CSCI-8450

2018 Spring

Yuchen Wang

You will find Chapter 4a posted at:

<http://www.nltk.org/book/ch04.html>

Let *TestText* be the text posted at URL[*https://www.cs.utexas.edu/~vl/notes/dijkstra.html*](https://www.cs.utexas.edu/~vl/notes/dijkstra.html)

To get text out of HTML, use a Python library called BeautifulSoup, available from <http://www.crummy.com/software/BeautifulSoup/> as prescribed in Dealing with HTML subsection in Chapter 3

Both of these variables are populated in the assignment template.

Complete the following of the exercises:

1. exercise 14.

◑ Write a function novel10(text) that prints any word that appeared in the last 10% of a text that had not been encountered earlier.

Use nltk.word\_tokenize() to tokenize the text. Assume each token returned by tokenizer to represent the word. Recall, that, for instance, “The” and “the” are the same words. In this question, use *TestText* for testing. Sort the resulting words alphabetically. Report first ten elements in your results.

Code:

**def** novel10(text):  
  
 cut = int(0.9 \* len(text))  
  
 first\_part, second\_part = text[:cut], text[cut:]  
 unique\_first\_part = set(first\_part)  
 unique\_second\_part = set(second\_part)  
  
 **return** [word **for** word **in** unique\_second\_part **if** word **not in** unique\_first\_part]  
  
  
**def** exercise14():  
 tokens = nltk.word\_tokenize(TestText)  
 lowerlist = [w.lower() **for** w **in** tokens]  
 result = novel10(lowerlist)  
 result = sorted(result)[:10]  
  
 print(**"first ten element: "**, result)

Result:

Exercise 14

first ten element: ['...', 'again', 'algorithm', 'allow', 'animation', 'apart', 'arrived', 'attention', 'banking', 'billington']

1. exercise 17.

◑ Write a function shorten(text, n) to process a text, omitting the *n* most frequently occurring words of the text. How readable is it?

a) How readable is the output for every n? (you don’t need to put entire outcome in your report. Instead, you can put few lines from each output.)

b) Use the function from part a on the following snippet text with n = 20. **Report the outcome** and describe your findings (how many words are taken off? What are those words?):

*to begin with I would like to thank the College of Natural Sciences for the most honouring Invitation to address its newest flock of Bachelors on this most festive day. I shall do my best.*

Code:

**def** shorten(text, n):  
   
 tokens = nltk.word\_tokenize(text)  
 fd = FreqDist(tokens)  
 charlist= [w **for** (w,n) **in** fd.most\_common(n)]  
 str = []  
 **for** t **in** tokens:  
 **if** t **not in** charlist:  
 str.append(t)  
 str = **' '**.join(str)  
 **return** str  
  
**def** exercise17():  
  
 print(**"Part a"**)  
 print(**"n = 20"**)  
 print (shorten(TestText, 20))  
 print(**"n = 35"**)  
 print (shorten(TestText, 35))  
 print(**"n = 50"**)  
 print (shorten(TestText, 50))  
 print(**"n = 65"**)  
 print (shorten(TestText, 65))  
  
 print(**"Part b"**)  
 print(shorten(snippet, 20))  
 print(**"length: "**,len(shorten(snippet,20)))

Result:

