Universal stuff

In [7]:
```python
import string
import nltk
import gensim
import numpy as np
from pathlib import Path

stopwords = nltk.corpus.stopwords.words("english")

lemmatizer = nltk.stem.WordNetLemmatizer()

def penntag_to_wordnettag(tag):
    if tag.startswith("NN"):
        return nltk.corpus.wordnet.NOUN
    elif tag.startswith("VB"):
        return nltk.corpus.wordnet.VERB
    elif tag.startswith("JJ"):
        return nltk.corpus.wordnet.ADJ
    elif tag.startswith("RB"):
        return nltk.corpus.wordnet.ADV
    else:
        return nltk.corpus.wordnet.NOUN

common_terms = ["America", "American", "Americans", "Applause.", "Applause", "applause.", "applause", "back",
                "Congress", "could", "country", "good", "government", "know", "like", "make", "must", "nation",
                "people", "President", "president", "United", "States", "They", "they", "this", "want", "would",
                "year","great", "national", "need", "many", "well", "take", "this", "trump", "think", "happen",
                "vote", "world", "time", "come", "life", "look", "never", "This", "upon", "purpose", "shall", "goal",
                "first", "every", "work", "help", "today", "meet", "tonight", "federal", "last", "right", "tell",
                "thing", "that", "much", "south", "South", "also", "believe", "north", "North", "east", "East",
                "program", "state", "policy", "increase", "thank", "give", "percent", "booing", "decision", "problem",
                "well", "begin", "APPLAUSE", "TRUMP", "HILLARY", "CLINTON", "office", "Well", "Hillary", "Clinton",
                "Trump", "audience", "bring", "BOOING", "Thank", "That", "that", "Ever", "ever", "AUDIENCE", "Even",
                "even", "continue", "folk", "Folk", "leave", "Leave", "talk", "Talk", "Lose", "lose", "Really",
                "really", "Million", "million", "Care", "care", "Place", "place", "mean", "Mean", "Love", "love",
                "Start", "start", "city"]
```

Dwight D. Eisenhower

In [2]:
```python
eisenhower_path = Path("presidential_speeches/eisenhower")

eisenhower_files = [ ]
for file in eisenhower_path.iterdir():
    if file.name != ".DS_Store":
        eisenhower_files.append(file)

with open("output_file", "w") as outfile:
    for fname in eisenhower_files:
        with open(fname) as infile:
            outfile.write(infile.read())

with open("output_file") as f:
    eisenhower_uncleaned = f.read()

eisenhower_tagged = nltk.pos_tag(nltk.word_tokenize(eisenhower_uncleaned))

eisenhower_lemmatized = [ ]
for word, tag in eisenhower_tagged:
    wntag = penntag_to_wordnettag(tag)
    eisenhower_lemmatized.append(lemmatizer.lemmatize(word, wntag))

eisenhower_text = [ ]
for word in eisenhower_lemmatized:
    if word not in stopwords and word not in common_terms and len(word) >= 4 and word.strip(string.punctuation) != "":
        eisenhower_text.append(word.lower())

eisenhower_corpus = [ ]
eisenhower_corpus.append(eisenhower_text)

eisenhower_dictionary = gensim.corpora.Dictionary(eisenhower_corpus)
eisenhower_gensim = [ ]
for document in eisenhower_corpus:
    eisenhower_gensim.append(eisenhower_dictionary.doc2bow(document))

eisenhower_model = gensim.models.ldamodel.LdaModel(eisenhower_gensim,
                                                   num_topics = 5,
                                                   id2word = eisenhower_dictionary,
                                                   passes = 20, iterations = 500,
                                                   alpha = "asymmetric",
                                                   random_state = 0)

for topic in eisenhower_model.print_topics(num_words = 10):
    print(topic)
```

```
(0, '0.000*"peace" + 0.000*"free" + 0.000*"soviet" + 0.000*"area" + 0.000*"faith" + 0.000*"future" + 0.000*"power" + 0.000*"security" + 0.000*"atomic" + 0.000*"hope"')
(1, '0.011*"peace" + 0.009*"free" + 0.006*"soviet" + 0.005*"hope" + 0.005*"future" + 0.005*"party" + 0.005*"atomic" + 0.005*"power" + 0.004*"security" + 0.004*"freedom"')
(2, '0.000*"free" + 0.000*"peace" + 0.000*"soviet" + 0.000*"power" + 0.000*"atomic" + 0.000*"faith" + 0.000*"party" + 0.000*"principle" + 0.000*"hope" + 0.000*"security"')
(3, '0.000*"peace" + 0.000*"free" + 0.000*"soviet" + 0.000*"nations" + 0.000*"atomic" + 0.000*"security" + 0.000*"power" + 0.000*"future" + 0.000*"military" + 0.000*"party"')
(4, '0.000*"peace" + 0.000*"free" + 0.000*"soviet" + 0.000*"hope" + 0.000*"power" + 0.000*"future" + 0.000*"party" + 0.000*"strength" + 0.000*"nations" + 0.000*"principle"')
```

Richard Nixon

In [8]:
```python
dtype = np.float64

nixon_path = Path("presidential_speeches/nixon")

nixon_files = [ ]
for file in nixon_path.iterdir():
    if file.name != ".DS_Store":
        nixon_files.append(file)

with open("output_file", "w") as outfile:
    for fname in nixon_files:
        with open(fname) as infile:
            outfile.write(infile.read())

with open("output_file") as f:
    nixon_uncleaned = f.read()

nixon_tagged = nltk.pos_tag(nltk.word_tokenize(nixon_uncleaned))

nixon_lemmatized = [ ]
for word, tag in nixon_tagged:
    wntag = penntag_to_wordnettag(tag)
    nixon_lemmatized.append(lemmatizer.lemmatize(word, wntag))

nixon_text = [ ]
for word in nixon_lemmatized:
    if word not in stopwords and word not in common_terms and len(word) >= 4 and word.strip(string.punctuation) != "":
        nixon_text.append(word.lower())

nixon_corpus = [ ]
nixon_corpus.append(nixon_text)

nixon_dictionary = gensim.corpora.Dictionary(nixon_corpus)
nixon_gensim = [ ]
for document in nixon_corpus:
    nixon_gensim.append(nixon_dictionary.doc2bow(document))

nixon_model = gensim.models.ldamodel.LdaModel(nixon_gensim,
                                              num_topics = 5,
                                              id2word = nixon_dictionary,
                                              passes = 20, iterations = 500,
                                              alpha = "asymmetric",
                                              random_state = 0)

for topic in nixon_model.print_topics(num_words = 10):
    print(topic)
```

```
(0, '0.000*"vietnam" + 0.000*"peace" + 0.000*"force" + 0.000*"action" + 0.000*"watergate" + 0.000*"history" + 0.000*"power" + 0.000*"hope" + 0.000*"concern" + 0.000*"negotiation"')
(1, '0.000*"vietnam" + 0.000*"peace" + 0.000*"force" + 0.000*"vietnamese" + 0.000*"watergate" + 0.000*"full" + 0.000*"action" + 0.000*"together" + 0.000*"house" + 0.000*"strong"')
(2, '0.000*"peace" + 0.000*"vietnam" + 0.000*"home" + 0.000*"force" + 0.000*"action" + 0.000*"become" + 0.000*"effort" + 0.000*"responsibility" + 0.000*"history" + 0.000*"power"')
(3, '0.000*"peace" + 0.000*"vietnam" + 0.000*"action" + 0.000*"house" + 0.000*"future" + 0.000*"power" + 0.000*"question" + 0.000*"support" + 0.000*"force" + 0.000*"responsibility"')
(4, '0.014*"peace" + 0.010*"vietnam" + 0.004*"force" + 0.004*"action" + 0.003*"responsibility" + 0.003*"vietnamese" + 0.003*"future" + 0.003*"watergate" + 0.003*"negotiation" + 0.003*"house"')
```

Gerald Ford

In [4]:
```python
ford_path = Path("presidential_speeches/ford")

ford_files = [ ]
for file in ford_path.iterdir():
    if file.name != ".DS_Store":
        ford_files.append(file)

with open("output_file", "w") as outfile:
    for fname in ford_files:
        with open(fname) as infile:
            outfile.write(infile.read())

with open("output_file") as f:
    ford_uncleaned = f.read()

ford_tagged = nltk.pos_tag(nltk.word_tokenize(ford_uncleaned))

ford_lemmatized = [ ]
for word, tag in ford_tagged:
    wntag = penntag_to_wordnettag(tag)
    ford_lemmatized.append(lemmatizer.lemmatize(word, wntag))

ford_text = [ ]
for word in ford_lemmatized:
    if word not in stopwords and word not in common_terms and len(word) >= 4 and word.strip(string.punctuation) != "":
        ford_text.append(word.lower())

ford_corpus = [ ]
ford_corpus.append(ford_text)

ford_dictionary = gensim.corpora.Dictionary(ford_corpus)
ford_gensim = [ ]
for document in ford_corpus:
    ford_gensim.append(ford_dictionary.doc2bow(document))

ford_model = gensim.models.ldamodel.LdaModel(ford_gensim,
                                             num_topics = 5,
                                             id2word = ford_dictionary,
                                             passes = 20, iterations = 500,
                                             alpha = "asymmetric",
                                             random_state = 0)

for topic in ford_model.print_topics(num_words = 10):
    print(topic)
```

```
(0, '0.000*"energy" + 0.000*"peace" + 0.000*"economic" + 0.000*"economy" + 0.000*"food" + 0.000*"foreign" + 0.000*"union" + 0.000*"future" + 0.000*"home" + 0.000*"inflation"')
(1, '0.000*"energy" + 0.000*"peace" + 0.000*"future" + 0.000*"foreign" + 0.000*"economic" + 0.000*"europe" + 0.000*"food" + 0.000*"action" + 0.000*"inflation" + 0.000*"price"')
(2, '0.000*"energy" + 0.000*"europe" + 0.000*"foreign" + 0.000*"achieve" + 0.000*"economic" + 0.000*"food" + 0.000*"future" + 0.000*"price" + 0.000*"action" + 0.000*"peace"')
(3, '0.000*"energy" + 0.000*"peace" + 0.000*"future" + 0.000*"economic" + 0.000*"power" + 0.000*"union" + 0.000*"foreign" + 0.000*"food" + 0.000*"economy" + 0.000*"security"')
(4, '0.008*"energy" + 0.005*"peace" + 0.005*"future" + 0.004*"economic" + 0.004*"foreign" + 0.003*"food" + 0.003*"security" + 0.003*"economy" + 0.003*"price" + 0.003*"union"')
```

Ronald Reagan

In [5]:
```python
reagan_path = Path("presidential_speeches/reagan")

reagan_files = [ ]
for file in reagan_path.iterdir():
    if file.name != ".DS_Store":
        reagan_files.append(file)

with open("output_file", "w") as outfile:
    for fname in reagan_files:
        with open(fname) as infile:
            outfile.write(infile.read())

with open("output_file") as f:
    reagan_uncleaned = f.read()

reagan_tagged = nltk.pos_tag(nltk.word_tokenize(reagan_uncleaned))

reagan_lemmatized = [ ]
for word, tag in reagan_tagged:
    wntag = penntag_to_wordnettag(tag)
    reagan_lemmatized.append(lemmatizer.lemmatize(word, wntag))

reagan_text = [ ]
for word in reagan_lemmatized:
    if word not in stopwords and word not in common_terms and len(word) >= 4 and word.strip(string.punctuation) != "":
        reagan_text.append(word.lower())

reagan_corpus = [ ]
reagan_corpus.append(reagan_text)

reagan_dictionary = gensim.corpora.Dictionary(reagan_corpus)
reagan_gensim = [ ]
for document in reagan_corpus:
    reagan_gensim.append(reagan_dictionary.doc2bow(document))

reagan_model = gensim.models.ldamodel.LdaModel(reagan_gensim,
                                               num_topics = 5,
                                               id2word = reagan_dictionary,
                                               passes = 20, iterations = 500,
                                               alpha = "asymmetric",
                                               random_state = 0)

for topic in reagan_model.print_topics(num_words = 10):
    print(topic)
```

```
(0, '0.000*"peace" + 0.000*"soviet" + 0.000*"freedom" + 0.000*"force" + 0.000*"union" + 0.000*"economic" + 0.000*"together" + 0.000*"family" + 0.000*"future" + 0.000*"hope"')
(1, '0.000*"freedom" + 0.000*"peace" + 0.000*"soviet" + 0.000*"force" + 0.000*"economic" + 0.000*"free" + 0.000*"human" + 0.000*"there" + 0.000*"hope" + 0.000*"family"')
(2, '0.007*"peace" + 0.006*"freedom" + 0.005*"soviet" + 0.004*"force" + 0.004*"economic" + 0.003*"free" + 0.003*"hope" + 0.003*"family" + 0.003*"union" + 0.003*"future"')
(3, '0.000*"peace" + 0.000*"freedom" + 0.000*"soviet" + 0.000*"force" + 0.000*"free" + 0.000*"economic" + 0.000*"hope" + 0.000*"future" + 0.000*"defense" + 0.000*"child"')
(4, '0.000*"peace" + 0.000*"freedom" + 0.000*"economic" + 0.000*"force" + 0.000*"soviet" + 0.000*"support" + 0.000*"child" + 0.000*"family" + 0.000*"together" + 0.000*"there"')
```

George H. W. Bush

In [6]:
```python
bush_path = Path("presidential_speeches/bush")

bush_files = [ ]
for file in bush_path.iterdir():
    if file.name != ".DS_Store":
        bush_files.append(file)
bush_files

with open("output_file", "w") as outfile:
    for fname in bush_files:
        with open(fname) as infile:
            outfile.write(infile.read())

with open("output_file") as f:
    bush_uncleaned = f.read()

bush_tagged = nltk.pos_tag(nltk.word_tokenize(bush_uncleaned))

bush_lemmatized = [ ]
for word, tag in bush_tagged:
    wntag = penntag_to_wordnettag(tag)
    bush_lemmatized.append(lemmatizer.lemmatize(word, wntag))

bush_text = [ ]
for word in bush_lemmatized:
    if word not in stopwords and word not in common_terms and len(word) >= 4 and word.strip(string.punctuation) != "":
        bush_text.append(word.lower())

bush_corpus = [ ]
bush_corpus.append(bush_text)

bush_dictionary = gensim.corpora.Dictionary(bush_corpus)
bush_gensim = [ ]
for document in bush_corpus:
    bush_gensim.append(bush_dictionary.doc2bow(document))
bush_gensim

bush_model = gensim.models.ldamodel.LdaModel(bush_gensim,
                                             num_topics = 5,
                                             id2word = bush_dictionary,
                                             passes = 20, iterations = 500,
                                             alpha = "asymmetric",
                                             random_state = 0)

for topic in bush_model.print_topics(num_words = 10):
    print(topic)
```

```
(0, '0.007*"force" + 0.004*"iraq" + 0.004*"peace" + 0.004*"soviet" + 0.004*"child" + 0.004*"future" + 0.004*"security" + 0.004*"freedom" + 0.004*"change" + 0.003*"union"')
(1, '0.000*"force" + 0.000*"security" + 0.000*"peace" + 0.000*"future" + 0.000*"soviet" + 0.000*"military" + 0.000*"freedom" + 0.000*"home" + 0.000*"iraq" + 0.000*"support"')
(2, '0.000*"force" + 0.000*"iraq" + 0.000*"free" + 0.000*"child" + 0.000*"soviet" + 0.000*"future" + 0.000*"peace" + 0.000*"union" + 0.000*"freedom" + 0.000*"security"')
(3, '0.000*"force" + 0.000*"change" + 0.000*"iraq" + 0.000*"stand" + 0.000*"soviet" + 0.000*"peace" + 0.000*"there" + 0.000*"together" + 0.000*"freedom" + 0.000*"economic"')
(4, '0.000*"force" + 0.000*"future" + 0.000*"iraq" + 0.000*"peace" + 0.000*"change" + 0.000*"friend" + 0.000*"freedom" + 0.000*"child" + 0.000*"free" + 0.000*"family"')
```

George W. Bush

In [7]:
```python
gwbush_path = Path("presidential_speeches/gwbush")

gwbush_files = [ ]
for file in gwbush_path.iterdir():
    if file.name != ".DS_Store":
        gwbush_files.append(file)
gwbush_files

with open("output_file", "w") as outfile:
    for fname in gwbush_files:
        with open(fname) as infile:
            outfile.write(infile.read())

with open("output_file") as f:
    gwbush_uncleaned = f.read()

gwbush_tagged = nltk.pos_tag(nltk.word_tokenize(gwbush_uncleaned))

gwbush_lemmatized = [ ]
for word, tag in gwbush_tagged:
    wntag = penntag_to_wordnettag(tag)
    gwbush_lemmatized.append(lemmatizer.lemmatize(word, wntag))

gwbush_text = [ ]
for word in gwbush_lemmatized:
    if word not in stopwords and word not in common_terms and len(word) >= 4 and word.strip(string.punctuation) != "":
        gwbush_text.append(word.lower())

gwbush_corpus = [ ]
gwbush_corpus.append(gwbush_text)

gwbush_dictionary = gensim.corpora.Dictionary(gwbush_corpus)
gwbush_gensim = [ ]
for document in gwbush_corpus:
    gwbush_gensim.append(gwbush_dictionary.doc2bow(document))
gwbush_gensim

gwbush_model = gensim.models.ldamodel.LdaModel(gwbush_gensim,
                                               num_topics = 5,
                                               id2word = gwbush_dictionary,
                                               passes = 20, iterations = 500,
                                               alpha = "asymmetric",
                                               random_state = 0)

for topic in gwbush_model.print_topics(num_words = 10):
    print(topic)
```

```
(0, '0.000*"iraq" + 0.000*"terrorist" + 0.000*"security" + 0.000*"iraqi" + 0.000*"freedom" + 0.000*"citizen" + 0.000*"health" + 0.000*"peace" + 0.000*"free" + 0.000*"border"')
(1, '0.000*"iraq" + 0.000*"terrorist" + 0.000*"freedom" + 0.000*"security" + 0.000*"woman" + 0.000*"force" + 0.000*"health" + 0.000*"regime" + 0.000*"citizen" + 0.000*"weapon"')
(2, '0.008*"iraq" + 0.006*"terrorist" + 0.005*"security" + 0.005*"freedom" + 0.004*"citizen" + 0.004*"iraqi" + 0.004*"child" + 0.003*"peace" + 0.003*"health" + 0.003*"force"')
(3, '0.000*"freedom" + 0.000*"iraq" + 0.000*"security" + 0.000*"terrorist" + 0.000*"child" + 0.000*"iraqi" + 0.000*"citizen" + 0.000*"health" + 0.000*"human" + 0.000*"enemy"')
(4, '0.000*"terrorist" + 0.000*"security" + 0.000*"iraq" + 0.000*"freedom" + 0.000*"citizen" + 0.000*"iraqi" + 0.000*"peace" + 0.000*"worker" + 0.000*"woman" + 0.000*"economy"')
```

Donald Trump

In [8]:
```python
trump_path = Path("presidential_speeches/trump")

trump_files = [ ]
for file in trump_path.iterdir():
    if file.name != ".DS_Store":
        trump_files.append(file)
trump_files

with open("output_file", "w") as outfile:
    for fname in trump_files:
        with open(fname) as infile:
            outfile.write(infile.read())

with open("output_file") as f:
    trump_uncleaned = f.read()

trump_tagged = nltk.pos_tag(nltk.word_tokenize(trump_uncleaned))

trump_lemmatized = [ ]
for word, tag in trump_tagged:
    wntag = penntag_to_wordnettag(tag)
    trump_lemmatized.append(lemmatizer.lemmatize(word, wntag))

trump_text = [ ]
for word in trump_lemmatized:
    if word not in stopwords and word not in common_terms and len(word) >= 4 and word.strip(string.punctuation) != "":
        trump_text.append(word.lower())

trump_corpus = [ ]
trump_corpus.append(trump_text)

trump_dictionary = gensim.corpora.Dictionary(trump_corpus)
trump_gensim = [ ]
for document in trump_corpus:
    trump_gensim.append(trump_dictionary.doc2bow(document))
trump_gensim

trump_model = gensim.models.ldamodel.LdaModel(trump_gensim,
                                              num_topics = 5,
                                              id2word = trump_dictionary,
                                              passes = 20, iterations = 500,
                                              alpha = "asymmetric",
                                              random_state = 0)

for topic in trump_model.print_topics(num_words = 10):
    print(topic)
```

```
(0, '0.006*"deal" + 0.005*"build" + 0.005*"money" + 0.005*"trade" + 0.005*"wall" + 0.004*"border" + 0.004*"remember" + 0.004*"change" + 0.004*"fight" + 0.004*"mexico"')
(1, '0.000*"money" + 0.000*"remember" + 0.000*"deal" + 0.000*"border" + 0.000*"campaign" + 0.000*"wall" + 0.000*"build" + 0.000*"mexico" + 0.000*"what" + 0.000*"trade"')
(2, '0.000*"deal" + 0.000*"trade" + 0.000*"build" + 0.000*"wall" + 0.000*"money" + 0.000*"border" + 0.000*"four" + 0.000*"keep" + 0.000*"thousand" + 0.000*"mexico"')
(3, '0.000*"money" + 0.000*"build" + 0.000*"trade" + 0.000*"deal" + 0.000*"border" + 0.000*"stop" + 0.000*"wall" + 0.000*"hear" + 0.000*"remember" + 0.000*"mexico"')
(4, '0.000*"money" + 0.000*"deal" + 0.000*"build" + 0.000*"change" + 0.000*"wall" + 0.000*"border" + 0.000*"remember" + 0.000*"trade" + 0.000*"fight" + 0.000*"everybody"')
```