SOU Similarities

September 14, 2020

[28]:

```
import copy
import string
import math
import pylab as py
import matplotlib.pyplot as plt
import numpy as np
import json
import re
import functools
import operator

speeches=[]
f = open("../hw1/speeches.json")
for line in f:
    speeches.append(json.loads(line))

#note that speeches is a dictionary with keys "president",
#"text", "year." there are 226 total speeches

#function that converts to lower case and removes punctuation
def clean_and_split(s):
    # encode to UTF-8, convert to lowercase and translate all hyphens and
    # punctuation to whitespace
    translator = str.maketrans
    (string.punctuation, ' '*len(string.punctuation)) #map punctuation to space
    s = s.lower().replace('-',' ').translate(translator)
    s = re.sub(r'(\x9d)', ' ', s)
    s = re.sub(r'(\x9c)', ' ', s)
    s = re.sub(r'(\x99)', ' ', s)
    s = re.sub(r'(\x97)', ' ', s)
    s = re.sub(r'(\x95)', ' ', s)
    s = re.sub(r'(\x94)', ' ', s)
    s = re.sub(r'(\x93)', ' ', s)
    s = re.sub(r'(\x80)', ' ', s)
    s = re.sub(r'(\n)', ' ', s)
    s = re.sub(r'(\r)', ' ', s)
    # replace whitespace substrings with one whitespace and remove
    # leading/trailing whitespaces
    s = re.sub(' +',' ',s.strip())
    return s.split(' ')

unique_words = []
cleaned = []
#get all unique words from speeches and store split arrays
for speech in speeches:
    txt = speech.get('text')
    split = clean_and_split(txt)
    cleaned.append(split)
    unique_words = set().union(unique_words, split)
```

[57]:

```
#vectors stores unique words in each speech
vectors = []
for speech in cleaned:
    unique_speech = set(speech)
    vectors.append(unique_speech)

#count indicators
num_appearances_array = []
for term in unique_words:
    num_appearances = sum((1 if term in speech else 0) for speech in vectors)
    num_appearances_array.append(num_appearances)

#dictionary mapping unique words to number of speeches they appear in
unique_words_dict = d = dict(zip(unique_words, num_appearances_array))
```

[109]:

```
#dictionary, word: appeared in more than in 50 different speeches
uncommon_words = dict()
for term in unique_words_dict:
    if unique_words_dict[term] < 50:
        uncommon_words.update({term: 0})
    else:
        uncommon_words.update({term: 1})

#make list of all words that will be in our vector
common_words_list = [term for term in uncommon_words if uncommon_words[term] ==␣
 ↪1]
```

(a) Compute the tf-idf vectors for each SOU address. I have chosen to ignore words that appear in less than 50 speeches.

[110]:

```
#create the tf_idf vector
tf_idf = []
D = len(speeches)
i = 0
```

```python
for speech in vectors:
    vec = []
    for term in common_words_list:
        n = cleaned[i].count(term)
        if n == 0:
            vec.append(n)
            continue
        num_appearances = unique_words_dict[term]
        w = 1*math.log(D/num_appearances)
#    for term in speech:
#        if uncommon_words[term] == 1:
#            num_appearances = unique_words_dict[term]
#            n = cleaned[i].count(term) #fix this term
#            w = 1*math.log(D/num_appearances)
#        else:
#            continue
        vec.append(w)
    i = i + 1
    tf_idf.append(vec)
```

[113]:

```python
pres_name = []
for speech in speeches:
    pres_name.append(speech["president"])

diff_presidents = set(pres_name)
index_speeches = []
for pres in diff_presidents:
    indices = [i for i, x in enumerate(pres_name) if x == pres]
    index_speeches.append(indices)
```

