package machado to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package machado is already up-to-date!
[nltk_data] | Downloading package masc_tagged to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package masc_tagged is already up-to-date!
[nltk_data] | Downloading package maxent_ne_chunker to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package maxent_ne_chunker is already up-to-date!
[nltk_data] | Downloading package maxent_treebank_pos_tagger to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package maxent_treebank_pos_tagger is already up-
[nltk_data] | to-date!
[nltk_data] | Downloading package moses_sample to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package moses_sample is already up-to-date!
[nltk_data] | Downloading package movie_reviews to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package movie_reviews is already up-to-date!
[nltk_data] | Downloading package mte_teip5 to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package mte_teip5 is already up-to-date!
[nltk_data] | Downloading package mwa_ppdb to

```

```

[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package mwa_ppdb is already up-to-date!
[nltk_data] | Downloading package names to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package names is already up-to-date!
[nltk_data] | Downloading package nombank.1.0 to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package nombank.1.0 is already up-to-date!
[nltk_data] | Downloading package nonbreaking_prefixes to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package nonbreaking_prefixes is already up-to-date!
[nltk_data] | Downloading package nps_chat to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package nps_chat is already up-to-date!
[nltk_data] | Downloading package omw to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package omw is already up-to-date!
[nltk_data] | Downloading package omw-1.4 to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package omw-1.4 is already up-to-date!
[nltk_data] | Downloading package opinion_lexicon to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package opinion_lexicon is already up-to-date!
[nltk_data] | Downloading package panlex_swadesh to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package panlex_swadesh is already up-to-date!
[nltk_data] | Downloading package paradigms to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package paradigms is already up-to-date!
[nltk_data] | Downloading package pe08 to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package pe08 is already up-to-date!
[nltk_data] | Downloading package perluniprops to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package perluniprops is already up-to-date!
[nltk_data] | Downloading package pil to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package pil is already up-to-date!
[nltk_data] | Downloading package pl196x to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package pl196x is already up-to-date!
[nltk_data] | Downloading package porter_test to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package porter_test is already up-to-date!
[nltk_data] | Downloading package ppattach to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package ppattach is already up-to-date!
[nltk_data] | Downloading package problem_reports to

```

```

[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package problem_reports is already up-to-date!
[nltk_data] | Downloading package product_reviews_1 to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package product_reviews_1 is already up-to-date!
[nltk_data] | Downloading package product_reviews_2 to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package product_reviews_2 is already up-to-date!
[nltk_data] | Downloading package propbank to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package propbank is already up-to-date!
[nltk_data] | Downloading package pros_cons to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package pros_cons is already up-to-date!
[nltk_data] | Downloading package ptb to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package ptb is already up-to-date!
[nltk_data] | Downloading package punkt to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package punkt is already up-to-date!
[nltk_data] | Downloading package qc to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package qc is already up-to-date!
[nltk_data] | Downloading package reuters to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package reuters is already up-to-date!
[nltk_data] | Downloading package rslp to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package rslp is already up-to-date!
[nltk_data] | Downloading package rte to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package rte is already up-to-date!
[nltk_data] | Downloading package sample_grammars to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package sample_grammars is already up-to-date!
[nltk_data] | Downloading package semcor to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package semcor is already up-to-date!
[nltk_data] | Downloading package senseval to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package senseval is already up-to-date!
[nltk_data] | Downloading package sentence_polarity to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package sentence_polarity is already up-to-date!
[nltk_data] | Downloading package sentiwordnet to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package sentiwordnet is already up-to-date!
[nltk_data] | Downloading package shakespeare to

```

```

[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package shakespeare is already up-to-date!
[nltk_data] | Downloading package sinica_treebank to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package sinica_treebank is already up-to-date!
[nltk_data] | Downloading package smultron to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package smultron is already up-to-date!
[nltk_data] | Downloading package snowball_data to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package snowball_data is already up-to-date!
[nltk_data] | Downloading package spanish_grammars to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package spanish_grammars is already up-to-date!
[nltk_data] | Downloading package state_union to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package state_union is already up-to-date!
[nltk_data] | Downloading package stopwords to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package stopwords is already up-to-date!
[nltk_data] | Downloading package subjectivity to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package subjectivity is already up-to-date!
[nltk_data] | Downloading package swadesh to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package swadesh is already up-to-date!
[nltk_data] | Downloading package switchboard to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package switchboard is already up-to-date!
[nltk_data] | Downloading package tagsets to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package tagsets is already up-to-date!
[nltk_data] | Downloading package timit to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package timit is already up-to-date!
[nltk_data] | Downloading package toolbox to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package toolbox is already up-to-date!
[nltk_data] | Downloading package treebank to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package treebank is already up-to-date!
[nltk_data] | Downloading package twitter_samples to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package twitter_samples is already up-to-date!
[nltk_data] | Downloading package udhr to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package udhr is already up-to-date!
[nltk_data] | Downloading package udhr2 to

```

```

[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package udhr2 is already up-to-date!
[nltk_data] | Downloading package unicode_samples to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package unicode_samples is already up-to-date!
[nltk_data] | Downloading package universal_tagset to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package universal_tagset is already up-to-date!
[nltk_data] | Downloading package universal_treebanks_v20 to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package universal_treebanks_v20 is already up-to-
[nltk_data] | date!
[nltk_data] | Downloading package vader_lexicon to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package vader_lexicon is already up-to-date!
[nltk_data] | Downloading package verbnet to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package verbnet is already up-to-date!
[nltk_data] | Downloading package verbnet3 to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package verbnet3 is already up-to-date!
[nltk_data] | Downloading package webtext to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package webtext is already up-to-date!
[nltk_data] | Downloading package wmt15_eval to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package wmt15_eval is already up-to-date!
[nltk_data] | Downloading package word2vec_sample to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package word2vec_sample is already up-to-date!
[nltk_data] | Downloading package wordnet to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package wordnet is already up-to-date!
[nltk_data] | Downloading package wordnet2021 to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package wordnet2021 is already up-to-date!
[nltk_data] | Downloading package wordnet2022 to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package wordnet2022 is already up-to-date!
[nltk_data] | Downloading package wordnet31 to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package wordnet31 is already up-to-date!
[nltk_data] | Downloading package wordnet_ic to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package wordnet_ic is already up-to-date!
[nltk_data] | Downloading package words to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package words is already up-to-date!

```



```
[nltk_data] | Downloading package ycoe to
[nltk_data] | /Users/pushpakulkarni/nltk_data...
[nltk_data] | Package ycoe is already up-to-date!
[nltk_data] |
[nltk_data] Done downloading collection all
```

[1]: True

```
[2]: import nltk
nltk.download('gutenberg')
nltk.corpus.gutenberg.fileids()
```

```
[nltk_data] Downloading package gutenberg to
[nltk_data] /Users/pushpakulkarni/nltk_data...
[nltk_data] Package gutenberg is already up-to-date!
```

```
[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']
```

```
[3]: emma=nltk.corpus.gutenberg.words('austen-emma.txt')
len(emma)
```

[3]: 192427

11 Concordance in corpus

```
[4]: emma = nltk.Text(nltk.corpus.gutenberg.words('austen-emma.txt'))
emma.concordance("surprise")
```

Displaying 25 of 37 matches:

er father , was sometimes taken by surprise at his being still able to pity `

hem do the other any good ." " You surprize me ! Emma must do Harriet good : a Knightley actually looked red with surprize and displeasure , as he stood up , r . Elton , and found to his great surprize , that Mr . Elton was actually on d aid ." Emma saw Mrs . Weston ' s surprize , and felt that it must be great , father was quite taken up with the surprize of so sudden a journey , and his f y , in all the favouring warmth of surprize and conjecture . She was , moreove he appeared , to have her share of surprize , introduction , and pleasure . Th ir plans ; and it was an agreeable surprize to her , therefore , to perceive t talking aunt had taken me quite by surprize , it must have been the death of m f all the dialogue which ensued of surprize , and inquiry , and congratulation the present . They might chuse to surprize her ." Mrs . Cole had many to agre the mode of it , the mystery , the surprize , is more like a young woman ' s s to her song took her agreeably by surprize -- a second , slightly but correct " " Oh ! no -- there is nothing to surprize one at all .-- A pretty fortune ; t to be considered . Emma ' s only surprize was that Jane Fairfax should accep of your admiration may take you by surprize some day or other ." Mr . Knightle ation for her will ever take me by surprize .-- I never had a thought of her i expected by the best judges , for surprize -- but there was great joy . Mr . sound of at first , without great surprize . " So unreasonably early !" she w d Frank Churchill , with a look of surprize and displeasure .-- " That is easy ; and Emma could imagine with what surprize and mortification she must be retu tled that Jane should go . Quite a surprize to me ! I had not the least idea ! . It is impossible to express our surprize . He came to speak to his father o g engaged !" Emma even jumped with surprize ;-- and , horror - struck , exclai

```
[5]: emma = nltk.Text(nltk.corpus.gutenberg.words('blake-poems.txt'))
emma.concordance("god")
```

Displaying 11 of 11 matches:

```
re called by His name . Little Lamb , God bless thee ! Little Lamb , God bless
Lamb , God bless thee ! Little Lamb , God bless thee ! THE LITTLE BLACK BOY My
ay : " Look on the rising sun : there God does live , And gives His light , and
te cloud free , And round the tent of God like lambs we joy I ' ll shade him fr
if he ' d be a good boy , He ' d have God for his father , and never want joy .
wandering light , Began to cry , but God , ever nigh , Appeared like his fathe
Mercy , Pity , Peace , and Love , Is God our Father dear ; And Mercy , Pity ,
Mercy , Love , and Pity dwell , There God is dwelling too . HOLY THURSDAY ' Twa
me no injury , And are gone to praise God and his priest and king , Who make up
ldren , nor fasting , nor birch . And God , like a father , rejoicing to see Hi
his , and therefore did I weep : That God would love a Worm I knew , and punish
```

12 To Do

-Displays three statistics for each text: average word length, average sentence length, and the number of times each vocabulary item appears in the text on average (our lexical diversity score). Anything specific you noticed???

```
[6]: from nltk.corpus import gutenberg
      gutenberg.fileids()
```

```
[6]: ['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']
```

```
[7]: for fileid in gutenberg.fileids():
      num_chars = len(gutenberg.raw(fileid))
      num_words = len(gutenberg.words(fileid))
      num_sents = len(gutenberg.sents(fileid))
      num_vocab = len(set(w.lower() for w in gutenberg.words(fileid)))
      print('Avg word length:', round(num_chars/num_words), 'Avg sentence length:',
      ↪round(num_words/num_sents), 'Lexical Diversity:', round(num_words/
      ↪num_vocab), 'Text name:', fileid)
```

```
Avg word length: 5 Avg sentence length: 25 Lexical Diversity: 26 Text name:
austen-emma.txt
Avg word length: 5 Avg sentence length: 26 Lexical Diversity: 17 Text name:
austen-persuasion.txt
Avg word length: 5 Avg sentence length: 28 Lexical Diversity: 22 Text name:
austen-sense.txt
Avg word length: 4 Avg sentence length: 34 Lexical Diversity: 79 Text name:
bible-kjv.txt
Avg word length: 5 Avg sentence length: 19 Lexical Diversity: 5 Text name:
blake-poems.txt
Avg word length: 4 Avg sentence length: 19 Lexical Diversity: 14 Text name:
bryant-stories.txt
Avg word length: 4 Avg sentence length: 18 Lexical Diversity: 12 Text name:
burgess-busterbrown.txt
Avg word length: 4 Avg sentence length: 20 Lexical Diversity: 13 Text name:
carroll-alice.txt
```

Avg word length: 5 Avg sentence length: 20 Lexical Diversity: 12 Text name: chesterton-ball.txt
 Avg word length: 5 Avg sentence length: 23 Lexical Diversity: 11 Text name: chesterton-brown.txt
 Avg word length: 5 Avg sentence length: 18 Lexical Diversity: 11 Text name: chesterton-thursday.txt
 Avg word length: 4 Avg sentence length: 21 Lexical Diversity: 25 Text name: edgeworth-parents.txt
 Avg word length: 5 Avg sentence length: 26 Lexical Diversity: 15 Text name: melville-moby_dick.txt
 Avg word length: 5 Avg sentence length: 52 Lexical Diversity: 11 Text name: milton-paradise.txt
 Avg word length: 4 Avg sentence length: 12 Lexical Diversity: 9 Text name: shakespeare-caesar.txt
 Avg word length: 4 Avg sentence length: 12 Lexical Diversity: 8 Text name: shakespeare-hamlet.txt
 Avg word length: 4 Avg sentence length: 12 Lexical Diversity: 7 Text name: shakespeare-macbeth.txt
 Avg word length: 5 Avg sentence length: 36 Lexical Diversity: 12 Text name: whitman-leaves.txt

13 Ans: We can observe that across all documents, word length on an average remains the same. Avg sentence length and lexical diversity varies.

