

An Application of Natural Language Processing on forex gold spot News Analysis การประยุกต์ใช้การประมวลผลภาษาธรรมชาติเพื่อการวิเคราะห์ข่าวตลาด forex gold spot

ผู้เสนอโครงงาน นายชวลชัย อภิชาติฐิติวรณ์ รหัสนักศึกษา 6210110646

> อาจารย์ที่ปรึกษาโครงงาน รศ.ดร.มนตรี กาญจนะเดชะ

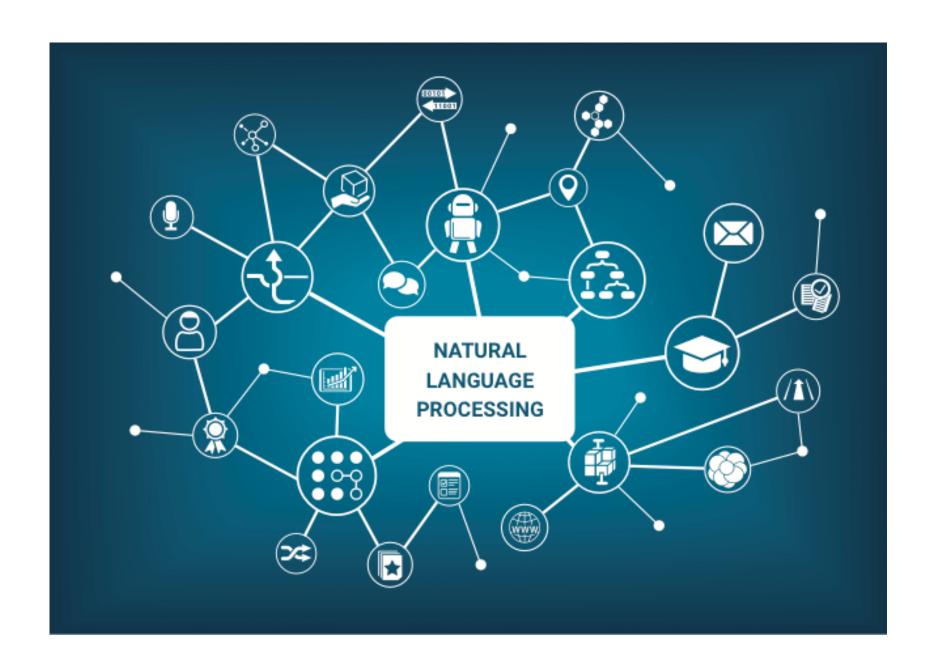
OUTLINE

- ที่มาและความสำคัญ
- แผนการดำเนินงาน
- รายละเอียดการดำเนินงาน
- ความก้าวหน้าการดำเนินงาน
- สรุป

