NER

March 11, 2023

[33]:
```
import pandas as pd

pd.set_option('display.max_rows', 100)
pd.set_option('display.max_columns', None)
pd.set_option('display.max_colwidth', 500)
```

[34]:
```
news_df = pd.read_parquet("news_v3.parquet")
news_df.head(2)
```

[34]:
```
   index  \
0      0
1      1

                                                                    url
\
0  http://auckland.scoop.co.nz/2020/01/aut-boosts-ai-expertise-with-new-ailab/
1                         http://en.people.cn/n3/2021/0318/c90000-9830122.html

        date language  \
0 2020-01-28       en
1 2021-03-18       en

           title  \
0                                  auckland.scoop.co.nz » AUT boosts AI
expertise with new AiLab
1  Artificial intelligence improves parking efficiency in Chinese cities -
People's Daily Online

                  text  \
0  \n\nauckland.scoop.co.nz » AUT boosts AI expertise with new AiLab\nScoop
Search\nContact\nNewsagent\nLogin\n \n
\nSupercity\nBusiness\nEducation\nEntertainment\nHealth\nPolice\nPolitics\n\n\n
\n\nTweet\nAUT boosts AI expertise with new AiLab\n\n\r\n       January 28,
2020Education, PressRelease0 comments  \n\nPress Release - Auckland University
of Technology\nThere is no question that artificial intelligence’s influence on
our daily life will only increase. AUT’s Faculty of Design and Creati…
1  \n\nArtificial intelligence improves parking efficiency in Chinese cities -
```

```
People's Daily Online\n\nHome\nChina Politics\nForeign Affairs\nOpinions\nVideo:
We Are China\nBusiness\nMilitary\nWorld\nSociety\nCulture\nTravel\nScience\nSpor
ts\nPhoto\n\nLanguages\n\nChinese\nJapanese\nFrench\nSpanish\nRussian\nArabic\nK
orean\nGerman\nPortuguese\nThursday, March 18, 2021\nHome>>\n\t\t\nArtificial
intelligence improves parking efficiency in Chinese cities\nBy Liu Shiyao
(People's Daily) 09:16, Mar…

                                                text_cleaned  \
0  aucklandscoopconz aut boost ai expertise new ailab scoop search contact
newsagent login supercity business education entertainment health police
politics tweet aut boost ai expertise new ailab january 2020education
pressrelease0 comment press release auckland university technology question
artificial intelligence influence daily life increase auts faculty design
creative technology partnered dr john flackett create artificial intelligence
lab ailab focused developing solution everyday life b…
1  artificial intelligence improves parking efficiency chinese city people daily
online home china politics foreign affair opinion video china business military
world society culture travel science sport photo language chinese japanese
french spanish russian arabic korean german portuguese thursday march home
artificial intelligence improves parking efficiency chinese city liushiyao
people daily march photo taken july show sign electronic toll collection etc
newly set roadside parking space yan…

   topic Sentiment
0      3  Positive
1      4  Negative
```

[35]:
```
import nltk
```

[36]:
```
nltk.download('punkt')
```

```
[nltk_data] Downloading package punkt to /home/jupyter/nltk_data…
[nltk_data]   Package punkt is already up-to-date!
```

[36]:
```
True
```

[37]:
```
nltk.download('averaged_perceptron_tagger')
```

```
[nltk_data] Downloading package averaged_perceptron_tagger to
[nltk_data]     /home/jupyter/nltk_data…
[nltk_data]   Package averaged_perceptron_tagger is already up-to-
[nltk_data]       date!
```

[37]:
```
True
```

[38]:
```
nltk.download('maxent_ne_chunker')
```

```
[nltk_data] Downloading package maxent_ne_chunker to
```

```
[nltk_data]     /home/jupyter/nltk_data…
[nltk_data]   Package maxent_ne_chunker is already up-to-date!
```

[38]:
```
True
```

[39]:
```
nltk.download('words')
```

```
[nltk_data] Downloading package words to /home/jupyter/nltk_data…
[nltk_data]   Package words is already up-to-date!
```

[39]:
```
True
```

Using Spacy

[11]:
```
import spacy
from spacy import displacy
```

```
2023-03-10 06:22:46.644445: I tensorflow/core/platform/cpu_feature_guard.cc:193]
This TensorFlow binary is optimized with oneAPI Deep Neural Network Library
(oneDNN) to use the following CPU instructions in performance-critical
operations:  AVX2 FMA
To enable them in other operations, rebuild TensorFlow with the appropriate
compiler flags.
2023-03-10 06:22:51.244643: W
tensorflow/compiler/xla/stream_executor/platform/default/dso_loader.cc:64] Could
not load dynamic library 'libcudart.so.11.0'; dlerror: libcudart.so.11.0: cannot
open shared object file: No such file or directory
2023-03-10 06:22:51.244705: I
tensorflow/compiler/xla/stream_executor/cuda/cudart_stub.cc:29] Ignore above
cudart dlerror if you do not have a GPU set up on your machine.
2023-03-10 06:23:00.248578: W
tensorflow/compiler/xla/stream_executor/platform/default/dso_loader.cc:64] Could
not load dynamic library 'libnvinfer.so.7'; dlerror: libnvinfer.so.7: cannot
open shared object file: No such file or directory
2023-03-10 06:23:00.250806: W
tensorflow/compiler/xla/stream_executor/platform/default/dso_loader.cc:64] Could
not load dynamic library 'libnvinfer_plugin.so.7'; dlerror:
libnvinfer_plugin.so.7: cannot open shared object file: No such file or
directory
2023-03-10 06:23:00.250833: W
tensorflow/compiler/tf2tensorrt/utils/py_utils.cc:38] TF-TRT Warning: Cannot
dlopen some TensorRT libraries. If you would like to use Nvidia GPU with
TensorRT, please make sure the missing libraries mentioned above are installed
properly.
2023-03-10 06:23:05.848576: W
tensorflow/compiler/xla/stream_executor/platform/default/dso_loader.cc:64] Could
not load dynamic library 'libcuda.so.1'; dlerror: libcuda.so.1: cannot open
shared object file: No such file or directory
```

```
2023-03-10 06:23:05.865707: W
tensorflow/compiler/xla/stream_executor/cuda/cuda_driver.cc:265] failed call to
cuInit: UNKNOWN ERROR (303)
2023-03-10 06:23:05.865794: I
tensorflow/compiler/xla/stream_executor/cuda/cuda_diagnostics.cc:156] kernel
driver does not appear to be running on this host (python-20230307-192621):
/proc/driver/nvidia/version does not exist
```