(b1) Find 50 most similar pairs of SOU's given by different presidents

[274]:

```python
#note there are 40+39+...+2+1 comparisons to make
#note each president has more than one speech

sim_array = [] #list of comparisons by president
#presidents that come earlier in "diff_presidents"
#will be the ones to search under

diff_pres_list = list(diff_presidents)

i = 0
for pres_1 in diff_pres_list:
    pres_1_sim = [] #pres_1_sim contains list of pres_2 (for j > i)
    j = 0
    for pres_2 in diff_pres_list:
        sim_measure = [] #contains lists by index of first speech
        #for two different presidents
        if j <= i:
            j = j + 1
            continue
        indices_1 = index_speeches[i]
        indices_2 = index_speeches[j]

        m = 0
        for p1 in indices_1:
            speech_sim = [] #contains sim(d,d') (index of second speech)
            n = 0
            for p2 in indices_2:
                if n == m:
                    n = n + 1
                    continue
                #calculate vector; p1, p2 indices in tf_idf
                dot_prod = np.dot(tf_idf[p1], tf_idf[p2])
                norm1 = np.linalg.norm(tf_idf[p1])
                norm2 = np.linalg.norm(tf_idf[p2])
                sim = dot_prod/(norm1*norm2)
                speech_sim.append(sim)
                n = n + 1
            m = m + 1
            sim_measure.append(speech_sim)
        pres_1_sim.append(sim_measure)
        j = j + 1
    sim_array.append(pres_1_sim)
    i = i + 1
```

[273]:

```
flat_sim = [sim for x in sim_array for sim in x]
flat_sim = [sim for x in flat_sim for sim in x]
flat_sim = [sim for x in flat_sim for sim in x]

ind = np.argpartition(flat_sim, 50)[:50]

#first index for first president
#second index for second president
#(1 + number of presidents behind first pres)
#third index for speech number of first pres
#fourth index for speech index of fourth pres
def find_coordinate(val):
    for a, pres_1 in enumerate(sim_array):
        for b, pres_2 in enumerate(pres_1):
            for c, pres_1_speech in enumerate(pres_2):
                for d, pres_2_speech in enumerate(pres_1_speech):
                    if pres_2_speech == val:
                        return [a, b, c, d]
```

```python
for i in ind:
    coords = find_coordinate(flat_sim[i])
    pres1idx = coords[0]
    pres2idx = coords[0] + 1 + coords[1]
    speech1idx = index_speeches[pres1idx][coords[2]]
    speech2idx = index_speeches[pres2idx][coords[3]]
    print("President 1:", diff_pres_list[pres1idx],
          "in", speeches[speech1idx]["year"])
    print("President 2:", diff_pres_list[pres2idx],
         "in", speeches[speech2idx]["year"])
```

```
President 1: George Washington in 1790
President 2: Jimmy Carter in 1980
President 1: George W. Bush in 2001
President 2: James Madison in 1809
President 1: George Bush in 1990
President 2: John Adams in 1800
President 1: John Adams in 1800
President 2: George W. Bush in 2002
President 1: John Adams in 1800
President 2: George W. Bush in 2007
President 1: John Adams in 1800
President 2: George W. Bush in 2001
President 1: Thomas Jefferson in 1804
President 2: Jimmy Carter in 1979
President 1: George Washington in 1792
President 2: George W. Bush in 2002
President 1: Richard M. Nixon in 1973
President 2: John Adams in 1797
President 1: Lyndon B. Johnson in 1964
President 2: Thomas Jefferson in 1808
President 1: Richard M. Nixon in 1973
President 2: John Adams in 1798
President 1: John Adams in 1800
President 2: Barack Obama in 2009
President 1: Richard M. Nixon in 1973
President 2: John Adams in 1800
President 1: John Adams in 1799
President 2: Franklin D. Roosevelt in 1944
President 1: John Adams in 1799
President 2: Franklin D. Roosevelt in 1940
President 1: Richard M. Nixon in 1973
President 2: George Washington in 1791
President 1: Richard M. Nixon in 1973
President 2: George Washington in 1793
President 1: George Bush in 1990
President 2: Thomas Jefferson in 1805
President 1: Richard M. Nixon in 1973
President 2: James Madison in 1809
President 1: John Adams in 1800
President 2: Ronald Reagan in 1985
President 1: Richard M. Nixon in 1973
President 2: James Madison in 1813
President 1: Richard M. Nixon in 1973
President 2: James Madison in 1812
President 1: John Adams in 1800
President 2: Jimmy Carter in 1980
President 1: Thomas Jefferson in 1804
President 2: Franklin D. Roosevelt in 1940
President 1: Richard M. Nixon in 1973
President 2: Thomas Jefferson in 1802
President 1: Richard M. Nixon in 1973
President 2: Thomas Jefferson in 1803
President 1: George W. Bush in 2002
President 2: James Madison in 1814
President 1: John Adams in 1800
President 2: Ronald Reagan in 1987
President 1: George Washington in 1790
President 2: Jimmy Carter in 1979
President 1: John Adams in 1800
President 2: William J. Clinton in 1997
President 1: Richard M. Nixon in 1971
President 2: John Adams in 1798
President 1: John Adams in 1800
President 2: Barack Obama in 2010
President 1: John Adams in 1800
President 2: Barack Obama in 2011
President 1: Richard M. Nixon in 1972
President 2: Thomas Jefferson in 1803
President 1: George Bush in 1990
President 2: George Washington in 1791
President 1: Thomas Jefferson in 1804
President 2: Franklin D. Roosevelt in 1945
President 1: Richard M. Nixon in 1973
President 2: Thomas Jefferson in 1805
President 1: George Washington in 1790
President 2: Ronald Reagan in 1984
President 1: John Adams in 1799
President 2: Ronald Reagan in 1986
President 1: George Washington in 1790
President 2: Ronald Reagan in 1983
President 1: John Adams in 1800
President 2: William J. Clinton in 1995
President 1: John Adams in 1800
President 2: Ronald Reagan in 1981
President 1: John Adams in 1800
President 2: Jimmy Carter in 1979
President 1: John Adams in 1799
President 2: Franklin D. Roosevelt in 1945
President 1: John Adams in 1799
President 2: Jimmy Carter in 1979
President 1: Richard M. Nixon in 1973
President 2: George Washington in 1794
President 1: Richard M. Nixon in 1973
President 2: George Washington in 1792
President 1: George Washington in 1791
President 2: Ronald Reagan in 1983
President 1: John Adams in 1800
President 2: William J. Clinton in 1994
President 1: George Bush in 1989
President 2: John Adams in 1798
```