ที่มาและความสำคัญ







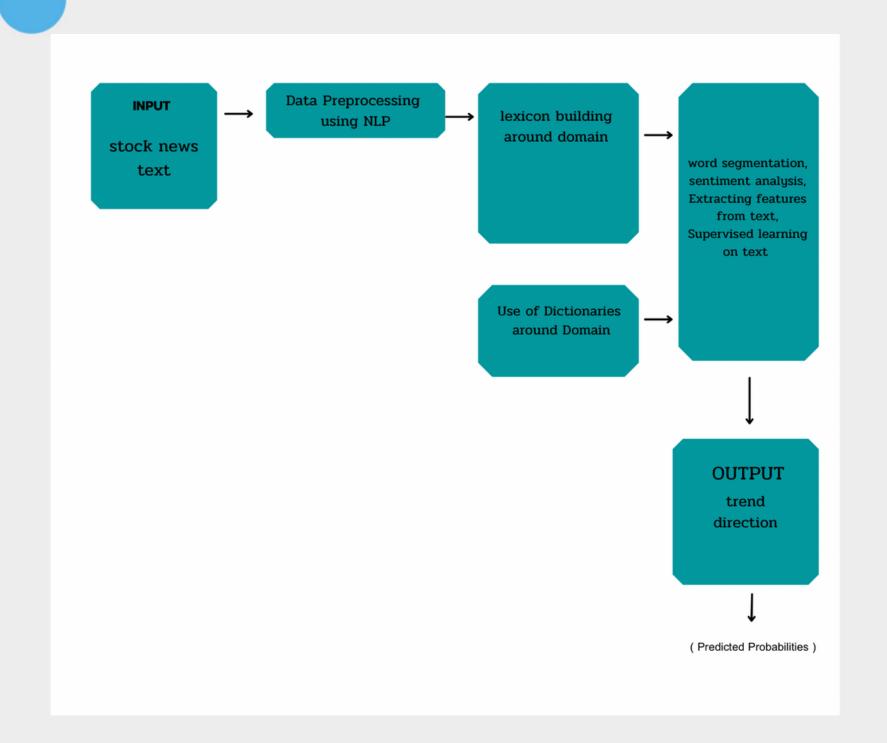


แผนการดำเนินงาน

เวลา	PROJECT 1					PROJECT 2			
	ີນ.ຍ. 65	ก.ค. 65	ส.ค. 65	ก.ข. 65	ต.ค. 65	ช.ค. 65	ม.ค. 66	ก.พ. 66	มี.ค. 66
กิจกรรม									
(1) พัฒนาอัลกอริทึม									
(2) พัฒนาเขียนโปรแกรมด้วย Python									
(3) เทรนสมองกล									
(4) ทคสอบระบบ									
(5) ทคลองนำไปใช้จริง กับบัญชี Forex demo									
(6) ปรับปรุงแก้ไขอุปกรณ์ให้สมบูรณ์									
(7) จัดทำ <u>เว็บ</u> ไซ <u>ต</u> ์									
(8) จัดทำรายงานให้สมบูรณ์									



รายละเอียดการดำเนินงาน



ขั้นตอน INPUT ข่าวสารtext

ขั้นตอน Process ใช้ algorithms

ขั้นตอน Output คาร์ดการณ์ล่วงหน้า สรุปข้อความ



ข้นตอน Input

	ข่าวสาร	ข่าวประเภท	อ้างอิงค์
1.	Census Bureau	กระทรวงพาณิชย์สหรัฐ	[7]
2.	Us Department of Labo	กระทรวงแรงงาน	[8]
3.	Energy information Administration	ข้อมูล <u>ด้านพลังงาน</u>	[9]
4.	กองทุน SPDR	การซื้อขายทองคำของ กองทุน SPDR	[10]
5.	Federal Reserve	ธนาคารกลางสหรัฐ	[11]
6.	Bloomberg	รายงานข่าวทั่วไปรอบโลก	[12]
7.	Twitter	ข่าวทั่วไป	[13]



ข้นตอน Process

- sentiment analysis
- Text Summarization
- Text extraction
- Text classification
- Machine translation



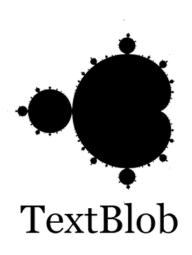
ขั้นตอน Output

- สามารถคาดการ์ณล่วงหน้าได้
- สามารถแปลภาษาได้
- สามารถสรุปบทความได้



ความก้าวหน้า 1 ศึกษา Library Python NLP และทดสอบ







ความก้าวหน้า 1 ศึกษา Library Python NLP และทดสอบ

```
🥏 stopwords.py 🤇 .
                             from nltk.corpus import stopwords
                             from nltk.tokenize import word_tokenize
                             example_sent = "This is a sample sentence, showing off the stop words filtration."
                            stop_words = set(stopwords.words('english'))
                            word_tokens = word_tokenize(example_sent)
                            filtered_sentence = [w for w in word_tokens if not w in stop_words]
                            filtered_sentence = []
                            for w in word tokens:
                                        if w not in stop_words:
                                                     filtered_sentence.append(w)
                            print("stop_words = ",stop_words)
                         print("word_tokens = ",word_tokens)
         19 print("filtered_sentence = ",filtered_sentence)
         PROBLEMS CUTPUT TERMINAL DEBUG CONSOLE
MacBook-Pro-khxng-Chualchai:project whatbest$ cd /Users/whatbest/project ; /usr/bin/env /usr/local/bin/python3 /Users/whatbest/.vscode/exten
sions/ns-python.python-2021.10.1365161279/pythonFiles/lib/python/debugpy/launcher 53097 -- /Users/whatbest/project/stopwords.py
stop_words = {'hadn', 'itself', 'just', "shan't", 'you', 'as', 'an', 'here', 'how', "that'll", 'below', 'very', 'where', 'the', "doesn't", '
theirs', 'won', "wouldn't", 'is', 'any', 'nor', "you've", 'he', 'them', 'yourselves', "aren't", 'their', 'above', 'my', "wasn't", 'while', "y
ou'll", 'too', 'over', 'some', 'isn', 'aren', 'your', 'been', 'were', 'do', 'have', 'no', "hasn't", 'why', 'will', 'if', 'these', 'having', '
few', 'ma', 'or', 'herself', 'yourself', 'we', "should've", 'nightn', 'she', 'ourselves', "won't", 'are', 'did', 'so', "couldn't", 'for', 'wa
s', 'off', 'our', 'only', 'who', 'up', 'it', 'when', 'can', "weren't", 'o', 'had', 'an', 'll', 'that', 'fron', 'they', 'abouldn't, 'ouldn', 'al
', 'weren', 'me', 'now', 'further', 'and', 'haven', 'hers', 'not', 'should', 'most', 'i, 'n', 're', 'wasn', 'once', 'during', 'myself', 'on
', "mustn't", 'her', 'than', 'needn', "shouldn't", 'in', 'again', 'between', "haven't", 'his', 'into', 'then', 'you'd", 'but', 'themselves',
'don', 'does', 'at', 'own', 's', "you're', 'because', 'himself', 'down', 'both', 'other', "needn't", 'isn't", 'has', 'after', "she's", 'of',
'each', "it's", 'this', "mightn't", 'to', 'its', 'through', 'down', 'both', 'other', "needn't", 'isn't", 'has', 'after', "she's", 'of',
'each', "it's", 'this', "mightn't", 'to', 'its', 'through', 'down', 'both', 'other', "needn't", 'isn't", 'has', 'after', "she's", 'of',
'each', "it's", 'this', "mightn't", 'to', 'its', 'through', 'down', 'both', 'other', "haedn't", 'whom', 'nore', 'd', 'ain', 'mustn', 'same',
'which', "didn't", 'out', "hadn't", 'before', "don't", 'against', 'under', 'shouldn', 'wouldn', 'ours', 'didn', 'until', 'there', 'yours', 'yo', 'be', 'be', 'sentence', ',', 'showing', 'off', 'the', 'stop', 'words', 'filtratio
```



ความก้าวหน้า 2 ทดสอบ Library NLTK

```
from nltk.stem import PorterStemmer
       from nltk.tokenize import sent_tokenize, word_tokenize
       example_words = ["python","pythoner","pythoning","pythoned","pythonely,feeling"]
 11 new_text = "It is important to by very pythonly while you are pythoning with python. All pythoners have pythoned poorly at least on
           print(ps.stem(w))
MacBook-Pro-khxng-Chualchai:project whatbest$ cd /Users/whatbest/project ; /usr/bin/env /usr/local/bin/python3 /Users/whatbest/.vscode/extensions/ms-python.python-2021.10.1365161279/pythonFiles/lib/python/debugpy/launcher 53446 -- /Users/whatbest/project/stem.py
all
python
have
python
poorli
at
least
```

퀒 part of speech.py > from nltk.corpus import state_union train text = state union.raw("2005-GWBush.txt") sample_text = state_union.raw("2006-GWBush.txt") 8 custom sent tokenizer = PunktSentenceTokenizer(train text) 10 def process content(): for i in tokenized[:5]: words = nltk.word_tokenize(i) tagged = nltk.pos_tag(words) print(tagged) except Exception as e: print(str(e)) 21 process_content() PROBLEMS OUTPUT TERMINAL DEBUG CONSOLE MacBook-Pro-khxng-Chualchai:project whatbest\$ cd /Users/whatbest/project; /usr/bin/env /usr/local/bin/python3 /Users/whatbest/.vscode/exten sions/ms-python.python-2021.10.1365161279/pythonfiles/lib/python/debugpy/launcher 57279 — "/Users/whatbest/project/part of speech.py" ('PRESIDENT', 'NNP'), ('GEGGE', 'NNP'), ('NNP'), ('GEGGE', 'NNP'), ('NNP'), ('SEGGGE', 'NNP'), ('NDRYSS', 'NNP'), ('BEGGGE', 'NNP'), ('NNP'), ('SEGGGE', 'NNP'), ('NNP'), ('SEGGGE', 'NNP'), ('NDRYSS', 'NNP'), ('NDRYSS', 'NNP'), ('NDRYSS', 'NNP'), ('NNP'), ('SEGGGE', 'NNP'), ('THE', 'NNP'), ('THE', 'NNP'), ('THE', 'NNP'), ('THE', 'NNP'), ('THE', 'NNP'), ('THE', 'NNP'), ('SEGGGE', 'NNP'), ('NNP'), ('NNP'), ('SEGGGE', 'NNP'), ('NNP'), ('SEGGGE', 'NNP'), ('NNP'), ('SEGGGE', 'NNP'), ('SEGGGE', 'NNP'), ('NNP'), ('SEGGGE', 'NNP'), ('NNP'), ('SEGGGE', 'NNP'), ('SEGGGGE', 'NNP'), ('SEGGGGE', 'NNP'), ('SEGGGGE', 'NNP

Part of Speech Tagging with NLTK



Stemming words with NLTK

ความก้าวหน้า 2 ทดสอบ Library NLTK

```
🔁 Lemmatizing.py > ...
      from nltk.stem import WordNetLemmatizer
      lemnatizer = WordNetLemmatizer()
      print(lemmatizer.lemmatize("cats"))
     print(lenmatizer.lenmatize("cacti"))
 7 print(lenmatizer.lemmatize("geese"))
 8 print(lenmatizer.lemmatize("rocks"))
 9 print(lenmatizer.lemmatize("python"))
10 print(lenmatizer.lenmatize("better", pos="a"))
11 print(lemmatizer.lemmatize("best", pos="a"))
12 print(lenmatizer.lenmatize("run"))
13 print(lemmatizer.lemmatize("run",'v'))
PROBLEMS OUTPUT TERMINAL DEBUG CONSOLE
MacBook-Pro-khxng-Chualchai:project whatbest$ cd /Users/whatbest/project ; /usr/bin/env /usr/local/bin/python3 /Users/whatbest/.vscode/exten
sions/ms-python.python-2021.10.1365161279/pythonFiles/lib/python/debugpy/launcher 57403 — /Users/whatbest/project/Lemmatizing.py
cactus
goose
python
good
best
run
MacBook-Pro-khxng-Chualchai:project whatbest$
```

Lemmatizing with NLTK



ความก้าวหน้า 3 ทดสอบ Library NLTK & TextBlob

```
🦰 Wordnet with NLTK.py > ..
                    from nltk.corpus import wordnet
                   syns = wordnet.synsets("program")
                    print(syns[0].name())
      4 print(syns[0].lemmas()[0].name())
     5 print(syns[0].definition())
     6 print(syns[0].examples())
                  synonyms = []
                  antonyms = []
    10 for syn in wordnet.synsets("good"):
                                for l in syn.lemmas():
                                               synonyms.append(l.name())
                                              if l.antonyms():
                                                              antonyms.append(l.antonyms()[0].name())
    16 print(set(synonyms))
    17 print(set(antonyms))
  PROBLEMS OUTPUT TERMINAL DEBUG CONSOLE
 MacBook-Pro-khxng-Chualchai:project whatbest$ cd /Users/whatbest/project; /usr/bin/env /usr/local/bin/python3 /Users/whatbest/.vscode/extensions/ms-python.python-2021.10.1365161279/pythonFiles/lib/python/debugpy/launcher 57441 -- "/Users/whatbest/project/Wordnet with NLTK.py"
  a series of steps to be carried out or goals to be accomplished
['they drew up a six-step plan', 'they discussed plans for a new bond issue']
{'near', 'effective', 'practiced', 'proficient', 'unspoiled', 'respectable', 'skilful', 'secure', 'unspoilt', 'estimable', 'beneficial', 'goo dness', 'upright', 'right', 'in_effect', 'expert', 'honorable', 'dependable', 'ripe', 'serious', 'soundly', 'just', 'undecomposed', 'safe', 'full', 'dear', 'in_force', 'thoroughly', 'adept', 'sound', 'salutary', 'skillful', 'trade_good', 'commodity', 'good', 'well', 'honest'}
{'evil', 'bad', 'evilness', 'badness', 'ill'}
**MacReak Pro khyra Chyralebair related to the state of the 
  MacBook-Pro-khxng-Chualchai:project whatbest$
```



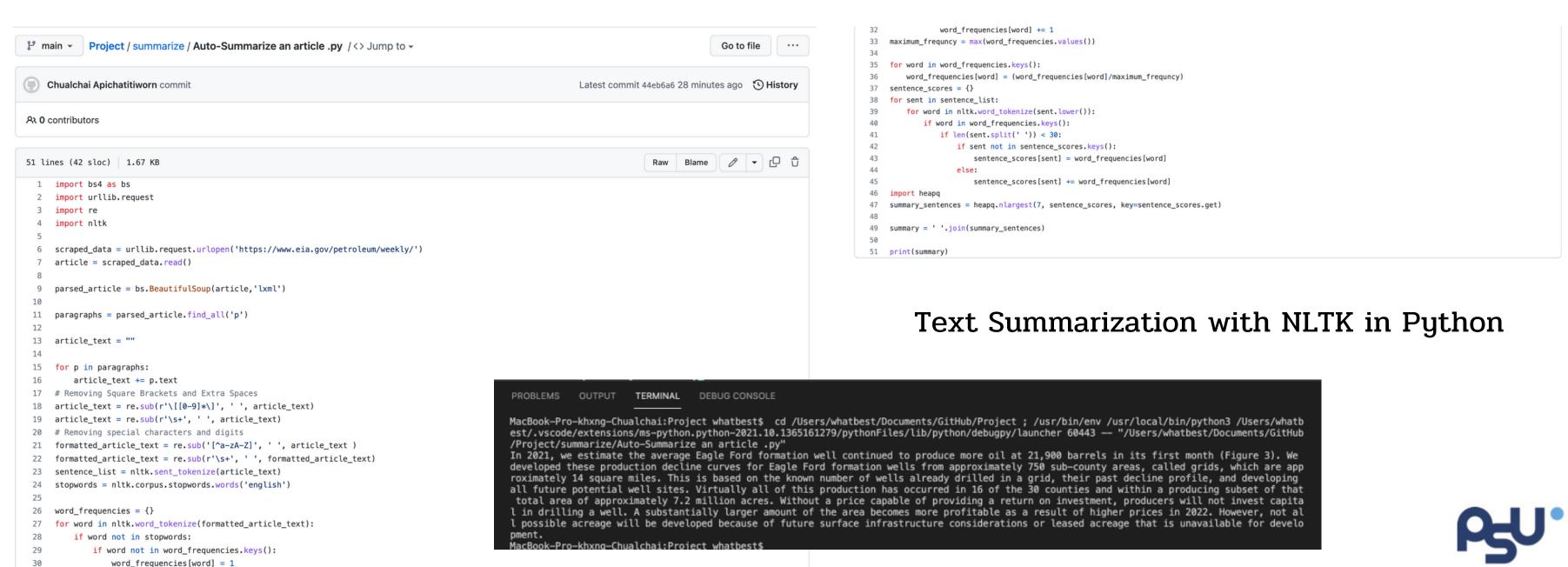
ความก้าวหน้า 3 ทดสอบ Library NLTK & TextBlob

```
textbox > 🔁 sentimentddd.py > ...
      from textblob import TextBlob
      # Preparing an input sentence
      sentence = '''The platform provides universal access to the world's best education, partnering with top universities and organization
      analysisPol = TextBlob(sentence).polarity
      analysisSub = TextBlob(sentence).subjectivity
      print(analysisPol)
      print(analysisSub)
PROBLEMS
            OUTPUT TERMINAL DEBUG CONSOLE
MacBook-Pro-khxng-Chualchai:Project whatbest$ cd /Users/whatbest/Documents/GitHub/Project ; /usr/bin/env /usr/local/bin/python3 /Users/whatbest
est/.vscode/extensions/ms-python.python-2021.10.1365161279/pythonFiles/lib/python/debugpy/launcher 61369 -- /Users/whatbest/Documents/GitHub/
Project/textbox/sentimentddd.py
0.5
0.2666666666666666
MacBook-Pro-khxng-Chualchai:Project whatbest$
```



ความก้าวหน้า 3 ทดสอบ Library NLTK & TextBlob

31



สรุป

สรุปผลการดำเนินงาน

- 1. สามารถวิเคราะห์ประโยคว่าเป็นเชิงบวกหรือเชิงลบได้โดยการใช้ sentiment analysis เบื้องต้น
- 2. สามารถสรุปบทความโดยนำลิ้งค์มาจากเว็บไซต์ต่างๆที่เราต้องการสรุปบทความโดยใช้ Text Summarization
- 3. ได้รู้หลักการเบื้องต้นการทำ sentiment analysis และ Text Summarization โดยการทดลองใช้ Library NLTK

ปัญหาและอุปสรรค

- 1. การวิเคราะห์หรือการเปลี่ยนคำ มีคำที่ผิดพลาดอยู่บ้างเนื่องจาก NLTK พจน์นานุกรมศัพท์ไม่ได้อัพเดท
- 2. การดึงข้อมูลจากเว็บข่าว เพื่อมาทำ Text Summarization บางเว็บไซต์ไม่อนุญาตให้ดึงข้อมูลมาได้

งานที่จะดำเนินการต่อ

- 1. ปรับปรุงแก้ไขการทำ Sentiment analysis เพื่อให้ทำนายหรือคาดการ์ณล่วงหน้าของกราฟ Forex
- 2. เทรนสมองกล เรื่องคำศัพท์เฉพาะทาง Forex



เอกสารอ้างอิง

- [1] S. Shalev-Shwartz, S. Ben-David, Understanding Machine Learning: From Theory to Algorithms (2014), Cambridge University Press , เข้าถึงล่าสุด 15 มกราคม 2565
- [2] T. Mikolov, I. Sutskever, K. Chen, G. S Corrado, and J. Dean. Distributed Representations of Words and Phrases and their Compositionality (2013), Advances in Neural Information Processing Systems 26 เข้าถึงล่าสุด 15 มกราคม 2565
- [3] J. Pennington, R. Socher, and C. D. Manning, GloVe: Global Vectors for Word Representation (2014), In EMNLP. เข้าถึงล่าสุด 16 มกราคม 2565
- [4] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov. Enriching word vectors with subword information (2016), arXiv preprint เข้าถึงล่าสุด 17 มกราคม 2565
- [5] NLP Implementations: URL:
- https://towardsdatascience.com/gentle-start-to-natural-language-processing-usingpython-6e46c07addf3 เข้าถึงล่าสุด 18มกราคม 2565

เอกสารอ้างอิง

- [6] The theory you need to know before you start an NLP : URL : https://towardsdatascience.com/the-theory-you-need-to-know-before-you-start-an-nlp-project-1890f5bbb793 เข้าถึงล่าสุด 12 มีนาคม 2565
 - [7] Us Department of labor : URL : https://www.dol.gov/ เข้าถึงล่าสุด 12 มีนาคม 2565
 - [8] Energy information Administration : URL : https://www.eia.gov/ เข้าถึงล่าสุด 12 มีนาคม 2565
 - [9] กองทุน SPDR : URL : https://traderider.com/forex/spdr-%E0%B8%81%E0%B8%AD%E0%B8%87%E0%B8%97%E0%B8%B8%E0%B8%99%E0%B8%97%E0%B8 %AD%E0%B8%87%E0%B8%84%E0%B8%B3%E0%B9%81%E0%B8%97%E0%B9%88%E0%B8%87 เข้า ถึงล่าสุด 12 มีนาคม 2565
 - [10] Federal Reserve : URL : https://www.federalreserve.gov/ เข้าถึงล่าสุด 12 มีนาคม 2565
 - [11] Bloomberg : URL : https://www.bloomberg.com/asia เข้าถึงล่าสุด 12 มีนาคม 2565
 - [12] Twitter : URL : https://twitter.com/ เข้าถึงล่าสุด 12 มีนาคม 2562

คำถามและข้อเสนอแนะ