```
[8]: #how many letters occur in the text, including the spaces between words.
from nltk.corpus import gutenberg
gutenberg.fileids()
len(gutenberg.raw('blake-poems.txt'))
```

[8]: 38153

```
[9]: nltk.download('punkt')
blake_sentences=gutenberg.sents('blake-poems.txt')
blake_sentences
```

```
[nltk_data] Downloading package punkt to
[nltk_data] /Users/pushpakulkarni/nltk_data...
[nltk_data] Package punkt is already up-to-date!
```

```
[9]: [['[', 'Poems', 'by', 'William', 'Blake', '1789', ']'], ['SONGS', 'OF',
'INNOCENCE', 'AND', 'OF', 'EXPERIENCE', 'and', 'THE', 'BOOK', 'of', 'THEL'],
...]
```

```
[10]: blake_sentencesii=gutenberg.raw('blake-poems.txt')
blake_sentencesii
```

[10]: '[Poems by William Blake 1789]\n\n \nSONGS OF INNOCENCE AND OF EXPERIENCE\nand THE BOOK of THEL\n\n\n SONGS OF INNOCENCE\n \n \n INTRODUCTION\n \n \n Piping down the valleys wild,\n \n Piping songs of pleasant glee,\n \n On a cloud I saw a child,\n \n And he laughing said to me:\n \n "Pipe a song about a Lamb!"\n \n So I piped with merry cheer.\n \n "Piper, pipe that song again;"\n \n So I piped: he wept to hear.\n \n \n "Drop thy pipe, thy happy pipe;\n \n Sing thy songs of happy cheer:!"\n \n So I sang the same again,\n \n While he wept with joy to hear.\n \n \n "Piper, sit thee down and write\n \n In a book, that all may read."\n \n So he vanish\'d from my sight;\n \n And I pluck\'d a hollow reed,\n \n \n And I made a rural pen,\n \n And I stain\'d the water clear,\n \n And I wrote my happy songs\n \n Every child may joy to hear.\n \n \n \n THE SHEPHERD\n \n \n How sweet is the Shepherd\'s sweet lot!\n \n From the morn to the evening he stays;\n \n He shall follow his sheep all the day,\n \n And his tongue shall be filled with praise.\n \n \n For he hears the lambs\' innocent call,\n \n And he hears the ewes\' tender reply;\n \n He is watching while they are in peace,\n \n For they know when their Shepherd is nigh.\n \n \n \n THE ECHOING GREEN\n \n \n The sun does arise,\n \n And make happy the skies;\n \n The merry bells ring\n \n To welcome the Spring;\n \n The skylark and thrush,\n \n The birds of the bush,\n \n Sing louder around\n \n To the bells\' cheerful sound;\n \n While our sports shall be seen\n \n On the echoing Green.\n \n \n \n Old John, with white hair,\n \n Does laugh away care,\n \n Sitting under the oak,\n \n Among the old folk.\n \n They laugh at our play,\n \n And soon they all say,\n \n "Such, such were the joys\n \n When we all -- girls and boys --\n \n In our youth-time were seen\n \n On the echoing Green." \n \n \n Till the little ones, weary,\n \n No more can be merry:\n \n The sun does descend,\n \n And our sports have an end.\n \n Round the laps of their mothers\n \n Many sisters and brothers,\n \n Like birds in their nest,\n \n Are ready for rest,\n \n And sport no more seen\n \n On the darkening green.\n \n \n \n THE LAMB\n \n \n Little Lamb, who make thee\n \n Dost thou know who made thee,\n \n Gave thee life, and bid thee feed\n \n By the stream and o\'er the mead;\n \n Gave thee clothing of delight,\n \n Softest clothing, woolly, bright;\n \n Gave thee such a tender voice,\n \n Making all the vales rejoice?\n \n Little Lamb, who made thee?\n \n Dost thou know who made thee?\n \n \n Little Lamb, I\'ll tell thee;\n \n Little Lamb, I\'ll tell thee:\n \n He is called by thy name,\n \n For He calls Himself a Lamb\n \n He is meek, and He is mild,\n \n He became a little child.\n \n I a child, and thou a lamb,\n \n We are called by His name.\n \n Little Lamb, God bless thee!\n \n Little Lamb, God bless thee!\n \n \n \n THE LITTLE BLACK BOY\n \n \n My mother bore me in the southern wild,\n \n And I am black, but oh my soul is white!\n \n White as an angel is the English child,\n \n But I am black, as if bereaved of light.\n \n \n \n My mother taught me underneath a tree,\n \n And, sitting down before the heat of day,\n \n She took me on her lap and kissed me,\n \n And, pointed to the east, began to say:\n \n \n "Look on the rising sun: there God does live,\n \n And gives His light, and gives His heat away,\n \n And flowers and trees and beasts and men receive\n \n Comfort in morning, joy in the noonday.\n \n \n "And we are put on earth a little space,\n \n That we may learn to bear the beams of love\n \n And these black bodies and this sunburnt face\n \n Is but a cloud, and like a shady grove.\n \n \n "For when our souls have learn\'d the heat to bear,\n \n The cloud will vanish, we shall hear His voice,\n \n Saying, \'Come out from the grove, my love and care\n \n And round my golden tent like lambs rejoice\'," \n \n \n Thus did

my mother say, and kissed me;\n And thus I say to little English boy.\n When I
 from black and he from white cloud free,\n And round the tent of God like lambs
 we joy\n \n I'll shade him from the heat till he can bear\n To lean in joy upon
 our Father's knee;\n And then I'll stand and stroke his silver hair,\n And be
 like him, and he will then love me.\n \n \n THE BLOSSOM\n \n Merry, merry
 sparrow!\n Under leaves so green\n A happy blossom \n Sees you, swift as
 arrow,\n Seek your cradle narrow,\n Near my bosom.\n Pretty, pretty robin!\n
 Under leaves so green\n A happy blossom\n Hears you sobbing, sobbing,\n Pretty,
 pretty robin,\n Near my bosom.\n \n \n THE CHIMNEY-SWEEPER\n \n When my mother
 died I was very young,\n And my father sold me while yet my tongue\n Could
 scarcely cry "Weep! weep! weep! weep!"\n So your chimneys I sweep, and in soot I
 sleep.\n \n There's little Tom Dacre, who cried when his head,\n That curled
 like a lamb's back, was shaved; so I said,\n "Hush, Tom! never mind it, for,
 when your head's bare,\n You know that the soot cannot spoil your white
 hair."\n \n And so he was quiet, and that very night,\n As Tom was a-sleeping,
 he had such a sight! --\n That thousands of sweepers, Dick, Joe, Ned, and
 Jack,\n Were all of them locked up in coffins of black.\n \n And by came an
 angel, who had a bright key,\n And he opened the coffins, and let them all
 free;\n Then down a green plain, leaping, laughing, they run,\n And wash in a
 river, and shine in the sun.\n \n Then naked and white, all their bags left
 behind,\n They rise upon clouds, and sport in the wind;\n And the Angel told
 Tom, if he'd be a good boy,\n He'd have God for his father, and never want
 joy.\n \n And so Tom awoke, and we rose in the dark,\n And got with our bags and
 our brushes to work.\n Though the morning was cold, Tom was happy and warm:\n So,
 if all do their duty, they need not fear harm.\n \n \n THE LITTLE BOY LOST\n
 \n "Father, father, where are you going?\n Oh do not walk so fast!\n Speak,
 father, speak to you little boy,\n Or else I shall be lost."\n \n The night
 was dark, no father was there,\n The child was wet with dew;\n The mire was
 deep, and the child did weep,\n And away the vapour flew.\n \n \n THE LITTLE
 BOY FOUND\n \n The little boy lost in the lonely fen,\n Led by the wandering
 light,\n Began to cry, but God, ever nigh,\n Appeared like his father, in
 white.\n \n He kissed the child, and by the hand led,\n And to his mother
 brought,\n Who in sorrow pale, through the lonely dale,\n The little boy
 weeping sought.\n \n \n LAUGHING SONG\n \n When the green woods laugh with the
 voice of joy,\n And the dimpling stream runs laughing by;\n When the air does
 laugh with our merry wit,\n And the green hill laughs with the noise of it;\n \n
 when the meadows laugh with lively green,\n And the grasshopper laughs in the
 merry scene,\n When Mary and Susan and Emily\n With their sweet round mouths
 sing "Ha, ha he!"\n \n When the painted birds laugh in the shade,\n Where our
 table with cherries and nuts is spread:\n Come live, and be merry, and join with
 me,\n To sing the sweet chorus of "Ha, ha, he!"\n \n \n A SONG\n \n Sweet
 dreams, form a shade\n O'er my lovely infant's head!\n Sweet dreams of
 pleasant streams\n By happy, silent, moony beams!\n \n Sweet Sleep, with soft
 down\n Weave thy brows an infant crown\n Sweet Sleep, angel mild,\n Hover o'er
 my happy child!\n \n Sweet smiles, in the night\n Hover over my delight!\n Sweet
 smiles, mother's smile,\n All the livelong night beguile.\n \n Sweet moans,
 dovelike sighs,\n Chase not slumber from thine eyes!\n Sweet moan, sweeter

smile,\n All the dovelike moans beguile.\n \n Sleep, sleep, happy child!\n All
 creation slept and smiled.\n Sleep, sleep, happy sleep,\n While o'er thee doth
 mother weep.\n \n Sweet babe, in thy face\n Holy image I can trace;\n Sweet
 babe, once like thee\n Thy Maker lay, and wept for me:\n \n Wept for me, for
 thee, for all,\n When He was an infant small.\n Thou His image ever see,\n
 Heavenly face that smiles on thee!\n \n Smiles on thee, on me, on all,\n Who
 became an infant small;\n Infant smiles are his own smiles;\n Heaven and earth
 to peace beguiles.\n \n \n DIVINE IMAGE\n \n To Mercy, Pity, Peace, and Love,\n
 All pray in their distress,\n And to these virtues of delight\n Return their
 thankfulness.\n \n For Mercy, Pity, Peace, and Love,\n Is God our Father
 dear;\n And Mercy, Pity, Peace, and Love,\n Is man, his child and care.\n \n
 For Mercy has a human heart\n Pity, a human face;\n And Love, the human form
 divine;\n And Peace, the human dress.\n \n Then every man, of every clime,\n
 That prays in his distress,\n Prays to the human form divine:\n Love, Mercy,
 Pity, Peace.\n \n And all must love the human form,\n In heathen, Turk, or
 Jew.\n Where Mercy, Love, and Pity dwell,\n There God is dwelling too.\n \n \n
 HOLY THURSDAY\n \n 'Twas on a Holy Thursday, their innocent faces clean,\n
 Came children walking two and two, in read, and blue, and green:\n Grey-headed
 beards walked before, with wands as white as snow,\n Till into the high dome of
 Paul's they like Thames waters flow.\n \n Oh what a multitude they seemed,
 these flowers of London town!\n Seated in companies they sit, with radiance all
 their own.\n The hum of multitudes was there, but multitudes of lambs,\n
 Thousands of little boys and girls raising their innocent hands.\n \n Now like a
 mighty wild they raise to heaven the voice of song,\n Or like harmonious
 thunderings the seats of heaven among:\n Beneath them sit the aged man, wise
 guardians of the poor.\n Then cherish pity, lest you drive an angel from your
 door.\n \n \n NIGHT\n \n The sun descending in the west,\n The evening star does
 shine;\n The birds are silent in their nest,\n And I must seek for mine.\n The
 moon, like a flower\n In heaven's high bower,\n With silent delight,\n
 Sits and smiles on the night.\n \n Farewell, green fields and happy grove,\n
 Where flocks have ta'en delight.\n Where lambs have nibbled, silent move\n The
 feet of angels bright;\n Unseen they pour blessing,\n And joy without
 ceasing,\n On each bud and blossom,\n And each sleeping bosom.\n \n They
 look in every thoughtless nest\n Where birds are covered warm;\n They visit
 caves of every beast,\n To keep them all from harm:\n If they see any
 weeping\n That should have been sleeping,\n They pour sleep on their head,\n
 And sit down by their bed.\n \n When wolves and tigers howl for prey,\n They
 pitying stand and weep;\n Seeking to drive their thirst away,\n And keep them
 from the sheep.\n But, if they rush dreadful,\n The angels, most heedful,\n
 Receive each mild spirit,\n New worlds to inherit.\n \n \n And there the
 lion's ruddy eyes\n Shall flow with tears of gold:\n And pitying the tender
 cries,\n And walking round the fold:\n Saying: "Wrath by His meekness,\n
 And, by His health, sickness,\n Are driven away\n From our immortal day.\n \n
 "And now beside thee, bleating lamb,\n I can lie down and sleep,\n Or think
 on Him who bore thy name,\n Graze after thee, and weep.\n For, washed in
 life's river,\n My bright mane for ever\n Shall shine like the gold,\n As
 I guard o'er the fold." \n \n \n SPRING\n \n Sound the flute!\n Now

it\'s mute!\n Bird\'s delight,\n Day and night,\n Nightingale,\n
 In the dale,\n Lark in sky,--\n Merrily,\n Merrily merrily, to welcome
 in the year.\n \n Little boy,\n Full of joy;\n Little girl,\n
 Sweet and small;\n Cock does crow,\n So do you;\n Merry voice,\n
 Infant noise;\n Merrily, merrily, to welcome in the year.\n \n Little
 lamb,\n Here I am;\n Come and lick\n My white neck;\n Let me
 pull\n Your soft wool;\n Let me kiss\n Your soft face;\n Merrily,
 merrily, to welcome in the year.\n \n \n NURSE\'S SONG\n \n When the voices of
 children are heard on the green,\n And laughing is heard on the hill,\n My
 heart is at rest within my breast,\n And everything else is still.\n "Then
 come home, my children, the sun is gone down,\n And the dews of night arise;\n
 Come, come, leave off play, and let us away,\n Till the morning appears in the
 skies." \n \n "No, no, let us play, for it is yet day,\n And we cannot go to
 sleep;\n Besides, in the sky the little birds fly,\n And the hills are all
 covered with sheep." \n "Well, well, go and play till the light fades away,\n
 And then go home to bed." \n The little ones leaped, and shouted, and laughed,\n
 And all the hills echoed.\n \n \n INFANT JOY\n \n "I have no name;\n I am but
 two days old." \n What shall I call thee?\n "I happy am,\n Joy is my name." \n
 Sweet joy befall thee!\n \n Pretty joy!\n Sweet joy, but two days old.\n Sweet
 Joy I call thee:\n Thou dost smile,\n I sing the while;\n Sweet joy befall
 thee!\n \n \n A DREAM\n \n Once a dream did weave a shade\n O\'er my angel-
 guarded bed,\n That an emmet lost its way\n Where on grass methought I lay.\n \n \n
 Troubled, wildered, and forlorn,\n Dark, benighted, travel-worn,\n Over many a
 tangle spray,\n All heart-broke, I heard her say:\n \n "Oh my children! do they
 cry,\n Do they hear their father sigh?\n Now they look abroad to see,\n Now
 return and weep for me." \n \n Pitying, I dropped a tear:\n But I saw a glow-worm
 near,\n Who replied, "What wailing wight\n Calls the watchman of the night?\n \n \n
 "I am set to light the ground,\n While the beetle goes his round:\n Follow now
 the beetle\'s hum;\n Little wanderer, hie thee home!" \n \n \n ON ANOTHER\'S
 SORROW\n \n Can I see another\'s woe,\n And not be in sorrow too?\n Can I see
 another\'s grief,\n And not seek for kind relief?\n \n Can I see a falling
 tear,\n And not feel my sorrow\'s share?\n Can a father see his child\n Weep,
 nor be with sorrow filled?\n \n Can a mother sit and hear\n An infant groan, an
 infant fear?\n No, no! never can it be!\n Never, never can it be!\n \n And can
 He who smiles on all\n Hear the wren with sorrows small,\n Hear the small
 bird\'s grief and care,\n Hear the woes that infants bear --\n \n And not sit
 beside the next,\n Pouring pity in their breast,\n And not sit the cradle
 near,\n Weeping tear on infant\'s tear?\n \n And not sit both night and day,\n
 Wiping all our tears away?\n Oh no! never can it be!\n Never, never can it be!\n \n
 \n He doth give his joy to all:\n He becomes an infant small,\n He becomes a man
 of woe,\n He doth feel the sorrow too.\n \n Think not thou canst sigh a sigh,\n
 And thy Maker is not by:\n Think not thou canst weep a tear,\n And thy Maker is
 not year.\n \n Oh He gives to us his joy,\n That our grief He may destroy:\n
 Till our grief is fled an gone\n He doth sit by us and moan.\n \n \n \n SONGS OF
 EXPERIENCE\n \n \n INTRODUCTION\n \n Hear the voice of the Bard,\n Who present,
 past, and future, sees;\n Whose ears have heard\n The Holy Word\n That walked
 among the ancient tree;\n \n Calling the lapsed soul,\n And weeping in the

evening dew;\n That might control\n The starry pole,\n And fallen, fallen light
 renew!\n \n "O Earth, O Earth, return!\n Arise from out the dewy grass!\n Night
 is worn,\n And the morn\n Rises from the slumbrous mass.\n \n "Turn away no
 more;\n Why wilt thou turn away?\n The starry floor,\n The watery shore,\n Are
 given thee till the break of day." \n \n \n EARTH'S ANSWER\n \n Earth raised up
 her head\n From the darkness dread and drear,\n Her light fled,\n Stony,
 dread,\n And her locks covered with grey despair.\n \n "Prisoned on watery
 shore,\n Starry jealousy does keep my den\n Cold and hoar;\n Weeping o're,\n I
 hear the father of the ancient men.\n \n "Selfish father of men!\n Cruel,
 jealous, selfish fear!\n Can delight,\n Chained in night,\n The virgins of youth
 and morning bear?\n \n \n "Does spring hide its joy,\n When buds and blossoms
 grow?\n Does the sower \n Sow by night,\n Or the plowman in darkness plough?\n
 \n "Break this heavy chain,\n That does freeze my bones around!\n Selfish,
 vain,\n Eternal bane,\n That free love with bondage bound." \n \n \n THE CLOUD AND
 THE PEBBLE\n \n "Love seeketh not itself to please,\n Nor for itself hath any
 care,\n But for another gives it ease,\n And builds a heaven in hell's
 despair." \n \n So sang a little cloud of clay,\n Trodden with the cattle's
 feet,\n But a pebble of the brook\n Warbled out these metres meet:\n \n "Love
 seeketh only Self to please,\n To bind another to its delight,\n Joys in
 another's loss of ease,\n And builds a hell in heaven's despite." \n \n \n
 HOLY THURSDAY\n \n Is this a holy thing to see\n In a rich and fruitful land,
 --\n Babes reduced to misery,\n Fed with cold and usurious hand?\n \n Is that
 trembling cry a song?\n Can it be a song of joy?\n And so many children
 poor?\n It is a land of poverty!\n \n And their sun does never shine,\n And
 their fields are bleak and bare,\n And their ways are filled with thorns:\n It
 is eternal winter there.\n \n For where'er the sun does shine,\n And
 where'er the rain does fall,\n Babes should never hunger there,\n Nor poverty
 the mind appall.\n \n \n THE LITTLE GIRL LOST\n \n In futurity\n I prophetic
 see\n That the earth from sleep\n (Grave the sentence deep)\n \n Shall arise,
 and seek\n for her Maker meek;\n And the desert wild\n Become a garden mild.\n
 \n In the southern clime,\n Where the summer's prime\n Never fades away,\n
 Lovely Lyca lay.\n \n Seven summers old\n Lovely Lyca told.\n She had wandered
 long,\n Hearing wild birds' song.\n \n "Sweet sleep, come to me\n Underneath
 this tree;\n Do father, mother, weep?\n Where can Lyca sleep?\n \n "Lost in
 desert wild\n Is your little child.\n How can Lyca sleep\n If her mother weep?\n
 \n "If her heart does ache,\n Then let Lyca wake;\n If my mother sleep,\n Lyca
 shall not weep.\n \n "Frowning, frowning night,\n O'er this desert bright\n Let
 thy moon arise,\n While I close my eyes." \n \n Sleeping Lyca lay\n While the
 beasts of prey,\n Come from caverns deep,\n Viewed the maid asleep.\n \n The
 kingly lion stood,\n And the virgin viewed:\n Then he gambolled round\n O'er
 the hallowed ground.\n \n Leopards, tigers, play\n Round her as she lay;\n While
 the lion old\n Bowed his mane of gold,\n \n And her breast did lick\n And upon
 her neck,\n From his eyes of flame,\n Ruby tears there came;\n \n While the
 lioness\n Loosed her slender dress,\n And naked they conveyed\n To caves the
 sleeping maid.\n \n \n THE LITTLE GIRL FOUND\n \n All the night in woe\n Lyca's
 parents go\n Over valleys deep,\n While the deserts weep.\n \n Tired and woe-
 begone,\n Hoarse with making moan,\n Arm in arm, seven days\n They traced the

desert ways.\n \n Seven nights they sleep\n Among shadows deep,\n And dream they
 see their child\n Starved in desert wild.\n \n Pale through pathless ways\n The
 fancied image strays,\n Famished, weeping, weak,\n With hollow piteous shriek.\n
 \n Rising from unrest,\n The trembling woman presse\n With feet of weary woe;\n
 She could no further go.\n \n In his arms he bore\n Her, armed with sorrow
 sore;\n Till before their way\n A couching lion lay.\n \n Turning back was
 vain:\n Soon his heavy mane\n Bore them to the ground,\n Then he stalked
 around,\n \n Smelling to his prey;\n But their fears allay\n When he licks their
 hands,\n And silent by them stands.\n \n They look upon his eyes,\n Filled with
 deep surprise;\n And wondering behold\n A spirit armed in gold.\n \n On his head
 a crown,\n On his shoulders down\n Flowed his golden hair.\n Gone was all their
 care.\n \n "Follow me," he said;\n "Weep not for the maid;\n In my palace
 deep,\n Lyca lies asleep." \n Then they followed\n Where the vision led,\n And
 saw their sleeping child\n Among tigers wild.\n \n To this day they dwell\n In a
