Assignment 1 kinz

October 10, 2024

```
[ ]: Assignment #1 - Sub Corpora 1 (Academic article on Jeffery Dahmer Academic
     ↪Article on Jeffery Dahmer - https://epublications.regis.edu/cgi/viewcontent.
     ↪cgi?article=1240&context=theses)
```

```
[9]: import os
     print(os.getcwd())
     print(os.listdir())
```

```
C:\Users\Kinzlee\SDA 250 Fall 2024\Assignment 1
['.ipynb_checkpoints', 'Assignment1_prep.ipynb', 'Jeffrey Dahmer_ Psychopathy
and Neglect.pdf', 'Jeffrey Dahmer_ Psychopathy and Neglect.txt',
'Untitled.ipynb']
```

```
[5]: import nltk
     from nltk import FreqDist
     import numpy
     import matplotlib
```

```
[15]: with open("./Jeffrey Dahmer_ Psychopathy and Neglect.txt", "r", encoding =
      ↪"utf8") as f:
          Jeffpsy = f.read()
      JeffTokens = nltk.word_tokenize(Jeffpsy)
```

```
[37]: # Question 1
      print("Question 1")
      print("Length:", len(JeffTokens))
      print()

      # Qusetion 2
      print("Question 2")
      JeffLength = len(JeffTokens)
      def lexical_diversity(JeffTokens):
          return len(set(JeffTokens)) / len(JeffTokens)
      fdist1 = FreqDist(JeffTokens)
      print("Lexical Diversity:", lexical_diversity(Jeffpsy))
      print()

      # Question 3
```

```python
print("Question 3")
print(fdist1.most_common(15))
print()

# Question 4
print("Question 4")
filtered_words = [word for word in JeffTokens if len(word) >= 10]
longdist = FreqDist(filtered_words)
print(longdist.most_common())
print()

# Question 5
print("Question 5")
from nltk.tokenize import sent_tokenize, word_tokenize
sentences = sent_tokenize(Jeffpsy)
sent1 = max(sentences, key=lambda sentence: len(word_tokenize(sentence)))
print(sent1)
sentlen = len(word_tokenize(longest_sentence))
print("The length of the longest sentence is:", sentlen)
print()

# Question 6
print("Question 6")
from nltk.stem.snowball import SnowballStemmer
stemmer = SnowballStemmer("english")
lsent1 = word_tokenize(longest_sentence)
stemmed_words = [stemmer.stem(word) for word in lsent1]
print(stemmed_words)
```

```
Question 1
Length: 23036

Question 2
Lexical Diversity: 0.0007104540519828852

Question 3
[(',', 1408), ('.', 794), ('the', 690), ('and', 614), ('of', 559), ('(', 489),
(')', 488), ('to', 454), ('a', 379), ('in', 226), (':', 202), ('or', 196),
('abuse', 178), ('child', 173), ('as', 167)]

Question 4
[('PSYCHOPATHY', 68), ('individuals', 53), ('development', 50), ('anti-social',
47), ('individual', 36), ('psychopathic', 35), ('tendencies', 30),
('adolescent', 30), ('experienced', 26), ('abuse/neglect', 25), ('antisocial',
21), ('information', 20), ('significant', 18), ('Maguire-Jack', 18),
('Dahmervideos', 17), ('researchers', 16), ('associated', 16), ('experiences',
16), ('interviews', 15), ('adolescents', 14), ('phenomenon', 14),
('psychological', 13), ('maltreatment', 12), ('understand', 11), ('witnessing',
```

11), ('Information', 11), ('additional', 11), ('Additionally', 10), ('aggressive', 10), ('participants', 10), ('feature=plcp', 10), ('physically', 9), ('aggression', 9), ('personality', 9), ('approximately', 9), ('intervention', 9), ('frequently', 8), ('interpersonal', 8), ('literature', 8), ('particular', 8), ('supervision', 8), ('//www.youtube.com/watch', 8), ('responsible', 7), ('fascination', 7), ('disorganized', 7), ('likelihood', 7), ('characteristics', 7), ('consequences', 7), ('frustration', 7), ('adolescence', 7), ('organizations', 7), ('Additional', 6), ('socioeconomic', 6), ('depression', 6), ('behavioral', 6), ('developing', 6), ('Dalnnervideos', 6), ('experience', 6), ('Anti-social', 6), ('activities', 6), ('participant', 6), ('questionnaire', 6), ('correlation', 6), ('interviewee', 6), ('ofInvestigation', 5), ('classified', 5), ('attachment', 5), ('hedonistic', 5), ('mutilation', 5), ('self-esteem', 5), ('interaction', 5), ('University', 5), ('considered', 5), ('neglect/abuse', 5), ('environment', 5), ('collection', 5), ('loneliness', 5), ('counseling', 5), ('Introduction', 4), ('Throughout', 4), ('enforcement', 4), ('Investigation', 4), ('biological', 4), ('environmental', 4), ('influenced', 4), ('previously', 4), ('abandonment', 4), ('overlooked', 4), ('definition', 4), ('population', 4), ('professionals', 4), ('discrepancies', 4), ('//www.youtube.comiwatch', 4), ('Development', 4), ('communities', 3), ('understanding', 3), ('exploratory', 3), ('motivation', 3), ('enviromnent', 3), ('gratification', 3), ('difficulties', 3), ('determining', 3), ('Prevention', 3), ('relationship', 3), ('Psychological', 3), ('internalize', 3), ('anti-anxiety', 3), ('altercations', 3), ('possibility', 3), ('neighborhood', 3), ('cannibalism', 3), ('detrimental', 3), ('Interpersonal', 3), ('medication', 3), ('displaying', 3), ('influential', 3), ('effectively', 3), ('similarities', 3), ('internalized', 3), ('externalizing', 3), ('conclusions', 3), ('Maltreatment', 3), ('intentional', 3), ('definitions', 3), ('caregivers', 3), ('supervised', 3), ('vulnerable', 3), ('developmental', 3), ('well-being', 3), ('protection', 3), ('independent', 3), ('experiment', 3), ('interactions', 3), ('Encyclopedia', 3), ('inter-parental', 3), ('longitudinal', 3), ('preconceived', 3), ('researching', 3), ('rationalize', 3), ('biographies', 3), ('government', 3), ('hypothesis', 3), ('International', 3), ('Comparative', 3), ('Criminology', 3), ('Description', 2), ('psychologically', 2), ('statements', 2), ('characteristic', 2), ('methodical', 2), ('successful', 2), ('prostitutes', 2), ('threatening', 2), ('conclusion', 2), ('victimization', 2), ('emotionally', 2), ('situations', 2), ('inappropriate', 2), ('self-worth', 2), ('themselves', 2), ('Individuals', 2), ('Department', 2), ('organization', 2), ('inevitably', 2), ('psychology', 2), ('eventually', 2), ('interviewed', 2), ('substantiated', 2), ('increasingly', 2), ('reprimanded', 2), ('misconduct', 2), ('exploiting', 2), ('particularly', 2), ('Dalmlervideos', 2), ('hitchhiker', 2), ('observational', 2), ('Beauregard', 2), ('criminological', 2), ('personalities', 2), ('necrophilia', 2), ('mistreated', 2), ('statistically', 2), ('Conclusions', 2), ('internalizing', 2), ('significantly', 2), ('domestically', 2), ('inpatients', 2), ('association', 2), ('beneficial', 2), ('destructive', 2), ('prevention', 2), ('Adolescents', 2), ('sufficient', 2), ('unsupervised', 2), ('disadvantages', 2), ('neglectful', 2), ('assessment', 2), ('background', 2), ('laboratory', 2), ('parent-child', 2), ('connection', 2), ('occasionally', 2), ('manipulation', 2), ('researched', 2), ('pm1icipants', 2), ('cOlmection', 2),

