

Assignment-3

February 17, 2023

0.0.1 WordNet

Wordnet is a database of nouns, verbs, adjectives, and adverbs. These words contain short definitions called glosses. Words are partitioned into synonym sets called synsets.

```
[ ]: import math
import nltk
from nltk.corpus import wordnet as wn
from nltk.wsd import lesk
from nltk.corpus import sentiwordnet as swn
from nltk.book import text2, text4
nltk.download('all')
```

```
[nltk_data] Downloading collection 'all'
[nltk_data] |
[nltk_data] | Downloading package abc to /root/nltk_data...
[nltk_data] |   Package abc is already up-to-date!
[nltk_data] | Downloading package alpino to /root/nltk_data...
[nltk_data] |   Package alpino is already up-to-date!
[nltk_data] | Downloading package averaged_perceptron_tagger to
[nltk_data] |   /root/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] |   /root/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] |   /root/nltk_data...
[nltk_data] |   Package basque_grammars is already up-to-date!
[nltk_data] | Downloading package bcp47 to /root/nltk_data...
[nltk_data] |   Package bcp47 is already up-to-date!
[nltk_data] | Downloading package biocreative_ppi to
[nltk_data] |   /root/nltk_data...
[nltk_data] |   Package biocreative_ppi is already up-to-date!
[nltk_data] | Downloading package bllip_wsj_no_aux to
[nltk_data] |   /root/nltk_data...
[nltk_data] |   Package bllip_wsj_no_aux is already up-to-date!
[nltk_data] | Downloading package book_grammars to
```

```

[nltk_data] | /root/nltk_data...
[nltk_data] | Package book_grammars is already up-to-date!
[nltk_data] | Downloading package brown to /root/nltk_data...
[nltk_data] | Package brown is already up-to-date!
[nltk_data] | Downloading package brown_tei to /root/nltk_data...
[nltk_data] | Package brown_tei is already up-to-date!
[nltk_data] | Downloading package cess_cat to /root/nltk_data...
[nltk_data] | Package cess_cat is already up-to-date!
[nltk_data] | Downloading package cess_esp to /root/nltk_data...
[nltk_data] | Package cess_esp is already up-to-date!
[nltk_data] | Downloading package chat80 to /root/nltk_data...
[nltk_data] | Package chat80 is already up-to-date!
[nltk_data] | Downloading package city_database to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package city_database is already up-to-date!
[nltk_data] | Downloading package cmudict to /root/nltk_data...
[nltk_data] | Package cmudict is already up-to-date!
[nltk_data] | Downloading package comparative_sentences to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package comparative_sentences is already up-to-
[nltk_data] | date!
[nltk_data] | Downloading package comtrans to /root/nltk_data...
[nltk_data] | Package comtrans is already up-to-date!
[nltk_data] | Downloading package conll2000 to /root/nltk_data...
[nltk_data] | Package conll2000 is already up-to-date!
[nltk_data] | Downloading package conll2002 to /root/nltk_data...
[nltk_data] | Package conll2002 is already up-to-date!
[nltk_data] | Downloading package conll2007 to /root/nltk_data...
[nltk_data] | Package conll2007 is already up-to-date!
[nltk_data] | Downloading package crubadan to /root/nltk_data...
[nltk_data] | Package crubadan is already up-to-date!
[nltk_data] | Downloading package dependency_treebank to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package dependency_treebank is already up-to-date!
[nltk_data] | Downloading package dolch to /root/nltk_data...
[nltk_data] | Package dolch is already up-to-date!
[nltk_data] | Downloading package europarl_raw to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package europarl_raw is already up-to-date!
[nltk_data] | Downloading package extended_omw to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package extended_omw is already up-to-date!
[nltk_data] | Downloading package floresta to /root/nltk_data...
[nltk_data] | Package floresta is already up-to-date!
[nltk_data] | Downloading package framenet_v15 to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package framenet_v15 is already up-to-date!
[nltk_data] | Downloading package framenet_v17 to