[12]:
```
!python -m spacy download en_core_web_lg
```

```
2023-03-10 06:23:08.822715: I tensorflow/core/platform/cpu_feature_guard.cc:193]
This TensorFlow binary is optimized with oneAPI Deep Neural Network Library
(oneDNN) to use the following CPU instructions in performance-critical
operations:  AVX2 FMA
To enable them in other operations, rebuild TensorFlow with the appropriate
compiler flags.
2023-03-10 06:23:09.034878: W
tensorflow/compiler/xla/stream_executor/platform/default/dso_loader.cc:64] Could
not load dynamic library 'libcudart.so.11.0'; dlerror: libcudart.so.11.0: cannot
open shared object file: No such file or directory
2023-03-10 06:23:09.034941: I
tensorflow/compiler/xla/stream_executor/cuda/cudart_stub.cc:29] Ignore above
cudart dlerror if you do not have a GPU set up on your machine.
2023-03-10 06:23:10.094235: W
tensorflow/compiler/xla/stream_executor/platform/default/dso_loader.cc:64] Could
not load dynamic library 'libnvinfer.so.7'; dlerror: libnvinfer.so.7: cannot
open shared object file: No such file or directory
2023-03-10 06:23:10.094357: W
tensorflow/compiler/xla/stream_executor/platform/default/dso_loader.cc:64] Could
not load dynamic library 'libnvinfer_plugin.so.7'; dlerror:
libnvinfer_plugin.so.7: cannot open shared object file: No such file or
directory
2023-03-10 06:23:10.094381: W
tensorflow/compiler/tf2tensorrt/utils/py_utils.cc:38] TF-TRT Warning: Cannot
dlopen some TensorRT libraries. If you would like to use Nvidia GPU with
TensorRT, please make sure the missing libraries mentioned above are installed
properly.
2023-03-10 06:23:11.469961: W
tensorflow/compiler/xla/stream_executor/platform/default/dso_loader.cc:64] Could
not load dynamic library 'libcuda.so.1'; dlerror: libcuda.so.1: cannot open
shared object file: No such file or directory
2023-03-10 06:23:11.470014: W
tensorflow/compiler/xla/stream_executor/cuda/cuda_driver.cc:265] failed call to
cuInit: UNKNOWN ERROR (303)
2023-03-10 06:23:11.470044: I
tensorflow/compiler/xla/stream_executor/cuda/cuda_diagnostics.cc:156] kernel
driver does not appear to be running on this host (python-20230307-192621):
/proc/driver/nvidia/version does not exist
```

```
Collecting en-core-web-lg==3.5.0
  Downloading https://github.com/explosion/spacy-
models/releases/download/en_core_web_lg-3.5.0/en_core_web_lg-3.5.0-py3-none-
any.whl (587.7 MB)
                              587.7/587.7
MB 1.4 MB/s eta 0:00:0000:0100:02
Requirement already satisfied: spacy<3.6.0,>=3.5.0 in
/opt/conda/lib/python3.7/site-packages (from en-core-web-lg==3.5.0) (3.5.0)
Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in
/opt/conda/lib/python3.7/site-packages (from spacy<3.6.0,>=3.5.0->en-core-web-
lg==3.5.0) (1.0.9)
Requirement already satisfied: spacy-legacy<3.1.0,>=3.0.11 in
/opt/conda/lib/python3.7/site-packages (from spacy<3.6.0,>=3.5.0->en-core-web-
lg==3.5.0) (3.0.12)
Requirement already satisfied: requests<3.0.0,>=2.13.0 in
/opt/conda/lib/python3.7/site-packages (from spacy<3.6.0,>=3.5.0->en-core-web-
lg==3.5.0) (2.28.2)
Requirement already satisfied: numpy>=1.15.0 in /opt/conda/lib/python3.7/site-
packages (from spacy<3.6.0,>=3.5.0->en-core-web-lg==3.5.0) (1.21.6)
Requirement already satisfied: setuptools in /opt/conda/lib/python3.7/site-
packages (from spacy<3.6.0,>=3.5.0->en-core-web-lg==3.5.0) (66.1.1)
Requirement already satisfied: catalogue<2.1.0,>=2.0.6 in
/opt/conda/lib/python3.7/site-packages (from spacy<3.6.0,>=3.5.0->en-core-web-
lg==3.5.0) (2.0.8)
Requirement already satisfied: typer<0.8.0,>=0.3.0 in
/opt/conda/lib/python3.7/site-packages (from spacy<3.6.0,>=3.5.0->en-core-web-
lg==3.5.0) (0.7.0)
Requirement already satisfied: jinja2 in /opt/conda/lib/python3.7/site-packages
(from spacy<3.6.0,>=3.5.0->en-core-web-lg==3.5.0) (3.1.2)
Requirement already satisfied: preshed<3.1.0,>=3.0.2 in
/opt/conda/lib/python3.7/site-packages (from spacy<3.6.0,>=3.5.0->en-core-web-
lg==3.5.0) (3.0.8)
Requirement already satisfied: srsly<3.0.0,>=2.4.3 in
/opt/conda/lib/python3.7/site-packages (from spacy<3.6.0,>=3.5.0->en-core-web-
lg==3.5.0) (2.4.6)
Requirement already satisfied: pathy>=0.10.0 in /opt/conda/lib/python3.7/site-
packages (from spacy<3.6.0,>=3.5.0->en-core-web-lg==3.5.0) (0.10.1)
Requirement already satisfied: smart-open<7.0.0,>=5.2.1 in
/opt/conda/lib/python3.7/site-packages (from spacy<3.6.0,>=3.5.0->en-core-web-
lg==3.5.0) (6.3.0)
Requirement already satisfied: pydantic!=1.8,!=1.8.1,<1.11.0,>=1.7.4 in
/opt/conda/lib/python3.7/site-packages (from spacy<3.6.0,>=3.5.0->en-core-web-
lg==3.5.0) (1.10.4)
Requirement already satisfied: tqdm<5.0.0,>=4.38.0 in
/opt/conda/lib/python3.7/site-packages (from spacy<3.6.0,>=3.5.0->en-core-web-
lg==3.5.0) (4.64.1)
Requirement already satisfied: spacy-loggers<2.0.0,>=1.0.0 in
```

```
/opt/conda/lib/python3.7/site-packages (from spacy<3.6.0,>=3.5.0->en-core-web-
lg==3.5.0) (1.0.4)
Requirement already satisfied: cymem<2.1.0,>=2.0.2 in
/opt/conda/lib/python3.7/site-packages (from spacy<3.6.0,>=3.5.0->en-core-web-
lg==3.5.0) (2.0.7)
Requirement already satisfied: packaging>=20.0 in /opt/conda/lib/python3.7/site-
packages (from spacy<3.6.0,>=3.5.0->en-core-web-lg==3.5.0) (23.0)
Requirement already satisfied: wasabi<1.2.0,>=0.9.1 in
/opt/conda/lib/python3.7/site-packages (from spacy<3.6.0,>=3.5.0->en-core-web-
lg==3.5.0) (1.1.1)
Requirement already satisfied: langcodes<4.0.0,>=3.2.0 in
/opt/conda/lib/python3.7/site-packages (from spacy<3.6.0,>=3.5.0->en-core-web-
lg==3.5.0) (3.3.0)
Requirement already satisfied: typing-extensions<4.5.0,>=3.7.4.1 in
/opt/conda/lib/python3.7/site-packages (from spacy<3.6.0,>=3.5.0->en-core-web-
lg==3.5.0) (4.4.0)
Requirement already satisfied: thinc<8.2.0,>=8.1.0 in
/opt/conda/lib/python3.7/site-packages (from spacy<3.6.0,>=3.5.0->en-core-web-
lg==3.5.0) (8.1.9)
Requirement already satisfied: zipp>=0.5 in /opt/conda/lib/python3.7/site-
packages (from catalogue<2.1.0,>=2.0.6->spacy<3.6.0,>=3.5.0->en-core-web-
lg==3.5.0) (3.11.0)
Requirement already satisfied: charset-normalizer<4,>=2 in
/opt/conda/lib/python3.7/site-packages (from
requests<3.0.0,>=2.13.0->spacy<3.6.0,>=3.5.0->en-core-web-lg==3.5.0) (2.1.1)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in
/opt/conda/lib/python3.7/site-packages (from
requests<3.0.0,>=2.13.0->spacy<3.6.0,>=3.5.0->en-core-web-lg==3.5.0) (1.26.14)
Requirement already satisfied: certifi>=2017.4.17 in
/opt/conda/lib/python3.7/site-packages (from
requests<3.0.0,>=2.13.0->spacy<3.6.0,>=3.5.0->en-core-web-lg==3.5.0) (2022.12.7)
Requirement already satisfied: idna<4,>=2.5 in /opt/conda/lib/python3.7/site-
packages (from requests<3.0.0,>=2.13.0->spacy<3.6.0,>=3.5.0->en-core-web-
lg==3.5.0) (3.4)
Requirement already satisfied: confection<1.0.0,>=0.0.1 in
/opt/conda/lib/python3.7/site-packages (from
thinc<8.2.0,>=8.1.0->spacy<3.6.0,>=3.5.0->en-core-web-lg==3.5.0) (0.0.4)
Requirement already satisfied: blis<0.8.0,>=0.7.8 in
/opt/conda/lib/python3.7/site-packages (from
thinc<8.2.0,>=8.1.0->spacy<3.6.0,>=3.5.0->en-core-web-lg==3.5.0) (0.7.9)
Requirement already satisfied: click<9.0.0,>=7.1.1 in
/opt/conda/lib/python3.7/site-packages (from
typer<0.8.0,>=0.3.0->spacy<3.6.0,>=3.5.0->en-core-web-lg==3.5.0) (8.1.3)
Requirement already satisfied: MarkupSafe>=2.0 in /opt/conda/lib/python3.7/site-
packages (from jinja2->spacy<3.6.0,>=3.5.0->en-core-web-lg==3.5.0) (2.1.2)
Requirement already satisfied: importlib-metadata in
/opt/conda/lib/python3.7/site-packages (from
click<9.0.0,>=7.1.1->typer<0.8.0,>=0.3.0->spacy<3.6.0,>=3.5.0->en-core-web-
```

```
lg==3.5.0) (6.0.0)
Installing collected packages: en-core-web-lg
Successfully installed en-core-web-lg-3.5.0
 Download and installation successful
You can now load the package via spacy.load('en_core_web_lg')
```

[13]:
```
nlp = spacy.load("en_core_web_lg")
```

[ ]:
```
Neg=news_df[news_df["Sentiment"]=="Negative"]
Pos=news_df[news_df["Sentiment"]=="Positive"]
print("Negative :" , Neg.shape)
print("Positive :" ,Pos.shape)
```

[ ]:
```
T1_Neg=Neg[Neg["topic"]==1]
T1_Neg.reset_index(inplace=True)

T1_Pos=Pos[Pos["topic"]==1]
T1_Pos.reset_index(inplace=True)
```

1 Topic 1 Neg Sentiment

[49]:
```
T1_Negentities_org=[]
T1_Negentities_person=[]
T1_Negentities_loc=[]

print("Length: ",len(T1_Neg))
for i in range(len(T1_Neg)):
    if i % 1000 == 0:
        print(i)
    doc = nlp(T1_Neg["text_cleaned"][i])
    entitiesO = []
    entitiesP = []
    entitiesL = []
    for ent in doc.ents:
        if ent.label_ == 'ORG':
            entitiesO.append(ent.text)
        elif ent.label_ == 'PERSON':
            entitiesP.append(ent.text)
        elif ent.label_ == 'GPE':
            entitiesL.append(ent.text)
    T1_Negentities_org.append(entitiesO)
    T1_Negentities_person.append(entitiesP)
    T1_Negentities_loc.append(entitiesL)
```

```
Length:  9659
0
1000
```

```
2000
3000
4000
5000
6000
7000
8000
9000
```

[50]:
```
import pickle

with open('T1_Negentities_org.pkl', 'wb') as f:
    pickle.dump(T1_Negentities_org, f)
with open('T1_Negentities_person.pkl', 'wb') as f:
    pickle.dump(T1_Negentities_person, f)
with open('T1_Negentities_loc.pkl', 'wb') as f:
    pickle.dump(T1_Negentities_loc, f)
```