 lonely dell,\n Nor fear the wolvish howl\n Nor the lion\'s growl.\n \n \n THE
 CHIMNEY SWEEPER\n \n A little black thing in the snow,\n Crying "weep! weep!" in
 notes of woe!\n "Where are thy father and mother? Say!"--\n "They are both gone
 up to the church to pray.\n \n "Because I was happy upon the heath,\n And smiled
 among the winter\'s snow,\n They clothed me in the clothes of death,\n And
 taught me to sing the notes of woe.\n \n "And because I am happy and dance and
 sing,\n They think they have done me no injury,\n And are gone to praise God and
 his priest and king,\n Who make up a heaven of our misery." \n \n \n NURSE\'S
 SONG\n \n When voices of children are heard on the green,\n And whisperings are
 in the dale,\n The days of my youth rise fresh in my mind,\n My face turns green
 and pale.\n \n Then come home, my children, the sun is gone down,\n And the dews
 of night arise;\n Your spring and your day are wasted in play,\n And your winter
 and night in disguise.\n \n \n \n THE SICK ROSE\n \n O rose, thou art sick!\n The
 invisible worm,\n That flies in the night,\n In the howling storm,\n \n Has
 found out thy bed\n Of crimson joy,\n And his dark secret love\n Does thy
 life destroy.\n \n \n \n THE FLY\n \n Little Fly,\n Thy summer\'s play\n My
 thoughtless hand\n Has brushed away.\n \n Am not I\n A fly like thee?\n Or art
 not thou\n A man like me?\n \n For I dance\n And drink, and sing,\n Till some
 blind hand\n Shall brush my wing.\n \n If thought is life\n And strength and
 breath\n And the want \n Of thought is death;\n \n Then am I\n A happy fly,\n If
 I live,\n Or if I die.\n \n \n \n THE ANGEL\n \n I dreamt a dream! What can it
 mean?\n And that I was a maiden Queen\n Guarded by an Angel mild:\n Witless woe
 was ne\'er beguiled!\n \n And I wept both night and day,\n And he wiped my tears
 away;\n And I wept both day and night,\n And hid from him my heart\'s delight.\n
 \n So he took his wings, and fled;\n Then the morn blushed rosy red.\n I dried
 my tears, and armed my fears\n With ten-thousand shields and spears.\n \n \n Soon
 my Angel came again;\n I was armed, he came in vain;\n For the time of youth was
 fled,\n And grey hairs were on my head.\n \n \n \n THE TIGER\n \n Tiger, tiger,
 burning bright\n In the forest of the night,\n What immortal hand or eye\n Could
 Frame thy fearful symmetry?\n \n In what distant deeps or skies\n Burnt the fire
 of thine eyes?\n On what wings dare he aspire?\n What the hand dare seize the
 fire?\n \n And what shoulder and what art\n Could twist the sinews of thy
 heart?\n And, when thy heart began to beat,\n What dread hand and what dread

feet?\n \n What the hammer? what the chain?\n In what furnace was thy brain?\n
 What the anvil? what dread grasp\n Dare its deadly terrors clasp?\n \n When the
 stars threw down their spears,\n And watered heaven with their tears,\n Did he
 smile his work to see?\n Did he who made the lamb make thee?\n \n Tiger, tiger,
 burning bright\n In the forests of the night,\n What immortal hand or eye\n Dare
 frame thy fearful symmetry?\n \n \n MY PRETTY ROSE TREE\n \n A flower was
 offered to me,\n Such a flower as May never bore;\n But I said "I've a pretty
 rose tree," \n And I passed the sweet flower o'er.\n \n Then I went to my
 pretty rose tree,\n To tend her by day and by night;\n But my rose turned away
 with jealousy,\n And her thorns were my only delight.\n \n \n AH SUNFLOWER\n
 \n Ah Sunflower, weary of time,\n Who countest the steps of the sun;\n Seeking
 after that sweet golden clime\n Where the traveller's journey is done;\n \n
 Where the Youth pined away with desire,\n And the pale virgin shrouded in
 snow,\n Arise from their graves, and aspire\n Where my Sunflower wishes to
 go!\n \n \n THE LILY\n \n The modest Rose puts forth a thorn,\n The humble sheep
 a threat'\ning horn:\n While the Lily white shall in love delight,\n Nor a thorn
 nor a threat stain her beauty bright.\n \n \n THE GARDEN OF LOVE\n \n I laid me
 down upon a bank,\n Where Love lay sleeping;\n I heard among the rushes dank\n
 Weeping, weeping.\n \n Then I went to the heath and the wild,\n To the
 thistles and thorns of the waste;\n And they told me how they were beguiled,\n
 Driven out, and compelled to the chaste.\n \n I went to the Garden of Love,\n
 And saw what I never had seen;\n A Chapel was built in the midst,\n Where I
 used to play on the green.\n \n And the gates of this Chapel were shut\n And
 "Thou shalt not," writ over the door;\n So I turned to the Garden of Love\n
 That so many sweet flowers bore.\n \n And I saw it was filled with graves,\n
 And tombstones where flowers should be;\n And priests in black gowns were
 walking their rounds,\n And binding with briars my joys and desires.\n \n \n
 THE LITTLE VAGABOND\n \n Dear mother, dear mother, the Church is cold;\n But the
 Alehouse is healthy, and pleasant, and warm.\n Besides, I can tell where I am
 used well;\n The poor parsons with wind like a blown bladder swell.\n \n But, if
 at the Church they would give us some ale,\n And a pleasant fire our souls to
 regale,\n We'd sing and we'd pray all the livelong day,\n Nor ever once wish
 from the Church to stray.\n \n Then the Parson might preach, and drink, and
 sing,\n And we'd be as happy as birds in the spring;\n And modest Dame Lurch,
 who is always at church,\n Would not have bandy children, nor fasting, nor
 birch.\n \n And God, like a father, rejoicing to see\n His children as pleasant
 and happy as he,\n Would have no more quarrel with the Devil or the barrel,\n
 But kiss him, and give him both drink and apparel.\n \n \n LONDON\n \n I
 wandered through each chartered street,\n Near where the chartered Thames does
 flow,\n A mark in every face I meet,\n Marks of weakness, marks of woe.\n \n
 In every cry of every man,\n In every infant's cry of fear,\n In every voice,
 in every ban,\n The mind-forged manacles I hear:\n \n How the chimney-
 sweeper's cry\n Every blackening church appals,\n And the hapless soldier's
 sigh\n Runs in blood down palace-walls.\n \n But most, through midnight
 streets I hear\n How the youthful harlot's curse\n Blasts the new-born
 infant's tear,\n And blights with plagues the marriage-hearse.\n \n \n THE
 HUMAN ABSTRACT\n \n Pity would be no more\n If we did not make somebody poor,\n

And Mercy no more could be\n If all were as happy as we.\n \n And mutual fear
 brings Peace,\n Till the selfish loves increase\n Then Cruelty knits a snare,\n
 And spreads his baits with care.\n \n He sits down with his holy fears,\n And
 waters the ground with tears;\n Then Humility takes its root\n Underneath his
 foot.\n \n Soon spreads the dismal shade\n Of Mystery over his head,\n And the
 caterpillar and fly\n Feed on the Mystery.\n \n And it bears the fruit of
 Deceit,\n Ruddy and sweet to eat,\n And the raven his nest has made\n In its
 thickest shade.\n \n The gods of the earth and sea\n Sought through nature to
 find this tree,\n But their search was all in vain:\n There grows one in the
 human Brain.\n \n \n INFANT SORROW\n \n My mother groaned, my father wept:\n
 Into the dangerous world I leapt,\n Helpless, naked, piping loud,\n Like a fiend
 hid in a cloud.\n \n Struggling in my father\'s hands,\n Striving against my
 swaddling-bands,\n Bound and weary, I thought best\n To sulk upon my mother\'s
 breast.\n \n \n A POISON TREE\n \n I was angry with my friend:\n I told my
 wrath, my wrath did end.\n I was angry with my foe:\n I told it not, my wrath
 did grow.\n \n And I watered it in fears\n Night and morning with my tears,\n
 And I sunned it with smiles\n And with soft deceitful wiles.\n \n And it grew
 both day and night,\n Till it bore an apple bright,\n And my foe beheld it
 shine,\n and he knew that it was mine, --\n \n And into my garden stole\n When
 the night had veiled the pole;\n In the morning, glad, I see\n My foe
 outstretched beneath the tree.\n \n \n A LITTLE BOY LOST\n \n "Nought loves
 another as itself,\n Nor venerates another so,\n Nor is it possible to
 thought\n A greater than itself to know.\n \n "And, father, how can I love you
 \n Or any of my brothers more?\n I love you like the little bird\n That
 picks up crumbs around the door."\n \n The Priest sat by and heard the child;\n
 In trembling zeal he seized his hair,\n He led him by his little coat,\n And
 all admired the priestly care. \n \n And standing on the altar high,\n "Lo,
 what a fiend is here! said he:\n "One who sets reason up for judge\n Of our
 most holy mystery."\n \n The weeping child could not be heard,\n The weeping
 parents wept in vain:\n They stripped him to his little shirt,\n And bound him
 in an iron chain,\n \n And burned him in a holy place\n Where many had been
 burned before;\n The weeping parents wept in vain.\n Are such thing done on
 Albion\'s shore?\n \n \n A LITTLE GIRL LOST\n \n Children of the future age,\n
 Reading this indignant page,\n Know that in a former time\n Love, sweet love,
 was thought a crime.\n \n In the age of gold,\n Free from winter\'s cold,\n
 Youth and maiden bright,\n To the holy light,\n Naked in the sunny beams
 delight.\n \n Once a youthful pair,\n Filled with softest care,\n Met in garden
 bright\n Where the holy light\n Had just removed the curtains of the night.\n \n
 Then, in rising day,\n On the grass they play;\n Parents were afar,\n Strangers
 came not near,\n And the maiden soon forgot her fear.\n \n Tired with kisses
 sweet,\n They agree to meet\n When the silent sleep\n Waves o\'er heaven\'s
 deep,\n And the weary tired wanderers weep.\n \n To her father white\n Came the
 maiden bright;\n But his loving look,\n Like the holy book\n All her tender
 limbs with terror shook.\n \n "Ona, pale and weak,\n To thy father speak!\n Oh
 the trembling fear!\n Oh the dismal care\n That shakes the blossoms of my hoary
 hair!"\n \n \n THE SCHOOLBOY\n \n I love to rise on a summer morn, \n When
 birds are singing on every tree;\n The distant huntsman winds his horn,\n And

the skylark sings with me:\n Oh what sweet company!\n \n But to go to school
 in a summer morn, --\n Oh it drives all joy away!\n Under a cruel eye
 outworn,\n The little ones spend the day\n In sighing and dismay.\n \n Ah
 then at times I drooping sit,\n And spend many an anxious hour;\n Nor in my
 book can I take delight,\n Nor sit in learning's bower,\n Worn through with
 the dreary shower.\n \n How can the bird that is born for joy\n Sit in a cage
 and sing?\n How can a child, when fears annoy,\n But droop his tender wing,\n
 And forget his youthful spring?\n \n Oh father and mother, if buds are nipped,\n
 And blossoms blown away;\n And if the tender plants are stripped\n Of their
 joy in the springing day,\n By sorrow and care's dismay, --\n \n How shall
 the summer arise in joy,\n Or the summer fruits appear?\n Or how shall we
 gather what griefs destroy,\n Or bless the mellowing year,\n When the blasts
 of winter appear?\n \n \n TO TERZAH\n \n Whate'er is born of mortal birth\n
 Must be consumed with the earth,\n To rise from generation free:\n Then what
 have I to do with thee?\n The sexes sprang from shame and pride,\n Blown in the
 morn, in evening died;\n But mercy changed death into sleep;\n The sexes rose to
 work and weep.\n \n Thou, mother of my mortal part,\n With cruelty didst mould
 my heart,\n And with false self-deceiving tears\n Didst bind my nostrils, eyes,
 and ears,\n \n Didst close my tongue in senseless clay,\n And me to mortal life
 betray.\n The death of Jesus set me free:\n Then what have I to do with thee?\n
 \n \n THE VOICE OF THE ANCIENT BARD\n \n Youth of delight! come hither\n And
 see the opening morn,\n Image of Truth new-born.\n Doubt is fled, and clouds of
 reason,\n Dark disputes and artful teasing.\n Folly is an endless maze;\n
 Tangled roots perplex her ways;\n How many have fallen there!\n They stumble all
 night over bones of the dead;\n And feel -- they know not what but care;\n And
 wish to lead others, when they should be led.\n \n \n APPENDIX\n \n A DIVINE IMAGE\n
 \n Cruelty has a human heart,\n And Jealousy a human face;\n Terror the human
 form divine,\n And Secresy the human dress.\n \n The human dress is forged
 iron,\n The human form a fiery forge,\n The human face a furnace sealed,\n
 The human heart its hungry gorge.\n \n NOTE: Though written and engraved by
 Blake, "A DIVINE IMAGE" was never\n included in the SONGS OF INNOCENCE AND OF
 EXPERIENCE.\n \n \n \n \n \n William Blake's \n \n THE BOOK of THEL\n \n \n THEL'S
 Motto\n \n Does the Eagle know what is in the pit?\n Or wilt thou go ask the
 Mole:\n Can Wisdom be put in a silver rod?\n Or Love in a golden bowl?\n \n \n THE
 BOOK of THEL\n \n The Author & Printer Willm. Blake. 1780\n \n \n THEL\n \n I\n \n The
 daughters of Mne Seraphim led round their sunny flocks,\n All but the youngest:
 she in paleness sought the secret air.\n To fade away like morning beauty from
 her mortal day:\n Down by the river of Adona her soft voice is heard;\n And thus
 her gentle lamentation falls like morning dew.\n \n O life of this our spring!
 why fades the lotus of the water?\n Why fade these children of the spring? born
 but to smile & fall.\n Ah! Thel is like a watry bow, and like a parting
 cloud,\n Like a reflection in a glass: like shadows in the water\n Like dreams of
 infants, like a smile upon an infants face.\n Like the doves voice, like
 transient day, like music in the air:\n Ah! gentle may I lay me down and gentle
 rest my head.\n And gentle sleep the sleep of death, and gently hear the voice
 \n Of him that walketh in the garden in the evening time.\n \n The Lilly of the
 valley breathing in the humble grass\n Answerd the lovely maid and said: I am a

watry weed,\nAnd I am very small and love to dwell in lowly vales:\nSo weak the
 gilded butterfly scarce perches on my head\nYet I am visited from heaven and he
 that smiles on all\nWalks in the valley, and each morn over me spreads his
 hand\nSaying, rejoice thou humble grass, thou new-born lily flower.\nThou gentle
 maid of silent valleys and of modest brooks:\nFor thou shall be clothed in
 light, and fed with morning manna:\nTill summers heat melts thee beside the
 fountains and the springs\nTo flourish in eternal vales: they why should Thel
 complain.\nWhy should the mistress of the vales of Har, utter a sigh.\n\nShe
 ceas'd & smild in tears, then sat down in her silver shrine.\n\nThel answerd, O
 thou little virgin of the peaceful valley.\nGiving to those that cannot crave,
 the voiceless, the o'er tired\nThe breath doth nourish the innocent lamb, he
 smells the milky garments\nHe crops thy flowers while thou sittest smiling in
 his face,\nWiping his mild and meek mouth from all contagious taints.\nThy
 wine doth purify the golden honey; thy perfume.\nWhich thou dost scatter on
 every little blade of grass that springs\nRevives the milked cow, & tames the
 fire-breathing steed.\nBut Thel is like a faint cloud kindled at the rising
 sun:\nI vanish from my pearly throne, and who shall find my place.\n\nQueen of
 the vales the Lily answered, ask the tender cloud,\nAnd it shall tell thee why
 it glitters in the morning sky.\nAnd why it scatters its bright beauty thro the
 humid air.\nDescend O little cloud & hover before the eyes of Thel.\n\nThe Cloud
 descended and the Lily bow'd her modest head:\nAnd went to mind her numerous
 charge among the verdant grass.\n\nII.\n\nO little Cloud the virgin said, I
 charge thee to tell me\nWhy thou complainest now when in one hour thou fade
 away:\nThen we shall seek thee but not find: ah Thel is like to thee.\nI pass
 away, yet I complain, and no one hears my voice.\n\nThe Cloud then shew'd his
 golden head & his bright form emerg'd.\nHovering and glittering on the air
 before the face of Thel.\n\nO virgin know'st thou not our steeds drink of the
 golden springs\nWhere Luvah doth renew his horses: lookst thou on my youth.\nAnd
 fearest thou because I vanish and am seen no more.\nNothing remains; O maid I
 tell thee, when I pass away.\nIt is to tenfold life, to love, to peace, and
 raptures holy:\nUnseen descending, weigh my light wings upon balmy flowers:\nAnd
 court the fair eyed dew, to take me to her shining tent\nThe weeping virgin,
 trembling kneels before the risen sun.\nTill we arise link'd in a golden band
 and never part:\nBut walk united bearing food to all our tender flowers.\n\nDost
 thou O little cloud? I fear that I am not like thee:\nFor I walk through the
 vales of Har, and smell the sweetest flowers:\nBut I feed not the little
 flowers: I hear the warbling birds,\nBut I feed not the warbling birds, they fly
 and seek their food:\nBut Thel delights in these no more because I fade
 away\nAnd all shall say, without a use this shining woman liv'd,\nOr did she
 only live to be at death the food of worms.\n\nThe Cloud reclind upon his airy
 throne and answerd thus.\n\nThen if thou art the food of worms, O virgin of the
 skies,\nHow great thy use, how great thy blessing, every thing that
 lives.\nLives not alone nor of itself: fear not and I will call,\nThe weak worm
 from its lowly bed, and thou shalt hear its voice.\nCome forth worm and the
 silent valley, to thy pensive queen.\n\nThe helpless worm arose and sat upon the
 Lillys leaf,\nAnd the bright Cloud sail'd on, to find his partner in the
 vale.\n\nIII.\n\nThen Thel astonish'd view'd the Worm upon its dewy

bed.\n\nArt thou a Worm? image of weakness. art thou but a Worm?\nI see thee
 like an infant wrapped in the Lillys leaf;\nAh weep not little voice, thou
 can\'st not speak, but thou can\'st weep:\nIs this a Worm? I see they lay
 helpless & naked: weeping\nAnd none to answer, none to cherish thee with mothers
 smiles.\n\nThe Clod of Clay heard the Worms voice & rais\'d her pitying
 head:\nShe bowd over the weeping infant, and her life exhald\nIn milky fondness,
 then on Thel she fix\'d her humble eyes\n\nO beauty of the vales of Har, we live
 not for ourselves,\nThou seest me the meanest thing, and so I am indeed:\nMy
 bosom of itself is cold, and of itself is dark,\n\nBut he that loves the lowly,
 pours his oil upon my head\nAnd kisses me, and binds his nuptial bands around my
 breast.\nAnd says; Thou mother of my children, I have loved thee\nAnd I have
 given thee a crown that none can take away.\nBut how this is sweet maid, I know
 not, and I cannot know\nI ponder, and I cannot ponder; yet I live and
 love.\n\nThe daughter of beauty wip\'d her pitying tears with her white
 veil,\nAnd said, Alas! I knew not this, and therefore did I weep:\nThat God
 would love a Worm I knew, and punish the evil foot\nThat wilful bruis\'d its
 helpless form: but that he cherish\'d it\nWith milk and oil I never knew, and
 therefore did I weep,\nAnd I complaind in the mild air, because I fade
 away.\nAnd lay me down in thy cold bed, and leave my shining lot.\n\nQueen of
 the vales, the matron Clay answered: I heard thy sighs.\nAnd all thy moans flew
 o\'er my roof, but I have call\'d them down:\nWilt thou O Queen enter my house,
 tis given thee to enter,\nAnd to return: fear nothing, enter with thy virgin
 feet.\n\n\nIV.\n\nThe eternal gates terrific porter lifted the northern
 bar:\nThel enter\'d in & saw the secrets of the land unknown;\nShe saw the
 couches of the dead, & where the fibrous roots\nOf every heart on earth infixes
 deep its restless twists:\nA land of sorrows & of tears where never smile was
 seen.\n\nShe wandered in the land of clouds thro\' valleys dark,
 listning\nDolors & lamentations: waiting oft beside the dewy grave\nShe stood in
 silence, listning to the voices of the ground,\nTill to her own grave plot she
 came, & there she sat down.\nAnd heard this voice of sorrow breathed from the
 hollow pit.\n\nWhy cannot the Ear be closed to its own destruction?\nOr the
 glistening Eye to the poison of a smile!\nWhy are Eyelids stord with arrows
 ready drawn,\nWhere a thousand fighting men in ambush lie!\nOr an Eye of gifts &
 graces showing fruits & coined gold!\n\nWhy a Tongue impress\'d with honey from
 every wind?\nWhy an Ear, a whirlpool fierce to draw creations in?\nWhy a Nostril
 wide inhaling terror trembling & affright\nWhy a tender curb upon the youthful
 burning boy?\nWhy a little curtain of flesh on the bed of our desire?\n\nThe
 Virgin started from her seat, & with a shriek,\nFled back unhinderd till she
 came into the vales of Har\n\n\n'