```

```

[nltk_data] | /root/nltk_data...
[nltk_data] | Package framenet_v17 is already up-to-date!
[nltk_data] | Downloading package gazetteers to /root/nltk_data...
[nltk_data] | Package gazetteers is already up-to-date!
[nltk_data] | Downloading package genesis to /root/nltk_data...
[nltk_data] | Package genesis is already up-to-date!
[nltk_data] | Downloading package gutenber to /root/nltk_data...
[nltk_data] | Package gutenber is already up-to-date!
[nltk_data] | Downloading package ieer to /root/nltk_data...
[nltk_data] | Package ieer is already up-to-date!
[nltk_data] | Downloading package inaugural to /root/nltk_data...
[nltk_data] | Package inaugural is already up-to-date!
[nltk_data] | Downloading package indian to /root/nltk_data...
[nltk_data] | Package indian is already up-to-date!
[nltk_data] | Downloading package jeita to /root/nltk_data...
[nltk_data] | Package jeita is already up-to-date!
[nltk_data] | Downloading package kimmo to /root/nltk_data...
[nltk_data] | Package kimmo is already up-to-date!
[nltk_data] | Downloading package knbc to /root/nltk_data...
[nltk_data] | Package knbc is already up-to-date!
[nltk_data] | Downloading package large_grammars to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package large_grammars is already up-to-date!
[nltk_data] | Downloading package lin_thesaurus to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package lin_thesaurus is already up-to-date!
[nltk_data] | Downloading package mac_morpho to /root/nltk_data...
[nltk_data] | Package mac_morpho is already up-to-date!
[nltk_data] | Downloading package machado to /root/nltk_data...
[nltk_data] | Package machado is already up-to-date!
[nltk_data] | Downloading package masc_tagged to /root/nltk_data...
[nltk_data] | Package masc_tagged is already up-to-date!
[nltk_data] | Downloading package maxent_ne_chunker to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package maxent_ne_chunker is already up-to-date!
[nltk_data] | Downloading package maxent_treebank_pos_tagger to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package maxent_treebank_pos_tagger is already up-
[nltk_data] | to-date!
[nltk_data] | Downloading package mooses_sample to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package mooses_sample is already up-to-date!
[nltk_data] | Downloading package movie_reviews to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package movie_reviews is already up-to-date!
[nltk_data] | Downloading package mte_teip5 to /root/nltk_data...
[nltk_data] | Package mte_teip5 is already up-to-date!
[nltk_data] | Downloading package mwa_ppdb to /root/nltk_data...

```

```

[nltk_data] | Package mwa_ppdb is already up-to-date!
[nltk_data] | Downloading package names to /root/nltk_data...
[nltk_data] | Package names is already up-to-date!
[nltk_data] | Downloading package nombank.1.0 to /root/nltk_data...
[nltk_data] | Package nombank.1.0 is already up-to-date!
[nltk_data] | Downloading package nonbreaking_prefixes to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package nonbreaking_prefixes is already up-to-date!
[nltk_data] | Downloading package nps_chat to /root/nltk_data...
[nltk_data] | Package nps_chat is already up-to-date!
[nltk_data] | Downloading package omw to /root/nltk_data...
[nltk_data] | Package omw is already up-to-date!
[nltk_data] | Downloading package omw-1.4 to /root/nltk_data...
[nltk_data] | Package omw-1.4 is already up-to-date!
[nltk_data] | Downloading package opinion_lexicon to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package opinion_lexicon is already up-to-date!
[nltk_data] | Downloading package panlex_swadesh to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package panlex_swadesh is already up-to-date!
[nltk_data] | Downloading package paradigms to /root/nltk_data...
[nltk_data] | Package paradigms is already up-to-date!
[nltk_data] | Downloading package pe08 to /root/nltk_data...
[nltk_data] | Package pe08 is already up-to-date!
[nltk_data] | Downloading package perluniprops to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package perluniprops is already up-to-date!
[nltk_data] | Downloading package pil to /root/nltk_data...
[nltk_data] | Package pil is already up-to-date!
[nltk_data] | Downloading package pl196x to /root/nltk_data...
[nltk_data] | Package pl196x is already up-to-date!
[nltk_data] | Downloading package porter_test to /root/nltk_data...
[nltk_data] | Package porter_test is already up-to-date!
[nltk_data] | Downloading package ppattach to /root/nltk_data...
[nltk_data] | Package ppattach is already up-to-date!
[nltk_data] | Downloading package problem_reports to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package problem_reports is already up-to-date!
[nltk_data] | Downloading package product_reviews_1 to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package product_reviews_1 is already up-to-date!
[nltk_data] | Downloading package product_reviews_2 to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package product_reviews_2 is already up-to-date!
[nltk_data] | Downloading package propbank to /root/nltk_data...
[nltk_data] | Package propbank is already up-to-date!
[nltk_data] | Downloading package pros_cons to /root/nltk_data...
[nltk_data] | Package pros_cons is already up-to-date!

```

```

[nltk_data] | Downloading package ptb to /root/nltk_data...
[nltk_data] | Package ptb is already up-to-date!
[nltk_data] | Downloading package punkt to /root/nltk_data...
[nltk_data] | Package punkt is already up-to-date!
[nltk_data] | Downloading package qc to /root/nltk_data...
[nltk_data] | Package qc is already up-to-date!
[nltk_data] | Downloading package reuters to /root/nltk_data...
[nltk_data] | Package reuters is already up-to-date!
[nltk_data] | Downloading package rslp to /root/nltk_data...
[nltk_data] | Package rslp is already up-to-date!
[nltk_data] | Downloading package rte to /root/nltk_data...
[nltk_data] | Package rte is already up-to-date!
[nltk_data] | Downloading package sample_grammars to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package sample_grammars is already up-to-date!
[nltk_data] | Downloading package semcor to /root/nltk_data...
[nltk_data] | Package semcor is already up-to-date!
[nltk_data] | Downloading package senseval to /root/nltk_data...
[nltk_data] | Package senseval is already up-to-date!
[nltk_data] | Downloading package sentence_polarity to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package sentence_polarity is already up-to-date!
[nltk_data] | Downloading package sentiwordnet to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package sentiwordnet is already up-to-date!
[nltk_data] | Downloading package shakespeare to /root/nltk_data...
[nltk_data] | Package shakespeare is already up-to-date!
[nltk_data] | Downloading package sinica_treebank to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package sinica_treebank is already up-to-date!
[nltk_data] | Downloading package smultron to /root/nltk_data...
[nltk_data] | Package smultron is already up-to-date!
[nltk_data] | Downloading package snowball_data to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package snowball_data is already up-to-date!
[nltk_data] | Downloading package spanish_grammars to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package spanish_grammars is already up-to-date!
[nltk_data] | Downloading package state_union to /root/nltk_data...
[nltk_data] | Package state_union is already up-to-date!
[nltk_data] | Downloading package stopwords to /root/nltk_data...
[nltk_data] | Package stopwords is already up-to-date!
[nltk_data] | Downloading package subjectivity to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package subjectivity is already up-to-date!
[nltk_data] | Downloading package swadesh to /root/nltk_data...
[nltk_data] | Package swadesh is already up-to-date!
[nltk_data] | Downloading package switchboard to /root/nltk_data...

```

```

[nltk_data] | Package switchboard is already up-to-date!
[nltk_data] | Downloading package tagsets to /root/nltk_data...
[nltk_data] | Package tagsets is already up-to-date!
[nltk_data] | Downloading package timit to /root/nltk_data...
[nltk_data] | Package timit is already up-to-date!
[nltk_data] | Downloading package toolbox to /root/nltk_data...
[nltk_data] | Package toolbox is already up-to-date!
[nltk_data] | Downloading package treebank to /root/nltk_data...
[nltk_data] | Package treebank is already up-to-date!
[nltk_data] | Downloading package twitter_samples to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package twitter_samples is already up-to-date!
[nltk_data] | Downloading package udhr to /root/nltk_data...
[nltk_data] | Package udhr is already up-to-date!
[nltk_data] | Downloading package udhr2 to /root/nltk_data...
[nltk_data] | Package udhr2 is already up-to-date!
[nltk_data] | Downloading package unicode_samples to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package unicode_samples is already up-to-date!
[nltk_data] | Downloading package universal_tagset to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package universal_tagset is already up-to-date!
[nltk_data] | Downloading package universal_treebanks_v20 to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package universal_treebanks_v20 is already up-to-
[nltk_data] | date!
[nltk_data] | Downloading package vader_lexicon to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package vader_lexicon is already up-to-date!
[nltk_data] | Downloading package verbnet to /root/nltk_data...
[nltk_data] | Package verbnet is already up-to-date!
[nltk_data] | Downloading package verbnet3 to /root/nltk_data...
[nltk_data] | Package verbnet3 is already up-to-date!
[nltk_data] | Downloading package webtext to /root/nltk_data...
[nltk_data] | Package webtext is already up-to-date!
[nltk_data] | Downloading package wmt15_eval to /root/nltk_data...
[nltk_data] | Package wmt15_eval is already up-to-date!
[nltk_data] | Downloading package word2vec_sample to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package word2vec_sample is already up-to-date!
[nltk_data] | Downloading package wordnet to /root/nltk_data...
[nltk_data] | Package wordnet is already up-to-date!
[nltk_data] | Downloading package wordnet2021 to /root/nltk_data...
[nltk_data] | Package wordnet2021 is already up-to-date!
[nltk_data] | Downloading package wordnet2022 to /root/nltk_data...
[nltk_data] | Package wordnet2022 is already up-to-date!
[nltk_data] | Downloading package wordnet31 to /root/nltk_data...
[nltk_data] | Package wordnet31 is already up-to-date!

```

```

[nltk_data] | Downloading package wordnet_ic to /root/nltk_data...
[nltk_data] | Package wordnet_ic is already up-to-date!
[nltk_data] | Downloading package words to /root/nltk_data...
[nltk_data] | Package words is already up-to-date!
[nltk_data] | Downloading package ycoe to /root/nltk_data...
[nltk_data] | Package ycoe is already up-to-date!
[nltk_data] |
[nltk_data] Done downloading collection all

```

```
[ ]: True
```

0.0.2 Finding Definition, Examples, and Lemmas of a Noun

Chosen synset is 'city.n.01'.

```
[ ]: wn.synsets("city")
```

```
[ ]: [Synset('city.n.01'), Synset('city.n.02'), Synset('city.n.03')]
```

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

```
[ ]: 'a large and densely populated urban area; may include several independent
administrative districts'
```

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

```
[ ]: ['Ancient Troy was a great city']
```

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

```
[ ]: [Lemma('city.n.01.city'),
      Lemma('city.n.01.metropolis'),
      Lemma('city.n.01.urban_center')]
```

0.0.3 WordNet Noun Hierarchy

This code section traverses up the WordNet hierarchy and prints synsets. WordNet organizes its nouns via hierarchical relations. At the top, synsets retain only abstract and general values. For instance, synsets such as “location” provide a vague but encompassing definition of a city. As you descend the hierarchy, synsets incorporate distinctive and specific features. In this case, the synset, “municipality”, provides a more concrete definition of a city.

```

[ ]: city = wn.synset('city.n.01')
     hyp = city.hypernyms()[0]
     top = wn.synset('entity.n.01')
     while hyp:
         print(hyp)
         if hyp == top:

```

```

    break
    if hyp.hypernyms():
        hyp = hyp.hypernyms()[0]

```

```

Synset('municipality.n.01')
Synset('administrative_district.n.01')
Synset('district.n.01')
Synset('region.n.03')
Synset('location.n.01')
Synset('object.n.01')
Synset('physical_entity.n.01')
Synset('entity.n.01')

```

0.0.4 Extracting Hypernyms, Hyponyms, Meronyms, Holonyms, and Antonyms from a Noun

```
[ ]: wn.synset('city.n.01').hypernyms()
```

```
[ ]: [Synset('municipality.n.01')]
```

```
[ ]: wn.synset('city.n.01').hyponyms()
```

```
[ ]: [Synset('national_capital.n.01'),
      Synset('provincial_capital.n.01'),
      Synset('state_capital.n.01')]
```

```
[ ]: wn.synset('city.n.01').part_meronyms()
```

```
[ ]: [Synset('city_center.n.01'),
      Synset('civic_center.n.02'),
      Synset('concrete_jungle.n.01'),
      Synset('financial_center.n.01'),
      Synset('inner_city.n.01'),
      Synset('medical_center.n.01')]
```

```
[ ]: wn.synset('city.n.01').part_holonyms()
```

```
[ ]: []
```

```
[ ]: for lemma in wn.synset('city.n.01').lemmas():
      print("Lemma:", lemma)
      print("Antonym(s):", lemma.antonyms())

```

```

Lemma: Lemma('city.n.01.city')
Antonym(s): []
Lemma: Lemma('city.n.01.metropolis')
Antonym(s): []

```



```
Lemma: Lemma('city.n.01.urban_center')
Antonym(s): []
```

0.0.5 Finding Definition, Examples, and Lemmas of a Verb

Chosen synset is 'solve.v.01'.

```
[ ]: wn.synsets("solve")

[ ]: [Synset('solve.v.01'), Synset('resolve.v.06'), Synset('clear.v.19')]

[ ]: wn.synset('solve.v.01').definition()

[ ]: 'find the solution to (a problem or question) or understand the meaning of'

[ ]: wn.synset('solve.v.01').examples()

[ ]: ['did you solve the problem?',
      'Work out your problems with the boss',
      "this unpleasant situation isn't going to work itself out",
      'did you get it?',
      'Did you get my meaning?',
      'He could not work the math problem']

[ ]: wn.synset('solve.v.01').lemmas()

[ ]: [Lemma('solve.v.01.solve'),
      Lemma('solve.v.01.work_out'),
      Lemma('solve.v.01.figure_out'),
      Lemma('solve.v.01.puzzle_out'),
      Lemma('solve.v.01.lick'),
      Lemma('solve.v.01.work')]
```

0.0.6 WordNet Verb Hierarchy

WordNet does not provide an abstract top level synset for verbs. The original verb synset yielded only one hypernym synset - "understand.v.01". That particular synset yielded no hypernyms.

```
[ ]: solve = wn.synset('solve.v.01')
      solve.hypernyms()

[ ]: [Synset('understand.v.01')]

[ ]: solve.hypernyms()[0].hypernyms()

[ ]: []
```

0.0.7 Morphy

Morphy returns the root forms of words. It can work with different parts of speech.

```
[ ]: wn.morphy('solved', wn.VERB)
```

```
[ ]: 'solve'
```

```
[ ]: wn.morphy('cities', wn.NOUN)
```

```
[ ]: 'city'
```

0.0.8 Word Similarity

The Wu-Palmer similarity metric yielded a low score for the words — “tree” and “forest”. This suggests the lack of shared common ancestor words.

Lesk algorithm can perform word sense disambiguation. Specifically, it looked at context words and inferred the original synset of the word — “forest”.

```
[ ]: tree = wn.synset("tree.n.01")  
     forest = wn.synset("forest.n.01")
```

```
[ ]: print(wn.wup_similarity(tree, forest))
```

0.125

```
[ ]: sentence = nltk.word_tokenize("I visited the trees in the forest.")  
     print(lesk(sentence, 'forest', 'n'))
```

Synset('forest.n.01')

0.0.9 SentiWordNet

SentiWordNet is a resource built on top of WordNet. For each synset, it can assign three sentiment scores: positivity, negativity, and objectivity. Regarding a possible use case, SentiWordNet can perform sentiment analysis on documents.

0.0.10 Observations and Thoughts

Not surprisingly, the word, “happy”, received a high positive score. I believe this word had a low objective score due to its subjectivity.

With the exception of the adjective, the sentence contained many neutral words. As a result, nearly all of these tokens had a high objective score.

In the broader context, these scores can assist NLP applications. Their utility enables the usage of sentiment analysis. The machine can conceptualize subjective information and assign an associated polarity score.

```
[ ]: happy_synsets = wn.synsets("happy")
senti_list = []
for synset in happy_synsets:
    senti_synset = swn.senti_synset(synset.name())
    print("Synset:", senti_synset.synset.name())
    print("Positive Score:", senti_synset.pos_score())
    print("Negative Score:", senti_synset.neg_score())
    print("Objective Score:", senti_synset.obj_score())
    print("\n")
```

Word: happy.a.01
 Positive Score: 0.875
 Negative Score: 0.0
 Objective Score: 0.125

Word: felicitous.s.02
 Positive Score: 0.75
 Negative Score: 0.0
 Objective Score: 0.25

Word: glad.s.02
 Positive Score: 0.5
 Negative Score: 0.0
 Objective Score: 0.5

Word: happy.s.04
 Positive Score: 0.125
 Negative Score: 0.0
 Objective Score: 0.875

```
[ ]: sentence = "The story had a happy ending"
tokens = sentence.split()
for token in tokens:
    syn_list = list(swn.senti_synsets(token))
    if syn_list:
        print("Token:", token)
        print("Positive Score:", syn_list[0].pos_score())
        print("Negative Score:", syn_list[0].neg_score())
        print("Objective Score:", syn_list[0].obj_score())
        print("\n")
```

Token: story
 Positive Score: 0.0

Negative Score: 0.0
Objective Score: 1.0

Token: had
Positive Score: 0.25
Negative Score: 0.0
Objective Score: 0.75

Token: a
Positive Score: 0.0
Negative Score: 0.0
Objective Score: 1.0

Token: happy
Positive Score: 0.875
Negative Score: 0.0
Objective Score: 0.125

Token: ending
Positive Score: 0.0
Negative Score: 0.0
Objective Score: 1.0

0.0.11 Collocations

Collocations refer to popular combinations of two or more words. These words occur together with a frequency greater than chance. One example is the phrase — “heavy rain”.

```
[ ]: text4.collocations()
```

```
United States; fellow citizens; years ago; four years; Federal  
Government; General Government; American people; Vice President; God  
bless; Chief Justice; one another; fellow Americans; Old World;  
Almighty God; Fellow citizens; Chief Magistrate; every citizen; Indian  
tribes; public debt; foreign nations
```

0.0.12 Mutual Information

This formula finds collocations in text.

```
[ ]: text2.collocations()
```

```
Colonel Brandon; Sir John; Lady Middleton; Miss Dashwood; every thing;  
thousand pounds; dare say; Miss Steeles; said Elinor; Miss Steele;  
every body; John Dashwood; great deal; Harley Street; Berkeley Street;  
Miss Dashwoods; young man; Combe Magna; every day; next morning
```

```
[ ]: text = ' '.join(text2.tokens)  
text[:500]
```

```
[ ]: '[ Sense and Sensibility by Jane Austen 1811 ] CHAPTER 1 The family of Dashwood  
had long been settled in Sussex . Their estate was large , and their residence  
was at Norland Park , in the centre of their property , where , for many  
generations , they had lived in so respectable a manner as to engage the general  
good opinion of their surrounding acquaintance . The late owner of this estate  
was a single man , who lived to a very advanced age , and who for many years of  
his life , had a constant com'
```

0.0.13 Commentary and Interpretation

PMI can range from negative to positive values. Because the pmi is positive, this suggests that the words—“every” and “day”— occur together at a frequency greater than chance. This indicates the presence of a collocation.

```
[ ]: text_size = len(set(text2))  
every_day = text.count('every day')/text_size  
print("p(every day) = ", every_day )  
every = text.count('every')/text_size  
print("p(every) = ", every)  
day = text.count('day')/text_size  
print('p(day) = ', day)  
pmi = math.log2(every_day / (every * day))  
print('pmi = ', pmi)
```

```
p(every day) = 0.004244109468754574  
p(every) = 0.05619786331040538  
p(day) = 0.03424557295477828  
pmi = 1.14095718629765
```