Assignment 10

Find your favorite news source and grab the article text.

1. Show the most common words in the article.
2. Show the most common words under a part of speech. (i.e. NOUN: {'Bob':12, 'Alice':4,})
3. Find a subject/object relationship through the dependency parser in any sentence.
4. Show the most common Entities and their types.
5. Find Entites and their dependency (hint: entity.root.head)
6. Find the most similar words (noun chunks) in the article

```
In [1]: import spacy
```

Washington Post Article:

Zuckerberg once wanted to sanction Trump. Then Facebook wrote rules that accommodated him. (https://www.washingtonpost.com/technology/2020/06/28/facebook-zuckerberg-trump-hate/)

```
In [2]: processor = spacy.load('en')
```

```python
text = """ Zuckerberg once wanted to sanction Trump. Then Facebook wrote rules th
Starting as early as 2015, Facebook executives started crafting exceptions for th

Hours after President Trump's incendiary post last month about sending the milita
The post put the company in a difficult position, Zuckerberg told Trump, accordir
To Facebook's executives in Washington, the post didn't appear to violate its pol
Eventually, Trump posted again, saying his comments were supposed to be a warning
The frenzied push-pull was just the latest incident in a five-year struggle by Fa
Bowing to pressure, Facebook will start labeling violating posts from politicians
Facebook has constrained its efforts against false and misleading news, adopted a
The concessions to Trump have led to a transformation of the world's information
And as Trump grew in power, the fear of his wrath pushed Facebook into more defer
Facebook is now confronting a mounting advertiser boycott that has pushed down it
Facebook is also facing a slow-burning crisis of morale, with more than 5,000 emp

Bowing to those pressures on Friday, Zuckerberg announced a rash of new policies
But the company said the post wouldn't have qualified.
As the United States heads into another presidential election while facing a pand
It also places Facebook in growing conflict with its counterparts in Silicon Vall
"The value of being in favor with people in power outweighs almost every other cc
Facebook contends the use of incendiary populist language predates social media.
Nick Clegg, Facebook's vice president of global affairs, said that populism wasn'
"From the Arab Spring to local candidates challenging political incumbents, socia
Facebook declined to make Zuckerberg available for an interview, although it poir
Zuckerberg talks frequently about making choices that stand the test of time, pre
At one point, however, he wanted a different approach to Trump.
Setting the stage
Before the 2016 election, the company largely saw its role in politics as courtir

But that started to change in 2015, as Trump's candidacy picked up speed. In Dece
Outrage over the video led to a companywide town hall, in which employees decried
At one of the meetings, Monika Bickert, Facebook's vice president for policy, dra
The document, which is previously unreported and obtained by The Post, weighed fc
Facebook spokesman Tucker Bounds said the latter option was never seriously consi
The document also listed possible "PR Risks" for each. For example, lowering the
Ultimately, Zuckerberg was talked out of his desire to remove the post in part by
That allowance was not formally written into the policies, even though it informe
In an interview, Bickert said the company ultimately made a call to maintain Trun
Facebook's Bounds added that the "newsworthiness" policy was added in 2016 after
In spring of 2016, Zuckerberg was also talked out of his desire to write a post s
The political speech carveout ended up setting the stage for how the company woul
"Though [Facebook] has cracked down on misinformation, the most problematic influ
Trump presented a unique challenge, she added. "Until then, no one would have cor
Protecting the right
After the election, it became clear Russia had used social media to sow disinforn
At the same time, GOP leaders stepped up criticism that platforms such as Facebo
"There's no credible research supporting Trump's claim that social platforms supp
As Facebook scrambled to tackle foreign interference and misinformation, its exec
Facebook's security engineers in December 2016 presented findings from a broad in
A year later, Facebook considered how to overhaul its scrolling news feed, the hc
In meetings about the change, Kaplan questioned whether the revamped algorithm wc

Isolated and divided
With the 2020 election on the horizon, Facebook and Zuckerberg's hands-off approa
In May 2019, Zuckerberg, citing free speech, refused to take down a doctored vide
That summer, company leaders held a meeting to revisit its newsworthiness excepti
```

```
    Clegg, the company's new head of global affairs and communications and a former E
    Facebook did not want to be an arbiter of truth in political debate, he said, ech
    The speech angered some employees, triggering more than 250 of them to sign a pet
    One former executive, Yael Eisenstat, who worked to improve the political ads pro
    She said that she routinely experienced how the company's efforts at integrity we
    Meanwhile, in October, as Facebook faced more potential regulation and political
    Tweetstorm
    As the pandemic and civil unrest dominated the first half of this year, Trump con
    Meanwhile, Facebook employees began challenging the company's decisions.

    Two months before Trump's "looting, shooting" post, the Brazilian president poste
    Thiel, the security engineer, and other employees argued internally that it viola
    Thiel quit in disgust.
    In May, following years of internal debate of its own, Twitter chose to go in the
    Trump responded two days later with an executive order that could hurt social med
    President Trump, with Attorney General William P. Barr, speaks in the Oval Office
    The next day, Trump tweeted about the Minnesota protests. Twitter quickly labeled
    Facebook, on the other hand, chose to haggle with the White House, asking for a c
    As employees raged on internal message boards and externally on Twitter, Zuckerbe
    In June, Facebook removed a swath of Trump campaign ads with Nazi symbolism, afte
    Last week, the advertiser boycott picked up steam. Hershey, Verizon, Unilever, Co
    On Friday, Zuckerberg told employees in a live-streamed town hall that he was cha

    "There are no exceptions for politicians in any of the policies that I'm announci

    """
```