```
[11]: blake_sentences[100:150]
```

```
[11]: [['Then',
        'cherish',
        'pity',
        ',',
        'lest',
```

'you',
'drive',
'an',
'angel',
'from',
'your',
'door',
'.''],
['NIGHT'],
['The',
'sun',
'descending',
'in',
'the',
'west',
'',
'The',
'evening',
'star',
'does',
'shine',
';',
'The',
'birds',
'are',
'silent',
'in',
'their',
'nest',
'',
'And',
'I',
'must',
'seek',
'for',
'mine',
'.''],
['The',
'moon',
'',
'like',
'a',
'flower',
'In',
'heaven',
'''],
's',

'high',
'bower',
,',
'With',
'silent',
'delight',
,',
'Sits',
'and',
'smiles',
'on',
'the',
'night',
'.''],
['Farewell',
,',
'green',
'fields',
'and',
'happy',
'grove',
,',
'Where',
'flocks',
'have',
'ta',
"''",
'en',
'delight',
'.''],
['Where',
'lambs',
'have',
'nibbled',
,',
'silent',
'move',
'The',
'feet',
'of',
'angels',
'bright',
,',
'Unseen',
'they',
'pour',
'blessing',

'',
'And',
'joy',
'without',
'ceasing',
'',
'On',
'each',
'bud',
'and',
'blossom',
'',
'And',
'each',
'sleeping',
'bosom',
'.''],
['They',
'look',
'in',
'every',
'thoughtless',
'nest',
'Where',
'birds',
'are',
'covered',
'warm',
';',
'They',
'visit',
'caves',
'of',
'every',
'beast',
'',
'To',
'keep',
'them',
'all',
'from',
'harm',
':',
'If',
'they',
'see',
'any',

'weeping',
'That',
'should',
'have',
'been',
'sleeping',
'',
'They',
'pour',
'sleep',
'on',
'their',
'head',
'',
'And',
'sit',
'down',
'by',
'their',
'bed',
'.''],
['When',
'wolves',
'and',
'tigers',
'howl',
'for',
'prey',
'',
'They',
'pitying',
'stand',
'and',
'weep',
';',
'Seeking',
'to',
'drive',
'their',
'thirst',
'away',
'',
'And',
'keep',
'them',
'from',
'the',

'sheep',
 '.'],
 ['But',
 ',',
 'if',
 'they',
 'rush',
 'dreadful',
 ',',
 'The',
 'angels',
 ',',
 'most',
 'heedful',
 ',',
 'Receive',
 'each',
 'mild',
 'spirit',
 ',',
 'New',
 'worlds',
 'to',
 'inherit',
 '.'],
 ['And',
 'there',
 'the',
 'lion',
 '"',
 's',
 'ruddy',
 'eyes',
 'Shall',
 'flow',
 'with',
 'tears',
 'of',
 'gold',
 ':',
 'And',
 'pitying',
 'the',
 'tender',
 'cries',
 ',',
 'And',

'walking',
'round',
'the',
'fold',
':',
'Saying',
':',
'"',
'Wrath',
'by',
'His',
'meekness',
'',
'And',
'',
'by',
'His',
'health',
'',
'sickness',
'',
'Are',
'driven',
'away',
'From',
'our',
'immortal',
'day',
'.''],
['"',
'And',
'now',
'beside',
'thee',
'',
'bleating',
'lamb',
'',
'I',
'can',
'lie',
'down',
'and',
'sleep',
'',
'Or',
'think',

'on',
 'Him',
 'who',
 'bore',
 'thy',
 'name',
 ',',
 'Graze',
 'after',
 'thee',
 ',',
 'and',
 'weep',
 '.'],
 ['For',
 ',',
 'washed',
 'in',
 'life',
 '"',
 's',
 'river',
 ',',
 'My',
 'bright',
 'mane',
 'for',
 'ever',
 'Shall',
 'shine',
 'like',
 'the',
 'gold',
 ',',
 'As',
 'I',
 'guard',
 'o',
 '"',
 'er',
 'the',
 'fold',
 '.'],
 ['SPRING'],
 ['Sound', 'the', 'flute', '!'],
 ['Now', 'it', '"', 's', 'mute', '!'],
 ['Bird',

'"',
 's',
 'delight',
 ', ',
 'Day',
 'and',
 'night',
 ', ',
 'Nightingale',
 ', ',
 'In',
 'the',
 'dale',
 ', ',
 'Lark',
 'in',
 'sky',
 ', --',
 'Merrily',
 ', ',
 'Merrily',
 'merrily',
 ', ',
 'to',
 'welcome',
 'in',
 'the',
 'year',
 '. '],
 ['Little',
 'boy',
 ', ',
 'Full',
 'of',
 'joy',
 '; ',
 'Little',
 'girl',
 ', ',
 'Sweet',
 'and',
 'small',
 '; ',
 'Cock',
 'does',
 'crow',
 ', ',

'So',
'do',
'you',
';',
'Merry',
'voice',
',',
'Infant',
'noise',
';',
'Merrily',
',',
'merrily',
',',
'to',
'welcome',
'in',
'the',
'year',
'.'],
['Little',
'lamb',
',',
'Here',
'I',
'am',
';',
'Come',
'and',
'lick',
'My',
'white',
'neck',
';',
'Let',
'me',
'pull',
'Your',
'soft',
'wool',
';',
'Let',
'me',
'kiss',
'Your',
'soft',
'face',