*Exercise 17*

*Part a*

*n = 20*

*Edsger Dijkstra Universities Edsger Dijkstra universities On December 8 1996 Edsger Dijkstra addressed graduates College Natural Sciences University Texas at Austin This text his speech Ladies Gentlemen begin with would like thank College Natural Sciences most honouring invitation address its newest flock Bachelors this most festive day shall do my best Your past your inalienable property can deny better or worse will shape rest your lives until death follows In sentimental moment may long lost innocence bliss ignorance but ignorance lost forever now Bachelors will have carry burden your college education every waking hour rest your lives nay even your sleep will be pressed have dreams full responsibility A main purpose this commencement speech therefore help live with your academic past 'll try give some advice how remember your university It only persons shaped their pasts so institutions so whole peoples am Dutch Dutch formative experience has been how they liberated themselves several centuries ago from Spanish occupation ( To give an example their national anthem dates from those days ) Since war dragged 80 years might feel Dutch were very efficient liberating themselves but they succeeded with tools time In 1574 Spanish tried conquer one our cities but they could because town was defended too well ; so they laid siege trying starve its population an effort which they partly succeeded But water our foe friend 3rd October 1574 some other Dutchmen caused flood simple device cutting few dikes Spanish soldiers surrounded city were just washed away Because lifting siege was turning point our struggle freedom William Silent wanted reward city 's population its bravery endurance gave them choice between two prizes either full year freedom from tax paying or right University citizens chose latter Within year 1575 University Leyden my Alma Mater was started former nunnery building thanks Reformation was market anyhow told this story because so neatly symbolizes tight link between concept University practice intellectual freedom In dozen years am now here have learnt appreciate UT place where link full vigour my first remark both advice request remember cherish your university place dedicated practice intellectual freedom where combination exercising your brain opening your mouth encouraged A next characteristic University which should be cherished protected its openness Let me explain referring briefly University 's place along whole spectrum techniques which one generation transmits its insights abilities next At one extreme we have techniques guilds which treat their insights abilities valuable property treasure be kept secret Their technique protecting secrecy keeping secret knowledge unformulated ; therefore apprentice has join master seven meagre years during which he can absorb craft osmosis so speak The university at other end spectrum professor's task bring relevant insights abilities into public domain explicit formulation It no accident universities we know them now started flourish after art book printing had been established There more be said about spectrum educational techniques but shall do so now mentioned remind why absence secrecy or more positively formulated openness honesty characteristics touch heart academic enterprise university hides or cheats can close its doors beg remember essential role academic openness when considering academic/industrial cooperation urge remember whenever government invents so-called reasons national security or prosperity prevention free publication results academic research Universities part nation 's security organisation they nation 's research laboratories either they nation's Universities In passing would like mention completely different reason such openness precondition academic survival Just being different doing things uneducated can understand academics hated feared vide Socrates executed 399 BC Archimedes killed 212 BC more recently Hypatia AD 415 barbariously murdered Christian mob Alexandria The original Oxford Colleges were buildings fortified order protect students against rabble if think all old hat refer recent histories DDR or People 's Republic China These days miracle whenever academic world tolerated at all personally am convinced what tolerance there would completely disappear were academic world become secretive The University with its intellectual life campus undoubtedly creation restless mind but more than its creation also its refuge The University unique campus being brilliant socially acceptable Furthermore fabric academic world so sturdy can absorb most revolutionary ideas But only refuge restless minds