In [4]:
```
processed_text = processor(text)
processed_text
```

Out[4]:
```
 Zuckerberg once wanted to sanction Trump. Then Facebook wrote rules that acc
ommodated him.
Starting as early as 2015, Facebook executives started crafting exceptions fo
r the then-candidate that transformed the world's information battlefield for
years to come.

Hours after President Trump's incendiary post last month about sending the mi
litary to the Minnesota protests, Trump called Facebook chief executive Mark
Zuckerberg.
The post put the company in a difficult position, Zuckerberg told Trump, acco
rding to people familiar with the discussions. The same message was hidden by
Twitter, the strongest action ever taken against a presidential post.
To Facebook's executives in Washington, the post didn't appear to violate its
policies, which allows leaders to post about government use of force if the m
essage is intended to warn the public — but it came right up to the line. The
deputies had already contacted the White House earlier in the day with an urg
ent plea to tweak the language of the post or simply delete it, the people sa
id.
Eventually, Trump posted again, saying his comments were supposed to be a war
```

1. Show the most common words in the article.

In [5]:
```
import re
from collections import Counter
```

```
In [6]:  words = [token.text for token in processed_text if token.is_stop != True
                and token.is_punct !=True]
         lower_words = [word.lower() for word in words]
```

```
In [7]:  Counter(lower_words).most_common(10)
```

```
Out[7]:  [('\n', 67),
          ('facebook', 45),
          ('trump', 33),
          ('company', 30),
          ('said', 26),
          ('zuckerberg', 22),
          ('post', 22),
          ('political', 17),
          ('people', 16),
          ('speech', 16)]
```

2. Show the most common words under a part of speech. (i.e. NOUN: {'Bob':12, 'Alice':4,})

```
In [8]:  nouns = [token.text for token in processed_text if token.is_stop != True
                and token.is_punct !=True and token.pos_ =='NOUN']
         print(Counter(nouns).most_common(15))
```

```
[('company', 30), ('people', 16), ('speech', 16), ('post', 15), ('media', 12),
('policies', 11), ('leaders', 10), ('president', 10), ('politicians', 10), ('em
ployees', 10), ('policy', 9), ('video', 9), ('content', 8), ('executives', 7),
('news', 7)]
```

```
In [9]:  verbs = [token.text for token in processed_text if token.is_stop != True
                and token.is_punct !=True and token.pos_ =='VERB']
         print(Counter(verbs).most_common(15))
```

```
[('said', 26), ('according', 14), ('posted', 6), ('told', 5), ('wanted', 4),
('went', 4), ('remove', 4), ('leaning', 4), ('added', 4), ('making', 4), ('cons
idered', 4), ('violated', 4), ('wrote', 3), ('violate', 3), ('allowing', 3)]
```

```
In [10]:  adj = [token.text for token in processed_text if token.is_stop != True
                and token.is_punct !=True and token.pos_ =='ADJ']
          print(Counter(adj).most_common(15))
```

```
[('political', 17), ('social', 13), ('familiar', 6), ('internal', 6), ('new',
4), ('senior', 4), ('incendiary', 3), ('presidential', 3), ('false', 3), ('misl
eading', 3), ('conservative', 3), ('civil', 3), ('dehumanizing', 3), ('global',
3), ('newsworthy', 3)]
```

```
In [11]: adv = [token.text for token in processed_text if token.is_stop != True
             and token.is_punct !=True and token.pos_ =='ADV']
         print(Counter(adv).most_common(15))

[('far', 4), ('previously', 3), ('right', 2), ('frequently', 2), ('largely',
2), ('Ultimately', 2), ('ultimately', 2), ('later', 2), ('increasingly', 2),
('ahead', 2), ('early', 1), ('earlier', 1), ('simply', 1), ('Eventually', 1),
('online', 1)]
```

3. Find a subject/object relationship through the dependency parser in any sentence.

```
In [12]: def pr_tree(word, level):
             if word.is_punct:
                 return
             for child in word.lefts:
                 pr_tree(child, level+1)
             print('\t'* level + word.text + ' - ' + word.dep_)
             for child in word.rights:
                 pr_tree(child, level+1)
```

```
In [13]: for sentence in processed_text.sents:
             pr_tree(sentence.root, 0)
             print('-------------------------------------------')
```

```
                    -
            Zuckerberg - nsubj
            once - advmod
    wanted - ROOT
                    to - aux
            sanction - xcomp
                    Trump - dobj
    -------------------------------------------
            Then - advmod
            Facebook - nsubj
    wrote - ROOT
            rules - dobj
                        that - nsubj
                accommodated - relcl
                        him - dobj
    -------------------------------------------
            Starting - advcl
                        as - advmod
                early - advmod
```

4. Show the most common Entities and their types.

```
In [14]:  counts = Counter()
          for entity in processed_text.ents:
              counts[entity.text.upper(), entity.label_] += 1
          counts.most_common(15)
```

```
Out[14]:  [(('FACEBOOK', 'ORG'), 39),
           (('ZUCKERBERG', 'PERSON'), 21),
           (('TRUMP', 'PERSON'), 13),
           (('TRUMP', 'ORG'), 13),
           (('WASHINGTON', 'GPE'), 5),
           (('ONE', 'CARDINAL'), 5),
           (('THE UNITED STATES', 'GPE'), 4),
           (('SILICON VALLEY', 'LOC'), 4),
           (('FACEBOOK', 'GPE'), 4),
           (('THE WHITE HOUSE', 'ORG'), 3),
           (('MUSLIMS', 'NORP'), 3),
           (('THREE', 'CARDINAL'), 3),
           (('KAPLAN', 'PERSON'), 3),
           (('THIEL', 'PERSON'), 3),
           (('POST', 'ORG'), 3)]
```

5. Find Entites and their dependency (hint: entity.root.head)

```
In [15]:  for entity in processed_text.ents:
              print(entity, entity.root.head)
```

```
Zuckerberg wanted
Trump sanction
Facebook wrote
as early as 2015 Starting
Facebook executives
years for
Trump post
last month post
Minnesota protests
Trump called
Facebook executive
Mark Zuckerberg called
Zuckerberg told
Trump told
Facebook To
Washington in
the White House contacted
earlier in the day contacted
Trump posted
```

6. Find the most similar words (noun chunks) in the article

```
In [16]: n=0
         for sentence in processed_text.sents:
             for noun_chunk in processed_text.noun_chunks:
                 print(n, noun_chunk, processed_text.similarity(noun_chunk))
                 n+=1
```

```
0  Zuckerberg 0.2961844530245486
1 Trump 0.31444014219688976
2 Facebook 0.29626347871979575
3 rules 0.11625004209008458
4 him 0.20298750481510144
5 Facebook executives 0.43100492450074884
6 exceptions 0.20333825537131975
7 the then-candidate 0.5612437630600484
8 the world 0.5136187624168242
9 information 0.43784747765295845
10 years 0.34093053476086066
11 President Trump's incendiary post 0.5119528730824511
12 the military 0.6631335229433752
13 the Minnesota protests 0.4743665555013331
14 Trump 0.4320341481329247
15 The post 0.5201830020290689
16 the company 0.48811568794405713
17 a difficult position 0.5857734633745992
18 Zuckerberg 0.2830270847416785
```