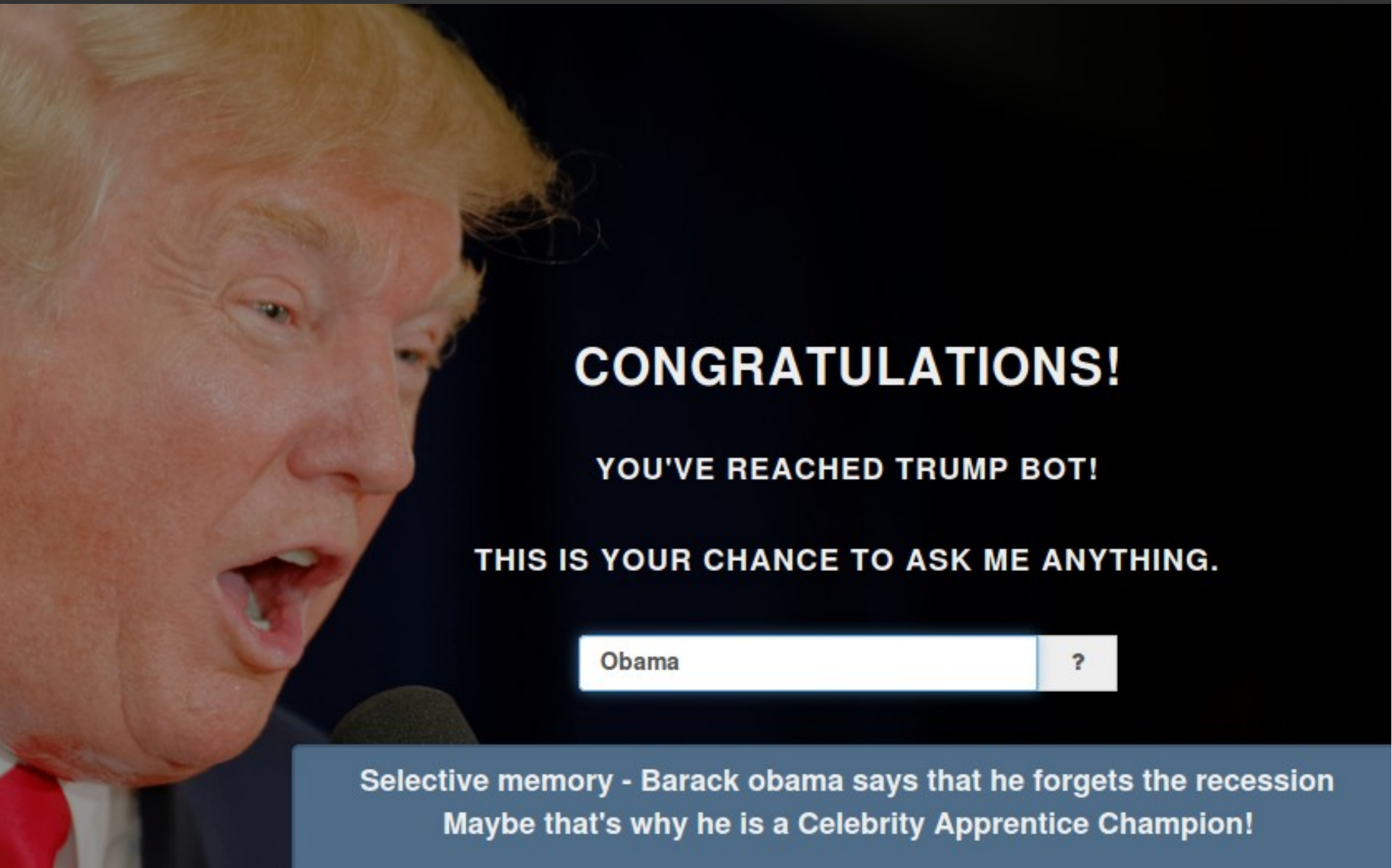
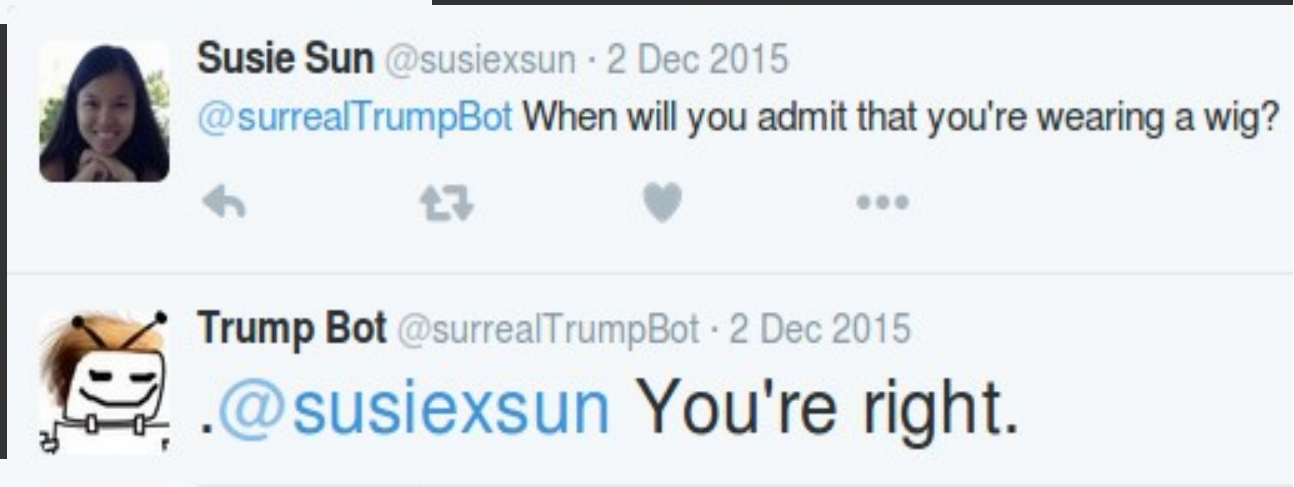
Trump Bot