('encouraged', 2), ('stereotyping', 2), ('misconception', 2), ('dimensions', 2), ('implemented', 2), ('respondent', 2), ('relatively', 2), ('occurrence', 2), ('memorabilia', 2), ('patiicular', 2), ('pertaining', 2), ('geographically', 2), ('ultimately', 2), ('social-emotional', 2), ('throughout', 2), ('qualitative', 2), ('patticipant', 2), ('communication', 2), ('critically', 2), ('specialize', 2), ('supplementary', 2), ('Organizations', 2), ('Interviews', 2), ('questioned', 2), ('specializes', 2), ('pro-social', 2), ('therapists', 2), ('projection', 2), ('volitional', 2), ('inconsistencies', 2), ('Researchers', 2), ('Herrenkohl', 2), ('//www.victimsofviolence.on.calrev2/index.php', 2), ('serialldllers', 2), ('captivated', 1), ('sociologists', 1), ('criminologists', 1), ('unsuspecting', 1), ('importance', 1), ('Statistically', 1), ('discussing', 1), ('comparison', 1), ('discernible', 1), ('investigations', 1), ('homophobic', 1), ('homophobia', 1), ('homosexual', 1), ('low-middle', 1), ('sociopathic', 1), ('contradictory', 1), ('strangling', 1), ('incorporate', 1), ('intelligent', 1), ('transportation', 1), ('unemployed', 1), ('categorize', 1), ('classifying', 1), ('powercontrol', 1), ('Hedonistic', 1), ('schizophrenic', 1), ('responsibility', 1), ('Power-control', 1), ('belongings', 1), ('cOlmnunities', 1), ('regressive', 1), ('delinquency', 1), ('involvement', 1), ('desensitized', 1), ('backgrounds', 1), ('orientations', 1), ('intimidation', 1), ('continuous', 1), ('occupation', 1), ('orientation', 1), ('Punishment', 1), ('commission', 1), ('traumatize', 1), ('deliberately', 1), ('availability', 1), ('self-perception', 1), ('employment', 1), ('persuasion', 1), ('inducement', 1), ('enticement', 1), ('simulation', 1), ('molestation', 1), ('prostitution', 1), ('exploitation', 1), ('especially', 1), ('borderline', 1), ('Emotionally', 1), ('discovered', 1), ('externalize', 1), ('unexplainable', 1), ('simultaneously', 1), ('potentially', 1), ('traumatizing', 1), ('castration', 1), ('compounded', 1), ('unsubstantiated', 1), ('subsequent', 1), ('influencing', 1), ('disturbing', 1), ('manifested', 1), ('conversation', 1), ('conversations', 1), ('superficial', 1), ('everything', 1), ('confessions', 1), ('Dalunervideos', 1), ('Dahmervideos,2012e', 1), ('committing', 1), ('continually', 1), ('unprovoked', 1), ('characteristically', 1), ('officially', 1), ('manipulating', 1), ('escalating', 1), ('unfortunate', 1), ('particulady', 1), ('collecting', 1), ('mutilating', 1), ('compulsion', 1), ('recollection', 1), ('cannibalize', 1), ('uncontrollable', 1), ('unavoidable', 1), ('television', 1), ('power-control', 1), ('Literature', 1), ('comlection', 1), ("analysis's", 1), ('child/parent', 1), ('Specifically', 1), ('subcategories', 1), ('governments', 1), ('psychopath', 1), ('variations', 1), ('combinations', 1), ('readability', 1), ('foundation', 1), ('opportunity', 1), ('attractive', 1), ('ineffective', 1), ('protecting', 1), ('illustrate', 1), ('childparents', 1), ('encouraging', 1), ('Individual', 1), ('differences', 1), ('connections', 1), ('upbringing', 1), ('victimology', 1), ('Correctional', 1), ('Facilities', 1), ('patiicularly', 1), ('encompassing', 1), ('disciplining', 1), ('misbehaves', 1), ('controversial', 1), ('misbehaving', 1), ('spouse/partner', 1), ('emotionality', 1), ('Behavioral', 1), ('externalized', 1), ('Vocabulary', 1), ('consequentially', 1), ('Emotionality', 1), ('Sociability', 1), ('Temperament', 1), ('Protective', 1), ('Classification', 1), ('interpretation', 1), ('circumcision', 1), ('culturally', 1), ('requirement', 1), ('adequately', 1), ('unsuitable', 1), ('inadequate', 1), ('predictive', 1), ('combination', 1), ('longitudinally', 1), ('disciplinary',

