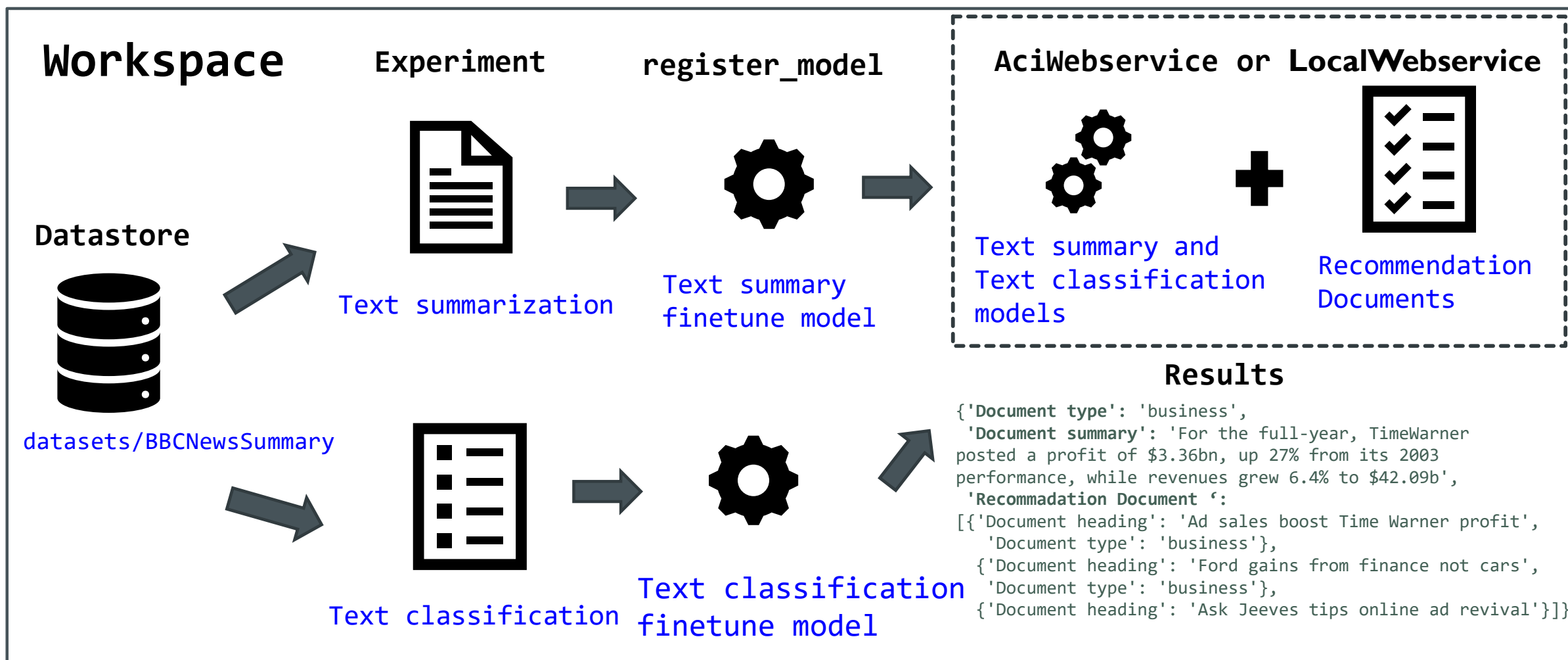




ACCENTURE:AI PROFESSOR

Aunchana pimpisal
16 May 2022

ACCENTURE:AI PROFESSOR



TEXT CLASSIFICATION MODEL

BBCNewsSummary

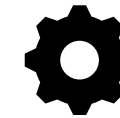


News_Articles

- business
- entertainment
- politics
- sport
- Tech

Document per line

Docum ent no.	Line no.	sentence	group_ document
1	1	Ad sales boost Time Warner profit	business
1	2	Quarterly profits at US media giant TimeWarner jumped 76% to \$1.13bn (£600m) for the three months to December, from \$639m year-earlier.	business
1	3	The firm, which is now one of the biggest investors in Google, benefited from sales of high-speed internet connections and higher advert sales. TimeWarner said fourth quarter sales rose 2% to \$11.1bn from \$10.9bn. Its profits were buoyed by one-off gains which offset a profit dip at Warner Bros, and less users for AOL.	business



Training model

- Tokenizer: BertTokenizerFast
- Model: BertForSequenceClassification
- Model name: bert-base-uncased



Prediction

- Predicted by line of documents.
- Voted the maximum prediction.

Document no.1

Original:[line no.1,line no.2,lineno.3]

Prediction:[business, business, tech]

Result prediction : business

TEXT SUMMARIZATION

BBCNewsSummary



News Articles

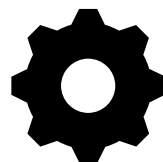
- business
- entertainment
- politics
- sport
- Tech

Summaries

- business
- entertainment
- politics
- sport
- Tech



Training model



- Tokenizer:AutoTokenizer
- Model:AutoModelForSeq2SeqLM
- Model name:t5-small
- Max input length :1024
- Max target length :300
- Metric :rouge



Prediction

`Document=''Ad sales boost Time Warner profit
Quarterly profits at US media giant TimeWarner jump
ed 76% to $1.13bn (£600m) for the three months to D
ecember, from $639m year-earlier.
The firm, which is now one of the biggest investors
in Google, benefited from sales of high-
speed internet connections and higher advert sales.
TimeWarner said fourth quarter sales rose 2% to $1
1.1bn from $10.9bn. Its profits were buoyed by one-
off gains which offset a profit dip at Warner Bros,
and less users for AOL.'''`

`'Document summary': 'Time Warner profitQuarterly
profits at US media giant TimeWarner jumped 76% to
$1.13bn (£600m) for the three months to December,
from $639m year-earlier.The firm, which is now one
of the biggest investors in Google, benefited from
sales of high-speed internet connections and higher
advert sales.Its profits were buoyed by one-off
gains which offset a profit dip at Warner Bros, and
less users for AOL.'`

RECOMMEDATION DOCUMENT

BBCNewsSummary



News_Articles

- business
- entertainment
- politics
- sport
- Tech



Text cleasing

- String lower
- Delete '\n', '\t' and '\s'
- Keep a-z and number

Tokenizer and pos tag selection

Tag	Description
CC	Coordinating conjunction
EX	Existential there
FW	Foreign word
NN	Noun, singular or mass
NNS	Noun, plural
NNP	Proper noun, singular
NNPS	Proper noun, plural
POS	Possessive ending
PRP	Personal pronoun
PRP\$	Possessive pronoun
RP	Particle
WDT	Wh-determiner
WRB	Wh-adverb

TFIDF of documents

dollar	TimeWarn	better	year	government
0.953704	0.225195	0	0.951664	0
0	0	0.42505	0.378727	0.668075305
0.449311	0	0.182369	0	0
0	0	0	0.711764	0.570577102
0.941265	0	0.072782	0	0
0.360879	0	0.498432	0.634757	0.627150549

Similarity
(sentence, documents)

sentence

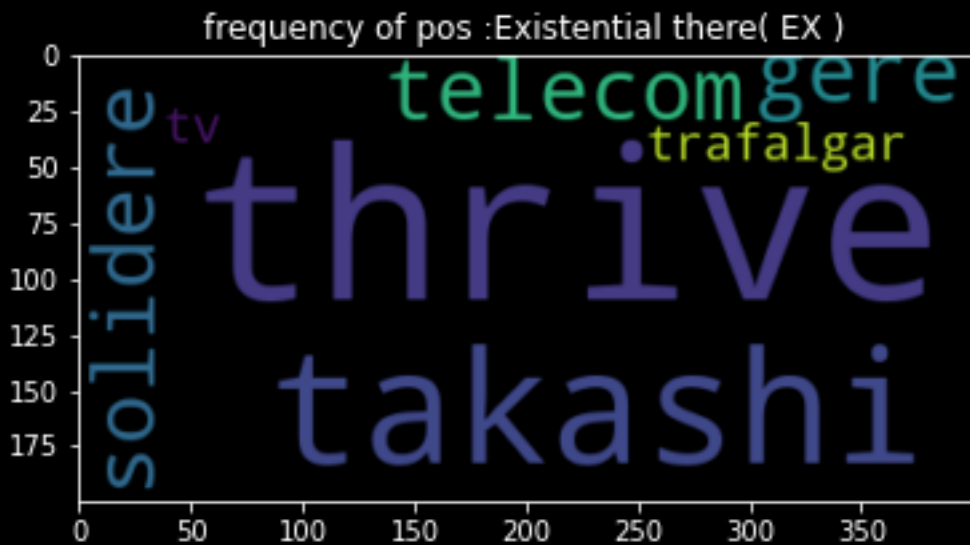
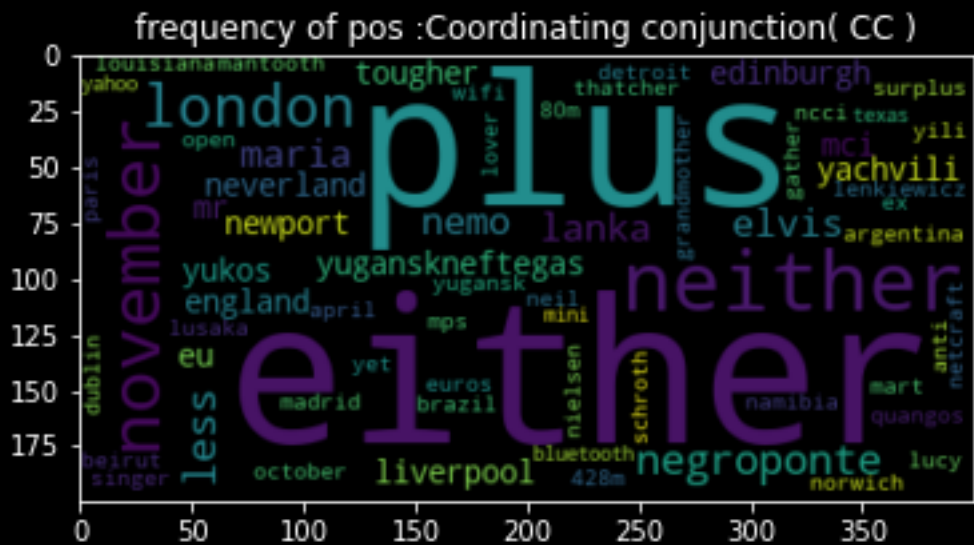
'Ad sales boost Time Warner profit'

Quarterly profits at US media giant Time Warner'

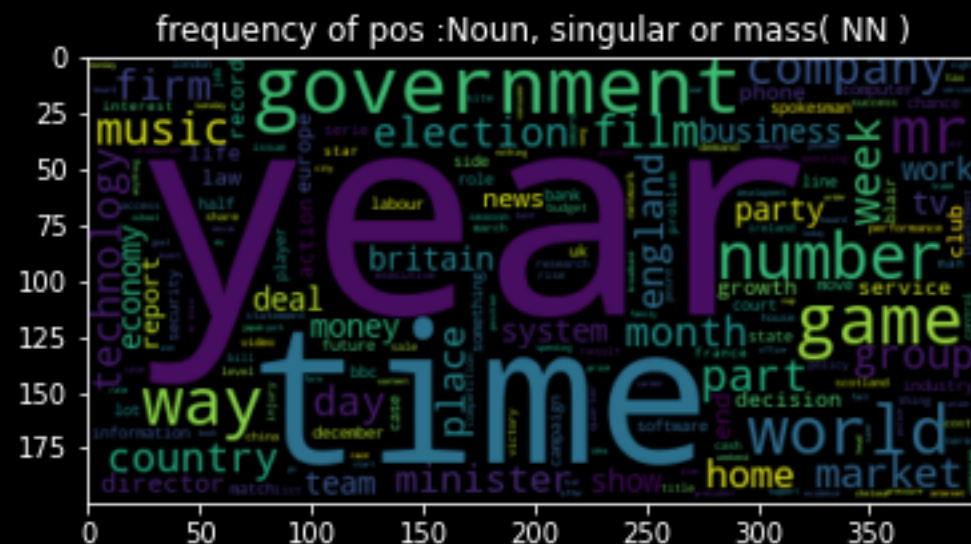
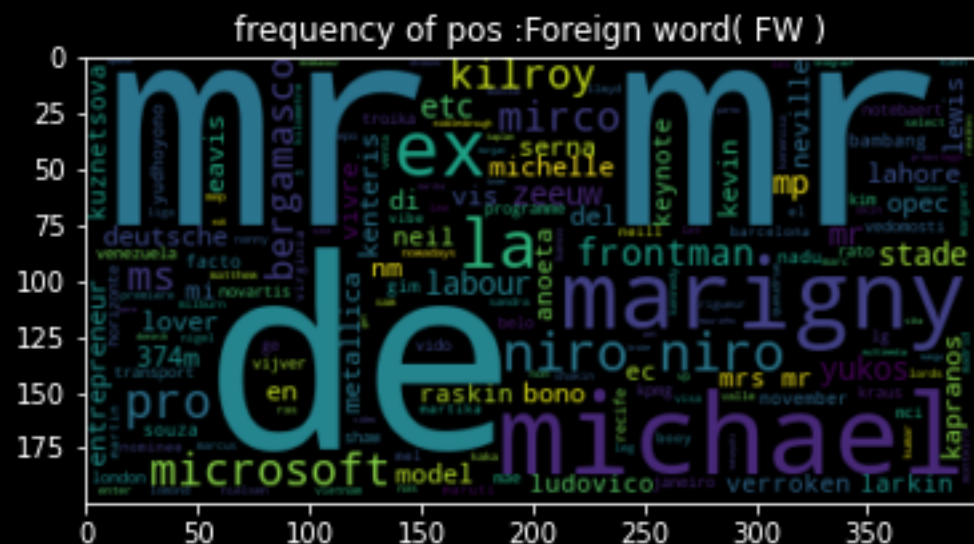
Recommendation documents

```
[{'Document heading':  
'Ad sales boost Time  
Warner profit',  
  'Document type':  
  'business'},  
 {'Document heading':  
'Ford gains from  
finance not cars',  
  'Document type':  
  'business'},  
 {'Document heading':  
'Ask Jeeves tips online  
ad revival'}]
```

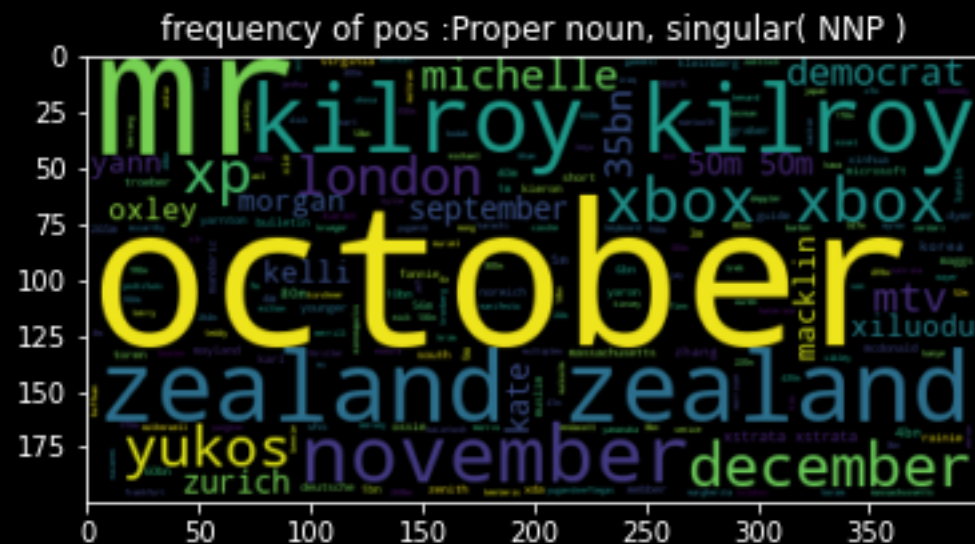
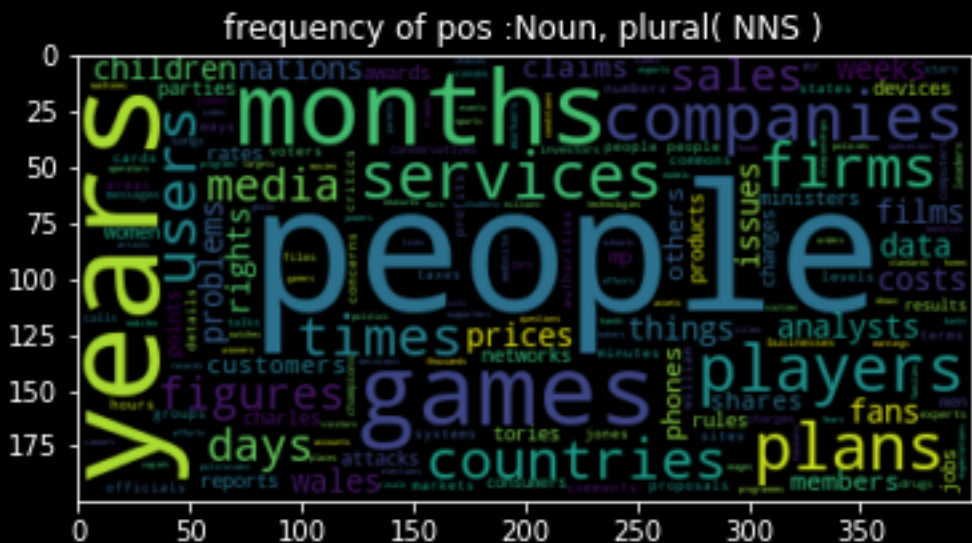
RECOMMENDATION DOCUMENT



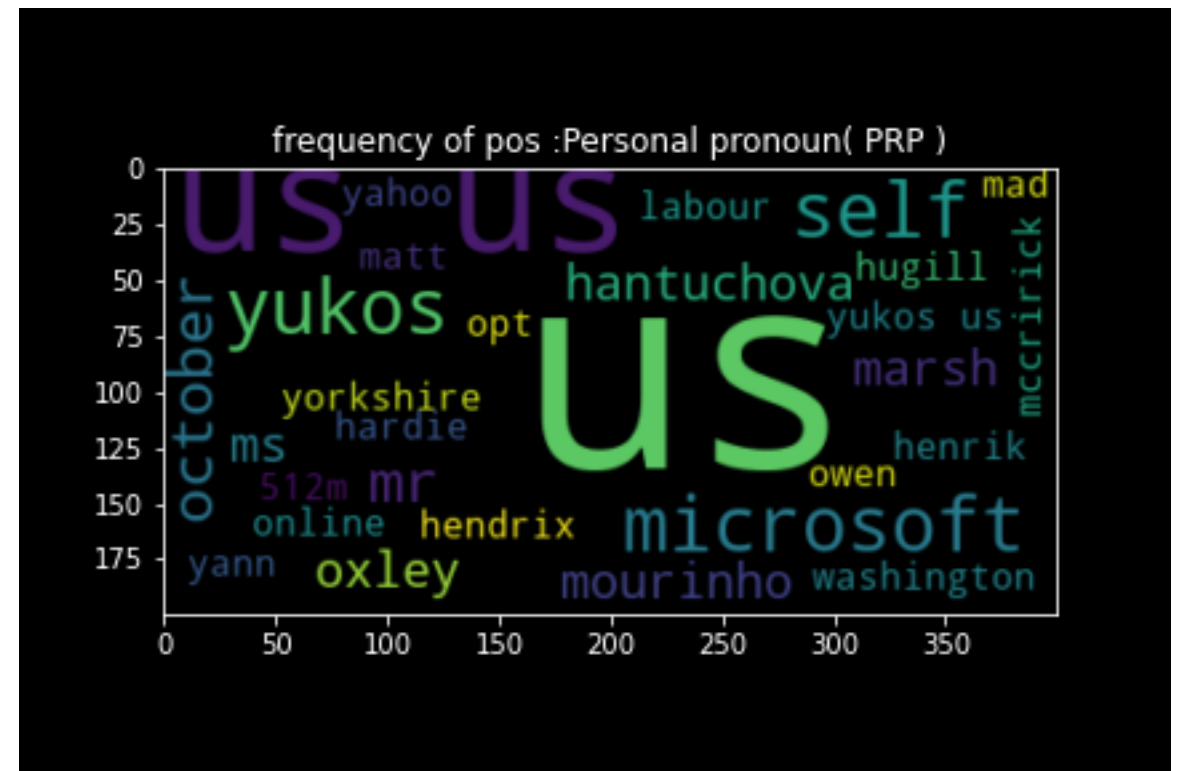
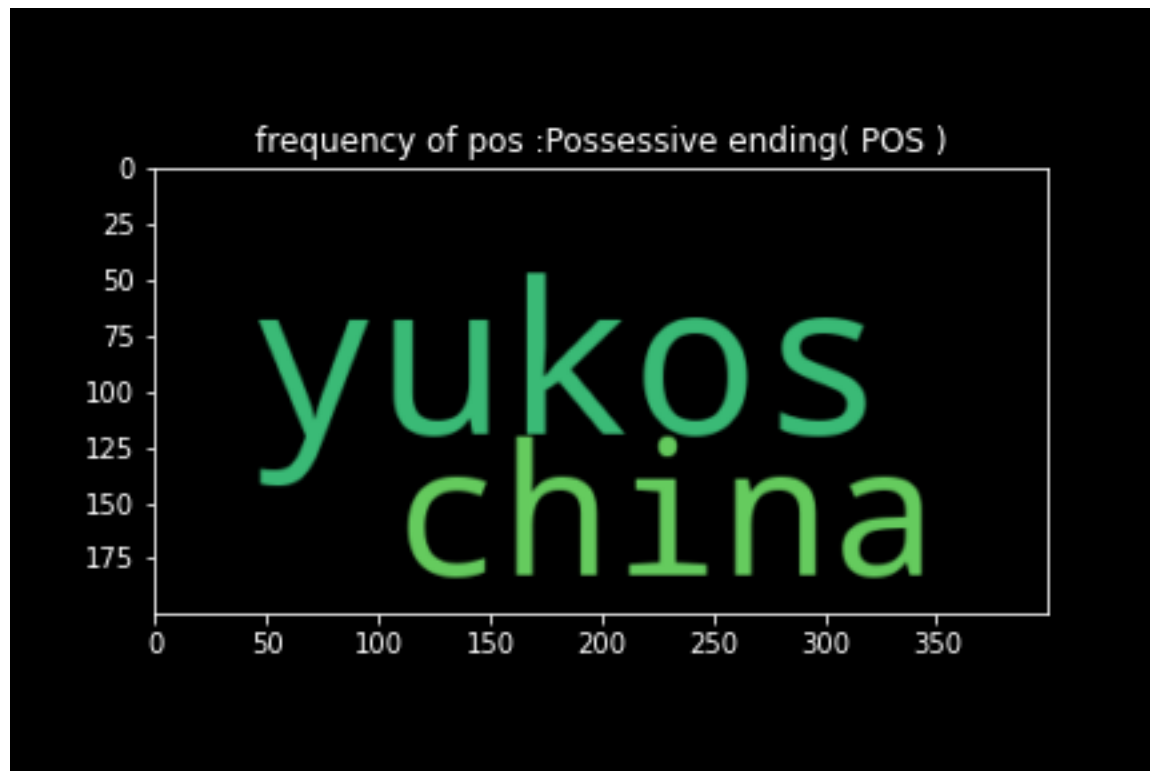
RECOMMENDATION DOCUMENT



RECOMMEDATION DOCUMENT



RECOMMEDATION DOCUMENT



RECOMMEDATION DOCUMENT