```

';',
'Merrily',
',',
'merrily',
',',
'to',
'welcome',
'in',
'the',
'year',
'.'],
['NURSE', '"', 'S', 'SONG'],
['When',
'the',
'voices',
'of',
'children',
'are',
'heard',
'on',
'the',
'green',
',',
'And',
'laughing',
'is',
'heard',
'on',
'the',
'hill',
',',
'My',
'heart',
'is',
'at',
'rest',
'within',
'my',
'breast',
',',
'And',
'everything',
'else',
'is',
'still',
'.'],
['"',

```

'Then',
'come',
'home',
'',
'my',
'children',
'',
'the',
'sun',
'is',
'gone',
'down',
'',
'And',
'the',
'dews',
'of',
'night',
'arise',
';',
'Come',
'',
'come',
'',
'leave',
'off',
'play',
'',
'and',
'let',
'us',
'away',
'',
'Till',
'the',
'morning',
'appears',
'in',
'the',
'skies',
'."'],
['',
'No',
'',
'no',
'',
'let',

'us',
'play',
'',
'for',
'it',
'is',
'yet',
'day',
'',
'And',
'we',
'cannot',
'go',
'to',
'sleep',
';',
'Besides',
'',
'in',
'the',
'sky',
'the',
'little',
'birds',
'fly',
'',
'And',
'the',
'hills',
'are',
'all',
'covered',
'with',
'sheep',
'."'],
['',
'Well',
'',
'well',
'',
'go',
'and',
'play',
'till',
'the',
'light',
'fades',

'away',
 ',',
 'And',
 'then',
 'go',
 'home',
 'to',
 'bed',
 '."'],
 ['The',
 'little',
 'ones',
 'leaped',
 ',',
 'and',
 'shouted',
 ',',
 'and',
 'laughed',
 ',',
 'And',
 'all',
 'the',
 'hills',
 'echoed',
 '.'],
 ['INFANT', 'JOY'],
 ['"',
 'I',
 'have',
 'no',
 'name',
 ';',
 'I',
 'am',
 'but',
 'two',
 'days',
 'old',
 '."'],
 ['What', 'shall', 'I', 'call', 'thee', '?'],
 ['"', 'I', 'happy', 'am', ',', 'Joy', 'is', 'my', 'name', '.'],
 ['Sweet', 'joy', 'befall', 'thee', '!'],
 ['Pretty', 'joy', '!'],
 ['Sweet', 'joy', ',', 'but', 'two', 'days', 'old', '.'],
 ['Sweet',
 'Joy',

'I',
'call',
'thee',
':',
'Thou',
'dost',
'smile',
'',
'I',
'sing',
'the',
'while',
';',
'Sweet',
'joy',
'befall',
'thee',
'!'],
['A', 'DREAM'],
['Once',
'a',
'dream',
'did',
'weave',
'a',
'shade',
'O',
""',
'er',
'my',
'angel',
'-',
'guarded',
'bed',
'',
'That',
'an',
'emmet',
'lost',
'its',
'way',
'Where',
'on',
'grass',
'methought',
'I',
'lay',

```

'.'],
['Troubled',
',',
'wildered',
',',
'and',
'forlorn',
',',
'Dark',
',',
'benighted',
',',
'travel',
'-',
'worn',
',',
'Over',
'many',
'a',
'tangle',
'spray',
',',
'All',
'heart',
'-',
'broke',
',',
'I',
'heard',
'her',
'say',
':'],
['"', 'Oh', 'my', 'children', '!'],
['do',
'they',
'cry',
',',
'Do',
'they',
'hear',
'their',
'father',
'sigh',
'?'],
['Now',
'they',
'look',

```

'abroad',
'to',
'see',
,',
'Now',
'return',
'and',
'weep',
'for',
'me',
'."'],
['Pitying',
,',
'I',
'dropped',
'a',
'tear',
,':',
'But',
'I',
'saw',
'a',
'glow',
'-',
'worm',
'near',
,',
'Who',
'replied',
,',
'"',
'What',
'wailing',
'wight',
'Calls',
'the',
'watchman',
'of',
'the',
'night',
'?'],
['"',
'I',
'am',
'set',
'to',
'light',

'the',
 'ground',
 ',',
 'While',
 'the',
 'beetle',
 'goes',
 'his',
 'round',
 ':',
 'Follow',
 'now',
 'the',
 'beetle',
 '"',
 's',
 'hum',
 ';',
 'Little',
 'wanderer',
 ',',
 'hie',
 'thee',
 'home',
 '!"'],
 ['ON', 'ANOTHER', '"', 'S', 'SORROW'],
 ['Can',
 'I',
 'see',
 'another',
 '"',
 's',
 'woe',
 ',',
 'And',
 'not',
 'be',
 'in',
 'sorrow',
 'too',
 '?'],
 ['Can',
 'I',
 'see',
 'another',
 '"',
 's',

'grief',
'',
'And',
'not',
'seek',
'for',
'kind',
'relief',
'?'],
['Can',
'I',
'see',
'a',
'falling',
'tear',
'',
'And',
'not',
'feel',
'my',
'sorrow',
''',
's',
'share',
'?'],
['Can',
'a',
'father',
'see',
'his',
'child',
'Weep',
'',
'nor',
'be',
'with',
'sorrow',
'filled',
'?'],
['Can',
'a',
'mother',
'sit',
'and',
'hear',
'An',
'infant',

```

'groan',
',',
'an',
'infant',
'fear',
'?'],
['No', ',', 'no', '!'],
['never', 'can', 'it', 'be', '!'],
['Never', ',', 'never', 'can', 'it', 'be', '!'],
['And',
'can',
'He',
'who',
'smiles',
'on',
'all',
'Hear',
'the',
'wren',
'with',
'sorrows',
'small',
',',
'Hear',
'the',
'small',
'bird',
'""',
's',
'grief',
'and',
'care',
',',
'Hear',
'the',
'woes',
'that',
'infants',
'bear',
'--']]

```

```

[12]: longest_len = max([len(s) for s in blake_sentences])
      print(longest_len)
      [s for s in blake_sentences if len(s) == longest_len]

```

```
[12]: [['I',  
      'fear',  
      'that',  
      'I',  
      'am',  
      'not',  
      'like',  
      'thee',  
      ':',  
      'For',  
      'I',  
      'walk',  
      'through',  
      'the',  
      'vales',  
      'of',  
      'Har',  
      ',',  
      'and',  
      'smell',  
      'the',  
      'sweetest',  
      'flowers',  
      ':',  
      'But',  
      'I',  
      'feed',  
      'not',  
      'the',  
      'little',  
      'flowers',  
      ':',  
      'I',  
      'hear',  
      'the',  
      'warbling',  
      'birds',  
      ',',  
      'But',  
      'I',  
      'feed',  
      'not',  
      'the',  
      'warbling',  
      'birds',  
      ',',  
      'they',
```

'fly',
'and',
'seek',
'their',
'food',
':',
'But',
'Thel',
'delights',
'in',
'these',
'no',
'more',
'because',
'I',
'fade',
'away',
'And',
'all',
'shall',
'say',
'',
'without',
'a',
'use',
'this',
'shining',
'women',
'liv',
'''',
'd',
'',
'Or',
'did',
'she',
'only',
'live',
'to',
'be',
'at',
'death',
'the',
'food',
'of',
'worms',
'.']]

#Web and Chat Text Corpus

```
[13]: from nltk.corpus import webtext
      nltk.download('webtext')
      nltk.corpus.webtext.fileids()
```

```
[nltk_data] Downloading package webtext to
[nltk_data]   /Users/pushpakulkarni/nltk_data...
[nltk_data]   Package webtext is already up-to-date!
```

```
[13]: ['firefox.txt',
      'grail.txt',
      'overheard.txt',
      'pirates.txt',
      'singles.txt',
      'wine.txt']
```

```
[14]: firefox_text=webtext.sents('firefox.txt')
      firefox_text
```

```
[14]: [['Cookie', 'Manager', ':', '"', 'Don', '"', 't', 'allow', 'sites', 'that',
      'set', 'removed', 'cookies', 'to', 'set', 'future', 'cookies', '"', 'should',
      'stay', 'checked', 'When', 'in', 'full', 'screen', 'mode', 'Pressing', 'Ctrl',
      '-', 'N', 'should', 'open', 'a', 'new', 'browser', 'when', 'only', 'download',
      'dialog', 'is', 'left', 'open', 'add', 'icons', 'to', 'context', 'menu', 'So',
      'called', '"', 'tab', 'bar', '"', 'should', 'be', 'made', 'a', 'proper',
      'toolbar', 'or', 'given', 'the', 'ability', 'collapse', '/', 'expand', '.'],
      [['', 'XUL', ''], 'Implement', 'Cocoa', '-', 'style', 'toolbar',
      'customization', '.'], ...]
```

```
[15]: pirates_sentences=webtext.sents('pirates.txt')
      pirates_sentences[:65]
```

