

Assignment week 2-3

ชื่อวิชา : Software Development Practic II

รหัสวิชา : 010123132

ผู้จัดทำ : นาย นนท์ช คงดี

รหัสนักศึกษา : 6201012610028

ชั้นปี : 2

อาจารย์ผู้สอน อาจารย์ ดำรงฤทธิ์ เศรษฐ์ศิริโชค

Presentation

Week 2

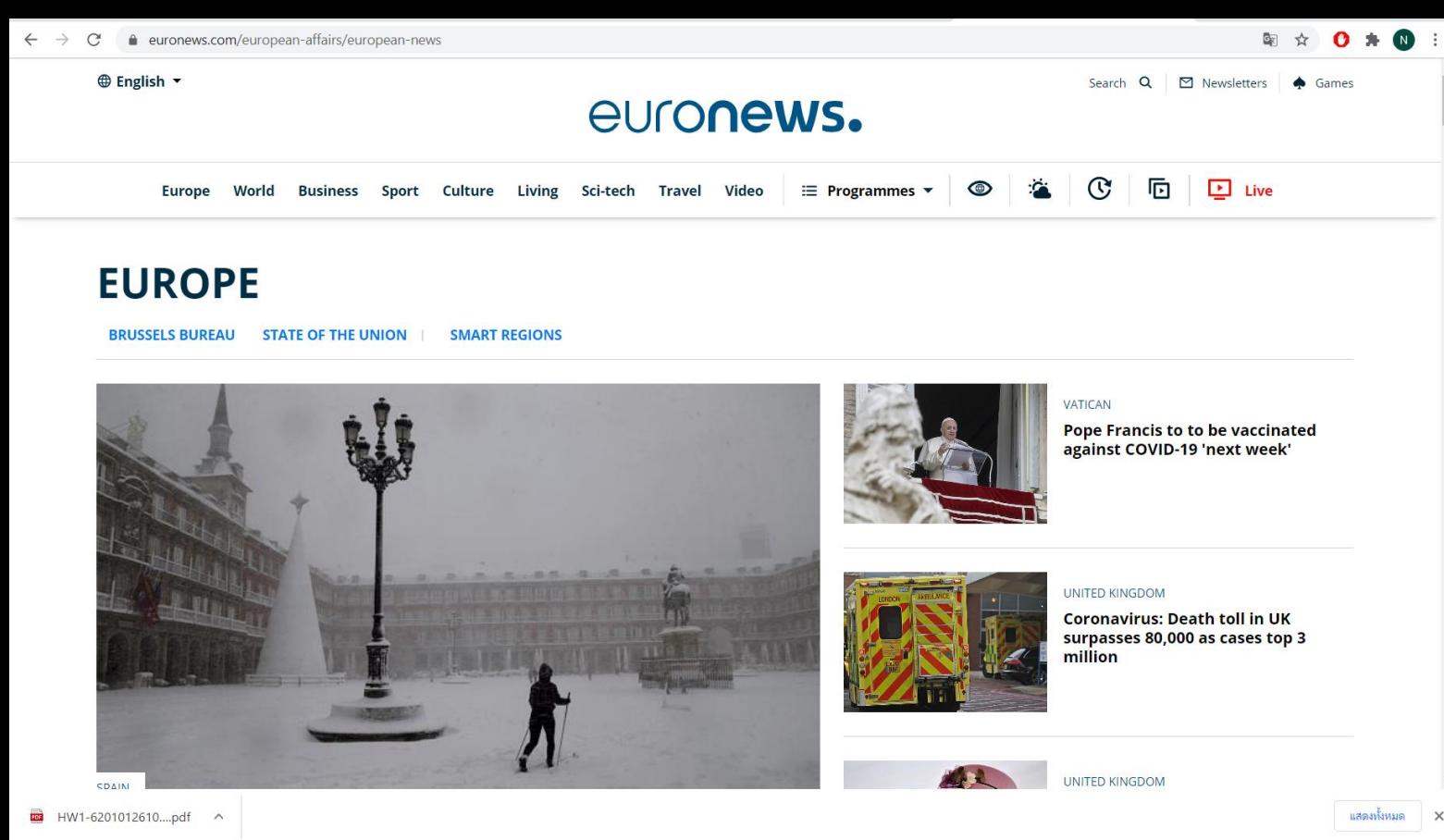
การนำหัวข้อข่าวต่างๆจาก **EURONEWS WEBSITE**

และ

หัวข้อที่พูดถึงกันมากที่สุดใน **TWITTER**

การนำหัวข้อข่าวต่างๆ จาก EURONEWS WEBSITE

การทำ
Data Scraping
HTML code
เป็น
python code



The screenshot shows the EuroNews website homepage. The header features the EuroNews logo and navigation links for English, Search, Newsletters, and Games. Below the header, there's a main menu with categories like Europe, World, Business, Sport, Culture, Living, Sci-tech, Travel, Video, Programmes, and a live video feed. The main content area is titled "EUROPE" and includes sections for BRUSSELS BUREAU, STATE OF THE UNION, and SMART REGIONS. There are several news articles with accompanying images: one about a snowy square in Madrid, another about Pope Francis being vaccinated, and one about the coronavirus death toll in the UK. The bottom of the page shows a footer with links to PDF files and a "马克思列寧學院" link.

CODE ที่เขียนได้

ใช้ **python** เขียน

Module ที่ใช้

- **requests**
- **BeautifulSoup**

```
1 import requests
2 from bs4 import BeautifulSoup
3
4 url = "https://www.euronews.com/european-affairs/european-news" # ตัวแปรสำหรับเก็บ link ของเว็บข่าว euronews
5 web = requests.get(url) # เป็นการรับ link เข้ามาใน python โดยใช้ module requests เว็บมาช่วย
6 soup = BeautifulSoup(web.content, 'html.parser') # เป็นการจัดเรียงข้อมูลที่ได้มาให้ดีขึ้นโดย module BeautifulSoup และ method สำหรับใช้หาข้อมูลเฉพาะได้
7
8 head = soup.find_all("a", {"rel" : "bookmark"}) # ต้นเหตุข้อมูลเฉพาะที่ต้องการ
9
10 news = []
11 for data in head :
12     temp = data.get('title') # นำชุดข้อมูลที่หามาแล้วมาดูเวลาเฉพาะส่วนที่เป็นหัวข้อข่าวที่จะดึง
13     if temp in news : # เป็นการป้องกันการเก็บหัวข้อซ้ำๆลงใน list news
14         pass
15     else :
16         news.append(temp)
17
18 for present in news :
19     print(present) # นำข้อมูลของหัวข้อข่าวที่คัดกรองมาแสดง
```

output

Is Serbia manipulating data to cover up its air pollution problem?
French winemakers 'stand to lose nearly €1 billion to US tariffs'
Golf is a giant board game damaging the planet: Time for it to go
Indonesia: Navy has located Sriwijaya Air Boeing 737 passenger plane crash site
WHO warns of 'vaccine nationalism' as the world rolls out COVID-19 immunisation jabs
'Attack on democracy': European leaders express shock at storming of US Capitol
Romania to introduce law against anti-Roma hate speech
Belgium to launch new coronavirus vaccine strategy amid criticism for a slow roll-out
This cross-border cycle path links the Atlantic Pyrenees to towns on the Spanish side of the border
Fresh tensions in Eastern Mediterranean after Greek and Turkish coast guard boats collide
Ukrainian President Zelenskyy calls for investigation into alleged COVID-19 vaccine 'smuggling'
Poland charges two suspects for preparing 'terror attack' on mosque
Spain records coldest ever temperature at -35.8°C
Europe at a 'tipping point' in COVID-19 pandemic as cases surge, WHO says
John Dillerman: Danish children's TV show gets mixed reviews of protagonist with a giant penis
Coronavirus: Latvia has sacked its health minister and the PM's interim choice said 'Né' - what now?
Swedish chief quits after ignoring his own agency's COVID holiday advice
France's bars, restaurants and ski resorts fear extension of COVID-19 restrictions
At least 200,000 ducks culled in France amid bird flu outbreak
Germany to set quota for women in top companies' boardrooms as progress stalls
Germany defends COVID vaccine rollout amid criticism it's too slow
EU's health chief defends COVID vaccine rollout after criticism
Nicola Sturgeon tells Donald Trump he is not allowed to come to Scotland to play golf
Italy's schools set to reopen as NGOs highlight risks of remote learning
Julian Assange: WikiLeaks founder denied bail by UK court due to continued concerns he could abscond
COVID vaccine: Moderna jab given the green light by European Medicines Agency

ส่วนขอเบย์โค้ด

ในส่วนของ code ใน line ที่ 4 และ 5
จะเป็นการนำ link ของ webside ข้าว
ที่ต้องการลงใน Python
ผ่าน request module

```
4 url = "https://www.euronews.com/european-affairs/european-news"  
5 web = requests.get(url) # เป็นการรับ link เข้าม
```

ในส่วนของ code ใน line ที่ 6 และ 8
จะเป็นการนำ BeautifulSoup module
นำมาช่วยในการเลือกข้อมูล
จาก web ผ่าน HTML โดยใช้
method ที่มีชื่อว่า find_all
ช่วยในการเลือก code
ส่วนที่เป็นหัวข้อของข่าว

```
6 soup = BeautifulSoup(web.content, 'html.parser')  
7  
8 head = soup.find_all("a", {"rel" : "bookmark"})
```

ສ່ວນອອົບປາຍໂຄດ

ໃນສ່ວນຂອງ code ໃນ line ທີ່ 10 ຊົ່ງ 19 ຈະເປັນການນຳ code ສ່ວນທີ່ເປັນຫົວໜ້ວຂອງໝາວຈາກຕົວແປຣທີ່ມີຊື່ວ່າ head ນຳມາແຍກບອກເລີພະສ່ວນທີ່ເປັນຫຼື້ອຂອງຫົວໜ້ວຂອງໝາວສ່ວນສາເຫດຖື່ຕ້ອງໃຊ້ loop ນັ້ນເພຣະຕົວແປຣ head ເກີນ HTML code ທີ່ເປັນສ່ວນຂອງຫົວໜ້ວຂອງໝາວຕ່າງ ຖໍ່ຈຶ່ງຕ້ອງແຍກສ່ວນຂອງຫຼື້ອຫົວໜ້ວຂອງໝາວອອກຈາກ HTML code ທີ່ລະໝາວແລະຕັ້ງເງື່ອນໄຂ if – else ເພື່ອປົ້ນກັນການໃຊ້ຫຼື້ອຫົວໜ້ວໜ້າແລະນຳຫຼື້ອຂອງຫົວໜ້ວໜ້າທີ່ໄດ້ມາເກີນ ໃນຕົວແປຣມີຊື່ວ່າ news ໃນ line ທີ່ 10 ສໍາຮັບແສດງຂອງມາເປັນ output ໃນ line ທີ່ 18 - 19

```
10 news = []
11 for data in head :
12     temp = data.get('title')
13     if temp in news :    # ເປົ້ນ
14         pass
15     else :
16         news.append(temp)
17
18 for present in news :
19     print(present) # ນໍາຫຼື້ອມູນຂອງ
```

การนำ HASHTAG จาก TWITTER มาวิเคราะห์

ในส่วนของ
GUI จะ
ใช้ **PyQt5**

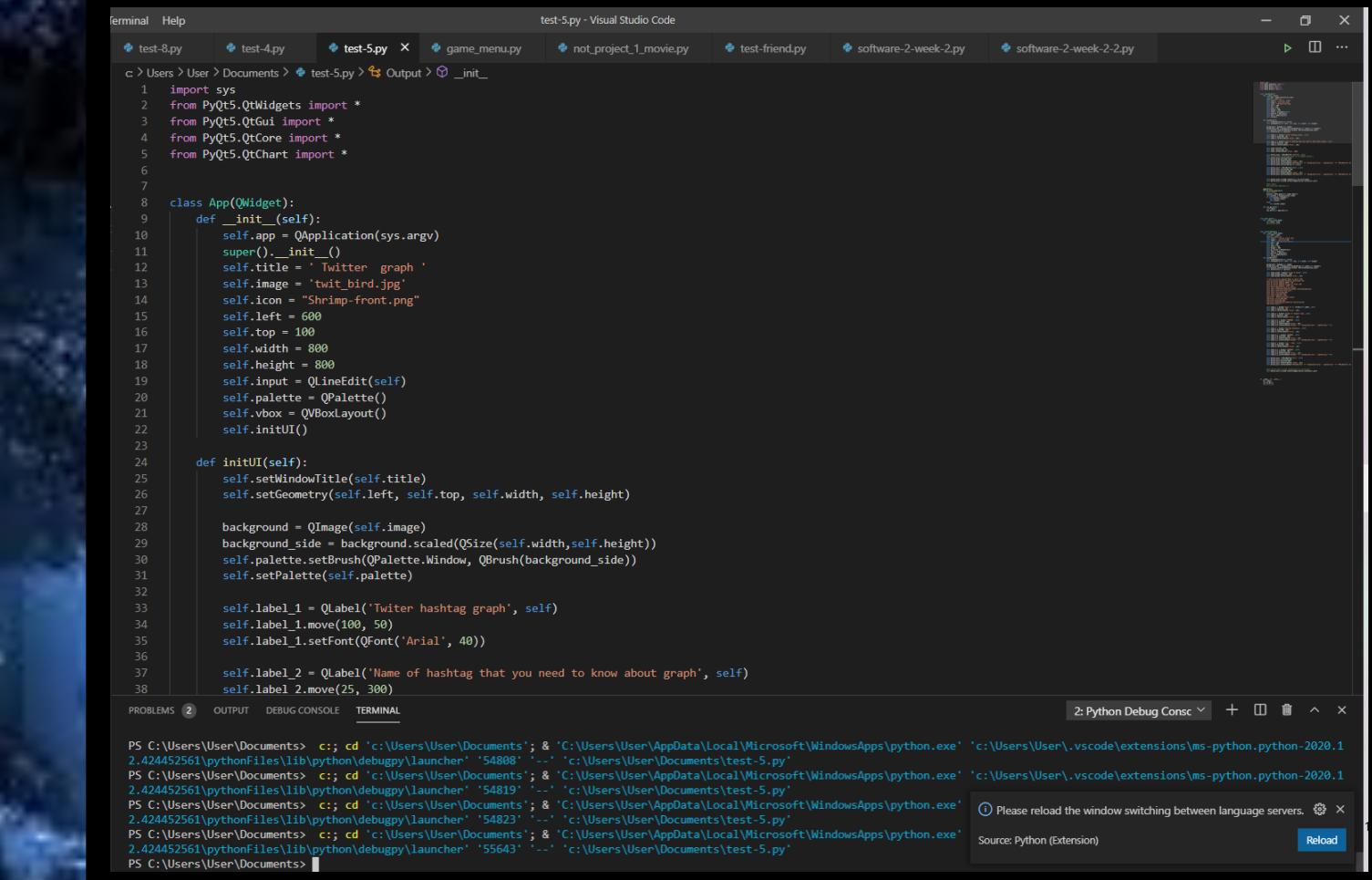
```
1 import sys
2 from PyQt5.QtWidgets import *
3 from PyQt5.QtGui import *
4 from PyQt5.QtCore import *
5 from PyQt5.QtChart import *
```

CODE ที่เขียนได้

โดยจะมี 2 ส่วนดังนี้

▪ ส่วนของหน้า
กรอก **input**

▪ ส่วนของ
การแสดง **output**



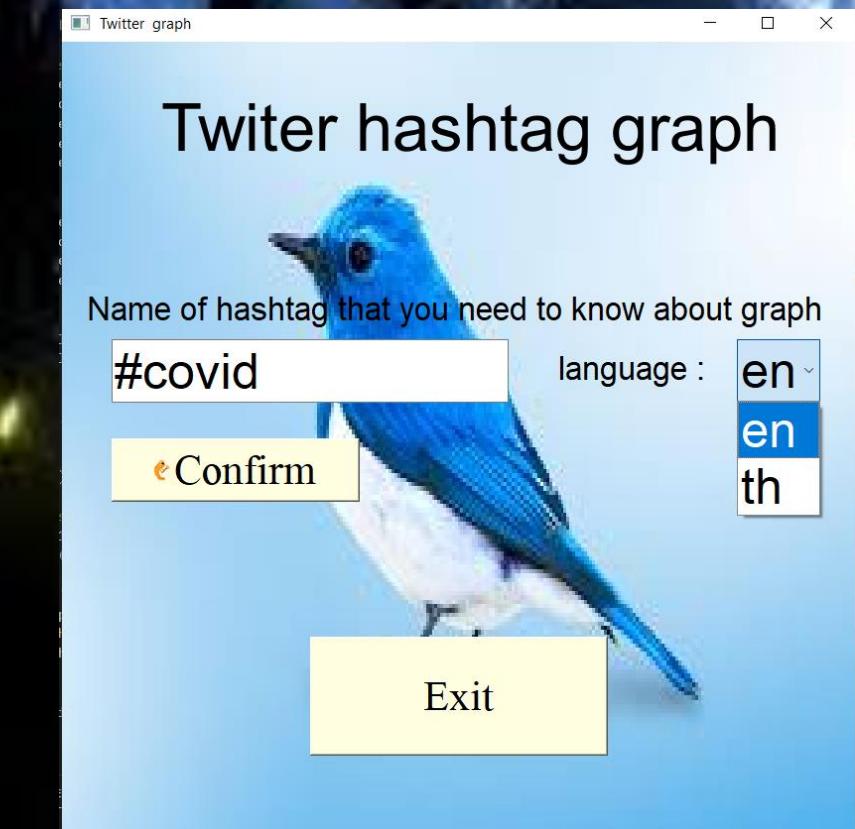
The screenshot shows a Visual Studio Code interface with a dark theme. The main area displays Python code for a PyQt5 application. The code defines a class `App` that inherits from `QWidget`. It initializes the application, sets the window title to 'Twitter graph', and specifies dimensions. It also sets a background image and palette, and adds two labels: `label_1` and `label_2`. The terminal below shows the command to run the script and its execution.

```
terminal Help test-5.py - Visual Studio Code
test-8.py test-4.py test-5.py game_menu.py not_project_1_movie.py test-friend.py software-2-week-2.py software-2-week-2-2.py
c > Users > User > Documents > test-5.py > Output > __init__
1 import sys
2 from PyQt5.QtWidgets import *
3 from PyQt5.QtGui import *
4 from PyQt5.QtCore import *
5 from PyQt5.QtChart import *
6
7
8 class App(QWidget):
9     def __init__(self):
10         super().__init__()
11         self.app = QApplication(sys.argv)
12         self.title = 'Twitter graph'
13         self.image = 'twit_bird.jpg'
14         self.icon = "Shrimp-front.png"
15         self.left = 600
16         self.top = 100
17         self.width = 800
18         self.height = 800
19         self.input = QLineEdit(self)
20         self.palette = QPalette()
21         self.vbox = QVBoxLayout()
22         self.initUI()
23
24     def initUI(self):
25         self.setWindowTitle(self.title)
26         self.setGeometry(self.left, self.top, self.width, self.height)
27
28         background = QImage(self.image)
29         background_side = background.scaled(QSize(self.width, self.height))
30         self.palette.setBrush(QPalette.Window, QBrush(background_side))
31         self.setPalette(self.palette)
32
33         self.label_1 = QLabel('Twitter hashtag graph', self)
34         self.label_1.move(100, 50)
35         self.label_1.setFont(QFont('Arial', 40))
36
37         self.label_2 = QLabel('Name of hashtag that you need to know about graph', self)
38         self.label_2.move(25, 300)

PROBLEMS ② OUTPUT DEBUG CONSOLE TERMINAL
2: Python Debug Consc + □ ^ x
PS C:\Users\User\Documents> c; cd 'c:\Users\User\Documents'; & 'C:\Users\User\AppData\Local\Microsoft\WindowsApps\python.exe' 'c:\Users\User\.vscode\extensions\ms-python.python-2020.12.424452561\pythonFiles\lib\python\debugpy\launcher' '54808' ... 'c:\Users\User\Documents\test-5.py'
PS C:\Users\User\Documents> c; cd 'c:\Users\User\Documents'; & 'C:\Users\User\AppData\Local\Microsoft\WindowsApps\python.exe' 'c:\Users\User\.vscode\extensions\ms-python.python-2020.12.424452561\pythonFiles\lib\python\debugpy\launcher' '54819' ... 'c:\Users\User\Documents\test-5.py'
PS C:\Users\User\Documents> c; cd 'c:\Users\User\Documents'; & 'C:\Users\User\AppData\Local\Microsoft\WindowsApps\python.exe' 'c:\Users\User\Documents\test-5.py'
PS C:\Users\User\Documents> c; cd 'c:\Users\User\Documents'; & 'C:\Users\User\AppData\Local\Microsoft\WindowsApps\python.exe' 'c:\Users\User\Documents\test-5.py'
PS C:\Users\User\Documents> c; cd 'c:\Users\User\Documents'; & 'C:\Users\User\AppData\Local\Microsoft\WindowsApps\python.exe' 'c:\Users\User\Documents\test-5.py'

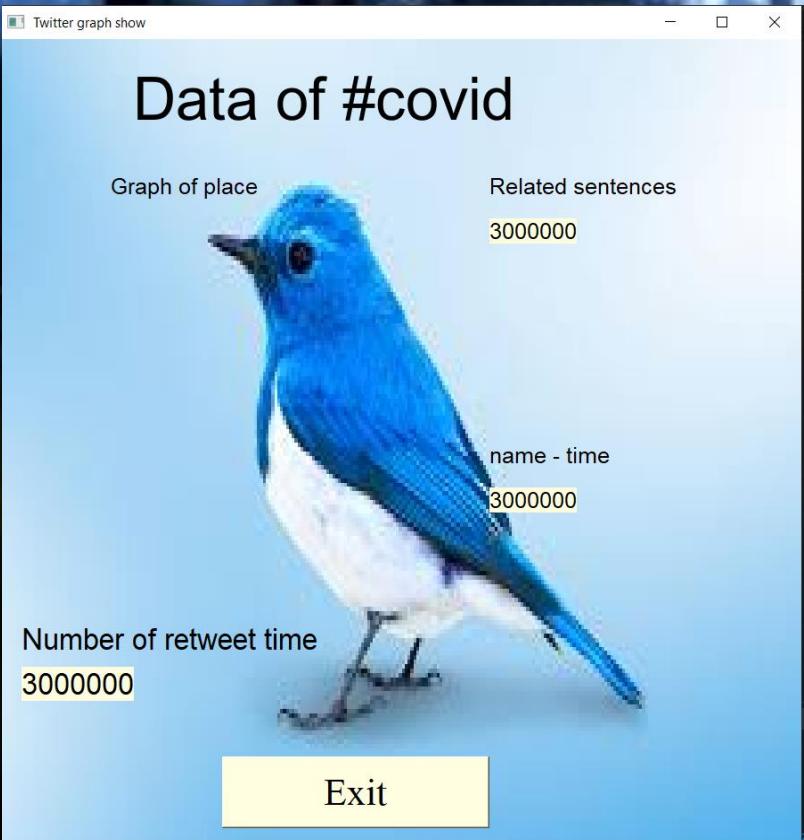
① Please reload the window switching between language servers. ⚙ x
Source: Python (Extension) Reload
```

ស៊ុនខែង input GUI



```
8  class App(QWidget):
9      def __init__(self):
10         self.app = QApplication(sys.argv)
11         super().__init__()
12         self.title = 'Twitter graph'
13         self.image = 'twit_bird.jpg'
14         self.icon = "Shrimp-front.png"
15         self.left = 600
16         self.top = 100
17         self.width = 800
18         self.height = 800
19         self.input = QLineEdit(self)
20         self.palette = QPalette()
21         self.vbox = QVBoxLayout()
22         self.initUI()
23
24     def initUI(self):
25         self.setWindowTitle(self.title)
26         self.setGeometry(self.left, self.top, self.width, self.height)
27
28         background = QImage(self.image)
29         background_side = background.scaled(QSize(self.width, self.height))
30         self.palette.setBrush(QPalette.Window, QBrush(background_side))
31         self.setPalette(self.palette)
32
33         self.label_1 = QLabel('Twiter hashtag graph', self)
34         self.label_1.move(100, 50)
35         self.label_1.setFont(QFont('Arial', 40))
36
37         self.label_2 = QLabel('Name of hashtag that you need to know about graph', self)
38         self.label_2.move(25, 300)
39         self.label_2.setFont(QFont('Arial', 20))
40
41         self.input.move(50, 350)
42         self.input.resize(400, 64)
43         self.input.setFont(QFont('Arial', 30))
44
45         self.button_play = QPushButton('Confirm', self)
46         self.button_play.setToolTip('This is an example button')
47         self.button_play.resize(256, 64)
48         self.button_play.move(500, 350)
49         self.button_play.setFont(QFont('Times', 25))
50         self.button_play.setStyleSheet("QPushButton" "{ "background-color : lightyellow; }" "QPushButton::hover"
51         self.button_play.setIcon(QIcon(self.icon))
52
53         self.button_exit = QPushButton('Exit', self)
54         self.button_exit.resize(300,120)
```

ສ່ວນຂອງ output GUI



```
92 class Output(QWidget) :
93     def __init__(self,name):
94         self.name = name
95         super().__init__()
96         self.title = ' Twitter graph show '
97         self.image = 'twit_bird.jpg'
98         self.icon = "Shrimp-front.png"
99         self.left = 600
100        self.top = 100
101        self.width = 900
102        self.height = 900
103        self.pi_series = QPieSeries()
104        self.chart = QChart()
105        self.palette = QPalette()
106        self.vbox = QVBoxLayout()
107        self.initUI()
108    def initUI(self):
109        self.setWindowTitle(self.title)
110        self.setGeometry(self.left, self.top, self.width, self.height)
111
112        background = QImage(self.image)
113        background_side = background.scaled(QSize(self.width,self.height))
114        self.palette.setBrush(QPalette.Window, QBrush(background_side))
115        self.setPalette(self.palette)
116
117        self.label_graph = QLabel('Graph of place', self)
118        self.label_graph.move(125, 150)
119        self.label_graph.setFont(QFont('Arial', 15))
120
121        """self.pi_series.append("Made in Abyss",90)
122        self.pi_series.append("Promises Neverland",70)
123        self.pi_series.append("ญี่ปุ่น",50)
124        self.pi_series.append("Attack on Titan",80)
125        self.pi_series.append("วิถี",30)
126        self.chart.addSeries(self.pi_series)
127        self.chart.setAnimationOptions(QChart.SeriesAnimations)
128        self.chart.setTitle("Fuck you")
129        self.chart.resize(100,100)
130        self.chart.move(100,100)
131        self.chart.legend().hide()
132        chartview = QChartView(self.chart)
133        chartview.resize(100,100)
134        chartview.move(100,100)
135        chartview.setRenderHint(QPainter.Antialiasing)
136        chartview.show()"""
137
138        self.label_1 = QLabel('Data of {}'.format(self.name), self)
```

ส่วนของ Twitter APIs

Module ที่ใช้



- CSV
- tweepy

The screenshot shows a Gmail inbox with several unread emails. One specific email is highlighted, which is from developer.twitter.com. The subject line reads "Your Twitter developer account application has been approved!". The body of the email contains a message of thanks for applying for access, stating that their review has been completed and their request has been approved. It encourages the user to sign in to their developer account to get started and expresses appreciation for building on Twitter. At the bottom, there is a footer with the URL "developer.twitter.com | @twitterdev" and the address "Twitter, Inc. 1355 Market Street, San Francisco, CA 94103".

The screenshot shows the Twitter Developer Portal dashboard. On the left, there is a sidebar with options for "Projects & Apps", "Products", and "Account". The main area is titled "Dashboard" and features a section for "Project 1" with "V1.1 ACCESS" and "V2 ACCESS" buttons. Below this, there is a chart for "MONTHLY TWEET CAP USAGE" showing 0 Tweets pulled of 500,000, with a note that it resets on February 13 at 00:00 UTC. Another section labeled "PROJECT APP" shows "Master.E app". To the right, there is a "New endpoints!" card with the text "We're rebuilding the Twitter API and have a new set of Early Access endpoints." and a "View endpoints" button. At the bottom, there are links for "Helpful docs", "Docs homepage", and "API reference index".

ส่วนของ database หรือ Twitter APIs

Class ของการดึงข้อมูลจาก twitter จะทำหน้าที่เป็น class นำข้อมูลจาก twitter มาจัดเรียง

```
89  class Data_twitter():
90      def __init__(self,keyword):
91          self.keyword = keyword
92          self.language = "en"
93          self.tweet_mode = "extended"
94          self.result_type = "mixed"
95          self.count = 100
96
97          self.tweet_key = [REDACTED]
98          self.tweet_secret_key = [REDACTED]
99          self.tweet_token = [REDACTED]
100         self.tweet_secret_token = [REDACTED]
101
102         self.auth = tweepy.OAuthHandler(self.tweet_key, self.tweet_secret_key)
103         self.auth.set_access_token(self.tweet_token, self.tweet_secret_token)
104
105         self.api = tweepy.API( self.auth,wait_on_rate_limit=True,wait_on_rate_limit_notify=True)
106         self.make_file = open('twitter_all_data.csv', 'w', newline='', encoding='utf-8')
107         fieldnames = ['places', 'time', 'tweet']
108         self.writer = csv.DictWriter( self.make_file, fieldnames=fieldnames )
109         self.writer.writeheader()
110
111     def get_data(self):
112         count_limit = 10
113         count = 0
114         maxip = 0
115         while count < count_limit :
116             try :
117                 data = self.api.search(q= self.keyword,
118                                         lang= self.language,
119                                         count= self.count,
120                                         tweet_mode= self.tweet_mode,
121                                         result_type= self.result_type,
122                                         max_id= str(maxip - 1))
123                 self.file_manage(data)
124                 if len(data) == 0 :
125                     continue
126
127                 maxip = data[-1].id
128                 count += 1
129
130             except tweepy.RateLimitError :
131                 if input("Do you want quit? y/n? ") == "y" :
132                     break
133                 else:
134                     pass
135
136             print("Done",count)
137
138     def file_manage(self,data):
139         for information in data:
140             if( (not information.retweeted) and ("RT @" not in information.full_text) ):
141                 self.writer.writerow( { 'places': information.user.location, 'time' : str(information.created_at), 'tweet' :information.full_text } )
```

CODE

ที่เขียนได้

โดยจะมี 2 ส่วนดังนี้

▪ ส่วนของการสร้าง API และสร้างไฟล์ excel ไว้เก็บข้อมูลที่เรียก

▪ ส่วนของการดึงข้อมูลมาจากการ twitter และนำ

```
89 class Data_twitter():
90     def __init__(self,keyword):
91         self.keyword = keyword
92         self.language = "en"
93         self.tweet_mode = "extended"
94         self.result_type = "mixed"
95         self.count = 100
96
97         self.tweet_key = [REDACTED]
98         self.tweet_secret_key = [REDACTED]
99         self.tweet_token = [REDACTED]
100        self.tweet_secret_token = [REDACTED]
101
102        self.auth = tweepy.OAuthHandler(self.tweet_key, self.tweet_secret_key)
103        self.auth.set_access_token(self.tweet_token, self.tweet_secret_token)
104
105        self.api = tweepy.API( self.auth,wait_on_rate_limit=True,wait_on_rate_limit_notify=True)
106        self.make_file = open('twitter_all_data.csv', 'w', newline='', encoding='utf-8')
107        fieldnames = ['places', 'time', 'tweet']
108        self.writer = csv.DictWriter( self.make_file, fieldnames=fieldnames )
109        self.writer.writeheader()
```

```
111    def get_data(self):
112        count_limit = 10
113        count = 0
114        maxip = 0
115        while count < count_limit :
116            try :
117                data = self.api.search(q= self.keyword,
118                                      lang= self.language,
119                                      count= self.count,
120                                      tweet_mode= self.tweet_mode,
121                                      result_type= self.result_type,
122                                      max_id= str(maxip - 1))
123                self.file_manage(data)
124                if len(data) == 0 :
125                    continue
126
127                maxip = data[-1].id
128                count += 1
129
130            except tweepy.RateLimitError :
131                if input("Do you want quit? y/n? ") == "y" :
132                    break
133                else:
134                    pass
135
136            print("Done",count)
137
138    def file_manage(self,data):
139        for information in data:
140            if( (not information.retweeted) and ("RT @" not in information.full_text) ):
141                self.writer.writerow( {'places':information.user.location, 'time': str(information.created_at), 'tweet':information.full_text} )
142
```

สอนขั้นตอนการนำข้อมูลจาก GUI มาเชื่อมกับ Twitter APIs

```
//  
78     @pyqtSlot()  
79     def on_click_play(self):  
80         #self.hide()  
81         #process = Data_base(self.input.text())  
82         run = Data_twitter(self.input.text(),self.choice_1.currentText())  
83         run.get_data()  
84         #read_file = open('twitter_all_data.csv', 'w', newline='', encoding='utf-8')  
85         self.another = Output(self.input.text())  
86         if self.another.isHidden():  
87             self.another.show()  
88             self.close()  
89         else:  
90             self.another.hide()  
91
```

Output ที่ออกมาจะเป็นตาราง

Screenshot of Microsoft Excel showing a table titled "twitter_all_data 2". The table has three columns: "places", "time", and "tweet". The "places" column contains various locations, and the "time" column shows dates and times. The "tweet" column displays tweets corresponding to each location and time. Three specific rows are highlighted with purple boxes and arrows pointing to them:

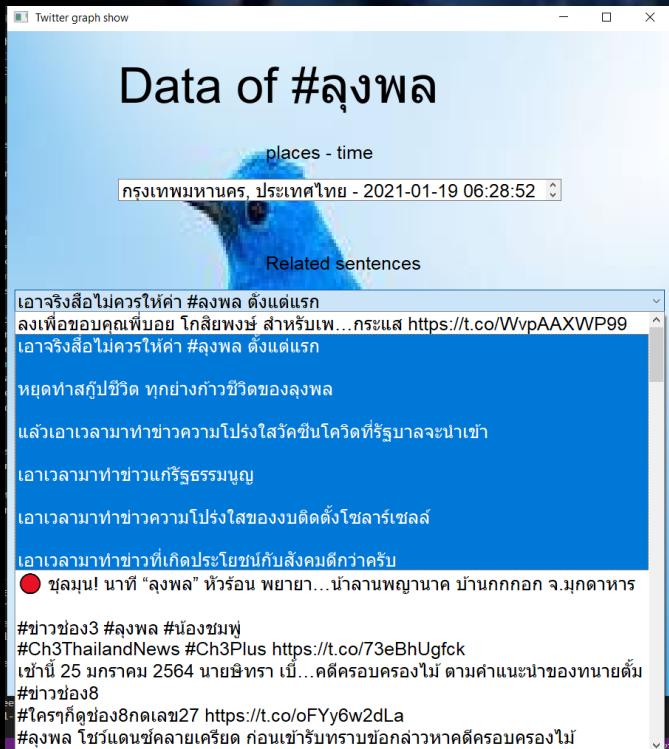
- Row 23: "Los Angeles, California" in the "places" column.
- Row 24: "France" in the "places" column.
- Row 25: "2606 W. Jefferson St. Joliet" in the "places" column.

The table continues with more rows of data below these highlights.

	places	time	tweet
1	places		
2	Washington, D.C.	1/16/2021 13:21	#Covid-19 READ: US Gov 🇺🇸 has reason to believe that several researchers inside the WIV (Wuhan Institute of Virology) became sick in autumn 2019. They now report 29 post-Covid mRNA vaccine deaths, a rate of 1 per 1,000 vaccinations - 10000x reported
3	New York	1/16/2021 19:21	UPDATE: Norway is growing seriously concerned. They now report 29 post-Covid mRNA vaccine deaths, a rate of 1 per 1,000 vaccinations - 10000x reported
4	London	1/17/2021 9:47	We - @IndependentSage - call for clarity on UK's COVID strategy in the Sunday Times. We can't just depend on the vaccine, important though it is https://t.co/
5	San Diego, California	1/18/2021 6:12	@BillFOXLA @LASDHQ @FOXLA One would figure LASD would not send anyone since there is an increase of violent crime throughout the county, but who
6		1/18/2021 6:12	#covid @BorisJohnson @Keir_Starmer Boris you berated keir for abstaining in votes but you are instructing your party to do the same how is that any different
7	East coast USA	1/18/2021 6:10	@meetlauderdale Covid hot spot.
8		1/18/2021 6:10	Coco Gauff gets it #WeDidntDoTheWork COVID https://t.co/Oa0Hp7vuqg
9	Punjab	1/18/2021 6:10	https://t.co/j2AxUq6Sp7
10	Punjab	1/18/2021 6:10	https://t.co/PLk6hOmry8
11	Indianapolis, IN	1/18/2021 6:10	https://t.co/Qp2RkNrWcD https://t.co/lL3XcriyD
12	Chicago, IL	1/18/2021 6:09	#yesecart #pulmonologist #respiratorycare #pulmonary #lungs #respiratorytherapy #covid #lunghealth #medicalscience #nurselife #medicaldevices #pulm
13		1/18/2021 6:09	Tyree (now 2-2) scored 692.9 points this week and somehow knocked off the previously unbeaten Dylan (now 3-1). That is a BL 2.0 record for least amount
14		1/18/2021 6:09	delta. https://t.co/0xks9nCzR
15	United Kingdom	1/18/2021 6:09	This Is the Best News Yet for Moderna's COVID Vaccine - The Motley Fool #health #covid-19 #coronavirus https://t.co/WcmVvbtVo4 https://t.co/QjnaMeD
16	United Kingdom	1/18/2021 6:09	Jameela Jamil wants influencers to stop promoting the keto diet - https://t.co/8MssxqjNEm #health #covid-19 #coronavirus https://t.co/kGkZjeVlkC https://t.co/
17	United Kingdom	1/18/2021 6:09	Sure Signs You've Already Had COVID, According to Experts - Yahoo Lifestyle #health #covid-19 #coronavirus https://t.co/XPOZxEi2pY https://t.co/6zfsKS2O
18	United Kingdom	1/18/2021 6:09	Sure Signs You've Already Had COVID, According to Experts Eat This Not That - Eat This, Not That #health #covid-19 #coronavirus https://t.co/7CpyfdiVd9
19	United Kingdom	1/18/2021 6:09	Highly transmissible COVID-19 variant first found in U.K. could be dominant in U.S. by March, CDC says - CBS News #health #covid-19 #coronavirus https://t.co/
20	United Kingdom	1/18/2021 6:09	Companies Race to Develop Drugs That Stay Ahead of Coronavirus Mutations - The Wall Street Journal #health #covid-19 #coronavirus https://t.co/YWHSf
21		1/18/2021 6:08	Guidelines you need to follow to stay safe, help reduce the spread of coronavirus (#COVID-19) and protect others during the celebrations. #COVID-19 precautions
22	United Kingdom	1/18/2021 6:08	#china is the only country in the world that has increased its economy despite #covid
23	Los Angeles, California	1/18/2021 6:08	I'm a bit reassured today. Apparently she's having no trouble breathing.
24	France	1/18/2021 6:08	#Nigeria: The House of Representatives has called on the federal government to postpone the reopening of schools by three months due to the increasing numbers
25	2606 W. Jefferson St. Joliet	1/18/2021 6:07	According to the most recent contact tracing data- 75% of #covid spread occurs in households, 8% in medical facilities, 2% on college campuses. Over 85% of
26	Polska / Poland	1/18/2021 6:07	How a Baltimore #design initiative is rethinking #cityscapes to promote economic #recovery from #COVID-19 https://t.co/y8OMPVhnfa
27		1/18/2021 6:06	#covid #corona #COVID19 #COVID-19 #Coronavirus #CoronaUpdate #Lockdown #StayHome #socialdistancing #pandemic #StaySafeStayHome

การแสดง Output โดย GUI

ส่วนของข้อความ tweet



ส่วนของสถานที่และเวลา

```
Done 5
tweet
dict_keys(['places', 'time', 'tweet'])
กรุงเทพมหานคร, ประเทศไทย - 2021-01-19 06:28:52
tweet
dict_keys(['places', 'time', 'tweet'])
กรุงเทพมหานคร, ประเทศไทย - 2021-01-19 06:28:52
tweet
dict_keys(['places', 'time', 'tweet'])
HOME - 2021-01-19 08:34:23
tweet
dict_keys(['places', 'time', 'tweet'])
กรุงเทพมหานคร, ประเทศไทย - 2021-01-19 06:28:52
PS C:\Users\User\Documents>
```

Code ส่วนในการนำ Output มาแสดงใน GUI

```
253     def read_file(self,block) :
254         self.dictionary = {}
255         data_file = pandas.read_csv('twitter_all_data.csv')
256         for column in data_file :
257             self.dictionary[column] = []
258             for inner in data_file[column] :
259                 self.dictionary[column].append(inner)
260
261         if block is self.show_3 :
262             for word in self.dictionary['tweet'] :
263                 self.show_3.addItem(word)
264
265         #print(dictionary.keys())
266
267
268     def sentences_support(self) :
269         data = self.show_3.currentText()
270         key_mark = self.find_keys_by_values(data)
271         print(key_mark)
272         tem_index = int()
273         for index in range(len(self.dictionary[key_mark])) :
274             if self.dictionary[key_mark][index] == data :
275                 tem_index = index
276         new_data = self.dictionary
277         #new_data.pop(key_mark)
278         print(new_data.keys())
279         result = []
280         for token,key in zip(new_data.values(),new_data.keys()) :
281             if key != key_mark :
282                 result.append(token[tem_index])
283         self.use_word = str(result[0]) + ' - ' + str(result[1])
284         print(self.use_word)
285
286
287     def find_keys_by_values(self,values) :
288         test = self.dictionary.items()
289         for item in test :
290             if values in item[1] :
291                 return item[0]
292                 break
293             else :
294                 pass
```

ส่วนของการอ่าน file
และเพิ่มข้อมูลลง GUI

```
253     def read_file(self,block) :
254         self.dictionary = {}
255         data_file = pandas.read_csv('twitter_all_data.csv')
256         for column in data_file :
257             self.dictionary[column] = []
258             for inner in data_file[column] :
259                 self.dictionary[column].append(inner)
260
261         if block is self.show_3 :
262             for word in self.dictionary['tweet'] :
263                 self.show_3.addItem(word)
264
265
266
267
268     def sentences_support(self) :
269         data = self.show_3.currentText()
270         key_mark = self.find_keys_by_values(data)
271         print(key_mark)
272         tem_index = int()
273         for index in range(len(self.dictionary[key_mark])) :
274             if self.dictionary[key_mark][index] == data :
275                 tem_index = index
276         new_data = self.dictionary
277         #new_data.pop(key_mark)
278         print(new_data.keys())
279         result = []
280         for token,key in zip(new_data.values(),new_data.keys()) :
281             if key != key_mark :
282                 result.append(token[tem_index])
283         self.use_word = str(result[0]) + ' - ' + str(result[1])
284         print(self.use_word)
285
286
287     def find_keys_by_values(self,values) :
288         test = self.dictionary.items()
289         for item in test :
290             if values in item[1] :
291                 return item[0]
292                 break
293             else :
294                 pass
```

ส่วนของการแสดง
สถานที่และเวลาจาก
ข้อความ

Output ในการถีที่ rate เกิน 180

จะไม่มีการ file update เต็ม
จะใช้ใน python แทน

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL

```
run = Data_twitter(self.input.tex
```

```
SyntaxError: invalid syntax
PS C:\Users\User\Documents> c;; cd '2.424452561\pythonFiles\lib\python\de
en
Over limit rate
Done 0
PS C:\Users\User\Documents>
```

ส่วนของ NLP

Module ที่ใช้

- spacy
- nltk
- pythainlp
- CSV
- pandas

ส่วนการสร้าง file และ main function

```
11 class NLP:  
12     def __init__(self):  
13         self.csvfile_output = open('tweet_NLP.csv', 'w', newline='', encoding="utf-8")  
14         self.input_file = pandas.read_csv('twitter_all_data.csv')  
15  
16         fieldnames = ['word', 'number']  
17         self.writer_output = csv.DictWriter( self.csvfile_output, fieldnames=fieldnames )  
18         self.writer_output.writeheader()  
19  
20  
21     def word_analysis(self, lang):  
22         dictionary_count = {}  
23         count = 0  
24  
25         for row in self.input_file['tweet']:  
26             print(count)  
27  
28             if lang == "th" :  
29                 temp = self.analyze_word_th(row)  
30             elif lang == "en" :  
31                 temp = self.analyze_word_en(row)  
32             for word in temp:  
33                 message = word.lower()  
34                 if( message not in dictionary_count ):  
35                     dictionary_count[message] = 1  
36                 elif( message in dictionary_count ):  
37                     dictionary_count[message] += 1  
38                 if count == 10:  
39                     break  
40                 count += 1  
41  
42             for temp in dictionary_count:  
43                 self.writer_output.writerow({ 'word':temp, 'number':dictionary_count[temp] })  
44  
45         self.csvfile_output.close()
```

FUNCTION ส่วนการตัดคำ

```
47 def analyze_word_en(self, data=""):
48     self.output = []
49     self.check = {}
50     self.nlp = spacy.load("en_core_web_md")
51
52     self.data = data
53     self.docs = self.nlp(self.data)
54
55     # =====
56     self.hashtag_filter()
57     self.add_filter()
58
59     for word in self.docs:
60         if ( word.text[0] != "#" and
61             word.text[0] != "@" and
62             "#" + word.text not in self.check and
63             "@" + word.text not in self.check and
64             word.text not in self.nlp.Defaults.stop_words and
65             word.text not in stopwords.words('english') and
66             not word.is_punct and
67             "https:" not in word.text and
68             self.word_type(word) ) :
69
70             self.output.append(word.text)
71
72     return self.output
73
74 def analyze_word_th(self, data):
75     proc = word_tokenize(data, engine='newmm')
76     return proc
77
```

```
80     def hashtag_filter(self):
81         matcher = Matcher(self.nlp.vocab)
82         matcher.add('HASHTAG', None, [{"ORTH': '#"}, {"IS_ASCII": True}])
83         matches = matcher(self.docs)
84         spans = []
85         for match_id, start, end in matches:
86             spans.append(self.docs[start:end])
87         for span in spans:
88             span.merge()
89         for word in self.docs:
90             if word.text[0] == "#" :
91                 self.output.append(word.text)
92                 self.check[word.text] = word.text
93
94     def add_filter(self):
95         for word in self.docs:
96             if word.text[0] == "@" :
97                 self.output.append(word.text)
98                 self.check[word.text] = word.text
99
100    def word_type(self, word):
101
102        type_word = ["ADJ", "INTJ", "NOUN", "PROPN", "VERB", "NUM"]
103
104        if word.pos_ in type_word :
105            return True
106        else:
```

ส่วนของ NLP (ภาษาไทย)

Module ที่ใช้เพิ่ม

- **re**
- **thai_stopwords**
- **pythainlp**
- **Pythainlp.corpus**
- **CSV**
- **pandas**

```
1 import pandas
2 import re
3 from pythainlp.corpus.common import thai_stopwords
4 from pythainlp import word_tokenize, Tokenizer
5 from pythainlp.corpus import *
6
7
8 stopwords = thai_stopwords()
9 input_file = pandas.read_csv('twitter_all_data.csv')
10 #print(stopwords)
11
12 for text in input_file['tweet'] :
13     text = text.replace("_", "")
14     text = text.replace(":", "")
15     text = text.replace("(", "")
16     text = text.replace(")", "")
17     text = text.replace("-", "")
18     text = text.replace("*", "")
19     text = text.replace(".", "")
20     text = text.replace("'", "")
21     text = text.replace('+', "")
22     text = text.replace('/', "")
23     text = text.replace('\n', "")
24     text = text.replace("!", "")
25     text = text.replace("?", "")
26     text = text.replace("“", "")
27     text = text.replace("”", "")
28     text = re.sub('http', '', text)
29     text = re.sub('https', '', text)
30 proc = word_tokenize(text, engine="newmm", keep_whitespace = False)
31 proc2 = [i for i in proc if i not in stopwords]
32 print(proc)
33 print(proc2)
```

ส่วนการทำงานของ Code

- ใน line ที่ 8 จะเป็นส่วนของการกำหนดตัวแปรที่เก็บ stopwords ของภาษาไทย หรือคำที่พบบ่อยในภาษาไทย
- ในส่วน line ที่ 9 จะเป็นการอ่าน file ที่เก็บข้อมูลที่ได้มาจากการ twitter

```
8 stopwords = thai_stopwords()  
9 input_file = pandas.read_csv('twitter_all_data.csv')  
10 #print(stopwords)
```

ส่วนการทำงานของ Code(ต่อ)

- ใน line ที่ 13 ถึง 29 จะเป็นส่วนของการตัดคำที่ไม่ต้องการคืออักขระพิเศษ
- ในส่วน line ที่ 30 จะเป็นการเลือกประเภท AI ที่ต้องการใช้ตัดคำและตัดส่วนที่มีตัวเป็นเว้นวรรคนั้นคือคำว่า `keep_whitespace = False`
- ส่วนใน line ที่ 31 คือการนำข้อมูลมากรองอีกทีว่ามีคำที่อยู่ใน stop word หรือไม่

```
--  
12 for text in input_file['tweet'] :  
13     text = text.replace("_", "")  
14     text = text.replace(":", "")  
15     text = text.replace("(", "")  
16     text = text.replace(")", "")  
17     text = text.replace("-", "")  
18     text = text.replace("*", "")  
19     text = text.replace(".", "")  
20     text = text.replace('\'', "")  
21     text = text.replace('+', "")  
22     text = text.replace('/', "")  
  
23     text = text.replace('\n', "")  
24     text = text.replace("!", "")  
25     text = text.replace("?", "")  
26     text = text.replace("\"", "")  
27     text = text.replace("'''", "")  
28     text = re.sub('http', ' ', text)  
29     text = re.sub('https', ' ', text)  
30     proc = word_tokenize(text, engine="newmm", keep_whitespace = False)  
31     proc2 = [i for i in proc if i not in stopwords]  
32     print(proc)  
33     print(proc2)
```

การแสดง Output

เป็นการนำข้อความจาก

twitter มาตัดทีละข้อความ

[ตีวนี้, 'ต่อง', 'ขออนุญาต', 'ก่อน', 'อย่า', 'ทบ', 'หลัง', 'พยายาม', 'จะ', 'ศรีบ', 'พยายาม', 'ลับ', '#', 'ใจ', 'ให้', 'กระแซ', 'stcoIvvpAAXWP', '99']
['ขออนุญาต', 'พี', 'น้อย', 'โภ', 'สี', 'ย', 'พชร', 'สำเร็จ', 'เพลง', 'ขออนุญาต', 'ลู', 'พด', 'สำเร็จ', 'พยายาม', '#', 'ใจ', 'ให้', 'กระแซ', 'stcoIvvpAAXWP', '99']
['ขออนุญาต', 'พี', 'น้อย', 'โภ', 'สี', 'ย', 'พชร', 'สำเร็จ', 'เพลง', 'ขออนุญาต', 'ลู', 'พด', 'สำเร็จ', 'พยายาม', '#', 'ใจ', 'ให้', 'กระแซ', 'stcoIvvpAAXWP', '99']
['เอกสาร', 'สือ', 'ໃນ', 'ผอ', 'ໃນ', 'ค่า', '#', 'ล', 'พด', 'ตั้งแต่แรก', 'หยด', 'ท', 'สปี', 'ชีด', 'หยด', 'ย่างก้าว', 'ชีด', 'ขอ', 'ลู', 'พด', 'แฉ', 'เอกสาร', 'ม',
'ท', 'ช้า', 'คำน', 'ไปช่องใส', 'วัคซีน', 'โควิด', 'ที', 'รัฐบาล', 'จะ', 'นำไปที', 'เอกสาร', 'เอกสาร', 'พด', 'ท', 'ช้า', 'แก', 'รัฐธรรมนูญ', 'เอกสาร', 'เอกสาร', 'พด', 'ช้า', 'คำน',
'ไปช่องใส', 'ช่อง', 'งบ', 'ติดตั้ง', 'ใช้ครั้', 'เชลล์', 'เอกสาร', 'เอกสาร', 'พด', 'ท', 'ช้า', 'ทีกิด', 'ประโยชน์', 'กบ', 'ลัค', 'ดีกว่า', 'ศรีบ']
['เอกสาร', 'สือ', 'ค่า', '#', 'ล', 'พด', 'ตั้งแต่แรก', 'หยด', 'ท', 'สปี', 'ชีด', 'หยด', 'ย่างก้าว', 'ชีด', 'ขอ', 'ลู', 'พด', 'เอกสาร', 'ท', 'ช้า', 'ไปช่องใส', 'วัคซีน', 'โควิด', 'รัฐบาล',
'นำไปที', 'เอกสาร', 'พด', 'ช้า', 'แก', 'รัฐธรรมนูญ', 'เอกสาร', 'พด', 'ช้า', 'ทีกิด', 'ประโยชน์', 'ลัค', 'ลัค', 'ดีกว่า', 'ศรีบ']
[●, 'ชลธุ', 'นาที', 'ล', 'พด', 'ร้อน', 'พยายาม', 'จะ', 'ขอ', 'ไมค์', 'หักก้าว', 'ก่อน', 'ปี', 'ได้เรื่อง', 'แฉ', 'เดิน', 'เข้าไป', 'ผลัก', 'ขณะ', 'กำลังจะ', 'ให้', 'ลัค',
'ชลธุ', 'ปี', 'เงินบริจาค', 'หน้า', 'ลาน', 'พยายาม', 'กอก', 'กอก', 'จ', 'มูกดาวกร', '#', 'ช้า', 'ช่อง', '3', '#', 'ลู', 'พด', 'แฉ', 'ชลธุ', '#', 'Ch', '3', 'ThailandNews'
, '#', 'Ch', '3', 'Plus', 'stco', '73', 'eBhUgfck']
[●, 'ชลธุ', 'นาที', 'ล', 'พด', 'ร้อน', 'ไมค์', 'หักก้าว', 'ปี', 'ได้เรื่อง', 'เดิน', 'เข้าไป', 'ผลัก', 'ลัค', 'พยายาม', 'ปี', 'เงินบริจาค', 'หน้า', 'ลาน', 'พยายาม', 'หน้า', 'กอก',
'กอก', 'จ', 'มูกดาวกร', '#', 'ช้า', 'ช่อง', '3', '#', 'ลู', 'พด', '#', 'แฉ', 'ชลธุ', '#', 'Ch', '3', 'ThailandNews', '#', 'Ch', '3', 'Plus', 'stco', '73', 'eBhUgfck']
['ชีคาก', 'ไคร', 'ท', 'น้อง', 'ชลธุ', 'ก้าว', 'เฉลย', 'เฉพาะ', 'กบ', 'น้ำที', 'คดี', 'น้อง', 'ชลธุ', 'คดี', 'ไคร', '#', 'นาก', '2', '#', 'พญานาค', '#', 'นาก', '#', 'ลู'
, 'พด', '#', 'ล', 'พด', 'ปี', 'เด่น', 'stcogvRA', '0', 'jb', '4', 'ec']
['ชีคาก', 'ช', 'น้อง', 'ชลธุ', 'เฉลย', 'เฉพาะ', 'น้ำที', 'คดี', 'น้อง', 'ชลธุ', '#', 'นาก', '#', 'ลู', 'พด', 'ปี', 'เด่น', 'stcogvRA', '0', 'jb', '4', 'ec']
['ชีคาก', 'ช', 'น้อง', 'ชลธุ', 'เฉลย', 'เฉพาะ', 'น้ำที', 'คดี', 'น้อง', 'ชลธุ', '#', 'นาก', '#', 'ลู', 'พด', 'ปี', 'เด่น', 'stcogvRA', '0', 'jb', '4', 'ec']
['ถึง', 'หากคน', 'คดี', 'ล', 'พด', 'ทีด', 'ให้', 'หมาย', 'ไป', 'แฉ', 'ต่อง', 'กักต', '14', 'วัน', 'ปี', 'ค่า', 'ไม', 'กักต', 'คดี', 'ว่า', 'คดี', 'เฉพาะ', 'พดเม', 'กบล',
'จ้า', 'กบล', 'ยัง', 'ต่อง', 'กักต', '14', 'วัน', 'ค่า', 'มก', 'ถ', 'ค', 'นี', 'ไป', 'เที่', 'เฉพาะ', 'ลั่นต้า', 'ค', 'ไม', 'ใช', 'ไป', 'ชอย', 'กู้ขกด', 'ค', 'ขอให', 'ช้า
บ้าน', 'ออกน', 'ปก', 'ปี', 'งก', 'ถ', 'นະ', '#', 'ลู', 'พด', '#', 'ลู', 'พด', 'ปี', 'เด่น', '#', 'นกน', 'กอก', 'กอก']
['คดี', 'ลู', 'พด', 'ทีด', 'หมาย', 'กักต', '14', 'กักต', 'พดเม', 'กักต', '14', 'ค่า', 'มก', 'ถ', 'ลั่นต้า', 'กู้ขกด', 'ขอให', 'ช้าบ้าน', 'ออกน', 'ปก', 'ปี', 'งก', '
ถ', 'นະ', '#', 'ล', 'พด', '#', 'บก', 'กอก']
['พด', 'อาจารย', 'สมยต', 'อคี', 'ลัค', 'พยายาม', 'สำเร็จ', 'พยายาม', 'จ', 'มูกดาวกร', 'พด', 'ชั้นเจน', 'ว', 'พด', 'จ', 'ช้า', 'เงินบริจาค', 'ร
รวมเง', 'กสิณ', 'หรือ', 'พาก', 'มา', 'พยายาม', 'พยายาม', 'ก', 'จะ', 'ไม', 'ขอรับ', 'ไว', 'เฉพาะ', 'ไม', 'ได', 'อยาก', 'ชีด', 'ลั่น', 'ตั้งแต่แรก', 'ขอให', '#', 'ลู', '
พด', 'ช้า', 'ไป', 'ชอย', 'ท', 'อีก', 'ที', 'เชา', 'ต่องกร', 'จะ', 'ดีกว่า', '#', 'รายกร', 'ช้าให', '#', 'ช่อง', '8', 'stco', '3', 'WunWSdcuP']
['พด', 'อาจารย', 'สมยต', 'อคี', 'ลัค', 'พยายาม', 'สำเร็จ', 'พยายาม', 'จ', 'มูกดาวกร', 'ชั้นเจน', 'พด', 'ช้า', 'เงินบริจาค', 'กสิณ', 'พาก', 'พยายาม',
'ขอรับ', 'ทีด', 'ลัค', 'ตั้งแต่แรก', 'ขอให', '#', 'ลู', 'พด', 'ต่องกร', 'ดีกว่า', '#', 'รายกร', 'ช้าให', '#', 'ช่อง', '8', 'stco', '3', 'WunWSdcuP']
[ล', 'ล', 'ล', 'ล', 'ล', 'ล', 'พด', 'stcogegMGNyP', '3', 'h']
['ลัค', 'หร', 'เอ', '#', 'ล', 'พด', 'stcogegMGNyP', '3', 'h']
['ลัค', 'หร', 'ส', 'ล', 'พด', 'stcogegMGNyP', '3', 'h']
['ปด', 'ให', 'ส', 'เหลือง', 'นรร', 'จ', 'ใส', 'ปี', 'ปี', 'กีด', 'พยายาม', 'ได', 'พด', 'ลู', 'เลย', 'จับ', '170', 'เทาพื้น', 'กอก', '3238', 'รอน', 'แขน', '81
6', '#', 'ลัค', 'เดือน', 'ลีส่อง', '#', 'ลัค', 'ส', 'กิน', 'แพร', '#', 'ลู', 'พด', '#', 'HowtoPerfect', '#', 'rip', 'la', 'ล', '#', 'เดือน', 'ลีส่อง', 'สภาพ', 'ดี', '#w',

Monday

01 / 02 / 2021

สิ่งที่จะ
UPDATE



การรวม **NLP** ของทั้งภาษาไทย
และอังกฤษให้เป็น **CLASS** เดียวกัน

ปรับปรุง **NLP** ของภาษาไทย

การจัดอันดับคำที่ใช้มากสุดในประโยค
และนำไปแสดงเป็น **PIE CHART**

การรวม NLP ของทั้งภาษาไทย

และอังกฤษให้เป็น class เดียวกัน

Function main ของ class NLP

Function ของ NLP สำหรับ
ภาษาไทย

```
23 def word_analysis(self, lang):
24     dictionary_count = {}
25     count = 0
26
27     for row in self.input_file['tweet']:
28         print(count)
29
30         if lang == "th" :
31             temp = self.analyze_word_th(row)
32         elif lang == "en" :
33             temp = self.analyze_word_en(row)
34         for word in temp:
35             message = word.lower()
36             if( message not in dictionary_count ):
37                 dictionary_count[message] = 1
38             elif( message in dictionary_count ):
39                 dictionary_count[message] += 1
40
41             if count == 10:
42                 break
43             count += 1
44
45             for temp in dictionary_count:
46                 self.writer_output.writerow({'word':temp, 'number':dictionary_count[temp]})
47
48 self.csvfile_output.close()
```

```
252     def analyze_word_th(self, data = ""):
253
254         stopwords = thai_stopwords()
255         ignor = []
256
257         data = re.sub('http', '', data)
258         data = re.sub('https', '', data)
259         data = re.sub('[a-z A-Z ]', '', data)
260         data = re.sub('[0-9 ]', '', data)
261         proc = word_tokenize(data, engine="newmm", keep_whitespace = False)
262         proc2 = [i for i in proc if i not in stopwords]
263
264         for word in proc2 :
265             ignor.append(re.sub('\w', '', word))
266
267         proc3 = [i for i in proc2 if i not in ignor]
268
269         return proc3
```

การจัดอันดับของคำที่ใช้มากที่สุด

Function สำหรับการจัด rank

ใช้ bubble sort สำหรับการเรียง

```
299 def word_rank(self, data) :  
300     sorted_word = {}  
301     key = [i for i in data.keys()]  
302     value = [i for i in data.values()]  
303     for first_number in range(len(value)) :  
304         for second_number in range(len(value) - first_number - 1) :  
305             if value[first_number] <= value[first_number + second_number] :  
306                 value[first_number] ,value[first_number + second_number] = value[first_number + second_number] ,value[first_number]  
307                 key[first_number] ,key[first_number + second_number] = key[first_number + second_number] ,key[first_number]  
308             else :  
309                 pass  
310  
311     for finish_sorted in range(len(data)) :  
312         sorted_word[key[finish_sorted]] = value[finish_sorted]  
313  
314 return sorted_word
```

Output ที่ออกมายจาก NLP

สถานที่

เวลา

C > Users < User Documents > tweet_NLP.csv

- 1 word, number
- 2 #covid, 8
- 3 news, 3
- 4 trump, 2
- 5 budget, 2
- 6 scheme, 2
- 7 experts, 2

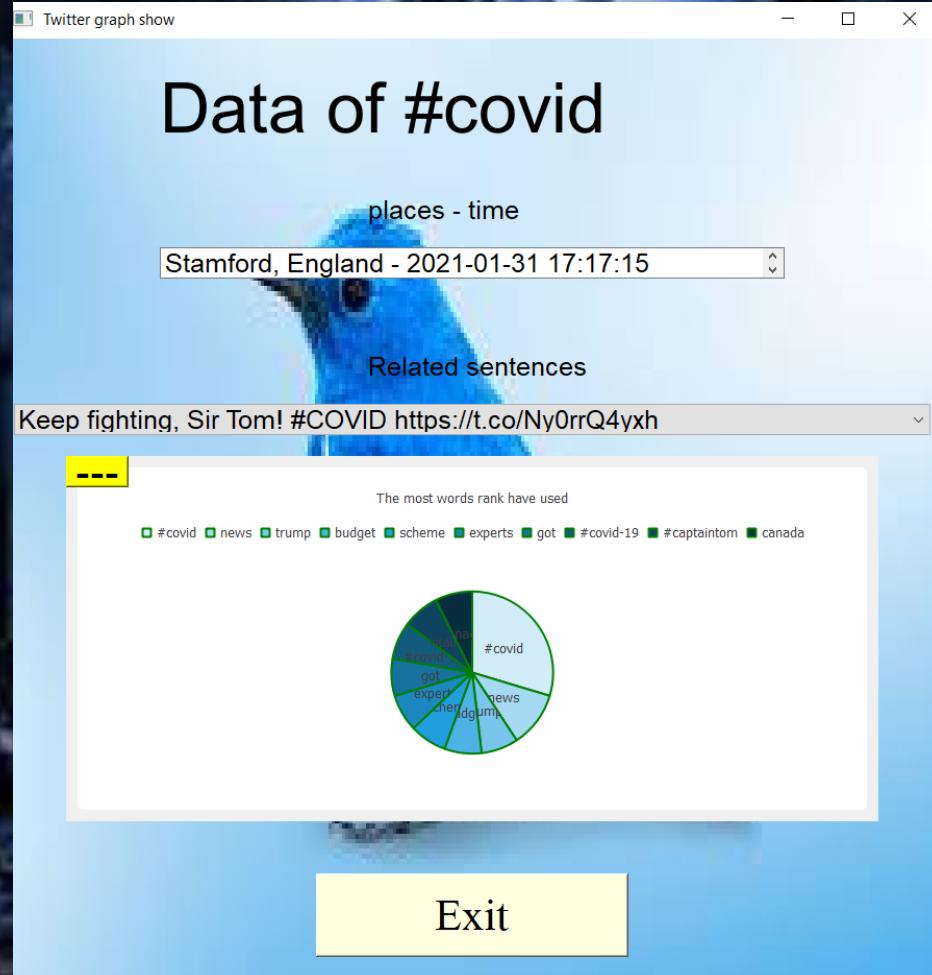
การแสดงอันดับของคำแบบ Pie Chart

การเลือกคำมา 10 อันดับแรก

การสร้าง Pie Chart ใน GUI

```
449     def word_ranging(self) :
450         data_file = pandas.read_csv('tweet_NLP.csv')
451
452         count = 0
453         count_limit = 10
454         for key,value in zip(data_file['word'] ,data_file['number']) :
455             if count < count_limit :
456                 self.pi_series.append(key ,value)
457                 count += 1
458             else :
459                 break
460
461         self.chart.addSeries(self.pi_series)
462         self.chart.setTitle("The most words rank have used")
463         self.pi_series.setLabelsVisible()
464         self.pi_series.setLabelsPosition(QtChart.QPieSlice.LabelInsideHorizontal)
465
466         for slice in self.pi_series.slices():
467             slice.setPen(QPen(Qt.darkGreen, 2))
468         chartview = QChartView(self.chart)
469         chartview.setGeometry(0,0,780,350)
470         chartview.setRenderHint(QPainter.Antialiasing)
471
472         self.pixmap = QPixmap(chartview.grab())
473         self.label_rank_1 = QLabel(self)
474         self.label_rank_1.setPixmap(self.pixmap)
475         self.label_rank_1.resize(chartview.width(), chartview.height())
476         self.label_rank_1.move(60,400)
```

Output ที่ออกแบบในรูปของ GUI



การเก็บข้อมูลของคำที่ search ล่าสุดสำหรับการใช้งานซ้ำ

```
90     def on_click_play(self):
91         tem_key = []
92         try :
93             keyword_check = pandas.read_csv('keyword_search.csv')
94             for key in keyword_check[ 'head' ] :
95                 tem_key.append(key)
96         except pandas.errors.EmptyDataError :
97             tem_key[0] = 'no data'
98
99         print(keyword_check[ 'head' ])
100        if tem_key[0] == 'no data' or self.input.text() not in tem_key :
101            fieldnames = [ 'head' ]
102            csvfile_for_head = open('keyword_search.csv', 'w', newline='', encoding="utf-8")
103            writer_file = csv.DictWriter( csvfile_for_head, fieldnames=fieldnames )
104            writer_file.writeheader()
105            writer_file.writerow({ 'head' : self.input.text() })
106            csvfile_for_head.close()
107
108        run = Data_twitter(self.input.text(),self.choice_1.currentText())
109        run.get_data()
110
111        npi = NLP()
112        npi.word_analysis(self.choice_1.currentText())
113
114    else :
115        pass
```

การเก็บข้อมูลของคำที่ search ล่าสุดสำหรับการใช้งานช้า(ต่อ)

ลำดับการเรียก

A	B
1	head
2	#covid

A	B	C	D	E	F
1	places	time	tweet		
2	Washington	#####	Relieved that @CDCgov has issued an emergency declaration. #Covid19		
3	New York	#####	1/ The media keeps saying not to worry about the numbers. #Covid19		
4	Pakistan	#####	Great news ດີເນີໂຫຼວດ ສະຫະກຳ will receive 17m @UniofPak		
5	United States	#####	bK2Qd		
6	Port Alber	#####	ZzngM		
7	Istanbul, Turkey	#####	#Breaking: COVID-19: Captain Sir Tom Moore in hospital after testing positive		
8	Naperville	#####	Too many #Covid-19 patients face death alone.		
9	Canada	#####	This is good news for Canada. Thanks @JustinTrudeau		
10	Pune, India	#####	oQPls8		
11	mainly here	#####	C'mon #CaptainTom you got this ດີເນີໂຫຼວດ .. #Covid19		
12	Bronx, NY	#####	TRUMP WAS RIGHT ABOUT HCQ but they always		
13	BLM	#####	journal_		
14	Geordie-D	#####	Interesting points about potential life, post-#COVID		
15	Clearwater	#####	kEsyd		
16	South of France	#####	Thoughts and prayers go to #CaptainTom who is		
17	#scl Chile	#####	Successful use of mild therapeutic hypothermia		
18	Chicago	#####	Just realized this coming Tuesday is #GroundhogDay		
19	New York, USA	#####	tter		
20	New Scotland	#####	48XX		
21		#####	9		
22	2354 Ocean	#####	crossed!		
23	United States	#####	VAERS data updates every Friday. This week the		
24	Parts Unknown	#####	Claudine, Joe and Derek can all cry me a fucking		

A	B	C	D	E	F
1	word	number			
2	#covid	8			
3	news	3			
4	trump	2			
5	budget	2			
6	scheme	2			
7	experts	2			
8	got	2			
9	#covid-19	2			
10	#captaintom	2			
11	canada	2			
12	australia	2			
13	business	2			
14	#economy	2			
15	people	2			
16	reports	2			
17	codes	1			
18	fake	1			
19	crack	1			
20	lines	1			
21	read	1			
22	ridiculous	1			
23	are	1			
24	â™€	1			
25	ດີເນີໂຫຼວດ	1			
26	ດີເນີໂຫຼວດ	1			
27		1			

Thursday

04 / 02 / 2021

สิ่งที่จะ
UPDATE



การดึงหัวข้อข่าวจาก **KEYWORD** ที่
ต้องการ

การดึงหัวข้อข่าวจากหลายๆ

WEBSITE

การนำเข้า **LINK** จากหัวข้อข่าวที่
SEARCH ได้เพื่ออ่านข้อมูลของ
ข่าว

Module ที่ใช้เพิ่ม

- **SoupStrainer**



- **re**



เป็น **METHOD** ที่ใช้ช่วยหา **CLASS**
ที่เราต้องการจาก **HTTP CODE**

ใช้ช่วยตัดคำของย่อหน้าของ
หัวข้อข่าว

Class ของ web crawler

```
7 class Web_crawler() :
8
9     def __init__(self, keyword = "") :
10         self.keyword = keyword
11
12         self.make_file = open('web_crawler.csv', 'w', newline='', encoding='utf-8')
13         fieldnames = ['head_news', 'link']
14         self.writer = csv.DictWriter( self.make_file, fieldnames=fieldnames )
15         self.writer.writeheader()
16
17         url = ["https://www.euronews.com/",
18                "https://www.themoscowtimes.com/",
19                "https://www.thairath.co.th/home"] # ตัวแปรสำหรับเก็บ link ของเว็บไซต์euronews
20
21         self.tem_web = {}
22
23         for link in url :
24             self.tem_web[link] = []
25
26             web = requests.get(link) # เป็นการรับ link ที่เข้ามายัง python โดยใช้ module
27             soup = BeautifulSoup(web.content, 'html.parser', parse_only = SoupStrainer("div")) # เป็นการจัด
28
29             all_head = soup.find_all(name ="a") # ค้นหาข้อมูลเฉพาะที่ต้องการ
30
31             for data in all_head :
32                 head = data.text
33                 head = re.sub('\n', ' ', head)
34                 if self.keyword == "" :
35                     try :
36                         link_head = self.link_analysis(data["href"],link)
37
38                         self.tem_web[link].append([head.lower(),link_head])
39                         self.writer.writerow( {'head_news': head.lower(), 'link': link_head} )
40                         #print(self.tem_web[link][-1][1])
41                         self.click_head_link(self.tem_web[link][-1][1])
42                     except KeyError :
43                         pass
44                 else :
45                     try :
46                         if self.keyword.lower() in head.lower() :
47                             link_head = self.link_analysis(data["href"],link)
48
49                             self.tem_web[link].append([head.lower(),link_head])
50                             self.writer.writerow( {'head_news': head.lower(), 'link': link_head} )
51
52                             self.click_head_link(self.tem_web[link][-1][1])
```

```
53
54
55
56
57
58
59         else :
60             pass
61         except KeyError :
62             pass
63
64
65
66
67         def click_head_link(self,data) :
68
69             web = requests.get(data)
70             soup = BeautifulSoup(web.content, 'html.parser', parse_only = SoupStrainer("div"))
71
72             all_text = soup.find_all(name = "p")
73             #print(all_text)
74
75
76             def link_analysis(self, link,main_link) :
77                 result = ""
78                 if link[0:6] == "https:" :
79                     result = link
80                 elif link[0:2] == "//" :
81                     result = "https:" + link
82                 else :
83                     result = main_link + link[1:]
84
85
86
87
88             if __name__ == '__main__':
89                 ex = Web_crawler()
```

Output ที่ได้

หัวข้อ

ลิงค์ของหัวข้อ

1	head_news,link		
2		english	, https://www.euronews.com
3		français	, https://fr.euronews.com
4		deutsch	, https://de.euronews.com
5		italiano	, https://it.euronews.com
6		español	, https://es.euronews.com
7		português	, https://pt.euronews.com
8		русский	, https://ru.euronews.com
9		türkçe	, https://tr.euronews.com
10		ελληνικά	, https://gr.euronews.com
11		magyar	, https://hu.euronews.com
12		فارسی	, https://per.euronews.com
13		العربية	, https://arabic.euronews.com
14		shqip	, https://www.euronews.al/al
15		română	, https://www.euronews.com/ro
16	newsletters,	https://www.euronews.com/follow-us	
17	games,	https://games.euronews.com/en?arkpromo=Site_TopNav	

ปេល្តាហានិយម



การเข้ามายังเว็บไซต์นี้ไม่ปลอดภัย

ผู้ใช้งานอาจพิจารณาข้อมูลจาก games.euronews.com (ตัวอย่างเช่น รหัสผ่าน ข้อมูลบัญชี
หรือบัตรเครดิต) **ดูข้อมูลเพิ่มเติม**

NET::ERR_CERT_REVOKED



[เปิดการปกป้องที่ปรับปรุงแล้วเพื่อให้ Chrome รักษาความปลอดภัยในระดับสูงสุด](#)

ขั้นสูง

ໂលດໃໝ່

Monday

08 / 02 / 2021

สิ่งที่จะ
UPDATE

- มีการ **STORE** ข้อมูลที่ **SEARCH** มาจาก **TWITTEET** และ **NLP**
- มีปุ่ม **UPDATE** ของ **GUI** ที่จะ **UPDATE** ข้อมูลใหม่ทั้งหมดใน **STORE** เก่า
- สามารถเก็บเนื้อข่าวที่ได้มาราจากแต่ละหัวข้อข่าวและวันเวลาที่ข่าวนั้นลงเก็บใน **FILE**

Store ข้อมูลที่ search มาจาก twittet และ NLP

ตัวอย่าง search คำว่า #covid

 keyword_search

1	date	keyword
2	2/7/2021 22:08	#covid
3	2/7/2021 22:17	#dota2
4	2/7/2021 22:52	#à,¥à,,à,‡à,žà,¥
5	2/8/2021 0:28	#china

 #covid_api

1	places	time	tweet
2	New York, NY	2/5/2021 20:10	JUST IN: NFL pledges to President Biden to ban the use of... Read more
3	New York	2/7/2021 11:41	HK3Sk
4	Washington, USA	2/6/2021 6:35	Sec Def Lloyd Austin states DOD will... Read more
5	DIRECTOR	2/7/2021 17:18	A new study confirms the CDC prediction that... Read more
6		2/7/2021 17:18	@js_edit The country has been thru... Read more
7	Montreal	2/7/2021 17:18	US, UK!
8	in transit	2/7/2021 17:18	Healthcare's supply-demand inefficiencies are... Read more
9	Montreal	2/7/2021 17:18	US, UK!
10	St Lucia	2/7/2021 17:18	@MbalulaFikile posted a picture of... Read more
11	Montreal	2/7/2021 17:18	US, UK!
12	San Fernando Valley,	2/7/2021 17:17	EHvht
13	planet earth	2/7/2021 17:17	This litany of all the mis-steps by the... Read more
14	Montreal	2/7/2021 17:17	US, UK!
15	NC	2/7/2021 17:17	Randâ€™s comments mean zilch- H... Read more
16	Montreal	2/7/2021 17:17	US, UK!
17		2/7/2021 17:17	#Covid-19 We just got our cases of... Read more
18	Nouvelle Aquitaine	2/7/2021 17:17	@k8nyquist @Julietilsen You do know... Read more
19	Online	2/7/2021 17:17	Classy Tennis Ball Pattern Sports Br... Read more

 #covid_NLP

1	word	number
2	days	6
3	2	6
4	#covid	6
5	amp	4
6	us	4
7	uk	4
8	#covid-19	4
9	canada	3
10	approved	3
11	day	3
12	protocol	3
13	webinar	3
14	1	3
15	button	2

Code ส่วนของการ store

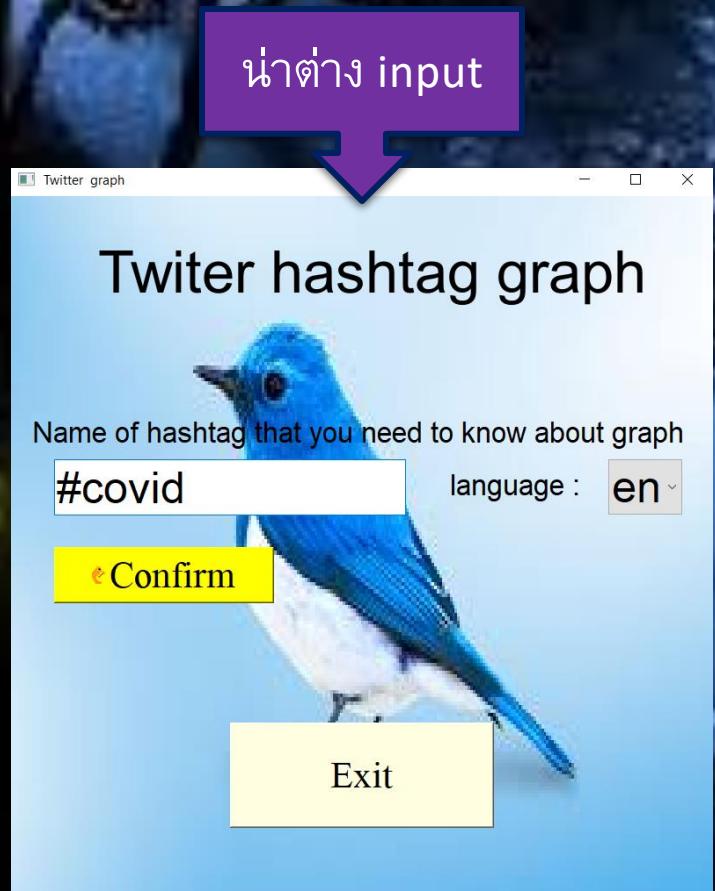
ส่วนของการเก็บ keyword
ที่เคย search

การ check ว่าคำที่ search มา
ล่าสุดมีอยู่ใน store หรือไม่

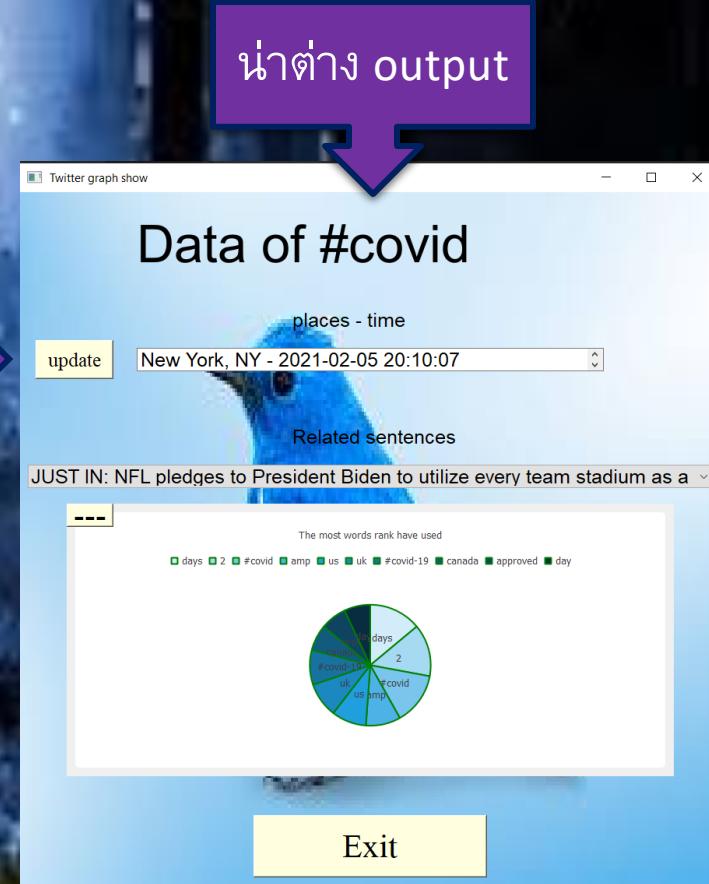
```
103 def make_data_base(self) :  
104     now = datetime.now()  
105     date_time = now.strftime("%Y-%m-%d %H:%M:%S")  
106     head = ["date", "keyword"]  
107  
108     tempfile = NamedTemporaryFile(mode='w', delete=False, newline='', encoding="utf-8")  
109     writer_re = csv.DictWriter(tempfile, fieldnames= head )  
110     writer_re.writeheader()  
111     tem_key = []  
112     try :  
113         keyword_check = pandas.read_csv('keyword_search.csv')  
114         for date,key in zip(keyword_check['date'],keyword_check['keyword']) :  
115             if key == self.input.text() :  
116                 writer_re.writerow({'date':date_time, 'keyword' : self.input.text() })  
117             else :  
118                 writer_re.writerow({'date':date, 'keyword' : key })  
119             tem_key.append(key)  
120     except FileNotFoundError :  
121         csvfile_for_head = open('keyword_search.csv', 'w', newline='', encoding="utf-8")  
122         writer_file = csv.DictWriter( csvfile_for_head, fieldnames=head )  
123         writer_file.writeheader()  
124         writer_file.writerow({'date':date_time, 'keyword' : self.input.text() })  
125         csvfile_for_head.close()
```

```
127     if self.input.text() not in tem_key :  
128         writer_re.writerow({'date':date_time, 'keyword' : self.input.text() })  
129         tempfile.close()  
130         shutil.move(tempfile.name, 'keyword_search.csv')  
131  
132         run = Data_twitter(self.input.text(),self.choice_1.currentText())  
133         run.get_data()  
134  
135         npi = NLP(self.input.text())  
136         npi.word_analysis(self.choice_1.currentText())  
137  
138     else :  
139         pass  
140
```

ມີບຸນ update ຂອງ GUI ທີ່ຈະ update ຂໍ້ມູລືໄໝ
ກັບໃນ store ແກ່



ປຸ່ມ
update



Code ส่วนของการ update

Function ของปุ่ม update

```
419     def update_button(self) :  
420         run = Data_twitter(self.name,self.lang)  
421         run.get_data()  
  
422         npi = NLP(self.name)  
423         npi.word_analysis(self.lang)  
  
425         self.show_3.clear()  
426         self.read_file(self.show_3)  
  
428         self.word_ranging()
```

เรียก API ใหม่

เรียก NLP ใหม่

GUI Update

ส่วนของ web crawler

1	head_news	link	content	date	time
2	english	https://wv	["We will	no data	no data
3	franÃ§ais	https://fr.e	["L'OMS a	no data	no data
4	deutsch	https://de	['Sturmtie	no data	no data
5	italiano	https://it.e	['Tra loro,	no data	no data
6	espaÃ±ol	https://es.	['Con la lle	no data	no data
7	portuguÃ‰s	https://pt.	['Mais trÃ©	no data	no data
8	Ã–Ã–fÃ–nÃ–dÃ–d,Ã–	https://ru.	['Watch Eu	no data	no data
9	tÃ¼rkÃ§e	https://tr.e	["ABD BaÅ	no data	no data
10	îµî»î»î·î½îµî»	https://gr.	['î î±î»î½	no data	no data
11	magyar	https://hu	['Az Embe	no data	no data
12	ÙøØ§Ø±Ø³ÙŒ	https://pe	['Ù...ØÙ...Ù	no data	no data
13	Ø§Ù„Ø¹Ø±ØºÙšØ©	https://ara	['Watch Eu	no data	no data
14	shqip	https://wv	[", '\r\n\t\\no data	no data	no data
15	romÃ¢nÃ¢f	https://wv	[] no data	no data	no data
16	newsletters	https://wv	['Fresh ne	no data	no data

1	head_newlink	link	content	date	time
2		https://wv	["We will	no data	no data
3		https://fr.e	["L'OMS a	no data	no data
4		https://de	['Sturmtie	no data	no data
5		https://it.e	['Tra loro,	no data	no data
6		https://es.	['Con la lle	no data	no data
7		https://pt.	['Mais trÃ©	no data	no data
8		https://ru.	['Watch Eu	no data	no data
9		https://tr.e	["ABD BaÅ	no data	no data
10		https://gr.	['î î±î»î½	no data	no data
11		https://hu	['Az Embe	no data	no data
12		https://pe	['Ù...ØÙ...Ù	no data	no data
13		https://ara	['Watch Eu	no data	no data
14		https://wv	[", '\r\n\t\\no data	no data	no data
15		https://wv	[] no data	no data	no data

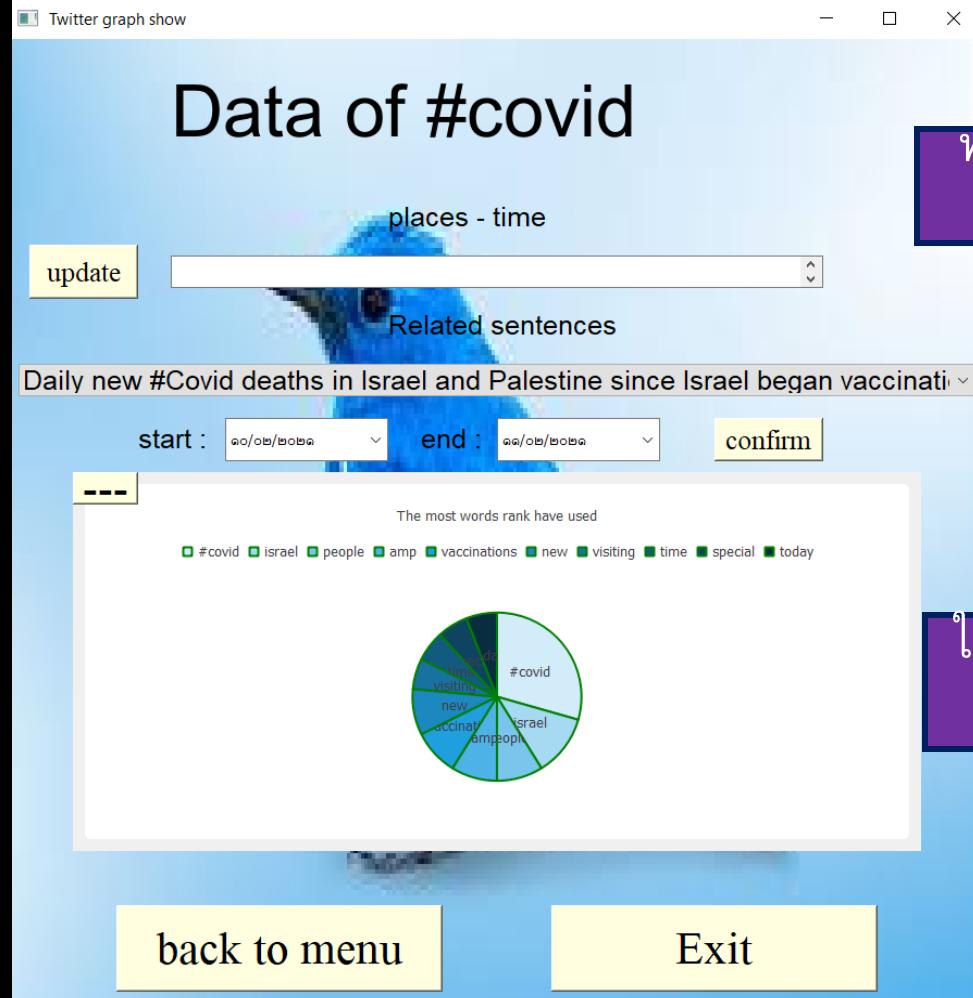
Thursday

11 / 02 / 2021

สิ่งที่จะ UPDATE

- เพิ่มการกำหนดการแสดงข้อมูลโดยการใช้ช่วงเวลา
- ปรับปรุง **GUI** ให้ทำงานได้ดีขึ้น และปรับปรุงการทำงาน **NLP** ให้เร็วขึ้น

การใช้ช่วงเวลาในการกำหนดการแสดงข้อมูล



พงกชันรับค่า

เวลา

ใช้ method ของ
pandas ช่วย

```
terminal  Help
test-8.py  test-9.py  test-friend.py  test-friend-2.py  test-5.py  test-4.py
c: > Users > User > Documents > test-5.py > Output > get_time
599
600
def get_time(self) :
    self.data_time_1 = self.time_label_1.date().toPyDate()
    self.data_time_2 = self.time_label_2.date().toPyDate()

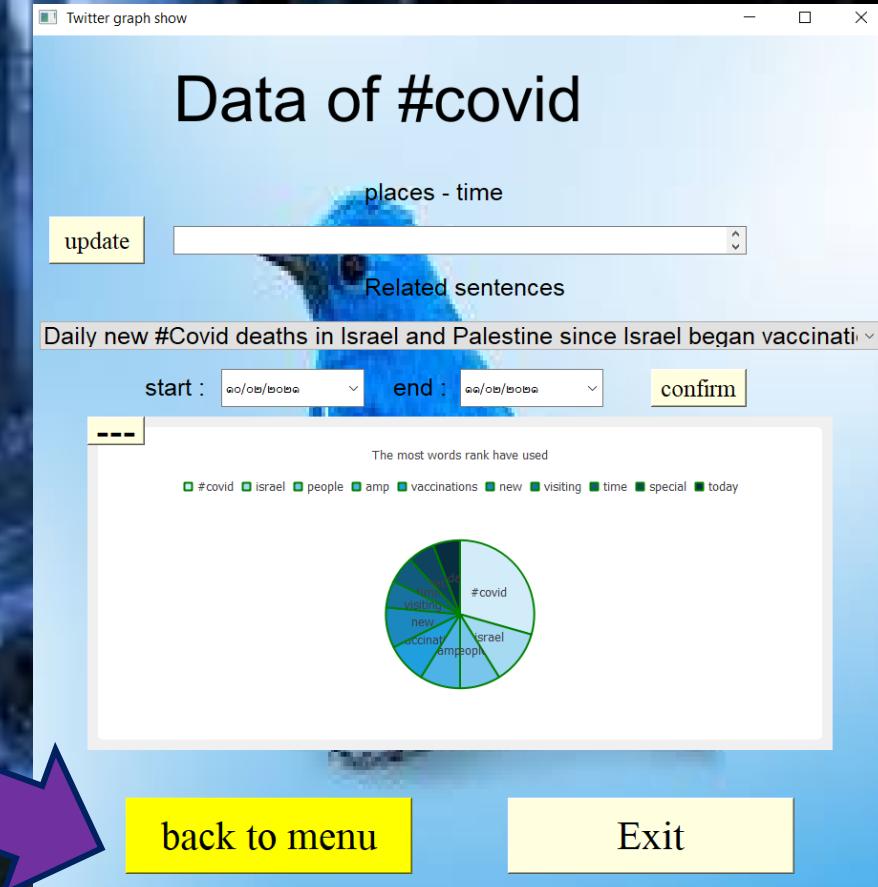
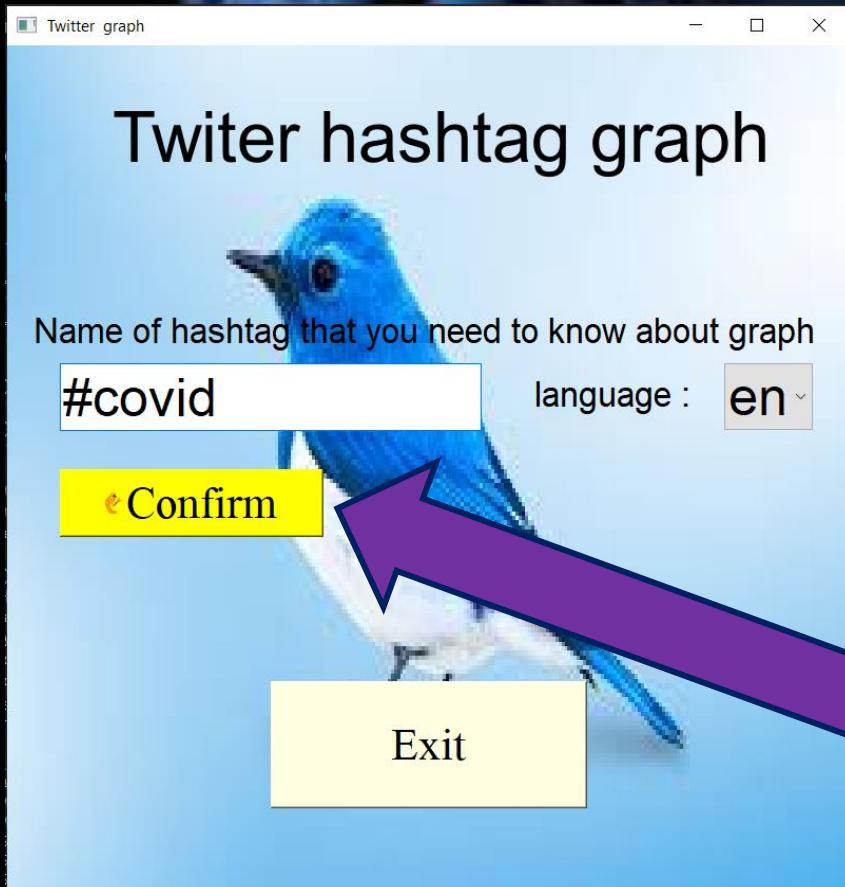
    day_1 ,day_2 = str(self.data_time_1.day) ,str(self.data_time_2.day)
    month_1 ,month_2 = str(self.data_time_1.month) ,str(self.data_time_2.month)
    year_1 ,year_2 = str(self.data_time_1.year) ,str(self.data_time_2.year)

    data_file = pandas.read_csv(str(self.name) + '_api.csv')
    if len(day_1) == 1 or len(day_2) == 1 :
        if len(day_1) == 1 :
            day_1 = '0' + day_1
        if len(day_2) == 1 :
            day_2 = '0' + day_2
    if len(month_1) == 1 or len(month_2) == 1 :
        if len(month_1) == 1 :
            month_1 = '0' + month_1
        if len(month_2) == 1 :
            month_2 = '0' + month_2

    con1 = data_file['time'] >= f'{year_1}-{month_1}-{day_1} 00:00:00'
    con2 = data_file['time'] <= f'{year_2}-{month_2}-{day_2} 23:59:59'
    compare = data_file[con1 & con2]

    self.show_3.clear()
    self.dictionary = {}
    for column in compare :
        self.dictionary[column] = []
        for inner in compare[column] :
            self.dictionary[column].append(inner)
    #print(self.dictionary)
    for word in self.dictionary['tweet'] :
        self.show_3.addItem(word)
```

การปรับปรุง GUI



การปรับปรุง NLP

ก่อนแก้ function

```
1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
0:00:47.506784
```

หลังแก้ function

```
1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
0:00:04.072378
```

Thursday
18 / 02 / 2021

สิ่งที่จะ **UPDATE**

- มีการเก็บข้อมูลตามช่วงวันที่เลือกคุ้มหากพบร่วงในช่วงวันที่คุณมีวันที่ไม่มีข้อมูล
- มีการนำข้อความ **TWEET** มาทำการประเมินแนวโน้มความรู้สึกโดย **SENTIMENT**
- ออกแบบ **GUI** สำหรับ **WEB CRAWLER** และแก้ไขปัญหาเรื่องการอ่านช่วงเวลา

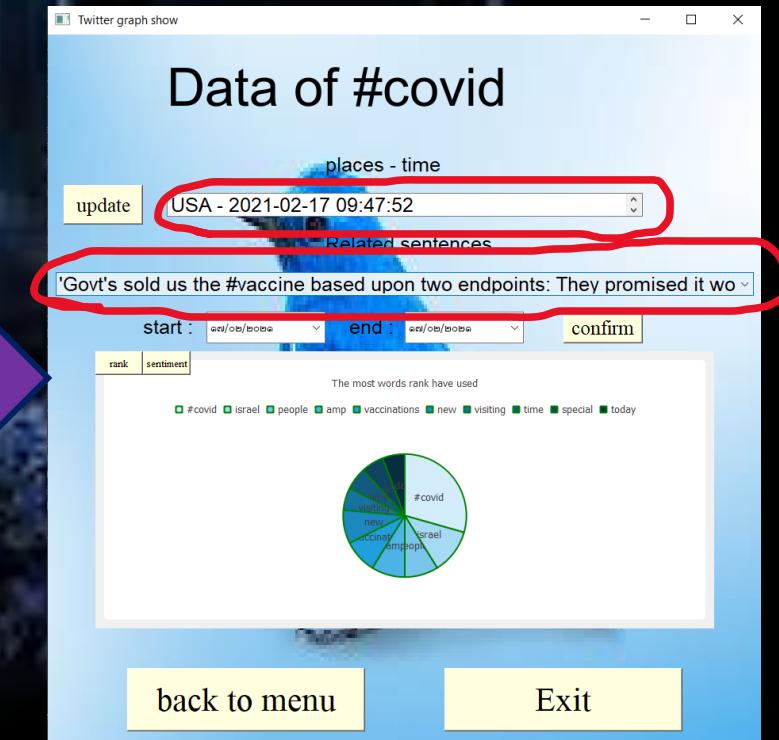
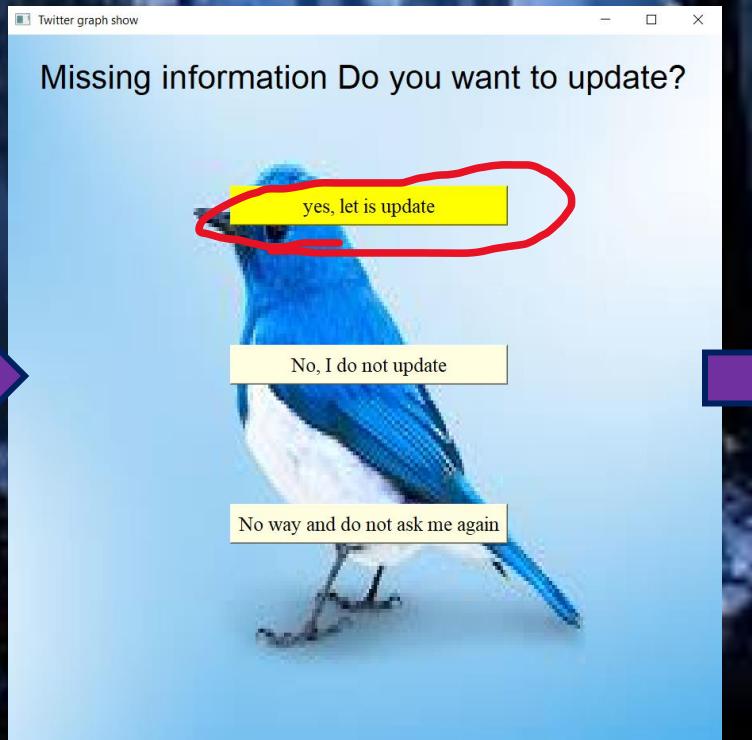
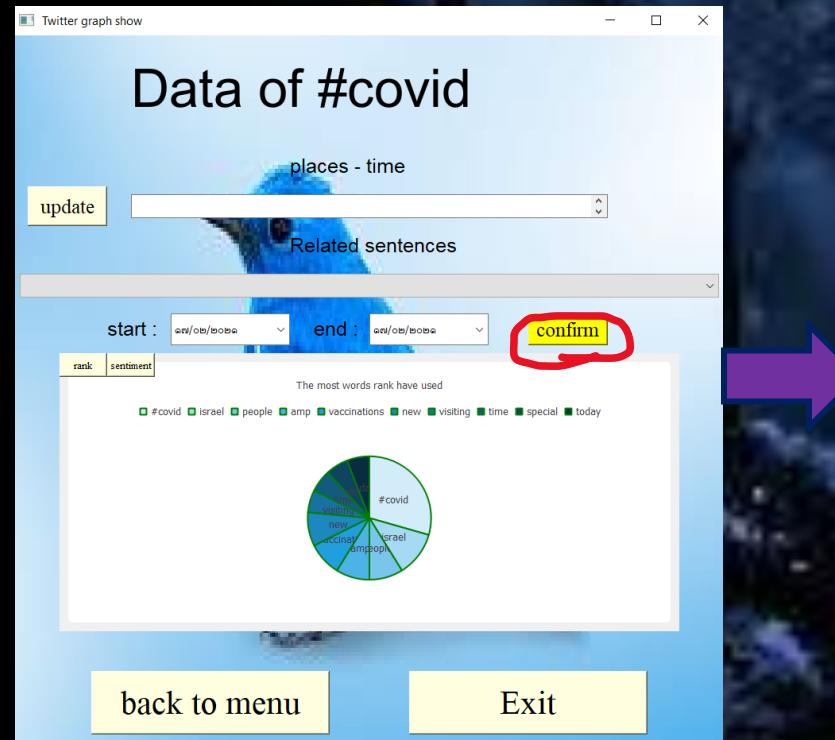
การใช้ช่วงเวลาในการกำหนดการแสดงข้อมูล

และเพิ่มข้อมูลที่ไม่มี

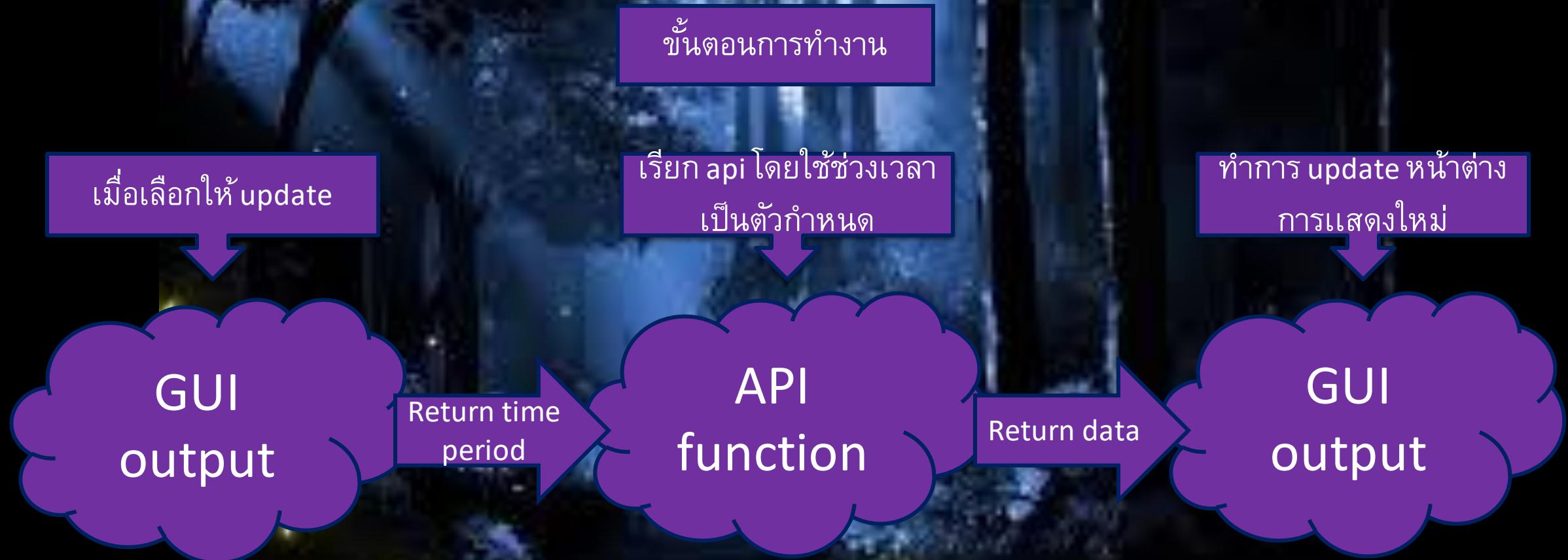
เลือกดูข้อมูลตามช่วงวัน

พบว่ามีวันที่ไม่มีข้อมูล

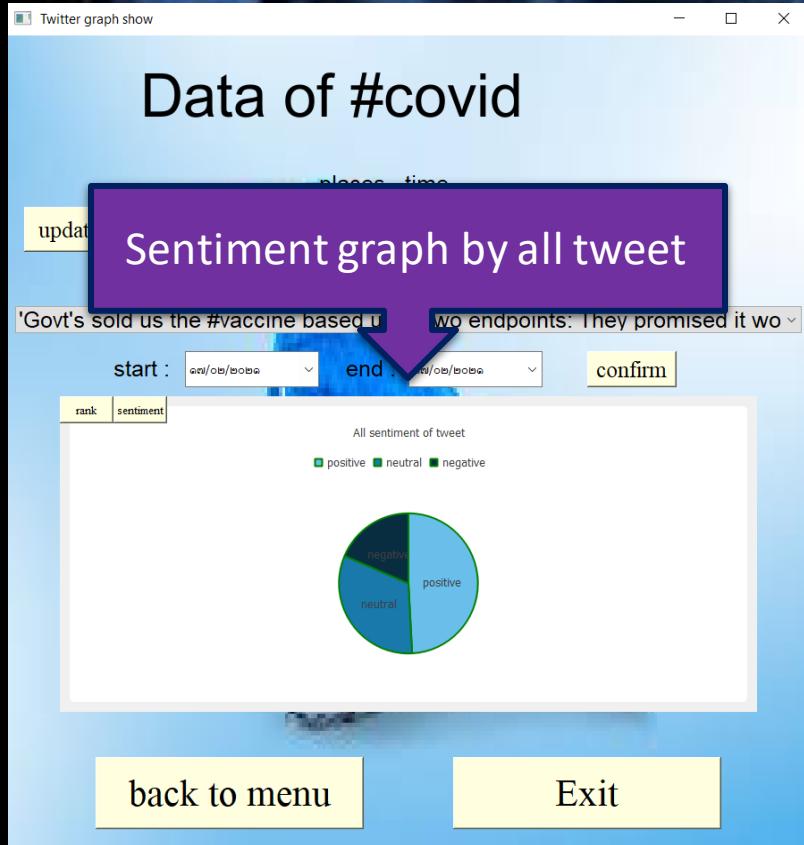
เพิ่มข้อมูลลงในวันที่ไม่มีข้อมูล



การใช้ช่วงเวลากำหนดการแสดงผลข้อมูล และเพิ่มข้อมูลที่ไม่มี(ต่อ)



การ sentiment ในการประเมินเนื้อหาของ tweet



Sentiment function

```
631 def tweet_sentiment(self) :
632     data_file = pandas.read_csv(str(self.name) + '_api.csv')
633     positive = 0
634     neutral = 0
635     negative = 0
636
637     for text in data_file['tweet'] :
638         analysis = TextBlob(text)
639         if analysis.sentiment[0] > 0 :
640             positive += 1
641         elif analysis.sentiment[0] < 0 :
642             negative += 1
643         else :
644             neutral += 1
645
646         #print('numbers of positive word is',positive)
647         #print('numbers of neutral word is',neutral)
648         #print('numbers of negative word is',negative)
649     self.pi_series_2 = QPieSeries()
650     self.chart_2 = QChart()
651
652     self.pi_series_2.append('positive' ,positive)
653     self.pi_series_2.append('neutral' ,neutral)
654     self.pi_series_2.append('negative' ,negative)
655
656     self.chart_2.addSeries(self.pi_series_2)
657     self.chart_2.setTitle("All sentiment of tweet")
658     self.pi_series_2.setLabelsVisible()
659     self.pi_series_2.setLabelsPosition(QChart.QPieSlice.LabelInsideHorizontal)
660
661     for slice in self.pi_series_2.slices():
662         slice.setPen(QPen(Qt.darkGreen, 2))
663     chartview = QChartView(self.chart_2)
664     chartview.setGeometry(0,0,780,350)
665     chartview.setRenderHint(QPainter.Antialiasing)
666
667     self.pixmap_2 = QPixmap(chartview.grab())
668     self.pixmap_2.save("sentiment_graph.png","PNG")
669     self.label_sentiment_1.setStyleSheet("border-image:url(sentiment_graph.png);")
```

การ sentiment ในการประเมินแนวโน้มของ tweet(ต่อ)

ทำการ import textblob มาใช้
สำหรับ sentiment function

```
from textblob import TextBlob
```

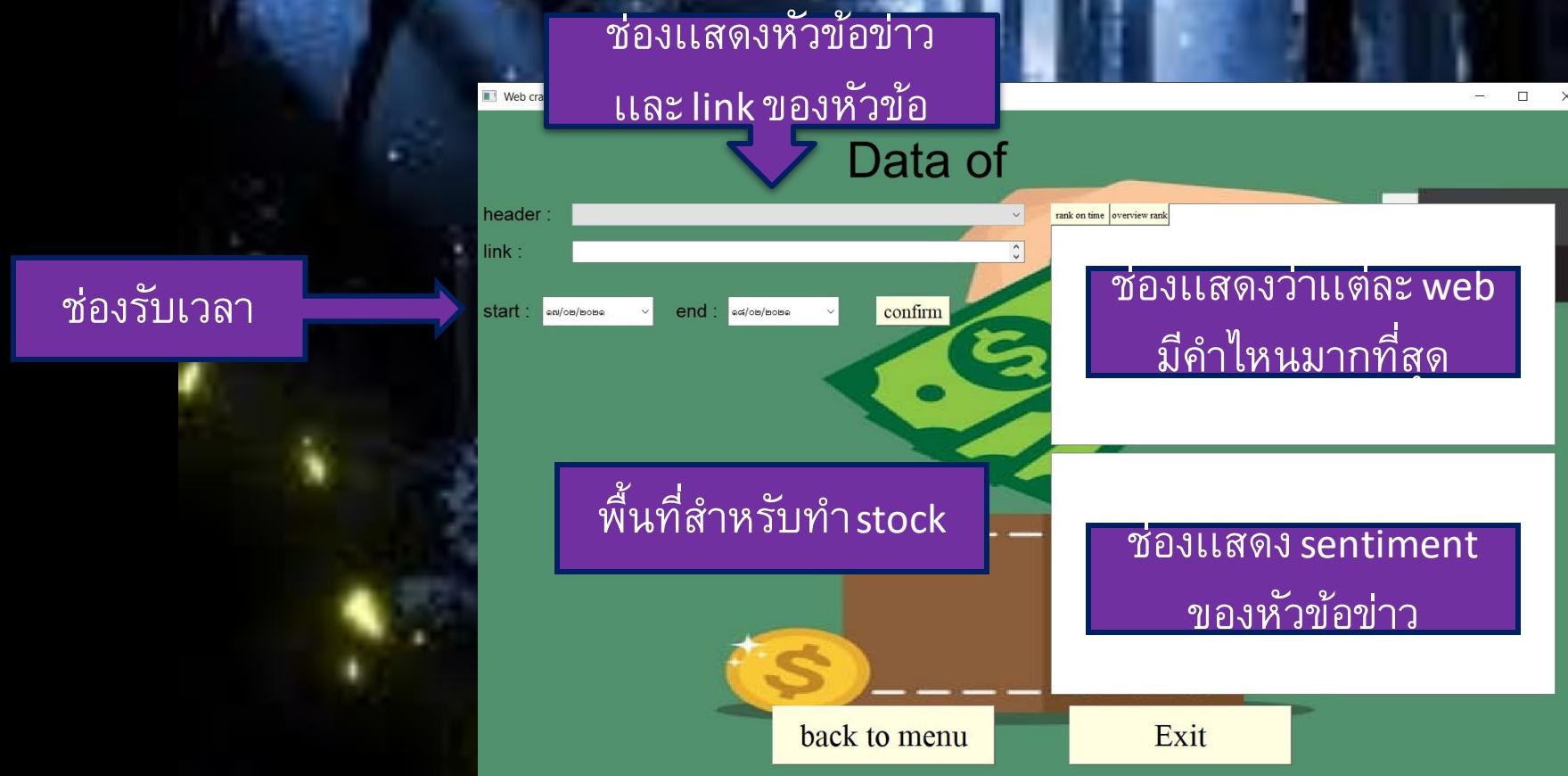
นำแต่ละประโยคมาดูว่าในประโยค
นั้นมีแนวโน้มของคำเท่าไหร่บ้าง

```
631     def tweet_sentiment(self) :  
632         data_file = pandas.read_csv(str(self.name) + '_api.csv')  
633         positive = 0  
634         neutral = 0  
635         negative = 0  
636         for text in data_file['tweet'] :  
637             analysis = TextBlob(text)  
638             if analysis.sentiment[0] > 0 :  
639                 positive += 1  
640             elif analysis.sentiment[0] < 0 :  
641                 negative += 1  
642             else :  
643                 neutral += 1
```

ส่วนของการสร้างกราฟ

```
647     self.pi_series_2 = QPieSeries()  
648     self.chart_2 = QChart()  
649  
650     self.pi_series_2.append('positive' ,positive)  
651     self.pi_series_2.append('neutral' ,neutral)  
652     self.pi_series_2.append('negative' ,negative)  
653  
654     self.chart_2.addSeries(self.pi_series_2)  
655     self.chart_2.setTitle("All sentiment of tweet")  
656     self.pi_series_2.setLabelsVisible()  
657     self.pi_series_2.setLabelsPosition(QtChart.QPieSlice.LabelInsideHorizontal)  
658  
659     for slice in self.pi_series_2.slices():  
660         slice.setPen(QPen(Qt.darkGreen, 2))  
661     chartview = QChartView(self.chart_2)  
662     chartview.setGeometry(0,0,780,350)  
663     chartview.setRenderHint(QPainter.Antialiasing)  
664  
665     self.pixmap_2 = QPixmap(chartview.grab())  
666     self.pixmap_2.save("sentiment_graph.png","PNG")  
667     self.label_sentiment_1.setStyleSheet("border-image:url(sentiment_graph.png);")
```

ออกแบบ GUI สำหรับ web crawler



แก้ไขปัญหาระบบการอ่านช่วงเวลา



The Moscow Times
INDEPENDENT NEWS FROM RUSSIA

Support The Moscow Times!
CONTRIBUTE TODAY

div.site-header.py-3.hide 940.8 x 116.31 ANWHILE ARTS AND LIFE PODCASTS VIDEOS IN-DEPTH RU Q

INTERNATIONAL PRESSURE

European Rights Court Demands Navalny's 'Immediate' Release

The ECHR document published on Navalny's website says its demand was made in "regard to the nature and extent of risk" to his life.

Details from the Chrome DevTools:

- Elements tab: Shows the DOM structure of the page, with a red box highlighting the URL in the address bar: `https://www.themoscowtimes.com/`.
- Console tab: Displays developer logs and errors.
- Styles tab: Shows the CSS styles applied to the elements on the page.
- Computed tab: Shows the final computed styles for selected elements.
- Event Listeners tab: Shows event listeners attached to the elements.
- DOM Breakpoints tab: Allows setting breakpoints in the DOM tree.
- Properties tab: Shows element properties.
- Accessibility tab: Provides accessibility-related information.

```
21 url = ["https://www.thairath.co.th/home/",  
22 "https://www.themoscowtimes.com/",  
23 "https://www.euronews.com/"]  
24  
25 #url = ["https://www.thairath.co.th/home/"]  
26  
27 function = ['h1', 'h2', 'h3', 'h4', 'h5', 'h6']  
28 self.tem_web = {}  
29  
30 for link in url :  
31     func_list = []  
32     self.tem_web[link] = []  
33  
34     web = requests.get(link)  
35     soup = BeautifulSoup(web.content, 'lxml', parse_only = SoupStrainer("div"))  
36  
37     for fuct in function :  
38         func_list += soup.find_all(name =fuct)  
39  
40     print('hey')  
41     print(len(func_list))  
42     #all_head = soup.find_all(name = "a")  
43  
44     for data in func_list :  
45         try :  
46             head = data.text  
47             #print(head)  
48             head = re.sub('\n', ' ', head)  
49             if self.keyword == "" :  
50                 print(data.find(name = "a")["href"])  
51                 link_head = self.link_analysis(data.find(name = "a")["href"], link)  
52  
53                 self.tem_web[link].append([head.lower(),link_head])  
54                 #self.writer.writerow( {'head_news': head.lower(), 'link': link_head})  
55  
56             context = self.click_head_link(self.tem_web[link][-1][1])  
57             created_at = self.get_time(self.tem_web[link][-1][1])  
58             #print(created_at)  
59  
60             self.writer.writerow( {'head_news': head.lower(), 'link': link_head,  
61             except KeyError :  
62                 pass}
```

Update week 8

Monday

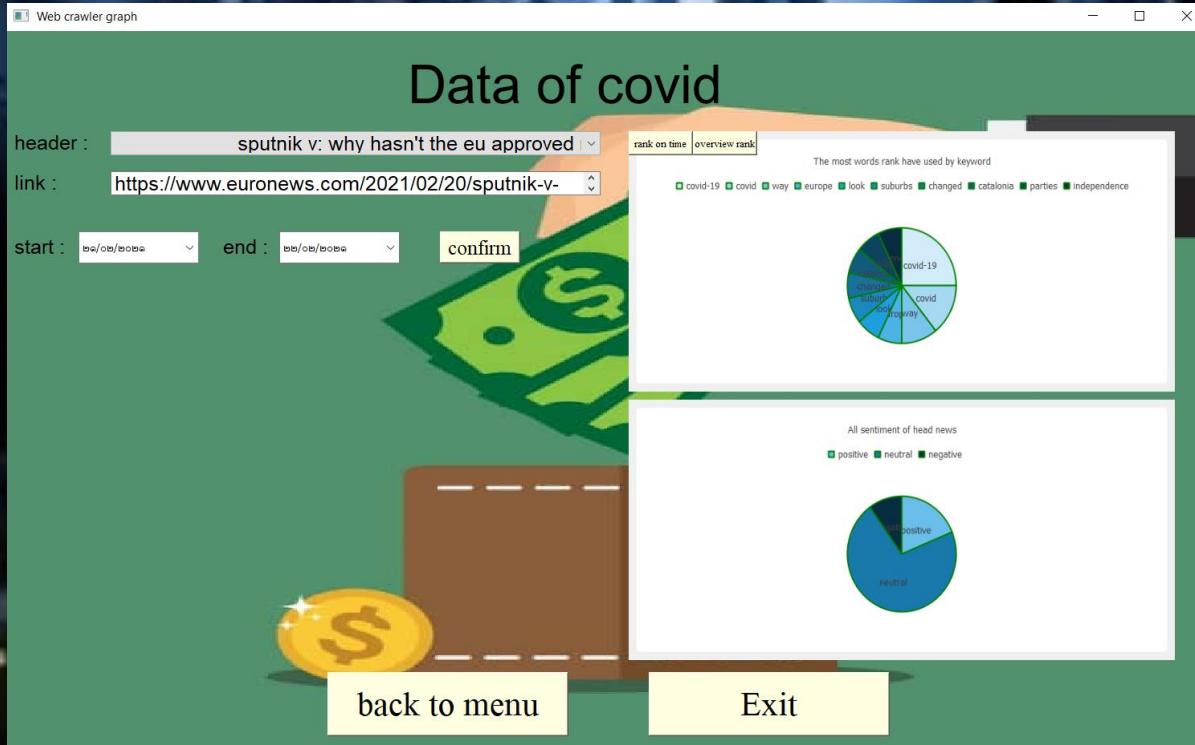
22 / 02 / 2021

สิ่งที่จะ
UPDATE ➤

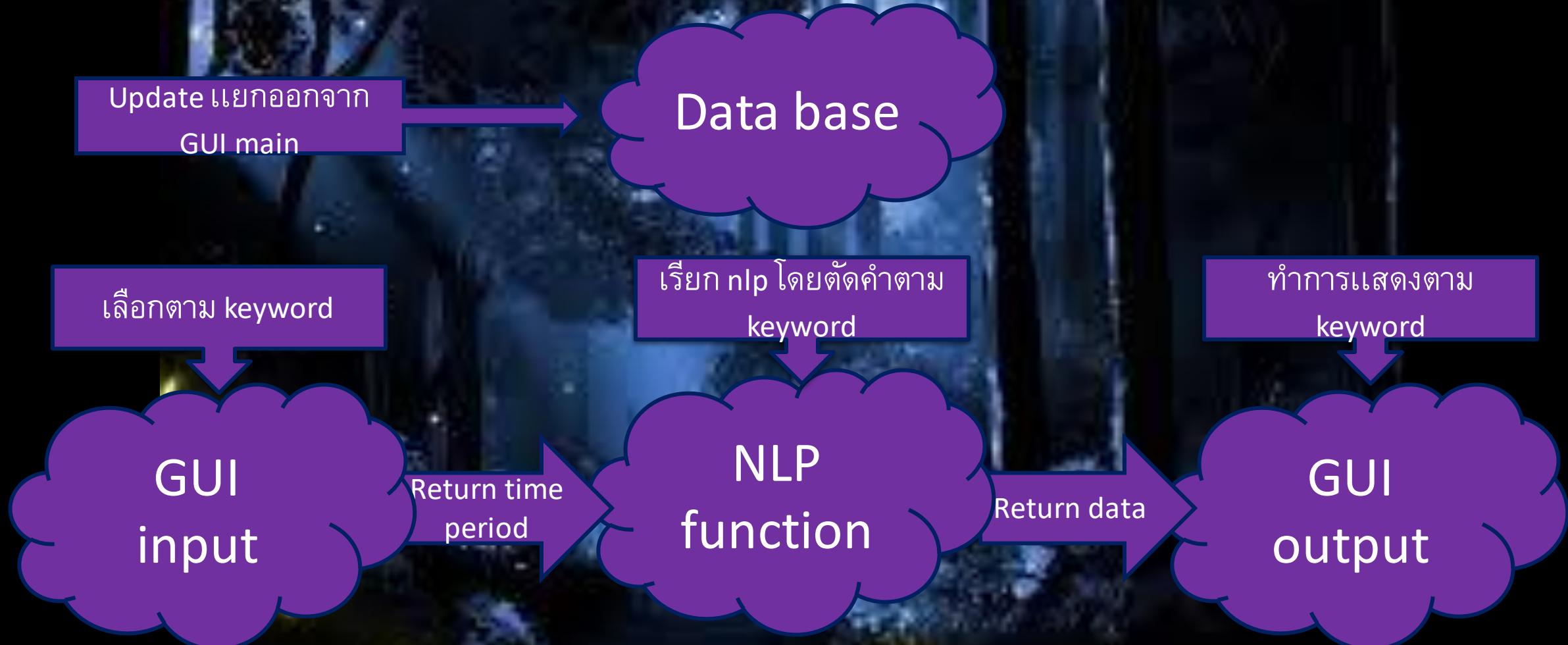
- ทำการเชื่อมในส่วนของ **GUI**
และ **DATA BASE**

ส่วนของการเชื่อม GUI กับ database

ดึงตาม keyword



ส่วนของการเชื่อม GUI กับ database(ต่อ)



Thursday

11 / 03 / 2021

สิ่งที่จะ UPDATE

- แก้ไข **NLP** ใหม่ให้ทำได้ทั้งภาษาไทยและอังกฤษ
- ปรับปรุง **CRAWLER** ให้รับข้อมูลได้สมบูรณ์มากขึ้น
- ทดลองการใช้ **STOCK**

แก้ไข NLP ใหม่ให้ทำได้ทั้งภาษาไทยและอังกฤษ

เป็นส่วนที่เอาไว้ set engine ของภาษาไทย

และกำหนด stop word ต่าง ๆ ของภาษาอังกฤษ

```
terminal
test-8.py test-9.py test-friend.py software-2-week-2.py test-friend-2.py test-5
C: > Users > User > Documents > test-5.py > NLP > _init_
404
405 class NLP() :
406     def __init__(self,keyword):
407         # -----read file-----
408         name_file = str(keyword) + '_NLP.csv'
409         self.csvfile_output = open(name_file, 'w', newline='', encoding="utf-8")
410         self.input_file = pandas.read_csv(str(keyword) + '_api.csv')
411
412         fieldnames = ['word', 'number']
413         self.writer_output = csv.DictWriter( self.csvfile_output, fieldnames=fieldnames )
414         self.writer_output.writeheader()
415         # -----NLP thai-----
416         self.stopwords_thai = pythainlp.corpus.common.thai_stopwords()
417
418         # -----NLP english-----
419         self.nlp = spacy.load("en_core_web_md")
420         self.STOP_WORD_1 = self.nlp.Defaults.stop_words
421         self.STOP_WORD_2 = stopwords.words('english') |
422         self.STOP_WORD_3 = STOP_WORDS
423
424         # -----
425         self.all_word = {}
426         for sentence in self.input_file['tweet'] :
427             self.main_function(sentence)
428
429         self.all_word = self.word_rank(self.all_word)
430         for temp in self.all_word :
431             self.writer_output.writerow({'word':temp, 'number': self.all_word[temp]})
```

แก้ไข NLP ใหม่ให้ทำได้ทั้งภาษาไทยและอังกฤษ

ส่วนจัดการในเรื่องคำที่มี # หรือ @ นำหน้า

ส่วนที่นำ stop มาตัดคำที่
ไม่จำเป็นออก

```
435 def main_function(self, datas):
436
437     output_list = []
438     out_STR = ""
439     # -----Filter all thing and return list of word "usefull"-----
440     try:
441         pattern = re.compile(r"(#[a-zA-Z0-9(_)|n-\w(_)]{1,})")
442         out_str_hashtags = pattern.sub("", datas)
443
444         pattern = re.compile(r"(@+[a-zA-Z0-9(_)|n-\w(_)]{1,})")
445         out_str_add = pattern.sub("", out_str_hashtags)
446
447         str_output = emoji.get_emoji_regexp().sub(u'', out_str_add)
448
449         pattern = re.compile(r"(?i)\b(?:https://|www\d{0,3}[.][a-zA-Z0-9.\-]+[.][a-zA-Z]{2,4})")
450         out_str_link = pattern.sub("", str_output)
451
452         out_STR += out_str_link
453
454         pattern = re.compile(r"(?P<out_list>#[a-zA-Z0-9(_)|n-\w(_)]{1,})")
455         out_list = re.search(pattern, datas)
456         output_list.append(out_list.group("out_list"))
457
458         pattern = re.compile(r"(?P<out_list>@[a-zA-Z0-9(_)|n-\w(_)]{1,})")
459         out_list = re.search(pattern, datas)
460         output_list.append(out_list.group("out_list"))
461         #print(str_output,end="\n\n")
462     except AttributeError:
463         pass
```

```
464 # -----Filter all thing and return list of word "usefull"-----
465
466 proc = word_tokenize(out_STR, engine="newmm", keep_whitespace=False)
467 for i in proc:
468     # -----special symbol-----
469     special = re.compile(r"\W+").sub("",i) # special symbol
470     if(special == "" or i.lower() == "https" or i.lower() == "http"):
471         continue
472     #
473
474     # ----- stop word thai and english -----
475     if(isthai(i)):
476         if( i not in self.stopwords_thai ):
477             output_list.append(i)
478     elif(i.isalpha()):
479         if( i.lower() not in self.STOP_WORD_1 and i.lower() not in self.STOP_WORD_2 and i.lower() not in self.STOP_WORD_3 ):
480             output_list.append(i)
481         #
482
483     #return output_list
484
485     for word in output_list :
486         if word not in self.all_word :
487             self.all_word[word] = 1
488         else :
489             self.all_word[word] += 1
```

ប្រុបក្រែង crawler

Main function

```
39 def main_function (self) :
40
41     for news in self.main_url :
42
43         temp_name = str(news).split("https://")
44         temp_name = temp_name[1][:-1].split("/")
45         name_file = temp_name[0] + '_crawler.csv'
46
47         self.back_up = self.file_manage(str(name_file))
48
49         self.each_file = open( name_file , 'a', newline='', encoding='utf-8')
50         fieldnames = ['head_news', 'link', 'content','time']
51         self.writer = csv.DictWriter( self.each_file, fieldnames = fieldnames )
52         self.writer.writeheader()
53
54         self.news = news
55         web = requests.get(news)
56         news_page = web.content
57         self.soup = BeautifulSoup( news_page , 'lxml', parse_only = SoupStrainer("div"))
58
59         head_link = self.click_head_link(news)
60
61         count = 0
62         print(len(head_link))
63         for each_link in head_link :
64             print("count : " + str(count))
65
66             if each_link[0:6] == "https:" :
67                 self.get_content(each_link)
68             elif each_link[0:2] == "/" :
69                 self.get_content("https:" + each_link)
70             else :
71                 self.get_content(str(news) + each_link[1:])
72
73             count += 1
74
75         self.each_file.close()
```

```
78
79
80     def click_head_link(self,link) :
81
82         not_head = self.check_not_head(link)
83         all_data = self.soup.find_all(name="a")
84
85         output = []
86         for data in all_data :
87             try:
88                 temp_link = data["href"]
89                 if ( self.link_analysis(temp_link, not_head) ) and ( temp_link not in output ) :
90                     output.append(temp_link)
91             except KeyError:
92                 pass
93
94         return output
```

Link function

Content function

```
157
158     def get_content(self, url):
159         print(url)
160
161         web = requests.get(url)
162         print('1')
163
164         news_page = web.content
165         print('2')
166         soup = BeautifulSoup(news_page, 'lxml', parse_only=SoupStrainer("div"))
167         print('3')
168         head = BeautifulSoup(news_page, 'lxml')
169
170         all_content = soup.find_all(name="p")
171         complete_content = ""
172         output = []
173
174         for content in all_content :
175             complete_content += content.text+ "\n"
176
177         print(complete_content)
178         time = self.get_time(url)
179         #print(time[0] + " " + time[1])
180
181         self.check_double = []
182         try:
183
184             print( head.title.string)
185             if (url not in self.check_double) and (url not in self.back_up) :
186                 self.writer.writerow( { 'head_news': head.title.string, 'link': url, 'content': complete_content, 'time': time[0] + " " + time[1] } )
187                 self.check_double.append( url )
188
189         except TypeError :
190             print('not write')
191             pass
192
193         except AttributeError :
194             print('not write')
195             pass
```

ប្របប្រុង crawer(ទៅ)

Link function

```
78     def click_head_link(self,link) :
79
80         not_head = self.check_not_head(link)
81         all_data = self.soup.find_all(name="a")
82
83         output = []
84         for data in all_data :
85             try:
86                 temp_link = data["href"]
87                 if ( self.link_analysis(temp_link, not_head) ) and ( temp_link not in output ) :
88                     output.append(temp_link)
89             except KeyError:
90                 pass
91
92         return output
93
94
95     def check_not_head(self, url) :
96
97         web = requests.get(url)
98         news_page = web.content
99         soup = BeautifulSoup(news_page, 'lxml', parse_only=SoupStrainer("nav"))
100        data = soup.find_all(name="ul")
101        storage = []
102        for i in data:
103            temp = i.find_all("li")
104            for j in temp:
105                try:
106                    g = j.find("a")["href"]
107                    storage.append(g)
108                except TypeError:
109                    #print(g)
110                    pass
111                except KeyError:
112                    #print(g)
113                    pass
114
115        soup1 = BeautifulSoup(news_page, 'lxml', parse_only=SoupStrainer("footer"))
116        sub_footer = []
117        for i in soup1.find_all("a"):
118            sub_footer.append(i.get("href"))
119
120        return storage+sub_footer
```

Function កើត link ទៅ
មិនឱ្យខ្សោយរាយ

Compare function

```
200     def link_analysis(self, link, not_link_head) :
201
202         cond_link = ""
203
204         condition_1 = True
205         if link[0:6] == "https:" :
206             cond_link = link
207         elif link[0:2] == "//" :
208             cond_link = "https:" + link
209         else :
210             cond_link = self.news + link[1:]
211
212         for each in not_link_head :
213             link_check = ""
214             if( each[0:6] == "https:" ):
215                 link_check = each
216             elif( each[0:2] == "//" ):
217                 link_check = "https:" + each
218             elif( each[0] == "/" ):
219                 link_check = self.news + each[1:]
220
221         if link_check == cond_link :
222             condition_1 = False
223
224         condition_2 = True
225         if( link[0:6] == "https:" ):
226             main_link = self.news.split(".co")
227             cond_link = link.split(".co")
228
229             if(main_link[0] != cond_link[0]):
230                 condition_2 = False
231
232         condition_3 = True
233         if self.news == "https://www.sanook.com/news/" :
234             try:
235                 reg = re.compile(r'(?P<name>/news/\d+)')
236                 ou = reg.search(str(link))
237                 name_output = ou.group("name")
238                 condition_3 = True
239             except AttributeError:
240                 condition_3 = False
241
242         result = condition_1 and condition_2 and condition_3
243
244         return result
```

ปรับปรุง crawler(ต่อ)

Content function

```
157
158     def get_content(self, url):
159         print(url)
160
161         web = requests.get(url)
162         print('1')
163
164         news_page = web.content
165         print('2')
166         soup = BeautifulSoup(news_page, 'lxml', parse_only=SoupStrainer("div"))
167         print('3')
168         head = BeautifulSoup(news_page, 'lxml')
169
170         all_content = soup.find_all(name="p")
171         complete_content = ""
172         output = []
173
174         for content in all_content :
175             complete_content += content.text+ "\n"
176
177         print([complete_content])
178         time = self.get_time(url)
179         #print(time[0] + " " + time[1])
180
181         self.check_double = []
182         try:
183
184             print( head.title.string)
185             if (url not in self.check_double) and (url not in self.back_up) :
186                 self.writer.writerow( { 'head_news': head.title.string, 'link': url, 'content': complete_content, 'time': time })
187                 self.check_double.append( url )
188
189         except TypeError :
190             print('not write')
191             pass
192
193         except AttributeError :
194             print('not write')
195             pass
```

time function

ทดลองการใช้ stock

ส่วน code

```
1 import sys
2 import csv
3 from pandas_datareader import data as pdr
4
5 stock = 'btc-usd'
6 start = "2021-01-01"
7 end = "2021-03-11"
8
9 try:
10     ptt = pdr.get_data_yahoo(stock, start=start, end=end)
11     array = str(ptt['Adj Close']).split("\n")
12     #print(array)
13     array_x = []
14     array_y = []
15     for i in array[1:-1]:
16         array_x.append(i.split(" ")[0])
17         array_y.append(i.split(" ")[2])
18 except:
19     print("Error:", sys.exc_info()[0])
20     print("Description:", sys.exc_info()[1])
21
22 for x,y in zip(array_x,array_y) :
23     print(x, y)
```

Output

วันที่ของหุ้น

ราคปิดท้ายของวัน

2020-12-31	29001.720703
2021-01-01	29374.152344
2021-01-02	32127.267578
2021-01-03	32782.023438
2021-01-04	31971.914062
2021-03-07	51206.691406
2021-03-08	52246.523438
2021-03-09	54824.117188
2021-03-10	56008.550781
2021-03-11	54686.085938

Monday

15 / 03 / 2021

สิ่งที่จะ
UPDATE



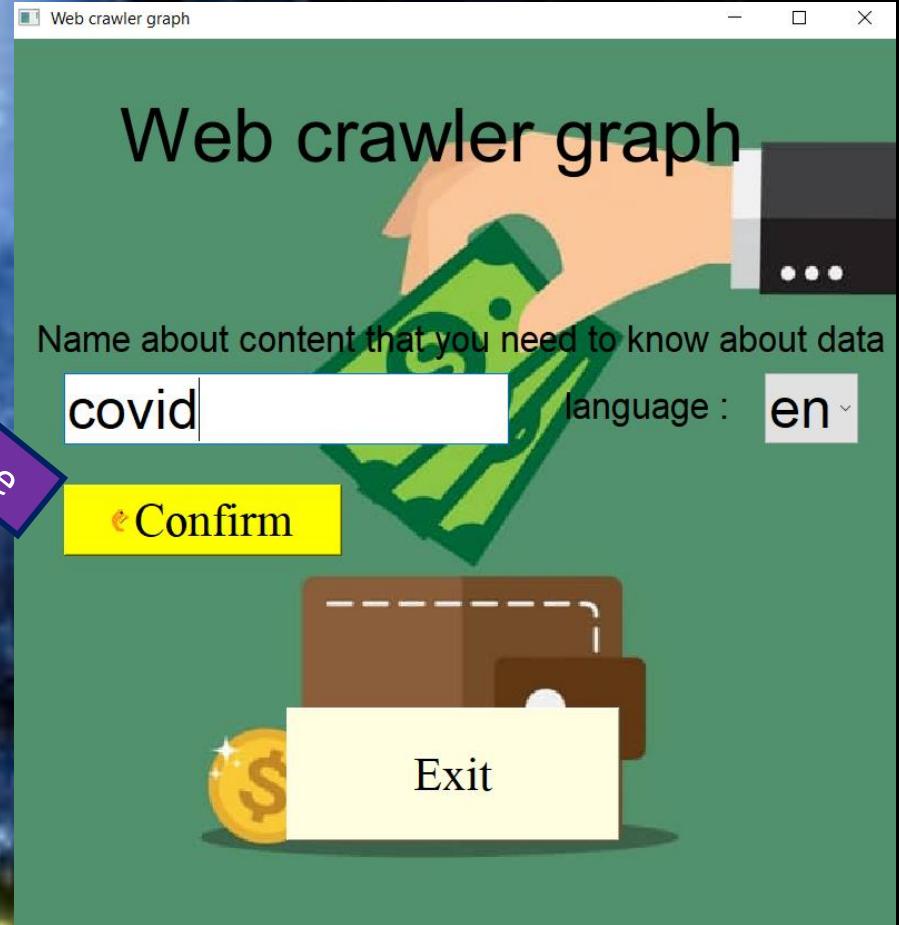
- ทำ **GUI** สำหรับ **WEB CRAWLER**
- เพิ่มฟังก์ชัน **STOCK** เข้ามาในส่วนของ **CRAWLER GUI**
- เชื่อมต่อระหว่าง **CRAWLER** และ **API** โดยการนำคำที่ผู้มากที่สุดจาก **CRAWLER** มาเป็น **KEYWORD** ของ **API**

ស៊ីនខែង GUI សំរាប់ web crawler

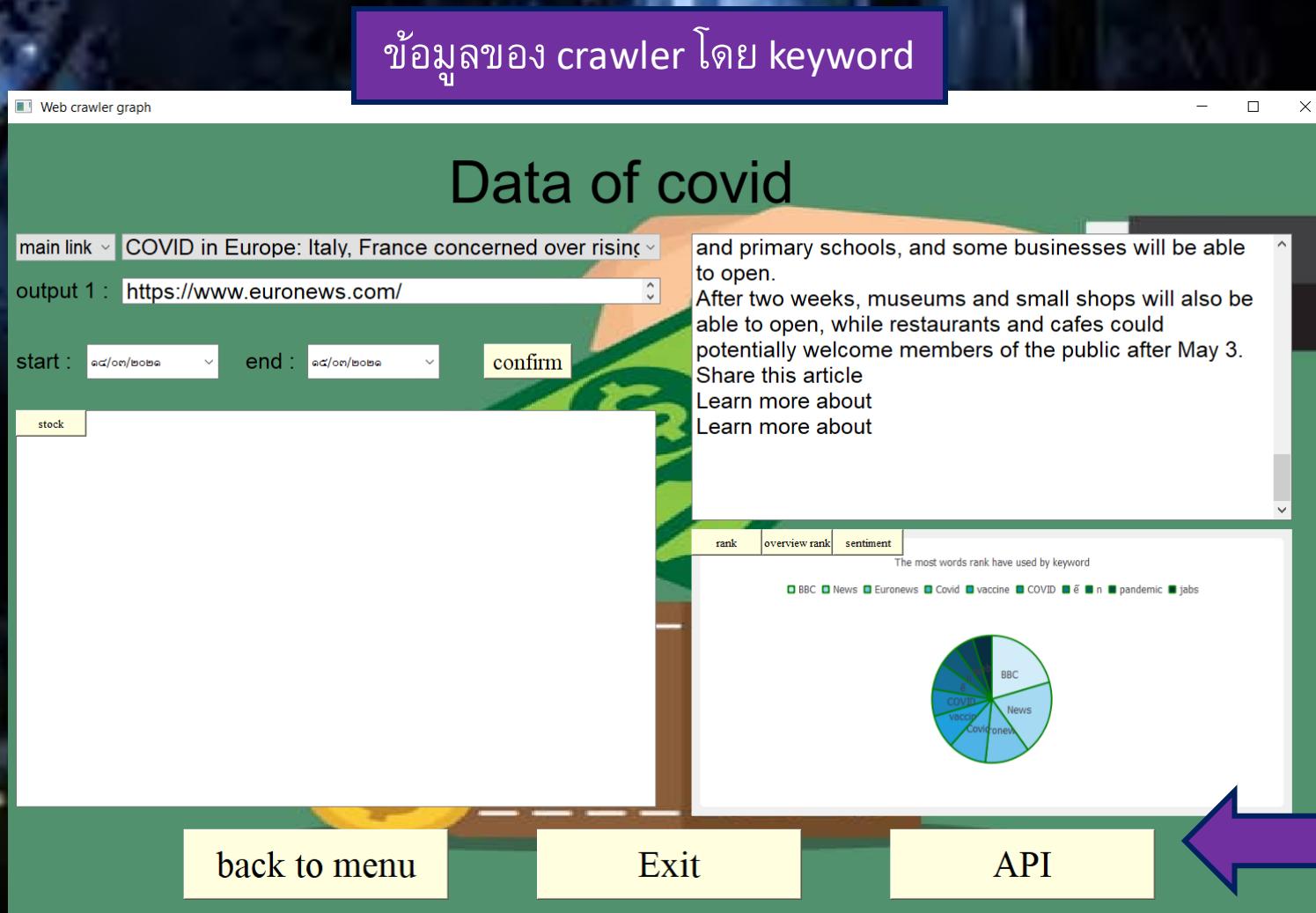
```
104     def on_click_play(self):
105         all_result = self.sreach_function()
106         #print(all_result)
107         Crawler_NLP(all_result)
108         self.switch_window.emit(self.input.text(),self.choice_1.currentText(),all_result)
109
110     def sreach_function(self) :
111         now = datetime.now()
112         date_time = now.strftime("%Y-%m-%d %H:%M:%S")
113         all_result = {}
114
115         news_file = pandas.read_csv('web_crawler.csv')
116         if self.choice_1.currentText() == "th" :
117             pass
118         elif self.choice_1.currentText() == "en" :
119             global_news = news_file['global'].dropna()
120             #print(global_news)
121             for news in global_news :
122                 #print(news)
123                 temp_name = str(news).split("https://")
124                 temp_name = temp_name[1][-1].split("/")
125                 name_file = temp_name[0] + '_crawler.csv'
126                 #print(name_file)
127
128                 data_file = open(name_file,'r', encoding="utf-8")
129                 read = csv.reader(data_file, delimiter = ',')
130                 data_frame = pandas.DataFrame(read)
131
132                 head_news = data_frame[0].str.lower()
133                 #print(head_news)
134                 head_keyword = data_frame[head_news.str.contains(self.input.text())]
135                 #print(head_keyword[0])
136                 all_result[name_file] = []
137                 for result in head_keyword[0] :
138                     all_result[name_file].append(result)
139                     #nlp = Crawler_NLP(head_keyword,0)
140                     #nlp.word_analysis('en')
141                     #print("____")
142
143         return all_result
```

ធា ពីរ លេក្ខាកាត់
ឱវមន

ដំឡើច្ចូនខាងក្រុង database



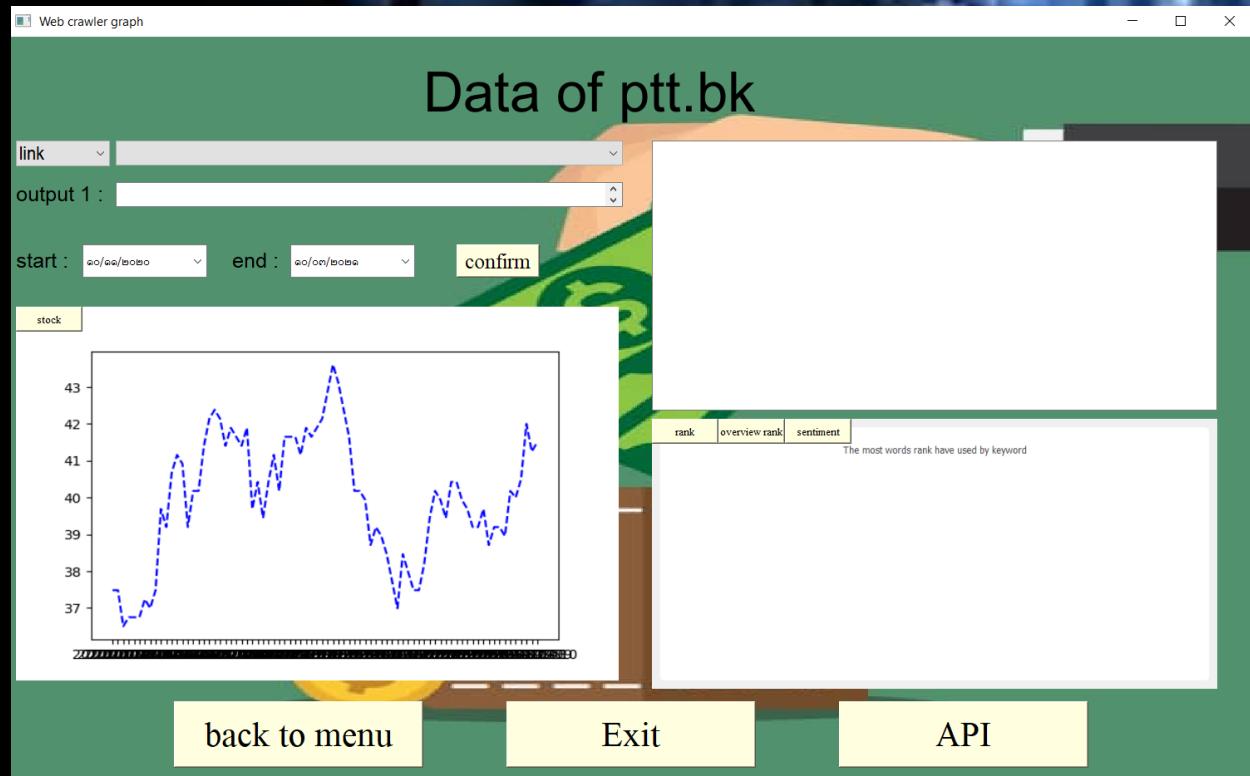
ส่วนของ GUI สำหรับ web crawler(ต่อ)



นำคำที่พูดมาก
ที่สุดมาเป็น
input ใน api

ສ້ານຂອງ stock function

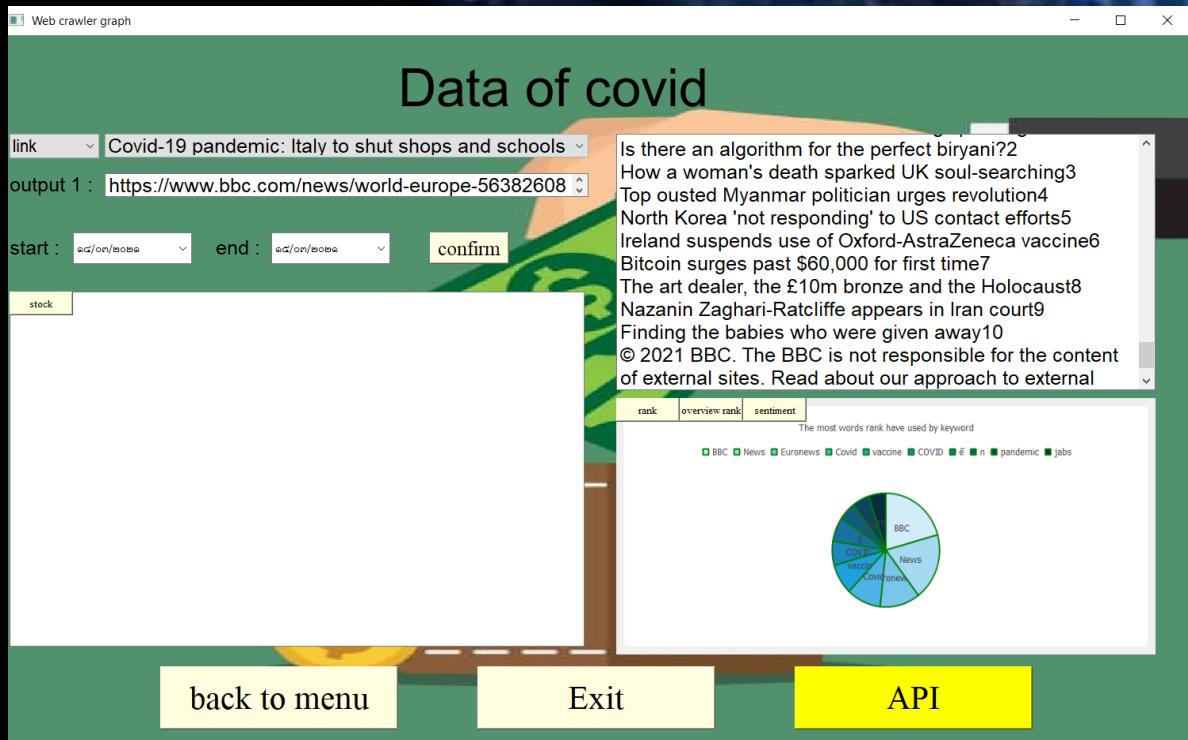
ຈະແສດງການມີເອກດັບນູ່



```
769     def crawler_stock(self) :
770         data_time_1 = self.time_label_1.date().toPyDate()
771         data_time_2 = self.time_label_2.date().toPyDate()
772
773         day_1 ,day_2 = str(data_time_1.day) ,str(data_time_2.day)
774         month_1 ,month_2 = str(data_time_1.month) ,str(data_time_2.month)
775         year_1 ,year_2 = str(data_time_1.year) ,str(data_time_2.year)
776
777         if len(day_1) == 1 or len(day_2) == 1 :
778             if len(day_1) == 1 :
779                 day_1 = '0' + day_1
780             if len(day_2) == 1 :
781                 day_2 = '0' + day_2
782         if len(month_1) == 1 or len(month_2) == 1 :
783             if len(month_1) == 1 :
784                 month_1 = '0' + month_1
785             if len(month_2) == 1 :
786                 month_2 = '0' + month_2
787
788         self.since = f'{year_1}-{month_1}-{day_1}'
789         self.until = f'{year_2}-{month_2}-{day_2}'
790
791         print(self.since,self.until)
792
793     try:
794         stock = pdr.get_data_yahoo(self.name, start = self.since, end = self.until)
795         save = open("crawler_stock.csv","w",newline = "")
796         save.write(stock.to_csv())
797         save.close()
798         ""array = str(ptt['Adj Close']).split("\n")
799         #print(array)
800         array_x = []
801         array_y = []
802         for i in array[1:-1]:
803             array_x.append(i.split(" ")[0])
804             array_y.append(i.split(" ")[2])
805
806     except:
807         print("Error:", sys.exc_info()[0])
808         print("Description:", sys.exc_info()[1])
809
810         """for x,y in zip(array_x,array_y) :
811             print(x, y)"""
812
813         output_file = pandas.read_csv("crawler_stock.csv")
```

สวนของการเชื่อมต่อระหว่าง crawler และ API

นำข้อมูลของ crawler ส่งไปให้ api



back to menu

Exit

API

```
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867

    def api_link(self):
        #print(type(self.top_rank_1))
        condi = self.make_data_base()
        if condi :
            self.switch_window_3.emit(self.top_rank_1,self.lang)
        else :
            pass

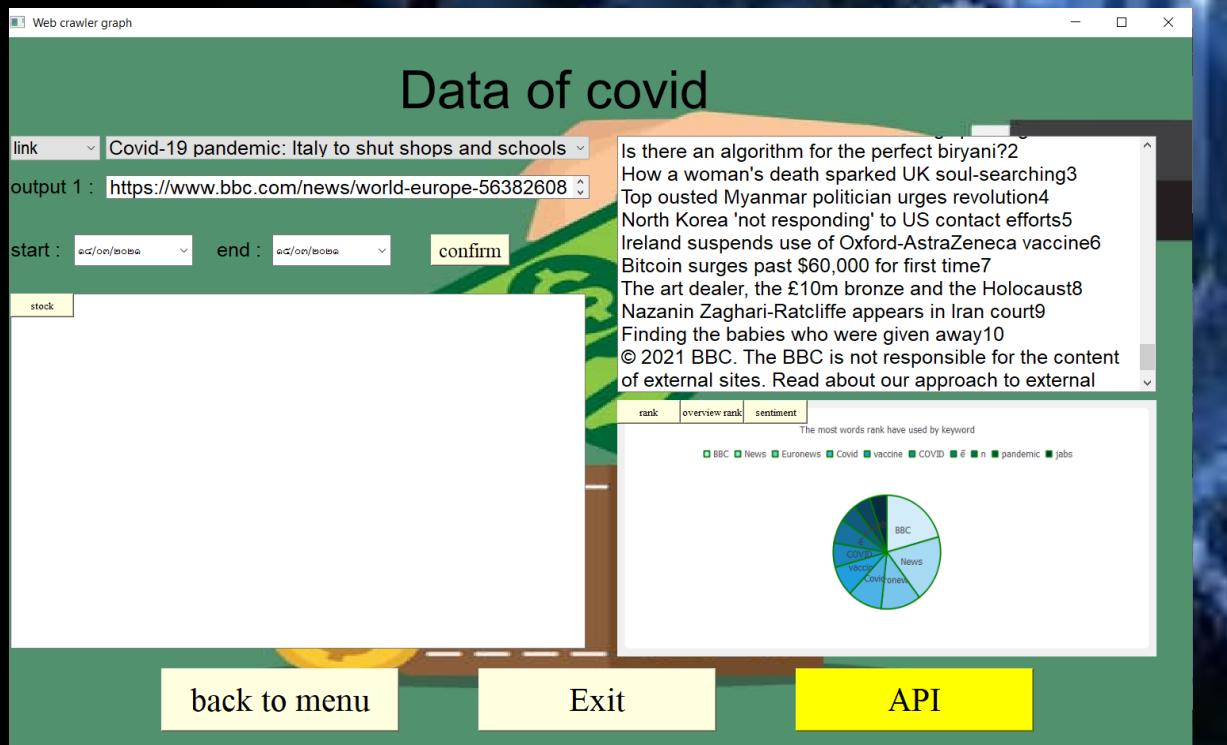
    def make_data_base(self) :
        now = datetime.now()
        date_time = now.strftime("%Y-%m-%d %H:%M:%S")
        head = ["date", "keyword"]

        tempfile = NamedTemporaryFile(mode='w', delete=False, newline='', encoding="utf-8")
        writer_re = csv.DictWriter(tempfile, fieldnames= head )
        writer_re.writeheader()
        tem_key = []
        try :
            keyword_check = pandas.read_csv('keyword_search.csv')
            for date,key in zip(keyword_check['date'],keyword_check['keyword']) :
                if key == self.top_rank_1 :
                    writer_re.writerow({'date':date_time, 'keyword' : self.top_rank_1 })
                else :
                    writer_re.writerow({'date':date, 'keyword' : key })
                tem_key.append(key)
        except FileNotFoundError :
            csvfile_for_head = open('keyword_search.csv', 'w', newline='', encoding="utf-8")
            writer_file = csv.DictWriter( csvfile_for_head, fieldnames=head )
            writer_file.writeheader()
            writer_file.writerow({'date':date_time, 'keyword' : self.top_rank_1 })
            csvfile_for_head.close()

        if self.top_rank_1 not in tem_key :
            run = Data_twitter(self.top_rank_1,self.lang)
            run.get_data()
            try :
                file_check = pandas.read_csv(str(self.top_rank_1) + '_api.csv')

                writer_re.writerow({'date':date_time, 'keyword' : self.top_rank_1 })
                tempfile.close()
                shutil.move(tempfile.name, 'keyword_search.csv')
```

ສ້າງຂອງການເຊື່ອມຕໍ່ອະຫວ່າງ crawler ແລະ API(ຕວ)



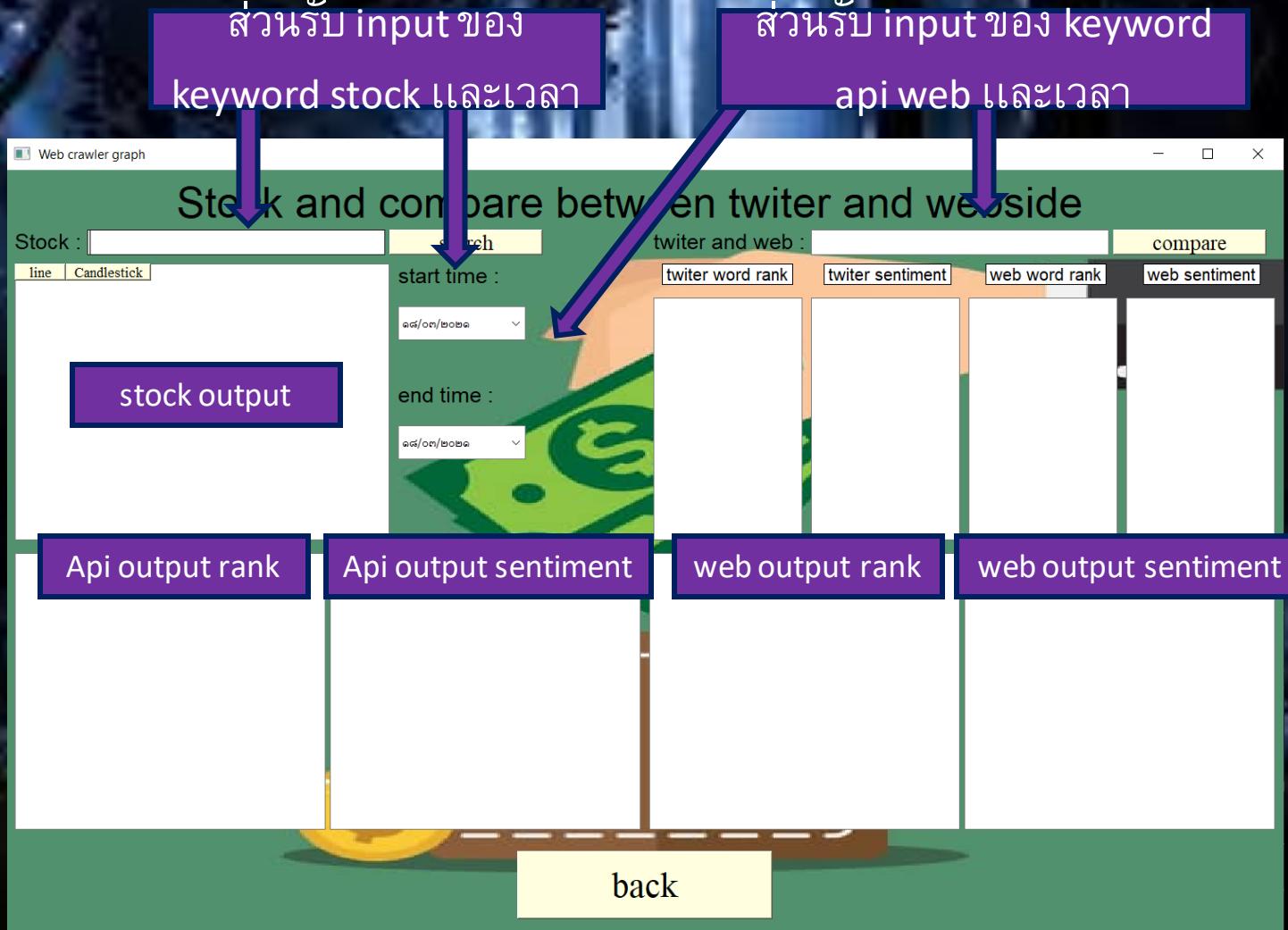
Thursday

18 / 03 / 2021

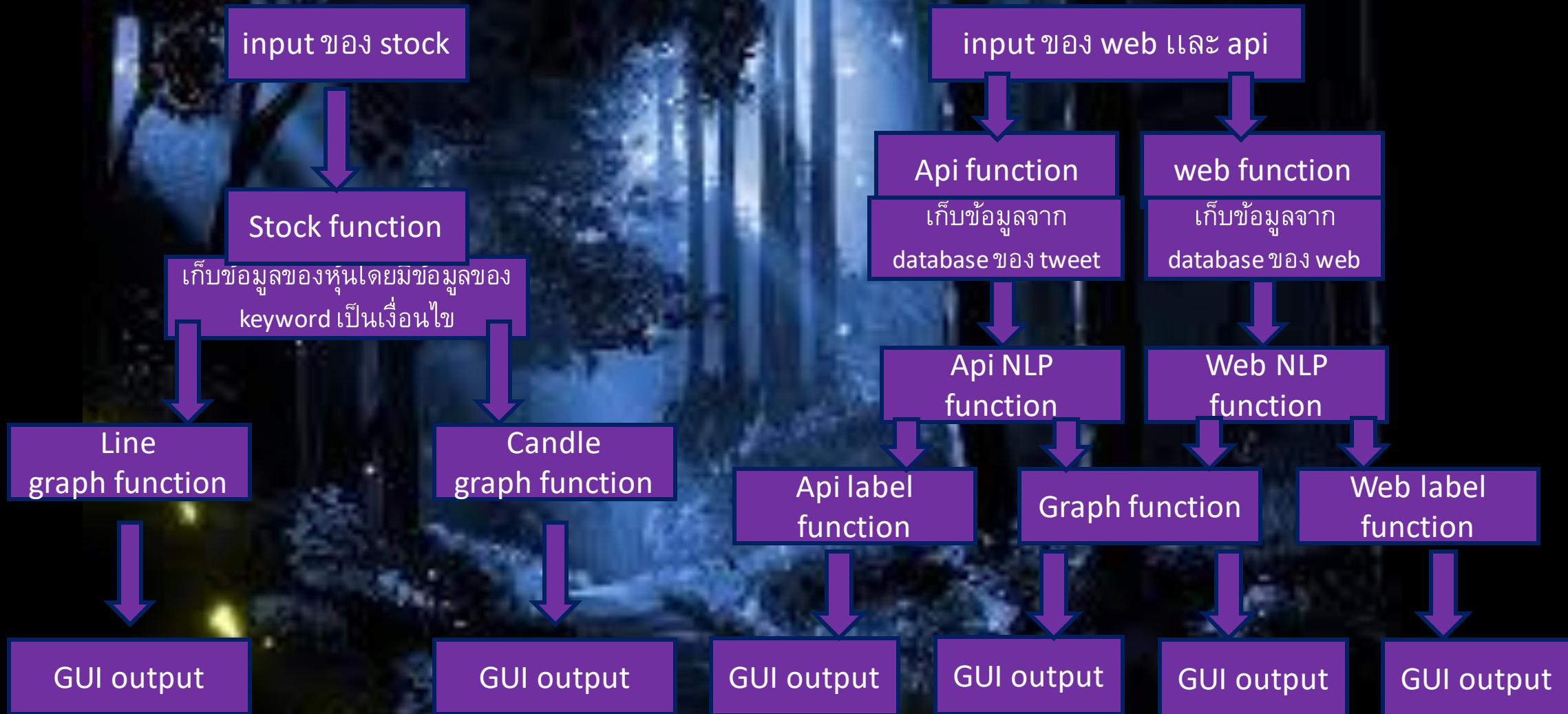
สิ่งที่จะ
UPDATE

- ทำ **GUI** สำหรับการเปรียบเทียบทั้งหมด
- เพิ่มกราฟ **CANDLE** สำหรับกราฟ
ของ **STOCK**

สวนของ GUI สำหรับการเปรียบเทียบทงหมด



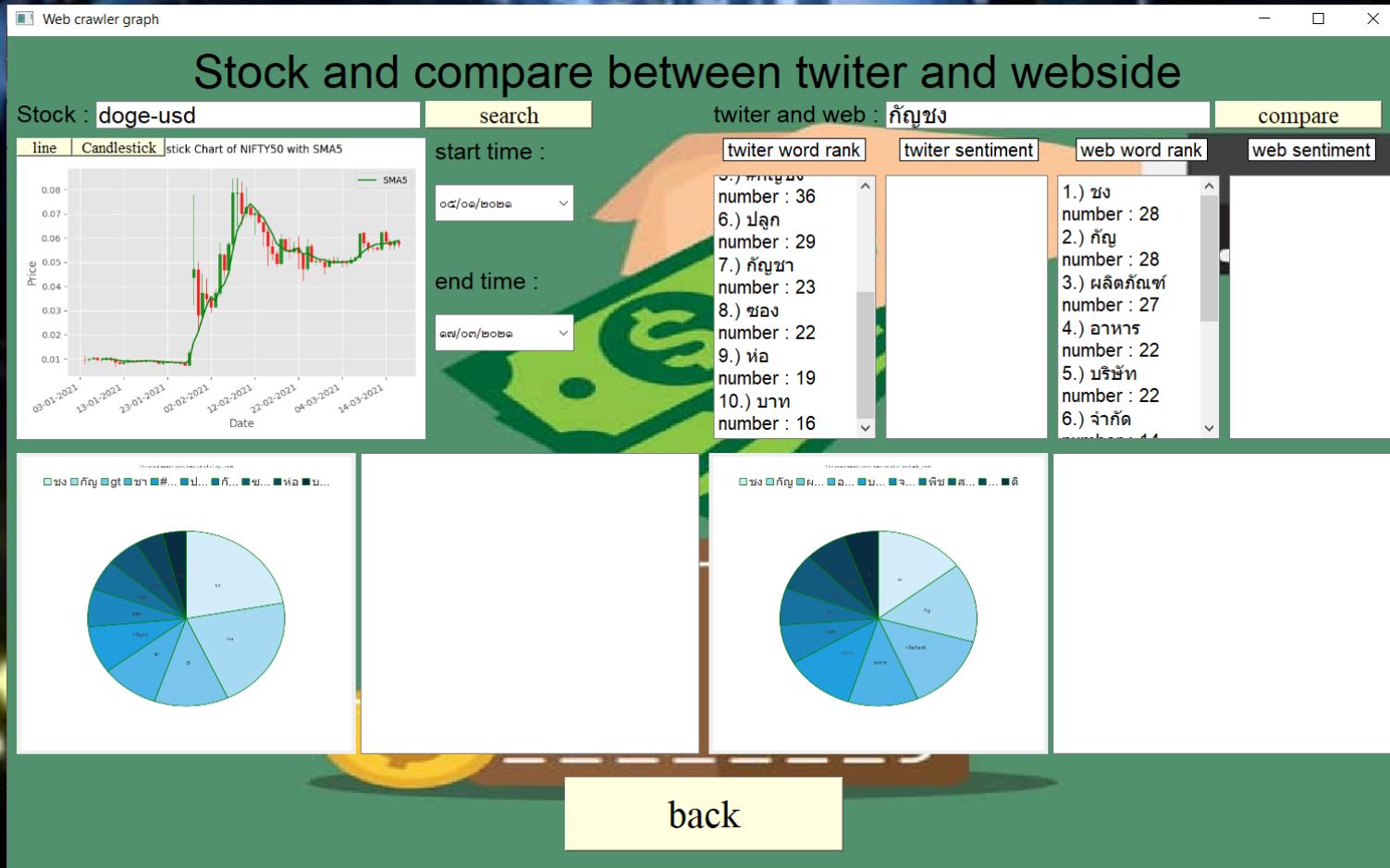
ส่วนของ GUI สำหรับการเปรียบเทียบทางหอด (ต่อ)



ສ່ວນຂອງ GUI ສໍາຮັບການປະເມີນເຖິງບທງໜົມດ(ຕອ)

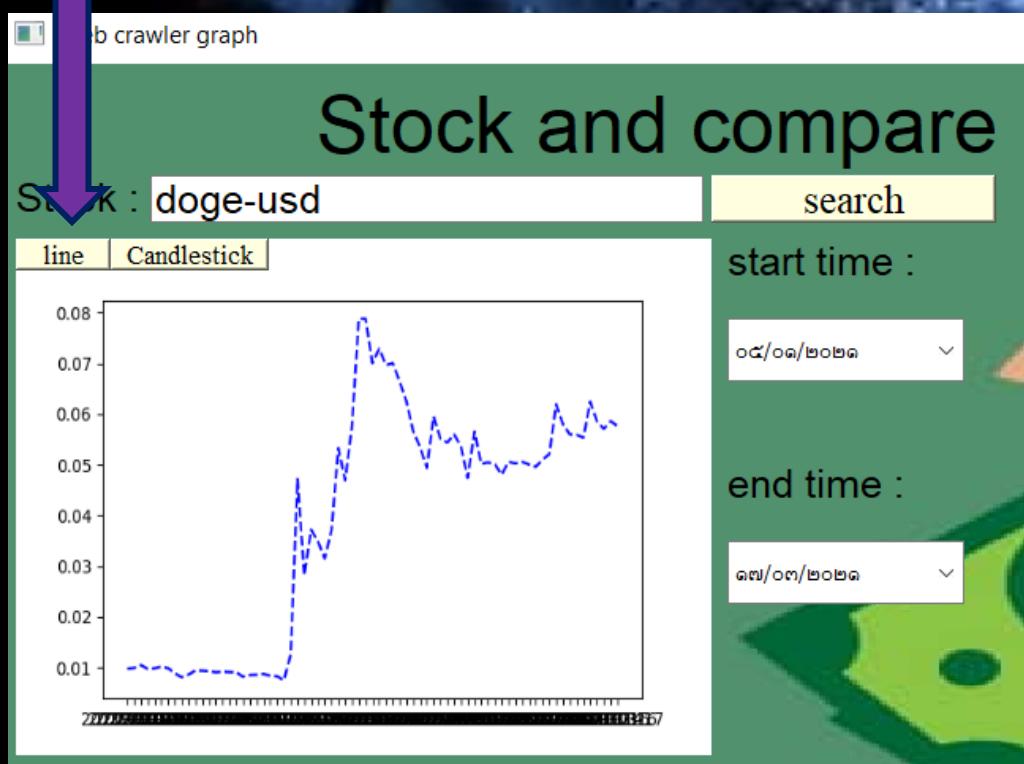
ສ່ວນຂອງ output ທີ່

ຈະເສັດງ

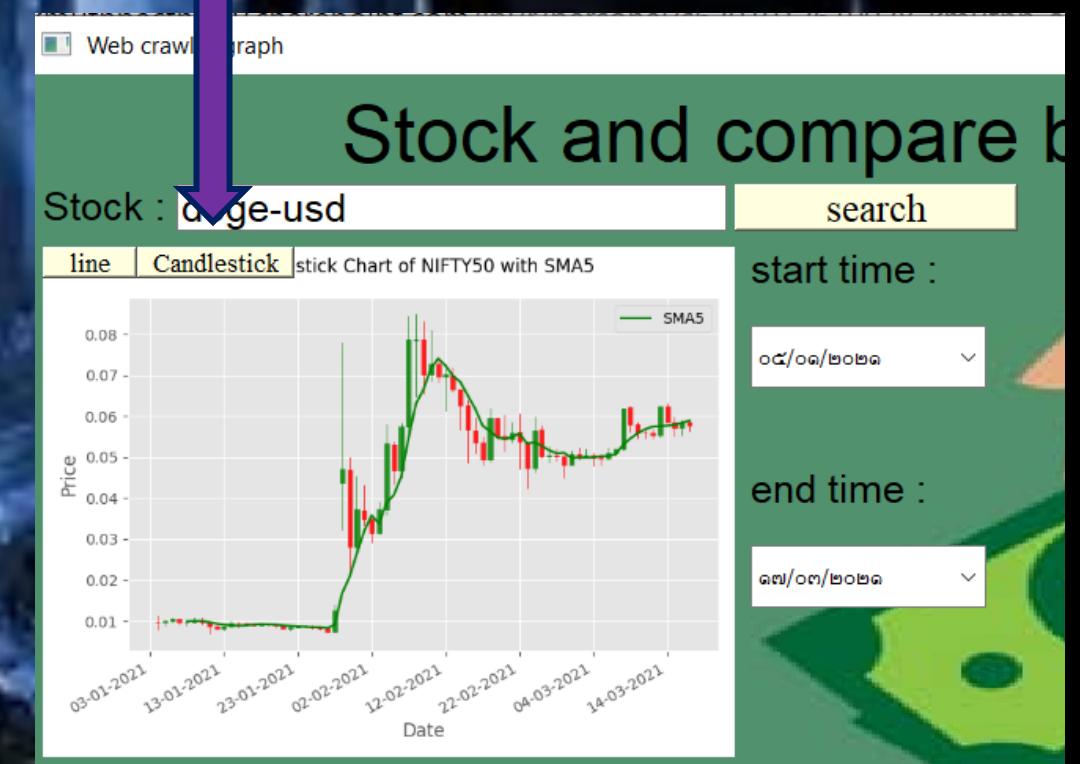


สวนของกราฟ candle สำหรับกราฟของ stock

ปุ่มเรียก function ของ
line graph



ปุ่มเรียก function ของ
candle graph



สวนของกราฟ candle สำหรับกราฟของ stock(ต่อ)

ปุ่มเรียก function ของ
line graph

ปุ่มเรียก function ของ
candle graph

Line
graph function

Candle
graph function

GUI output

GUI output

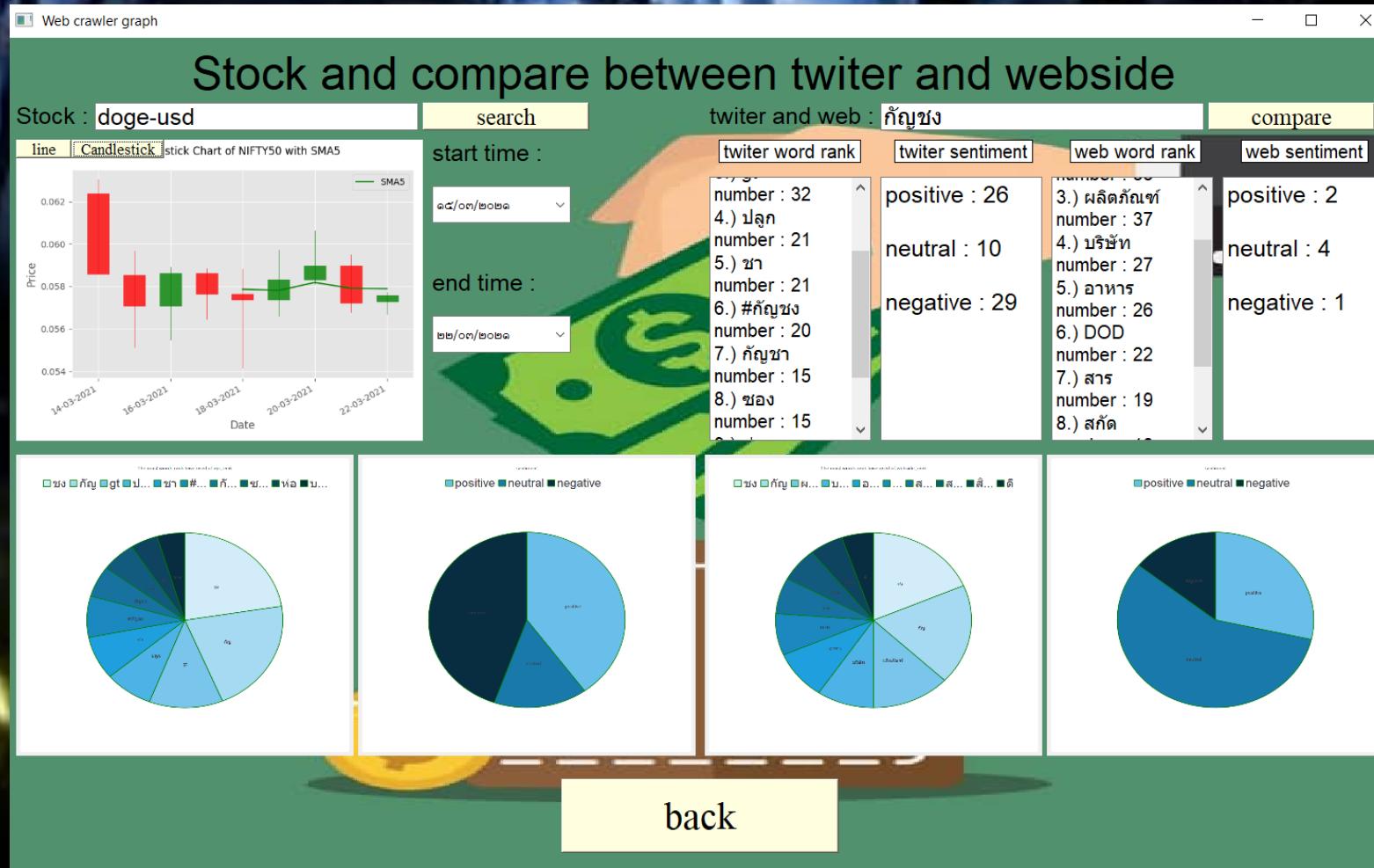
Monday

22 / 03 / 2021

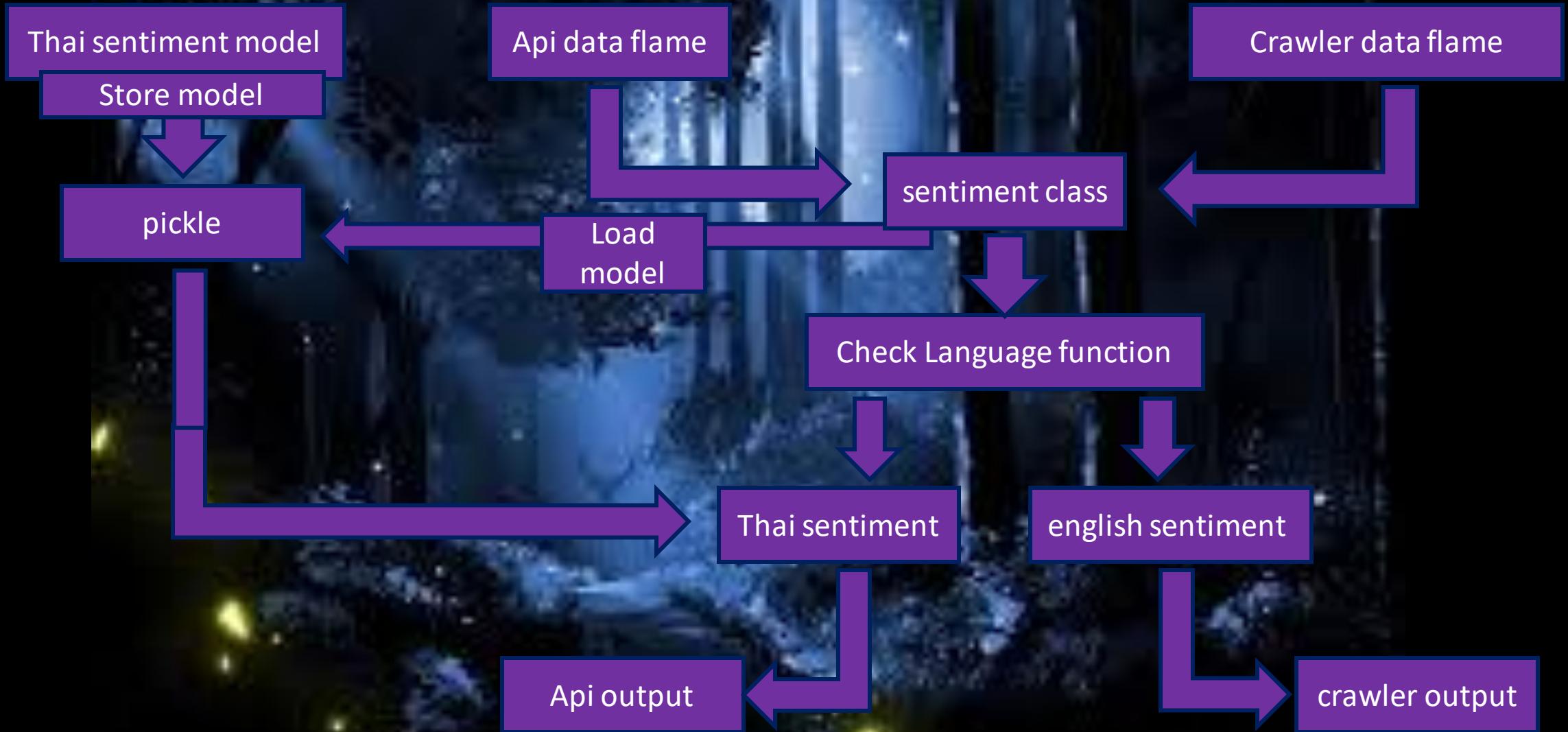
สิ่งที่จะ
UPDATE >

- ทำ **SENTIMENT** ห้องภาษาไทยและอังกฤษโดยที่ทำงานร่วมกันได้

สวนของ sentiment



ສ່ວນຂອງ sentiment



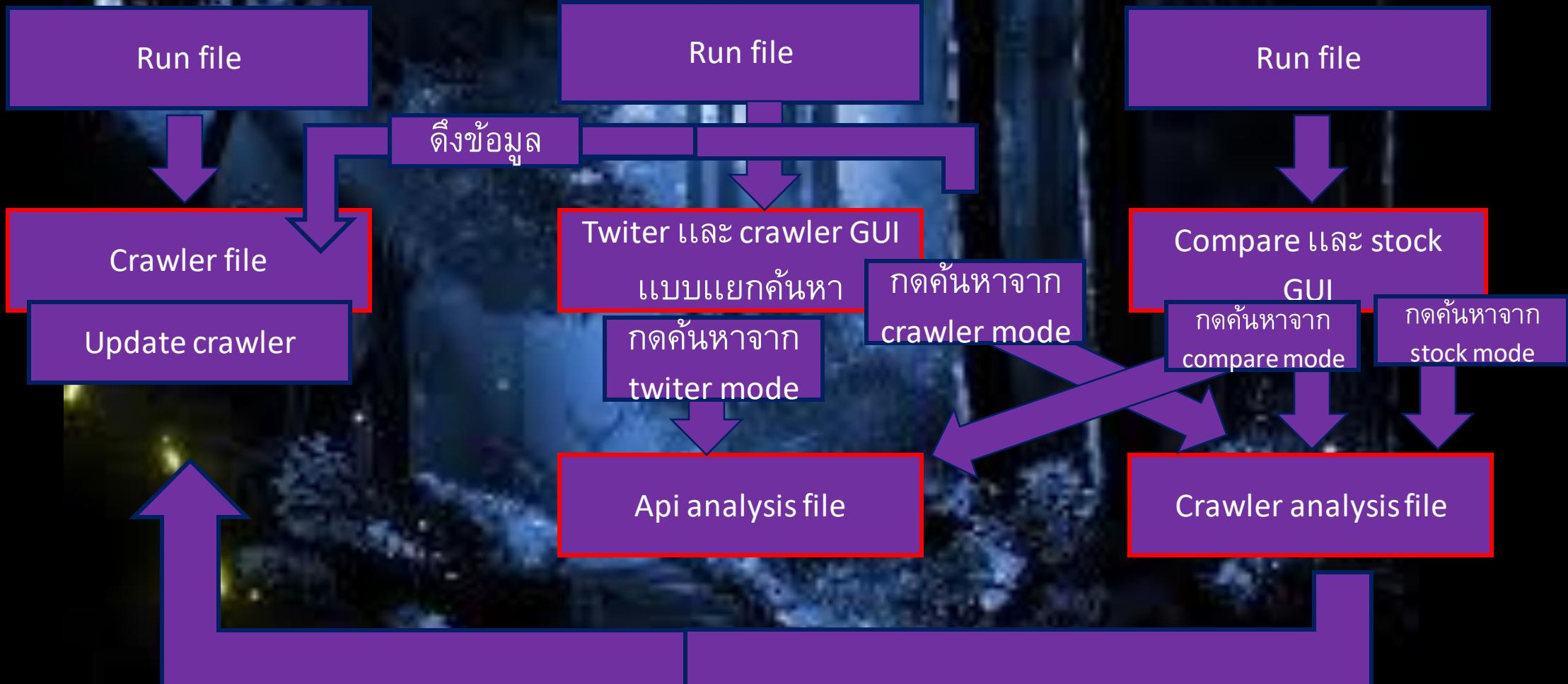
Monday
29 / 03 / 2021

Update week 13

สิ่งที่จะ
UPDATE >

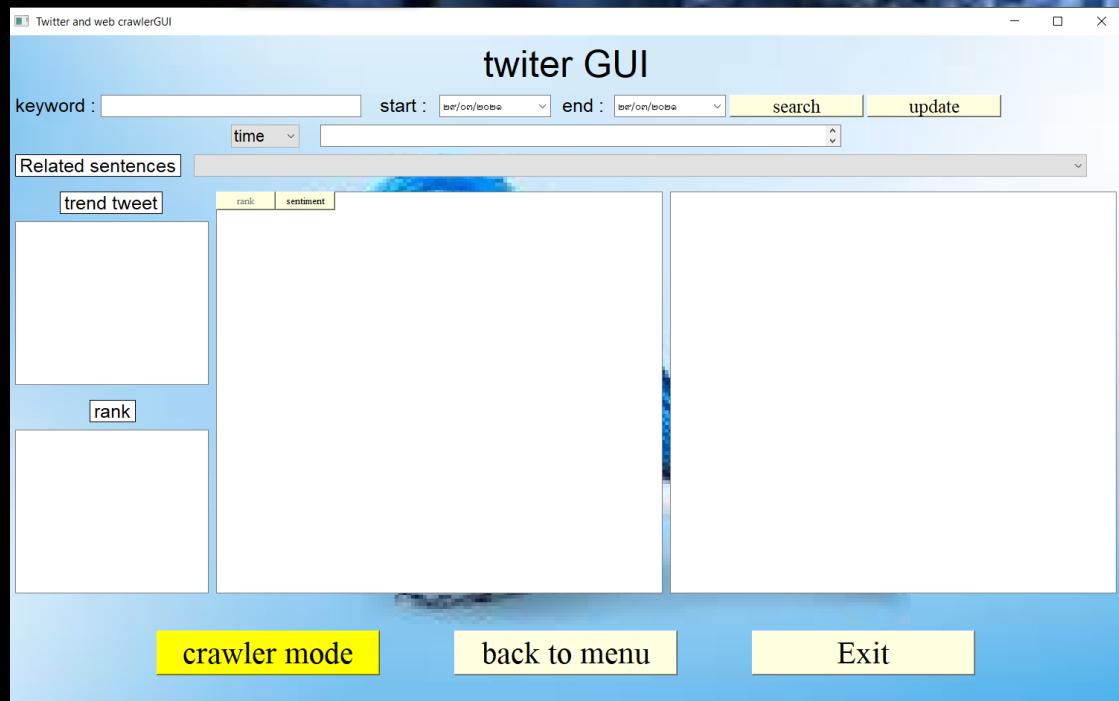
- แก้ไขการทำงานของโค๊ดและ **GUI**
ทั้งหมด

การทำงานของโค้ดทั้งหมด

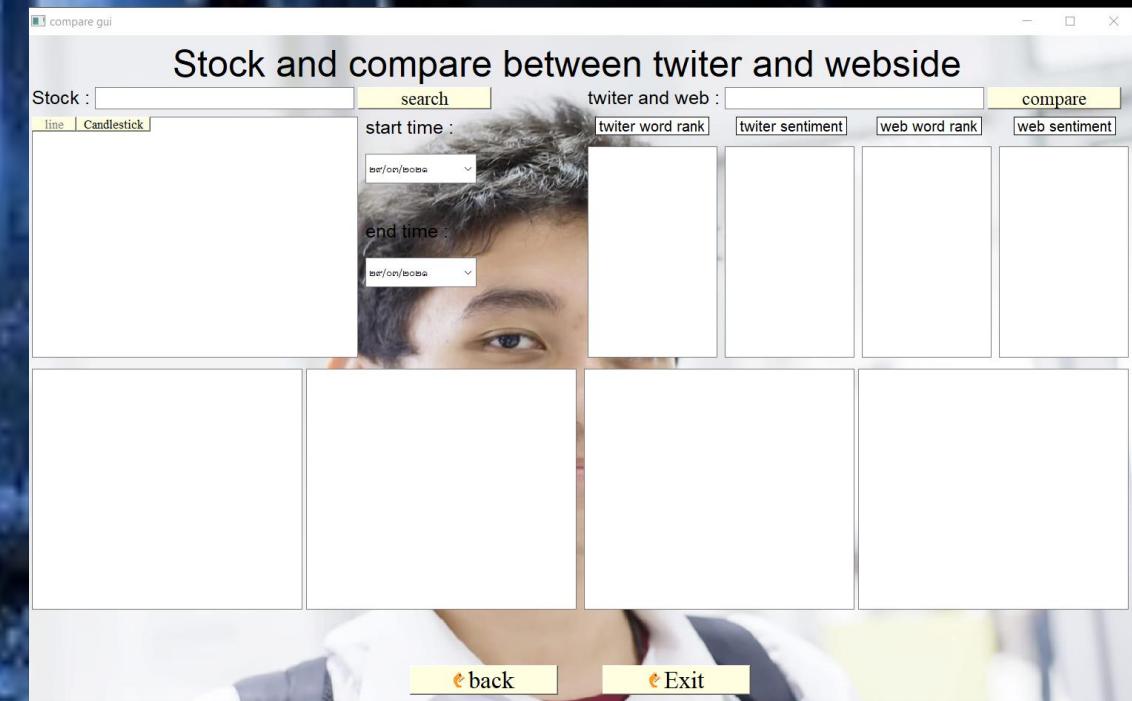


ภาพรวมของการทำงานของทงหมด

Twiter และ crawler GUI
แบบแยกคันหา



Compare และ stock
GUI



Update week 13

Thursday

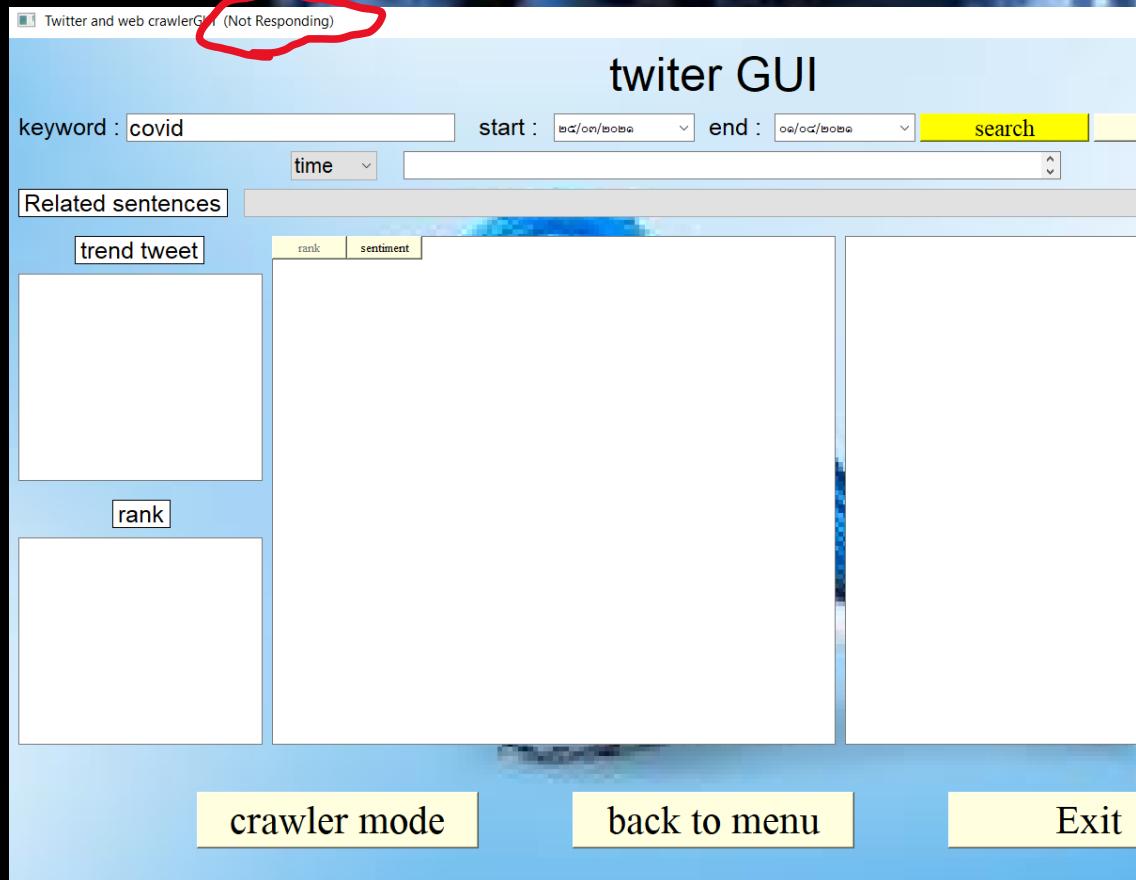
01 / 04 / 2021

ถึงที่จะ
UPDATE >

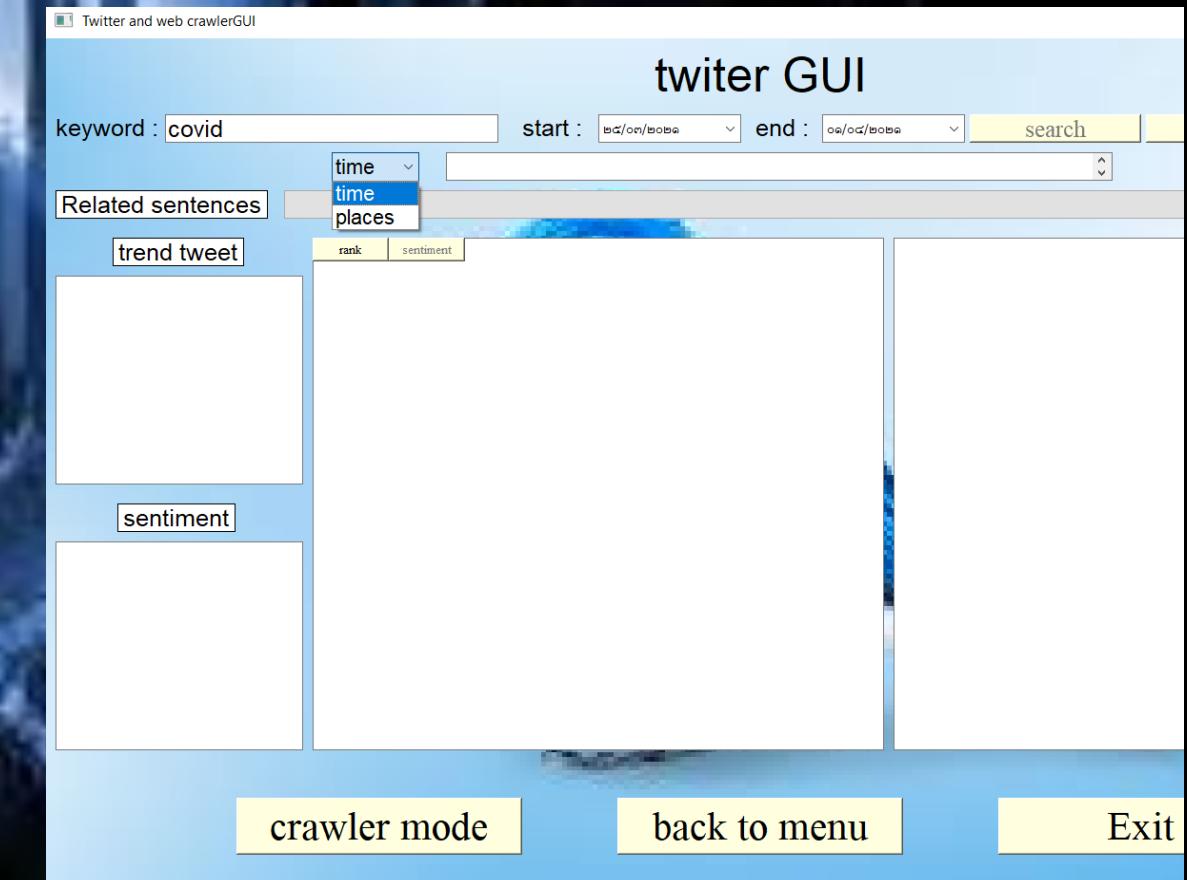
- นำ **QTHREAD** มาปรับใช้กับ **GUI** ไม่ให้ **NOT RESPOND** ตอนกำลังทำฟังก์ชันอื่น

การเปรียบเทียบการทำงานของ GUI

แบบไม่มี thread



แบบมี thread



การทำงานของ threading กับ GUI



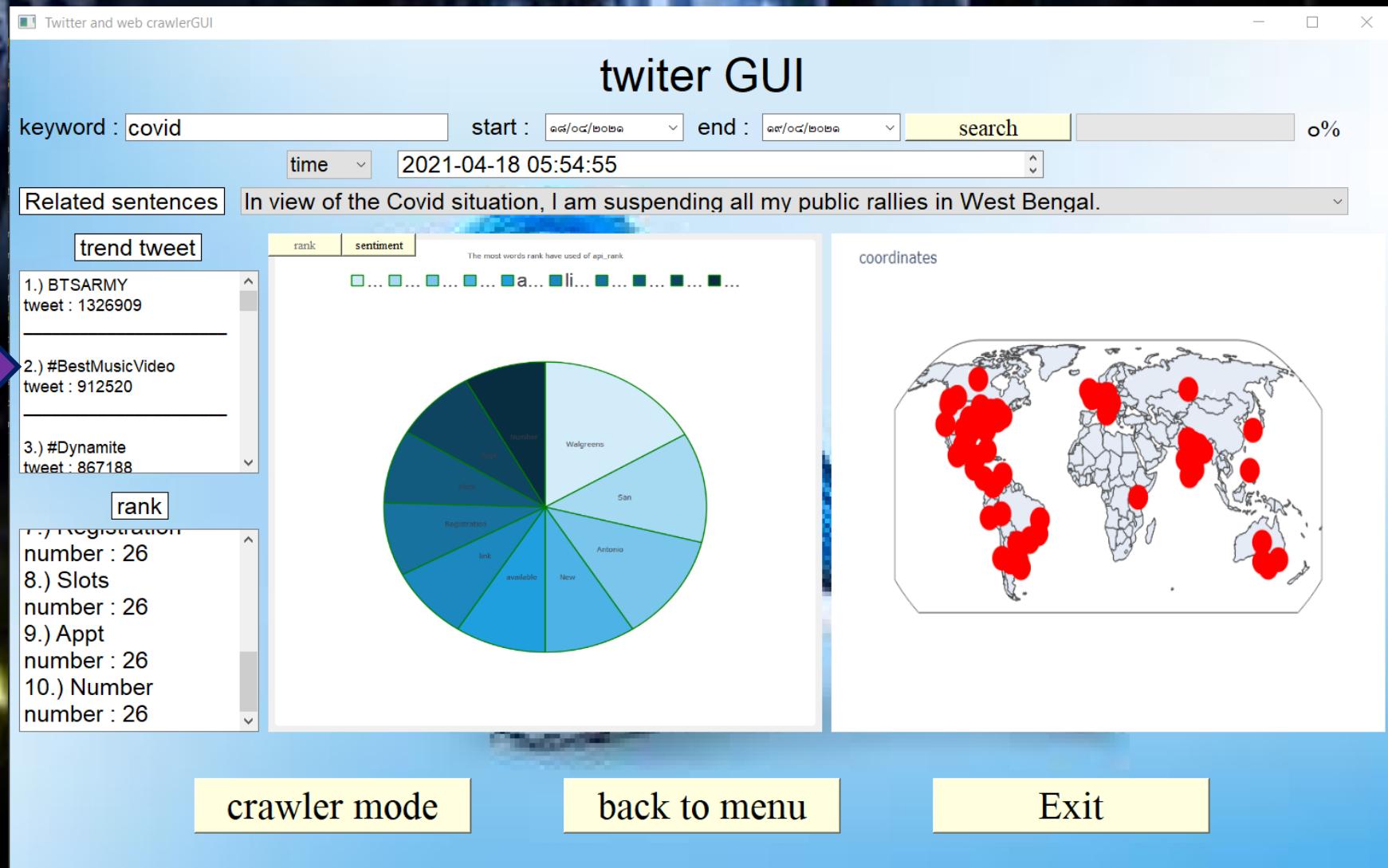
Monday
19 / 04 / 2021

สิ่งที่จะ
UPDATE >

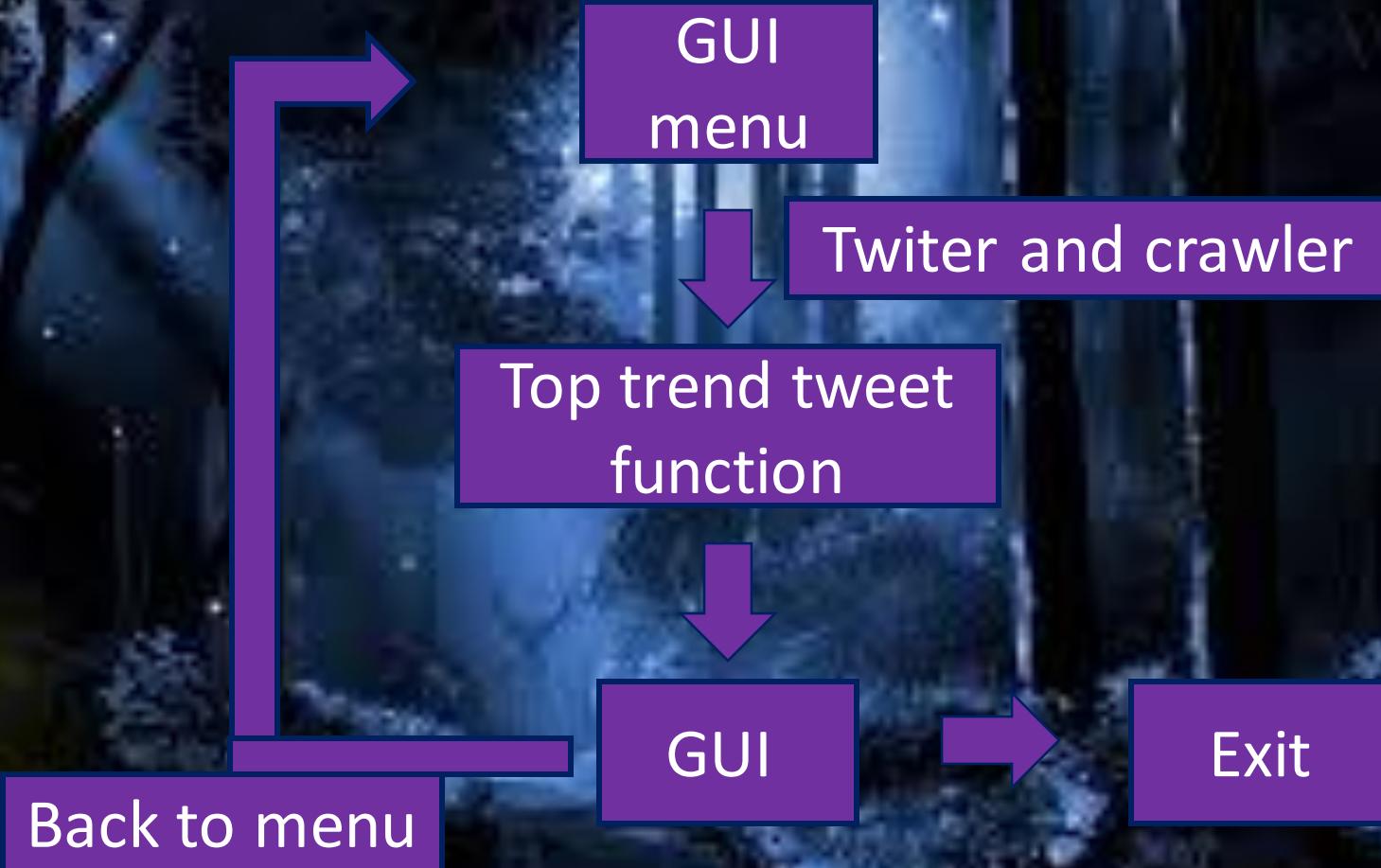
- เพิ่ม **TOP TRENDS TWEET** เข้ามาในหน้าต่าง **GUI**
- เพิ่ม **PROGRESS BAR** เข้ามาใน **GUI**
- สำหรับบวกกว่า **GUI** และการทำงานจาก
หลังยังทำงานอยู่
- เพิ่ม **MAP** เข้าไปใน **GUI** สำหรับแสดง

LOCATION ของผู้ส่ง **TWEET**

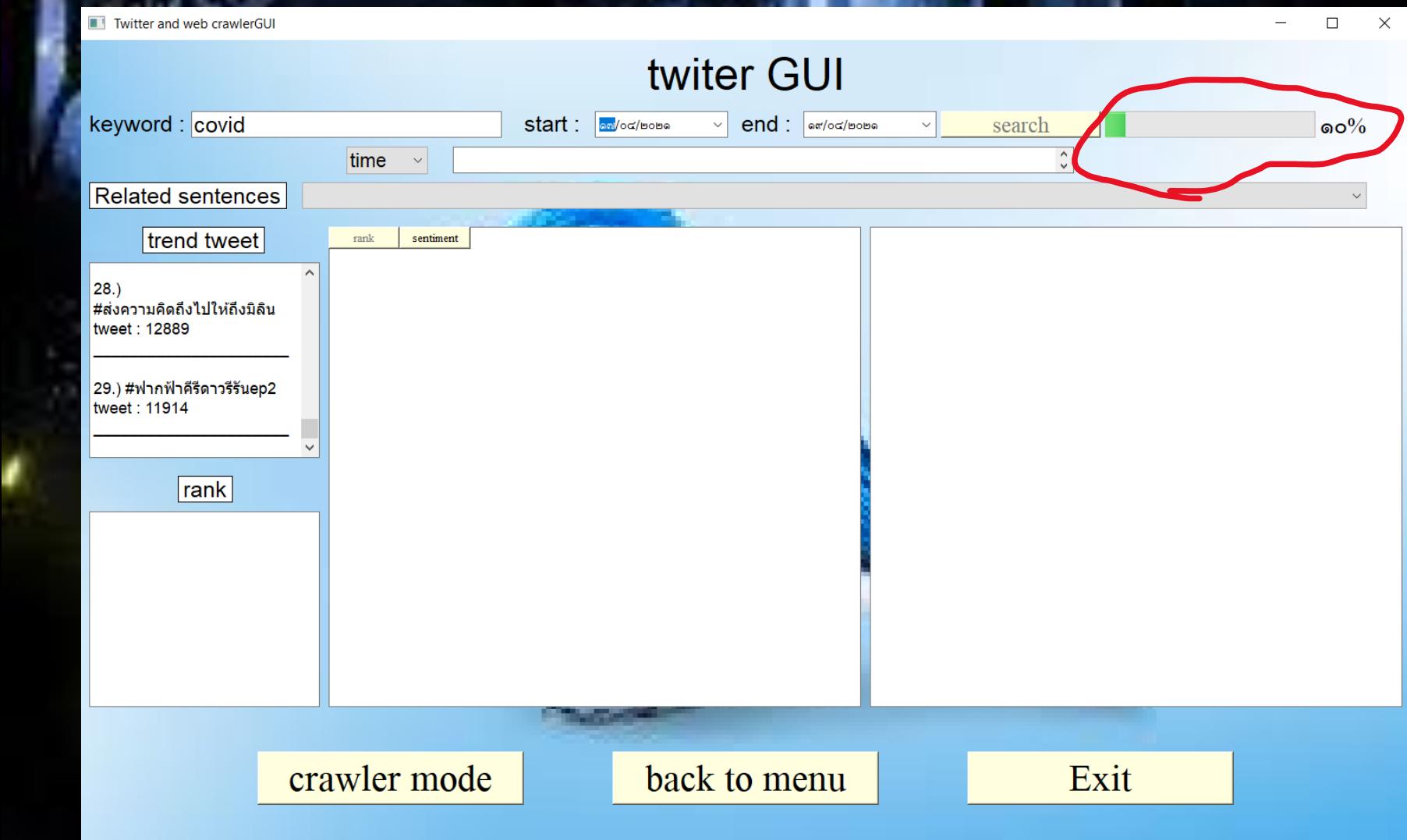
top trends tweet ຂອງ GUI



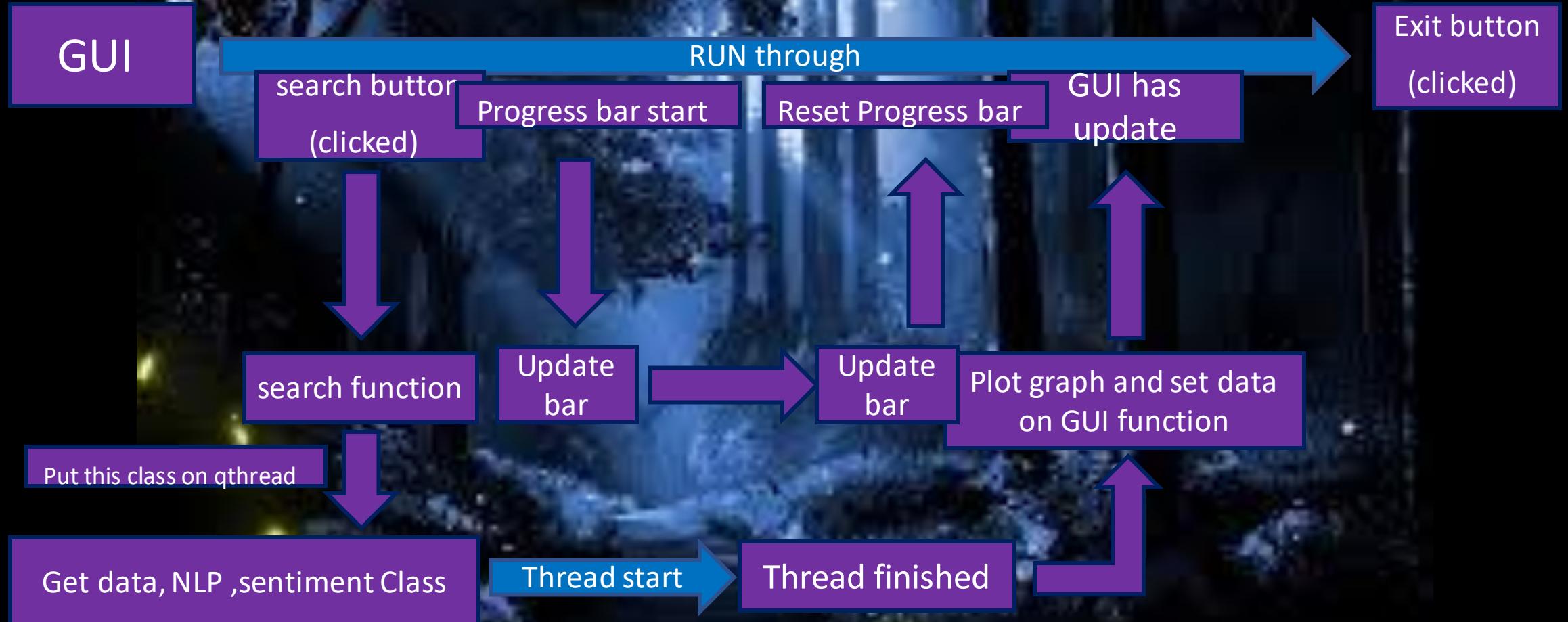
top trends tweet ของ GUI(ต่อ)



Progress bar



Progress bar(ទឹក)



การแสดง location บน map

Twitter and web crawlerGUI

twiter GUI

keyword : covid start : ၂၇/၀၄/၂၀၂၁ end : ၂၈/၀၄/၂၀၂၁ search ၀%

time 2021-04-18 05:54:55

Related sentences In view of the Covid situation, I am suspending all my public rallies in West Bengal.

trend tweet

- 1.) BTSARMY
tweet : 1326909
- 2.) #BestMusicVideo
tweet : 912520
- 3.) #Dynamite
tweet : 867188

rank

- 1.) Registration
number : 26
- 8.) Slots
number : 26
- 9.) Appt
number : 26
- 10.) Number
number : 26

sentiment

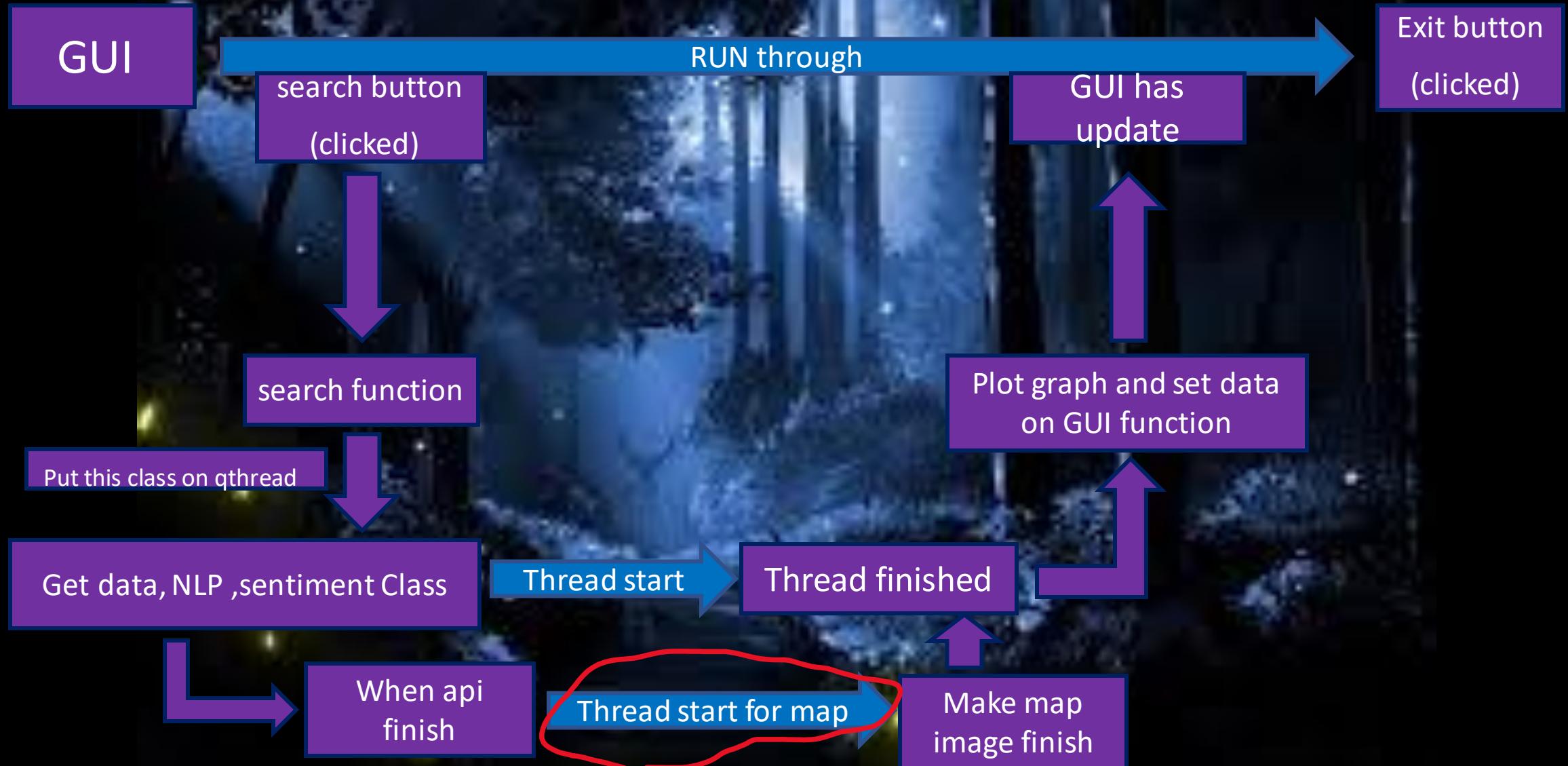
The most words rank have used of epi_rank

coordinates

Map ဗုံး GUI

crawler mode back to menu Exit

การแสดง location บน map(ต่อ)



การแสดง location บน map(ต่อ)

Thread
start

Find coordinate
function

Plot map and save
image function

Thursday

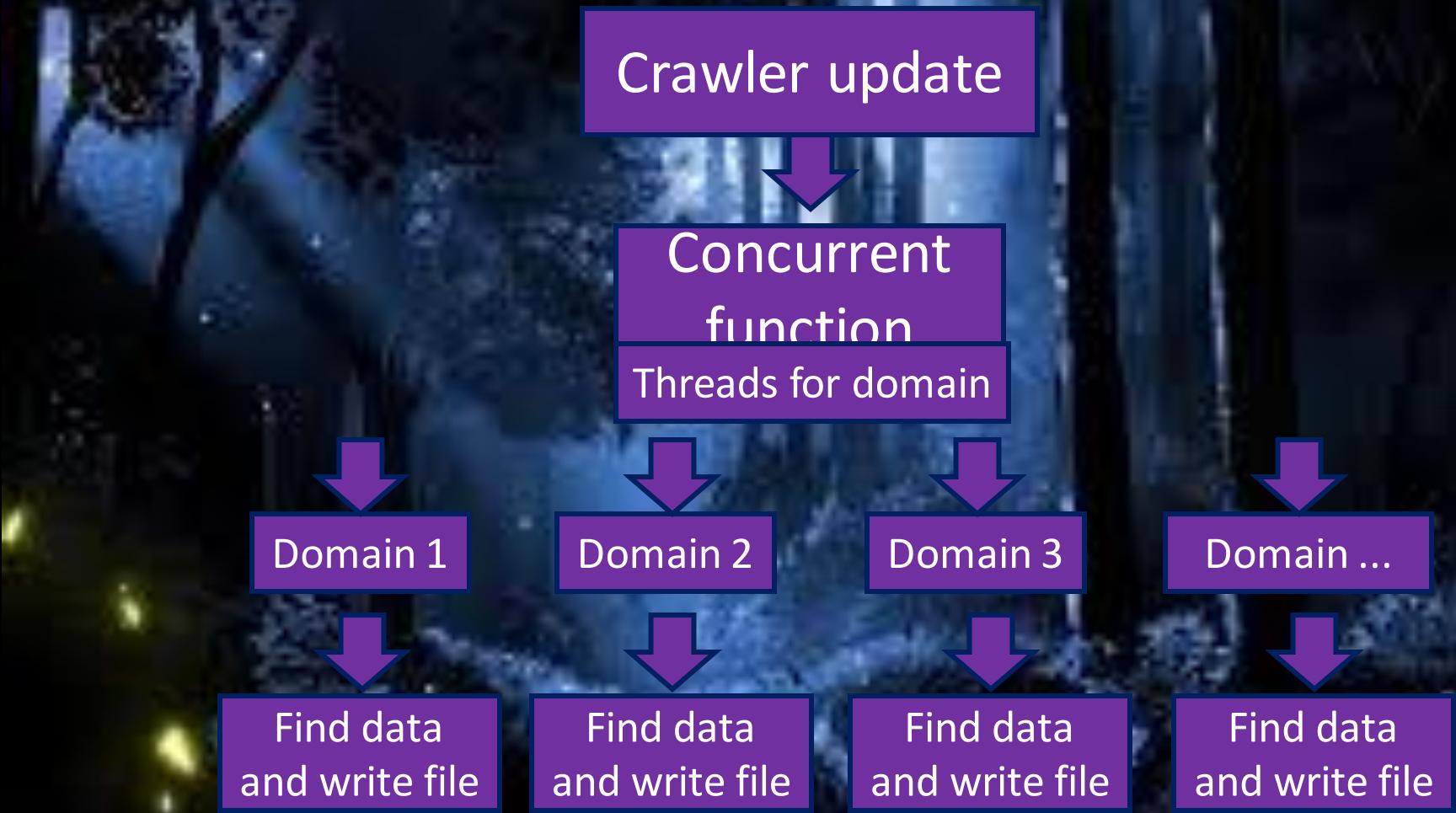
22 / 04 / 2021

Update last week

สิ่งที่จะ
UPDATE >

- แก้ไข **WEB CRAWLER** และเก็บ **FILE** ในรูปแบบวันที่
- แก้ไขการอ่าน **FILE** ของ **GUI** ให้เข้ากับรูปแบบ **FILE** ใหม่

แก้ไข web crawler และเก็บ file ในรูปแบบวันที่



ឧក្រិន់ការកែប file

before

 www.bangkokbizne...	19/04/2564 12:16	Microsoft Excel Co...	4,314 KB
 www.bbc.com_craw...	19/04/2564 12:41	Microsoft Excel Co...	2,802 KB
 www.euronews.com...	29/03/2564 09:32	Microsoft Excel Co...	1,013 KB
 www.nbcnews.com_...	29/03/2564 09:27	Microsoft Excel Co...	1,083 KB
 www.posttoday.co...	19/04/2564 12:36	Microsoft Excel Co...	3,132 KB
 www.reuters.com_cr...	19/04/2564 12:43	Microsoft Excel Co...	780 KB
 www.thairath.co.th_...	19/04/2564 12:15	Microsoft Excel Co...	861 KB
 www.themoscowti...	29/03/2564 09:30	Microsoft Excel Co...	608 KB

after

 crawler_2021-03-28_	22/04/2564 01:01	Microsoft Excel Co...	11 KB
 crawler_2021-03-29_	22/04/2564 10:13	Microsoft Excel Co...	5 KB
 crawler_2021-03-30_	22/04/2564 01:02	Microsoft Excel Co...	5 KB
 crawler_2021-03-31_	22/04/2564 10:15	Microsoft Excel Co...	14 KB
 crawler_2021-04-01_	22/04/2564 10:29	Microsoft Excel Co...	30 KB
 crawler_2021-04-02_	22/04/2564 10:15	Microsoft Excel Co...	19 KB
 crawler_2021-04-03_	22/04/2564 01:01	Microsoft Excel Co...	9 KB
 crawler_2021-04-04_	22/04/2564 10:32	Microsoft Excel Co...	43 KB
 crawler_2021-04-05_	22/04/2564 10:14	Microsoft Excel Co...	10 KB
 crawler_2021-04-06_	22/04/2564 10:32	Microsoft Excel Co...	34 KB
 crawler_2021-04-07_	22/04/2564 10:50	Microsoft Excel Co...	75 KB
 crawler_2021-04-08_	22/04/2564 10:36	Microsoft Excel Co...	54 KB
 crawler_2021-04-09_	22/04/2564 10:49	Microsoft Excel Co...	55 KB
 crawler_2021-04-10_	22/04/2564 10:30	Microsoft Excel Co...	26 KB
 crawler_2021-04-11_	22/04/2564 10:16	Microsoft Excel Co...	15 KB
 crawler_2021-04-12_	22/04/2564 10:35	Microsoft Excel Co...	64 KB
 crawler_2021-04-13_	22/04/2564 10:30	Microsoft Excel Co...	13 KB
 crawler_2021-04-14_	22/04/2564 10:49	Microsoft Excel Co...	68 KB
 crawler_2021-04-15_	22/04/2564 10:54	Microsoft Excel Co...	60 KB
 crawler_2021-04-16_	22/04/2564 14:24	Microsoft Excel Co...	195 KB
 crawler_2021-04-17_	22/04/2564 10:47	Microsoft Excel Co...	131 KB
 crawler_2021-04-18_	22/04/2564 10:46	Microsoft Excel Co...	140 KB
 crawler_2021-04-19_	22/04/2564 10:54	Microsoft Excel Co...	433 KB
 crawler_2021-04-20_	22/04/2564 14:32	Microsoft Excel Co...	873 KB
 crawler_2021-04-21_	22/04/2564 14:32	Microsoft Excel Co...	2,584 KB
 crawler_2021-04-22_	22/04/2564 14:35	Microsoft Excel Co...	4,472 KB

การอ่าน file ของ GUI

Crawler function

Keyword and time

Find time range

Read all file on time
range and sentiment

Return data

Thursday

29 / 04 / 2021

Summary week

ALL
FEATURE ➤

- ระบบการทำงานของ **GUI**
- การทำงานของ **TWITTER API** และการจัดการข้อมูล
- การจัดการข้อมูลเกี่ยวกับ **CRAWLER**
- การหาข้อมูลจาก **100 WEBSITE** และเก็บเป็น **DATABASE**
- **UNIT TEST** สำหรับตรวจสอบการทำงานของ **FUNCTION** ต่างๆ

ระบบการทำงานของ GUI

การทำงานทั้งหมด
ของ GUI ประกอบ
ไปด้วย

Main GUI

Twitter and
crawler GUI

Compare GUI

ทำหน้าที่เป็นหน้า menu

สำหรับเชื่อมระหว่าง

twitter and crawler GUI
กับ compare GUI

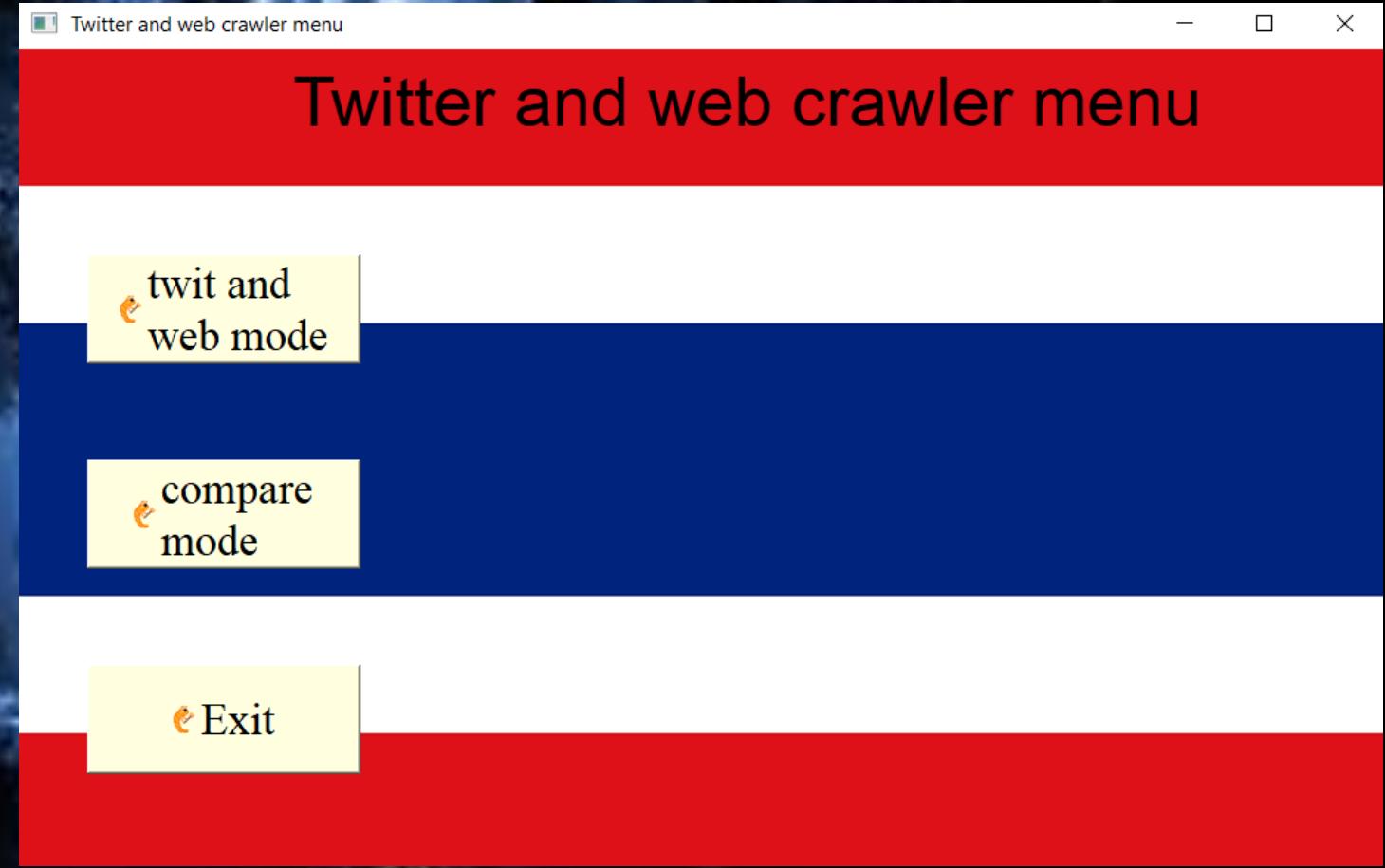
ทำหน้าที่สำหรับแสดง
ข้อมูลต่างๆ ของ crawler
และ twitter

ทำหน้าที่สำหรับการ
แสดงผลของ rank word
และ sentiment ของทั้ง
twitter และ crawler

ส่วนของ main menu

- > ประกอบด้วย <
- ปุ่ม twit and web mode
สำหรับเรียก
twit and web GUI
- ปุ่ม compare
สำหรับเรียก compare
GUI
- ปุ่ม Exit สำหรับปิดตัว
Menu GUI

ระบบการทำงานของ GUI(ต่อ)

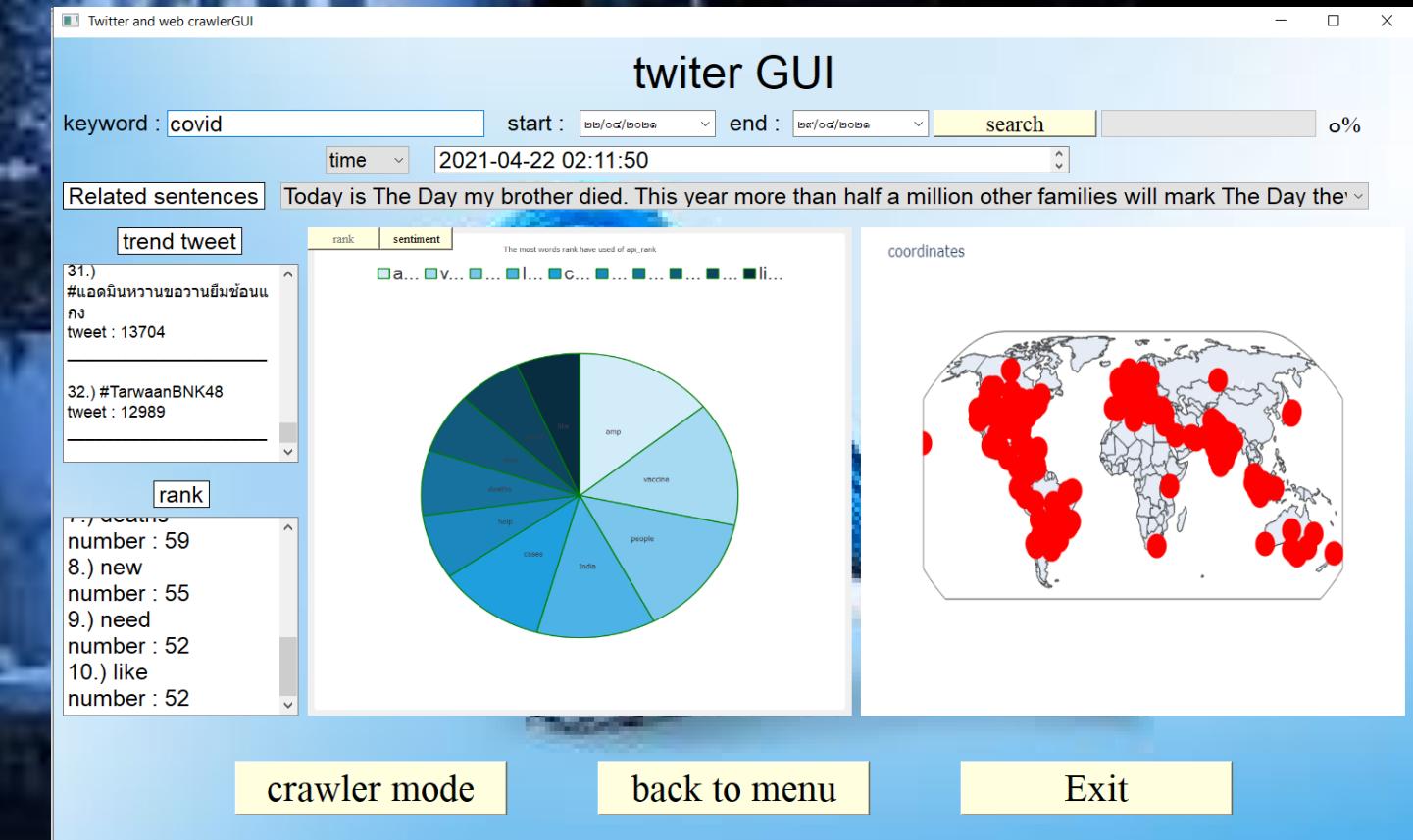


ส่วนของ twitter

> ประกอบด้วย <

- ส่วนกรอก keyword
- ส่วนเลือกช่วงเวลา
- ส่วนดูข้อความ tweet ต่างๆที่ search ได้
- ส่วนดูข้อมูลต่างๆของ Tweet ที่เลือก
- ส่วนดู top trend tweet
- ส่วนดูอันดับคำที่ใช้มากที่สุดและกราฟ
- ส่วนดู sentiment และกราฟ
- ส่วนที่ใช้ดู map
- ปุ่มสำหรับ web mode
- ปุ่มกลับไปยังหน้า menu
- ปุ่มปิด GUI

ระบบการทำงานของ GUI(ต่อ)

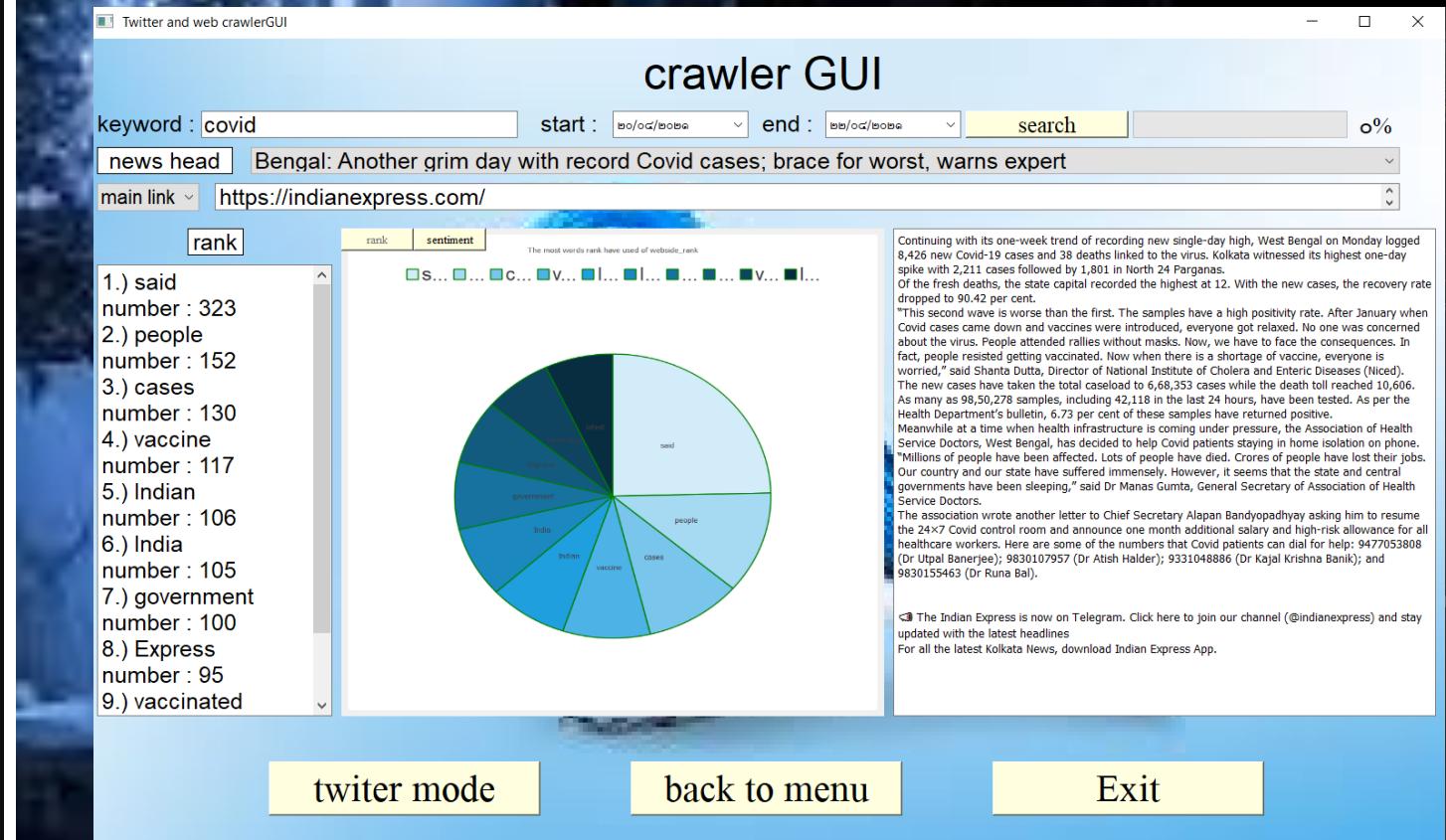


ส่วนของ webside

> ประกอบด้วย <

- ส่วนกรอก keyword
- ส่วนเลือกช่วงเวลา
- ส่วนดูข้อมูลต่างๆที่ search ได้
- ส่วนเลือกประเภทของข้อมูลที่จะดู
- ส่วนดูอันดับคำที่ใช้มากที่สุดและกราฟ
- ส่วนดู sentiment และกราฟ
- ปุ่มสำหรับ twitter mode
- ปุ่มกลับไปยังหน้า menu
- ปุ่มปิด GUI

ระบบการทำงานของ GUI(ต่อ)

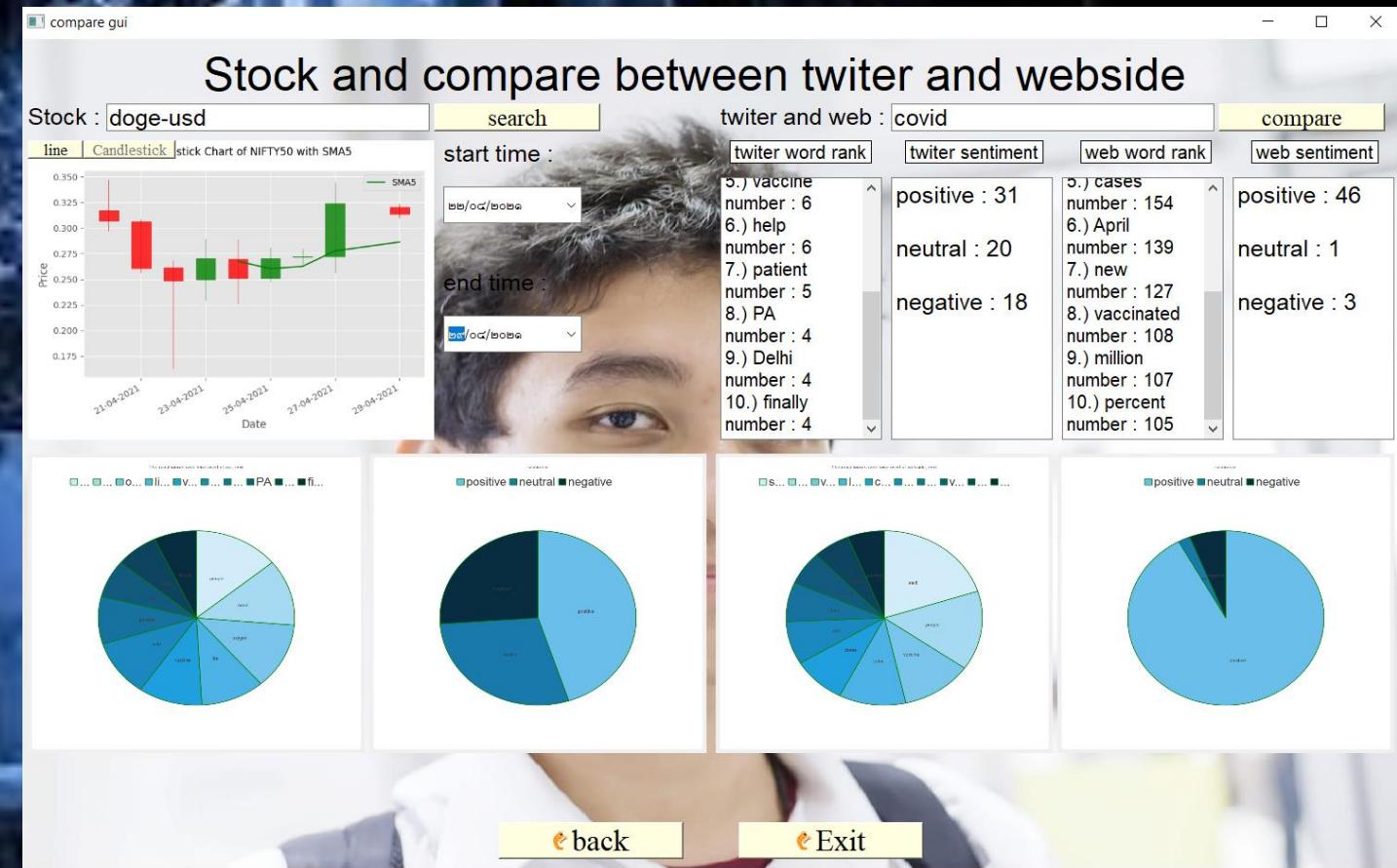


ส่วนของ compare

> ประกอบด้วย <

- ส่วนกรอก keyword ของ compare
- ส่วนกรอก keyword ของ stock
- ส่วนเลือกช่วงเวลา
- ส่วนดูกราฟ stock
- ส่วนดูอันดับคำที่ใช้มากที่สุดและกราฟ
- ส่วนดู sentiment และกราฟ
- ปุ่มกลับไปยังหน้า menu
- ปุ่มปิด GUI

ระบบการทำงานของ GUI(ต่อ)



การทำงานของ twitter api และการจัดการข้อมูล

โดยจะมี function ในการทำงานดังนี้

- 1.) api class คือ class ที่ทำหน้าที่ดึงข้อมูลมาจากการ twitter มาเก็บข้อมูลลงใน database file ตาม keyword
- 2.) NLP class(twit ver.) คือ class ที่ทำหน้าที่ตัดคำที่มาจากการ tweet ต่างๆ และนับหาว่ามีการใช้คำให้มากที่สุด 10 อันดับ
- 3.) top trend tweet function คือ function ที่ทำหน้าที่ดึงข้อมูล top tweet มาเก็บไว้ใน file สำหรับเรียกใช้ภายหลัง
- 4.) geopy function คือ function ที่ทำหน้าที่แปลงชื่อสถานที่เป็นพิกัดบนโลกและเก็บไว้ใน file สำหรับเรียกใช้ภายหลัง
- 5.) plotly function คือ function ที่ทำหน้าที่นำพิกัดสถานที่ต่างๆ มา plot ลงในแผนที่โลกและ save เก็บเป็น file
- 6.) sentiment class คือ class ที่ทำหน้าที่วิเคราะห์ประโยคแต่ละประโยคว่าเป็นประโยคลักษณะใดเป็นบวก, ลบหรือกลาง

การทำงานของ twitter api และการจัดการข้อมูล(ต่อ)

ส่วนการดึงข้อมูลมาจาก data base และ
การ update(search and update function)

โดยตัว function จะนำ keyword และช่วงเวลาไปตรวจสอบใน data base ก่อนว่ามีข้อมูลครบทุกวันหรือไม่
ถ้าไม่จะทำการเรียก api class เพื่อดึงข้อมูลตามช่วงเวลาที่ขาดไปมาจาก twitter เลยและ update เพิ่มลงไปใน
Data base file ตามชื่อ keyword ที่ search

การจัดการข้อมูลเกี่ยวกับ crawler

โดยจะมี function ในการทำงานดังนี้

1.) NLP class(web ver.) คือ class ที่ทำหน้าที่ตัดคำที่มาจากการประযุกต์ต่างๆ และนับหาว่ามีการใช้คำที่ไหนมากที่สุด 10 อันดับ

2.) search function คือ function ที่ทำหน้าที่ดึงข้อมูลใน data base ของ crawler โดยดึงมาเฉพาะช่วงเวลา และ keyword ที่ต้องการ

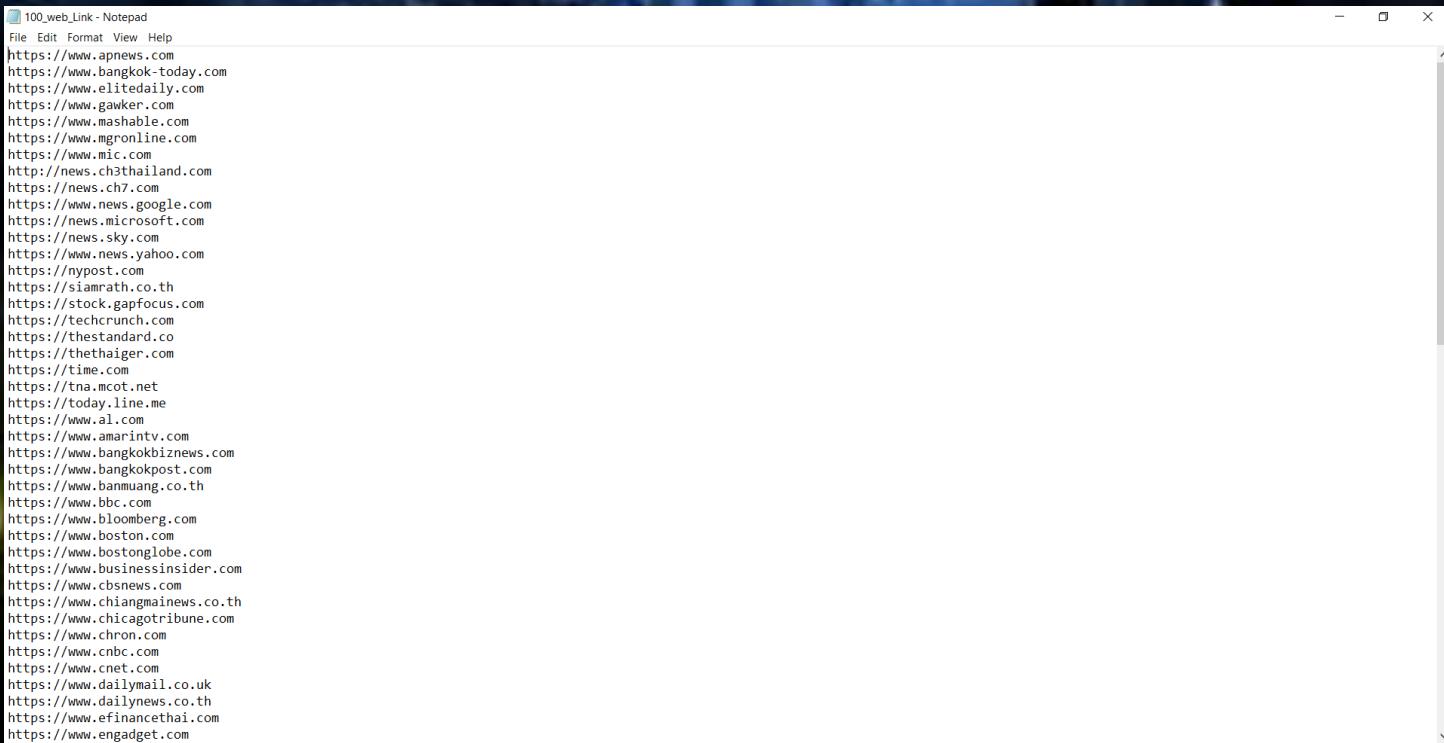
3.) sentiment function จะใช้ function เดียวกันกับของ tweet

4.) stock function คือ function ที่ทำหน้าที่ดึงข้อมูล stock จาก yahoo การ save ไว้เป็น file สำหรับจัดการข้อมูล

การหาข้อมูลจาก 100 webside

File ที่เก็บ link ของ webside ข่าวต่างๆ

ลงใน text file สำหรับเรียกใน crawler
update class



```
100_web_Link - Notepad
File Edit Format View Help
https://www.apnews.com
https://www.bangkok-today.com
https://www.elitedaily.com
https://www.gawker.com
https://www.mashable.com
https://www.mgronline.com
https://www.mic.com
http://news.ch3thailand.com
https://news.ch7.com
https://www.news.google.com
https://news.microsoft.com
https://news.sky.com
https://news.yahoo.com
https://nypost.com
https://siamrath.co.th
https://stock.gapfocus.com
https://techcrunch.com
https://thestandard.co
https://thethaiger.com
https://time.com
https://tma.mcot.net
https://today.line.me
https://www.al.com
https://www.amarintv.com
https://www.bangkokbiznews.com
https://www.bangkokpost.com
https://www.bannuang.co.th
https://www.bb.com
https://www.bloomberg.com
https://www.boston.com
https://www.bostonglobe.com
https://www.businessinsider.com
https://www.cbsnews.com
https://www.chiangmainews.co.th
https://www.chicagotribune.com
https://www.chron.com
https://www.cnbc.com
https://www.cnet.com
https://www.dailymail.co.uk
https://www.dailynews.co.th
https://www.efinancethai.com
https://www.engadget.com
```

การหาข้อมูลจาก 100 webside(ต่อ)

ส่วนของ crawler update class

โดยตัว function จะทำหน้าที่นำ link ของเต็ลล์ webside จำนวน 100 link มาหาว่าใน webside นั้นมีหัวข้อข่าวอะไรบ้างจากนั้นก็จะกดเข้าไปในหัวข้อข่าวนั้นๆเพื่อที่จะไปหารายละเอียดต่างๆ เช่น เนื้อข่าว วันที่ลงข่าว เมื่อข้อมูลครบแล้วก็จะทำการเก็บไว้ใน Data base file โดยเก็บข้อมูลแยกเป็นวันละ file

Unit test

test ในส่วนของ twitter

test ในส่วนของ crawler

ตรวจสอบ output ของเต'
และ function ที่เกี่ยวกับการ
จัดการข้อมูลของ twitter

ตรวจสอบ output ของเต'
และ function ที่เกี่ยวกับการ
จัดการข้อมูลของ crawler

Reference ที่ใช้งานมาม

<https://www.geeksforgeeks.org/python-programming-language/>

<https://stackoverflow.com/>

<https://www.kaggle.com/drvaibhavkumar/twitter-data-analysis-using-tweepy?fbclid=IwAR2zxKUWji0NNJANG2iKIfyZj2xWZ7zhrWtIRbfY6XnkzNTgKuCSnh7lbLg>

<https://docs.tweepy.org/en/latest/index.html>