(b2) Find 50 most similar pairs of SOU's given by same president

[278]:

```python
same_sim_array = []

i = 0
for pres in diff_pres_list:
    pres_sim = [] #pres_sim contains a list of speeches for pres
    indices = index_speeches[i] #indexes of speeches given by pres

    m = 0
    for p1 in indices:
        speech_sim = [] #contains sim(d,d') (index of second speech)
        n = 0
        for p2 in indices:
            if n <= m:
                n = n + 1
                continue
            #calculate vector; p1, p2 indices in tf_idf
            dot_prod = np.dot(tf_idf[p1], tf_idf[p2])
            norm1 = np.linalg.norm(tf_idf[p1])
            norm2 = np.linalg.norm(tf_idf[p2])
            sim = dot_prod/(norm1*norm2)
            speech_sim.append(sim)
            n = n + 1
        m = m + 1
        pres_sim.append(speech_sim)
    same_sim_array.append(pres_sim)
    i = i + 1
```

[297]:

```
flat_same_sim = [sim for x in same_sim_array for sim in x]
flat_same_sim = [sim for x in flat_same_sim for sim in x]

ind_2 = np.argpartition(flat_same_sim, 50)[:50]

#first index for first president
#second index for speech number of pres
#third index for second speech index of pres
def find_coordinate_2(val):
    for a, pres in enumerate(same_sim_array):
        for b, pres_speech_1 in enumerate(pres):
            for c, pres_speech_2 in enumerate(pres_speech_1):
                if pres_speech_2 == val:
                    return [a, b, c]

for i in ind_2:
    coords = find_coordinate_2(flat_same_sim[i])
    presidx = coords[0]
    speech1idx = index_speeches[presidx][coords[1]]
    speech2idx = index_speeches[presidx][coords[2]]
    print("President:", diff_pres_list[presidx],
          "in years", speeches[speech1idx]["year"],
          "and", speeches[speech2idx]["year"])
```

```
President: Woodrow Wilson in years 1920 and 1918
President: George Washington in years 1790 and 1796
President: Franklin D. Roosevelt in years 1941 and 1943
President: George Washington in years 1791 and 1793
President: Franklin D. Roosevelt in years 1942 and 1938
President: Franklin D. Roosevelt in years 1944 and 1941
President: George Washington in years 1790 and 1790
President: James Madison in years 1809 and 1811
President: George Washington in years 1790 and 1793
President: James Madison in years 1814 and 1816
President: Franklin D. Roosevelt in years 1945 and 1940
President: Franklin D. Roosevelt in years 1944 and 1936
President: James Madison in years 1814 and 1814
President: George Washington in years 1796 and 1793
President: James Madison in years 1809 and 1816
President: Franklin D. Roosevelt in years 1935 and 1939
President: Woodrow Wilson in years 1916 and 1916
President: Franklin D. Roosevelt in years 1939 and 1939
President: George Washington in years 1791 and 1796
President: George Washington in years 1794 and 1791
President: George Washington in years 1794 and 1793
President: Woodrow Wilson in years 1916 and 1920
President: George Washington in years 1791 and 1791
```

```
President: George Washington in years 1793 and 1791
President: George Washington in years 1792 and 1791
President: Thomas Jefferson in years 1802 and 1803
President: George Washington in years 1790 and 1791
President: John Adams in years 1797 and 1798
President: Woodrow Wilson in years 1916 and 1914
President: George Washington in years 1793 and 1793
President: John Adams in years 1798 and 1797