1), ('consecutively', 1), ('Psychiatry', 1), ('threatened', 1), ('reoccurring', 1), ('intimidating', 1), ('substantial', 1), ('maladaptive', 1), ('Anti-Social', 1), ('intentionally', 1), ('Lindenburg', 1), ('overprotective', 1), ('domineering', 1), ('overprotected', 1), ('privileges', 1), ('restriction', 1), ('independence', 1), ('free-living', 1), ('friendships', 1), ('miscommunication', 1), ('protective', 1), ('monitoring', 1), ('Questionnaire', 1), ('administering', 1), ('Upbringing', 1), ('maturation', 1), ('acknowledge', 1), ('wellspecified', 1), ('operational', 1), ('accurately', 1), ('disadvantaged', 1), ('southeastern', 1), ('urbanization', 1), ('stlUctured', 1), ('assessments', 1), ('Achievement', 1), ('communicate', 1), ('identified', 1), ('professional', 1), ('replicated', 1), ('snpervisory', 1), ('neglectfhl', 1), ('discipline', 1), ('disadvantage', 1), ('inconsistent', 1), ('assembling', 1), ('demonstrate', 1), ('Theoretically', 1), ('encourages', 1), ('fundamental', 1), ('populations', 1), ('researcher', 1), ('OCCUITence', 1), ('perpetrated', 1), ('Association', 1), ('permission', 1), ('elementary', 1), ('Socioeconomic', 1), ('Tormenting', 1), ('preadolescents', 1), ('determined', 1), ('supportive', 1), ('automatically', 1), ('relationships', 1), ('pmiicipants', 1), ('Queensland', 1), ('selfadministered', 1), ('questiOimaires', 1), ('covertness', 1), ('externalization', 1), ('Checklist-Revised-Parent', 1), ('Antisocial', 1), ('Environment', 1), ('cOlmections', 1), ('preventative', 1), ('jurisdictions', 1), ('legislation', 1), ('investigators', 1), ('investigation', 1), ('universities', 1), ('California', 1), ('Demographic', 1), ('retrospective', 1), ('Computer-Assisted', 1), ('Animal-Related', 1), ('Experiences', 1), ('abuselneglect', 1), ('necessities', 1), ('confinement', 1), ('cleanliness', 1), ('infliction', 1), ('domesticated', 1), ('participated', 1), ('infiltrated', 1), ('unidentified', 1), ('fulfilling', 1), ('perpetrators', 1), ('imagination', 1), ('challenging', 1), ('bureaucracy', 1), ('underlying', 1), ('mesmerized', 1), ('memorialized', 1), ('realization', 1), ('generations', 1), ('predilections', 1), ('enlightenment', 1), ('scientitic', 1), ('generation', 1), ('dehumanizing', 1), ('statistics', 1), ('advancements', 1), ('mainstream', 1), ('profitable', 1), ('glorification', 1), ('glamorization', 1), ('conlll0dities', 1), ('Ultimately', 1), ('transparent', 1), ('superstars', 1), ('consumerism', 1), ('Samaritans', 1), ('celebrities', 1), ('Christopher', 1), ('predilection', 1), ("individual's", 1), ('humiliating', 1), ('examinations', 1), ('humiliation', 1), ('international', 1), ('constitutes', 1), ('requirements', 1), ('separately', 1), ('non-sexually', 1), ('individually', 1), ('examination', 1), ('represented', 1), ('dismembering', 1), ('sodomizing', 1), ('grandfather', 1), ('remarriage', 1), ('interested', 1), ('indifference', 1), ('incapacities', 1), ('indifferent', 1), ('considerable', 1), ('Institutional', 1), ('45CFR46.l01.b', 1), ('subsections', 1), ('instrumentation', 1), ('Consequentially', 1), ('intemalized', 1), ('Instrumentation', 1), ('integrating', 1), ('interviewing', 1), ('precaution', 1), ('questiomlaire', 1), ('open-ended', 1), ('interviewees', 1), ('palticipants', 1), ("interviewee's", 1), ('patticular', 1), ('Collection', 1), ('documented', 1), ('unobtrusive', 1), ('methodology', 1), ('rehabilitation', 1), ('throughjournal', 1), ('constmcted', 1), ('constructed', 1), ('Government', 1), ('assistance', 1), ('attempting', 1), ('statistical', 1), ('invaluable', 1), ('motivations', 1), ('heartbreaking', 1), ('revelations', 1), ('paliicipating',

1), ('respondents', 1), ('SpecialistlInterventionist', 1), ('genetically', 1),
('undiagnosed', 1), ('diagnosable', 1), ("sUll'ounding", 1), ('untreatable', 1),
('functional', 1), ('manifestations', 1), ('unadulterated', 1), ('dissecting',
1), ('anger/frustration', 1), ('mistreatment', 1), ('transference', 1),
('Abused/neglected', 1), ('discussion', 1), ('explaining', 1), ('Trickettet',
1), ('questiOlmaire', 1), ('childhood/adolescent', 1), ('dysfunctional', 1),
('repressing', 1), ('compulsive/addictive', 1), ('contributing', 1), ('free-
agent', 1), ("Reseal'ell", 1), ('SimplyPsychology.org', 1), ('conditioning', 1),
('reinforcement', 1), ('guaranteed', 1), ('contrasting', 1), ('impervious', 1),
('acceptable', 1), ('abused/neglected', 1), ('contributors', 1), ('boundaries',
1), ('increasing', 1), ('predicting', 1), ('preventing', 1), ('child-parent',
1), ('suppOllive', 1), ('Suggestions', 1), ('confronted', 1), ('enlightened',
1), ('Dalnnervideo', 1), ('productive', 1), ('discrepancy', 1), ('significmlt',
1), ('environments', 1), ('demographics', 1), ('summarized', 1),
('abused/neglect', 1), ('abuse/neglected', 1), ('REFERENCES', 1),
('//www.ncbi.n1m.nih.gov/pubmedhealthlPMHOOOI917/', 1),
('//www.acadv.org/abusers.html', 1), ('ofCriminal', 1), ('007/BF02885707', 1),
('interparental', 1), ('ofinterpersonal', 1), ('l0.1177/0886260502250081', 1),
('//www.trutv.com/library/crime/seriaIJillers/notorious/gacy/gacy_l.html', 1),
('Biography.com', 1), ('//www.biography.com/people/tedbundy-9231165', 1),
('protection.', 1), ('compassion', 1), ('Definitions', 1),
('//www.cdc.govNiolencePreventionichildmaltreatmentidefinitions.html', 1),
('integrated', 1), ('socialleaming', 1), ('doi:1O.1177/0306624X10361317', 1),
('//www.childwelfare', 1), ('gov/pubs/factsheets/long_tenu_consequences', 1),
('//www.childwelfare.gov/pubs/factsheets/whatiscan.cfm', 1), ('Sunuuaryof', 1),
('//www.childwelfare.gov/systemwide/laws~olicies/statutes/witnessdv.cfiu', 1),
('Associations', 1), ('ofInterpersonal', 1), ('doi:10.1177/0886260505283341',
1), ('v=vjIyhvnB9TM', 1), ('feature=picp', 1), ('Dalnllervideos', 1),
('v=eMpsudjaPk', 1), ('v=vdXIHPtkw88', 1), ('v=iKkyZXOfZOs', 1),
('v=9A719vAbzQ8', 1), ('v=QiHUAi8xwh4', 1), ('v=wMP850ARWXU', 1),
('v=_upLGnsle5I', 1), ('v=mROaHILumjI', 1), ('v=V2nTBXWI4Ko', 1),
('feature=relmfu', 1), ('Investigating', 1), ('co-occurring', 1),
('intelpersonal', 1), ('1036-1056.', 1), ('doi:10.1177/0886260508319362', 1),
('anti-sociality', 1), ('l0.1007/s10519-010-9336-2', 1),
('//www.fbi.gov/statsservices/publica', 1), ('tions/serial-murder', 1),
('serial-murder-l', 1), ('Aggressive', 1), ('doi:10.1002/ab.20322', 1),
('comprehensive', 1), ('//www.ncjrs.gov/pdffiles1/ojjdp/227744.pdf', 1),
('uklnews/worldnews/', 1), ('O/Serial-killer-8-charged', 1), ('-with-
threemurders.html', 1), ('doi:10.l177/1741659009335714', 1),
('//socyberty.com/Crime/jolm-waynegacy/', 1), ('hlternational', 1),
('IO.1177/1741659007082469', 1), ('//www.merckmanuals.com/home/childrens_', 1),
('health_issues/mental_health_disorders_i', 1), ('children/conduct-', 1),
('disorder.html', 1), ('supervisory', 1), ('replication', 1),
('077559504273684', 1), ('Bangdiwala', 1), ('Importance', 1), ('Pediatrics', 1),
('10.1542/peds.2006-3622', 1), ('psychopathy', 1), ('BriejTreatment', 1),
('Intervention', 1), ('O.1093/brieftreatmentlmlnn004', 1),
('lll1j.14678624.2012.01820.x', 1), ('Loneliness', 1), ("Intel'llational", 1),
('177/0306624X05274898', 1), ('Conditioning', 1),

('//www.simplypsychology.org/operant-conditioning.htmI', 1),
('//www.simp1ypsychology.orglbandura.html', 1), ('Punislunent', 1),
('llorcashealingarts.org/poisonous-pedagogy-punislmlent-followed-on-a-grand-
scalel', 1), ('psychiatric', 1), ('10.1007/s10578-011-0222-9', 1),
('Punislmlent', 1),
('llwww.crimemuseum.org/library/seriaIKillers/seriaIKillerVictim.html', 1),
('10.1177/14680173030032003', 1), ('Lindenberg', 1), ("adolescents'", 1),
('ofBehavioral', 1), ('10.1177/0165025409350949', 1), ('firesetting', 1),
('10.117710306624X04265087', 1), ('Cornerstone', 1), ('civilization', 1),
('//urantiabook.org/archive/ww', 1), ('Longitudinal', 1), ("children's", 1),
('attachments', 1), ("JOlll'l1al", 1), ('10.1177/0886260510362883', 1),
('ofResearch', 1), ('Adolescence', 1), ('.15327795.2010.0071l.x', 1),
('//www.ncbi.n1m.nih.gov/pubmedhealthIPMH0001919/', 1), ('option=com_content',
1), ('200ption=com_conte', 1), ('schizophrenia', 1), ('jill1ctional', 1),
('antisocial/psychopathic', 1), ('dysjill1ctional', 1), ('physical/sexual', 1),
('manifesting', 1), ('ojabuse/neglect', 1), ('expressing', 1),
('anger/jhlstration', 1), ('retaliation', 1), ('conjilsion-', 1),
('questioning', 1), ('hurt/bully/abuse', 1), ('inflnenced', 1), ('passionate',
1), ('11/19/2012', 1), ('addictions', 1), ('ofdevelopment', 1), ('Pertaining',
1), ('wellirnown', 1), ('contributed', 1), ("i'ee-agent", 1), ('11118/2012', 1),
('11119/2012', 1)]

Question 5
The nine dimensions of cruelty
consisted of: severity or degree of pain inflicted upon the animal, frequency of
the animal abuse,
duration or average time spent abusing the animal in one session, when the last
act of abuse was
committed (recency), diversity of animals abused (example: dogs, cats, birds …
), level of
concern for the abused animal (sentience), whether the child attempts to hide
the act
(covertness), whether the act was committed alone or with others (isolation),
and whether the
child regrets the acts of abuse(empathy) (Dadds et aI., 2006).
The length of the longest sentence is: 119

Question 6
['the', 'nine', 'dimens', 'of', 'cruelti', 'consist', 'of', ':', 'sever', 'or',
'degre', 'of', 'pain', 'inflict', 'upon', 'the', 'anim', ',', 'frequenc', 'of',
'the', 'anim', 'abus', ',', 'durat', 'or', 'averag', 'time', 'spent', 'abus',
'the', 'anim', 'in', 'one', 'session', ',', 'when', 'the', 'last', 'act', 'of',
'abus', 'was', 'commit', '(', 'recenc', ')', ',', 'divers', 'of', 'anim',
'abus', '(', 'exampl', ':', 'dog', ',', 'cat', ',', 'bird', '…', ')', ',',
'level', 'of', 'concern', 'for', 'the', 'abus', 'anim', '(', 'sentienc', ')',
',', 'whether', 'the', 'child', 'attempt', 'to', 'hide', 'the', 'act', '(',
'covert', ')', ',', 'whether', 'the', 'act', 'was', 'commit', 'alon', 'or',
'with', 'other', '(', 'isol', ')', ',', 'and', 'whether', 'the', 'child',

```
'regret', 'the', 'act', 'of', 'abus', '(', 'empathi', ')', '(', 'dadd', 'et',
'ai.', ',', '2006', ')', '.']
```

[ ]:

Assignment1_Porime

October 10, 2024

Assignment #1 - Sub Corpora 2 (Transcript of a YouTube documentary on Jeffery Dahmer - https://www.youtube.com/watch?v=Y1EWmrzD2Mk)

```
[8]:  #loading packages
      import nltk
      import numpy
      import matplotlib

      #loading data
      with open("/Users/catalinaporime/Desktop/SDA250/Assignment_1/Dahmer.txt", "r",␣
        ↪encoding = "utf8") as f:
          Dahmer = f.read()

      #tokenizing the data
      DahmerTokens = nltk.word_tokenize(Dahmer)

      #printing text (commented due to length)
      #print(Dahmer)


      #Q1:
      print("Question #1")
      lengthDahmer = len(DahmerTokens)
      print(lengthDahmer)
      print("The number of words in the Jeffrey Dahmer documentary on YouTube:",␣
        ↪lengthDahmer)
      print()

      #Q2:
      print("Question #2")
      def lexical_diversity(DahmerTokens):
          return len(set(DahmerTokens)) / len(DahmerTokens)
      lex_div = lexical_diversity(DahmerTokens)
      print(lex_div)
      print("The lexical diversity of the text is:", lex_div)
      print()

      #Q3:
```

```
print("Question #3")
from nltk.probability import FreqDist
fdist1 = FreqDist(DahmerTokens)
fdist1.most_common(13)
print("These are the 10 most common words in the text and their count:")
print(fdist1.most_common(13))
print()


#Q4:
print("Question #4")
dahmerLong = [word for word in DahmerTokens if len(word) >= 10]
longDist = FreqDist(dahmerLong)
print("These are the words that are at least 10 characters long and their count:
  ↪", longDist.most_common())
print()


#Q5:
print("Question #5")
from nltk.tokenize import sent_tokenize, word_tokenize
sentences = sent_tokenize(Dahmer)
longest_sentence = max(sentences, key=lambda sentence:␣
  ↪len(word_tokenize(sentence)))
word_count = len(word_tokenize(longest_sentence))
print("Longest sentence:", longest_sentence)
print("Number of words:", word_count)
print()


#Q6:
print("Question #6")
from nltk.stem.snowball import SnowballStemmer
stemmer = SnowballStemmer ("english")
long1 = word_tokenize(longest_sentence)
stemmed_word = [stemmer.stem(word) for word in long1]
print("This is the stemmed version of the longest sentence:")
print(stemmed_word)
```

```
Question #1
8084
The number of words in the Jeffrey Dahmer documentary on YouTube: 8084

Question #2
0.19594260267194458
The lexical diversity of the text is: 0.19594260267194458

Question #3
These are the 10 most common words in the text and their count:
[(',', 420), ('.', 388), ('the', 301), ('of', 199), ('to', 194), ('was', 189),
```

('he', 171), ('a', 165), ('and', 144), ('that', 134), ('in', 132), ('-', 127),
('Dahmer', 115)]

Question #4
These are the words that are at least 10 characters long and their count:
[('television', 6), ('themselves', 5), ('confession', 5), ('necrophilia', 4),
('detectives', 4), ('cannibalism', 3), ('constantly', 3), ('completely', 3),
('interested', 3), ('remembered', 3), ('grandmother', 3), ('14-year-old', 3),
('photographs', 3), ('essentially', 3), ('immediately', 3), ('one-bedroom', 2),
('investigate', 2), ('31-year-old', 2), ('strangling', 2), ('psychological', 2),
('classmates', 2), ('collection', 2), ('apparently', 2), ('neighborhood', 2),
('hitchhiker', 2), ('ultimately', 2), ('eventually', 2), ('dismembered', 2),
('homosexual', 2), ('opportunity', 2), ('predominantly', 2), ('atrocities', 2),
('13-year-old', 2), ('unconscious', 2), ('necrophiliac', 2), ('refrigerator',
2), ('potentially', 2), ('circumstances', 2), ('incredible', 2), ('controlling',
2), ('assistants', 2), ('whispering', 1), ('synonymous', 1), ('revelations', 1),
('journalist', 1), ('antiseptic', 1), ('throughout', 1), ('playground', 1),
('ostracized', 1), ('surrounding', 1), ('progressed', 1), ('fascinated', 1),
('decapitating', 1), ('disturbing', 1), ('intoxicated', 1), ('experiences', 1),
('ruminating', 1), ('fantasizing', 1), ('19-year-old', 1), ('abandonment', 1),
('pulverized', 1), ('sledgehammer', 1), ('experienced', 1), ('accustomed', 1),
('apprehended', 1), ('relentless', 1), ('Culpability', 1), ('compromise', 1),
('desecrating', 1), ('University', 1), ('colleagues', 1), ('dependency', 1),
('21-year-old', 1), ('discharged', 1), ('approaching', 1), ('frequenting', 1),
('bathhouses', 1), ('stigmatized', 1), ('underground', 1), ('companionship', 1),
('conversations', 1), ('prospective', 1), ('supportive', 1), ('presumably', 1),
('Dissolving', 1), ('identifying', 1), ('committing', 1), ('complaining', 1),
('benzodiazepines', 1), ('sentencing', 1), ('unrelenting', 1), ('unbeknownst',
1), ('authorities', 1), ('29-year-old', 1), ('dismembering', 1),
('constructing', 1), ('diabolical', 1), ('performing', 1), ('sacrificial', 1),
('approached', 1), ('cannibalize', 1), ('metaphorical', 1), ('incorporate', 1),
('experimenting', 1), ('semi-conscious', 1), ('subservient', 1), ('attempting',
1), ('incredibly', 1), ('unpleasant', 1), ('personality', 1), ('manufacturing',
1), ('Midwestern', 1), ('collecting', 1), ('cannibalizing', 1), ('successful',
1), ('African-American', 1), ('infrequently', 1), ('agonizingly', 1),
('Sinthasomphone', 1), ('attractive', 1), ('remarkable', 1), ('characteristic',
1), ('psychopaths', 1), ('extraordinarily', 1), ('high-pressure', 1),
('decomposing', 1), ('minorities', 1), ('continuous', 1), ('fulfillment', 1),
('insatiable', 1), ('dismemberment', 1), ('speculation', 1), ('struggling', 1),
('officially', 1), ('resistance', 1), ('investigation', 1), ('unbelievable', 1),
('fantastical', 1), ('international', 1), ('incredulous', 1), ('everything', 1),
('confessions', 1), ('involvement', 1), ('interesting', 1), ('differentiate',
1), ('investigators', 1), ('preliminary', 1), ('appearance', 1), ('good-
looking', 1), ('responsible', 1), ('psychiatry', 1), ('definition', 1),
('paraphilias', 1), ('implications', 1), ('consultant', 1), ('necrophiliacs',
1), ('accomplish', 1), ('necrophilous', 1), ('specialist', 1), ('psychiatrists',
1), ('psychiatric', 1), ('traditional', 1), ('imprisonment', 1), ('especially',
1), ('hoodwinked', 1), ('Correctional', 1), ('disconcerting', 1), ('respectful',

3

1), ('forgiveness', 1), ('Christopher', 1), ('bludgeoned', 1), ('demolished',
1), ('impossible', 1), ('Ultimately', 1), ('inexplicably', 1), ('absolutely',
1), ('perversions', 1), ('nightmares', 1), ('unquestionably', 1)]

Question #5
Longest sentence: - And then when the guy said he wanted to leave, Dahmer
clubbed him on the back of the head with a barbell
and then strangled him, then ultimately disposed of the body,
removed all the flesh, and eventually dissolved it in acid
and pulverized the bones with a sledgehammer.
Number of words: 55

Question #6
This is the stemmed version of the longest sentence:
['-', 'and', 'then', 'when', 'the', 'guy', 'said', 'he', 'want', 'to', 'leav',
',', 'dahmer', 'club', 'him', 'on', 'the', 'back', 'of', 'the', 'head', 'with',
'a', 'barbel', 'and', 'then', 'strangl', 'him', ',', 'then', 'ultim', 'dispos',
'of', 'the', 'bodi', ',', 'remov', 'all', 'the', 'flesh', ',', 'and', 'eventu',
'dissolv', 'it', 'in', 'acid', 'and', 'pulver', 'the', 'bone', 'with', 'a',
'sledgehamm', '.']

[ ]:

assignment1Emma

October 10, 2024

Assignment #1 - Sub Corpora 3 (A Reddit Post that discusses: "What Do You Think of Jeffery Dahmer?" - https://www.reddit.com/r/AskALiberal/comments/xr64cy/what_do_you_think_of_jeffery_dahmer/

```python
[37]: import nltk
      nltk.download()

      from nltk.book import *
      import numpy
      import matplotlib

      import os
      from nltk.corpus import PlaintextCorpusReader
      corpus_root = "C:/Users/emmad/SDA250_labs/Notebook/assignment1/redditData/
        ↪Dahmer_Reddit.txt"
      corpus_root

      # Question 1
      print("Question 1")
      with open(corpus_root, "r", encoding = "utf8") as f:
          Dahmer = f.read()
      DahmerTokens = nltk.word_tokenize(Dahmer)
      lengthDahmer = len(DahmerTokens)

      print("The number of words in the 'Dahmer Reddit Post' is:" ,lengthDahmer)
      print()
      print()

      # Question 2
      print("Question 2")
      def lexical_diversity(DahmerTokens):
          return len(set(DahmerTokens)) / len(DahmerTokens)
      lexical_diversity(DahmerTokens)

      print("The lexical diversity in the 'Dahmer Reddit Post' is:"␣
        ↪,lexical_diversity(DahmerTokens))
      print()
      print()
```

```python
print("Question 3")

# Question 3
fdist3 = FreqDist(DahmerTokens)
fdist3.most_common(20)
#excluding punctuation and '2y' because it's not a proper word
print("Top 10 most frequent words and their counts in the 'Dahmer Reddit Post'␣
 ↪are:" ,fdist3.most_common(20))

print("This excludes any punctuation and the word'2y' becasue it is not a␣
 ↪proper word.")
print()
print()

print("Question 4")

# Question 4
Dahmerlong = [word for word in DahmerTokens if len(word) >= 10]

longDist = FreqDist(Dahmerlong)

print("All words  in the 'Dahmer Reddit Post' that are at least 10 characters␣
 ↪long and their counts inclide the following:" ,longDist.most_common())
print()
print()

print("Question 5")

# Question 5
from nltk.tokenize import sent_tokenize, word_tokenize
sentences = sent_tokenize(Dahmer)
longest_sentence = max(sentences, key=lambda sentence:␣
 ↪len(word_tokenize(sentence)))
word_count = len(word_tokenize(longest_sentence))
print("The longest sentence in the 'Dahmer Reddit Post' is:" ,longest_sentence)
print("word count:" ,word_count)
print()
print()


print("Question 6")

#Question 6
from nltk.stem.snowball import SnowballStemmer
stemmer = SnowballStemmer ("english")
long1 = word_tokenize(longest_sentence)
```

```
stemmed_word = [stemmer.stem(word) for word in long1]
print("A stemmed version of the longest sentence is:" ,stemmed_word)
```

showing info https://raw.githubusercontent.com/nltk/nltk_data/gh-pages/index.xml
Question 1
The number of words in the 'Dahmer Reddit Post' is: 9356

Question 2
The lexical diversity in the 'Dahmer Reddit Post' is: 0.1974134245404019

Question 3
Top 10 most frequent words and their counts in the 'Dahmer Reddit Post' are:
[('.', 324), ('¥', 263), (',', 262), ('Share', 250), ('the', 246), ('to', 223),
('of', 186), ('a', 164), ('and', 155), ('I', 136), ('ago', 133), ('2y', 131),
('Downvote', 125), ('Upvote', 124), ('that', 115), ('he', 101), ('is', 98),
('was', 93), ('in', 93), ('for', 65)]
This excludes any punctuation and the word'2y' becasue it is not a proper word.

Question 4
All words in the 'Dahmer Reddit Post' that are at least 10 characters long and
their counts inclide the following: [('Progressive', 12), ('Libertarian', 11),
('conditions', 7), ('Independent', 6), ('rehabilitate', 6), ('understand', 5),
('u/almightywhacko', 5), ('almightywhacko', 5), ('consequences', 5),
('government', 5), ('Sinthasomphone', 5), ('problematic', 4), ('u/tie-
dyed_dolphin', 4), ('tie-dyed_dolphin', 4), ('confinement', 4), ('committing',
4), ('spidersinterweb', 3), ('Christopher', 3), ('rehabilitated', 3),
('u/Indrigotheir', 3), ('Indrigotheir', 3), ('effectively', 3), ('individual',
3), ('justification', 3), ('inevitably', 3), ('reasonably', 3),
('realistically', 3), ('department', 3), ('automatically', 2), ('recognized',
2), ('acceptable', 2), ('u/carlse20', 2), ('absolutely', 2), ('Hot_Dog_Cobbler',
2), ('particular', 2), ('wrongfully', 2), ('intentional', 2), ('definitely', 2),
('premeditated', 2), ('accidental', 2), ('u/Barbados_slim12', 2),
('Barbados_slim12', 2), ('retribution', 2), ('reasonable', 2), ('sanctioned',
2), ('predisposition', 2), ('difference', 2), ('rehabilitation', 2),
('rehabilitative', 2), ('u/Kerplonk', 2), ('incarcerated', 2), ('incarceration',
2), ('individuals', 2), ('afflictions', 2), ('epicgrilledchees', 2),
('areyouseriousdotard', 2), ('exceptions', 2), ('crossroads', 2), ('Democratic',
2), ('segregated', 2), ('meticulously', 2), ('ultimately', 2), ('civilization',
2), ('increasingly', 2), ('completely', 2), ('compromised', 2), ('potentially',
2), ('likelihood', 2), ('reprehensible', 2), ('compulsive', 2),
('dysregulation', 2), ('unsupervised', 2), ('paramedics', 2),
('u/AutoModerator', 1), ('AutoModerator', 1), ('originally', 1), ('moderators',
1), ('UpvoteVote', 1), ('NicklAAAAs', 1), ('alcoholism', 1), ('needlessly', 1),
('disappointed', 1), ('PM_ME_YOUR_DARKNESS', 1), ('conversation', 1),
('pathological', 1), ('questionable', 1), ('perpetrated', 1), ('disgusting', 1),
```

('tysontysontyson1', 1), ('collapsingrebel', 1), ('cannibalistic', 1),
('commitment', 1), ('compulsion', 1), ('crazy-bisquit', 1), ('intentionally',
1), ('imprisoned', 1), ('confessing', 1), ('u/imveganbro', 1), ('imveganbro',
1), ('u/ImInOverMyHead95', 1), ('ImInOverMyHead95', 1), ('everywhere', 1),
('sympathetic', 1), ('perspective', 1), ('lyssthebitchcalore', 1),
('disturbances', 1), ('remorseful', 1), ('encountered', 1),
('//youtu.be/HP5SQDdcgDk', 1), ('childhoods', 1), ('sociopathic', 1),
('preferential', 1), ('benefitted', 1), ('carelessly', 1), ('regardless', 1),
('restitution', 1), ('manslaughter', 1), ('personally', 1), ('alternative', 1),
('essentially', 1), ('cosmicnitwit', 1), ('u/Kitchen_Agency4375', 1),
('Kitchen_Agency4375', 1), ('dehumanising', 1), ('punishments', 1),
('justifying', 1), ('prisoners/people', 1), ('legislature', 1), ('republicans',
1), ('unconstitutional', 1), ('inadequacy', 1), ('inaccuracy', 1),
('identified', 1), ('preferable', 1), ('administration', 1), ('biological', 1),
('biological/chemical', 1), ('incredibly', 1), ('SoÉ.thereÕs', 1),
('defensively', 1), ('considered', 1), ('permanently', 1), ('propensity', 1),
('beforehand', 1), ('proportional', 1), ('dryadbride', 1), ('Late-Bar-8498', 1),
('interesting', 1), ('Regardless', 1), ('possibility', 1), ('re-offending', 1),
('discourage', 1), ('clarification', 1), ('eventually', 1), ('circulation', 1),
('ÒdisappearÓ', 1), ('psychopathy', 1), ('circulationÓ', 1), ('insinuating', 1),
('euphemisms', 1), ('BlueCollarBeagle', 1), ('Mississippi', 1),
('u/ExplorersxMuse', 1), ('ExplorersxMuse', 1), ('parameters', 1),
('conceivable', 1), ('whatsoever', 1), ('everything', 1), ('miscarriage', 1),
('comsumption', 1), ('institutions', 1), ('population', 1), ('preserving', 1),
('displaying', 1), ('accountable', 1), ('LoopyMercutio', 1), ('Unfortunately',
1), ('grammanarchy', 1), ('exonerated', 1), ('technicality', 1), ('exceptional',
1), ('RandomGrasspass', 1), ('u/ill-independent', 1), ('ill-independent', 1),
('euthanasia', 1), ('interpersonal', 1), ('implication', 1), ('coexisting', 1),
('imprisonment', 1), ('u/Stormlight1984', 1), ('Stormlight1984', 1),
('immediately', 1), ('Understanding', 1), ('laws/rules/policies', 1),
('existentially', 1), ('individualÕs', 1), ('fuckedupness', 1), ('illegalize-
it', 1), ('girlfriend_pregnant', 1), ('u/Scalage89', 1), ('retaliation', 1),
('prevention', 1), ('noticeable', 1), ('u/Heyoteyo', 1), ('Conservative', 1),
('rehabilitatableÉ', 1), ('participant', 1), ('fundamental', 1), ('perpetuity',
1), ('healthcare', 1), ('thoroughly', 1), ('determined', 1), ('viscerally', 1),
('u/TarnishedVictory', 1), ('TarnishedVictory', 1), ('musicalpants999', 1),
('Complicated', 1), ('surprising', 1), ('u/onikaizoku11', 1), ('onikaizoku11',
1), ('documentaries', 1), ('scientific', 1), ('individually', 1), ('unassisted',
1), ('body-specifically', 1), ('brain-should', 1), ('occurrences', 1),
('proclivities', 1), ('atrocities', 1), ('vulnerable', 1), ('betterment', 1),
('u/OpenMindTulsaBill', 1), ('OpenMindTulsaBill', 1), ('AskALiberal', 1),
('devastated', 1), ('philosophies', 1), ('introspection', 1), ('bobbyjames1986',
1), ('TheTarkovskyParadigm', 1), ('trilobright', 1), ('distinction', 1),
('misleading', 1), ('criminally', 1), ('themselves', 1), ('necessarily', 1),
('facilitated', 1), ('u/SwagLord5002', 1), ('SwagLord5002', 1), ('predicament',
1), ('considering', 1), ('reoriented', 1), ('rehabilitating', 1), ('reoffended',
1), ('conviction', 1), ('eye-for-an-eye', 1), ('executioner', 1), ('clinically',
1), ('developmentally', 1), ('singlehandedly', 1), ('authoritarian', 1),

('fearmongering', 1), ('especially', 1), ('admittedly', 1), ('nonexistent', 1),
('reoffending', 1), ('negligible', 1), ('Personally', 1), ('metaphorical', 1),
('therapeutic', 1), ('impulsive/compulsive', 1), ('redirecting', 1),
('subsequent', 1), ('combination', 1), ('obsessive-compulsive/impulsive', 1),
('internally', 1), ('fantasizing', 1), ('complicated', 1), ('frequently', 1),
('characterized', 1), ('disproportionate', 1), ('percentage', 1),
('psychotherapy', 1), ('previously', 1), ('conspiracies', 1), ('virulently', 1),
('Additionally', 1), ('supposedly', 1), ('coincidence', 1), ('universally', 1),
('not_a_flying_toy_', 1), ('Homegrownscientist', 1), ('homophobic', 1),
('24_Elsinore', 1), ('equivalent', 1), ('hydrochloric', 1), ('authorities', 1),
('threatening', 1), ('semi-naked', 1), ('masterminds', 1), ('QuixoticMarten',
1), ('deliberate', 1), ('adjustments', 1), ('enforcement', 1), ('intimidated',
1), ('BenMullen2', 1), ('prescribed', 1), ('failedtalkshowhost', 1),
('danceswithronin', 1), ('u/Tron_1981', 1), ('demographic', 1),
('Just_a_reddit_duck', 1), ('NimishApte', 1)]


Question 5
The longest sentence in the 'Dahmer Reddit Post' is: On the other hand, I've
come to believe that the prison system itself is so innately flawed compared to
other countries, that we would never be in the predicament of even considering
having to use the death penalty if we reoriented our prisons more towards
rehabilitating criminals rather than punishing them, which is not to say that I
think that the crimes that many criminals commit are acceptable by any metric,
but more that the system inevitably fucks over people who could otherwise
reasonably be released back into civilized society (namely those with drug abuse
charges) and ultimately puts them in a position where the state is forced to
divide between whether or not they should be given a life sentence or a death
sentence when realistically, if they had focused on trying to rehabilitate the
person instead of punishing them, they may have never reoffended or committed
crimes more grievous than their original conviction.
word count: 165


Question 6
A stemmed version of the longest sentence is: ['on', 'the', 'other', 'hand',
',', 'i', 've', 'come', 'to', 'believ', 'that', 'the', 'prison', 'system',
'itself', 'is', 'so', 'innat', 'flaw', 'compar', 'to', 'other', 'countri', ',',
'that', 'we', 'would', 'never', 'be', 'in', 'the', 'predica', 'of', 'even',
'consid', 'have', 'to', 'use', 'the', 'death', 'penalti', 'if', 'we',
'reorient', 'our', 'prison', 'more', 'toward', 'rehabilit', 'crimin', 'rather',
'than', 'punish', 'them', ',', 'which', 'is', 'not', 'to', 'say', 'that', 'i',
'think', 'that', 'the', 'crime', 'that', 'mani', 'crimin', 'commit', 'are',
'accept', 'by', 'ani', 'metric', ',', 'but', 'more', 'that', 'the', 'system',
'inevit', 'fuck', 'over', 'peopl', 'who', 'could', 'otherwis', 'reason', 'be',
'releas', 'back', 'into', 'civil', 'societi', '(', 'name', 'those', 'with',
'drug', 'abus', 'charg', ')', 'and', 'ultim', 'put', 'them', 'in', 'a', 'posit',
'where', 'the', 'state', 'is', 'forc', 'to', 'divid', 'between', 'whether',

```
    'or', 'not', 'they', 'should', 'be', 'given', 'a', 'life', 'sentenc', 'or', 'a',
    'death', 'sentenc', 'when', 'realist', ',', 'if', 'they', 'had', 'focus', 'on',
    'tri', 'to', 'rehabilit', 'the', 'person', 'instead', 'of', 'punish', 'them',
    ',', 'they', 'may', 'have', 'never', 'reoffend', 'or', 'commit', 'crime',
    'more', 'grievous', 'than', 'their', 'origin', 'convict', '.']
```

[ ]: