

# Testimony Classifier and Fact Dispenser

TO ASSIST LEGAL MINDS AND NEWS MEDIA IN THEIR QUICK SEARCH FOR  
RELEVANT TESTIMONY OF FACTS

# Dataset:

- Congressional testimony transcripts from
  - <https://www.congress.gov/congressional-record>
- Files were downloaded in pdf format then converted into txt using Adobe.



# Problem:

Unreliable speech from public officials and news networks have reached a head in our nation.

# How can this be helped?

- News outlets, law offices, researchers, as well as others have an innate desire to be able to easily find relevant information within current, past, and future cases on the fly (immediate basis).
- The purpose is to allow cross-referencing multiple testimonies and other public speech for inconsistencies and inaccuracies over time.



# Solution:

- ▶ Collect testimony, commentary, etc. from a target witness or person over years in the past to feed into our classifier. The A.I. will classify the data and allow for future comments and testimony to update the database as well as compare to existing testimony.

# Why?

- ▶ As one can imagine, having such a tool at your disposal during an interview can be powerful in challenging testimony. This could broaden out to public statements made over news networks, social media, etc. to combat the misinformation propagated as of late.

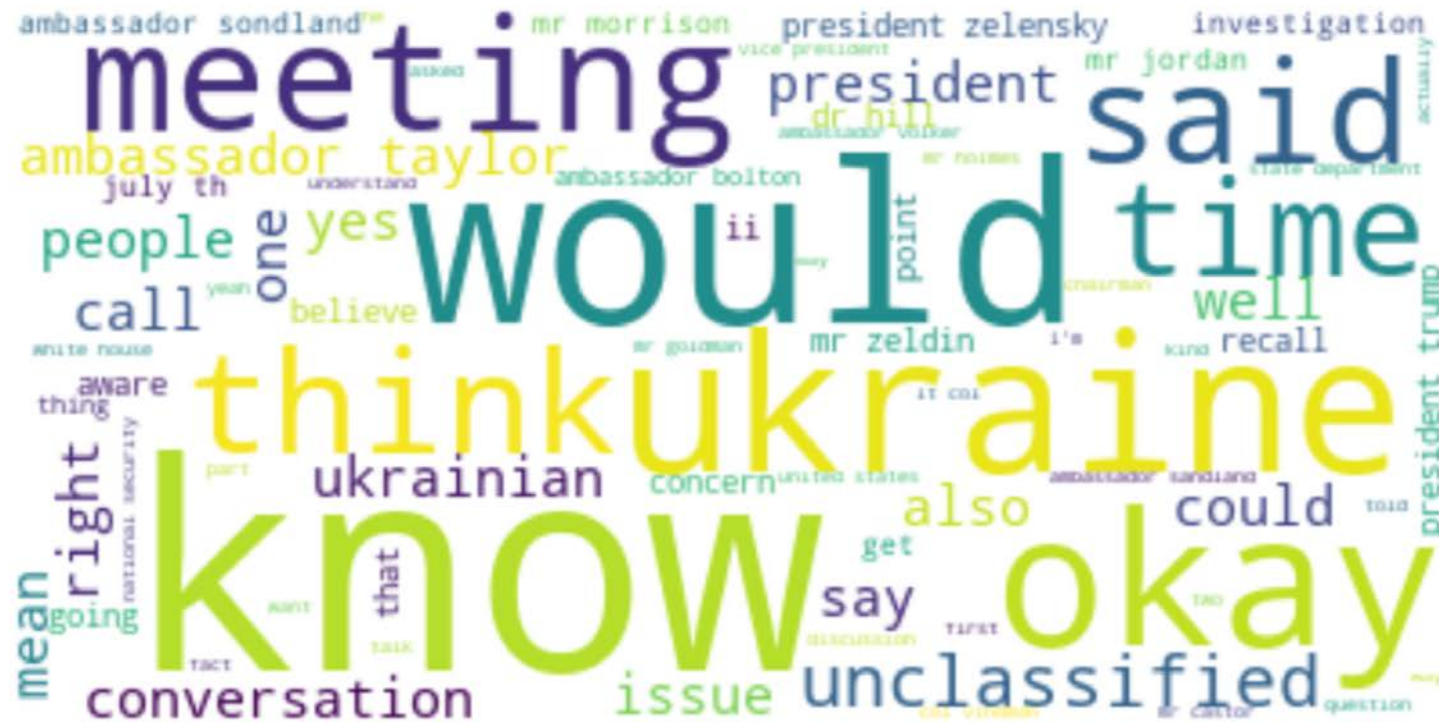


# Data Cleaning:

Utilizing functions as well as stopwords (not shown) was key to help reduce the text into easier to process information.

```
def testpreptxt(FILE):  
    # Perform some basic cleaning and character removal.  
  
    # Make everything lower case.  
    text = FILE.lower()  
  
    # Remove non-text characters.  
    text = text.replace(r'\.|\\!|\\?|\\'|, *|-|\\(|\\)', '')  
    text = re.sub(r'\\d+', '', text)  
    text = re.sub('\\n', '', text)  
    text = re.sub('[\'\"'-]', '', text)  
  
    return text  
  
#cleaning  
alltestclean = testpreptxt(alltest)
```

```
#creating the wordcloud from all testimony input
plt.figure(figsize=(16,13))
wordcloud(alltestclean)
```





# Fact Machine:

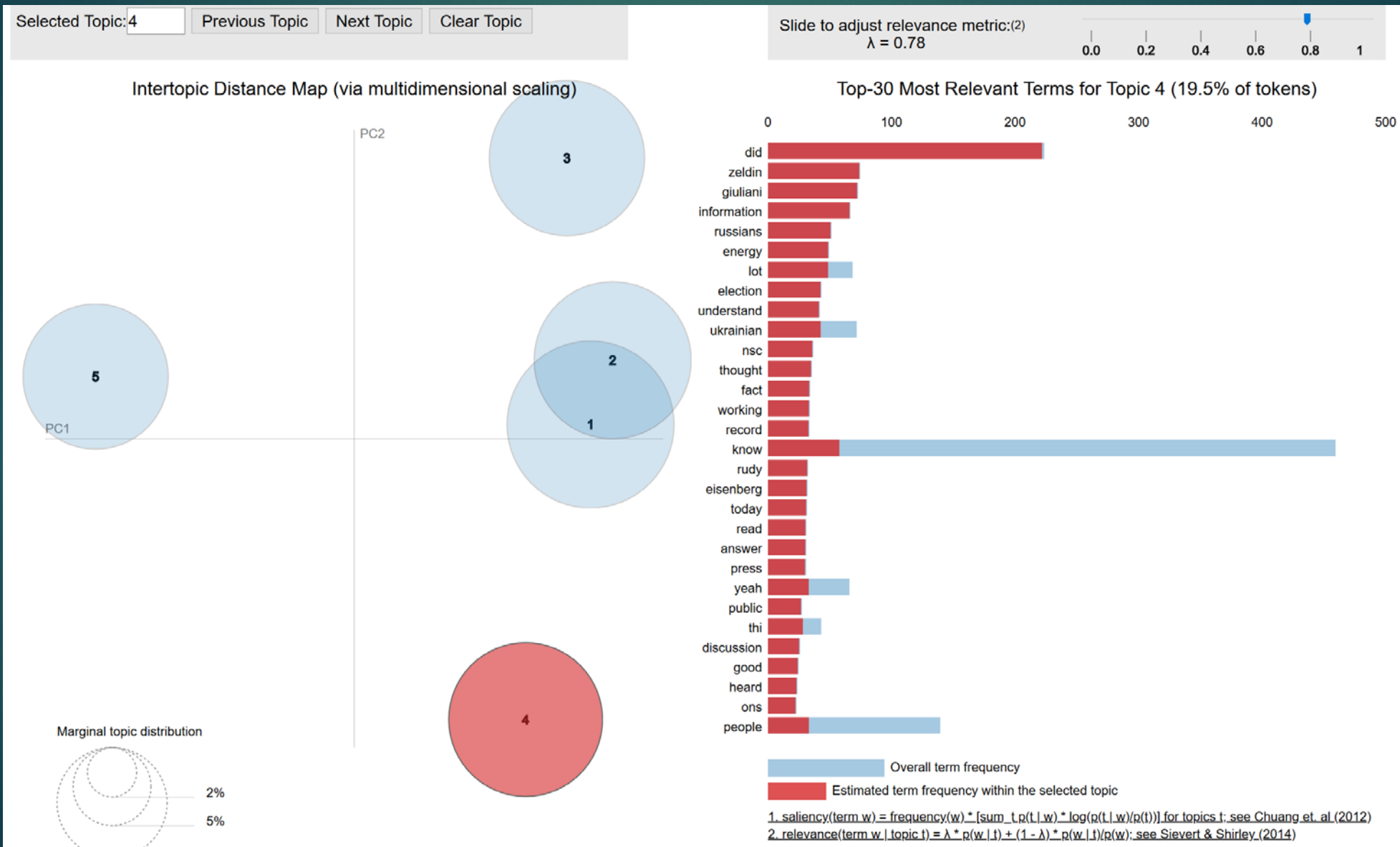
```
» def topicinfo(sourcefile, topic):
    statements = textacy.extract.semistructured_statements(sourcefile, topic)
    # Print the results
    print("Here are the things I know about", topic, ":")
    for statement in statements:
        subject, verb, fact = statement
        print(f" - {fact}")
```

```
» topicinfo(alltestnlp, 'biden')
```

Here are the things I know about biden :

- on the board of
- the point man for the obama admi ni strati on wi th ukrai ne.
- part of the board of burisma
- a board member of burisma
- qualified for that position
- on its board
- pushing the ukrainians very hard on corruption and the allegation you know the allegation
- the top cover.
- the one that communicated
- on the board
- on the board of a company owned by somebody
- on the board of an energy company

# Topics and Visualizing:





# In the Future...

- ▶ Not only could AI be an effective tool in ensuring the accountability of truth to supersede all else but, we could create a subscription service to which news outlets and other researchers subscribe to have access to the neural network classifier utilizing datasets collected. Future upgrade the subscription service could use the large datasets to run personality analyses on anyone with a public social account. The benefit of this could target human resource prospecting applicants in hiring, negotiations between individuals/companies/foreign leaders, as well as other beneficial applications not listed here.



Thank you!