```
[15]: [['PIRATES',
      'OF',
      'THE',
      'CARRIBEAN',
      ':',
      'DEAD',
      'MAN',
      '"',
      'S',
      'CHEST',
      ', ',
      'by',
      'Ted',
      'Elliott',
      '&'],
```

'Terry',
'Rossio',
['',
'view',
'looking',
'straight',
'down',
'at',
'rolling',
'swells',
'',
'sound',
'of',
'wind',
'and',
'thunder',
'',
'then',
'a',
'low',
'heartbeat',
'],
'Scene',
':',
'PORT',
'ROYAL',
['',
'teacups',
'on',
'a',
'table',
'in',
'the',
'rain',
'],
['',
'sheet',
'music',
'on',
'music',
'stands',
'in',
'the',
'rain',
'],
['',
'bouquet',

'of',
'white',
'orchids',
'',
'Elizabeth',
'sitting',
'in',
'the',
'rain',
'holding',
'the',
'bouquet',
']',
'[',
'men',
'rowing',
'',
'men',
'on',
'horseback',
'',
'to',
'the',
'sound',
'of',
'thunder',
']',
'[',
'EITC',
'logo',
'on',
'flag',
'blowing',
'in',
'the',
'wind',
']',
'[',
'many',
'rowboats',
'are',
'entering',
'the',
'harbor',
']',
'[',
'Elizabeth',

'sitting',
'alone',
,',
'at',
'a',
'distance',
'],'
'[',
'marines',
'running',
,',
'kick',
'a',
'door',
'in',
'],'
'[',
'a',
'mule',
'is',
'seen',
'on',
'the',
'left',
'in',
'the',
'barn',
'where',
'the',
'marines',
'enter',
'],'
'[',
'Liz',
'looking',
'over',
'her',
'shoulder',
'],'
'[',
'Elizabeth',
'drops',
'her',
'bouquet',
'],'
'[',
'Will',

'is',
 'in',
 'manacles',
 ',',
 'being',
 'escorted',
 'by',
 'red',
 'coats',
 ']',
 'ELIZABETH',
 'SWANN',
 ':',
 'Will',
 '...!'],
 ['[',
 'Elizabeth',
 'runs',
 'to',
 'Will',
 ']',
 'ELIZABETH',
 'SWANN',
 ':',
 'Why',
 'is',
 'this',
 'happening',
 '?'],
 ['WILL', 'TURNER', ':', 'I', 'don', '"', 't', 'know', '.'],
 ['You', 'look', 'beautiful', '.'],
 ['ELIZABETH',
 'SWANN',
 ':',
 'I',
 'think',
 'it',
 '"',
 's',
 'bad',
 'luck',
 'for',
 'the',
 'groom',
 'to',
 'see',
 'the',

'bride',
'before',
'the',
'wedding',
'.''],
['[',
'marines',
'cross',
'their',
'long',
'axes',
'to',
'bar',
'Governor',
'from',
'entering',
']',
'[',
'Beckett',
,',
'in',
'white',
'hair',
'and',
'curls',
,',
'is',
'standing',
'with',
'Mercer',
']',
'LORD',
'CUTLER',
'BECKETT',
,':',
'Governor',
'Weatherby',
'Swann',
,',
'it',
'"',
's',
'been',
'too',
'long',
'.''],
['LORD',

'CUTLER',
 'BECKETT',
 ':',
 'His',
 'Lord',
 'now',
 '...',
 'actually',
 '.'],
 ['LORD', 'CUTLER', 'BECKETT', ':', 'In', 'fact', ',', 'I', '*', 'do', '*.'],
 ['Mister', 'Mercer', '!'],
 ['The',
 'warrant',
 'for',
 'the',
 'arrest',
 'of',
 'one',
 'William',
 'Turner',
 '.'],
 ['LORD', 'CUTLER', 'BECKETT', ':', 'Oh', ',', 'is', 'it', '?'],
 ['That', '"', 's', 'annoying', '.'],
 ['My', 'mistake', '.'],
 ['Arrest', 'her', '.'],
 ['ELIZABETH', 'SWANN', ':', 'On', 'what', 'charges', '?'],
 ['WILL', 'TURNER', ':', 'No', '!'],
 ['[',
 'Beckett',
 'takes',
 'another',
 'document',
 'from',
 'Mercer',
 ',',
 'who',
 'is',
 'standing',
 'with',
 'Beckett',
 ',',
 'craggy',
 'face',
 'and',
 'pony',
 'tail',
 ']',

'LORD',
 'CUTLER',
 'BECKETT',
 ':',
 'Ah',
 '-',
 'ha',
 '!',
 ['Here', '"', 's', 'the', 'one', 'for', 'William', 'Turner', '.'],
 ['And',
 'I',
 'have',
 'another',
 'one',
 'for',
 'a',
 'Mister',
 'James',
 'Norrington',
 '.'],
 ['Is', 'he', 'present', '?'],
 ['ELIZABETH', 'SWANN', ':', '*', 'What', 'are', 'the', 'charges', '?'],
 ['*',
 'LORD',
 'CUTLER',
 'BECKETT',
 ':',
 'I',
 'don',
 '"',
 't',
 'believe',
 'that',
 '"',
 's',
 'the',
 'answer',
 'to',
 'the',
 'question',
 'I',
 'asked',
 '.'],
 ['WILL', 'TURNER', ':', 'Lord', 'Beckett', '!'],
 ['In',
 'the',
 'category',

'of',
'questions',
'*',
'not',
'*',
'answered',
'...',
'ELIZABETH',
'SWANN',
':',
'We',
'are',
'under',
'the',
'jurisdiction',
'of',
'the',
'King',
" ",
's',
'governor',
'of',
'Port',
'Royal',
'',
'and',
'you',
'will',
'tell',
'us',
'what',
'we',
'are',
'charged',
'with',
'.'],
['LORD',
'CUTLER',
'BECKETT',
':',
'For',
'which',
'the',
'punishment',
'',
'regrettably',
'',

'is',
 '*',
 'also',
 '*',
 'death',
 '.'],
 ['Perhaps',
 'you',
 'remember',
 'a',
 'certain',
 'pirate',
 'named',
 'Jack',
 'Sparrow',
 '.'],
 ['ELIZABETH', 'SWANN', ':', '*', 'Captain', '*', 'Jack', 'Sparrow', '.'],
 ['LORD', 'CUTLER', 'BECKETT', ':', 'Captain', 'Jack', 'Sparrow', '.'],
 ['Yes', ',', 'I', 'thought', 'you', 'might', '.'],
 ['Scene',
 ':',
 'BLACK',
 'PEARL',
 '[',
 'views',
 'of',
 'rigging',
 'ropes',
 'and',
 'blocks',
 'aboard',
 'a',
 'ship',
 'at',
 'night',
 ']',
 '[',
 'Gibbs',
 'walks',
 'the',
 'deck',
 'alone',
 'at',
 'night',
 ',',
 'singing',
 'and',

'drinking',
'from',
'a',
'bottle',
']',
'GIBBS',
':',
'Fifteen',
'men',
'on',
'a',
'dead',
'man',
'''',
's',
'chest',
'.'],
['Yo', 'ho', 'ho', ',', 'and', 'a', '*', 'bottle', '*', 'of', 'rum', '.'],
['Drink', 'and', 'the', 'devil', 'had', 'done', 'for', 'the', 'rest', '.'],
['Yo', 'ho', 'ho', ',', 'and', 'a', '*', 'bottle', '*', 'of', 'rum', '.'],
['Ha', '-', 'ha', '-', 'ha', '-', 'ha', '-', 'ha', '!'],
['[',
'a',
'bell',
'tolls',
',',
'Gibbs',
'is',
'interrupted',
'from',
'his',
'drinking',
'by',
'the',
'sound',
']',
'[',
'Gibbs',
'looks',
'up',
'to',
'see',
'a',
'flock',
'of',
'crows',
'flying',

'in',
'the',
'foggy',
'night',
'sky',
']',
'Scene',
':',
'TURKISH',
'PRISON',
'[',
'tolling',
'of',
'bells',
'',
'camera',
'pans',
'to',
'show',
'the',
'crows',
'flying',
'toward',
'a',
'coastal',
'island',
']',
'[',
'pinnacle',
'-',
'like',
'island',
'off',
'the',
'coast',
'at',
'night',
'',
'connected',
'to',
'the',
'mainland',
'by',
'thin',
'',
'lighted',
'bridge',

']',
'[',
'cages',
'with',
'men',
'inside',
'line',
'the',
'bridge',
']',
'[',
'prisoners',
'are',
'being',
'marched',
'across',
'the',
'bridge',
']',
'[',
'a',
'male',
'prisoner',
'with',
'bloody',
'legs',
'is',
'muttering',
,',
'and',
'is',
'dragged',
'across',
'the',
'bridge',
'by',
'guards',
']',
'[',
'a',
'crow',
'is',
'seen',
'reflected',
'in',
'the',
'eye',

'of',
'one',
'male',
'prisoner',
'in',
'a',
'cage',
,',
'the',
'crow',
'attacks',
'and',
'pecks',
'his',
'eye',
'out',
,',
'he',
'screams',
']',
'[',
'the',
'prisoner',
'who',
'was',
'being',
'dragged',
'is',
'taken',
'down',
'a',
'hatch',
'-',
'like',
'door',
'in',
'the',
'ground',
,',
'he',
'also',
'screams',
']',
'[',
'rocky',
'shoreline',
,',

'lit',
'with',
'torches',
']',
'[',
'men',
'toss',
'caskets',
'into',
'the',
'sea',
']',
'[',
'numerous',
'caskets',
'are',
'floating',
'out',
'to',
'sea',
'',
'vertical',
'streaks',
'in',
'the',
'distant',
'clouds',
'are',
'seen',
']',
'[',
'a',
'crow',
'lands',
'on',
'one',
'casket',
'',
'begins',
'pecking',
'at',
'the',
'top',
']',
'[',
'sudden',
'gunblast',

'from',
'inside',
'the',
'casket',
'blows',
'the',
'crow',
'away',
,',
'along',
'with',
'much',
'wood',
'],'
'['
'hand',
'holding',
'flintlock',
'pistol',
'emerges',
,',
'pans',
'about',
,',
'points',
'it',
'at',
'audience',
'for',
'a',
'moment',
'],'
'['
'Jack',
'breaks',
'out',
'of',
'the',
'top',
'of',
'the',
'casket',
,',
'puts',
'his',
'captain',
""",

's',
'hat',
'on',
']',
'JACK',
'SPARROW',
':',
'Sorry',
',',
'mate',
'.'],
['[',
'Jack',
'breaks',
'the',
'skeleton',
'''',
's',
'leg',
'off',
']',
'JACK',
'SPARROW',
':',
'Mind',
'if',
'we',
'make',
'a',
'little',
'side',
'trip',
'?'],
['I', 'didn', ''', 't', 'think', 'so', '.'],
['[',
'Jack',
'uses',
'the',
'skeleton',
'leg',
'as',
'an',
'oar',
',',
'rows',
'toward',
'the',

'full',
'moon',
'and',
'a',
'waiting',
'ship',
'in',
'the',
'distance',
'to',
'the',
'right',
'of',
'the',
'moon',
']',
'Scene',
':',
'BLACK',
'PEARL',
'[',
'Jack',
'arrives',
'at',
'the',
'Black',
'Pearl',
'in',
'his',
'coffin',
'rowboat',
']',
'[',
'Gibbs',
'holds',
'out',
'his',
'hand',
'to',
'help',
'Jack',
'aboard',
'',
'Jack',
'puts',
'the',
'skeleton',

'leg',
'in',
'Gibbs',
''',
'hand',
']',
'[',
'Cotton',
'drapes',
'a',
'coat',
'onto',
'Jack',
''',
's',
'back',
']',
'GIBBS',
':',
'Not',
'*',
'quite',
'*',
'according',
'to',
'plan',
'.'],
['JACK',
'SPARROW',
':',
'Complications',
'arose',
'',
'ensued',
'',
'were',
'overcome',
'.'],
['GIBBS',
':',
'You',
'got',
'what',
'you',
'went',
'in',
'for',

',',
 'then',
 '?'],
 ['JACK', 'SPARROW', ':', 'Mm', '-', 'hmm', '!'],
 ['GIBBS',
 ':',
 'Captain',
 ',',
 'I',
 'think',
 'the',
 'crew',
 ',',
 'meaning',
 'me',
 'as',
 'well',
 ',',
 'were',
 'expecting',
 'something',
 'a',
 'bit',
 'more',
 '...',
 '*',
 'shiny',
 '*.'],
 ['What',
 'with',
 'the',
 'Isla',
 'de',
 'Muerta',
 'going',
 'all',
 'pear',
 'shaped',
 ',',
 'reclaimed',
 'by',
 'the',
 'sea',
 ',',
 'and',
 'the',
 'treasure',

'with',
 'it',
 '.',
 ['LEECH',
 ':',
 'And',
 'the',
 'Royal',
 'Navy',
 'chasing',
 'us',
 'all',
 'around',
 'the',
 'Atlantic',
 '.'],
 ['MARTY', ':', 'And', 'the', 'hurricane', '!'],
 ['Aye', '.'],
 ['CREW', ':', 'Aye', '.'],
 ['Aye', '.'],
 ['GIBBS',
 ':',
 'All',
 'in',
 'all',
 ',',
 'it',
 '"',
 's',
 'seems',
 'some',
 'time',
 'since',
 'we',
 'did',
 'a',
 'speck',
 'of',
 'honest',
 'pirating',
 '.'],
 ['JACK', 'SPARROW', ':', 'Shiny', '?'],
 ['GIBBS', ':', 'Aye', ', ', 'shiny', '.'],
 ['JACK',
 'SPARROW',
 ':',
 'Is',

'that',
 'how',
 'you',
 '"',
 're',
 'all',
 'feeling',
 ',',
 'then',
 '?'],
 ['Perhaps',
 'dear',
 'old',
 'Jack',
 'is',
 'not',
 'serving',
 'your',
 'best',
 'interests',
 'as',
 'captain',
 '?'],
 ['COTTON', '"', 'S', 'PARROT', ':', 'Awk', '!'],
 ['Walk', 'the', 'plank', '!'],
 ['[',
 'Jack',
 'grabs',
 'his',
 'pistol',
 'and',
 'points',
 'it',
 'at',
 'the',
 'bird',
 ']',
 'JACK',
 'SPARROW',
 ':',
 'What',
 'did',
 'the',
 'bird',
 'say',
 '?'],
 ['LEECH', ':', 'Do', 'not', 'blame', 'the', 'bird', '.'],

['Show',
 'us',
 'what',
 'is',
 'on',
 'that',
 'piece',
 'of',
 'cloth',
 'there',
 '.'],
['[',
 'the',
 'monkey',
 'snarls',
 ',',
 'seizes',
 'the',
 'cloth',
 ',',
 'scampers',
 'off',
 'with',
 'it',
 ']',
 'JACK',
 'SPARROW',
 ':',
 'Ohhh',
 '!'],
['[',
 'Jack',
 'attempts',
 'to',
 'shoot',
 'the',
 'monkey',
 'with',
 'his',
 'pistol',
 ',',
 'his',
 'pistol',
 'misfires',
 ']',
 '[',
 'Jack',

'grabs',
'a',
'pistol',
'from',
'another',
'crewman',
'and',
'blasts',
'the',
'monkey',
'while',
'it',
''',
's',
'on',
'the',
'deck',
']',
'[',
'the',
'monkey',
'drops',
'the',
'cloth',
,',
'but',
'is',
'unharmed',
,',
'and',
'scurries',
'into',
'the',
'rigging',
']',
'GIBBS',
':',
'Know',
'that',
'don',
''',
't',
'do',
'no',
'good',
'.''],
['JACK', 'SPARROW', ':', 'It', 'does', 'me', '.'],

```
[['',
  'Marty',
  'picks',
  'up',
  'the',
  'cloth',
  'that',
  'the',
  'monkey',
  'dropped',
  ''],
  'MARTY',
  ':',
  'It',
  '"',
  's',
  'a',
  'key',
  '!'],
  ['JACK', 'SPARROW', ':', 'No', '!'],
  ['Much', 'more', 'better', '.'],
  ['It', 'is', 'a', '*', 'drawing', '*', 'of', 'a', 'key', '.']]
```

14 Brown Corpus

The Brown Corpus was the first million-word electronic corpus of English, created in 1961 at Brown University. This corpus contains text from 500 sources, and the sources have been categorized by genre, such as news, editorial, and so on. for a complete list, see <http://icame.uib.no/brown/bcm-los.html>).

Import brown corpus and try to explore the categories into it

```
[16]: from nltk.corpus import brown
      nltk.download('brown')
      brown.categories()
```

```
[nltk_data] Downloading package brown to
[nltk_data] /Users/pushpakulkarni/nltk_data...
[nltk_data] Package brown is already up-to-date!
```

```
[16]: ['adventure',
      'belles_lettres',
      'editorial',
      'fiction',
      'government',
      'hobbies',
      'humor',
```

```
'learned',
'lore',
'mystery',
'news',
'religion',
'reviews',
'romance',
'science_fiction']
```

```
[17]: brown.words(categories='news')
```

```
[17]: ['The', 'Fulton', 'County', 'Grand', 'Jury', 'said', ...]
```

```
[18]: brown.words(fileids=['cg22'])
```

```
[18]: ['Does', 'our', 'society', 'have', 'a', 'runaway', ',', ...]
```

```
[19]: brown.sents(categories=['news', 'editorial', 'reviews'])
```

```
[19]: [['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', 'over-
all', 'charge', 'of', 'the', 'election', ',', 'deserves', 'the', 'praise',
'and', 'thanks', 'of', 'the', 'City', 'of', 'Atlanta', '', 'for', 'the',
'manner', 'in', 'which', 'the', 'election', 'was', 'conducted', '.'], ...]
```

compare genres in their usage of modal verbs

```
[20]: news_text=brown.words(categories='news')
fdist = nltk.FreqDist([w.lower() for w in news_text])
print(fdist)
modals = ['can', 'could', 'may', 'might', 'must', 'will']
for m in modals:
    print(m + ': ', fdist[m])
```

```
<FreqDist with 13112 samples and 100554 outcomes>
```

```
can: 94
could: 87
may: 93
might: 38
must: 53
will: 389
```

15 To Do:

obtain counts for each genre of interest. use NLTK's support for conditional frequency distributions.

```
[21]: cfd = nltk.ConditionalFreqDist(
        (genre, word)
        for genre in brown.categories()
        for word in brown.words(categories=genre))
genres = ['news', 'religion', 'hobbies', 'science_fiction', 'romance', 'humor']
modals = ['can', 'could', 'may', 'might', 'must', 'will']
cfd.tabulate(conditions=genres, samples=modals)
```

	can	could	may	might	must	will
news	93	86	66	38	50	389
religion	82	59	78	12	54	71
hobbies	268	58	131	22	83	264
science_fiction	16	49	4	12	8	16
romance	74	193	11	51	45	43
humor	16	30	8	8	9	13

Ans: The most frequent modal in the news genre is 'will', while the most frequent modal in the romance genre is 'could'. Using this method we can make models to analyse word counts and hence distinguish genres

other corpus in NLTK are Reuter's corpus(news headline),genesis,udhr,inaugral etc. Unfortunately, for many languages, substantial corpora are not yet available. Often there is insufficient government or industrial support for developing language resources, and individual efforts are piecemeal and hard to discover or reuse. Some languages have no established writing system, or are endangered.

#Common structures for text corpora: The simplest kind of corpus is a collection of isolated texts with no particular organization; some corpora are structured into categories, such as genre(Brown Corpus); some categorizations overlap, such as topic categories (Reuters Corpus); other corpora represent language use over time (Inaugural Address Corpus).

16 Generating random text:

obtain all bigrams from the text of the book of Genesis, then constructs a conditional frequency distribution to record which words are most likely to follow a given word; e.g., after the word living, the most likely word is creature

```
[22]: nltk.download('genesis')
```

```
[nltk_data] Downloading package genesis to
[nltk_data] /Users/pushpakulkarni/nltk_data...
[nltk_data] Package genesis is already up-to-date!
```

```
[22]: True
```

```
[23]: def generate_text(cfdist,word,num=15):
        for i in range(num):
            print(word),
            word=cfdist[word].max()
        text = nltk.corpus.genesis.words('english-kjv.txt')
```

```
bigrams = nltk.bigrams(text)
cfd = nltk.ConditionalFreqDist(bigrams)
```

```
[24]: cfd['living']
```

```
[24]: FreqDist({'creature': 7, 'thing': 4, 'substance': 2, 'soul': 1, '.': 1, ',': 1})
```

```
[25]: generate_text(cfd, 'living')
```

```
living
creature
that
he
said
,
and
the
land
of
the
land
of
the
land
```

17 write a python program to work out the plural form of any English noun e.g., if the word ends with y, delete the y and add ies.

```
[26]: import inflect

p = inflect.engine()
plural_word = p.plural("revolutionary")
print(plural_word)
```

```
revolutionaries
```

18 Wordlist Corpora

NLTK includes some corpora that are nothing more than wordlists. We can use it to find unusual or misspelled words in a text corpus

Filtering a text: This program computes the vocabulary of a text, then removes all items that occur in an existing wordlist, leaving just the uncommon or misspelled words.


```
[27]: def unusual_words(text):
      text_vocab=set(w.lower() for w in text if w.isalpha())
      english_vocab=set(w.lower() for w in nltk.corpus.words.words())
      unusual=text_vocab.difference(english_vocab)
      return sorted(unusual)
```

```
[28]: print(nltk.corpus.words.words()[:25])
```

```
['A', 'a', 'aa', 'aal', 'aalii', 'aam', 'Aani', 'aardvark', 'aardwolf', 'Aaron',
'Aaronic', 'Aaronical', 'Aaronite', 'Aaronitic', 'Aaru', 'Ab', 'aba', 'Ababdeh',
'Ababua', 'abac', 'abaca', 'abacate', 'abacay', 'abacinate', 'abacination']
```

```
[29]: unusual_words(nltk.corpus.gutenberg.words('austen-sense.txt'))
```

```
[29]: ['abbeyland',
      'abhorred',
      'abilities',
      'abounded',
      'abridgement',
      'abused',
      'abuses',
      'accents',
      'accepting',
      'accommodations',
      'accompanied',
      'accounted',
      'accounts',
      'accustomary',
      'aches',
      'acknowledging',
      'acknowledgment',
      'acknowledgments',
      'acquaintances',
      'acquiesced',
      'acquitted',
      'acquitting',
      'acted',
      'actions',
      'adapted',
      'adding',
      'additions',
      'addressed',
      'addresses',
      'addressing',
      'adhering',
      'adieu',
      'adjusting',
```

'administering',
'admirers',
'admires',
'admitting',
'adorned',
'advances',
'advantages',
'affairs',
'affections',
'affects',
'affixed',
'afflictions',
'afforded',
'affording',
'ages',
'agitated',
'agonies',
'ailments',
'aimed',
'alarms',
'alienated',
'alighted',
'alleged',
'allenham',
'allowances',
'allowed',
'allowing',
'alluded',
'alterations',
'altered',
'altering',
'amended',
'amounted',
'amusements',
'ankles',
'annamaria',
'annexed',
'announced',
'announcing',
'annuities',
'annum',
'answered',
'answering',
'answers',
'anticipated',
'anticipating',
'anticipations',

'anymore',
'apartments',
'apologies',
'apologising',
'apologized',
'appearances',
'appeared',
'appearing',
'appeased',
'appetites',
'applauded',
'applying',
'appointed',
'apprehended',
'apprehensions',
'approached',
'approved',
'arbour',
'ardour',
'arguments',
'arranged',
'arrangements',
'arranging',
'arrived',
'arrives',
'arriving',
'ascended',
'ascertained',
'asked',
'asking',
'assembled',
'assemblies',
'asserted',
'assertions',
'assiduities',
'assisted',
'assisting',
'associating',
'assurances',
'astonished',
'atoned',
'atoning',
'attaching',
'attachments',
'attacked',
'attacks',
'attained',

'attempted',
'attempting',
'attempts',
'attendants',
'attended',
'attending',
'attentions',
'attracted',
'attractions',
'attributed',
'attributing',
'auditors',
'augmenting',
'austen',
'authorised',
'authors',
'availed',
'avignon',
'avoided',
'avoiding',
'awaited',
'awakened',
'awaking',
'bags',
'balls',
'banished',
'barouches',
'bathed',
'bears',
'beasts',
'beauties',
'became',
'bedrooms',
'beds',
'befallen',
'befalls',
'befell',
'began',
'begged',
'begins',
'behaved',
'beings',
'believed',
'believes',
'belonged',
'belongs',
'benefited',

'bequeathed',
'berkeley',
'bestowed',
'betrayed',
'betraying',
'biased',
'blackest',
'blameable',
'blessings',
'blights',
'blossoms',
'blundered',
'blushed',
'blushes',
'bolder',
'bones',
'bonomi',
'books',
'booksellers',
'borrowed',
'bottoms',
'boys',
'breakfasting',
'bribing',
'brightened',
'brighter',
'bringing',
'brings',
'broader',
'brothers',
'bruised',
'buildings',
'bursts',
'buying',
'called',
'calls',
'calming',
'candles',
'candour',
'canvassing',
'cards',
'cares',
'caresses',
'careys',
'carriages',
'carries',
'cases',

'casts',
'cats',
'caused',
'ceased',
'ceasing',
'censured',
'centre',
'certainties',
'chagrined',
'chairs',
'chambers',
'chanced',
'changed',
'changes',
'changing',
'characters',
'charged',
'charmed',
'charms',
'cheated',
'checking',
'cheeks',
'cheerfuller',
'cherished',
'cherries',
'children',
'choked',
'chuse',
'chusing',
'circles',
'circumstances',
'civilities',
'claimed',
'claiming',
'claims',
'clarke',
'cleared',
'cleveland',
'clogged',
'closing',
'clouds',
'coats',
'collecting',
'coloured',
'colouring',
'combe',
'comforted',

'comforts',
'comings',
'commanded',
'commands',
'commended',
'comments',
'commissioned',
'commonest',
'communicated',
'companions',
'compared',
'compares',
'comparisons',
'complained',
'complaining',
'complaints',
'completed',
'compliments',
'comprehended',
'concealing',
'concerns',
'concessions',
'concluded',
'conclusions',
'conditions',
'conducted',
'confessed',
'confidante',
'conforming',
'congratulated',
'congratulating',
'congratulations',
'conjectured',
'conjectures',
'conjecturing',
'connections',
'conquests',
'consented',
'consequences',
'considerations',
'considers',
'consisted',
'consists',
'consoled',
'conspired',
'constantia',
'consulted',

'contained',
'containing',
'contend',
'contenting',
'continuing',
'contradicted',
'contrasted',
'contributed',
'contributing',
'contrived',
'contrives',
'contriving',
'controlled',
'conveniences',
'conversations',
'conversed',
'conversing',
'conveyed',
'conveying',
'copying',
'cordials',
'cottages',
'counsellor',
'counteracted',
'couples',
'courted',
'courting',
'courtland',
'cousins',
'cowper',
'cows',
'coxcombs',
'cramps',
'created',
'creating',
'creatures',
'cries',
'crimsoned',
'curtsying',
'cutlets',
'danced',
'dances',
'dared',
'darker',
'dartford',
'dashwood',
'dashwoods',

'daughters',
'davies',
'dawdled',
'dawlish',
'dawned',
'dearer',
'dearest',
'debated',
'debts',
'deceived',
'deciding',
'decisions',
'declares',
'declaring',
'declining',
'deemed',
'deeper',
'deepest',
'defects',
'defended',
'deficiencies',
'degrees',
'delaford',
'delayed',
'delays',
'deliberating',
'delicacies',
'delighful',
'delineated',
'delivered',
'demanded',
'demands',
'demonstrations',
'demur',
'denied',
'dennison',
'denoted',
'denoting',
'departing',
'depended',
'depends',
'deprived',
'described',
'describing',
'deserts',
'deserves',
'designs',

'desiring',
'despatch',
'despatching',
'despised',
'despising',
'destroyed',
'destroys',
'detaining',
'detected',
'detecting',
'determining',
'deterred',
'detested',
'devolved',
'died',
'dies',
'differed',
'differing',
'difficulties',
'dimensions',
'diminished',
'dined',
'dinners',
'directing',
'directions',
'disagreements',
'disappeared',
'disappointments',
'disapproved',
'disapproves',
'disapproving',
'discarded',
'discharged',
'disclaiming',
'disclosing',
'discontents',
'discovering',
'discussions',
'disgraced',
'disinherited',
'disliked',
'dismissed',
'dismounted',
'dispatched',
'dispatches',
'dispersing',
'disposing',

'disputes',
'disqualifications',
'disregarded',
'dissembling',
'dissented',
'distresses',
'distrusts',
'diverted',
'doatingly',
'donavan',
'doomed',
'dooming',
'doors',
'dorsetshire',
'doubted',
'doubts',
'douceur',
'downs',
'dr',
'drains',
'drawings',
'draws',
'dreaded',
'dreading',
'dreaming',
'dresses',
'drives',
'dropped',
'drops',
'drury',
'duets',
'duties',
'earlier',
'earliest',
'earned',
'ears',
'echoed',
'editions',
'editions',
'effected',
'effecting',
'effusions',
'ellison',
'ellisons',
'eloping',
'eluded',
'embellishments',

'embraced',
'embraces',
'employments',
'enabled',
'enamoured',
'encouraged',
'encouragements',
'encroachments',
'encumbered',
'endeavoring',
'endeavors',
'endeavour',
'endeavoured',
'endeavouring',
'endeavours',
'endowed',
'ends',
'endured',
'enfeebled',
'enforcing',
'engagements',
'england',
'enjoyed',
'enjoyments',
'enquired',
'enquiries',
'enquiring',
'ensued',
'ensured',
'entered',
'entertained',
'entitled',
'entreated',
'entreaties',
'entrusted',
'equalled',
'equals',
'erred',
'errors',
'escaped',
'esq',
'establishing',
'esteemed',
'esteeming',
'esteems',
'estimating',
'estranged',

'evenings',
'events',
'evils',
'examined',
'exceeded',
'excellencies',
'exchanged',
'exclaimed',
'exclamations',
'excused',
'excuses',
'exercised',
'exercising',
'exerted',
'exertions',
'exeter',
'exhilarated',
'existed',
'expectations',
'expected',
'expecting',
'expects',
'expenses',
'experiencing',
'explained',
'explanations',
'expressing',
'expressions',
'extolling',
'extorted',
'extorting',
'extremest',
'eyeing',
'eyes',
'faces',
'facts',
'failed',
'falls',
'familiarized',
'families',
'fancying',
'fates',
'fatigued',
'fatigues',
'faults',
'favour',
'favourable',

'favourite',
'favourites',
'fearing',
'fears',
'features',
'feelings',
'feels',
'feet',
'felicitations',
'females',
'ferrars',
'fetches',
'fettered',
'finds',
'finest',
'fingers',
'flattered',
'flatteries',
'flowed',
'fluctuating',
'flushed',
'foibles',
'followed',
'follows',
'fond',
'footsteps',
'forebodings',
'foreplanned',
'foresaw',
'foreseeing',
'foreseen',
'forfeited',
'forfeiting',
'forgave',
'forgiven',
'forms',
'forsaking',
'fortunes',
'forwarded',
'foundations',
'founded',
'fowls',
'friendliest',
'friends',
'frightens',
'froid',
'frosts',

'fulfil',
'fulfilled',
'fullest',
'gained',
'gales',
'gardens',
'garrets',
'gates',
'gathered',
'generations',
'gentlemen',
'gigs',
'gilberts',
'girls',
'gives',
'glances',
'gloried',
'gloves',
'godby',
'goings',
'goodby',
'governed',
'gowns',
'graces',
'grandmothers',
'granted',
'greatest',
'grieves',
'grows',
'guardians',
'guessed',
'guests',
'guided',
'guineas',
'habits',
'hallooing',
'hands',
'handsomer',
'handsomest',
'hang',
'hanover',
'happened',
'happens',
'hardened',
'hardships',
'harley',
'has',

'hastened',
'hastening',
'hated',
'hates',
'hating',
'having',
'hazarded',
'hazarding',
'heads',
'heard',
'hears',
'heightened',
'heightening',
'heights',
'heirs',
'held',
'hens',
'henshawe',
'hesitated',
'hiding',
'hills',
'hinted',
'hints',
'hoarded',
'holborn',
'holburn',
'holds',
'holidays',
'homes',
'honeysuckles',
'honiton',
'honour',
'honourable',
'honourably',
'honoured',
'honours',
'hopes',
'hoping',
'horrors',
'horses',
'hours',
'houses',
'howsoever',
'humbled',
'humiliations',
'humored',
'humoured',

'humouring',
'hunted',
'hunters',
'hunts',
'hurrying',
'husbands',
'huswives',
'ideas',
'idled',
'idolized',
'ii',
'imaginations',
'imagined',
'imagining',
'imbibed',
'immoveable',
'imparted',
'imperfections',
'implied',
'implies',
'impoverished',
'impoverishing',
'improved',
'improvements',
'imputed',
'inclinations',
'inclined',
'inclosing',
'including',
'incommoded',
'inconveniences',
'increased',
'incurred',
'incurring',
'indulged',
'infants',
'inflicted',
'inflicting',
'influenced',
'inforce',
'inforced',
'informing',
'inhabitants',
'inhabiting',
'inheriting',
'injuries',
'inquired',

'inquiries',
'insinuations',
'insisted',
'installed',
'instigated',
'instructions',
'insulted',
'intends',
'intentions',
'intents',
'interests',
'interposed',
'interspersed',
'intervals',
'interviews',
'intimated',
'introduced',
'introducing',
'intruded',
'invented',
'inventing',
'invitations',
'invited',
'irritated',
'irritates',
'issued',
'jealousies',
'jenning',
'jennings',
'jewels',
'jilting',
'joined',
'joked',
'jokes',
'joking',
'joys',
'judged',
'judging',
'judgments',
'jumbled',
'justified',
'keeps',
'keys',
'kicked',
'kinder',
'kindest',
'kingham',

'kissed',
'kisses',
'knees',
'knives',
'knows',
'laboured',
'lamentations',
'lamps',
'lanes',
'languages',
'larger',
'largest',
'lasted',
'laughed',
'laughs',
'leagued',
'legacies',
'lengthened',
'lengths',
'lessened',
'lessening',
'letters',
'letting',
'lies',
'lifted',
'lightened',
'liked',
'likes',
'limbs',
'limits',
'lines',
'lingered',
'lingering',
'lips',
'listened',
'lives',
'livings',
'll',
'lodges',
'loitered',
'lombardy',
'london',
'longed',
'longest',
'longstaple',
'looked',
'looks',

'loved',
'lovers',
'loves',
'lowered',
'lurking',
'magna',
'maids',
'maintained',
'makes',
'mama',
'managed',
'marlborough',
'marriages',
'marries',
'matters',
'maxims',
'meadows',
'meals',
'means',
'meantime',
'measures',
'medicines',
'meditated',
'meditations',
'meetings',
'mentioned',
'mentioning',
'merest',
'merits',
'merrier',
'messages',
'middleton',
'middletons',
'militated',
'minds',
'minutes',
'misapplied',
'misinformed',
'missed',
'misses',
'mistakes',
'mixing',
'modestest',
'mohrs',
'moments',
'months',
'mosquitoes',

'mothers',
'motives',
'moved',
'murmurings',
'muttered',
'nabobs',
'named',
'names',
'natured',
'nearer',
'needed',
'neglected',
'neighbour',
'neighbourhood',
'neighbouring',
'neighbourly',
'neighbours',
'nerves',
'nests',
'nettles',
'newer',
'newspapers',
'nicest',
'nieces',
'nipped',
'nodded',
'nods',
'noisier',
'notes',
'noticed',
'noticing',
'notions',
'nt',
'nurses',
'obeyed',
'objected',
'objections',
'objects',
'obligations',
'observations',
'observed',
'obstacles',
'obstructed',
'obtained',
'obtaining',
'obviated',
'obviating',

```
'occasioned',  
'occasions',  
'occupations',  
'occupied',  
'occurred',  
'oddest',  
'offence',  
'offences',  
'offending',  
'offered',  
'offices',  
'oftener',  
'oftenest',  
'oldest',  
'olives',  
'omitted',  
'ones',  
'opened',  
'opinions',  
'opportunities',  
'ordained',  
'orders',  
'originated',  
'ornamented',  
'ornaments',  
'others',  
'outdone',  
...]
```

19 stopwords corpus

```
[32]: import nltk  
      from nltk.corpus import stopwords  
      stop_words_nltk=stopwords.words('english')
```

20 Write a function in python to compute what fraction of words in a text are not in the stopwords list and check the fraction in gutenbergr, reuters, brown and genesis corpus.

```
[33]: import string  
      from nltk.tokenize import word_tokenize  
      import re  
      from nltk.corpus import stopwords
```

```

def check(raw):
    stop_words_nltk=stopwords.words('english')
    print('Number of stopwords:',len(stop_words_nltk))
    print()
    corpus_lower_3 = raw
    corpus_new_3 = re.sub(r'\d+', ' ', corpus_lower_3)
    corpus_new_3 = corpus_new_3.translate(str.maketrans("", "", string.
→punctuation))
    corpus_new_3 = ' '.join([token for token in corpus_new_3.split()])
    tokenized_corpus_nltk_3 = word_tokenize(corpus_new_3)
    print("Total Number of words in given Corpus: ",len(tokenized_corpus_nltk_3))
    print()
    tokenized_corpus_without_stopwords_3 = [i for i in tokenized_corpus_nltk_3 if
→not i in stop_words_nltk]
    print("Number of words in corpus which are not in stopword list:
→",len(tokenized_corpus_without_stopwords_3))
    print("Content fraction is:",(len(tokenized_corpus_without_stopwords_3)/
→len(raw)))

```

```

[34]: raw1=nltk.corpus.gutenberg.raw()
      check(raw1)

```

Number of stopwords: 179

Total Number of words in given Corpus: 2103363

Number of words in corpus which are not in stopword list: 1133359

Content fraction is: 0.09610179255744651

```

[35]: raw2=nltk.corpus.reuters.raw()
      check(raw2)

```

Number of stopwords: 179

Total Number of words in given Corpus: 1291328

Number of words in corpus which are not in stopword list: 890757

Content fraction is: 0.1006863118444491

```

[36]: raw3=nltk.corpus.brown.raw()
      check(raw3)

```

Number of stopwords: 179

Total Number of words in given Corpus: 1015080

Number of words in corpus which are not in stopword list: 1010546

Content fraction is: 0.10141682031543862

```
[37]: raw4=nltk.corpus.genesis.raw()  
      check(raw4)
```

Number of stopwords: 179

Total Number of words in given Corpus: 255324

Number of words in corpus which are not in stopword list: 210582

Content fraction is: 0.1482529160307063

A word puzzle: A grid of randomly chosen letters with rules for creating words out of the letters; this puzzle is known as “Target.” Hint: Use Freqdist to check frequency of words

21 Ans: We got 19 words.

```
[38]: puzzle = nltk.FreqDist('egivrvonl')  
      central = 'r'  
      wordlist = nltk.corpus.words.words()  
      [w for w in wordlist if len(w) >= 6 and central in w and nltk.FreqDist(w) <= 10  
      → puzzle]
```

```
[38]: ['glover',  
      'gorlin',  
      'govern',  
      'grovel',  
      'ignore',  
      'involver',  
      'lienor',  
      'linger',  
      'longer',  
      'loving',  
      'noiler',  
      'overling',  
      'region',  
      'renvoi',  
      'revolving',  
      'ringle',  
      'roving',  
      'violer',  
      'virole']
```

22 wordnet

WordNet is a semantically oriented dictionary of English, similar to a traditional thesaurus but with a richer structure. NLTK includes the English WordNet, with 155,287 words and 117,659 synonym

sets. We'll begin by looking at synonyms and how they are accessed in WordNet.

#Senses and Synonyms Consider the sentence in (1a). If we replace the word motorcar in (1a) with automobile, to get (1b), the meaning of the sentence stays pretty much the same: (1) a. Benz is credited with the invention of the motorcar. b. Benz is credited with the invention of the automobile. Since everything else in the sentence has remained unchanged, we can conclude that the words motorcar and automobile have the same meaning, i.e., they are synonyms.

```
[39]: from nltk.corpus import wordnet as wn
      wn.synsets('motorcar')
```

```
[39]: [Synset('car.n.01')]
```

```
[40]: wn.synsets('hi')
```

```
[40]: [Synset('hello.n.01'), Synset('hawaii.n.01')]
```

23 Thus, motorcar has just one possible meaning and it is identified as car.n.01, the first

noun sense of car. The entity car.n.01 is called a synset, or “synonym set,” a collection of synonymous words (or “lemmas”).

```
[41]: wn.synset('car.n.01').lemma_names()
```

```
[41]: ['car', 'auto', 'automobile', 'machine', 'motorcar']
```

```
[42]: wn.synset('hello.n.01').lemma_names()
```

```
[42]: ['hello', 'hullo', 'hi', 'howdy', 'how-do-you-do']
```

```
[43]: wn.synset('car.n.01').definition()
```

```
[43]: 'a motor vehicle with four wheels; usually propelled by an internal combustion
      engine'
```

```
[44]: wn.synset('hello.n.01').definition()
```

```
[44]: 'an expression of greeting'
```

```
[45]: wn.synset('car.n.01').examples()
```

```
[45]: ['he needs a car to get to work']
```

```
[46]: wn.synset('hello.n.01').examples()
```

```
[46]: ['every morning they exchanged polite hellos']
```

```
[47]: wn.synset('car.n.01').lemmas()
```

```
[47]: [Lemma('car.n.01.car'),  
      Lemma('car.n.01.auto'),  
      Lemma('car.n.01.automobile'),  
      Lemma('car.n.01.machine'),  
      Lemma('car.n.01.motorcar')]
```

```
[48]: wn.synset('hello.n.01').lemmas()
```

```
[48]: [Lemma('hello.n.01.hello'),  
      Lemma('hello.n.01.hullo'),  
      Lemma('hello.n.01.hi'),  
      Lemma('hello.n.01.howdy'),  
      Lemma('hello.n.01.how-do-you-do')]
```

```
[49]: wn.synsets('car')  
for synset in wn.synsets('car'):  
    print(synset.lemma_names())
```

```
['car', 'auto', 'automobile', 'machine', 'motorcar']  
['car', 'railcar', 'railway_car', 'railroad_car']  
['car', 'gondola']  
['car', 'elevator_car']  
['cable_car', 'car']
```

```
[50]: wn.synsets('hello')  
for synset in wn.synsets('hello'):  
    print(synset.lemma_names())
```

```
['hello', 'hullo', 'hi', 'howdy', 'how-do-you-do']
```

24 Write down all the senses of the word dish that you can think of. Now, explore this word with the help of WordNet

```
[51]: wn.synsets('dish')
```

```
[51]: [Synset('dish.n.01'),  
      Synset('dish.n.02'),  
      Synset('dish.n.03'),  
      Synset('smasher.n.02'),  
      Synset('dish.n.05'),  
      Synset('cup_of_tea.n.01'),  
      Synset('serve.v.06'),  
      Synset('dish.v.02')]
```

```
[52]: wn.synset('dish.n.03').lemma_names()
```

```
[52]: ['dish', 'dishful']
```

```
[53]: wn.synset('dish.n.03').definition()
```

```
[53]: 'the quantity that a dish will hold'
```

```
[54]: wn.synset('dish.n.03').examples()
```

```
[54]: ['they served me a dish of rice']
```

```
[55]: wn.synsets('dish')
for synset in wn.synsets('dish'):
    print(synset.lemma_names())
```

```
['dish']
['dish']
['dish', 'dishful']
['smasher', 'stunner', 'knockout', 'beauty', 'ravisher', 'sweetheart', 'peach',
'lulu', 'looker', 'mantrap', 'dish']
['dish', 'dish_aerial', 'dish_antenna', 'saucer']
['cup_of_tea', 'bag', 'dish']
['serve', 'serve_up', 'dish_out', 'dish_up', 'dish']
['dish']
```

```
[56]: eg = wn.synset('dish.n.03')
a = eg.hyponyms()
a
```

```
[56]: []
```

```
[57]: for synset in wn.synsets('dish', wn.NOUN):
    print(synset.name() + ': ', synset.definition())
```

```
dish.n.01: a piece of dishware normally used as a container for holding or
serving food
dish.n.02: a particular item of prepared food
dish.n.03: the quantity that a dish will hold
smasher.n.02: a very attractive or seductive looking woman
dish.n.05: directional antenna consisting of a parabolic reflector for microwave
or radio frequency radiation
cup_of_tea.n.01: an activity that you like or at which you are superior
```

25 Solved Examples:

```
[58]: motorcar = wn.synset('car.n.01')
      types_of_motorcar = motorcar.hyponyms()
      types_of_motorcar
```

```
[58]: [Synset('ambulance.n.01'),
      Synset('beach_wagon.n.01'),
      Synset('bus.n.04'),
      Synset('cab.n.03'),
      Synset('compact.n.03'),
      Synset('convertible.n.01'),
      Synset('coupe.n.01'),
      Synset('cruiser.n.01'),
      Synset('electric.n.01'),
      Synset('gas_guzzler.n.01'),
      Synset('hardtop.n.01'),
      Synset('hatchback.n.01'),
      Synset('horseless_carriage.n.01'),
      Synset('hot_rod.n.01'),
      Synset('jeep.n.01'),
      Synset('limousine.n.01'),
      Synset('loaner.n.02'),
      Synset('minicar.n.01'),
      Synset('minivan.n.01'),
      Synset('model_t.n.01'),
      Synset('pace_car.n.01'),
      Synset('racer.n.02'),
      Synset('roadster.n.01'),
      Synset('sedan.n.01'),
      Synset('sport_utility.n.01'),
      Synset('sports_car.n.01'),
      Synset('stanley_steamer.n.01'),
      Synset('stock_car.n.01'),
      Synset('subcompact.n.01'),
      Synset('touring_car.n.01'),
      Synset('used-car.n.01')]
```

```
[59]: motorcar = wn.synset('computer.n.01')
      types_of_motorcar = motorcar.hyponyms()
      types_of_motorcar
```

```
[59]: [Synset('analog_computer.n.01'),
      Synset('digital_computer.n.01'),
      Synset('home_computer.n.01'),
      Synset('node.n.08'),
      Synset('number_cruncher.n.02'),
```

```
Synset('pari-mutuel_machine.n.01'),
Synset('predictor.n.03'),
Synset('server.n.03'),
Synset('turing_machine.n.01'),
Synset('web_site.n.01')]
```

```
[60]: motorcar = wn.synset('car.n.01')
      types_of_motorcar = motorcar.hypernyms()
      types_of_motorcar
```

```
[60]: [Synset('motor_vehicle.n.01')]
```

```
[61]: for synset in wn.synsets('mint', wn.NOUN):
      print(synset.name() + ': ', synset.definition())
```

```
batch.n.02: (often followed by `of') a large number or amount or extent
mint.n.02: any north temperate plant of the genus Mentha with aromatic leaves
and small mauve flowers
mint.n.03: any member of the mint family of plants
mint.n.04: the leaves of a mint plant used fresh or candied
mint.n.05: a candy that is flavored with a mint oil
mint.n.06: a plant where money is coined by authority of the government
```

26 Conclusion:

- 1) Brown and Genesis corpus have least stop words where as Gutenberg has most number of stop words.
- 2) Difference() gives the difference between 2 strings or lists.
- 3) A Conditional Frequency is a collection of frequency distributions, computed based on a condition. For computing a conditional frequency, you have to attach a condition to every occurrence of an event.