President: Woodrow Wilson in years 1913 and 1914
President: Woodrow Wilson in years 1913 and 1920
President: Thomas Jefferson in years 1802 and 1806
President: Woodrow Wilson in years 1913 and 1918
President: Woodrow Wilson in years 1913 and 1915
President: Franklin D. Roosevelt in years 1937 and 1938
President: Woodrow Wilson in years 1914 and 1913
President: Woodrow Wilson in years 1914 and 1914
President: George Washington in years 1793 and 1794
President: George Washington in years 1793 and 1796
President: Woodrow Wilson in years 1920 and 1913
President: Woodrow Wilson in years 1920 and 1914
President: Franklin D. Roosevelt in years 1944 and 1943
President: Franklin D. Roosevelt in years 1943 and 1941
President: John Adams in years 1800 and 1797
President: Franklin D. Roosevelt in years 1940 and 1943
President: Woodrow Wilson in years 1916 and 1913
President: Franklin D. Roosevelt in years 1943 and 1944
President: Thomas Jefferson in years 1806 and 1801
```

(b3) Find the 25 most similar presidents

[317]:

```python
#sim_array contains the information we need here
#first index for first pres
#second index for second pres
#third index for all speeches between two presidents
avgs = []
similar_pres = []
for a, pres1 in enumerate(sim_array):
    for b, pres2 in enumerate(pres1):
        sum = 0
        n = 0
        for speech1 in pres2:
            for angle in speech1:
                sum = sum + angle
                n = n + 1
        avg = sum/n
        avgs.append(avg)
        similar_pres.append([a, a + b + 1])
```

```
ind_3 = np.argpartition(avgs, 25)[:25]
for i in ind_3:
    coords = similar_pres[i]
    print("Presidents:", diff_pres_list[coords[0]],
          "and", diff_pres_list[coords[1]])
```

```
Presidents: Richard M. Nixon and John Adams
Presidents: John Adams and Barack Obama
Presidents: George Bush and John Adams
Presidents: John Adams and Ronald Reagan
Presidents: Richard M. Nixon and George Washington
Presidents: Richard M. Nixon and Thomas Jefferson
Presidents: John Adams and Franklin D. Roosevelt
Presidents: John Adams and John F. Kennedy
Presidents: John Adams and William J. Clinton
Presidents: George Bush and James Madison
Presidents: George Bush and Thomas Jefferson
Presidents: Gerald R. Ford and John Adams
Presidents: James Madison and Ronald Reagan
Presidents: George Washington and George W. Bush
Presidents: Richard M. Nixon and James Madison
Presidents: Lyndon B. Johnson and John Adams
Presidents: John Adams and Jimmy Carter
Presidents: George Washington and Ronald Reagan
Presidents: George Washington and Barack Obama
Presidents: George Bush and George Washington
Presidents: John Adams and George W. Bush
Presidents: William J. Clinton and George Washington
Presidents: Gerald R. Ford and Thomas Jefferson
Presidents: Lyndon B. Johnson and George Washington
Presidents: Gerald R. Ford and George Washington
```

The speeches do not seem very similar. This is likely because, in order to work with the data well, I eliminated all vocabulary that does not appear in more than 50 speeches. In fact, most of the speeches given are not very similar, even in terms of vocabulary, because they all use many words that do not appear across many speeches. In order to construct a better similarity measure, we might want to consider all vocabulary, or measure the weights with some measure that considers the similarity between different words.

(c) Cluster the speeches using k-means.

[341]:

```
from sklearn.cluster import KMeans

model = KMeans(n_clusters = 8, max_iter = 50, init = "random")
sou_clust = model.fit(tf_idf)
```

[357]:

```
all_indices = []
for i in range(0, 8):
    indices = [x for x, label in enumerate(sou_clust.labels_) if label == i]
    all_indices.append(indices)

for i, cluster in enumerate(all_indices):
    print("Cluster", i + 1, ":")
    for j, index in enumerate(cluster):
        print(pres_name[index], "in", speeches[index]["year"])
    print("\n")
```

```
Cluster 1 :
Theodore Roosevelt in 1905
Theodore Roosevelt in 1908
Theodore Roosevelt in 1907
Grover Cleveland in 1895
Theodore Roosevelt in 1901
Harry S Truman in 1946
Grover Cleveland in 1894
Grover Cleveland in 1885
Theodore Roosevelt in 1903
William Howard Taft in 1909
William Howard Taft in 1911
Theodore Roosevelt in 1906
William Howard Taft in 1912
William Howard Taft in 1910
Grover Cleveland in 1896
Theodore Roosevelt in 1904
Benjamin Harrison in 1891
William McKinley in 1900
William McKinley in 1899
Jimmy Carter in 1981
William McKinley in 1898
Grover Cleveland in 1888


Cluster 2 :
James Monroe in 1821
John Quincy Adams in 1827
Abraham Lincoln in 1861
James Monroe in 1822
Ulysses S. Grant in 1871
Franklin Pierce in 1853
Abraham Lincoln in 1863
Andrew Johnson in 1866
John Quincy Adams in 1828
John Tyler in 1843
Andrew Jackson in 1833
```

```
Ulysses S. Grant in 1874
James Monroe in 1819
John Quincy Adams in 1826
Abraham Lincoln in 1864
James Monroe in 1824
John Tyler in 1841
Andrew Jackson in 1829
Millard Fillmore in 1852
Zachary Taylor in 1849
Ulysses S. Grant in 1869
John Quincy Adams in 1825
James Monroe in 1818
Ulysses S. Grant in 1876
James Monroe in 1823
Andrew Jackson in 1831


Cluster 3 :
Andrew Johnson in 1865
Andrew Jackson in 1835
John Tyler in 1844
Andrew Jackson in 1832


Cluster 4 :
Calvin Coolidge in 1923
Dwight D. Eisenhower in 1955
Herbert Hoover in 1930
Theodore Roosevelt in 1902
Calvin Coolidge in 1927
Herbert Hoover in 1931
Dwight D. Eisenhower in 1956
Warren G. Harding in 1921
Calvin Coolidge in 1928
Warren G. Harding in 1922
Herbert Hoover in 1929
Calvin Coolidge in 1924
Dwight D. Eisenhower in 1954
Calvin Coolidge in 1925
Calvin Coolidge in 1926


Cluster 5 :
Dwight D. Eisenhower in 1960
Lyndon B. Johnson in 1965
George Bush in 1991
Franklin D. Roosevelt in 1944
Lyndon B. Johnson in 1968
```

```
Harry S Truman in 1950
Barack Obama in 2010
George W. Bush in 2002
Lyndon B. Johnson in 1964
Lyndon B. Johnson in 1966
Ronald Reagan in 1988
John F. Kennedy in 1962
Franklin D. Roosevelt in 1938
Barack Obama in 2009
Harry S Truman in 1949
Richard M. Nixon in 1971
Harry S Truman in 1952
William J. Clinton in 1998
William J. Clinton in 1995
Franklin D. Roosevelt in 1940
George W. Bush in 2003
Jimmy Carter in 1979
William J. Clinton in 1999
Ronald Reagan in 1981
Barack Obama in 2013
Ronald Reagan in 1984
William J. Clinton in 1997
Ronald Reagan in 1983
Franklin D. Roosevelt in 1939
William J. Clinton in 1994
Franklin D. Roosevelt in 1941
Dwight D. Eisenhower in 1959
Ronald Reagan in 1986
Franklin D. Roosevelt in 1935
Barack Obama in 2011
Franklin D. Roosevelt in 1943
Jimmy Carter in 1980
Dwight D. Eisenhower in 1958
Gerald R. Ford in 1977
Jimmy Carter in 1978
George Bush in 1989
Richard M. Nixon in 1973
Dwight D. Eisenhower in 1957
John F. Kennedy in 1963
Barack Obama in 2012
Ronald Reagan in 1987
George W. Bush in 2006
George W. Bush in 2007
Gerald R. Ford in 1975
Lyndon B. Johnson in 1967
Herbert Hoover in 1932
Franklin D. Roosevelt in 1937
Franklin D. Roosevelt in 1945
```

```
Harry S Truman in 1953
Franklin D. Roosevelt in 1942
Ronald Reagan in 1985
Ronald Reagan in 1982
Lyndon B. Johnson in 1969
Franklin D. Roosevelt in 1936
George W. Bush in 2001
Richard M. Nixon in 1972
Harry S Truman in 1947
George Bush in 1992
Richard M. Nixon in 1974
Dwight D. Eisenhower in 1953
Richard M. Nixon in 1970
Harry S Truman in 1948
Dwight D. Eisenhower in 1961
George W. Bush in 2004
George Bush in 1990
William J. Clinton in 1993
George W. Bush in 2008
Franklin D. Roosevelt in 1934
George W. Bush in 2005
William J. Clinton in 1996
Harry S Truman in 1951
Gerald R. Ford in 1976
John F. Kennedy in 1961
William J. Clinton in 2000


Cluster 6 :
James Madison in 1816
Woodrow Wilson in 1913
Woodrow Wilson in 1914
John Adams in 1797
George Washington in 1791
Thomas Jefferson in 1802
George Washington in 1793
Grover Cleveland in 1887
George Washington in 1790
Thomas Jefferson in 1801
Thomas Jefferson in 1806
John Adams in 1798
Woodrow Wilson in 1920
Thomas Jefferson in 1808
James Madison in 1809
Woodrow Wilson in 1916
James Madison in 1814
James Madison in 1810
James Madison in 1813
```

```
George Washington in 1796
John Adams in 1800
Thomas Jefferson in 1803
George Washington in 1794
James Monroe in 1817
Woodrow Wilson in 1918
George Washington in 1792
James Madison in 1811
James Madison in 1812
Woodrow Wilson in 1919
George Washington in 1795
Thomas Jefferson in 1804
Thomas Jefferson in 1805
James Monroe in 1820
John Adams in 1799
James Madison in 1815
Thomas Jefferson in 1807
Woodrow Wilson in 1915
Woodrow Wilson in 1917


Cluster 7 :
William McKinley in 1897
Grover Cleveland in 1886
Ulysses S. Grant in 1873
Benjamin Harrison in 1889
Rutherford B. Hayes in 1877
Rutherford B. Hayes in 1879
Chester A. Arthur in 1883
Rutherford B. Hayes in 1878
Grover Cleveland in 1893
Benjamin Harrison in 1892
Rutherford B. Hayes in 1880
Benjamin Harrison in 1890
Chester A. Arthur in 1884
Chester A. Arthur in 1882
Chester A. Arthur in 1881
Ulysses S. Grant in 1872


Cluster 8 :
James K. Polk in 1848
Franklin Pierce in 1856
Andrew Johnson in 1867
James Buchanan in 1860
Ulysses S. Grant in 1875
Martin Van Buren in 1839
Martin Van Buren in 1837
```

```
James K. Polk in 1845
James Buchanan in 1859
Andrew Jackson in 1836
Millard Fillmore in 1851
Millard Fillmore in 1850
Andrew Jackson in 1830
Andrew Jackson in 1834
Martin Van Buren in 1838
Franklin Pierce in 1854
Abraham Lincoln in 1862
James K. Polk in 1846
James Buchanan in 1857
James K. Polk in 1847
Andrew Johnson in 1868
Ulysses S. Grant in 1870
James Buchanan in 1858
Franklin Pierce in 1855
John Tyler in 1842
Martin Van Buren in 1840
```

It could be interpreted that speeches of the same cluster have similar vocabulary because they are in similar year ranges. For example, cluster 8 are all speeches given in the mid-1800s. However, this trend does not seem to be true for all of the clusters. The results are not fully interpretable, but it can be interpreted that the smaller clusters are likely to be speeches that are different from other speeches (Andrew Jackson's speeches in cluster 3, for example).

[ ]: