

Gutenberg

```
In [5]: import nltk

In [2]: nltk.__file__
Out[2]: '/usr/local/lib/python3.7/site-packages/nltk/_init_.py'

In [3]: nltk.download('popular')

[nltk_data] Downloading collection 'popular'
[nltk_data] |
[nltk_data] |   Downloading package cmudict to /root/nltk_data...
[nltk_data] |   Unzipping corpora/cmudict.zip.
[nltk_data] |
[nltk_data] |   Downloading package gazetteers to /root/nltk_data...
[nltk_data] |   Unzipping corpora/gazetteers.zip.
[nltk_data] |
[nltk_data] |   Downloading package genesis to /root/nltk_data...
[nltk_data] |   Unzipping corpora/genesis.zip.
[nltk_data] |
[nltk_data] |   Downloading package gutenberg to /root/nltk_data...
[nltk_data] |   Unzipping corpora/gutenberg.zip.
[nltk_data] |
[nltk_data] |   Downloading package inaugural to /root/nltk_data...
[nltk_data] |   Unzipping corpora/inaugural.zip.
[nltk_data] |
[nltk_data] |   Downloading package movie_reviews to
[nltk_data] |   /root/nltk_data.
[nltk_data] |   Unzipping corpora/movie_reviews.zip.
[nltk_data] |
[nltk_data] |   Downloading package names to /root/nltk_data...
[nltk_data] |   Unzipping corpora/names.zip.
[nltk_data] |
[nltk_data] |   Downloading package shakespeare to /root/nltk_data...
[nltk_data] |   Unzipping corpora/shakespeare.zip.
[nltk_data] |
[nltk_data] |   Downloading package stopwords to /root/nltk_data...
[nltk_data] |   Unzipping corpora/stopwords.zip.
[nltk_data] |
[nltk_data] |   Downloading package treebank to /root/nltk_data...
[nltk_data] |   Unzipping corpora/treebank.zip.
[nltk_data] |
[nltk_data] |   Downloading package twitter_samples to
[nltk_data] |   /root/nltk_data.
[nltk_data] |   Unzipping corpora/twitter_samples.zip.
[nltk_data] |
[nltk_data] |   Downloading package omt to /root/nltk_data...
[nltk_data] |   Unzipping corpora/omt.zip.
[nltk_data] |
[nltk_data] |   Downloading package omt-1.4 to /root/nltk_data...
[nltk_data] |   Unzipping corpora/omt-1.4.zip.
[nltk_data] |
[nltk_data] |   Downloading package wordnet to /root/nltk_data...
[nltk_data] |   Unzipping corpora/wordnet.zip.
[nltk_data] |
[nltk_data] |   Downloading package wordnet2001 to /root/nltk_data...
[nltk_data] |   Unzipping corpora/wordnet2001.zip.
[nltk_data] |
[nltk_data] |   Downloading package wordnet31 to /root/nltk_data...
[nltk_data] |   Unzipping corpora/wordnet31.zip.
[nltk_data] |
[nltk_data] |   Downloading package wordnet_lic to /root/nltk_data...
[nltk_data] |   Unzipping corpora/wordnet_lic.zip.
[nltk_data] |
[nltk_data] |   Downloading package words to /root/nltk_data...
[nltk_data] |   Unzipping corpora/words.zip.
[nltk_data] |
[nltk_data] |   Downloading package maxent_ne_chunker to
[nltk_data] |   /root/nltk_data.
[nltk_data] |   Unzipping chunker/./maxent_ne_chunker.zip.
[nltk_data] |
[nltk_data] |   Downloading package punkt to /root/nltk_data...
[nltk_data] |   Unzipping tokenizers/punkt.zip.
[nltk_data] |
[nltk_data] |   Downloading package averaged_perceptron_tagger to
[nltk_data] |   /root/nltk_data.
[nltk_data] |   Unzipping taggers/averaged_perceptron_tagger.zip.
[nltk_data] |
[nltk_data] | Done downloading collection popular

Out[3]: True

In [4]: nltk.download('gutenberg')

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

Out[4]: True

In [5]: from nltk.corpus import gutenberg

In [6]: #read single file
gutenberg.raw('austen-emma.txt')

Out[6]: '[Emma by Jane Austen 1818]\n\nVOLUME I\n\nCHAPTER I\n\nIn\n\nEmma Woodhouse, handsome, clever, and rich, with a comfortable home\n\nand happy disposition, seemed to\n\nbe the best blessings\n\nof existence; and had lived nearly twenty-one years in the world\n\nwith very little to distress or vex her.\n\nShe was the\n\nyoungest of the two daughters of a most affectionate, indulgent father; and had, in consequence of her sister's marriage, been mistress of his house from a\n\nvery early period. Her mother\n\nhad died too long ago for her to have more than an indistinct remembrance of her caresses; and her place had been supplied\n\nby an excellent woman, who had fallen little short of her mother in affection.\n\nSixteen years had Miss Taylor been in Mr. Woodhouse's family, and\n\ness as a governess than a friend, very fond of both daughters, but particularly of Emma. Between them it was more the intimacy of sisters. Even before M\n\niss Taylor had ceased to hold the nominal\n\noffice of governess, the mildness of her temper had hardly allowed\n\nher to impose any restraint; and the shadow of\n\nauthority being\n\nnow laid aside, they had been living together as friend and friend, very mutually attached, and Emma doing just what she liked; highly\n\nesteeming Miss Taylor's judgment, but directed chiefly by her own.\n\nThe real evils, indeed, of Emma's situation were the power of having\n\nrather too much her own way, and a disposition to think a little\n\ntoo well of herself; these were the disadvantages which threatened\n\nall to her many enjoyments. The danger, however, was at present unperceived, that they did not by any means rank as misfortunes\n\nwith her. No sorrow came a gentle sorrow—but not at all in the\n\nshape of any\n\nundiscreet consciousness.—Miss Taylor\n\nmourned. It was Miss Taylor's loss which first brought grief. It was on the wedding-day\n\nof this beloved friend that Emma first sat in mournful thought\n\nof any continuance.'

In [8]: #read multiple files
gutenberg.raw(['austen-emma.txt', 'austen-sense.txt'])

Out[8]: '[Emma by Jane Austen 1818]\n\nVOLUME I\n\nCHAPTER I\n\nIn\n\nEmma Woodhouse, handsome, clever, and rich, with a comfortable home\n\nand happy disposition, seemed to\n\nbe the best blessings\n\nof existence; and had lived nearly twenty-one years in the world\n\nwith very little to distress or vex her.\n\nShe was the\n\nyoungest of the two daughters of a most affectionate, indulgent father; and had, in consequence of her sister's marriage, been mistress of his house from a\n\nvery early period. Her mother\n\nhad died too long ago for her to have more than an indistinct remembrance of her caresses; and her place had been supplied\n\nby an excellent woman, who had fallen little short of her mother in affection.\n\nSixteen years had Miss Taylor been in Mr. Woodhouse's family, and\n\ness as a governess than a friend, very fond of both daughters, but particularly of Emma. Between them it was more the intimacy of sisters. Even before M\n\niss Taylor had ceased to hold the nominal\n\noffice of governess, the mildness of her temper had hardly allowed\n\nher to impose any restraint; and the shadow of\n\nauthority being\n\nnow laid aside, they had been living together as friend and friend, very mutually attached, and Emma doing just what she liked; highly\n\nesteeming Miss Taylor's judgment, but directed chiefly by her own.\n\nThe real evils, indeed, of Emma's situation were the power of having\n\nrather too much her own way, and a disposition to think a little\n\ntoo well of herself; these were the disadvantages which threatened\n\nall to her many enjoyments. The danger, however, was at present unperceived, that they did not by any means rank as misfortunes\n\nwith her. No sorrow came a gentle sorrow—but not at all in the\n\nshape of any\n\nundiscreet consciousness.—Miss Taylor\n\nmourned. It was Miss Taylor's loss which first brought grief. It was on the wedding-day\n\nof this beloved friend that Emma first sat in mournful thought\n\nof any continuance.'
```

```
In [9]: nltk.corpus.gutenberg.categories()

AttributeError                                Traceback (most recent call last)
<ipython-input-9-7dcdf5898ce9> in <module>()
----> 1 nltk.corpus.gutenberg.categories()

AttributeError: 'PlainTextCorpusReader' object has no attribute 'categories'

In [10]: nltk.corpus.gutenberg.fileids()

Out[10]: ['austen-emma.txt',
'austen-persuasion.txt',
'austen-sense.txt',
'bible-kjv.txt',
'blake-poes.txt',
'bryant-stories.txt',
'burgess-busterbrown.txt',
'carroll-alice.txt',
'chesterton-bail.txt',
'chesterton-brown.txt',
'chesterton-thursday.txt',
'edgeworth-paddy.txt',
'fable-moby-dick.txt',
'hallion-paradise.txt',
'hawthorne-cathedral.txt',
'shakespeare-hamlet.txt',
'shakespeare-macbeth.txt',
'whitman-leaves.txt']

In [14]: nltk.corpus.gutenberg.words(fileids=['austen-emma.txt', 'whitman-leaves.txt'])

Out[14]: ['I', 'Emma', 'by', 'Jane', 'Austen', '1816', ''], ...]

In [16]: gutenberg.raw(fileids=['austen-emma.txt', 'whitman-leaves.txt'])

Out[16]: '[Emma by Jane Austen 1818]\n\nVOLUME I\n\nCHAPTER I\n\nIn\n\nEmma Woodhouse, handsome, clever, and rich, with a comfortable home\n\nand happy disposition, seemed to\n\nbe the best blessings\n\nof existence; and had lived nearly twenty-one years in the world\n\nwith very little to distress or vex her.\n\nShe was the\n\nyoungest of the two daughters of a most affectionate, indulgent father; and had, in consequence of her sister's marriage, been mistress of his house from a\n\nvery early period. Her mother\n\nhad died too long ago for her to have more than an indistinct remembrance of her caresses; and her place had been supplied\n\nby an excellent woman, who had fallen little short of her mother in affection.\n\nSixteen years had Miss Taylor been in Mr. Woodhouse's family, and\n\ness as a governess than a friend, very fond of both daughters, but particularly of Emma. Between them it was more the intimacy of sisters. Even before M\n\niss Taylor had ceased to hold the nominal\n\noffice of governess, the mildness of her temper had hardly allowed\n\nher to impose any restraint; and the shadow of\n\nauthority being\n\nnow laid aside, they had been living together as friend and friend, very mutually attached, and Emma doing just what she liked; highly\n\nesteeming Miss Taylor's judgment, but directed chiefly by her own.\n\nThe real evils, indeed, of Emma's situation were the power of having\n\nrather too much her own way, and a disposition to think a little\n\ntoo well of herself; these were the disadvantages which threatened\n\nall to her many enjoyments. The danger, however, was at present unperceived, that they did not by any means rank as misfortunes\n\nwith her. No sorrow came a gentle sorrow—but not at all in the\n\nshape of any\n\nundiscreet consciousness.—Miss Taylor\n\nmourned. It was Miss Taylor's loss which first brought grief. It was on the wedding-day\n\nof this beloved friend that Emma first sat in mournful thought\n\nof any continuance.'
```

```
In [17]: len(gutenberg.raw(fileids=['austen-emma.txt', 'whitman-leaves.txt']))

Out[17]: 158286

In [18]: len(nltk.corpus.gutenberg.words(fileids=['austen-emma.txt', 'whitman-leaves.txt']))

Out[18]: 347310

In [20]: len(gutenberg.sents())

Out[20]: 98503

In [23]: gutenberg.root

FileNotFoundError: [Errno 2] No such file or directory: '/root/nltk_data/corpora/gutenberg'
```

```
In [24]: gutenberg.encoding('austen-emma.txt')

Out[24]: 'latin1'

In [25]: gutenberg.readme()

Project Gutenberg Selections
http://gutenberg.net/

This corpus contains texts from From Project Gutenberg, by the following authors:

(1) Jane Austen
(2) William Blake
(3) Milton
(4) Thornton W. Burgess
(5) Sarah Coar
(6) Lewis Carroll
(7) G. K. Chesterton
(8) Maria Edgeworth
(9) King James Bible
(10) Herman Melville
(11) John Milton
(12) William Shakespeare
(13) Walt Whitman
(14) The beginning of the body of each book could not be identified automatically, so the
general header of each file has been removed, and included below. Some source files ended with a line "End of The Project Gutenberg eBook...", and this h
as been deleted.

Information about Project Gutenberg (one page)

We produce about two million files for each hour we work.
```

brown

```
In [26]: nltk.download('brown')

[nltk_data] Downloading package brown to /root/nltk_data...
[nltk_data] Unzipping corpora/brown.zip.

Out[26]: True

In [27]: from nltk.corpus import brown

In [28]: nltk.corpus.brown.fileids()

Out[28]: ['ca01',
'ca02',
'ca03',
'ca04',
'ca05',
'ca06',
'ca07',
'ca08',
'ca09',
'ca10',
'ca11',
'ca12',
'ca13',
'ca14',
'ca15',
'ca16',
'ca17',
'ca18',
'ca19',
'ca20',
'ca21',
'ca22',
'ca23',
'ca24',
'ca25',
'ca26',
'ca27',
'ca28',
'ca29',
'ca30',
'ca31',
'ca32',
'ca33',
'ca34',
'ca35',
'ca36',
'ca37',
'ca38',
'ca39',
'ca40',
'ca41',
'ca42',
'ca43',
'ca44',
'ca45',
'cb01',
'cb02',
'cb03',
'cb04',
'cb05',
'cb06',
'cb07',
'cb08',
'cb09',
'cb09',]

In [29]: nltk.corpus.brown.categories()

Out[29]: ['adventure',
'dellentes',
'letters',
'editorial',
'fiction',
'government',
'lectures',
'humor',
'learned',
'news',
'mystery',
'reviews',
'romance',
'science-fiction']

In [30]: brown.raw('ca01')

Out[30]: '\n\nThe/at Fulton/np-tl County/mn-tl Grand/jj-tl Jury/mn-tl said/vbd Friday/rn an/at investigation/mn of/in Atlanta/nps recent/jj primary/m election/mn
produced/vbd /:/ not evidence/m /:/ that/cs any/dti irregularities/mns took/vbd place/mn /:/\n\n\nThe/at Jury/m further/rbr said/vbd in/in term-end/
n present/mns that/cs the/at City/mn-tl Executive/jj-tl Committee/mn-tl /:/, which/wdt had/vbd over-all/jj charge/mn off/in the/at election/mn /:/,
d eserves/vbz the/at praise/m and/cc thanks/mns off/in the/at City/mn-tl of/in-tl Atlanta/mn-tl /:/ for/in the/at manner/m in/in which/wdt the/at election/mn
was/bdz conducted/vbd /:/\n\n\nThe/at September-October/m term/mn Jury/m had/vbd been/bn charged/vbn by/in Fulton/np-tl Superior/jj-tl Court/mn-tl Judg
e/mn-tl Burwood/np Pye/np to/to investigate/vb reports/mns off/in possible/jj /:/ Irregularities/mns /:/ in/in the/at hard-fought/rj primary/m which/wdt w
as/bdz won/vbn by/in Mayor-nominate/mn-tl Ivan/np Allen/np Jr./np /:/\n\n\nThe/at /:/ not/rb a/at relative/jj handful/m off/in such/jj

In [32]: brown.raw(['ca01', 'ca22'])

Out[32]: '\n\nThe/at Fulton/np-tl County/mn-tl Grand/jj-tl Jury/mn-tl said/vbd Friday/rn an/at investigation/mn of/in Atlanta/nps recent/jj primary/m election/mn
produced/vbd /:/ not evidence/m /:/ that/cs any/dti irregularities/mns took/vbd place/mn /:/\n\n\nThe/at Jury/m further/rbr said/vbd in/in term-end/
n present/mns that/cs the/at City/mn-tl Executive/jj-tl Committee/mn-tl /:/, which/wdt had/vbd over-all/jj charge/mn off/in the/at election/mn /:/,
d eserves/vbz the/at praise/m and/cc thanks/mns off/in the/at City/mn-tl of/in-tl Atlanta/mn-tl /:/ for/in the/at manner/m in/in which/wdt the/at election/mn
was/bdz conducted/vbd /:/\n\n\nThe/at September-October/m term/mn Jury/m had/vbd been/bn charged/vbn by/in Fulton/np-tl Superior/jj-tl Court/mn-tl Judg
e/mn-tl Burwood/np Pye/np to/to investigate/vb reports/mns off/in possible/jj /:/ Irregularities/mns /:/ in/in the/at hard-fought/rj primary/m which/wdt w
as/bdz won/vbn by/in Mayor-nominate/mn-tl Ivan/np Allen/np Jr./np /:/\n\n\nThe/at /:/ not/rb a/at relative/jj handful/m off/in such/jj reports/mns was/bdz rec
eived/vbn /:/, /:/, the/at Jury/m said/vbd /:/, /:/ considering/in the/at widespread/jj

In [33]: nltk.corpus.brown.words(fileids=['ca01', 'ca02'])

Out[33]: ['The', 'Fulton', 'County', 'Grand', 'Jury', 'said', ...]

In [34]: len(brown.raw(['ca01', 'ca22']))

Out[34]: 40792

In [35]: len(nltk.corpus.brown.words(fileids=['ca01', 'ca02']))

Out[35]: 4519

In [36]: len(brown.sents())

Out[36]: 57348

In [37]: brown.raw(categories=['adventure', 'action'])

Out[37]: '\n\nThe/at Morgan/np told/vbd himself/vgtl he/pps would/nd forget/vb Ann/np Turner/np /:/\n\nThe/pps was/bdz well/rbr rid/jj of/in her/ppo /:/\n\nThe/pps certainl
y/sbg did/nd t/dod want/vb a/at wife/m who/pps was/bdz fickle/jj as/cs Ann/np /:/\n\nIf/cs he/pps had/vbd married/vbn her/ppo /:/, he/d/pps had/hv been/bn a
s/rng vdg for/in trouble/m /:/\n\n\nThe/at but/cc all/abn off/in this/dt was/bdz rationalization/m /:/\n\nSometimes/rb he/pps woked/vbd up/rp in/in the/at middle/m
of/in the/at night/m thinking/vb of/in Ann/np /:/, and/cc the/rb could/nd no/t get/vb back/rb to/in sleep/m /:/\n\nHis/pps plans/mns and/cc dreams/mns had/v
d evolved/vbn around/in her/ppo so/ql much/rb and/cc for/in so/ql long/jj that/cs now/rb he/pps felt/vbd as/cs if/cs he/pps had/vbd nothing/np /:/\n\nThe/at e
asiest/jjt thing/mn would/nd be/be to/to sell/vb out/rp to/in Al/np Budd/np and/cc

In [38]: brown.categories(['ca01'])

Out[38]: ['news']

In [39]: brown.fileids(['adventure'])

Out[39]: ['cb01',
'cb02',
'cb03',
'cb04',
'cb05',
'cb06',
'cb07',
'cb08',
'cb09',
'cb10',
'cb11',
'cb12',
'cb13',
'cb14',
'cb15',
'cb16',
'cb17',
'cb18',
'cb19',
'cb20',
'cb21',
'cb22',
'cb23',
'cb24',
'cb25',
'cb26',
'cb27',
'cb28',
'cb29',
'cb30',
'cb31',
'cb32',
'cb33',
'cb34',
'cb35',
'cb36',
'cb37',
'cb38',
'cb39',
'cb40',
'cb41',
'cb42',
'cb43',
'cb44',
'cb45',
'cb46',
'cb47',
'cb48',
'cb49',
'cb50',
'cb51',
'cb52',
'cb53',
'cb54',
'cb55',
'cb56',
'cb57',
'cb58',
'cb59',
'cb60',
'cb61',
'cb62',
'cb63',
'cb64',
'cb65',
'cb66',
'cb67',
'cb68',
'cb69',
'cb70',
'cb71',
'cb72',
'cb73',
'cb74',
'cb75',
'cb76',
'cb77',
'cb78',
'cb79',
'cb80',
'cb81',
'cb82',
'cb83',
'cb84',
'cb85',
'cb86',
'cb87',
'cb88',
'cb89',
'cb90',
'cb91',
'cb92',
'cb93',
'cb94',
'cb95',
'cb96',
'cb97',
'cb98',
'cb99',
'cb100',
'cb101',
'cb102',
'cb103',
'cb104',
'cb105',
'cb106',
'cb107',
'cb108',
'cb109',
'cb110',
'cb111',
'cb112',
'cb113',
'cb114',
'cb115',
'cb116',
'cb117',
'cb118',
'cb119',
'cb120',
'cb121',
'cb122',
'cb123',
'cb124',
'cb125',
'cb126',
'cb127',
'cb128',
'cb129',
'cb130',
'cb131',
'cb132',
'cb133',
'cb134',
'cb135',
'cb136',
'cb137',
'cb138',
'cb139',
'cb140',
'cb141',
'cb142',
'cb143',
'cb144',
'cb145',
'cb146',
'cb147',
'cb148',
'cb149',
'cb150',
'cb151',
'cb152',
'cb153',
'cb154',
'cb155',
'cb156',
'cb157',
'cb158',
'cb159',
'cb160',
'cb161',
'cb162',
'cb163',
'cb164',
'cb165',
'cb166',
'cb167',
'cb168',
'cb169',
'cb170',
'cb171',
'cb172',
'cb173',
'cb174',
'cb175',
'cb176',
'cb177',
'cb178',
'cb179',
'cb180',
'cb181',
'cb182',
'cb183',
'cb184',
'cb185',
'cb186',
'cb187',
'cb188',
'cb189',
'cb190',
'cb191',
'cb192',
'cb193',
'cb194',
'cb195',
'cb196',
'cb197',
'cb198',
'cb199',
'cb200',
'cb201',
'cb202',
'cb203',
'cb204',
'cb205',
'cb206',
'cb207',
'cb208',
'cb209',
'cb210',
'cb211',
'cb212',
'cb213',
'cb214',
'cb215',
'cb216',
'cb217',
'cb218',
'cb219',
'cb220',
'cb221',
'cb222',
'cb223',
'cb224',
'cb225',
'cb226',
'cb227',
'cb228',
'cb229',
'cb230',
'cb231',
'cb232',
'cb233',
'cb234',
'cb235',
'cb236',
'cb237',
'cb238',
'cb239',
'cb240',
'cb241',
'cb242',
'cb243',
'cb244',
'cb245',
'cb246',
'cb247',
'cb248',
'cb249',
'cb250',
'cb251',
'cb252',
'cb253',
'cb254',
'cb255',
'cb256',
'cb257',
'cb258',
'cb259',
'cb260',
'cb261',
'cb262',
'cb263',
'cb264',
'cb265',
'cb266',
'cb267',
'cb268',
'cb269',
'cb270',
'cb271',
'cb272',
'cb273',
'cb274',
'cb275',
'cb276',
'cb277',
'cb278',
'cb279',
'cb280',
'cb281',
'cb282',
'cb283',
'cb284',
'cb285',
'cb286',
'cb287',
'cb288',
'cb289',
'cb290',
'cb291',
'cb292',
'cb293',
'cb294',
'cb295',
'cb296',
'cb297',
'cb298',
'cb299',
'cb300',
'cb301',
'cb302',
'cb303',
'cb304',
'cb305',
'cb306',
'cb307',
'cb308',
'cb309',
'cb310',
'cb311',
'cb312',
'cb313',
'cb314',
'cb315',
'cb316',
'cb317',
'cb318',
'cb319',
'cb320',
'cb321',
'cb322',
'cb323',
'cb324',
'cb325',
'cb326',
'cb327',
'cb328',
'cb329',
'cb330',
'cb331',
'cb332',
'cb333',
'cb334',
'cb335',
'cb336',
'cb337',
'cb338',
'cb339',
'cb340',
'cb341',
'cb342',
'cb343',
'cb344',
'cb345',
'cb346',
'cb347',
'cb348',
'cb349',
'cb350',
'cb351',
'cb352',
'cb353',
'cb354',
'cb355',
'cb356',
'cb357',
'cb358',
'cb359',
'cb360',
'cb361',
'cb362',
'cb363',
'cb364',
'cb365',
'cb366',
'cb367',
'cb368',
'cb369',
'cb370',
'cb371',
'cb372',
'cb373',
'cb374',
'cb375',
'cb376',
'cb377',
'cb378',
'cb379',
'cb380',
'cb381',
'cb382',
'cb383',
'cb384',
'cb385',
'cb386',
'cb387',
'cb388',
'cb389',
'cb390',
'cb391',
'cb392',
'cb393',
'cb394',
'cb395',
'cb396',
'cb397',
'cb398',
'cb399',
'cb400',
'cb401',
'cb402',
'cb403',
'cb404',
'cb405',
'cb406',
'cb407',
'cb408',
'cb409',
'cb410',
'cb411',
'cb412',
'cb413',
'cb414',
'cb415',
'cb416',
'cb417',
'cb418',
'cb419',
'cb420',
'cb421',
'cb422',
'cb423',
'cb424',
'cb425',
'cb426',
'cb427',
'cb428',
'cb429',
'cb430',
'cb431',
'cb432',
'cb433',
'cb434',
'cb435',
'cb436',
'cb437',
'cb438',
'cb439',
'cb440',
'cb441',
'cb442',
'cb443',
'cb444',
'cb445',
'cb446',
'cb447',
'cb448',
'cb449',
'cb450',
'cb451',
'cb452',
'cb453',
'cb454',
'cb455',
'cb456',
'cb457',
'cb458',
'cb459',
'cb460',
'cb461',
'cb462',
'cb463',
'cb464',
'cb465',
'cb466',
'cb467',
'cb468',
'cb469',
'cb470',
'cb471',
'cb472',
'cb473',
'cb474',
'cb475',
'cb476',
'cb477',
'cb478',
'cb479',
'cb480',
'cb481',
'cb482',
'cb483',
'cb484',
'cb485',
'cb486',
'cb487',
'cb488',
'cb489',
'cb490',
'cb491',
'cb492',
'cb493',
'cb494',
'cb495',
'cb496',
'cb497',
'cb498',
'cb499',
'cb500',
'cb501',
'cb502',
'cb503',
'cb504',
'cb505',
'cb506',
'cb507',
'cb508',
'cb509',
'cb510',
'cb511',
'cb512',
'cb513',
'cb514',
'cb515',
'cb516',
'cb517',
'cb518',
'cb519',
'cb520',
'cb521',
'cb522',
'cb523',
'cb524',
'cb525',
'cb526',
'cb527',
'cb528',
'cb529',
'cb530',
'cb531',
'cb532',
'cb533',
'cb534',
'cb535',
'cb536',
'cb537',
'cb538',
'cb539',
'cb540',
'cb541',
'cb542',
'cb543',
'cb544',
'cb545',
'cb546',
'cb547',
'cb548',
'cb549',
'cb550',
'cb551',
'cb552',
'cb553',
'cb554',
'cb555',
'cb556',
'cb557',
'cb558',
'cb559',
'cb560',
'cb561',
'cb562',
'cb563',
'cb564',
'cb565',
'cb566',
'cb567',
'cb568',
'cb569',
'cb570',
'cb571',
'cb572',
'cb573',
'cb574',
'cb575',
'cb576',
'cb577',
'cb578',
'cb579',
'cb580',
'cb581',
'cb582',
'cb583',
'cb584',
'cb585',
'cb586',
'cb587',
'cb588',
'cb589',
'cb590',
'cb591',
'cb592',
'cb593',
'cb594',
'cb595',
'cb596',
'cb597',
'cb598',
'cb599',
'cb600',
'cb601',
'cb602',
'cb603',
'cb604',
'cb605',
'cb606',
'cb607',
'cb608',
'cb609',
'cb610',
'cb611',
'cb612',
'cb613',
'cb614',
'cb615',
'cb616',
'cb617',
'cb618',
'cb619',
'cb620',
'cb621',
'cb622',
'cb623',
'cb624',
'cb625',
'cb626',
'cb627',
'cb628',
'cb629',
'cb630',
'cb631',
'cb632',
'cb633',
'cb634',
'cb635',
'cb636',
'cb637',
'cb638',
'cb639',
'cb640',
'cb641',
'cb642',
'cb643',
'cb644',
'cb645',
'cb646',
'cb647',
'cb648',
'cb649',
'cb650',
'cb651',
'cb652',
'cb653',
'cb654',
'cb655',
'cb656',
'cb657',
'cb658',
'cb659',
'cb660',
'cb661',
'cb662',
'cb663',
'cb664',
'cb665',
'cb666',
'cb667',
'cb668',
'cb669',
'cb670',
'cb671',
'cb672',
'cb673',
'cb674',
'cb675',
'cb676',
'cb677',
'cb678',
'cb679',
'cb680',
'cb681',
'cb682',
'cb683',
'cb684',
'cb685',
'cb686',
'cb687',
'cb688',
'cb689',
'cb690',
'cb691',
'cb692',
'cb693',
'cb694',
'cb695',
'cb696',
'cb697',
'cb698',
'cb699',
'cb700',
'cb701',
'cb702',
'cb703',
'cb704',
'cb705',
'cb706',
'cb707',
'cb708',
'cb709',
'cb710',
'cb711',
'cb712',
'cb713',
'cb714',
'cb715',
'cb716',
'cb717',
'cb718',
'cb719',
'cb720',
'cb721',
'cb722',
'cb723',
'cb724',
'cb725',
'cb726',
'cb727',
'cb728',
'cb729',
'cb730',
'cb731',
'cb732',
'cb733',
'cb734',
'cb735',
'cb736',
'cb737',
'cb738',
'cb739',
'cb740',
'cb741',
'cb742',
'cb743',
'cb744',
'cb745',
'cb746',
'cb747',
'cb748',
'cb749',
'cb750',
'cb751',
'cb752',
'cb753',
'cb754',
'cb755',
'cb756',
'cb757',
'cb758',
'cb759',
'cb760',
'cb761',
'cb762',
'cb763',
'cb764',
'cb765',
'cb766',
'cb767',
'cb768',
'cb769',
'cb770',
'cb771',
'cb772',
'cb773',
'cb774',
'cb775',
'cb776',
'cb777',
'cb778',
'cb779',
'cb780',
'cb781',
'cb782',
'cb783',
'cb784',
'cb785',
'cb786',
'cb787',
'cb788',
'cb789',
'cb790',
'cb791',
'cb792',
'cb793',
'cb794',
'cb795',
'cb796',
'cb797',
'cb798',
'cb799',
'cb800',
'cb801',
'cb802',
'cb803',
'cb804',
'cb805',
'cb806',
'cb807',
'cb808',
'cb809',
'cb810',
'cb811',
'cb812',
'cb813',
'cb814',
'cb815',
'cb816',
'cb817',
'cb818',
'cb819',
'cb820',
'cb821',
'cb822',
'cb823',
'cb824',
'cb825',
'cb826',
'cb827',
'cb828',
'cb829',
'cb830',
'cb831',
'cb832',
'cb833',
'cb834',
'cb835',
'cb836',
'cb837',
'cb838',
'cb839',
'cb840',
'cb841',
'cb842',
'cb843',
'cb844',
'cb845',
'cb846',
'cb847',
'cb848',
'cb849',
'cb850',
'cb851',
'cb852',
'cb853',
'cb854',
'cb855',
'cb856',
'cb857',
'cb858',
'cb859',
'cb860',
'cb861',
'cb862',
'cb863',
'cb864',
'cb865',
'cb866',
'cb867',
'cb868',
'cb869',
'cb870',
'cb871',
'cb872',
'cb873',
'cb874',
'cb875',
'cb876',
'cb877',
'cb878',
'cb879',
'cb880',
'cb881',
'cb882',
'cb883',
'cb884',
'cb885',
'cb886',
'cb887',
'cb888',
'cb889',
'cb890',
'cb891',
'cb892',
'cb893',
'cb894',
'cb895',
'cb896',
'cb897',
'cb898',
'cb899',
'cb900',
'cb901',
'cb902',
'cb903',
'cb904',
'cb905',
'cb906',
'cb907',
'cb908',
'cb909',
'cb910',
'cb911',
'cb912',
'cb913',
'cb914',
'cb915',
'cb916',
'cb917',
'cb918',
'cb919',
'cb920',
'cb921',
'cb922',
'cb923',
'cb924',
'cb925',
'cb926',
'cb927',
'cb928',
'cb929',
'cb930',
'cb931',
'cb932',
'cb933',
'cb934',
'cb935',
'cb936',
'cb937',
'cb938',
'cb939',
'cb940',
'cb941',
'cb942',
'cb943',
'cb944',
'cb945',
'cb946',
'cb947',
'cb948',
'cb949',
'cb950',
'cb951',
'cb952',
'cb953',
'cb954',
'cb955',
'cb956',
'cb957',
'cb958',
'cb959',
'cb960',
'cb961',
'cb962',
'cb963',
'cb964',
'cb965',
'cb966',
'cb967',
'cb968',
'cb969',
'cb970',
'cb971',
'cb972',
'cb973',
'cb974',
'cb975',
'cb976',
'cb977',
'cb978',
'cb979',
'cb980',
'cb981',
'cb982',
'cb983',
'cb984',
'cb985',
'cb986',
'cb987',
'cb988',
'cb989',
'cb990',
'cb991',
'cb992',
'cb993',
'cb994',
'cb995',
'cb996',
'cb997',
'cb998',
'cb999',
'cb1000']

In [40]: brown.root

FileNotFoundError: [Errno 2] No such file or directory: '/root/nltk_data/corpora/brown'
```

```
In [41]: brown.encoding('ca01')

Out[41]: 'ascii'

In [42]: brown.readme()

BROWN CORPUS
The Brown Corpus is a Standard Corpus of Present-Day Edited American English, for use with Digital Computers.
Compiled by Kenneth W. B. Brown, University of Texas at Austin, 1961.
Revised and Amplified 1979 by John F. Lincoln, University of Texas at Austin.
The
```