[85]:
```
from collections import Counter
```

[87]:
```
counter = Counter()
for sublist in T1_Negentities_org:
    for item in set(sublist):
        counter[item] += 1

top_20_T1_Negentities_org = counter.most_common(20)

print('The top 20 organisations talking about Topic 1 Negatively:',␣
 ↪top_20_T1_Negentities_org)


counter = Counter()
for sublist in T1_Negentities_person:
    for item in set(sublist):
        counter[item] += 1

top_20_T1_Negentities_person = counter.most_common(20)

print('\n\nThe top 20 people talking about Topic 1 Negatively:',␣
 ↪top_20_T1_Negentities_person)


counter = Counter()
for sublist in T1_Negentities_loc:
    for item in set(sublist):
        counter[item] += 1
```

```
top_20_T1_Negentities_loc = counter.most_common(20)

print('\n\nThe top 20 locations of Topic 1 Negative news:',␣
 ↪top_20_T1_Negentities_loc)
```

```
The top 20 organisations talking about Topic 1 Negatively: [('npr', 2447),
('fcc', 1878), ('youtube', 1396), ('chatgpt', 1327), ('android', 1142),
('samsung', 1105), ('microsoft', 1032), ('npr news', 930), ('google', 901),
('sony', 584), ('xbox', 582), ('strategiesbest', 560),
('cryptocurrencydeficrypto', 560), ('defidecentralized exchangesbest defi',
560), ('brokerscrypto brokersoptions brokersetf brokersmutual fund', 560),
('brokersstock brokersforex brokersfutures', 560), ('scannersbest business
crypto accountsbest crypto', 560), ('estateinvest wineinvest', 560),
('brokersstock appsall broker', 560), ('brokersindex', 560)]


The top 20 people talking about Topic 1 Negatively: [('premarketafter
hoursmoversetfsforexcannabiscommoditiesoptionsbinary optionsbondsfuturescme',
560), ('nftscryptopunks watchlistare', 560), ('reviewsinsuranceautohomemedicarel
ifevisiondentalbusinesspethealthmotorcyclerentersworkers comptop stockspenny
stocksstocks', 560), ('blockchainbest altcoinshow', 560), ('startedis bitcoin',
554), ('crowdfundinghow', 522), ('calendarearnings calendareconomic calendarfda
calendarguidance', 513), ('groupglobal
economicspreviewssmallcapcryptocurrencypenny stocksdigital', 480),
('earningsguidancedividendsmabuybackslegalinterviewsmanagementretail
salesofferingsiposinsider', 474), ('bobby allyn', 349), ('jerseynew mexiconew
yorknorth carolinanorth dakotaohiooklahomaoregonpennsylvaniarhode islandsouth',
334), ('micronesiaguampalaualberta canadabritish columbia', 332), ('thenew
caledonianew zealandnicaragua republic', 332), ('ofpapua', 332), ('ofcayman
islandscentral african republicchad republic ofchile republic', 332),
('ofkyrgyz', 332), ('ofchristmas islandcocos', 332), ('ofitaly italian', 332),
('americaus virgin islandsunited state', 332), ('republicjapanjordan hashemite',
332)]


The top 20 locations of Topic 1 Negative news: [('new york', 1029), ('india',
917), ('california', 737), ('china', 736), ('washington', 726), ('canada', 665),
('australia', 564), ('brokersshort', 560), ('japan', 525), ('america', 512),
('san francisco', 472), ('uk', 462), ('brunswick', 437), ('britain', 403),
('france', 390), ('russia', 377), ('new york city', 377), ('georgia', 358),
('pennsylvania', 355), ('ofhong kong', 332)]
```

2 Topic 1 Pos Sentiment

[29]:
```
T1_Posentities_org=[]
T1_Posentities_person=[]
T1_Posentities_loc=[]

print("Length: ",len(T1_Pos))
for i in range(len(T1_Pos)):
    if i % 1000 == 0:
        print(i)
    doc = nlp(T1_Pos["text_cleaned"][i])
    entitiesO = []
    entitiesP = []
    entitiesL = []
    for ent in doc.ents:
        if ent.label_ == 'ORG':
            entitiesO.append(ent.text)
        elif ent.label_ == 'PERSON':
            entitiesP.append(ent.text)
        elif ent.label_ == 'GPE':
            entitiesL.append(ent.text)
    T1_Posentities_org.append(entitiesO)
    T1_Posentities_person.append(entitiesP)
    T1_Posentities_loc.append(entitiesL)
```

```
Length:  4302
0
1000
2000
3000
4000
```

[48]:
```
with open('T1_Posentities_org.pkl', 'wb') as f:
    pickle.dump(T1_Posentities_org, f)
with open('T1_Posentities_person.pkl', 'wb') as f:
    pickle.dump(T1_Posentities_person, f)
with open('T1_Posentities_loc.pkl', 'wb') as f:
    pickle.dump(T1_Posentities_loc, f)
```

[88]:
```
counter = Counter()
for sublist in T1_Posentities_org:
    for item in set(sublist):
        counter[item] += 1

top_20_T1_Posentities_org = counter.most_common(20)
```

```
print('The top 20 organisations talking about Topic 1 Positvely:',␣
 ↪top_20_T1_Posentities_org)


counter = Counter()
for sublist in T1_Posentities_person:
    for item in set(sublist):
        counter[item] += 1

top_20_T1_Posentities_person = counter.most_common(20)

print('\n\nThe top 20 people talking about Topic 1 Positvely:',␣
 ↪top_20_T1_Posentities_person)


counter = Counter()
for sublist in T1_Posentities_loc:
    for item in set(sublist):
        counter[item] += 1

top_20_T1_Posentities_loc = counter.most_common(20)

print('\n\nThe top 20 locations of Topic 1 Positve news:',␣
 ↪top_20_T1_Posentities_loc)
```

```
The top 20 organisations talking about Topic 1 Positvely: [('samsung', 720),
('android', 699), ('microsoft', 408), ('xbox', 396), ('gpu', 371), ('sony',
326), ('england wale', 322), ('google', 305), ('quay house ambury', 298),
('youtube', 276), ('amazon', 261), ('utc', 255), ('gb', 244), ('chatgpt', 236),
('npr', 236), ('ps5', 221), ('united state', 220), ('intel', 205), ('nvidia',
202), ('u inc international medium group', 202)]


The top 20 people talking about Topic 1 Positvely: [('tom', 180), ('thenew
caledonianew zealandnicaragua republic', 151), ('ofpapua', 151), ('ofkyrgyz',
151), ('ofcayman islandscentral african republicchad republic ofchile republic',
151), ('ofchristmas islandcocos', 151), ('ofitaly italian', 151), ('americaus
virgin islandsunited state', 151), ('republicjapanjordan hashemite', 151),
('ofheard mcdonald islandsholy', 151), ('ofviet nam', 151),
('ofmauritiusmayottemicronesia', 151), ('republicsloveniasolomon
islandssomalia', 151), ('kitts nevisst luciast pierre miquelonst', 151),
('ofburkina fasoburundi', 151), ('vincent grenadinessudan', 151), ('united
republic ofcape verde', 151), ('islandnorthern mariana islandsnorway', 151),
('ofturkmenistanturks caicos islandstuvaluuganda republic ofukraineunited',
151), ('virgin islandsbrunei darussalambulgaria people republic', 151)]
```

```
The top 20 locations of Topic 1 Positve news: [('india', 362), ('australia',
361), ('new york', 343), ('japan', 291), ('canada', 277), ('china', 269),
('brunswick', 226), ('california', 215), ('ny', 211), ('uk', 207), ('london',
176), ('washington', 173), ('germany', 164), ('britain', 160), ('ofhong kong',
151), ('ofsaudi arabia', 151), ('socialist republic ofwallis futuna', 151),
('islandscosta rica republic', 151), ('ofcomoros', 151),
('chinatajikistantanzania united republic', 151)]
```

3 Topic 2 Neg Sentiment

[51]:
```
T2_Neg=Neg[Neg["topic"]==2]
T2_Neg.reset_index(inplace=True)

T2_Pos=Pos[Pos["topic"]==2]
T2_Pos.reset_index(inplace=True)
```

[52]:
```
T2_Negentities_org=[]
T2_Negentities_person=[]
T2_Negentities_loc=[]

print("Length: ",len(T2_Neg))
for i in range(len(T2_Neg)):
    if i % 2000 == 0:
        print(i)
    doc = nlp(T2_Neg["text_cleaned"][i])
    entitiesO = []
    entitiesP = []
    entitiesL = []
    for ent in doc.ents:
        if ent.label_ == 'ORG':
            entitiesO.append(ent.text)
        elif ent.label_ == 'PERSON':
            entitiesP.append(ent.text)
        elif ent.label_ == 'GPE':
            entitiesL.append(ent.text)
    T2_Negentities_org.append(entitiesO)
    T2_Negentities_person.append(entitiesP)
    T2_Negentities_loc.append(entitiesL)
```

```
Length:  33689
0
2000
4000
6000
8000
10000
12000
```

```
14000
16000
18000
20000
22000
24000
26000
28000
30000
32000
```

[53]:
```
with open('T2_Negentities_org.pkl', 'wb') as f:
    pickle.dump(T2_Negentities_org, f)
with open('T2_Negentities_person.pkl', 'wb') as f:
    pickle.dump(T2_Negentities_person, f)
with open('T2_Negentities_loc.pkl', 'wb') as f:
    pickle.dump(T2_Negentities_loc, f)
```

[89]:
```
counter = Counter()
for sublist in T2_Negentities_org:
    for item in set(sublist):
        counter[item] += 1

top_20_T2_Negentities_org = counter.most_common(20)

print('The top 20 organisations talking about Topic 2 Negatively:',␣
 ↪top_20_T2_Negentities_org)


counter = Counter()
for sublist in T2_Negentities_person:
    for item in set(sublist):
        counter[item] += 1

top_20_T2_Negentities_person = counter.most_common(20)

print('\n\nThe top 20 people talking about Topic 2 Negatively:',␣
 ↪top_20_T2_Negentities_person)


counter = Counter()
for sublist in T2_Negentities_loc:
    for item in set(sublist):
        counter[item] += 1

top_20_T2_Negentities_loc = counter.most_common(20)
```

```
print('\n\nThe top 20 locations of Topic 2 Negative news:',␣
 ↪top_20_T2_Negentities_loc)
```

```
The top 20 organisations talking about Topic 2 Negatively: [('gray television
inc', 15421), ('music lifestylegray dc', 11153), ('fda', 6568), ('american
african american asian american', 4446), ('espaol', 4355), ('united kingdom
service new', 4353), ('esko danmark deutschland espaa france italia nederland',
4353), ('gray medium group inc station gray television inc', 4111), ('cision
communication cloud marketer public', 3969), ('consumer technologyall consumer
technology', 3923), ('cision communication cloud cision ir', 3706), ('profnet
cision distribution helpline', 3670), ('cision cision', 3636), ('u state',
3581), ('cision communication cloud cision ir product general inquiry', 3291),
('suomi sverige overview distribution pr', 2950), ('serviceprivacy policyeeo
statementfcc', 2877), ('newscastspress', 2825), ('datadriven', 2413), ('cision
ir', 2225)]


The top 20 people talking about Topic 2 Negatively: [('captioningaudio
descriptiona gray medium group inc', 6836), ('greta van susterencircle country',
5175), ('r gdpr', 4862), ('norge polska', 4353), ('bureauinvestigate', 2079),
('captioningaudio descriptionadvertisinga gray medium group inc', 1293), ('greta
van susterengray dc', 1161), ('serviceprivacy policyfcc applicationseeo', 817),
('greta van', 749), ('captioningaudio', 711), ('gdpr', 676), ('pm gmt', 645),
('bureauinvestigate tvlatest', 628), ('filefcc applicationseeo', 591),
('tvlatest newscastspress', 574), ('tvgray', 563), ('bureauinvestigate
tvcircle', 475), ('applicationsterms serviceprivacy policyeeo
statementadvertisinga', 464), ('captioningaudio descriptionprivacy policyterms
serviceadvertisinga gray medium group inc', 455), ('schoolcmuathlete weekjuco',
450)]


The top 20 locations of Topic 2 Negative news: [('israel', 5996), ('india',
5881), ('canada', 5799), ('germany', 5558), ('france', 5523), ('sweden', 5241),
('spain', 5208), ('netherlands', 5180), ('russia', 5151), ('brazil', 5149),
('poland', 5087), ('denmark', 5062), ('norway', 5052), ('slovakia', 5030),
('portugal', 5026), ('italy', 4725), ('slovensko', 4353), ('serviceprivacy',
2433), ('china', 1818), ('california', 1550)]
```

4 Topic 2 Pos Sentiment

[56]:
```
T2_Posentities_org=[]
T2_Posentities_person=[]
T2_Posentities_loc=[]

print("Length: ",len(T2_Pos))
for i in range(len(T2_Pos)):
```

```
    if i % 2000 == 0:
        print(i)
    doc = nlp(T2_Pos["text_cleaned"][i])
    entitiesO = []
    entitiesP = []
    entitiesL = []
    for ent in doc.ents:
        if ent.label_ == 'ORG':
            entitiesO.append(ent.text)
        elif ent.label_ == 'PERSON':
            entitiesP.append(ent.text)
        elif ent.label_ == 'GPE':
            entitiesL.append(ent.text)
    T2_Posentities_org.append(entitiesO)
    T2_Posentities_person.append(entitiesP)
    T2_Posentities_loc.append(entitiesL)
```

```
Length:  8300
0
2000
4000
6000
8000
```

[57]:
```
with open('T2_Posentities_org.pkl', 'wb') as f:
    pickle.dump(T2_Posentities_org, f)
with open('T2_Posentities_person.pkl', 'wb') as f:
    pickle.dump(T2_Posentities_person, f)
with open('T2_Posentities_loc.pkl', 'wb') as f:
    pickle.dump(T2_Posentities_loc, f)
```

[90]:
```
counter = Counter()
for sublist in T2_Posentities_org:
    for item in set(sublist):
        counter[item] += 1

top_20_T2_Posentities_org = counter.most_common(20)

print('The top 20 organisations talking about Topic 2 Positvely:',␣
 ↪top_20_T2_Posentities_org)


counter = Counter()
for sublist in T2_Posentities_person:
    for item in set(sublist):
        counter[item] += 1
```

```
top_20_T2_Posentities_person = counter.most_common(20)

print('\n\nThe top 20 people talking about Topic 2 Positvely:',␣
 ↪top_20_T2_Posentities_person)


counter = Counter()
for sublist in T2_Posentities_loc:
    for item in set(sublist):
        counter[item] += 1

top_20_T2_Posentities_loc = counter.most_common(20)

print('\n\nThe top 20 locations of Topic 2 Positve news:',␣
 ↪top_20_T2_Posentities_loc)
```

```
The top 20 organisations talking about Topic 2 Positvely: [('gray television
inc', 4001), ('music lifestylegray dc', 3001), ('gray medium group inc station
gray television inc', 1191), ('newscastspress', 861), ('serviceprivacy policyeeo
statementfcc', 722), ('fda', 707), ('ml', 642), ('bureauinvestigate
tvpowernationlatest', 610), ('serviceprivacy policyeeo statementfcc
applicationsadvertisinga gray medium group inc station gray television inc',
575), ('espaol', 542), ('united kingdom service new', 537), ('esko danmark
deutschland espaa france italia nederland', 537), ('bureaucircle', 532),
('american african american asian american', 524), ('profnet cision distribution
helpline', 511), ('cision cision', 509), ('consumer technologyall consumer
technology', 508), ('videoscircle country', 506), ('u state', 503),
('datadriven', 487)]


The top 20 people talking about Topic 2 Positvely: [('captioningaudio
descriptiona gray medium group inc', 1683), ('greta van susterencircle country',
1583), ('r gdpr', 554), ('bureauinvestigate', 539), ('norge polska', 537),
('captioningaudio descriptionadvertisinga gray medium group inc', 349), ('greta
van susterengray dc', 332), ('hailo', 238), ('serviceprivacy policyfcc
applicationseeo', 234), ('austin', 226), ('greta van', 220), ('ltdaxiomtek',
218), ('korea china', 199), ('inperson', 195), ('filefcc applicationseeo', 190),
('captioningaudio', 180), ('jetson xavier', 171), ('zach shelby', 170),
('tvlatest newscastspress', 166), ('tvgray', 151)]


The top 20 locations of Topic 2 Positve news: [('israel', 948), ('germany',
747), ('india', 734), ('france', 720), ('serviceprivacy', 649), ('texas', 645),
('brazil', 637), ('canada', 633), ('netherlands', 624), ('denmark', 623),
('slovakia', 618), ('spain', 613), ('russia', 608), ('sweden', 605),
('portugal', 604), ('norway', 604), ('poland', 604), ('italy', 599),
('slovensko', 537), ('china', 517)]
```

5 Topic 3 Neg Sentiment

[58]:
```
T3_Neg=Neg[Neg["topic"]==3]
T3_Neg.reset_index(inplace=True)

T3_Pos=Pos[Pos["topic"]==3]
T3_Pos.reset_index(inplace=True)
```

[59]:
```
T3_Negentities_org=[]
T3_Negentities_person=[]
T3_Negentities_loc=[]

print("Length: ",len(T3_Neg))
for i in range(len(T3_Neg)):
    if i % 3000 == 0:
        print(i)
    doc = nlp(T3_Neg["text_cleaned"][i])
    entitiesO = []
    entitiesP = []
    entitiesL = []
    for ent in doc.ents:
        if ent.label_ == 'ORG':
            entitiesO.append(ent.text)
        elif ent.label_ == 'PERSON':
            entitiesP.append(ent.text)
        elif ent.label_ == 'GPE':
            entitiesL.append(ent.text)
    T3_Negentities_org.append(entitiesO)
    T3_Negentities_person.append(entitiesP)
    T3_Negentities_loc.append(entitiesL)
```

```
Length:  22754
0
3000
6000
9000
12000
15000
18000
21000
```

[60]:
```
with open('T3_Negentities_org.pkl', 'wb') as f:
    pickle.dump(T3_Negentities_org, f)
with open('T3_Negentities_person.pkl', 'wb') as f:
    pickle.dump(T3_Negentities_person, f)
with open('T3_Negentities_loc.pkl', 'wb') as f:
    pickle.dump(T3_Negentities_loc, f)
```

[91]:
```
counter = Counter()
for sublist in T3_Negentities_org:
    for item in set(sublist):
        counter[item] += 1

top_20_T3_Negentities_org = counter.most_common(20)

print('The top 20 organisations talking about Topic 3 Negatively:',␣
 ↪top_20_T3_Negentities_org)


counter = Counter()
for sublist in T3_Negentities_person:
    for item in set(sublist):
        counter[item] += 1

top_20_T3_Negentities_person = counter.most_common(20)

print('\n\nThe top 20 people talking about Topic 3 Negatively:',␣
 ↪top_20_T3_Negentities_person)


counter = Counter()
for sublist in T3_Negentities_loc:
    for item in set(sublist):
        counter[item] += 1

top_20_T3_Negentities_loc = counter.most_common(20)

print('\n\nThe top 20 locations of Topic 3 Negative news:',␣
 ↪top_20_T3_Negentities_loc)
```

```
The top 20 organisations talking about Topic 3 Negatively: [('fcc', 4887),
('chatgpt', 3775), ('android', 3383), ('google', 2988), ('microsoft', 2219),
('nexstar medium inc right', 2136), ('congress', 1979), ('bestreviews', 1909),
('white house', 1710), ('nfl', 1651), ('cnn', 1631), ('hill newsnation
bestreviews', 1468), ('nasa', 1393), ('reuters', 1288), ('united state', 1248),
('youtube', 1178), ('gamescomputerselectronicscell', 1163),
('courtselectionslawpresidential electionimmigrationpotusforeign
policyhealthhealthpublic', 1163), ('leaguebasketballcombat', 1163),
('designwildlifeworkoutslotterycrime safetypublic safetyaccidentslaw
enforcementtraffic violationsvandalismsocietysocietyenvironmentreligionadvocacyr
elationshipscharitieskidsrelationship', 1163)]


The top 20 people talking about Topic 3 Negatively: [('donald trump', 1443),
('joe biden', 1251), ('biden', 1232), ('servicesmental healthdiseases
```

```
treatmentspharmaceuticalswomens', 1163),
('safetynutritionsportssportsfootballnflsoccerbaseballmlbcollege sportspremier',
1163), ('advicefamily relationshipsprotestsfestivalhomelesshouse rentlabor
issuestrouble', 1160), ('drinkspetsbeauty fashionanimalsshoppinghome', 1135),
('matt obrien', 785), ('trump', 713), ('playbutton cancelcircle', 665),
('literaturecomicstheater dancebehind', 657), ('sundar pichai', 656), ('sam
altman', 653), ('joe bidens', 634), ('jones', 631), ('chevronright chevronleft
chevronup', 602), ('william', 568), ('rebekah jones', 555), ('mark zuckerberg',
538), ('obama', 506)]
```

```
The top 20 locations of Topic 3 Negative news: [('california', 5227), ('china',
4745), ('new york', 3894), ('washington', 3328), ('india', 3274), ('russia',
3178), ('america', 2972), ('florida', 2534), ('london', 2277), ('uk', 2102),
('texas', 2096), ('new york city', 1799), ('france', 1735), ('japan', 1662),
('san francisco', 1451), ('australia', 1428), ('beijing', 1384), ('pakistan',
1369), ('paris', 1328), ('los angeles', 1315)]
```

