

Lab 4 report

Lab4_1

Did not understand what was meant with top 20 results so we just printed the first 20

As done in the nltk book.

What we did was make a nltk regexparser that tags the format a verb followed by a sequence of noun phrases and prepositions. NP_PP

```
The/AT
Fulton/NP-TL
County/NN-TL
Grand/JJ-TL
Jury/NN-TL
(NP_PP said/VBD Friday/NR)
an/AT
investigation/NN
of/IN
Atlanta's/NP$
recent/JJ
primary/NN
election/NN
produced/VBD
```

Lab4_2

Our predict function looks at prior tweets and at 3grams. And based on what has been said before predicts the next word. I takes in the trigram to start with and predict based on and the max length of the returning string.

Lab4_3

```
['What happened to the Tesla that Elon Musk shot into space?\n', "There we  
re plenty of spectacular moments during the maiden launch of SpaceX's Falc  
on Heavy rocket on Tuesday.\n", '\n', 'But perhaps the most dramatic scene  
occurred about four minutes after liftoff: The second stage of the rocket  
, headed deeper into space, discarded the white nose cone at its tip.\n',  
'\n']
```

```
occurred/VBD
about/IN
four/CD
(NP minutes/NNS)
after/IN
liftoff/NN
```

Task three takes the first 5 sentences in the text to make a chunker, then we use the output to run on the rest of the text. We did not understand the difference between task A and B. the result is expected.

Lab4_4

In a i assumed that the acceptance rate is the number of accepted submissions from the poster presentations

Spotlight presentations and the oral presentations combined, divided by the withdrawn/rejected submissions.

By keywords i looked at the text in the title of each submission except for the rejected ones.

And the same with the authors with the most amount of papers.

```
the amount of accepted requests is: 687
The amount of rejected requests is: 1907
The rate of accepted to rejected is accepted /rejected = 0.3602517042475092
The 10 authors with the most number of papers: [('Sergey Levine', 13), ('Jun Zhu', 7), ('Le Song', 7), ('Cho-Jui Hsieh', 6), ('Liwei Wang', 6), ('Jimmy Ba', 6), ('Pushmeet Kohli', 6), ('Tom Goldstein', 6), ('Abhinav Gupta', 6), ('Sanjiv Kumar', 5)]
The 10 most common words with stopwords: [('for', 220), ('learning', 198), (':', 198), ('of', 148), ('neural', 135), ('networks', 116), ('with', 111), ('and', 104), ('deep', 75), ('the', 75)]

The 10 most common words without stopwords: [('learning', 198), ('neural', 135), ('networks', 116), ('deep', 75), ('a', 52), ('reinforcement', 44), ('training', 43), ('adversarial', 43), ('via', 42), ('graph', 40)]
```