[Code] Tweet Data Collection

Problem Solving

[1] Those who use this code in Jupyter (IPython)

- RuntimeError: This event loop is already running - asyncio event loop problem

⇒ Downgrade the version of asyncio related package

⇒ (shaun) I recommend not using the below line: a problem accured on loading Jupyer in my case

```
!pip install tornado==4.5.3 --force-reinstall --user
```

[2] Fail to run the code when crawling large data

- TimeoutError - Twitter blocked our access (cuz Twint use 10 torr VPN IP's ⇒ very suspicious..)

⇒ Just sleep the crawling code 15 min + Add resume option

```
c.Resume = "filename.txt"
twint.run.Search(c)
time.sleep(900) # sleep 15 min after twint.run.Search(c)
```

[3] Those who do not want to print out the result tweets:

⇒ Just add `Hide_output` options

```
c.Hide_output=True
```

[4] Fail to perfectly collect the tweets for each day

- Only collected 23:59:59 ~ 13:24:50 // the left part (13:24:49 ~ 0:00:00) is gone

⇒ Add the code that can check already collected parts

Code Snippet

```
import twint # python 3.6
import pandas as pd
from datetime import timedelta
from os import mkdir, path
```

```
def twint_search(dirname, searchterm, since, until, json_name, language_code):

    c = twint.Config()
    c.Search = searchterm
    c.Since = since
    c.Until = until
    c.Hide_output = True
    c.Store_json = True
    c.Output = json_name
    c.Debug = True
    c.Resume = f'{dirname}/save_endpoint/save_endpoint_{since}.txt'
    c.Lang = language_code # language option

    try:
```

```
        twint.run.Search(c)

    except (KeyboardInterrupt, SystemExit):
        raise

    except:
        print(f"Problem with {since}.")
```

```
def twint_loop(dirname, searchterm, since, until, language_code):

    # Create directory
    try:
        mkdir(dirname)
        mkdir(f'{dirname}/save_endpoint')
        print("Directory" , dirname ,  "created ")
    except FileExistsError:
        print("Directory" , dirname ,  "already exists")

    # Loop
    daterange = pd.date_range(since, until)
    for start_date in daterange:

        since= start_date.strftime("%Y-%m-%d")
        until = (start_date + timedelta(days=1)).strftime("%Y-%m-%d")

        json_name = "".join(since.split("-")) + ".json"
        json_name = path.join(dirname, json_name)

        print(f'Getting {since} ')
        twint_search(dirname, searchterm, since, until, json_name, language_code)
```

```
# RUN!!
## include "until" datetime
### language code: https://github.com/twintproject/twint/wiki/Langauge-codes

twint_loop("Japanese", 'コロナ OR 武漢', '2020-02-01', '2020-02-29', 'ja')
```