6 Topic 3 Pos Sentiment

[61]:
```
T3_Posentities_org=[]
T3_Posentities_person=[]
T3_Posentities_loc=[]

print("Length: ",len(T3_Pos))
for i in range(len(T3_Pos)):
    if i % 3000 == 0:
        print(i)
    doc = nlp(T3_Pos["text_cleaned"][i])
    entitiesO = []
    entitiesP = []
    entitiesL = []
    for ent in doc.ents:
        if ent.label_ == 'ORG':
            entitiesO.append(ent.text)
        elif ent.label_ == 'PERSON':
            entitiesP.append(ent.text)
        elif ent.label_ == 'GPE':
            entitiesL.append(ent.text)
    T3_Posentities_org.append(entitiesO)
    T3_Posentities_person.append(entitiesP)
    T3_Posentities_loc.append(entitiesL)
```

```
Length:  5142
0
3000
```

[62]:
```
with open('T3_Posentities_org.pkl', 'wb') as f:
    pickle.dump(T3_Posentities_org, f)
with open('T3_Posentities_person.pkl', 'wb') as f:
    pickle.dump(T3_Posentities_person, f)
with open('T3_Posentities_loc.pkl', 'wb') as f:
    pickle.dump(T3_Posentities_loc, f)
```

[92]:
```
counter = Counter()
for sublist in T3_Posentities_org:
    for item in set(sublist):
        counter[item] += 1

top_20_T3_Posentities_org = counter.most_common(20)

print('The top 20 organisations talking about Topic 3 Positvely:',␣
 ↪top_20_T3_Posentities_org)


counter = Counter()
for sublist in T3_Posentities_person:
    for item in set(sublist):
        counter[item] += 1

top_20_T3_Posentities_person = counter.most_common(20)

print('\n\nThe top 20 people talking about Topic 3 Positvely:',␣
 ↪top_20_T3_Posentities_person)


counter = Counter()
for sublist in T3_Posentities_loc:
    for item in set(sublist):
        counter[item] += 1

top_20_T3_Posentities_loc = counter.most_common(20)

print('\n\nThe top 20 locations of Topic 3 Positve news:',␣
 ↪top_20_T3_Posentities_loc)
```

```
The top 20 organisations talking about Topic 3 Positvely: [('connectstoriestech
news', 606), ('chatgpt', 540), ('everydayweb3keep', 437),
('gamescomputerselectronicscell', 364), ('courtselectionslawpresidential
electionimmigrationpotusforeign policyhealthhealthpublic', 364),
('leaguebasketballcombat', 364), ('scienceearth scienceeducationhigh
schoolcollegesworldworldunited', 364), ('designwildlifeworkoutslotterycrime
safetypublic safetyaccidentslaw enforcementtraffic violationsvandalismsocietysoc
ietyenvironmentreligionadvocacyrelationshipscharitieskidsrelationship', 364),
```

```
('infohelp', 336), ('android', 321), ('nationsmiddle eastafghanistaneuropeindiau
kamericasnewsbreakcontributorspublishersadvertisershomelocalcoronavirusoriginals
earch', 313), ('zooplacouk prime location', 295), ('uk news sport u', 295),
('ltd part', 295), ('netflix', 293), ('huntproductsbest productsdiscover', 275),
('congress', 271), ('nfl', 262), ('worldchangelogrelease', 261), ('samsung',
258)]
```

```
The top 20 people talking about Topic 3 Positvely: [('servicesmental
healthdiseases treatmentspharmaceuticalswomens', 364),
('safetynutritionsportssportsfootballnflsoccerbaseballmlbcollege sportspremier',
364), ('advicefamily relationshipsprotestsfestivalhomelesshouse rentlabor
issuestrouble', 359), ('drinkspetsbeauty fashionanimalsshoppinghome', 357),
('kim kardashian', 302), ('topicsupcoming productssee', 298), ('donald trump',
274), ('kylie jenner', 252), ('startuplaunch guidechecklists', 239),
('literaturecomicstheater dancebehind', 227), ('taylor', 225), ('jennifer
lopez', 223), ('harry', 223), ('britney', 218), ('joe biden', 214), ('kanye
west', 206), ('kendall jenner', 202), ('khloe kardashian', 201), ('meghan
markle', 196), ('tom', 187)]
```

```
The top 20 locations of Topic 3 Positve news: [('new york', 843), ('china',
821), ('india', 800), ('california', 766), ('america', 633), ('london', 604),
('los angeles', 565), ('uk', 531), ('australia', 515), ('paris', 514),
('washington', 458), ('new york city', 432), ('policydo', 426), ('florida',
424), ('hollywood', 419), ('russia', 408), ('japan', 390), ('miami', 373),
('healthhealth', 364), ('texas', 358)]
```

7 Topic 4 Neg Sentiment

[63]:
```
T4_Neg=Neg[Neg["topic"]==4]
T4_Neg.reset_index(inplace=True)

T4_Pos=Pos[Pos["topic"]==4]
T4_Pos.reset_index(inplace=True)
```

[64]:
```
T4_Negentities_org=[]
T4_Negentities_person=[]
T4_Negentities_loc=[]

print("Length: ",len(T4_Neg))
for i in range(len(T4_Neg)):
    if i % 1000 == 0:
        print(i)
    doc = nlp(T4_Neg["text_cleaned"][i])
    entitiesO = []
    entitiesP = []
```

```
    entitiesL = []
    for ent in doc.ents:
        if ent.label_ == 'ORG':
            entitiesO.append(ent.text)
        elif ent.label_ == 'PERSON':
            entitiesP.append(ent.text)
        elif ent.label_ == 'GPE':
            entitiesL.append(ent.text)
    T4_Negentities_org.append(entitiesO)
    T4_Negentities_person.append(entitiesP)
    T4_Negentities_loc.append(entitiesL)
```

```
Length:  40490
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
10000
11000
12000
13000
14000
15000
16000
17000
18000
19000
20000
21000
22000
23000
24000
25000
26000
27000
28000
29000
30000
31000
32000
33000
```

```
34000
35000
36000
37000
38000
39000
40000
```

[65]:
```
with open('T4_Negentities_org.pkl', 'wb') as f:
    pickle.dump(T4_Negentities_org, f)
with open('T4_Negentities_person.pkl', 'wb') as f:
    pickle.dump(T4_Negentities_person, f)
with open('T4_Negentities_loc.pkl', 'wb') as f:
    pickle.dump(T4_Negentities_loc, f)
```

[93]:
```
counter = Counter()
for sublist in T4_Negentities_org:
    for item in set(sublist):
        counter[item] += 1

top_20_T4_Negentities_org = counter.most_common(20)

print('The top 20 organisations talking about Topic 4 Negatively:',␣
 ↪top_20_T4_Negentities_org)


counter = Counter()
for sublist in T4_Negentities_person:
    for item in set(sublist):
        counter[item] += 1

top_20_T4_Negentities_person = counter.most_common(20)

print('\n\nThe top 20 people talking about Topic 4 Negatively:',␣
 ↪top_20_T4_Negentities_person)


counter = Counter()
for sublist in T4_Negentities_loc:
    for item in set(sublist):
        counter[item] += 1

top_20_T4_Negentities_loc = counter.most_common(20)

print('\n\nThe top 20 locations of Topic 4 Negative news:',␣
 ↪top_20_T4_Negentities_loc)
```

```
The top 20 organisations talking about Topic 4 Negatively: [('google', 3086),
('ibm', 2862), ('chatgpt', 2664), ('microsoft', 2659), ('ml', 2466),
('congress', 2057), ('united state', 1980), ('datadriven', 1886), ('nasa',
1681), ('eu', 1591), ('gpt3', 1451), ('scitech', 1431), ('healthcare', 1419),
('fed', 1369), ('reuters', 1333), ('medium llc', 1227), ('amazon', 1163),
('brandvoice', 1133), ('youtube', 1101), ('g', 1072)]
```

```
The top 20 people talking about Topic 4 Negatively: [('internet1', 765),
('analyticsibeat1', 722), ('toris bibtex', 707), ('websiteprivacy policyterms
conditionsguestpost guidelinessitemap', 680), ('deepmind', 678), ('russell
rabichev marketing', 652), ('gurgu cphone', 652), ('ani', 620), ('tamil nadu',
518), ('servicesmental healthdiseases treatmentspharmaceuticalswomens', 514),
('safetynutritionsportssportsfootballnflsoccerbaseballmlbcollege sportspremier',
514), ('donald trump', 512), ('advicefamily
relationshipsprotestsfestivalhomelesshouse rentlabor issuestrouble', 510),
('drinkspetsbeauty fashionanimalsshoppinghome', 504), ('sam altman', 465),
('biden', 462), ('sundar pichai', 457), ('byclicking heremore', 444), ('watson',
441), ('onlinelos angeles', 431)]
```

```
The top 20 locations of Topic 4 Negative news: [('india', 9465), ('china',
6012), ('california', 3979), ('japan', 3756), ('america', 3682), ('uk', 3175),
('australia', 3161), ('singapore', 2854), ('new york', 2812), ('washington',
2305), ('london', 2152), ('taiwan', 2071), ('indonesia', 2015), ('russia',
1978), ('israel', 1951), ('germany', 1938), ('malaysia', 1936), ('canada',
1882), ('korea', 1841), ('france', 1667)]
```

8 Topic 4 Pos Sentiment

[66]:
```
T4_Posentities_org=[]
T4_Posentities_person=[]
T4_Posentities_loc=[]

print("Length: ",len(T4_Pos))
for i in range(len(T4_Pos)):
    if i % 1000 == 0:
        print(i)
    doc = nlp(T4_Pos["text_cleaned"][i])
    entitiesO = []
    entitiesP = []
    entitiesL = []
    for ent in doc.ents:
        if ent.label_ == 'ORG':
            entitiesO.append(ent.text)
        elif ent.label_ == 'PERSON':
            entitiesP.append(ent.text)
```

```
        elif ent.label_ == 'GPE':
            entitiesL.append(ent.text)
    T4_Posentities_org.append(entitiesO)
    T4_Posentities_person.append(entitiesP)
    T4_Posentities_loc.append(entitiesL)
```

```
Length:  9273
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
```

[69]:
```
with open('T4_Posentities_org.pkl', 'wb') as f:
    pickle.dump(T4_Posentities_org, f)
with open('T4_Posentities_person.pkl', 'wb') as f:
    pickle.dump(T4_Posentities_person, f)
with open('T4_Posentities_loc.pkl', 'wb') as f:
    pickle.dump(T4_Posentities_loc, f)
```

[94]:
```
counter = Counter()
for sublist in T4_Posentities_org:
    for item in set(sublist):
        counter[item] += 1

top_20_T4_Posentities_org = counter.most_common(20)

print('The top 20 organisations talking about Topic 4 Positvely:',␣
 ↪top_20_T4_Posentities_org)


counter = Counter()
for sublist in T4_Posentities_person:
    for item in set(sublist):
        counter[item] += 1

top_20_T4_Posentities_person = counter.most_common(20)

print('\n\nThe top 20 people talking about Topic 4 Positvely:',␣
 ↪top_20_T4_Posentities_person)
```

```
counter = Counter()
for sublist in T4_Posentities_loc:
    for item in set(sublist):
        counter[item] += 1

top_20_T4_Posentities_loc = counter.most_common(20)

print('\n\nThe top 20 locations of Topic 4 Positve news:',␣
 ↪top_20_T4_Posentities_loc)
```

```
The top 20 organisations talking about Topic 4 Positvely: [('ibm', 664), ('ml',
632), ('microsoft', 513), ('medium llc', 454), ('google', 450), ('chatgpt',
436), ('brandvoice', 434), ('datadriven', 417), ('nasa', 390), ('g', 379),
('anywhere coursera inc right', 368), ('gpt3', 366), ('analystibm data
analytics', 356), ('degreesdegrees', 356), ('intelligencec
programmingcommunication', 356), ('analystibm', 356), ('degreesdata analytics
degreespublic health degreessocial science', 356), ('writingfull', 356),
('guideproject', 356), ('articlesskills data', 356)]


The top 20 people talking about Topic 4 Positvely: [('designersearn degree',
356), ('skillsskills financepopular', 356), ('certificationadvance
careerbrowse', 356), ('onlinemastertrack certificatesprofessional
certificatesuniversity certificatesmba business', 356), ('languagepythonjavaweb
designsqlcursos gratismicrosoft', 356), ('engineeringibm', 356), ('schoolsee
certificatescourseraaboutwhat offerleadershipcareerscatalogcoursera', 356),
('degreescomputer', 356), ('certificatesmastertrack certificatesdegreesfor',
356), ('skillsblockchainsee coursespopular', 356), ('certificationspopular
certificationspopular sql certificationsmarketing', 356), ('degreesmanagement
degreesdegrees', 356), ('takeis', 356), ('deepmind', 218), ('kyle
wiggerskylelwiggers', 209), ('byclicking heremore', 190), ('exploreonline
degreesfind', 183), ('venturebeat flipboard', 176),
('safetynutritionsportssportsfootballnflsoccerbaseballmlbcollege sportspremier',
169), ('servicesmental healthdiseases treatmentspharmaceuticalswomens', 169)]


The top 20 locations of Topic 4 Positve news: [('india', 1774), ('china', 1355),
('japan', 1007), ('australia', 943), ('america', 736), ('singapore', 726),
('california', 689), ('uk', 589), ('malaysia', 586), ('taiwan', 569),
('indonesia', 525), ('korea', 498), ('thailand', 497), ('new york', 478),
('philippine', 427), ('hong kongs', 410), ('brandvoice', 381), ('pathwaybsc',
356), ('skillssoft', 356), ('topicsfree', 356)]
```

9 Topic 5 Neg Sentiment

[68]:
```
T5_Neg=Neg[Neg["topic"]==5]
T5_Neg.reset_index(inplace=True)

T5_Pos=Pos[Pos["topic"]==5]
T5_Pos.reset_index(inplace=True)
```

[70]:
```
T5_Negentities_org=[]
T5_Negentities_person=[]
T5_Negentities_loc=[]

print("Length: ",len(T5_Neg))
for i in range(len(T5_Neg)):
    if i % 1000 == 0:
        print(i)
    doc = nlp(T5_Neg["text_cleaned"][i])
    entitiesO = []
    entitiesP = []
    entitiesL = []
    for ent in doc.ents:
        if ent.label_ == 'ORG':
            entitiesO.append(ent.text)
        elif ent.label_ == 'PERSON':
            entitiesP.append(ent.text)
        elif ent.label_ == 'GPE':
            entitiesL.append(ent.text)
    T5_Negentities_org.append(entitiesO)
    T5_Negentities_person.append(entitiesP)
    T5_Negentities_loc.append(entitiesL)
```

```
Length:  10471
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
10000
```

[71]:
```
with open('T5_Negentities_org.pkl', 'wb') as f:
    pickle.dump(T5_Negentities_org, f)
with open('T5_Negentities_person.pkl', 'wb') as f:
```

```
    pickle.dump(T5_Negentities_person, f)
with open('T5_Negentities_loc.pkl', 'wb') as f:
    pickle.dump(T5_Negentities_loc, f)
```

[95]:
```
counter = Counter()
for sublist in T5_Negentities_org:
    for item in set(sublist):
        counter[item] += 1

top_20_T5_Negentities_org = counter.most_common(20)

print('The top 20 organisations talking about Topic 5 Negatively:',␣
 ↪top_20_T5_Negentities_org)


counter = Counter()
for sublist in T5_Negentities_person:
    for item in set(sublist):
        counter[item] += 1

top_20_T5_Negentities_person = counter.most_common(20)

print('\n\nThe top 20 people talking about Topic 5 Negatively:',␣
 ↪top_20_T5_Negentities_person)


counter = Counter()
for sublist in T5_Negentities_loc:
    for item in set(sublist):
        counter[item] += 1

top_20_T5_Negentities_loc = counter.most_common(20)

print('\n\nThe top 20 locations of Topic 5 Negative news:',␣
 ↪top_20_T5_Negentities_loc)
```

```
The top 20 organisations talking about Topic 5 Negatively: [('fcc', 2576),
('wsi', 2527), ('east rock road', 2526), ('stroudsburg philadelphia jim thorpe',
2526), ('allentown pa', 2526), ('lehigh valley', 2526), ('bethlehem easton',
2526), ('programming lehigh valley berk county southeastern pa poconos western
nj', 2526), ('wdpntv retro tv', 2526), ('phillipsburg', 2526), ('wfmztv news',
2526), ('wilmington newark', 2460), ('maranatha broadcasting company inc',
2460), ('etf', 2352), ('android', 2268), ('io app', 2179), ('berk regional
school u world', 1941), ('nba', 1720), ('berk regional poconos coal region
southeastern pa western new jersey pennsylvania school u world', 1627), ('nhl',
1624)]
```

```
The top 20 people talking about Topic 5 Negatively: [('kutztown', 2526),
('wfmztv metv', 2526), ('wfmztv wfmztv', 2526), ('freddy', 1055), ('andy
serwer', 770), ('andy serwerinfluencers', 746), ('michael kelleymichael', 746),
('khemlanianjalee khemlani ben', 746), ('serwer anjalee', 746), ('etf
screeneretf screener', 710), ('wolffmannethan wolffmann', 703), ('belmonte akiko
fujitaakiko fujita', 682), ('adriana belmonteadriana', 682), ('st luke', 653),
('streamingnew', 599), ('tohow gamesgames videovideo', 589), ('smithseana
smith', 589), ('herhealth', 516), ('transactionsnew contractsprofit
warningsappointmentspress', 509), ('warren buffett', 498)]
```

```
The top 20 locations of Topic 5 Negative news: [('wdpntv', 2526), ('doylestown',
2526), ('allentown', 2526), ('china', 1383), ('america', 906), ('new york',
849), ('canada', 743), ('pennsylvania', 713), ('india', 712), ('berkshire',
626), ('japan', 595), ('california', 591), ('bethlehem', 590), ('herhealth',
556), ('uk', 529), ('russia', 520), ('syscoin', 438), ('france', 422), ('hong
kong', 403), ('london', 365)]
```

10 Topic 5 Pos Sentiment

[72]:
```
T5_Posentities_org=[]
T5_Posentities_person=[]
T5_Posentities_loc=[]

print("Length: ",len(T5_Pos))
for i in range(len(T5_Pos)):
    if i % 1000 == 0:
        print(i)
    doc = nlp(T5_Pos["text_cleaned"][i])
    entitiesO = []
    entitiesP = []
    entitiesL = []
    for ent in doc.ents:
        if ent.label_ == 'ORG':
            entitiesO.append(ent.text)
        elif ent.label_ == 'PERSON':
            entitiesP.append(ent.text)
        elif ent.label_ == 'GPE':
            entitiesL.append(ent.text)
    T5_Posentities_org.append(entitiesO)
    T5_Posentities_person.append(entitiesP)
    T5_Posentities_loc.append(entitiesL)
```

```
Length:  2638
0
```

```
1000
2000
```

[73]:
```
with open('T5_Posentities_org.pkl', 'wb') as f:
    pickle.dump(T5_Posentities_org, f)
with open('T5_Posentities_person.pkl', 'wb') as f:
    pickle.dump(T5_Posentities_person, f)
with open('T5_Posentities_loc.pkl', 'wb') as f:
    pickle.dump(T5_Posentities_loc, f)
```

[96]:
```
counter = Counter()
for sublist in T5_Posentities_org:
    for item in set(sublist):
        counter[item] += 1

top_20_T5_Posentities_org = counter.most_common(20)

print('The top 20 organisations talking about Topic 5 Positvely:',␣
 ↪top_20_T5_Posentities_org)


counter = Counter()
for sublist in T5_Posentities_person:
    for item in set(sublist):
        counter[item] += 1

top_20_T5_Posentities_person = counter.most_common(20)

print('\n\nThe top 20 people talking about Topic 5 Positvely:',␣
 ↪top_20_T5_Posentities_person)


counter = Counter()
for sublist in T5_Posentities_loc:
    for item in set(sublist):
        counter[item] += 1

top_20_T5_Posentities_loc = counter.most_common(20)

print('\n\nThe top 20 locations of Topic 5 Positve news:',␣
 ↪top_20_T5_Posentities_loc)
```

```
The top 20 organisations talking about Topic 5 Positvely: [('etf', 470), ('llc',
412), ('android', 377), ('fcc', 372), ('wsi', 362), ('east rock road', 362),
('stroudsburg philadelphia jim thorpe', 362), ('allentown pa', 362), ('lehigh
valley', 362), ('bethlehem easton', 362), ('programming lehigh valley berk
county southeastern pa poconos western nj', 362), ('wdpntv retro tv', 362),
('phillipsburg', 362), ('wfmztv news', 362), ('sec', 360), ('marketbeatcoms',
```

```
360), ('wilmington newark', 353), ('maranatha broadcasting company inc', 353),
('io app', 339), ('berk regional school u world', 319)]
```

```
The top 20 people talking about Topic 5 Positvely: [('wfmztv metv', 362),
('kutztown', 362), ('wfmztv wfmztv', 362), ('usdt', 195), ('freddy', 190),
('andy serwer', 184), ('btc filecoin fil', 182), ('andy serwerinfluencers',
181), ('michael kelleymichael', 181), ('khemlanianjalee khemlani ben', 181),
('serwer anjalee', 181), ('steth steth', 180), ('btc lido', 180),
('wolffmannethan wolffmann', 175), ('belmonte akiko fujitaakiko fujita', 172),
('etf screeneretf screener', 172), ('adriana belmonteadriana', 172), ('tohow
gamesgames videovideo', 151), ('warren buffett', 148), ('btc steth lido steth',
147)]
```

```
The top 20 locations of Topic 5 Positve news: [('wdpntv', 362), ('doylestown',
362), ('allentown', 362), ('china', 260), ('japan', 222), ('california', 194),
('america', 193), ('new york', 160), ('london', 143), ('canada', 143),
('berkshire', 142), ('india', 131), ('syscoin', 113), ('pennsylvania', 100),
('reportetf', 97), ('san mateo', 94), ('uk', 90), ('australia', 83),
('marketspremarket hour', 81), ('russia', 78)]
```

11 Topic 6 Neg Sentiment

[74]:
```
T6_Neg=Neg[Neg["topic"]==6]
T6_Neg.reset_index(inplace=True)

T6_Pos=Pos[Pos["topic"]==6]
T6_Pos.reset_index(inplace=True)
```

[75]:
```
T6_Negentities_org=[]
T6_Negentities_person=[]
T6_Negentities_loc=[]

print("Length: ",len(T6_Neg))
for i in range(len(T6_Neg)):
    if i % 1000 == 0:
        print(i)
    doc = nlp(T6_Neg["text_cleaned"][i])
    entitiesO = []
    entitiesP = []
    entitiesL = []
    for ent in doc.ents:
        if ent.label_ == 'ORG':
            entitiesO.append(ent.text)
        elif ent.label_ == 'PERSON':
            entitiesP.append(ent.text)
```

```
        elif ent.label_ == 'GPE':
            entitiesL.append(ent.text)
    T6_Negentities_org.append(entitiesO)
    T6_Negentities_person.append(entitiesP)
    T6_Negentities_loc.append(entitiesL)
```

```
Length:  20344
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
10000
11000
12000
13000
14000
15000
16000
17000
18000
19000
20000
```

[76]:
```
with open('T6_Negentities_org.pkl', 'wb') as f:
    pickle.dump(T6_Negentities_org, f)
with open('T6_Negentities_person.pkl', 'wb') as f:
    pickle.dump(T6_Negentities_person, f)
with open('T6_Negentities_loc.pkl', 'wb') as f:
    pickle.dump(T6_Negentities_loc, f)
```

[97]:
```
counter = Counter()
for sublist in T6_Negentities_org:
    for item in set(sublist):
        counter[item] += 1

top_20_T6_Negentities_org = counter.most_common(20)

print('The top 20 organisations talking about Topic 6 Negatively:',␣
 ↪top_20_T6_Negentities_org)
```

```
counter = Counter()
for sublist in T6_Negentities_person:
    for item in set(sublist):
        counter[item] += 1

top_20_T6_Negentities_person = counter.most_common(20)

print('\n\nThe top 20 people talking about Topic 6 Negatively:',␣
 ↪top_20_T6_Negentities_person)


counter = Counter()
for sublist in T6_Negentities_loc:
    for item in set(sublist):
        counter[item] += 1

top_20_T6_Negentities_loc = counter.most_common(20)

print('\n\nThe top 20 locations of Topic 6 Negative news:',␣
 ↪top_20_T6_Negentities_loc)
```

```
The top 20 organisations talking about Topic 6 Negatively: [('gray television
inc', 4046), ('music lifestylegray dc', 2947), ('u', 2543), ('sierra leone
singapore', 2507), ('estonia ethiopia falkland island', 2494), ('kansa', 2416),
('u state', 2408), ('ghana gibraltar', 2404), ('district columbia', 2403),
('mali malta maryland', 2400), ('u romania', 2400), ('u costa rica', 2400), ('u
mississippi', 2400), ('mexico michigan u', 2400), ('wisconsin u', 2400),
('iceland idaho u illinois u india indiana u', 2400), ('u moldova', 2400),
('helpsupport', 2400), ('morocco mozambique myanmar namibia', 2400), ('u
luxembourg madagascar', 2400)]


The top 20 people talking about Topic 6 Negatively: [('monaco mongolia', 2508),
('verde cayman', 2501), ('rico qatar', 2490), ('u nepal', 2400), ('jordan
kansa', 2400), ('u peru', 2400), ('turkey turkmenistan', 2400), ('greece guam
guatemala', 2400), ('captioningaudio descriptiona gray medium group inc', 1654),
('comoros congobrazzaville congokinshasa', 1587), ('greta van susterencircle
country', 1423), ('globenewswire', 713), ('anlegenwas', 681), ('blickpunktad
hocmitteilungenbestbewertete newsmeistgelesene newskonjunktur', 664), ('bevor
sie', 603), ('begriff ein produkt', 603), ('anlageentscheidung treffen', 603),
('sich mglichst umfassend', 603), ('erwerben da nicht einfach', 603), ('al ihre
befrwortung der angebotenen', 602)]


The top 20 locations of Topic 6 Negative news: [('japan', 3881), ('china',
3715), ('israel', 3311), ('california', 3279), ('australia', 3267), ('canada',
```

```
3258), ('new york', 3252), ('france', 3242), ('hong kong', 3096), ('taiwan',
3077), ('switzerland', 2979), ('poland', 2932), ('malaysia', 2932),
('netherlands', 2926), ('texas', 2922), ('spain', 2884), ('brazil', 2852),
('washington', 2849), ('russia', 2829), ('sweden', 2828)]
```

12 Topic 6 Pos Sentiment

[77]:
```
T6_Posentities_org=[]
T6_Posentities_person=[]
T6_Posentities_loc=[]

print("Length: ",len(T6_Pos))
for i in range(len(T6_Pos)):
    if i % 1000 == 0:
        print(i)
    doc = nlp(T6_Pos["text_cleaned"][i])
    entitiesO = []
    entitiesP = []
    entitiesL = []
    for ent in doc.ents:
        if ent.label_ == 'ORG':
            entitiesO.append(ent.text)
        elif ent.label_ == 'PERSON':
            entitiesP.append(ent.text)
        elif ent.label_ == 'GPE':
            entitiesL.append(ent.text)
    T6_Posentities_org.append(entitiesO)
    T6_Posentities_person.append(entitiesP)
    T6_Posentities_loc.append(entitiesL)
```

```
Length:  5956
0
1000
2000
3000
4000
5000
```

[78]:
```
with open('T6_Posentities_org.pkl', 'wb') as f:
    pickle.dump(T6_Posentities_org, f)
with open('T6_Posentities_person.pkl', 'wb') as f:
    pickle.dump(T6_Posentities_person, f)
with open('T6_Posentities_loc.pkl', 'wb') as f:
    pickle.dump(T6_Posentities_loc, f)
```

[98]:
```
counter = Counter()
for sublist in T6_Posentities_org:
    for item in set(sublist):
        counter[item] += 1

top_20_T6_Posentities_org = counter.most_common(20)

print('The top 20 organisations talking about Topic 6 Positvely:',␣
 ↪top_20_T6_Posentities_org)


counter = Counter()
for sublist in T6_Posentities_person:
    for item in set(sublist):
        counter[item] += 1

top_20_T6_Posentities_person = counter.most_common(20)

print('\n\nThe top 20 people talking about Topic 6 Positvely:',␣
 ↪top_20_T6_Posentities_person)


counter = Counter()
for sublist in T6_Posentities_loc:
    for item in set(sublist):
        counter[item] += 1

top_20_T6_Posentities_loc = counter.most_common(20)

print('\n\nThe top 20 locations of Topic 6 Positve news:',␣
 ↪top_20_T6_Posentities_loc)
```

```
The top 20 organisations talking about Topic 6 Positvely: [('endgltigen
bedingungen zu lesen', 757), ('chancen de wertpapiers sie sind im', 757),
('aktuell', 745), ('hocmitteilungennachrichten
brsenaktienempfehlungenbranchenmediennachrichtenarchivrssnews von
finanznachrichtendepresseimpressum agb disclaimer datenschutzmediadaten', 745),
('sengbranchenbranchenbersichtmedienmedienbersichtarchiverweiterte
sucheaktienkurseaktienkurserealtimeaktienkursliste', 724),
('rstungmaschinenbaumediennanotechnologienahrungsmittel', 711), ('gray
television inc', 710), ('fnalle newsrubrikenaktien i', 563), ('music
lifestylegray dc', 544), ('ml', 448), ('ibm', 390), ('datadriven', 357),
('united state', 309), ('microsoft', 295), ('gpu', 280), ('bewerten sie', 275),
('themavensterms andpolicysign upalready', 263), ('gray medium group inc station
gray television inc', 240), ('erhalten auf', 232), ('linkedin venturebeat',
225)]
```

```
The top 20 people talking about Topic 6 Positvely: [('bevor sie', 757),
('begriff ein produkt', 757), ('anlageentscheidung treffen', 757), ('sich
mglichst umfassend', 757), ('al ihre befrwortung der angebotenen', 757),
('erwerben da nicht einfach', 757), ('anlegenwas', 743), ('seitesehr
gut123456schlechtproblem meldenwerbehinweise', 709),
('tourismusimmobilienindustrie
mischkonzerneinternetitdienstleistungenkonsumgterkosmetikkunststoffe
verpackungenlogistik', 708), ('agrarnetzwerktechnikl
gaspharmarohstoffesoftwaresonstige technologietelekommunikationunterhaltungverso
rgerfondsanleihenderivaterohstoffedevisenkryptowhrungenwatchlistwatchlist',
706), ('infrastrukurbekleidung textilbiotechnologiechemiedienstleistungeneisen
stahlelektrotechnologieerneuerbare
energienfahrzeugefinanzdienstleistungenfreizeitproduktegesundheitswesengetrnke
tabakhalbleiterhandel ecommercehardwareholz', 705), ('blickpunktad
hocmitteilungenbestbewertete newsmeistgelesene newskonjunktur', 703),
('taufrufe7', 669), ('captioningaudio descriptiona gray medium group inc', 278),
('morei', 263), ('herejoin usreceive', 261), ('kyle wiggerskylelwiggers', 245),
('greta van susterencircle country', 231), ('wertpapiere zu verstehen', 225),
('jim cramerinvestingpersonal', 214)]
```

```
The top 20 locations of Topic 6 Positve news: [('india', 764), ('den
basisprospekt', 757), ('schwer', 757), ('angezeigte', 745), ('china', 733),
('japan', 620), ('singapore', 452), ('london', 404), ('australia', 402),
('france', 366), ('hong kong', 352), ('germany', 334), ('uk', 333), ('new york',
326), ('malaysia', 318), ('israel', 318), ('italy', 284), ('california', 276),
('taiwan', 274), ('america', 271)]
```

13 Topic 7 Neg Sentiment¶

[79]:
```
T7_Neg=Neg[Neg["topic"]==7]
T7_Neg.reset_index(inplace=True)

T7_Pos=Pos[Pos["topic"]==7]
T7_Pos.reset_index(inplace=True)
```

[80]:
```
T7_Negentities_org=[]
T7_Negentities_person=[]
T7_Negentities_loc=[]

print("Length: ",len(T7_Neg))
for i in range(len(T7_Neg)):
    if i % 1000 == 0:
        print(i)
    doc = nlp(T7_Neg["text_cleaned"][i])
    entities0 = []
```

```
    entitiesP = []
    entitiesL = []
    for ent in doc.ents:
        if ent.label_ == 'ORG':
            entitiesO.append(ent.text)
        elif ent.label_ == 'PERSON':
            entitiesP.append(ent.text)
        elif ent.label_ == 'GPE':
            entitiesL.append(ent.text)
    T7_Negentities_org.append(entitiesO)
    T7_Negentities_person.append(entitiesP)
    T7_Negentities_loc.append(entitiesL)
```

```
Length:  24302
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
10000
11000
12000
13000
14000
15000
16000
17000
18000
19000
20000
21000
22000
23000
24000
```

[81]:
```
with open('T7_Negentities_org.pkl', 'wb') as f:
    pickle.dump(T7_Negentities_org, f)
with open('T7_Negentities_person.pkl', 'wb') as f:
    pickle.dump(T7_Negentities_person, f)
with open('T7_Negentities_loc.pkl', 'wb') as f:
    pickle.dump(T7_Negentities_loc, f)
```

[99]:
```
counter = Counter()
for sublist in T7_Negentities_org:
    for item in set(sublist):
        counter[item] += 1

top_20_T7_Negentities_org = counter.most_common(20)

print('The top 20 organisations talking about Topic 7 Negatively:',␣
 ↪top_20_T7_Negentities_org)


counter = Counter()
for sublist in T7_Negentities_person:
    for item in set(sublist):
        counter[item] += 1

top_20_T7_Negentities_person = counter.most_common(20)

print('\n\nThe top 20 people talking about Topic 7 Negatively:',␣
 ↪top_20_T7_Negentities_person)


counter = Counter()
for sublist in T7_Negentities_loc:
    for item in set(sublist):
        counter[item] += 1

top_20_T7_Negentities_loc = counter.most_common(20)

print('\n\nThe top 20 locations of Topic 7 Negative news:',␣
 ↪top_20_T7_Negentities_loc)
```

```
The top 20 organisations talking about Topic 7 Negatively: [('ibm', 12999),
('microsoft', 3264), ('orbisresearchcom', 2816), ('intel', 2805), ('apac',
2692), ('united state', 2524), ('wordpressorg', 2407), ('samsung', 1659),
('nasa', 1650), ('nvidia', 1635), ('u uk', 1611), ('usa ind', 1265), ('canada
mexico europe', 1172), ('dallas texas usa phone usa ind', 1082), ('asean india
rest asia pacific latin america', 1024), ('ci asia pacific', 1020), ('gmbh',
1015), ('basf', 1013), ('wwwreportsandmarketscom', 1013), ('abb', 1005)]


The top 20 people talking about Topic 7 Negatively: [('costello', 2482),
('supernews acme', 1643), ('murphy', 1376), ('expresswaysuite dallastexas',
1145), ('sanjay jain', 1080), ('u orbis', 818), ('anitaadroit', 797), ('rohan',
785), ('bfsi', 724), ('sameer joshi', 715), ('baidu', 598), ('anita', 562),
('johnson johnson', 558), ('costellosenior', 554), ('albert', 534),
('brighterion kittai iflytek megvii technology', 409), ('galus', 401), ('topdown
```

```
bottomup', 383), ('r wordpressorg', 378), ('watson', 351)]


The top 20 locations of Topic 7 Negative news: [('japan', 13272), ('china',
13205), ('india', 10796), ('germany', 9183), ('brazil', 9000), ('france', 8963),
('uk', 8037), ('east africa', 7220), ('russia', 6431), ('italy', 6395),
('america', 5743), ('australia', 4988), ('spain', 4902), ('argentina', 4838),
('south africa', 4648), ('south korea', 4363), ('egypt', 4360), ('korea', 4100),
('saudi arabia', 3662), ('indonesia', 2915)]
```

14 Topic 7 Pos Sentiment¶

[82]:
```
T7_Posentities_org=[]
T7_Posentities_person=[]
T7_Posentities_loc=[]

print("Length: ",len(T7_Pos))
for i in range(len(T7_Pos)):
    if i % 1000 == 0:
        print(i)
    doc = nlp(T7_Pos["text_cleaned"][i])
    entitiesO = []
    entitiesP = []
    entitiesL = []
    for ent in doc.ents:
        if ent.label_ == 'ORG':
            entitiesO.append(ent.text)
        elif ent.label_ == 'PERSON':
            entitiesP.append(ent.text)
        elif ent.label_ == 'GPE':
            entitiesL.append(ent.text)
    T7_Posentities_org.append(entitiesO)
    T7_Posentities_person.append(entitiesP)
    T7_Posentities_loc.append(entitiesL)
```

```
Length:  2518
0
1000
2000
```

[83]:
```
with open('T7_Posentities_org.pkl', 'wb') as f:
    pickle.dump(T7_Posentities_org, f)
with open('T7_Posentities_person.pkl', 'wb') as f:
    pickle.dump(T7_Posentities_person, f)
with open('T7_Posentities_loc.pkl', 'wb') as f:
    pickle.dump(T7_Posentities_loc, f)
```

[100]:
```
counter = Counter()
for sublist in T7_Posentities_org:
    for item in set(sublist):
        counter[item] += 1

top_20_T7_Posentities_org = counter.most_common(20)

print('The top 20 organisations talking about Topic 7 Positvely:',␣
 ↪top_20_T7_Posentities_org)


counter = Counter()
for sublist in T7_Posentities_person:
    for item in set(sublist):
        counter[item] += 1

top_20_T7_Posentities_person = counter.most_common(20)

print('\n\nThe top 20 people talking about Topic 7 Positvely:',␣
 ↪top_20_T7_Posentities_person)


counter = Counter()
for sublist in T7_Posentities_loc:
    for item in set(sublist):
        counter[item] += 1

top_20_T7_Posentities_loc = counter.most_common(20)

print('\n\nThe top 20 locations of Topic 7 Positve news:',␣
 ↪top_20_T7_Posentities_loc)
```

```
The top 20 organisations talking about Topic 7 Positvely: [('ibm', 1286),
('apac', 351), ('microsoft', 314), ('united state', 286), ('nasa', 251),
('intel', 239), ('u uk', 234), ('samsung', 210), ('nvidia', 189), ('stumbleupon
tumblr pinterest', 172), ('orbisresearchcom', 162), ('wordpressorg', 153),
('mirror home press release', 146), ('scitech', 145), ('gmbh', 134), ('search
business health science technology world', 125), ('healthcare', 120), ('uk
italy', 119), ('abb', 111), ('basf', 109)]


The top 20 people talking about Topic 7 Positvely: [('supernews acme', 236),
('costello', 131), ('expresswaysuite dallastexas', 91), ('sameer joshi', 79),
('marketresearchfuture', 74), ('johnson johnson', 73), ('murphy', 72), ('bfsi',
60), ('galus', 59), ('sanjay jain', 58), ('freewebsite', 58), ('peachtree street
ne', 58), ('baidu', 55), ('dragon spacex', 54), ('robert bosch gmbh', 53),
('costellosenior', 48), ('watson', 45), ('u orbis', 43), ('smes
```

```
marketsandmarkets', 42), ('anita', 40)]


The top 20 locations of Topic 7 Positve news: [('japan', 1487), ('china', 1465),
('india', 1265), ('germany', 1094), ('france', 1065), ('brazil', 966), ('uk',
899), ('east africa', 796), ('russia', 794), ('italy', 756), ('australia', 689),
('america', 644), ('south africa', 563), ('spain', 547), ('argentina', 542),
('egypt', 537), ('south korea', 506), ('korea', 500), ('saudi arabia', 425),
('indonesia', 343)]
```

[ ]:

[ ]:

[ ]: