

Assignment 3

Chapter 4. Writing Structured Programs

```
In [1]: import re
import nltk
from nltk.book import *
from nltk import word_tokenize
from nltk import FreqDist
from urllib import request
from bs4 import BeautifulSoup
from nltk.corpus import stopwords
from nltk.corpus import wordnet as wn
```

```
*** Introductory Examples for the NLTK Book ***
Loading text1, ..., text9 and sent1, ..., sent9
Type the name of the text or sentence to view it.
Type: 'texts()' or 'sents()' to list the materials.
text1: Moby Dick by Herman Melville 1851
text2: Sense and Sensibility by Jane Austen 1811
text3: The Book of Genesis
text4: Inaugural Address Corpus
text5: Chat Corpus
text6: Monty Python and the Holy Grail
text7: Wall Street Journal
text8: Personals Corpus
text9: The Man Who Was Thursday by G . K . Chesterton 1908
```

1. Create a list of words = ['is', 'it', 'good', '?']. a) Use a series of assignment statements (e.g. words[1] = words[2]) and a temporary variable tmp to transform this list into the list ['it', 'is', 'good', '!']. b) Now do the same transformation using tuple assignment.

```
In [2]: words_1 = ['is', 'it', 'good', '?']
tmp = words_1[1]
words_1[1] = words_1[0]
words_1[0] = tmp
words_1[3] = '!'
words_1
```

```
Out[2]: ['it', 'is', 'good', '!']
```

```
In [3]: words_2 = ['is', 'it', 'good', '?']
(words_2[0], words_2[1], words_2[3]) = (words_2[1], words_2[0], '!')
words_2
```

```
Out[3]: ['it', 'is', 'good', '!']
```

2. Write code that removes whitespace at the beginning and end of a string (' this is a sample sentence '), and normalizes whitespace between words to be a single space character.

a) do this task using split() and join()

b) do this task using regular expression substitutions

```
In [4]: ##### Using split() and join()
sample = ' the red balloon was green '
sample_split = sample.split()
sample_word = [word for word in sample_split]
' '.join(sample_word)
```

```
Out[4]: 'the red balloon was green'
```

```
In [5]: ##### Using regular expression substitutions
reg_sample = ' the red balloon was green '
#removes leading and trailing whitespace
lt_ws_remove = re.sub('^\s+|\s+$','',reg_sample)
#removes the multiple whitespaces
sentence_fix = re.sub(r'\s+', ' ',lt_ws_remove)
sentence_fix
```

```
Out[5]: 'the red balloon was green'
```

I tried to get this all on one line of regex sub() but couldn't figure it out

3. sent1=['The', 'dog', 'gave', 'John', 'the', 'newspaper']. Now assign sent2=sent1. Modify sent1[1]='monkey'. Please review section 4.1 -Assignment in Chapter 4 to answer the following questions:

a) verify that sent2 has changed

b) Now try the same exercise but instead assign sent2=sent1[:]. Modify sent1[1]='monkey' and see what happens to sent2. Explain.

c) Now define text1=[['The', 'dog', 'gave', 'John', 'the', 'newspaper'], ['John', 'is', 'happy']]. Now assign text2=text1[:], assign a new value to one of the words (text1[0][1]='monkey'). Check what happens to text2. Explain.

d) Extract successive overlapping 4-grams from ['The', 'dog', 'gave', 'John', 'the', 'newspaper'].

```
In [6]: sent1=['The', 'dog', 'gave', 'John', 'the', 'newspaper']
```

```
In [7]: sent2 = sent1
```

```
In [8]: sent1[1] = 'monkey'
```

```
In [9]: ##### a) verify that sent2 has changed
sent2
```

```
Out[9]: ['The', 'monkey', 'gave', 'John', 'the', 'newspaper']
```

```
In [10]: ##### b) Now try the same exercise but instead assign sent2=sent1[:].
#####Modify sent1[1]='monkey' and see what happens to sent2. Explain.
sent1=['The', 'dog', 'gave', 'John', 'the', 'newspaper']
sent2 = sent1[:]
sent1[1] = 'monkey'
sent2
```

```
Out[10]: ['The', 'dog', 'gave', 'John', 'the', 'newspaper']
```

b) When using the [:] we copy the object references inside the list. This allows you to copy the contents to a new list without them changing.

```
In [11]: ##### c) Now define text1=[['The', 'dog', 'gave', 'John', 'the', 'newspaper'], ['John', 'is', 'happy']]
##### Now assign text2=text1[:], assign a new value to one of the words (text1[0][1]='monkey').
##### Check what happens to text2. Explain.
text1=[['The', 'dog', 'gave', 'John', 'the', 'newspaper'], ['John', 'is', 'happy']]
text2=text1[:]
text1[0][3]='Mikey'
text2
```

```
Out[11]: [['The', 'dog', 'gave', 'Mikey', 'the', 'newspaper'], ['John', 'is', 'happy']]
```

c) In this case the objects within the list did change because we were changing the list within the last which was not copied - only the text1 list was copied. The contents within are still able to be changed.

```
In [12]: ##### d) Extract successive overlapping 4-grams from ['The', 'dog', 'gave', 'John', 'the', 'newspaper']
sent1 = ['The', 'dog', 'gave', 'John', 'the', 'newspaper']
n = 4
[sent1[i:i+n] for i in range(len(sent1)-n+1)]
```

```
Out[12]: [['The', 'dog', 'gave', 'John'],
['dog', 'gave', 'John', 'the'],
['gave', 'John', 'the', 'newspaper']]
```

4. Write a function that prints any word that appeared in the last 20% of a text that had not been encountered earlier. Use text1 from nltk.book to call this function.

```
In [13]: text1 = nltk.book.text1
```

```
In [14]: def word_20(text1):
        cut = int(.8 * len(text1))
        for w in text1[cut:]:
            if w not in text1[:cut]:
                print(w)

word_20(text1)
```

```
geologist
mason
digger
ditches
canals
wells
cisterns
remind
fossils
Tertiary
formations
intercepted
antichronical
Ark
Tertiary
formations
justify
fossils
Detached
fossils
adamite
Scotland
Louisiana
1779
disinterred
Rue
Dauphine
Paris
disinterred
Antwerp
1842
Creagh
Basilosaurus
Anatomist
rechristened
Zeuglodon
Geological
mutations
blotted
characterized
partial
resemblances
```

affinities
antichronical
seniors
Saturn
bastions
Tropics
inhabitable
Himmalehs
pedigree
older
Pharaoh
Methuselah
Shem
antemosaic
unsourced
adamite
stereotype
limestone
marl
bequeathed
print
Denderah
planisphere
griffins
grotesque
moderns
Gliding
planisphere
cradled
attestation
osseous
diluvian
Rafters
Beams
Bones
Common
Power
Matter
Rocks
Miles
Rib
Miracle
Camel
Rib
layn
Historians
affirm
Mahomet
Base
105
Magnitude
Diminish
Perish
Inasmuch
Eternities
inquired
degenerated
investigation
fossil
Tertiary
Tertiary
formations
adamite
exhumed
tape
degenerated
Assuredly
Aldrovandus
Walks
Tunnels

Solander
Cooke
Academy
Sciences
reydan
siskur
Wrinkled
Bellies
metres
ancestors
coffins
stall
Smithfield
fattest
Pharaoh
kine
degenerated
omniscient
whaleships
penetrating
exterminated
Comparing
overspread
manes
scowled
sites
extinction
extermination
forbids
Canadian
moccasined
statistically
gradual
extinction
remunerative
influenced
yokes
fallacious
haunt
declining
cape
unfamiliar
invasion
Swiss
savannas
citadels
barriers
floes
defiance
cachalot
annually
populousness
Harto
Goa
droves
Semiramis
Hannibal
survive
outlast
Americas
Isles
longevity
imagining
perishable
continents
Kremlin
despised
Ark
Netherlands
rats

survive
crest
frothed
defiance
106
precipitating
thwart
splintering
urgent
inflexibly
wrench
lusty
prone
insensible
displaced
groin
cured
marsh
perpetuates
songster
ancestry
Grief
ancestry
Joy
canonic
enjoyments
childlessness
fertilely
progressive
inequality
analysis
felicities
unsignifying
pettiness
heartwoes
archangelic
tracings
deduction
sourceless
primogenitures
cymballing
moons
ineffaceable
signers
Unwittingly
Lama
exclusiveness
senate
revelation
partook
explanatory
recluseness
banned
moodily
unaccounted
underived
wails
conspired
muffle
transpire
synod
potentates
procedures
functionary
studs
clearest
fittings
forge
blacksmith
forging

107
Carpenter
Seat
sultanically
moons
Saturn
abstracted
duplicates
abstraction
callings
outbranching
handicrafts
efficient
repairing
reforming
bladed
unhesitatingly
expert
conflicting
aptitudes
enacted
sizes
athwartships
belaying
strays
sprains
concocts
lotion
longed
vermillion
drills
pincers
bids
unmanageably
winces
unconcluded
capstans
expertness
vivacity
impersonal
stolidity
impersonal
stolidity
pauselessly
ignores
stolidity
involving
ramifying
heartlessness
oddly
wheezing
unstreaked
wittiness
wanderer
clingings
pertained
abstract
unfractioned
babe
premeditated
uncompromisedness
unintelligence
intermixture
uneven
spontaneous
literal
manipulator
oozed
MULTUM
PARVO

Sheffield
sizes
drivers
screws
tweezers
awls
rulers
filers
countersinkers
superiors
tweezers
omnitooled
machine
automaton
quicksilver
abided
soliloquizing
hummingly
soliloquizes
soliloquizer
108
Carpenter
CARPENTER
VICE
BENCH
LIGHT
LANTERNS
BUSILY
FILING
IVORY
JOIST
LEG
JOIST
IS
FIRMLY
FIXED
VICE
SLABS
IVORY
LEATHER
STRAPS
PADS
SCREWS
TOOLS
SORTS
ABOUT
BENCH
FORWARD
RED
FLAME
FORGE
IS
SEEN
WHERE
BLACKSMITH
IS
WORK
Drat
drat
SNEEZES
SNEEZES
SNEEZES
SNEEZES
Saw
amputate
SNEEZES
Smut
Lucky
SNEEZES
shinbone

finish
SNEEZES
shop
soak
doctored
SNEEZES
lotions
Mogulship
DURING
ENSUING
SCENE
CARPENTER
CONTINUES
SNEEZING
TIMES
manmaker
pleases
Measured
grip
pinch
grip
blacksmith
forging
partnership
Um
blacksmith
Africans
Carpenter
forge
pedlar
crushing
pattern
Imprimis
modelled
Tunnel
outwards
architecture
thief
catcher
Thrusted
Carpenter
Clay
mud
ditchers
sneezing
Bone
callest
Truly
humbly
feelest
riddle
humbly
poser
uninterpenetratingly
standest
eavesdroppers
dissolved
mayst
Hah
Truly
Bungle
TURNS
debtor
blockhead
Cursed
inter
indebtedness
ledgers
wealthiest
Praetorians

auction
owe
crucible
dissolve
compendious
CARPENTER
RESUMING
WORK
dinning
thoughted
haphazard
undertake
heron
chucks
heron
slim
lasts
roly
poly
spavined
Smut
screws
finish
collecting
core
oval
slate
smoothed
109
unfavourable
semiweekly
drench
damply
leakage
Formosa
Bashee
Isles
Niphon
Matsmai
Sikoke
braced
pruning
wrinkling
footstep
Begone
Burtons
Burtons
nearing
Either
saving
Begone
aleak
Burtons
outyell
Typhoons
Owners
prating
miserly
manifestation
resent
Devils
critically
forbearing
musket
mastering
outraged
waxes
obeys
musket
plaits

Furl
honesty
imperiously
disaffection
Burtons
110
moles
lowermost
puncheons
copies
posted
placards
Tierce
shooks
bundles
treading
catacombs
demijohn
dinnerless
Typhoons
companion
sinecures
clumsiest
stowage
holders
disembowelled
dampness
lizard
chill
lapsed
thinned
fuller
fuller
softness
sickness
fainter
bystanders
Zoroaster
impresses
revelation
holier
embalming
horizons
uncontinented
interflow
adown
Lackaday
promptitude
chalking
ejaculated
transferringly
extremities
Overhearing
tyrannical
Leaning
attentive
woody
entreated
comforts
Rarmai
sobblings
rover
Antilles
Seek
Antilles
tongues
probed
tongues
sweetness
vouchers

homes
Antilles
runaway
shame
cowards
shame
Shame
shame
thereupon
convalescence
sickness
ungovernable
destroyer
convalescing
whimsiness
emptying
grotesque
riddle
unfold
moulder
inscribed
unsolved
surveying
tantalization
111
Bashee
thanks
supplication
eastwards
stirrings
Ephesian
sod
Evangelist
Potters
Fields
continents
ebb
somnambulisms
slumberers
restlessness
Magian
rover
adoption
midmost
moles
Californian
recentest
lave
older
Archipelagoes
Japans
Lifted
Pan
Pan
nostril
sugary
Bashee
consciously
Launched
Delta
112
Blacksmith
Availing
Perth
blistered
blacksmith
forge
contributory
bowsmen
repairing

jealously
murmur
Silent
chronically
yawing
gait
importunity
questionings
Belated
innocently
blacksmith
stupidly
numbness
extremities
revelation
uncatastrophied
drama
postponedly
technicals
artisan
blithe
disguisement
blacksmith
Conjuror
economic
blacksmith
shop
basement
listened
unhappy
reverberations
unsweetly
Labor
lullaby
blacksmith
Hadst
blacksmith
competency
virtuous
elder
responsibilities
basement
fainter
glitteringly
bellows
forge
cinders
houseless
familyless
crape
unreverenced
curls
Untried
Remote
Watery
Unshored
compunctions
suicide
contributed
receptive
alluringly
mermaids
guilt
abhorring
oblivious
THY
gravestone
Hearkening
blacksmith
Perth

113
Forge
swathed
bristling
apron
Perth
forge
forge
forge
Perth
hammering
Mother
Carey
chickens
Perth
liv
scorch
Perth
scorch
shrunk
woeful
blacksmith
wert
Welding
seams
blacksmith
smoothe
seams
blacksmith
seams
passionately
Perth
smoothe
blacksmith
blacksmith
smoothe
Said
seams
blacksmith
unsmoothable
gaffs
Perth
blacksmith
stubbs
racing
Horse
stubbs
stubbornest
stubbs
weld
glue
forge
forge
shank
flaw
Work
Perth
Perth
welding
weld
hems
Perth
forge
blessing
lucifers
fusee
musket
shank
Perth
temper

Perth
wincing
pain
forging
branding
razors
blacksmith
razors
shave
sup
Fashioned
arrowy
Perth
shank
blacksmith
tempering
temper
punctures
Ego
baptizo
te
nomine
patris
sed
nomine
diaboli
scorchingly
baptismal
hickory
unwound
Pressing
harp
strandings
unstranded
intertwistings
moodily
hickory
hollowly
mummeries
unmeaningly
mocked
114
Gilder
Penetrating
paddling
birch
mixing
cats
purr
forgets
conceals
fang
rover
filial
vales
mixes
interpenetrate
seamless
treasuries
tarnishing
landscapes
parched
earthy
clover
dew
mingling
unretracing
gradations
boyhood
thoughtless

adolescence
disbelief
Ifs
unmoor
ether
weariest
foundling
unwedded
paternity
Loveliness
lover
tiered
oust
oust
115
Bachelor
Bachelor
apparel
gloriously
streamers
bunting
Signals
basketed
brazen
Bachelor
bartered
kindling
centrepiece
caulked
plugged
complacent
barbarian
drums
POKE
hued
eloped
Isles
negroes
Bastille
drama
jubilations
forebodings
impersonated
Bachelor
islanders
aloud
sayst
Forward
Set
receding
Bachelor
Bachelor
homewardbound
116
Dying
favourites
Bachelor
spearings
stilly
sweetness
plaintiveness
inwreathing
orisons
convent
wantonly
vesper
hymns
Soothed
sterned
wanings

sunwards
placid
wondrousness
vassal
favouring
favouring
candid
unspoken
Niger
sunwards
whirls
unverdured
speakest
slaughtering
Typhoon
sunwards
trebly
hooped
aspiring
rainbowed
strivest
jettest
intercedings
unnamable
imminings
suckled
mothered
foster
117
flickering
chafed
spectrally
moaning
squadrons
Asphaltites
unforgiven
Started
hooped
hearses
couldst
hearses
plumes
pall
Believe
befall
pledges
survive
derision
;--"
Immortal
118
Quadrant
determine
unblinkingly
focus
immeasurable
splendors
quadrant
astrological
passionlessness
revery
AM
SHALL
quadrant
cabalistical
Foolish
toy
plaything
Admirals
Commodores

brags
happenest
impotence
insultest
Science
Curse
toy
vividness
scorches
Level
Curse
quadrant
guide
deadreckoning
THESE
trample
feebly
pointest
trampled
sneering
fatalistic
Unobserved
awestruck
clustered
troubledly
erectly
Horatii
piouetting
lurching
dumbest
charcoal
actest
119
Candles
Warmest
cruellest
crouches
Skies
Typhoon
cloudless
exploding
bomb
dazed
poled
Typhoon
blazed
disabled
fluttering
firmer
teetering
sieve
Bad
;"--(
flourishin
sporty
gamy
jesty
joky
hoky
poky
flyin
foamin
';
spicin
'--
sporty
gamy
jesty
joky
hoky

poky
splits
smacks
tastin
sporty
gamy
jesty
joky
hoky
poky
Typhoon
harp
Madman
markest
shortest
lightens
volley
peals
conductor
mishaps
interfering
impeding
flambeaux
weaker
Himmalehs
corpusants
corpusants
tipped
tri
sulphurous
corpusants
teeter
Mene
Mene
Tekel
Upharsin
pallidness
gleaming
phosphorescence
Relieved
tipped
corpusants
Satanic
tableau
pallidness
pall
thinkest
corpusants
chock
Glancing
supernaturalness
corpusants
cohered
numbed
wasps
Herculaneum
upcast
tri
trinity
sacramental
defiance
placeless
dispute
unintegral
mastery
impersonal
personality
whencesoe
wheresoe
queenly

personality
pain
supernal
lauchest
maiest
SUDDEN
REPEATED
FLASHES
LIGHTNING
NINE
FLAMES
LENGTHWISE
THRICE
PREVIOUS
HEIGHT
REST
CLOSES
EYES
PRESSED
HARD
UPON
placeless
wrung
shutter
stunning
blindfold
Light
leapest
genealogy
callest
unbegotten
callest
unbegun
unsuffusing
creativeness
foundling
riddle
unparticipated
Leap
Perth
Overhearing
torch
transfix
Petrified
shrinking
binding
sweeps
neighborhood
elm
120
End
HELM
APPROACHING
inboard
coasting
gluepots
Loftiest
cowards
hooroosh
colic
malady
121
Bulwarks
FLASK
MOUNTED
PASSING
ADDITIONAL
LASHINGS
OVER
ANCHORS

HANGING
lucifers
ARE
lucifers
lucifers
afire
drenching
afire
Shake
pitchers
risks
Marine
guarantees
hydrants
militia
skewered
trailing
sash
Tying
fists
wring
togs
Same
cocked
eave
troughs
nasty
122
Lightning
MAIN
TOP
SAIL
YARD
PASSING
LASHINGS
AROUND
Um
um
um
Plenty
Um
um
um
Um
um
um
123
Musket
Typhoon
reelingly
preventer
shuttlecock
anyone
Typhoon
strenuous
exertions
trysail
HO
FAIR
OH
HO
CHEERLY
MEN
falsified
compliance
trimmed
reluctantly
apprise
fitfully
fitful

panels
subterraneousness
hooped
shiningly
evolved
musket
Loaded
spill
musket
quadrant
gropes
Typhoon
suffered
Sleeping
remonstrance
scornest
Flat
breathest
Ahabs
forbid
lawful
caged
howlings
tindering
stealthily
musket
survive
musket
drunkard
panel
wrestling
124
Needle
outspread
unstaggering
abounded
outbellying
boomed
Muffled
Emblazonings
Babylonian
crucible
molten
bubblingly
tetering
loweringly
rearward
Yoke
hallo
reined
huskily
liest
Heading
palpableness
Thrusting
electricity
Instances
loadstone
recovers
reaches
lowermost
Deliberately
transpointed
unmurmuringly
acquiesced
rumbled
unimpressed
magnetism
tubes
quadrant

loadstone
Accessory
dictating
transpointed
clumsily
shudderings
Abashed
servile
exchanged
blunted
magnetizing
loadstone
slunk
125
Log
fated
reliance
determining
neglect
slate
railing
Rains
damped
warped
quadrant
plungingly
Forward
Heave
hued
Tahitian
grizzly
spindle
spool
unwinding
mistrust
truer
spool
subservient
wert
Excellent
unmanned
jerkingly
reelman
Snap
overstrained
sagged
tugging
quadrant
Tahitian
loosening
Tahitian
fished
Jerk
Jerk
cowards
cowards
loon
scolds
lesser
holiness
Astern
sieve
ding
dong
ding
cowardly
Ding
dong
ding
creative

libertines
touchest
strings
weak
Perth
rivet
believers
omniscient
oblivious
idiotic
gratitude
daft
daft
daft
rotten
Mend
126
Buoy
Making
unfrequented
monotonously
plaintively
wailings
Herod
Innocents
transfixedly
mermaids
unappalled
hollowly
dams
dams
sobbing
peeringly
bodings
plausible
confirmation
waked
shrunk
parched
pore
sooth
portent
foreshadowing
fulfilment
presaged
lightness
feverish
inuendoes
arrange
Bring
caulk
seams
caulking
huff
baulks
bandbox
cobbling
undignified
tinkers
brats
tinkerings
betters
jobs
cobbler
cobbling
jobs
tinkers
shop
heigh
Nail

caulk
seams
batten
knotty
Aroostook
hemlock
Cruppered
Sailing
tray
workers
bedsteads
coffins
hearses
cobbling
caulking
127
LAID
UPON
LINE
TUBS
BETWEEN
VICE
BENCH
OPEN
HATCHWAY
CARPENTER
CAULKING
ITS
SEAMS
STRING
TWISTED
OAKUM
SLOWLY
UNWINDING
ROLL
PLACED
BOSOM
FROCK
COMES
SLOWLY
GANGWAY
HEARS
FOLLOWING
complies
genially
aisle
Art
shop
ferrule
undertaker
arrant
intermeddling
scamp
coffins
coffins
Titans
chipping
craters
digger
sings
digger
caulking
Carpenter
bier
Um
um
Art
Despatch
Albemarle
musical

HIMSELF
woodpecker
Blind
envied
Rat
tat
seconds
tick
immaterial
imponderable
Carpenter
suck
philosophies
conduits
128
Rachel
Rachel
clustering
boastful
bladders
Bad
Throttling
negatively
pulls
exchanged
swiftest
gleam
kindling
conjure
icily
petition
charter
roundly
SHALL
Manx
Rachel
cruellest
adopting
iciness
unmisgiving
hardihood
paternal
initiate
unenervated
display
partiality
undue
boon
nestling
relent
prolongingly
Gardiner
forgive
Hurriedly
transfixed
rejection
enchantment
Gardiner
tack
clustered
cherry
cherrying
woeful
wept
Rachel
129
MOVING
CATCHES
Lad
malady

cures
malady
wert
bigot
fadeless
cures
grows
speakest
Weep
Met
befall
GOES
STEPS
STEP
FORWARD
Ding
dong
ding
fours
lieutenants
epaulets
epaulets
epaulets
monsieurs
Monsieurs
cowardly
Jumped
shame
cowards
Shame
Put
Shame
cowards
bulge
130
Hat
contrastingly
concurrent
sinning
sinned
unsettling
sustains
domineered
bodings
foreshadowing
check
Alike
powdered
clamped
machines
despot
awed
awed
shuddering
wan
watchmen
limits
slouching
dew
sunshine
gnarled
unearthed
intermission
unmomentous
asunder
verbal
forethrown
commandingly
subordinates
tyrant

siding
glimmering
slided
Rachel
basketed
reeved
attaching
shunning
relying
arranging
foothold
relations
discerned
fastenings
watchman
carelessness
swooping
ventured
select
watchman
billed
incommodiously
untrackably
spiralized
Sicilian
posted
scream
hawk
Tarquin
replace
thereupon
Tanaquil
Tarquin
hawk
disappearance
discerned
131
misnamed
shears
unrigged
disabled
shears
bleaching
sewing
Perth
Tempered
temper
triply
Brace
baptism
dejected
132
Symphony
firmaments
separable
pensive
transparently
Samson
Hither
unspeckled
masculine
sex
czar
groom
gnarled
haggardly
unyielding
untottering
clearness
innocency

Invisible
oblivious
elves
heedlessly
gambol
singed
marge
aromas
dispel
cankorous
caress
erring
wee
sobbing
sweetness
privation
forsaken
exclusiveness
heaviness
slavery
poorest
crusts
frenzy
foamingly
brush
mockery
mockery
mockery
Wife
sisterly
paternal
blighted
cindered
lovings
longings
Albicore
fang
Sleeping
Sleep
blanched
Mate
motionlessly
133
Chase
inspecting
vane
ascertaining
shortened
acute
dictating
vindicated
bordering
metallic
rip
Thundering
flattening
stunsails
Fired
jetting
moonlit
methodic
attuned
gradual
prolongings
stunsails
Helm
luff
Shiver
Boats
gnawed

nautilus
unsuspecting
rugged
musical
playfully
interchangeably
feathering
alternate
fitful
skimming
streaming
pennons
joyousness
ravished
Europa
bewitching
nuptial
Crete
Supreme
surpass
coincident
enticings
namelessly
ventured
assail
enticing
glidest
bejuggled
tranquillities
clappings
rapture
withholding
marbleized
warningly
bannered
Hoveringly
apeak
stilly
wooing
freshened
herons
fluttering
expectant
keener
weasel
undiscoverable
scrolled
doored
Perth
stratagem
sidelingly
thrilled
scrolled
unastonished
dallied
frenzied
wrench
frail
shears
crafty
upraising
loosed
tilting
Ripplingly
spindled
vindictively
Channel
overleap
designation
poise

comprehensively
madden
mulberries
Antiochus
Maccabees
insolent
incuriously
revolvingly
planetarily
swooping
!--"
crest
Dragged
caking
trodden
wails
ravines
prostration
abbreviate
pang
feebler
decree
intensities
pointless
centres
circumferences
draggingly
Help
Set
banked
treble
banked
vicissitude
overtaking
stacking
outstretching
jointed
methodic
allotted
.--"
Whose
clouded
intending
thistle
pricked
soulless
wert
Groan
Omen
honourably
darkling
Begone
peopled
Cold
rustling
unset
slower
Helm
."--
slouching
rousing
134
Chase
punctually
pertinacious
prescience
invincible
geniuses
foretell
trending

cape
obscures
proverbial
evanescence
railway
acuteness
windbound
Inferable
furrow
missent
plough
creeps
tingles
launch
gait
blister
miller
shuts
watergate
frenzies
bubblingly
forebodings
hares
bounding
suspense
enslaved
maple
individualities
guilt
guiltiness
oneness
outspreadingly
Clinging
wavings
shading
blueness
Sway
belayed
orchestra
vibrate
discharges
rifles
halloo
breaching
breaching
defiance
immeasurable
bravadoes
salmon
bluer
glittered
glacier
fading
fading
mistiness
Unmindful
ladders
dartingly
excludes
annihilating
manoeuvred
trained
chargers
slogan
untraceable
evolutions
recrossed
foreshortened
warped
hoping

disencumber
snarls
Caught
corkscrewed
bristling
critically
inboard
intercepted
fagot
tangles
husks
grated
nutmeg
aslope
bobbied
twitching
lustily
unstricken
struggled
modifying
crumb
trailing
intertangled
methodic
attentive
contusions
inextricable
splinter
leaner
ferrule
Dead
muster
splinter
gores
Accursed
wrench
tangles
Gone
knell
blistered
!--'
collect
unmeasured
Jesus

```
-----  
KeyboardInterrupt                                Traceback (most recent call last)  
<ipython-input-14-45f40a9cc86c> in <module>  
      5         print(w)  
      6  
----> 7 word_20(text1)  
  
<ipython-input-14-45f40a9cc86c> in word_20(text1)  
      2     cut = int(.8 * len(text1))  
      3     for w in text1[cut:]:  
----> 4         if w not in text1[:cut]:  
      5         print(w)  
      6
```

KeyboardInterrupt:

5. Write a program that takes the sentence ("we have seen two kinds of two sequence objects") expressed as a single string, splits it and counts up the tokens. Get it to print out each token and the token's frequency, one per line, in alphabetical order. You should write a function and call that function to process the sentence.

```
In [16]: sent_5 = "we have seen two kinds of two sequence objects"
```

```
In [17]: def token_counter(raw):
```

```

text = word_tokenize(raw)
fdist = nltk.FreqDist(text)
for key in sorted(fdist):
    print((key) + ': ', fdist[key])

```

```
token_counter(sent_5)
```

```

have: 1
kinds: 1
objects: 1
of: 1
seen: 1
sequence: 1
two: 2
we: 1

```

6. Write a function shorten(word, n) to process a text ("big big big world today tomorrow good Today good"), omitting the n most frequently occurring words of the text. You should use w.lower() to normalize the text first. Please call this shorten function.

```
In [18]: sent_6 = 'big big big world today tomorrow good Today good'
```

```
In [19]: def shorten(sent, n):
          text = sent.split()
          most_freq = set(w for (w,f) in nltk.FreqDist(text).most_common(n))
          return sorted([w.lower() for w in text if w not in most_freq])

          shorten(sent_6, 2)
```

```
Out[19]: ['today', 'today', 'tomorrow', 'world']
```

7. Write a list comprehension that sorts a list of WordNet synsets for proximity to a given synset. For example, given the synsets lesser_rorqual.n.01, killer_whale.n.01, novel.n.01, and tortoise.n.01, sort them according to their shortest_path_distance() from right_whale.n.01. You should use lambda in the sorted() function.

```
In [21]: synset_samples = 'lesser_rorqual.n.01, killer_whale.n.01, novel.n.01, tortoise.n.01'
```

```
In [22]: sorted([wn.synset(w) for w in synset_samples.split(',')],
               key = lambda x: wn.synset('right_whale.n.01').shortest_path_distance(x))
```

```
Out[22]: [Synset('lesser_rorqual.n.01'),
          Synset('killer_whale.n.01'),
          Synset('tortoise.n.01'),
          Synset('novel.n.01')]
```

8. Write a function that takes a list of words (containing duplicates) (i.e. words= ['table', 'chair', 'desk', 'table', 'table', 'chair']) and returns a list of words (with no duplicates) sorted by decreasing frequency. E.g. if the input list contained 10 instances of the word table and 9 instances of the word chair, then table would appear before chair in the output list. You should use lambda in the sorted() function.

```
In [23]: words_8= ['table', 'chair', 'desk', 'table', 'table', 'chair']
```

```
In [24]: def list_sorter(list_8):
          fq = FreqDist(list_8)
          freq = dict(sorted(fq.items(), key = lambda x: x[1], reverse = True))
          return list(freq.keys())
```

```
In [25]: list_sorter(words_8)
```

```
Out[25]: ['table', 'chair', 'desk']
```

9. Write a function that takes a text (e.g. text3 from nltk.book) and a vocabulary (e.g. nltk.corpus.words.words()) as its arguments and returns the set of words that appear in the text but not in the vocabulary. Both arguments can be represented as lists of strings. Can you do this in a single line, using set.difference()?

```
In [26]: text_9 = nltk.book.text3
        voc_9 = nltk.corpus.words.words()
```

```
In [27]: def text_diff_finder(text,voc):
        return set(text).difference(voc)

        text_diff_finder(text_9,voc_9)
```

```
Out[27]: {'awaked',
'pieces',
'Shinar',
'Spake',
'Shalem',
'Admah',
'Midianites',
'borders',
'Riphath',
'Moab',
'gro',
'handfuls',
'Because',
'Isui',
'Behold',
'repented',
'breasts',
'beasts',
'Zar',
'commanded',
'Calneh',
'Abidah',
'Ezer',
'Manasseh',
'spe',
'Terah',
'destroyed',
'catt',
'fle',
'From',
'shoelatchet',
'named',
'Zibeon',
'Siddim',
'marvelled',
'dreamed',
'?',
'counted',
'barr',
'overtook',
'My',
'lambs',
'messengers',
'Drink',
'Sitnah',
'Gaham',
'faileth',
'That',
'Arise',
'Aram',
'communing',
'burdens',
'verified',
'Tebah',
'cities',
```

'Haran',
'kissed',
'stories',
'Abrah',
'Moreh',
'sepulchre',
'fowls',
'Fifteen',
'yielded',
'tents',
'healed',
'Ephrath',
'refused',
'Zoar',
'eatest',
'hang',
'Asenath',
'Why',
'shoulders',
'lighted',
'Oh',
'Even',
'rested',
'Jimnah',
'journeyed',
'strakes',
'changes',
'Heber',
'circumcised',
'Hinder',
'Assyria',
'lords',
'Pinon',
'Surely',
'Horites',
'chode',
'journeys',
'fatfleshed',
'aileth',
'Sered',
'hunter',
'hairs',
'firmame',
'firstborn',
'Horite',
'Karnaim',
'Sod',
'Nineveh',
'Sarai',
'Angel',
'bulls',
'Er',
'comest',
'serva',
'thi',
'anoointedst',
'seekest',
'grisl',
'Shepho',
'Aran',
'subtil',
'sheweth',
'Gerar',
'Send',
'Amorites',
'sinners',
'bakemeats',
'thoughts',
'Shillem',
'tru',

'cometh',
'Cursed',
'magnified',
'Isra',
'rewarded',
'En',
'interpretations',
'Epher',
'maidservants',
'fearest',
'Potiphar',
'Zerah',
'Jahleel',
'sceptre',
'repenteth',
'Abimelech',
'troubled',
'prisoners',
'firstlings',
'saidst',
'Anah',
'Gatam',
'Methusael',
'Happy',
'decreased',
'spee',
'comforted',
'offered',
'names',
'Mahalaleel',
'Nod',
'Hushim',
'Unstable',
'Ehi',
'handmaidens',
'rulers',
'Almighty',
'pluckt',
'Hazarmaveth',
'EleloheIsrael',
'seemed',
'kings',
'Mahalath',
'Tubal',
'Peniel',
'halted',
'They',
'hadst',
'garments',
'Swear',
'morter',
'mandrakes',
'concubi',
'stronger',
'How',
'generations',
'suffered',
'tillest',
'Is',
'played',
'),'',
'Mehetabel',
'nati',
'Gilead',
'Elbethel',
'Eldaah',
'fath',
'daughters',
'Do',
'Our',

'officers',
'blessings',
'When',
'Avith',
'Ephron',
'These',
'began',
'became',
'Jahzeel',
'hanged',
'boys',
'Am',
'Zarah',
'moveth',
'grisled',
'Perizzite',
'Jokshan',
'interpreted',
'talked',
'Zepho',
'priests',
'Joktan',
'Jaalam',
'Jachin',
'ewes',
'Sabtech',
'Sichem',
'Therefore',
'Shobal',
'hor',
'Philistim',
'lovest',
'Methusa',
'having',
'honour',
'flocks',
'Fulfil',
'walked',
'perceived',
'Ebal',
)',
'Wilt',
'travailed',
'Naphish',
'husba',
'rods',
'kindled',
'dreams',
''',
'carcases',
'erected',
'Can',
'In',
'.)',
'comi',
'Fear',
'ceased',
'Which',
'Edom',
'Until',
'Cast',
'Are',
'Arbah',
'bou',
'waited',
'Both',
'forgot',
'prevailed',
'Jac',
'pitched',

'Reumah',
'Eri',
'Amalek',
'planted',
'Sidon',
'Beerlahairoi',
'bodies',
'pressed',
'rained',
'defiledst',
'giveth',
'measures',
'camest',
'served',
'bak',
'leanfleshed',
'Jetur',
'Thy',
'She',
'avenged',
'Zeboiim',
'progenitors',
'obeyed',
'Madai',
'bakers',
'Mesopotamia',
'princes',
'Zillah',
'Serah',
'crieth',
'Earth',
'asketh',
'Lotan',
'Also',
'breaketh',
'Some',
'Erech',
'whosoever',
'statutes',
'whales',
'lentiles',
'Perizzit',
'sacks',
'Kirjatharba',
'hou',
'twel',
'Assyr',
'Diklah',
'Hanoah',
'Calah',
'dwe',
'walketh',
'Jabbok',
'booths',
'pursued',
'oth',
'images',
'Escape',
'youngest',
'Abr',
'Shaveh',
'lands',
'Eno',
'bundles',
'Blessed',
'discerned',
'slayeth',
'shrubs',
'increased',
'Ur',

'colts',
'Mizz',
'Lamech',
'aprons',
'numbered',
'Cush',
'handmaids',
'Timnath',
'While',
'Bilhan',
'heav',
'Mesha',
'reproa',
'Pildash',
'Tarshish',
'Gaza',
'hil',
'Enmishpat',
'rul',
'guiding',
'Egyptia',
'bones',
'joined',
'windows',
'sto',
'wentest',
'circumcis',
'conceived',
'Husham',
'Havilah',
'Zebul',
'Zaavan',
'appe',
'needeth',
'Jubal',
'badest',
'Jezer',
'Be',
'Rebek',
'Hadoram',
'favoured',
'presented',
'Abelmizraim',
'At',
'clothed',
'Se',
'sevens',
'hills',
'Elah',
'.',
'devoured',
'lives',
'Gentiles',
'damsels',
'reigned',
'wiv',
'Sheba',
'mouths',
'wrestlings',
'Two',
'judged',
'clusters',
'Yet',
'Masrekah',
'goest',
'hastened',
'creepeth',
'ki',
'wells',
'Ham',

'deeds',
'Nebajoth',
'Every',
'droves',
'Arphaxad',
'Day',
'Ophir',
'Muppm',
'hast',
'trees',
'Now',
'daughte',
'Ananim',
'Rephaims',
'Wherefore',
'faults',
'Ezbon',
'Give',
'Dishon',
'places',
'rebuked',
'boug',
'gard',
'Arodi',
'sinning',
'coats',
'arrayed',
'years',
'Penuel',
'hundredfo',
'hid',
'Japhe',
'storehouses',
'Mezahab',
'Lie',
'Of',
'Jobab',
'shew',
'asswaged',
'skins',
'selfwill',
'twins',
'Eshcol',
'Birsha',
'believed',
'hith',
'slimepits',
'nostrils',
'Neither',
'dunge',
'biteth',
'gr',
'placed',
'liest',
'Say',
'Casluhim',
'Uz',
'Milcah',
'Magdiel',
'Who',
'Egy',
'Kenizzites',
'Togarmah',
'Zemarite',
'visions',
'fo',
'knees',
'appointed',
'Malchiel',
'Jehovahjireh',

('',
'Jerah',
'Bilhah',
'sawest',
'captives',
'spake',
'bands',
'songs',
'tarried',
'loved',
'Put',
'Accad',
'wrestled',
'instruments',
'lights',
'asses',
'Moriah',
'laws',
'kn',
'committed',
'Naphtuhim',
'offerings',
'Emins',
'merchantmen',
'This',
'delivered',
'Padanaram',
'pulled',
'seeth',
'sheepshearers',
'things',
'shouldest',
'Lift',
'Yea',
'?')',
'Whereas',
'Reu',
'heels',
'Elam',
'Ye',
'Mehujael',
'Leummim',
'Es',
'Mibzar',
'Isa',
'Hirah',
'Shuah',
'spices',
'wagons',
'commandments',
'Except',
'Twelve',
'Bera',
'Philistines',
'Shaul',
'So',
'riv',
'Abimael',
'Go',
'kine',
'friends',
'words',
'Zidon',
'failed',
'shekels',
'Massa',
'troughs',
'Only',
'Timnah',
'Machpelah',

'Jegarsahadutha',
'Aner',
'With',
'bondmen',
'Ammon',
'cru',
'mercies',
'befell',
'plains',
'speckl',
'Tiras',
'Lud',
'butlers',
'Hiddekel',
'hands',
'Keturah',
'Temani',
'Sheleph',
'doth',
'dainties',
'Mishma',
'Tola',
'Gether',
'Pau',
'Shammah',
'oa',
'attained',
'chariots',
';)',
'Hori',
'womenservants',
'voi',
'After',
'traffick',
'Peleg',
'Gather',
'faces',
'Chaldees',
'shewed',
'fainted',
'vestures',
'nourished',
'All',
'saith',
'inhabitants',
'ruled',
'Up',
'mayest',
'Esau',
'Zohar',
'herds',
'Shimron',
'Hadar',
'deprived',
'Carmi',
'shepherds',
'gifts',
'His',
'isles',
'Lahairoi',
'meanest',
'Huppim',
'Lest',
'feebler',
'Ephra',
'Fill',
'Kittim',
'Jemuel',
'waters',
'towns',

'prayed',
'espied',
'Belah',
'nations',
'plagues',
'Come',
'ste',
'!',
'Zebulun',
'Get',
'darkne',
'findeth',
'But',
'pillows',
'Salem',
'Haste',
'concubines',
'Mamre',
'Discern',
'Mash',
'See',
'generatio',
'denied',
'Beor',
'westwa',
'thousands',
'elders',
'proceedeth',
'compassed',
'Perizzites',
'ou',
'If',
'dukes',
'Hebrews',
'ribs',
'oversig',
'Medan',
'mocked',
'ears',
'Galeed',
'reached',
'possessions',
'Bless',
'Here',
'Binding',
'fai',
'Hemam',
'Bela',
'Eliphaz',
'Iscah',
'Jeush',
'Tubalcain',
'sacrifices',
'Shuni',
'Upon',
'Amal',
'Merari',
'LO',
'pleaseth',
'sle',
'Euphrat',
'fountains',
'famished',
'Bered',
'tr',
'struggled',
'Peace',
'families',
'birds',
'Cherubims',

'remembered',
'Manahath',
'observed',
'rams',
'drinketh',
'Potipherah',
'Canaanites',
'gathered',
'Thorns',
'overthrew',
'branches',
'physicians',
'imagined',
'communed',
'lingered',
'grap',
'Edar',
'purchased',
'Unto',
'Almodad',
'Elparan',
'Whose',
'consumed',
'Abide',
'sojourned',
'Shalt',
'Bring',
'It',
'cubits',
'stretched',
'Lasha',
'Chesed',
'Phichol',
'nuts',
'Take',
'By',
'Guni',
';',
'Have',
'ringstraked',
'LORD',
'wombs',
'ones',
'We',
'Hitti',
'wor',
'seas',
'refrained',
'countries',
'sheddeth',
'Succoth',
'Beersheba',
'Hazezontamar',
'Gera',
'Beno',
'Beriah',
'longedst',
'Ashteroth',
'baskets',
'An',
'Said',
'Kedemah',
'Forgive',
'stooped',
'lamentati',
'Feed',
'dost',
'womenservan',
'Kadmonites',
'heard',

'goeth',
'lieth',
'likene',
'Nahath',
'Atad',
'stars',
'Lehabim',
'caused',
'changed',
'Ohad',
'asked',
'Eshban',
'garmen',
'separated',
'habitations',
'wives',
'trembled',
'Kenites',
'Mahanaïm',
'Where',
'Hereby',
'Forasmuch',
'ri',
'putting',
'weaned',
'emptied',
'fleddest',
'fathers',
'bre',
'beguiled',
'badne',
'mountains',
'charged',
'To',
'Pass',
'Sojourn',
'Ludim',
'Eber',
'honourable',
'spies',
'Jabal',
'Bozrah',
'There',
'mightier',
'escaped',
'seest',
'loins',
'tithes',
'Esek',
'entreated',
'doeth',
'Padan',
'poured',
'souls',
'Hul',
'Deliver',
'Adbeel',
'Rosh',
'Shinab',
'Raamah',
'Phallu',
'Serug',
'Tell',
'mocking',
'reproved',
'blessi',
'signs',
'servants',
'commended',
'looked',

'seasons',
'gutters',
'Kenaz',
'Hittites',
'conspired',
'Heaven',
'Cause',
'Dothan',
'builded',
'flo',
'rebelled',
'kids',
'Jetheth',
'Zeboim',
'Ephah',
'Kohath',
'tongues',
'pris',
'pleased',
'Egyptians',
'waxed',
'loveth',
'fetcht',
'reviv',
'abated',
'angels',
'Ishbak',
'eyes',
'mules',
'deceived',
'Naamah',
'Dumah',
'magicians',
':',
'Luz',
'Hamor',
'children',
'redeemed',
'created',
'called',
'laughed',
'Zimran',
'Then',
'Set',
'Zuzims',
'fulfilled',
'He',
'males',
'chesnut',
'anoth',
'horses',
'embalmed',
'Thirty',
'Make',
'Ashbel',
'sepulchres',
'hous',
'Moreover',
'fishes',
'Machir',
'menservants',
'Shur',
'months',
'Salah',
'kinds',
'Whoso',
'Matred',
'laded',
'lads',
'sheaves',

'Hebron',
'savoury',
'Ajah',
'Phuvah',
'peop',
'Ishuah',
'Resen',
'herdmen',
'Bedad',
'Heth',
'messes',
'Irad',
'Tamar',
'Me',
'Samlah',
'cakes',
'Iram',
'Eliezer',
'foals',
'Should',
'findest',
'horsemen',
'co',
'gavest',
'Chezib',
'begettest',
'despised',
'thistles',
'Thus',
'Maachah',
'Midian',
'answered',
'wandered',
'Achbor',
'Tidal',
'Thahash',
'intreated',
'Whence',
'archers',
'urged',
'Seeing',
'women',
'longeth',
'Gihon',
'Remain',
'younge',
'fathe',
'Hadad',
'Hear',
'Know',
'remained',
'Dodanim',
'Huz',
'knoweth',
'Bethel',
'threshingfloor',
'spi',
'castles',
'compasseth',
'beari',
'rooms',
'required',
'supplanted',
'Melchizedek',
'Bashemath',
'Thou',
'Man',
'Zilpah',
'ev',
'Becher',


```

'departing',
'Ithran',
'Uzal',
'hasted',
'Cainan',
'stones',
'appeared',
'Ishmeelites',
'enemies',
'knowest',
'Sabtah',
'Speak',
'vowedst',
'entered',
'worshipped',
'Shall',
'Kor',
'Jidlaph',
'searched',
'tak',
'Moabites',
'hated',
'curseth',
'Cheran',
'Edomites',
'Look',
'Din',
'endued',
'possessi',
'childr',
'Save',
'Phut',
'On',
'Slay',
'numbering',
...}

```

10. Choose your own webpage (in html format) and output the 20 most common words in this web page. You should use `w.lower()` to normalize the text. Please get rid of stop words, numbers, and punctuations. Please define a function and call that function.

```

In [28]: url = 'https://www.theguardian.com/music/2021/jan/01/mf-doom-a-hip-hop-genius-who-built-his-own-un
html = request.urlopen(url).read().decode('utf8')
text = BeautifulSoup(html, 'html.parser').get_text()
fd = nltk.FreqDist(word.lower() for word in nltk.word_tokenize(text) if word.isalpha() and word not
[word.lower() for (word, _) in fd.most_common(20)]

```

```

Out[28]: ['dumile',
'doom',
'like',
'album',
'lines',
'mf',
'guardian',
'rap',
'new',
'madlib',
'genius',
'back',
'this',
'voice',
'i',
'black',
'likes',
'the',
'content',
'designstv']

```

In []:

