Combining the power of Python with R

Contents

R + Py

This is a simple (could be called, `naive` as well) attempt to show how we can combine the power of `Python` with `R` and create a new superpower.

Like this one, If you have watched **The Incredibles** before!

The R Code

```
#loading required R libraries


library(tidyverse)
tweets <- read_csv("justdoit_tweets_2018_09_07_2.csv")
text <- tweets$tweet_full_text
text_10 <- text[1:10]
```

The Python Code

```
import spacy
import pandas as pd
nlp = spacy.load('en_core_web_sm')
doc = nlp(str(r.text_10))
pos_df = pd.DataFrame(columns = ["text","pos","lemma"])
print(pos_df)
```

```
## Empty DataFrame
## Columns: [text, pos, lemma]
## Index: []
```

Figure 1: Jack-Jack Parr

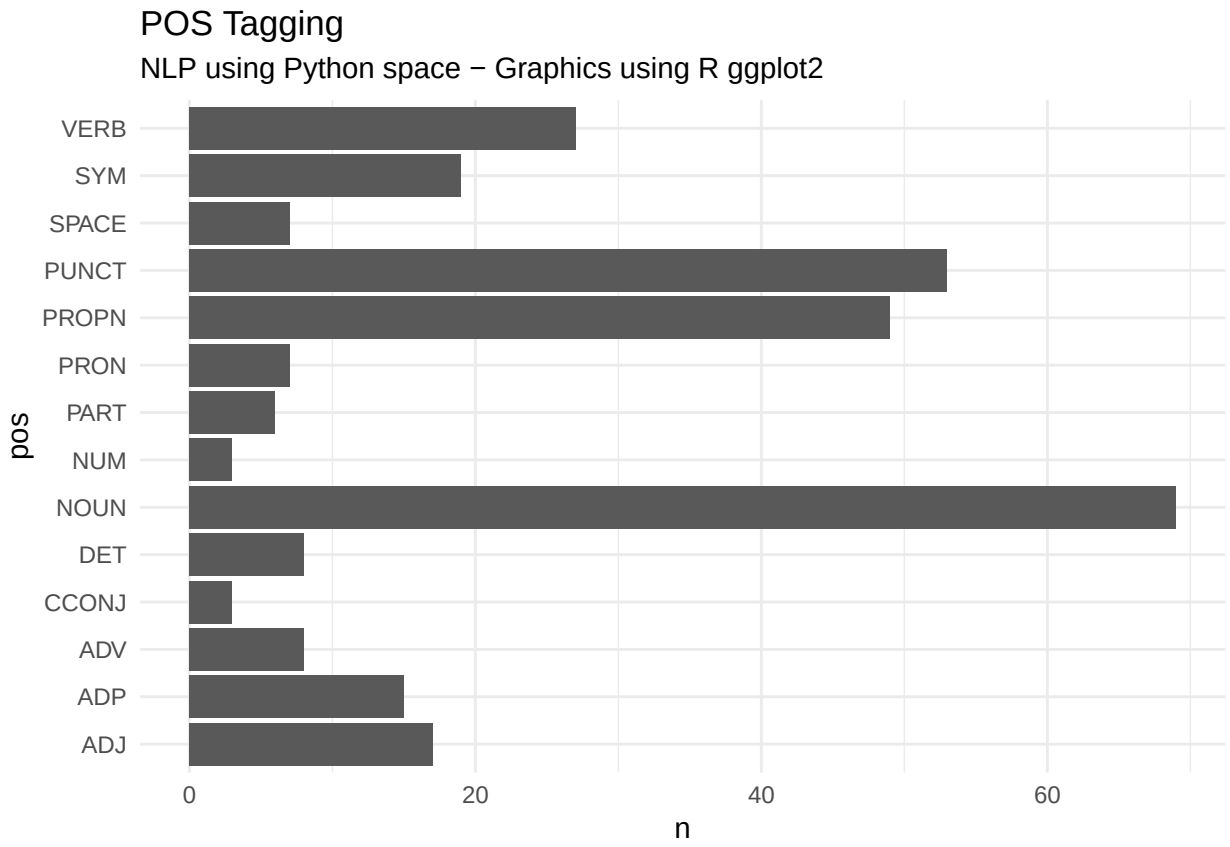
```
for token in doc:
    df1 = pd.DataFrame({"text" : token.text, "pos" : token.pos_, "lemma" : token.lemma_}, index = [0])
    #print(token.text, token.pos_)
    #print(df1)
    pos_df = pd.concat([pos_df,df1])
#print(pos_df)
```

Now, Again the R Code

```
#data.frame(token = as.vector(py$tokens)) %>% count(token) %>% arrange(desc(n))

py$pos_df %>%
  count(pos) %>%
  ggplot() + geom_bar(aes(pos,n), stat = "identity") +
  coord_flip() +
  theme_minimal() +
  labs(title = "POS Tagging",
       subtitle = "NLP using Python space - Graphics using R ggplot2")
```

```
## Warning in py_to_r.pandas.core.frame.DataFrame(x): index contains
## duplicated values: row names not set
```


POS Tagging
NLP using Python space − Graphics using R ggplot2
VERB
SYM
SPACE
PUNCT
PROPN
PRON
PART
NUM
NOUN
DET
CCONJ
ADV
ADP
ADJ
pos
0
20
40
60
n