also reservation them It does only protect restless minds also protects rest world where they would create havoc if they were let loose To put another way fence around campus essential because separates two worlds otherwise would harm each other The fence ensures we have relatively little direct influence world `` out there '' but we would be foolish complain our freedom be radical we like based fact at least first 25 years industry world-at-large ignore our work anyhow Currently there seems be world-wide tendency try lower fence ; effort strikes me ill-directed As very minor but recent example may have heard efforts with catchword `` post-tenure review '' change contract between university its faculty We do discuss merits case but would like see symptom lowered fence one underlying feelings campus hiring firing should be possible like everywhere else thus ignoring university should be unlike anything else can think think all students should remember University very special place but graduates College Natural Sciences should do so most all because School hard sciences have found roof above their heads It puts heavy responsibility those who have gone through school others have no idea what going there hence naturally assume all science ephemeral transitory fashion-dependent rest It this respect reassuring so many legislatures other governing bodies hardly contain any true scientists Consequently strange things can happen After WW II Dutch industries felt need Dutch business school Because its purely vocational calling but more so because perceived lack intellectual content none existing Dutch universities wanted have anything do with separate business school was founded So far so good but number decades later need was felt raise institute 's status Dutch government which did know or did care what was doing agreed raise business school officially level university And Austin American-Statesman November 8 we could read how newfangled `` Nijenrode University '' used its academic status bestowed an Honorary Doctorate Bill Gates all scholars have n't figured out yet how Dutchman am going atone Fortunately encouraging things sometimes happen well we could read frontpage article International Herald Tribune November 27 which reported Oxford University had just turned down gift $ 34 million because strings attached gift was intended new business school but Oxford did want new business school Universities can be Very Special Places indeed By way if got impression have my doubts about business management scientific discipline right seems too fickle be taken seriously In 70s creed was diversification 80s gospel said concentrate your core business 90s world-wide credo seems be intellectual cleansing high-tech industries There seems be one thing independently height fence surrounds campus any graduating student can take with him into world outside healthy scepticism goes with well-kept immunity hype slogans fads fashions And important latter seems pass ever increasing frequency It quite amazing -- bit sadenning -- how gullible desperate willing expect salvation from next gadget remember how TV was promoted theory daily dose Shakespeare every living room would elevate culturally deprived unfathomed heights thus curing all ills society etc And what did we get ? Soap operas quizzes remember how overhead projector was welcomed greatest educational innovation since Socrates allowed much more detailed preparation how new `` audio-visual aids '' they were called those days would revolutionize class room would bring modern teaching each little Eskimo his igloo Well what overhead projector did teaching 'm afraid better judge than I. remember how with advent terminals interactive debugging was supposed solve all our programming problems how with advent colour screens `` algorithm animation '' was supposed do same And what did we get ? Commercial software with disclaimer explicitly states fool if rely what just bought And now we have multimedia/communication hype best bits those just arrived from far away if `` line '' `` Net '' just do n't count this world ( which virtual anyhow ... ) Apart from change vocabulary same hype same snake oil over over again can do me favour getting excited all time supposed save switching `` home banking '' Recently James H. Billington current Librarian Congress remarked instead knowledge-based democracy we may end up with an information-inundated democracy share his concern so allow me end with this simple wish May spite all distractions generated technology all succeed turning information into knowledge knowledge into understanding understanding into wisdom thank your your attention To Lifschitz 's web page*

*n = 35*

*Edsger Dijkstra Universities Edsger Dijkstra universities On December 8 1996 Edsger Dijkstra addressed graduates College Natural Sciences Texas at Austin This text his speech Ladies Gentlemen begin like thank College Natural Sciences most honouring invitation address newest flock Bachelors this most festive day shall do my best Your past inalienable property deny better or worse will shape rest lives until death follows In sentimental moment may long lost innocence bliss ignorance ignorance lost forever now Bachelors will carry burden college education every waking hour rest lives nay even sleep will pressed dreams full responsibility A main purpose this commencement speech therefore help live academic past 'll try give some advice remember university It only persons shaped their pasts institutions whole peoples am Dutch Dutch formative experience has been liberated themselves several centuries ago from Spanish occupation ( To give an example their national anthem dates from those days ) Since war dragged 80 years might feel Dutch were very efficient liberating themselves succeeded tools time In 1574 Spanish tried conquer one our cities could because town defended too well ; laid siege trying starve population an effort which partly succeeded But water our foe friend 3rd October 1574 some other Dutchmen caused flood simple device cutting few dikes Spanish soldiers surrounded city were just washed away Because lifting siege turning point our struggle freedom William Silent wanted reward city 's population bravery endurance gave them choice between two prizes either full year freedom from tax paying or right citizens chose latter Within year 1575 Leyden my Alma Mater started former nunnery building thanks Reformation market anyhow told this story because neatly symbolizes tight link between concept practice intellectual freedom In dozen years am now here learnt appreciate UT place where link full vigour my first remark both advice request remember cherish university place dedicated practice intellectual freedom where combination exercising brain opening mouth encouraged A next characteristic which should cherished protected openness Let me explain referring briefly 's place along whole spectrum techniques which one generation transmits insights abilities next At one extreme techniques guilds which treat their insights abilities valuable property treasure kept secret Their technique protecting secrecy keeping secret knowledge unformulated ; therefore apprentice has join master seven meagre years during which he absorb craft osmosis speak The university at other end spectrum professor's task bring relevant insights abilities into public domain explicit formulation It no accident universities know them now started flourish after art book printing had been established There more said about spectrum educational techniques shall do now mentioned remind why absence secrecy or more positively formulated openness honesty characteristics touch heart academic enterprise university hides or cheats close doors beg remember essential role academic openness when considering academic/industrial cooperation urge remember whenever government invents so-called reasons national security or prosperity prevention free publication results academic research Universities part nation 's security organisation nation 's research laboratories either nation's Universities In passing like mention completely different reason such openness precondition academic survival Just being different doing things uneducated understand academics hated feared vide Socrates executed 399 BC Archimedes killed 212 BC more recently Hypatia AD 415 barbariously murdered Christian mob Alexandria The original Oxford Colleges were buildings fortified order protect students against rabble if think old hat refer recent histories DDR or People 's Republic China These days miracle whenever academic world tolerated at personally am convinced what tolerance there completely disappear were academic world become secretive The intellectual life campus undoubtedly creation restless mind more than creation also refuge The unique campus being brilliant socially acceptable Furthermore fabric academic world sturdy absorb most revolutionary ideas But only refuge restless minds also reservation them It does only protect restless minds also protects rest world where create havoc if were let loose To put another way fence around campus essential because separates two worlds otherwise harm each other The fence ensures relatively little direct influence world `` out there '' foolish complain our freedom radical like based fact at least first 25 years industry world-at-large ignore our work anyhow Currently there seems world-wide tendency try lower fence ; effort strikes me ill-directed As very minor recent example may heard efforts catchword `` post-tenure review '' change contract between university faculty We do discuss merits case like see symptom lowered fence one underlying feelings campus hiring firing should possible like everywhere else thus ignoring university should unlike anything else think think students should remember very special place graduates College Natural Sciences should do most because School hard sciences found roof above their heads It puts heavy responsibility those who gone through school others no idea what going there hence naturally assume science ephemeral transitory fashion-dependent rest It this respect reassuring many legislatures other governing bodies hardly contain any true scientists Consequently strange things happen After WW II Dutch industries felt need Dutch business school Because purely vocational calling more because perceived lack intellectual content none existing Dutch universities wanted anything do separate business school founded So far good number decades later need felt raise institute 's status Dutch government which did know or did care what doing agreed raise business school officially level university And Austin American-Statesman November 8 could read newfangled `` Nijenrode '' used academic status bestowed an Honorary Doctorate Bill Gates scholars n't figured out yet Dutchman am going atone Fortunately encouraging things sometimes happen well could read frontpage article International Herald Tribune November 27 which reported Oxford had just turned down gift $ 34 million because strings attached gift intended new business school Oxford did want new business school Universities Very Special Places indeed By way if got impression my doubts about business management scientific discipline right seems too fickle taken seriously In 70s creed diversification 80s gospel said concentrate core business 90s world-wide credo seems intellectual cleansing high-tech industries There seems one thing independently height fence surrounds campus any graduating student take him into world outside healthy scepticism goes well-kept immunity hype slogans fads fashions And important latter seems pass ever increasing frequency It quite amazing -- bit sadenning -- gullible desperate willing expect salvation from next gadget remember TV promoted theory daily dose Shakespeare every living room elevate culturally deprived unfathomed heights thus curing ills society etc And what did get ? Soap operas quizzes remember overhead projector welcomed greatest educational innovation since Socrates allowed much more detailed preparation new `` audio-visual aids '' were called those days revolutionize class room bring modern teaching each little Eskimo his igloo Well what overhead projector did teaching 'm afraid better judge than I. remember advent terminals interactive debugging supposed solve our programming problems advent colour screens `` algorithm animation '' supposed do same And what did get ? Commercial software disclaimer explicitly states fool if rely what just bought And now multimedia/communication hype best bits those just arrived from far away if `` line '' `` Net '' just do n't count this world ( which virtual anyhow ... ) Apart from change vocabulary same hype same snake oil over over again do me favour getting excited time supposed save switching `` home banking '' Recently James H. Billington current Librarian Congress remarked instead knowledge-based democracy may end up an information-inundated democracy share his concern allow me end this simple wish May spite distractions generated technology succeed turning information into knowledge knowledge into understanding understanding into wisdom thank attention To Lifschitz 's web page*

*n = 50*

*Edsger Dijkstra Universities Edsger Dijkstra universities On December 8 1996 Edsger Dijkstra addressed graduates College Natural Sciences Texas at Austin This text his speech Ladies Gentlemen begin like thank College Natural Sciences most honouring invitation address newest flock Bachelors most festive day shall my best Your past inalienable property deny better worse will shape rest lives until death follows In sentimental moment may long lost innocence bliss ignorance ignorance lost forever now Bachelors will carry burden college education every waking hour rest lives nay even sleep will pressed dreams full responsibility A main purpose commencement speech therefore help live past 'll try give some advice only persons shaped their pasts institutions whole peoples am formative experience has been liberated themselves several centuries ago from Spanish occupation ( To give an example their national anthem dates from those days ) Since war dragged 80 years might feel were very efficient liberating themselves succeeded tools time In 1574 Spanish tried conquer one our cities could because town defended too well ; laid siege trying starve population an effort partly succeeded But water our foe friend 3rd October 1574 some other Dutchmen caused flood simple device cutting few dikes Spanish soldiers surrounded city were just washed away Because lifting siege turning point our struggle freedom William Silent wanted reward city population bravery endurance gave them choice between two prizes either full year freedom from tax paying right citizens chose latter Within year 1575 Leyden my Alma Mater started former nunnery building thanks Reformation market anyhow told story because neatly symbolizes tight link between concept practice intellectual freedom In dozen years am now here learnt appreciate UT place where link full vigour my first remark both advice request cherish place dedicated practice intellectual freedom where combination exercising brain opening mouth encouraged A next characteristic should cherished protected openness Let me explain referring briefly place along whole spectrum techniques one generation transmits insights abilities next At one extreme techniques guilds treat their insights abilities valuable property treasure kept secret Their technique protecting secrecy keeping secret knowledge unformulated ; therefore apprentice has join master seven meagre years during he absorb craft osmosis speak The at other end spectrum professor's task bring relevant insights abilities into public domain explicit formulation no accident universities know them now started flourish after art book printing had been established There more said about spectrum educational techniques shall now mentioned remind why absence secrecy more positively formulated openness honesty characteristics touch heart enterprise hides cheats close doors beg essential role openness when considering academic/industrial cooperation urge whenever government invents so-called reasons national security prosperity prevention free publication results research Universities part nation security organisation nation research laboratories either nation's Universities In passing like mention completely different reason such openness precondition survival Just being different doing things uneducated understand academics hated feared vide Socrates executed 399 BC Archimedes killed 212 BC more recently Hypatia AD 415 barbariously murdered Christian mob Alexandria The original Oxford Colleges were buildings fortified order protect students against rabble if think old hat refer recent histories DDR People Republic China These days miracle whenever tolerated at personally am convinced tolerance there completely disappear were become secretive The intellectual life campus undoubtedly creation restless mind more than creation also refuge The unique campus being brilliant socially acceptable Furthermore fabric sturdy absorb most revolutionary ideas But only refuge restless minds also reservation them does only protect restless minds also protects rest where create havoc if were let loose To put another way fence around campus essential because separates two worlds otherwise harm each other The fence ensures relatively little direct influence out there foolish complain our freedom radical like based fact at least first 25 years industry world-at-large ignore our work anyhow Currently there seems world-wide tendency try lower fence ; effort strikes me ill-directed As very minor recent example may heard efforts catchword post-tenure review change contract between faculty We discuss merits case like see symptom lowered fence one underlying feelings campus hiring firing should possible like everywhere else thus ignoring should unlike anything else think think students should very special place graduates College Natural Sciences should most because School hard sciences found roof above their heads puts heavy responsibility those who gone through school others no idea going there hence naturally assume science ephemeral transitory fashion-dependent rest respect reassuring many legislatures other governing bodies hardly contain any true scientists Consequently strange things happen After WW II industries felt need school Because purely vocational calling more because perceived lack intellectual content none existing universities wanted anything separate school founded So far good number decades later need felt raise institute status government did know did care doing agreed raise school officially level And Austin American-Statesman November 8 could read newfangled Nijenrode used status bestowed an Honorary Doctorate Bill Gates scholars n't figured out yet Dutchman am going atone Fortunately encouraging things sometimes happen well could read frontpage article International Herald Tribune November 27 reported Oxford had just turned down gift $ 34 million because strings attached gift intended new school Oxford did want new school Universities Very Special Places indeed By way if got impression my doubts about management scientific discipline right seems too fickle taken seriously In 70s creed diversification 80s gospel said concentrate core 90s world-wide credo seems intellectual cleansing high-tech industries There seems one thing independently height fence surrounds campus any graduating student take him into outside healthy scepticism goes well-kept immunity hype slogans fads fashions And important latter seems pass ever increasing frequency quite amazing -- bit sadenning -- gullible desperate willing expect salvation from next gadget TV promoted theory daily dose Shakespeare every living room elevate culturally deprived unfathomed heights thus curing ills society etc And did get ? Soap operas quizzes overhead projector welcomed greatest educational innovation since Socrates allowed much more detailed preparation new audio-visual aids were called those days revolutionize class room bring modern teaching each little Eskimo his igloo Well overhead projector did teaching 'm afraid better judge than I. advent terminals interactive debugging supposed solve our programming problems advent colour screens algorithm animation supposed same And did get ? Commercial software disclaimer explicitly states fool if rely just bought And now multimedia/communication hype best bits those just arrived from far away if line Net just n't count ( virtual anyhow ... ) Apart from change vocabulary same hype same snake oil over over again me favour getting excited time supposed save switching home banking Recently James H. Billington current Librarian Congress remarked instead knowledge-based democracy may end up an information-inundated democracy share his concern allow me end simple wish May spite distractions generated technology succeed turning information into knowledge knowledge into understanding understanding into wisdom thank attention To Lifschitz web page*

*n = 65*

*Edsger Dijkstra Universities Edsger Dijkstra universities On December 8 1996 Edsger Dijkstra addressed graduates College Natural Sciences Texas at Austin This text his speech Ladies Gentlemen begin thank College Natural Sciences most honouring invitation address newest flock Bachelors most festive day shall my best Your past inalienable property deny better worse will shape rest lives until death follows sentimental moment may long lost innocence bliss ignorance ignorance lost forever Bachelors will carry burden college education every waking hour rest lives nay even sleep will pressed dreams full responsibility A main purpose commencement speech therefore help live past 'll try give some advice only persons shaped their pasts institutions whole peoples am formative experience has been liberated themselves several centuries ago Spanish occupation ( To give an example their national anthem dates those days ) Since war dragged 80 years might feel very efficient liberating themselves succeeded tools time 1574 Spanish tried conquer cities could town defended too well ; laid siege trying starve population an effort partly succeeded But water foe friend 3rd October 1574 some other Dutchmen caused flood simple device cutting few dikes Spanish soldiers surrounded city washed away Because lifting siege turning point struggle William Silent wanted reward city population bravery endurance gave them choice between two prizes either full year tax paying right citizens chose latter Within year 1575 Leyden my Alma Mater started former nunnery building thanks Reformation market anyhow told story neatly symbolizes tight link between concept practice dozen years am here learnt appreciate UT place where link full vigour my first remark both advice request cherish place dedicated practice where combination exercising brain opening mouth encouraged A next characteristic cherished protected openness Let me explain referring briefly place along whole spectrum techniques generation transmits insights abilities next At extreme techniques guilds treat their insights abilities valuable property treasure kept secret Their technique protecting secrecy keeping secret knowledge unformulated ; therefore apprentice has join master seven meagre years during he absorb craft osmosis speak The at other end spectrum professor's task bring relevant insights abilities into public domain explicit formulation no accident universities know them started flourish after art book printing had been established There said about spectrum educational techniques shall mentioned remind why absence secrecy positively formulated openness honesty characteristics touch heart enterprise hides cheats close doors beg essential role openness when considering academic/industrial cooperation urge whenever government invents so-called reasons national security prosperity prevention free publication results research Universities part nation security organisation nation research laboratories either nation's Universities passing mention completely different reason such openness precondition survival Just being different doing things uneducated understand academics hated feared vide Socrates executed 399 BC Archimedes killed 212 BC recently Hypatia AD 415 barbariously murdered Christian mob Alexandria The original Oxford Colleges buildings fortified order protect students against rabble if think old hat refer recent histories DDR People Republic China These days miracle whenever tolerated at personally am convinced tolerance there completely disappear become secretive The life campus undoubtedly creation restless mind than creation also refuge The unique campus being brilliant socially acceptable Furthermore fabric sturdy absorb most revolutionary ideas But only refuge restless minds also reservation them does only protect restless minds also protects rest where create havoc if let loose To put another way fence around campus essential separates two worlds otherwise harm each other The fence ensures relatively little direct influence out there foolish complain radical based fact at least first 25 years industry world-at-large ignore work anyhow Currently there seems world-wide tendency try lower fence ; effort strikes me ill-directed As very minor recent example may heard efforts catchword post-tenure review change contract between faculty We discuss merits case see symptom lowered fence underlying feelings campus hiring firing possible everywhere else thus ignoring unlike anything else think think students very special place graduates College Natural Sciences most School hard sciences found roof above their heads puts heavy responsibility those who gone through others no idea going there hence naturally assume science ephemeral transitory fashion-dependent rest respect reassuring many legislatures other governing bodies hardly contain any true scientists Consequently strange things happen After WW II industries felt need Because purely vocational calling perceived lack content none existing universities wanted anything separate founded So far good number decades later need felt raise institute status government know care doing agreed raise officially level And Austin American-Statesman November 8 could read newfangled Nijenrode used status bestowed an Honorary Doctorate Bill Gates scholars n't figured out yet Dutchman am going atone Fortunately encouraging things sometimes happen well could read frontpage article International Herald Tribune November 27 reported Oxford had turned down gift $ 34 million strings attached gift intended new Oxford want new Universities Very Special Places indeed By way if got impression my doubts about management scientific discipline right seems too fickle taken seriously 70s creed diversification 80s gospel said concentrate core 90s world-wide credo seems cleansing high-tech industries There seems thing independently height fence surrounds campus any graduating student take him into outside healthy scepticism goes well-kept immunity hype slogans fads fashions And important latter seems pass ever increasing frequency quite amazing -- bit sadenning -- gullible desperate willing expect salvation next gadget TV promoted theory daily dose Shakespeare every living room elevate culturally deprived unfathomed heights thus curing ills society etc And get ? Soap operas quizzes overhead projector welcomed greatest educational innovation since Socrates allowed much detailed preparation new audio-visual aids called those days revolutionize class room bring modern teaching each little Eskimo his igloo Well overhead projector teaching 'm afraid better judge than I. advent terminals interactive debugging supposed solve programming problems advent colour screens algorithm animation supposed same And get ? Commercial software disclaimer explicitly states fool if rely bought And multimedia/communication hype best bits those arrived far away if line Net n't count ( virtual anyhow ... ) Apart change vocabulary same hype same snake oil over over again me favour getting excited time supposed save switching home banking Recently James H. Billington current Librarian Congress remarked instead knowledge-based democracy may end up an information-inundated democracy share his concern allow me end simple wish May spite distractions generated technology succeed turning information into knowledge knowledge into understanding understanding into wisdom thank attention To Lifschitz web page*

*Explaination:*

*When we increase the n number, the amounts of vocabulary are decreased. Therefore, the readability of the text is decreased.*

*Part b*

*The result is: Bachelors on this festive day . shall do my best*

*We only left the words mentioned above. Other words are taken off from the text: “***to begin with I would like to thank the College of Natural Sciences for the most honouring Invitation to address its newest flock of I***”*

*Total 25 words are taken off .*

1. exercise 30.

★ With the help of the trie data structure, write a recursive function that processes text, locating the uniqueness point in each word, and discarding the remainder of each word. How much compression does this give? How readable is the resulting text?

Print the output of your function for *SimpleText*. Compute compression using the following formula: length of the text resulting from the application of your function (considered as a string) divided by the length of the original text (considered as a string).

To clarify the notion of uniqueness point, consider a text of five words “all apes ate apples”. By [] we denote the uniqueness point of each word in this text: “ a[l]l ap[e]s a[t]e ap[p]les”. Thus, the output of your function for this text should be “al ape at app”.

A trie for this text follows:

a l l

p e s

p l e s

t e

Also, https://en.wikipedia.org/wiki/Cohort\_model

has a paragraph, which provides more verbal description of where a "uniqueness point" stems from. There a term "recognition point" is used for this concept at first. Below is the extract from that paragraph:

The cohort model consists of three stages: access, selection, and integration.[[7]](https://en.wikipedia.org/wiki/Cohort_model#cite_note-Gaskell_1997_613.E2.80.93656-7) Under this model, auditory lexical retrieval begins with the first one or two speech segments, or [phonemes](https://en.wikipedia.org/wiki/Phoneme), reach the hearer's ear, at which time the mental lexicon activates every possible word that begins with that speech segment.[[8]](https://en.wikipedia.org/wiki/Cohort_model#cite_note-packard288-8) This occurs during the "access stage" and all of the possible words are known as the cohort.[[9]](https://en.wikipedia.org/wiki/Cohort_model#cite_note-harley-9) The words that are activated by the speech signal but are not the intended word are often called "competitors."[[10]](https://en.wikipedia.org/wiki/Cohort_model#cite_note-ibrahim-10)  As more speech segments enter the ear and stimulate more neurons, causing the competitors that no longer match the input to be "kicked out" or to decrease in activation.[[8]](https://en.wikipedia.org/wiki/Cohort_model#cite_note-packard288-8)[[12]](https://en.wikipedia.org/wiki/Cohort_model#cite_note-12) The processes by which words are activated and competitors rejected in the cohort model are frequently called "activation and selection" or "recognition and competition." These processes continue until an instant, called the *recognition point*,[[8]](https://en.wikipedia.org/wiki/Cohort_model#cite_note-packard288-8) at which only one word remains activated and all competitors have been kicked out. This is also known as the *uniqueness point.*

Code:

**def** insert(trie, key, value):  
 **if** key:  
 first, rest = key[0], key[1:]  
 **if** first **not in** trie:  
 trie[first] = {}  
 insert(trie[first], rest, value)  
 **else**:  
 trie[**'value'**] = value  
  
  
**def** compressedword(trie, token):  
  
 step = 0  
 subtrie = trie  
  
 **for** character **in** token:  
 subtrie = subtrie.get(character)  
 subtrie\_keys = list(subtrie.keys())  
 **if**(len(subtrie\_keys) == 1):  
 step += 1  
 **break  
 else**:  
 step += 1  
 **return** step  
  
  
  
**def** exercise30():  
 tokens = nltk.word\_tokenize(AppleText)  
 trie = {}  
  
 **for** word **in** tokens:  
 insert(trie, word, **''**)  
 trie = dict(trie) *# for nicer printing* pprint.pprint(trie)  
  
 unique\_chars = [] *# define a list of unique character output* **for** word **in** tokens:  
 lens = compressedword(trie, word)  
 print (word, **'=>'**, lens)  
  
 output = word[:lens]  
 unique\_chars.append(output)  
 print (word, **'=>'**, **'output'**, output)  
 unique = **' '**.join(unique\_chars)  
 compression = len(unique) / len(AppleText)  
 print (unique\_chars)  
 print (**'compression'**, compression)

Result:

Exercise 30

{'a': {'l': {'l': {'value': ''}},

'p': {'e': {'s': {'value': ''}},

'p': {'l': {'e': {'s': {'value': ''}}}}},

't': {'e': {'value': ''}}}}

['al', 'ape', 'at', 'app']

compression 0.6842105263157895

Change the AppleText to SimpleText.

Code:

**def** exercise30():  
 tokens = nltk.word\_tokenize(SimpleText)  
 trie = {}  
  
 **for** word **in** tokens:  
 insert(trie, word, **''**)  
 trie = dict(trie) *# for nicer printing* pprint.pprint(trie)  
  
 unique\_chars = [] *# define a list of unique character output* **for** word **in** tokens:  
 lens = compressedword(trie, word)  
 print (word, **'=>'**, lens)  
  
 output = word[:lens]  
 unique\_chars.append(output)  
 print (word, **'=>'**, **'output'**, output)  
 unique = **' '**.join(unique\_chars)  
 compression = len(unique) / len(SimpleText)  
 print (unique\_chars)  
 print (**'compression'**, compression)

Result:

Exercise 30

{'!': {'value': ''},

"'": {"'": {'value': ''}, 's': {'value': ''}},

',': {'value': ''},

'.': {'value': ''},

':': {'value': ''},

'?': {'value': ''},

'B': {'a': {'d': {'value': ''}}},

'G': {'o': {'o': {'d': {'value': ''}}}},

'H': {'o': {'w': {'value': ''}}},

'I': {'n': {'value': ''}},

'N': {'o': {'w': {'value': ''}}},

'O': {'h': {'value': ''}, 'n': {'e': {'value': ''}}},

'T': {'h': {'e': {'value': ''}}},

'W': {'h': {'o': {'value': ''}}},

'`': {'`': {'value': ''}},

'a': {'n': {'d': {'value': ''},

'o': {'t': {'h': {'e': {'r': {'value': ''}}}}}},

'r': {'e': {'value': ''}},

's': {'value': ''},

'value': '',

'w': {'a': {'y': {'value': ''}}}},

'b': {'a': {'c': {'k': {'value': ''}}, 'd': {'value': ''}},

'e': {'f': {'o': {'r': {'e': {'value': ''}}}}},

'r': {'o': {'u': {'g': {'h': {'t': {'value': ''}}}}}}},

'c': {'a': {'m': {'e': {'value': ''}}, 'n': {'value': ''}},

'o': {'n': {'c': {'e': {'r': {'n': {'value': ''}}}}}}},

'd': {'a': {'y': {'s': {'value': ''}, 'value': ''}}, 'o': {'value': ''}},

'e': {'x': {'p': {'r': {'e': {'s': {'s': {'value': ''}}}}}}},

'f': {'a': {'r': {'m': {'e': {'r': {'value': ''}}}}},

'e': {'w': {'value': ''}},

'i': {'e': {'l': {'d': {'s': {'value': ''}}}}}},

'g': {'l': {'a': {'d': {'value': ''}}},

'o': {'i': {'n': {'g': {'value': ''}}}}},

'h': {'e': {'r': {'value': ''}},

'i': {'s': {'value': ''}},

'o': {'r': {'s': {'e': {'value': ''}}}, 'w': {'value': ''}}},

'k': {'n': {'o': {'w': {'s': {'value': ''}}}}},

'l': {'u': {'c': {'k': {'y': {'value': ''}}}}},

'm': {'u': {'c': {'h': {'value': ''}}}},

'n': {'e': {'i': {'g': {'h': {'b': {'o': {'r': {'s': {'value': ''}}}}}}}},

'o': {'w': {'value': ''}}},

'r': {'a': {'n': {'value': ''}},

'e': {'p': {'l': {'i': {'e': {'d': {'value': ''}}}}}}},

't': {'h': {'a': {'t': {'value': ''}},

'e': {'i': {'r': {'value': ''}}, 'value': ''},

'i': {'n': {'g': {'value': ''}}}},

'o': {'o': {'value': ''}, 'value': ''},

'w': {'i': {'c': {'e': {'value': ''}}}}},

'w': {'e': {'r': {'e': {'value': ''}}},

'i': {'t': {'h': {'value': ''}}},

'o': {'r': {'k': {'value': ''}}}},

'y': {'o': {'u': {'value': ''}}}}

['On', 'da', ',', 'hi', 'hor', 'ra', 'aw', '.', 'T', 'ne', 'cam', 'to', 'e', 'thei', 'co', ':', '`', 'Oh', ',', 'tha', "'s", 'too', 'bad', '.', 'H', 'ar', 'y', 'go', 'to', 'wo', 'the', 'fi', 'no', '?', "''", 'T', 'fa', 're', ':', '`', 'G', 'thi', ',', 'B', 'thi', ',', 'W', 'k', '?', "''", 'I', 'a', 'fe', 'da', ',', 'hi', 'hor', 'cam', 'bac', 'and', 'br', 'ano', 'hor', 'wi', 'he', '.', 'N', ',', 'the', 'ne', 'we', 'gl', ':', '`', 'Oh', ',', 'how', 'l', '!', 'N', 'y', 'can', 'do', 'tw', 'as', 'm', 'wo', 'as', 'be', '!', "''", 'T', 'fa', 're', ':', '`', 'G', 'thi', ',', 'B', 'thi', ',', 'W', 'k', '?', "''"]

compression 0.7087378640776699