Speaks just like the real Trump!

Web App


CONGRATULATIONS!
YOU'VE REACHED TRUMP BOT!
THIS IS YOUR CHANCE TO ASK ME ANYTHING.
Obama
?
Selective memory - Barack obama says that he forgets the recession
Maybe that's why he is a Celebrity Apprentice Champion!

@surrealTrumpBot



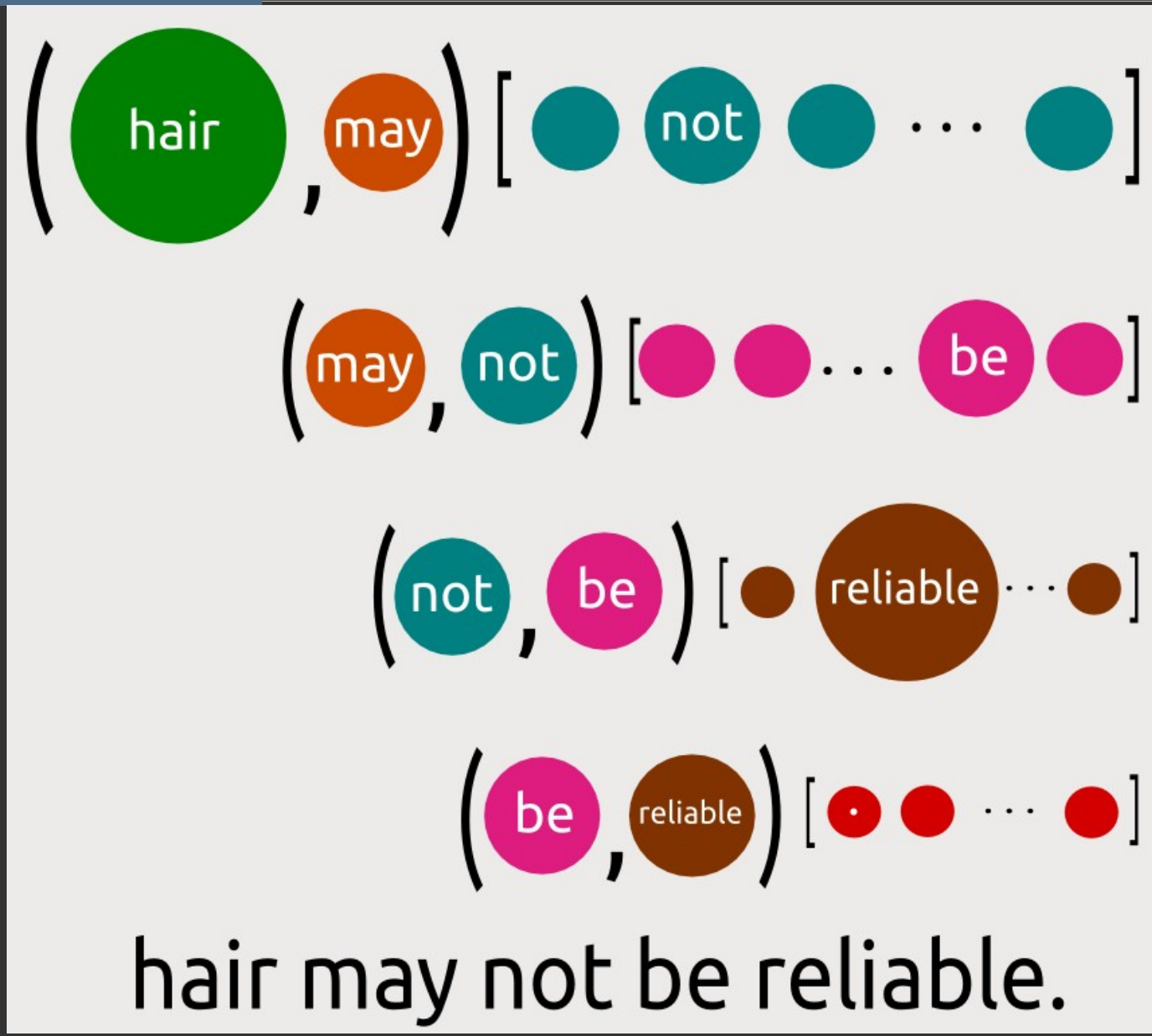
How does it work?

How does it work?

Bi-gram Markov Model

hair ?

My hair may not be reliable.



The Model

- Stores probabilities of all possible words following a bi-gram in the corpus.
- For the English language, this would be over 100,000 probabilities per bi-gram.
- Fortunately, Trump's vocabulary is way smaller than that!

Trump's Vocabulary

~250,000 words

~14,000 unique words

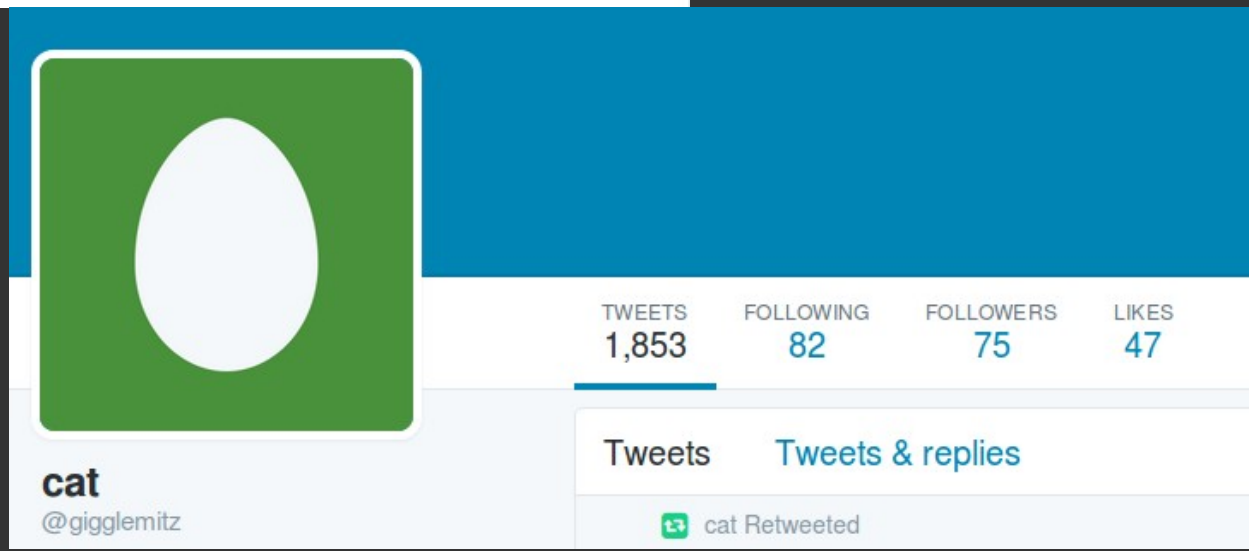
Collected from:

- Tweets
- Speeches
- Debates
- Essays

Cleaning Tweets

- Tweet completion: Longer tweets end in "(cont) some.link.to.twitlonger"
- Discarded re-tweets
- Removed hyperlinks
- Handle acronyms/titles, such as “U.S.” or “Jr.”.
- Added spaces between words and punctuation

Dealing with @usernames



Dealing with @usernames

- Keep only “important” people and organizations
- Replace the @usernames by the name as listed on Twitter.
- Tweets to “unimportant” people of a personal nature – discard.

Model Summary

- Input word or sentence
  - Select a *seed* from sentence
  - Search dictionary
  - If key found, generate sentence
  - Output sentence

What if key not found?

What about words outside Vocabulary?

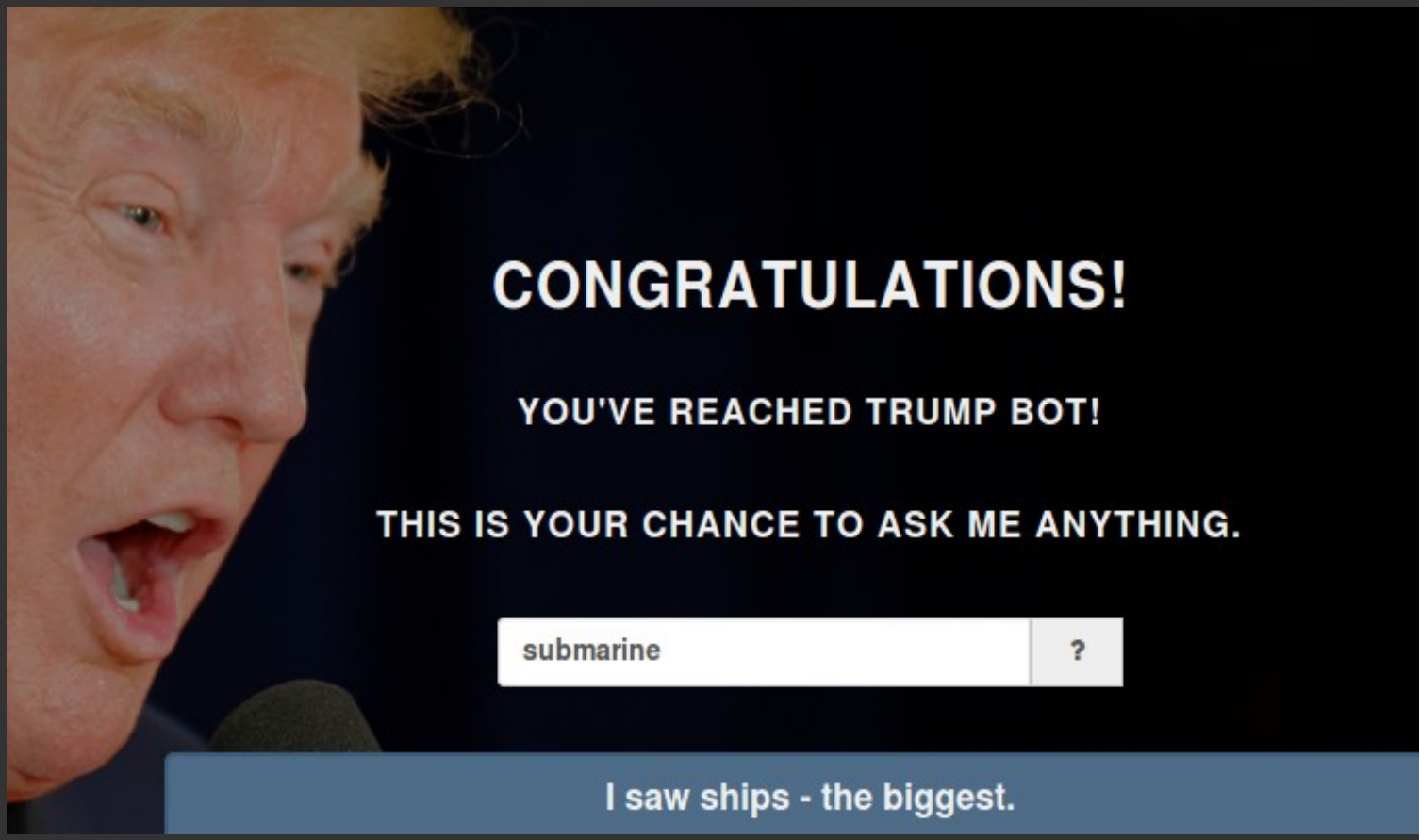
Word vectors!

- 400,000 pre-trained word vectors (GloVe) from Stanford NLP project
- For each word, find ten most similar words (cosine-similarity)
- Use one of the words as seed

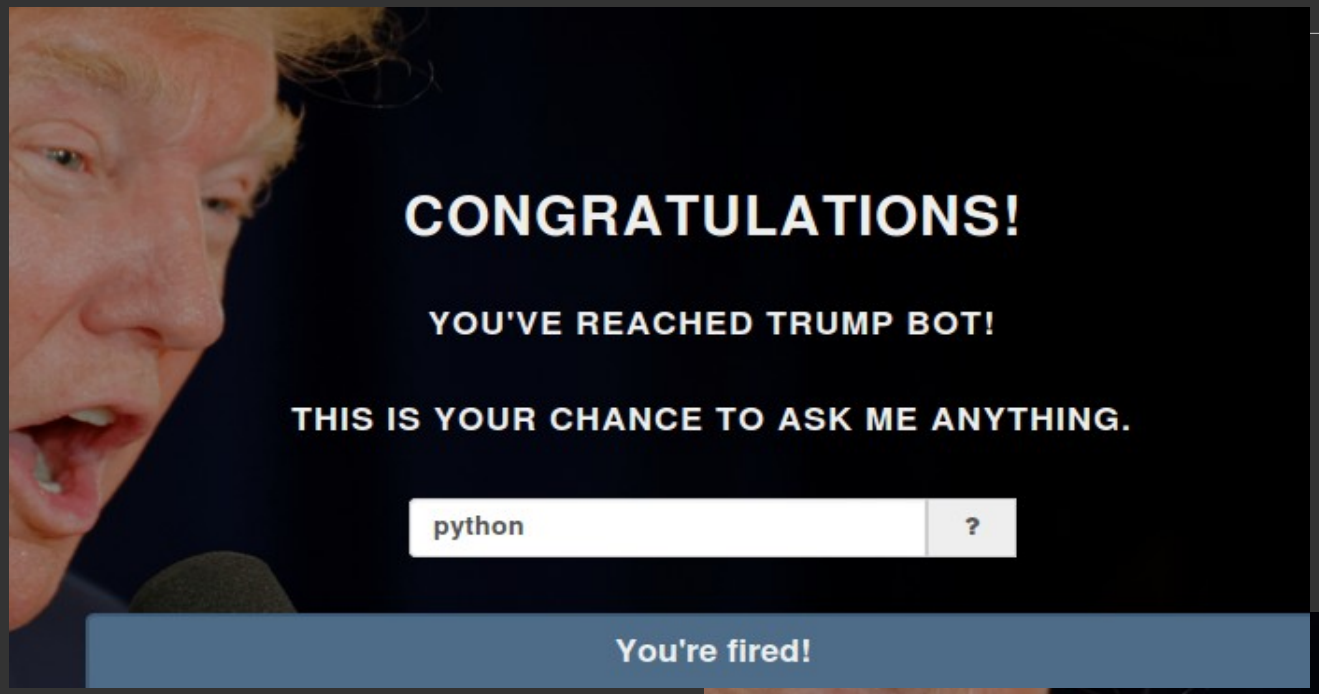
What about words outside Vocabulary?

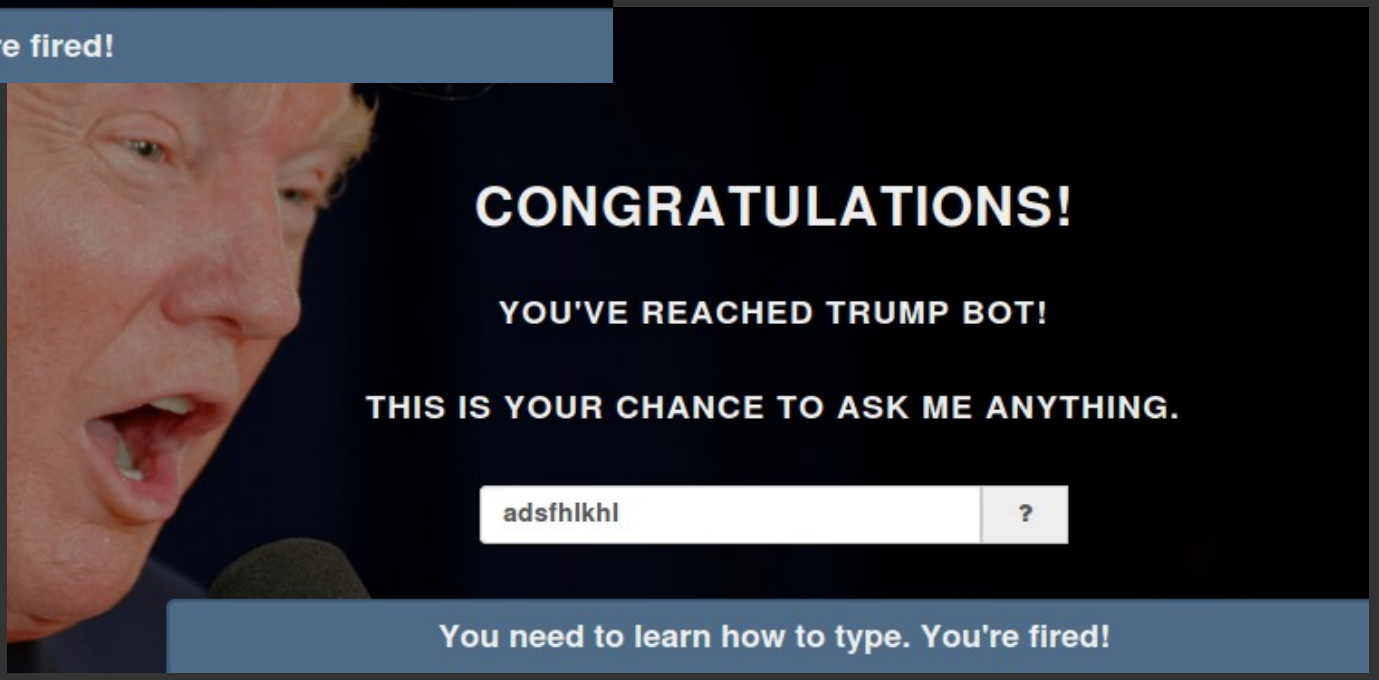
SUBMARINE

submarines
warship
ship
torpedo
destroyer
destroyers
vessel
ships
u-boat
navy



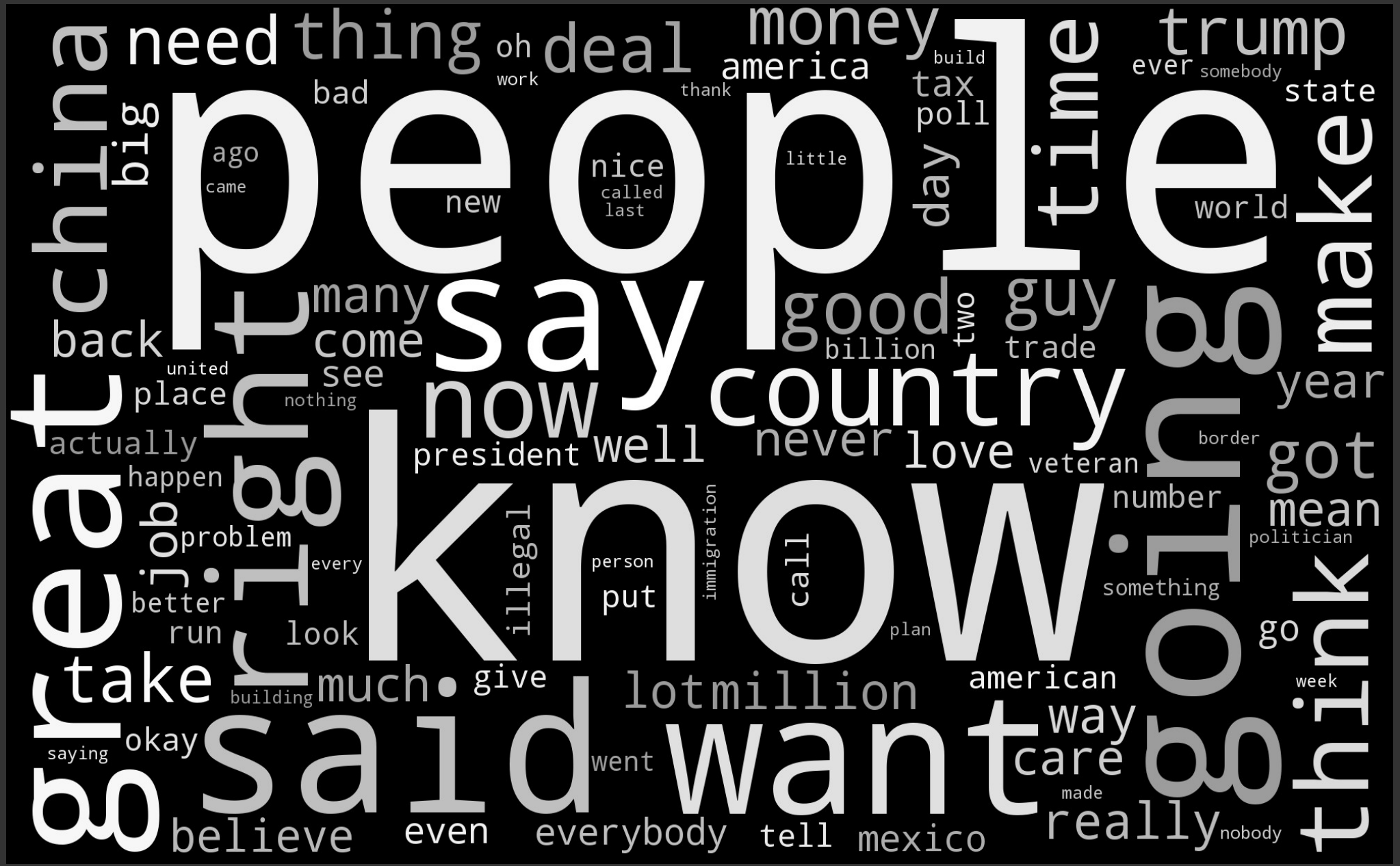
Canned responses





We are what we speak ...

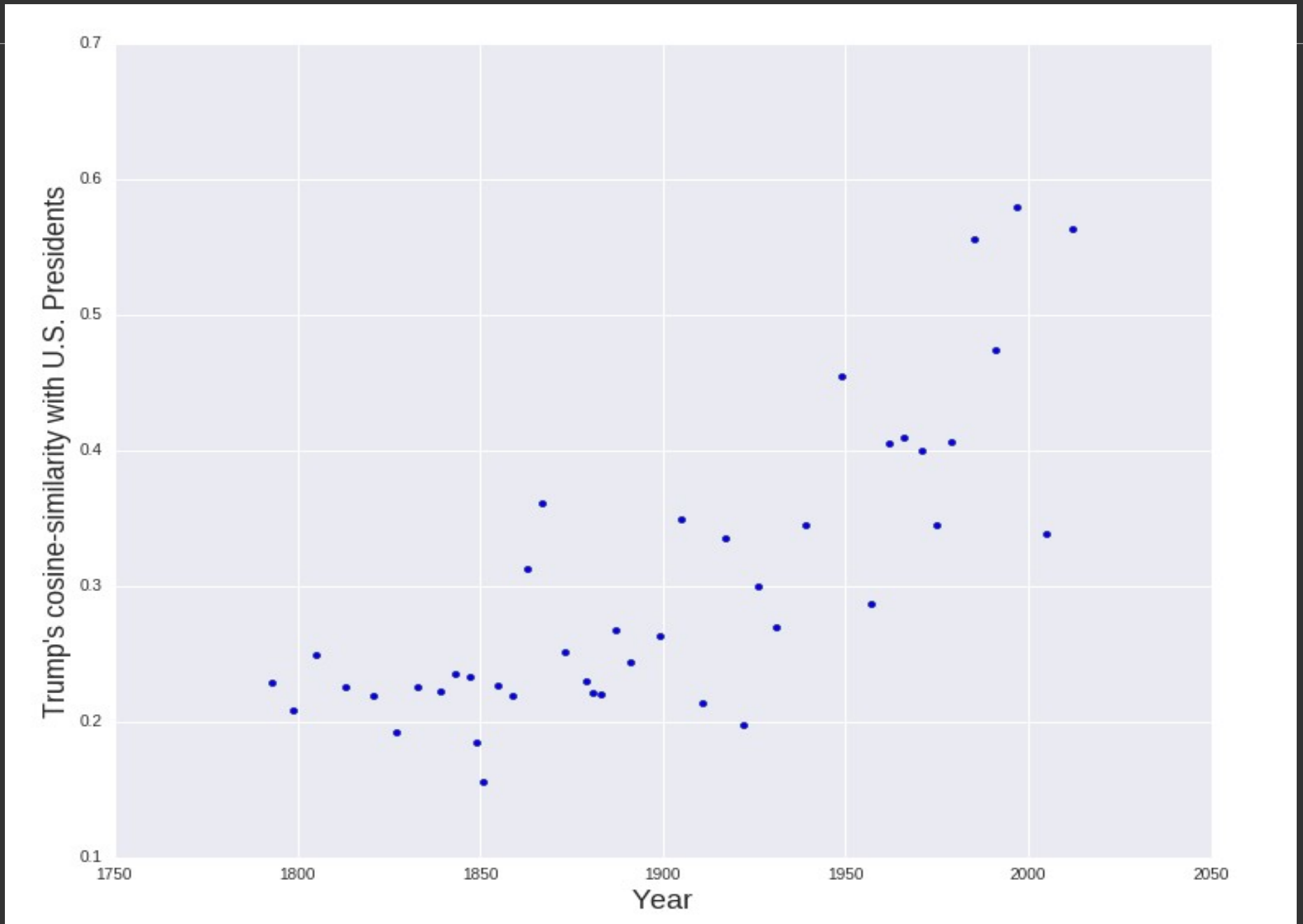
Trump's Favorite Words

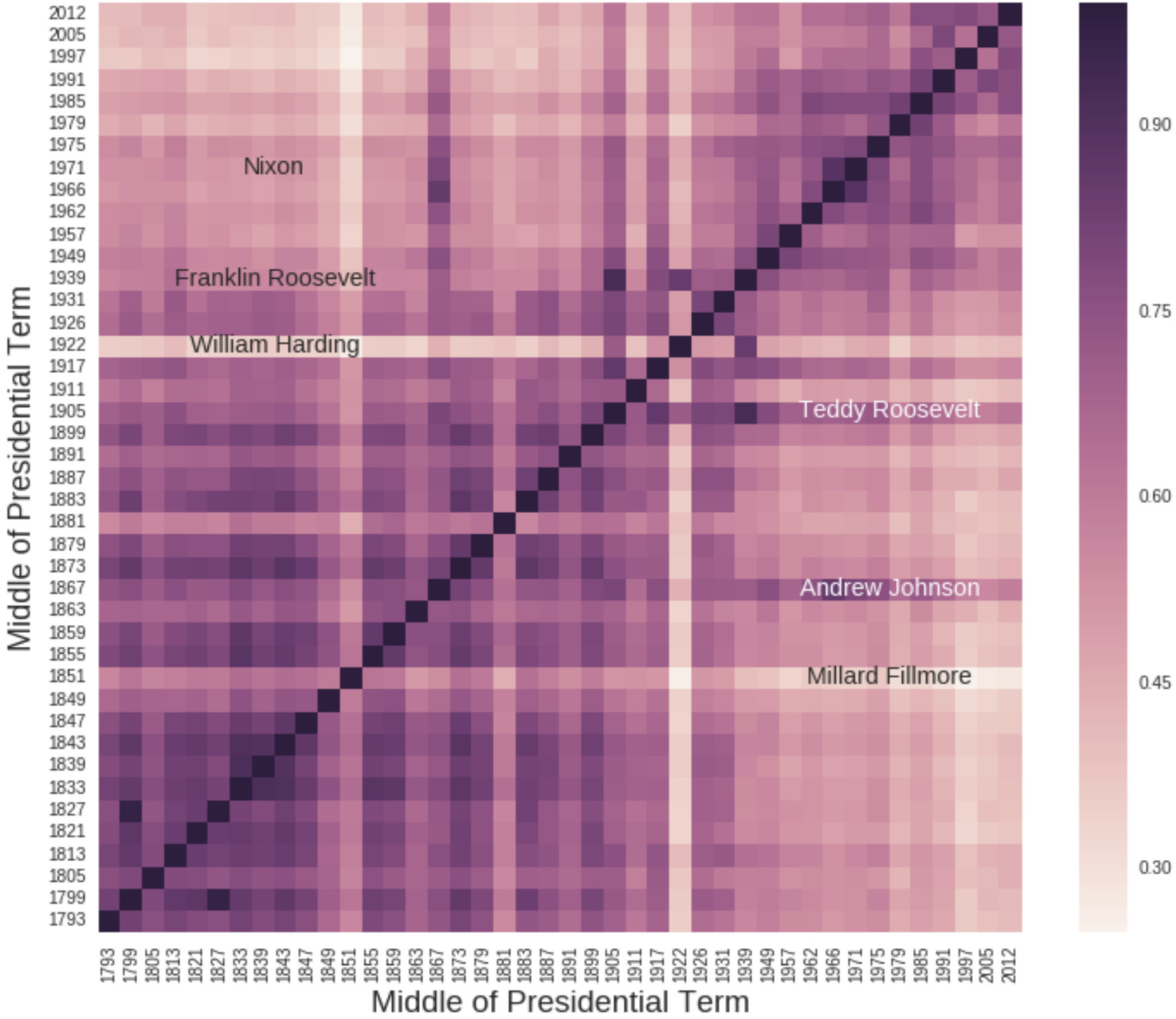


“Average” U.S. President's Favorite Words

How similar or different is he really?

Trump vs Presidents




Middle of Presidential Term
2012
2005
1997
1991
1985
1979
1975
1971
1966
1962
1957
1949
1939
1931
1926
1922
1917
1911
1905
1899
1891
1887
1883
1881
1879
1873
1867
1863
1859
1855
1851
1849
1847
1843
1839
1833
1827
1821
1813
1805
1799
1793
Nixon
Franklin Roosevelt
William Harding
Teddy Roosevelt
Andrew Johnson
Millard Fillmore
1793
1799
1805
1813
1821
1827
1833
1839
1843
1847
1849
1851
1855
1859
1863
1867
1873
1879
1881
1883
1887
1891
1899
1905
1911
1917
1922
1926
1931
1939
1949
1957
1962
1966
1971
1975
1979
1985
1991
1997
2005
2012
Middle of Presidential Term
0.90
0.75
0.60
0.45
0.30

Further Exploration

- Look for semantic similarity
  - document vectors
  - sentiment analysis
- For sentence generation
  - RNNs
  - Hidden Markov Models

Aditi Raval

---

www.twitter.com/surrealTrumpBot

www.whatwouldtrumpsay.elasticbeanstalk.com

www.github.com/araval/trump_bot