

# Sentiment Analysis Using Word Cloud

หัวข้อ : การเมืองไทย ในไทยและต่างประเทศ



ภาพ [1]

ผู้จัดทำ

61605041 ลดा พรวีวงศ์  
61605053 อานันท์ มนีพินิจ  
61605113 นนทกร เลิศชาณุวนิ



ภาพ [2]



ภาพ [3]

ภาพ [1] : <https://www.poolsawat.com/mongod-กับ-query-และคำย่อของเครื่อ/>

ภาพ [2] : <https://www.prachachat.net/politics/news-294964>

ภาพ [3] : <https://storylog.co/story/56d72140b5047a630cb8b44d>

# Two Most Popular APIs

<b>Streaming API</b>	<b>REST API</b>
a sample of public tweets and events as they published on Twitter (can specify search terms or users)	<ul style="list-style-type: none"><li>- search</li><li>- trends</li><li>- read author profile and follower data</li><li>- post / modify</li></ul>
<b>only</b> real-time data	historical data up to a week
continuous net connection	one-time request
no limit	rate limit (varies for different requests)

# Search

A screenshot of a Twitter search results page for the hashtag #nlproc. The search bar at the top contains the text "#nlproc". Below the search bar, there is a blue header bar with the hashtag "#nlproc" in white. The main content area shows search results for "Top" tweets. The first result is a tweet from Todd Smith (@Cisco\_Mobile) posted 4 hours ago, which reads: "Google Taps Neural Network Tech to Bolster Anti-Spam Efforts goo.gl/RmtZ4b #nlproc". To the left of the main content, there is a sidebar titled "Who to follow" with a "Follow" button for "It Can Wait". A red arrow points to the search bar at the top right.

Home Notifications Messages

#nlproc

#nlproc

Top | Live | Accounts | Photos | Videos | More options ▾

Who to follow · Refresh · View all

It Can Wait @ItCanWait Follow Promoted

Todd Smith @Cisco\_Mobile · 4h  
Google Taps Neural Network Tech to Bolster Anti-Spam Efforts goo.gl/RmtZ4b  
#nlproc

View summary

## body tweet

```
    print(status.text)  
RT @andi_staub: [insightful]  
  
#BigData sources and methods for social and economic analyses  
  
#fintech #Insurtech #DeepLearning #AI #Predic...
```

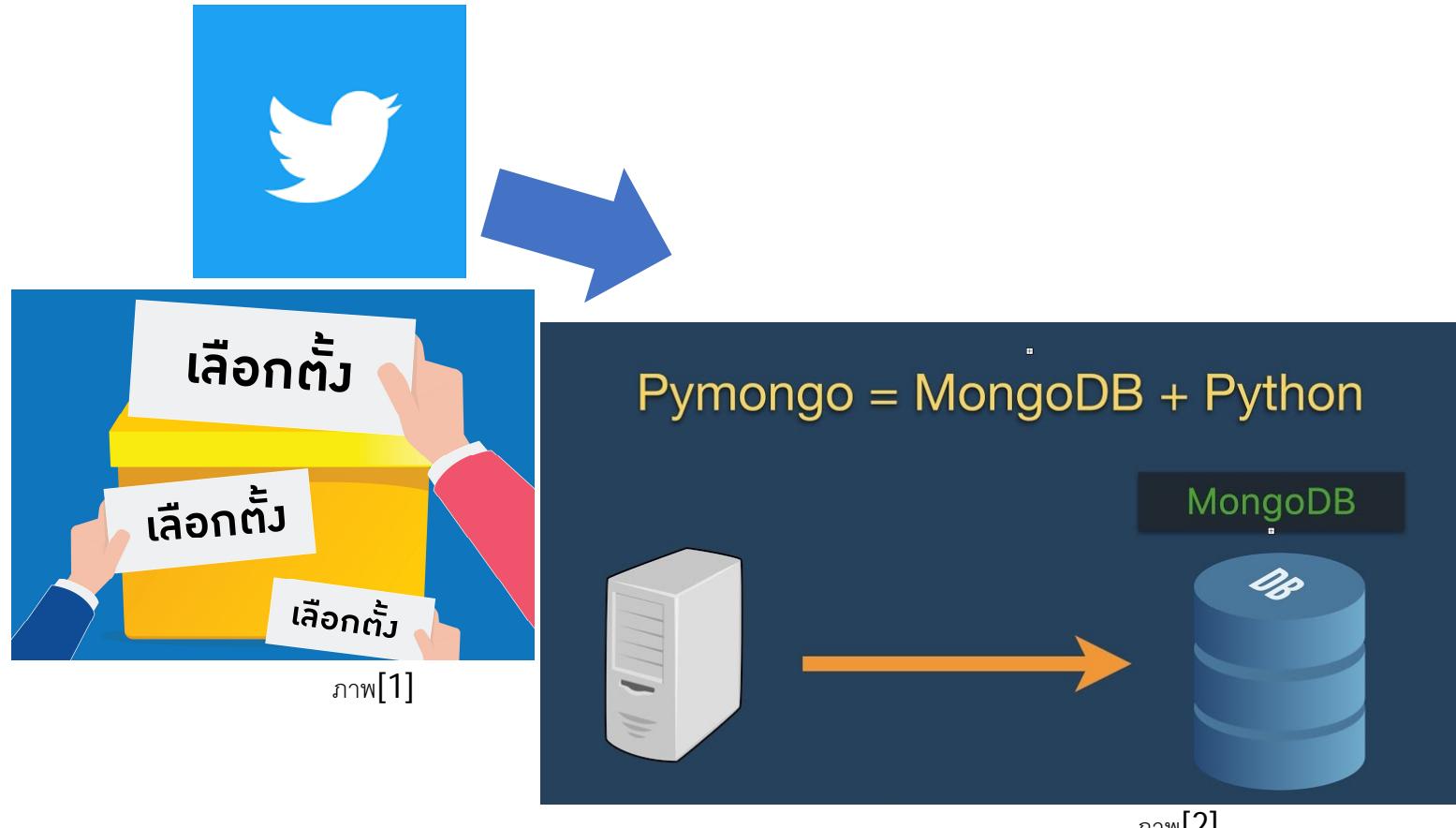
print(status.\_json)

JSON

```
{'created_at': 'Mon Feb 25 08:08:49 +0000 2019', 'id': 109994428872766251}
```

```
pprint.pprint(mycol.find_one({}, { "_id": 0, "text": 1}))
```

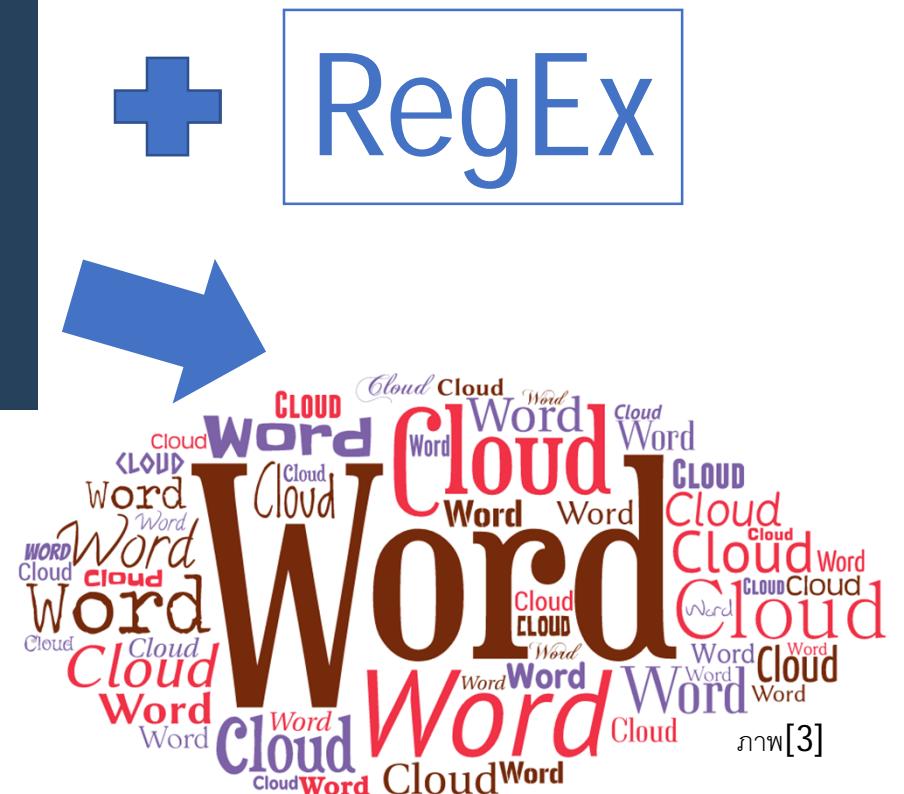
```
{'text': 'RT @TheEIU: Thailand has introduced new legislat  
cyber-security, data protection and digital iden  
nature and...'}  
{'text': 'RT @TheEIU: Thailand has introduced new legislat  
cyber-security, data protection and digital iden  
nature and...'}
```



ภาพ [1] : <https://www.prachachat.net/politics/news-294964>

ภาพ [2] : [https://www.youtube.com/watch?v=P1Qbf3A\\_E6U](https://www.youtube.com/watch?v=P1Qbf3A_E6U)

ภาพ [3] : <http://aptgadget.com/word-cloud-makers/>



keyword = "#เลือกตั้ง #การเมือง #ไทยรัฐ#เลือกตั้ง62

```
for status in tweepy.Cursor(api.search, q=keyword , count=100, lang='th').items(i):
    with open('thai_election_th.txt', 'a',encoding='utf-8') as outfile:
        json.dump(status._json, outfile,ensure_ascii=False)
    mycol.insert_one(status._json)
    print(mycol.count_documents({}))
```



สุดารัตน์ เกยุราพันธุ์



ประยุทธ์ จันทร์โอชา



ชนารัต จึงรุ่งเรืองกิจ



เสรีพิศุทธิ์ เทเมียวเวส



อภิสิทธิ์ เวชชาชีวะ



# regular expression and word cloud

"อนาคตใหม่\*"  
"เพื่อไทย"  
"ประชาธิปไตย"  
"พลังประชารัฐ"  
"เสรีรวมไทย"



"ประยุทธ์\*" | ตุ \*

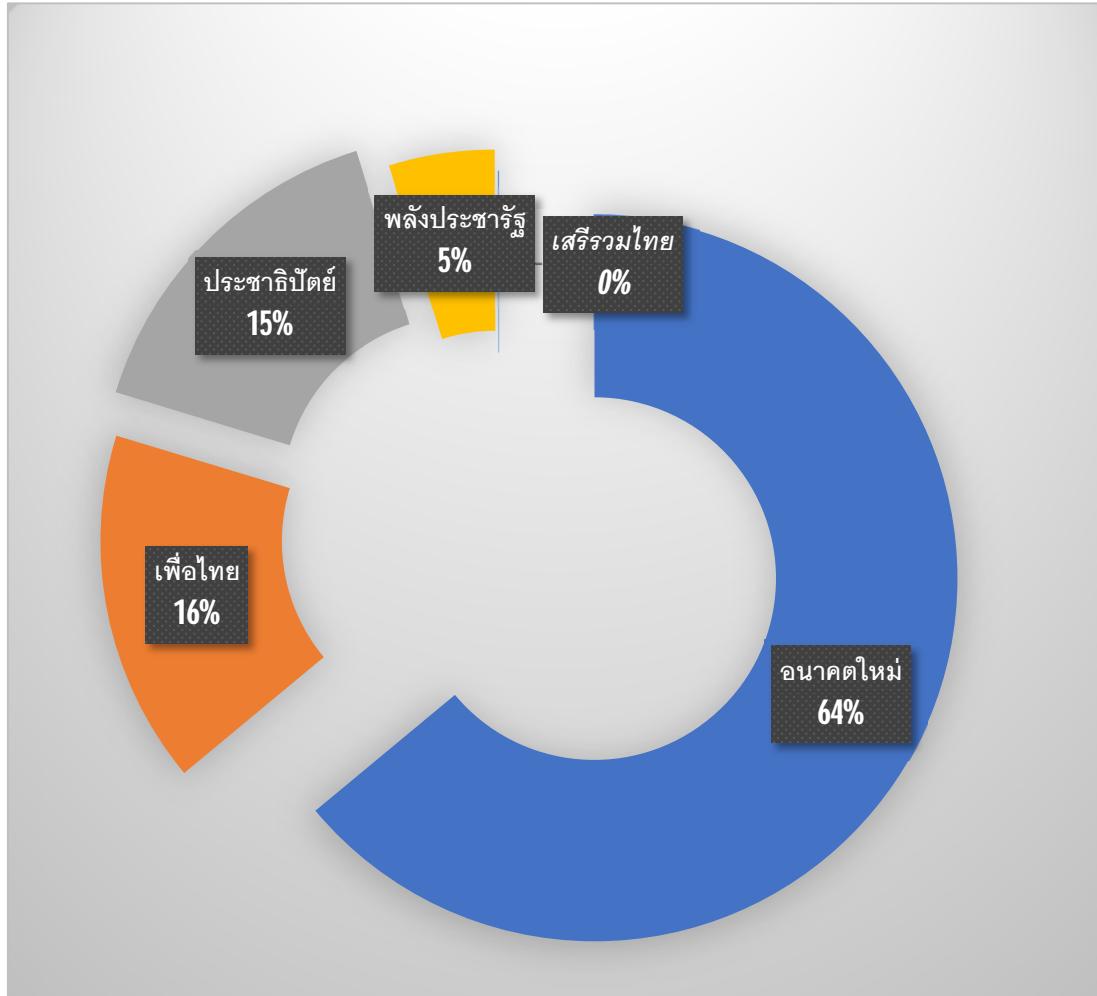
"ธนาธร | พาร์กพ่อ \*"

"เสรีพิศุทธิ์"

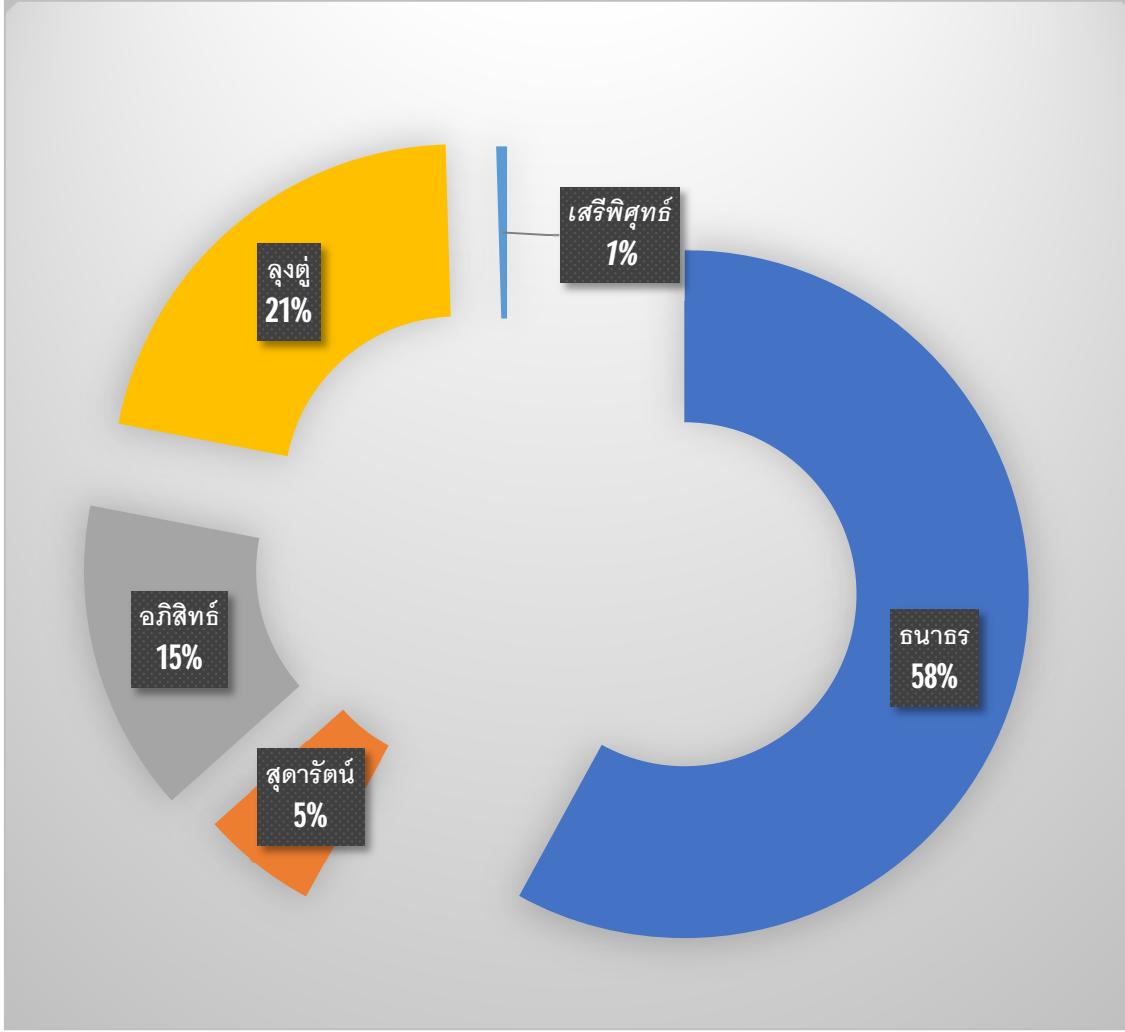
"อภิสิทธิ์ | มาร์ค \*"

"สุดารัตน์ \* | เมย์าย  
แห่งชาติ \*"

## ความนิยมพรรค



## ความนิยมผู้นำพรรค



## จํานวนผู้สนับสนุน





procurement  
การเมือง และการเมืองไทย  
ในมุมมองของสื่อต่างประเทศ  
???

```
query = {}
view = { "text": 1}
y = mycol.find(query,view)
for x in y :
    a = x["text"]
    a = re.sub("'text':", " ", a)
    b = re.sub("http\S*", " ", a)
    c = re.sub("\W", " ", b)
    d = re.sub("[^a-zA-Z]", " ", c)
    e = re.sub("\s+", " ", d)
    e = e.lower()
    e = str(e)
    myquery = {"_id": x["_id"]}
    newvalues = { "$set" : { "text_clean": e } }
    mycol.update_one(myquery, newvalues)
```



```
'text': 'RT @TheEIU: Thailand has introduced new led
entity. However, the nature and...',
'text_clean': 'rt theeiu thailand has introduced new
identity however the nature and '}
```

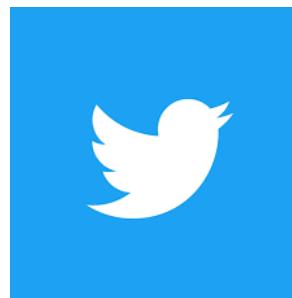
keyword =

'(thai OR thaïs OR thailand ) AND

(election OR elections OR politician OR politicians OR policy OR politics OR political OR government)'



```
for status in tweepy.Cursor(api.search, q=keyword , count=100, lang='en').items(i):
    with open('thai_election_tt.txt', 'a',encoding='utf-8') as outfile:
        json.dump(status._json, outfile,ensure_ascii=False)
    mycol.insert_one(status._json)
    print(mycol.count_documents({}))
```



mongoDB®

หา 4 คำหน้า 4 คำหลัง ของ "party"

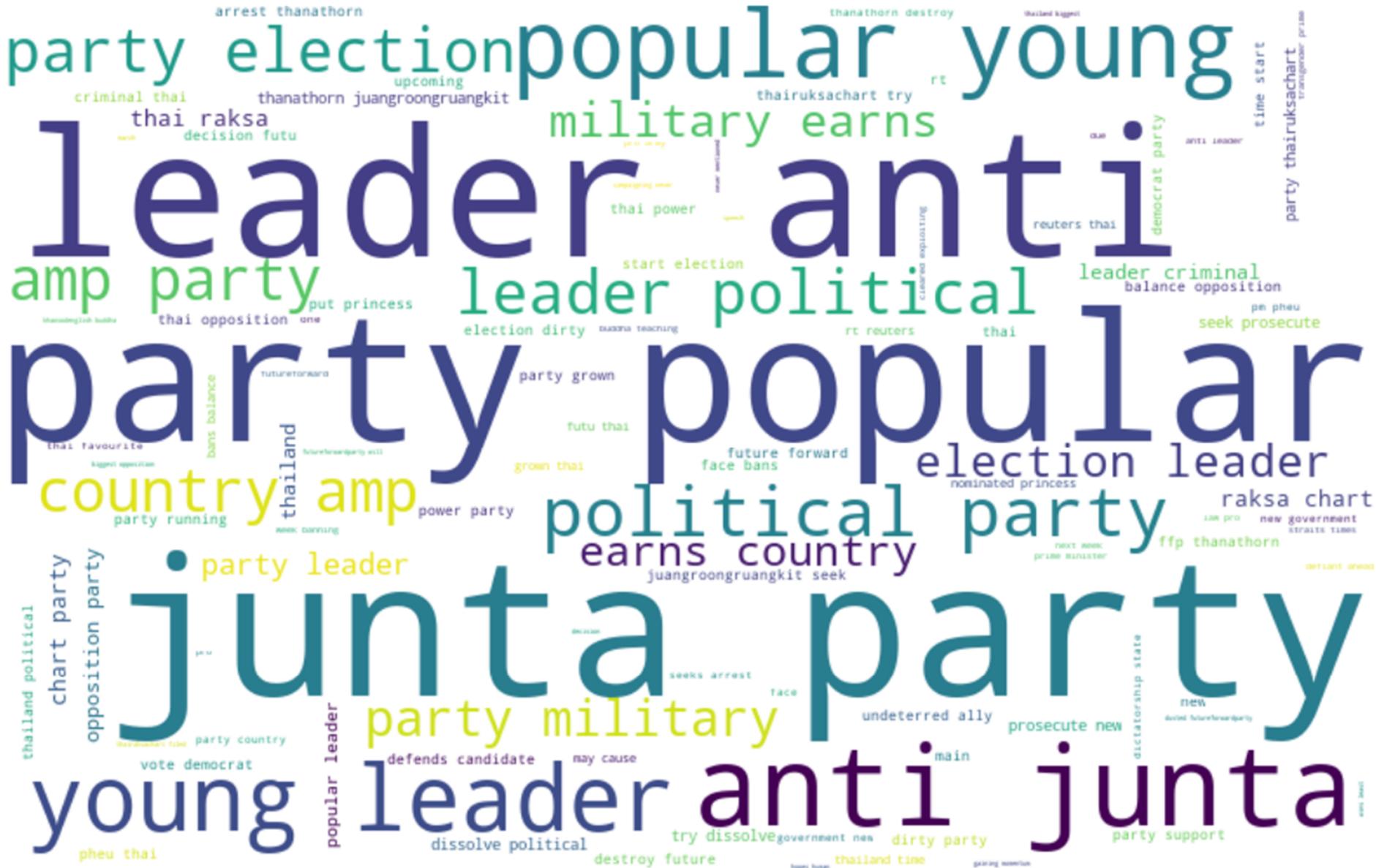


```
query = {"text_clean": {"$regex": "party"} }  
view = {"_id": 0, "text_clean": 1}  
  
y = mycol.find(query, view)  
text = ""  
  
for x in y:  
    x = str(x)  
    x = re.sub("'text_clean':", " ", x)  
    x = re.findall("[a-z]*\s*[a-z]*\s*[a-z]*\s*[a-z]*\s*  
                  \s*party\s*  
                  [a-z]*\s*[a-z]*\s*[a-z]*\s*[a-z]*\s*", x)  
  
    text = text + str(x)
```

```
query = {"text_clean": "party" }  
print(mycol.find_one(query,view))
```

None

"\$regex": "party"



# ไทยชื่อ政党

party list name from

<http://www.khaosodenglish.com/politics/2019/02/27/thai-election-for-dummies-guide-to-the-parties/>

Pheu Thai

Phalang Pracharat

Bhumjaithai

Democrat

Chartpattana

Future Forward

Chartthaipattana

Action Coalition for Thailand

Puea Chat

Peoples Reform

Prachachart

Thai Raksa Chart

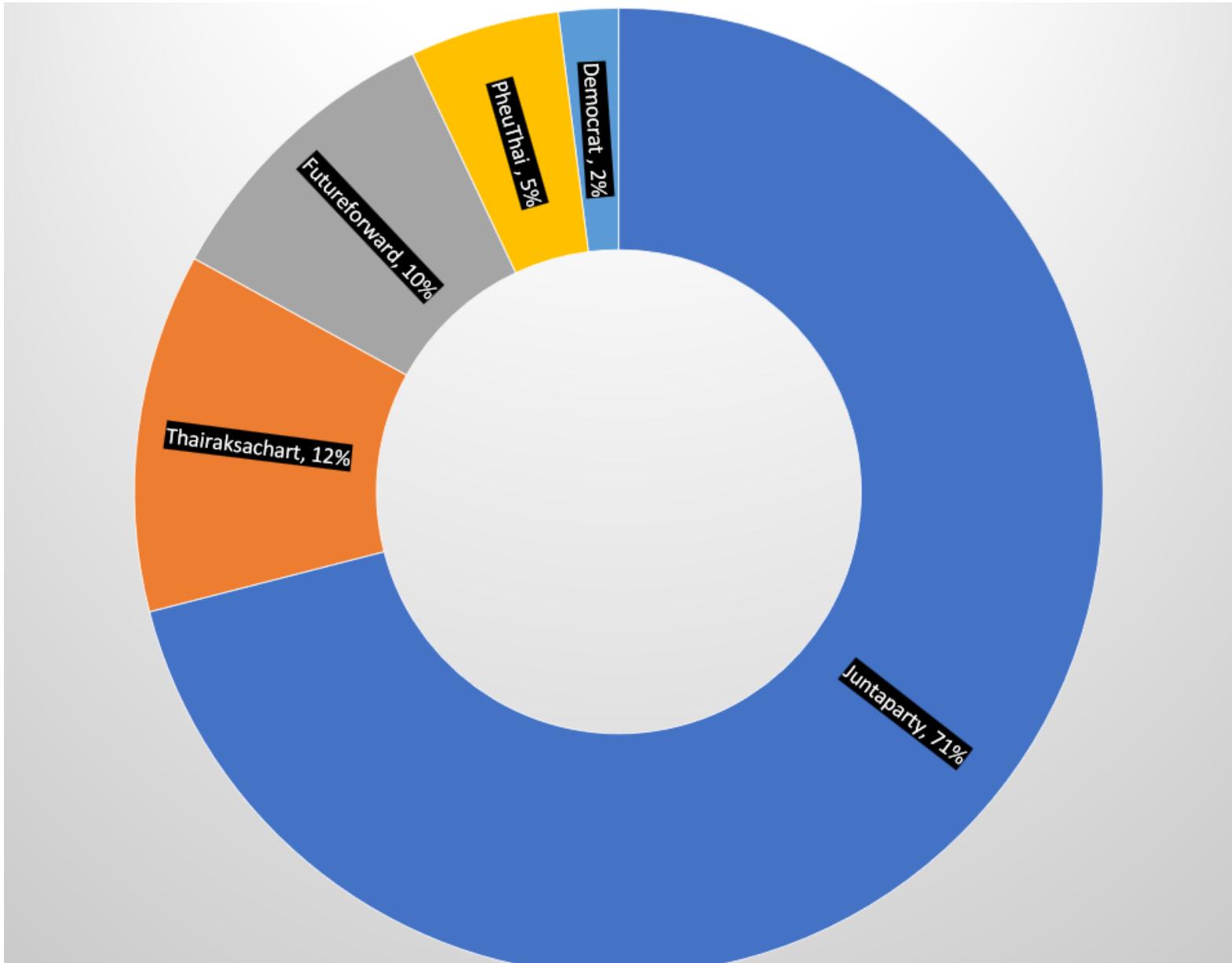
Seri Ruam Thai

Thai Local Power



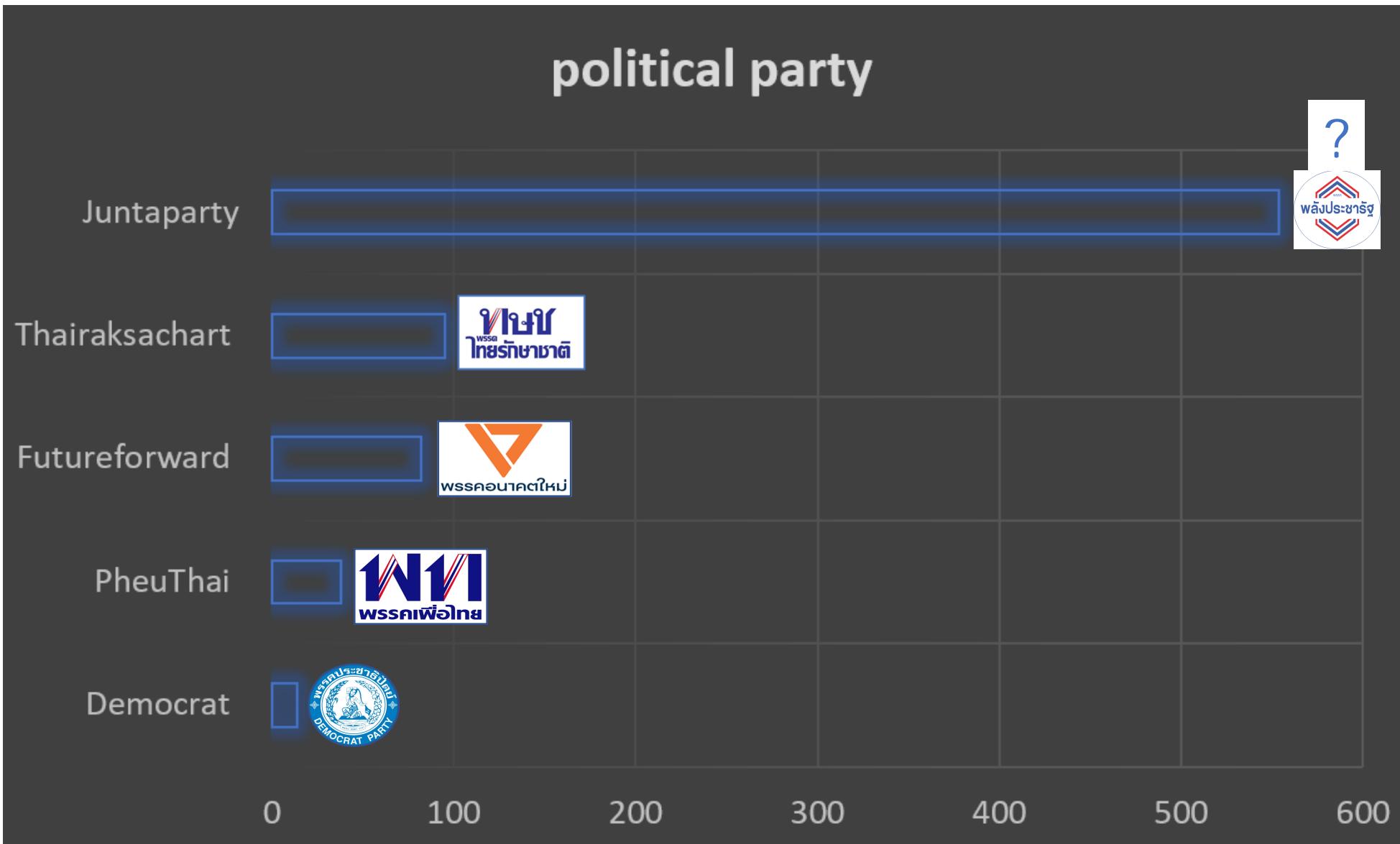
juntaparty

## นับจำนวนและเรียงลำดับ



นับจำนวนและเรียงลำดับ

## political party



## แต่ละพร็อค หาคำที่ปรากฏร่วมใน tweet 10 คำแรก

```
wordfreq = {}

wordslen = []
for raw_word in words:
    if len(raw_word) > 2 : wordslen.append(raw_word)

words    = wordslen

for raw_word in words:
    word = raw_word.strip(unwanted_chars)
    if word not in wordfreq:
        wordfreq[word] = 0
    wordfreq[word] += 1
pprint.pprint(wordfreq)
```

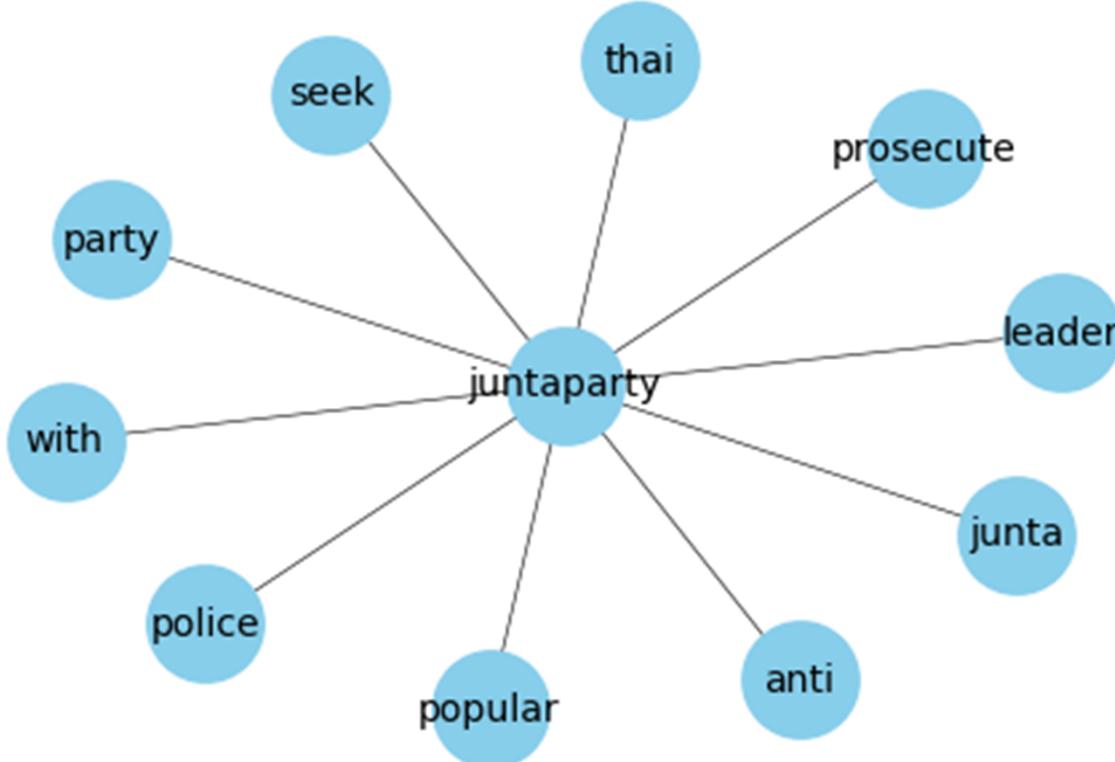
```
mydict = wordfreq
mynode = []
j = 0
for i in sorted(mydict, key=mydict.get, reverse = True):
    j = j+1
    if j <= 10 :
        mynode.append(i)
```



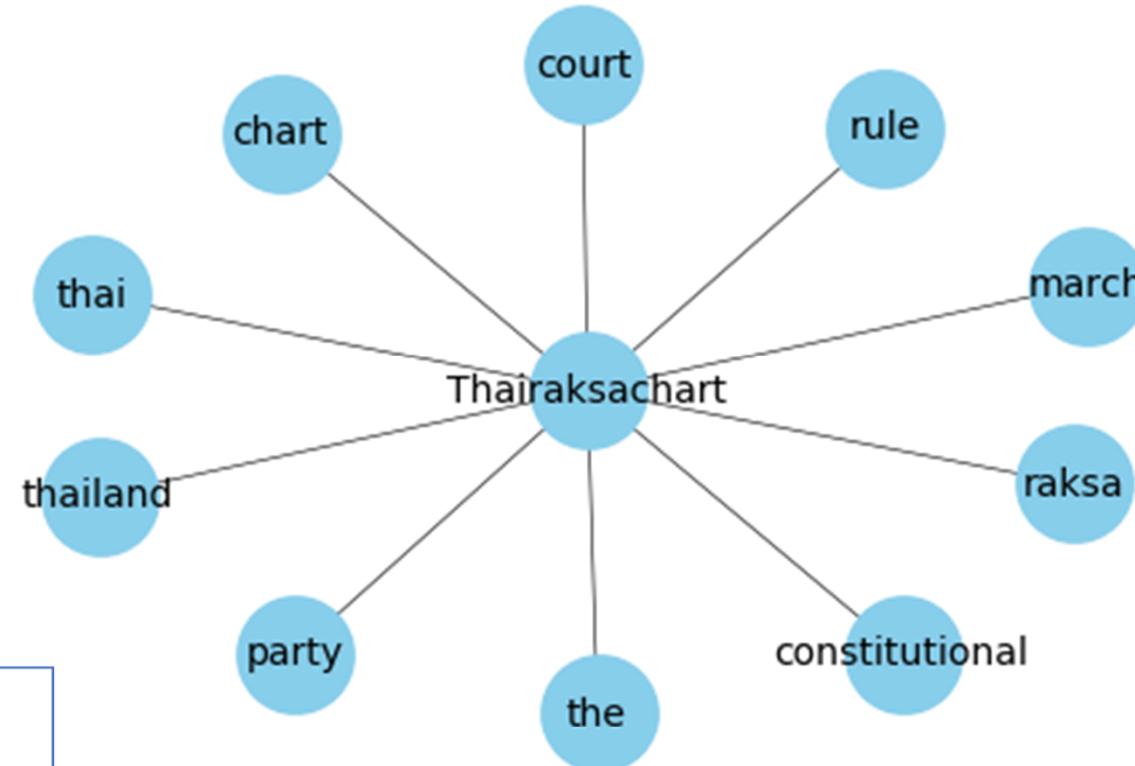
NetworkX  
สร้างกราฟอนด์

# RegEx : (junta party)|(juntaparty)

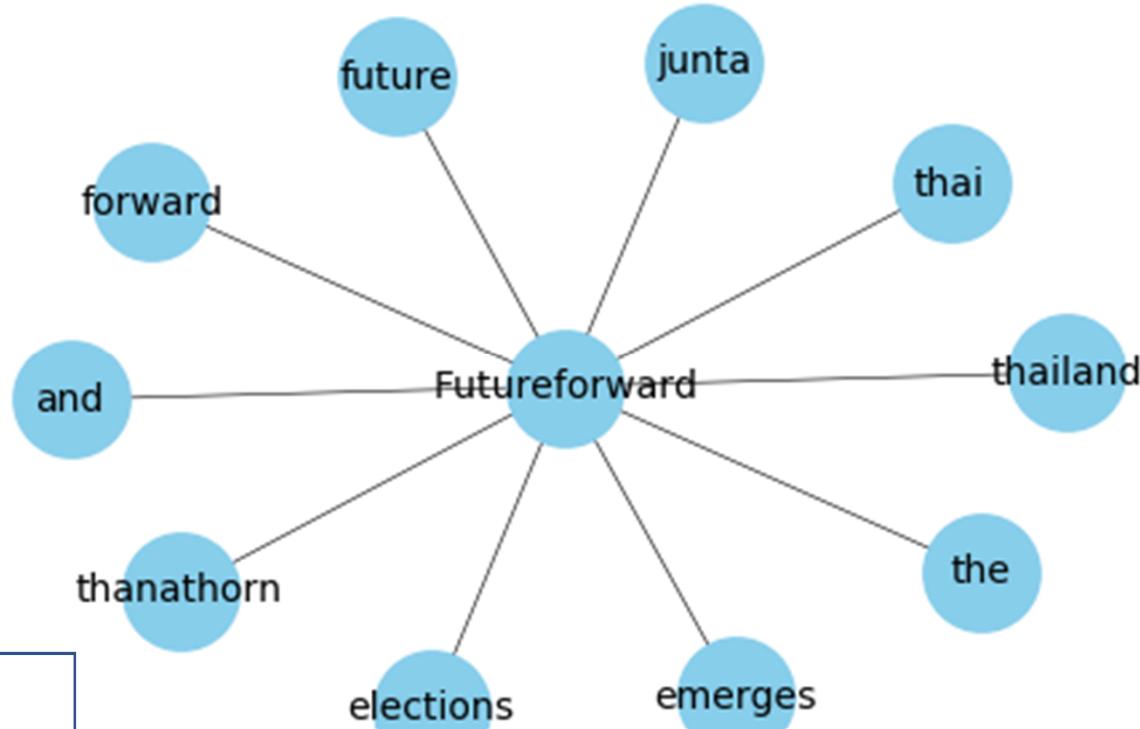
?



RegEx : (thai raksa chart)|(thairaksachart)|(thairaksachartparty)

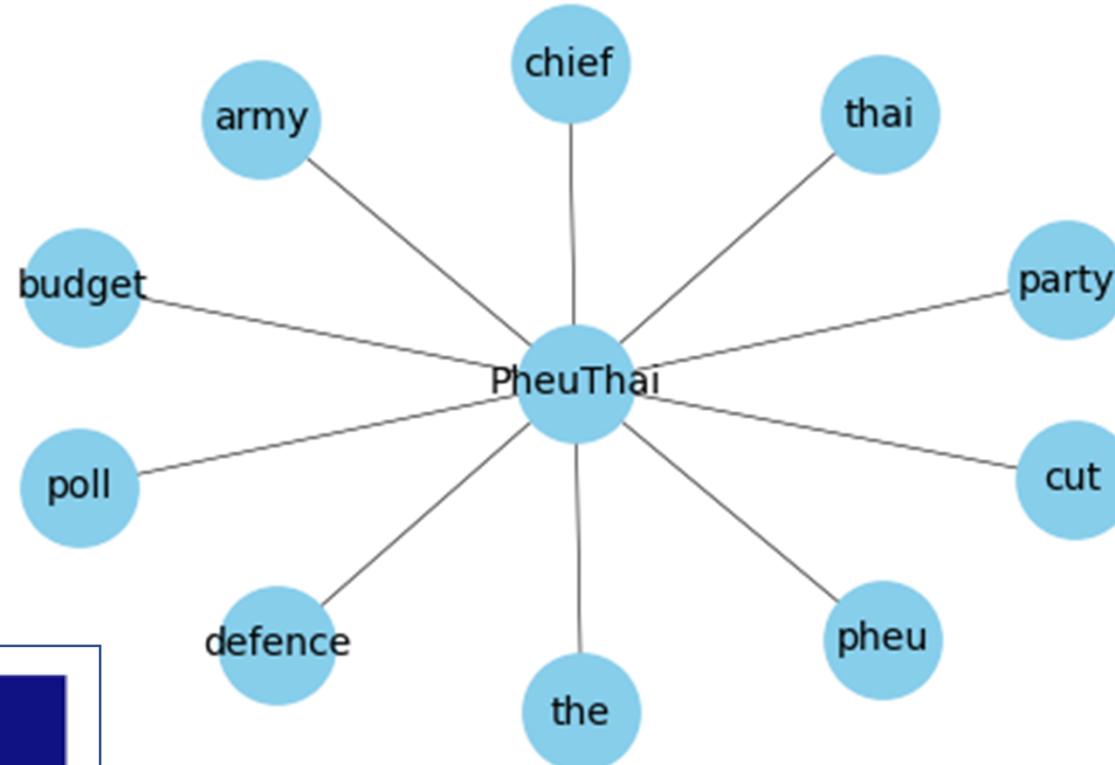


RegEx : (future forward)|(futureforward)|(futureforwardparty)

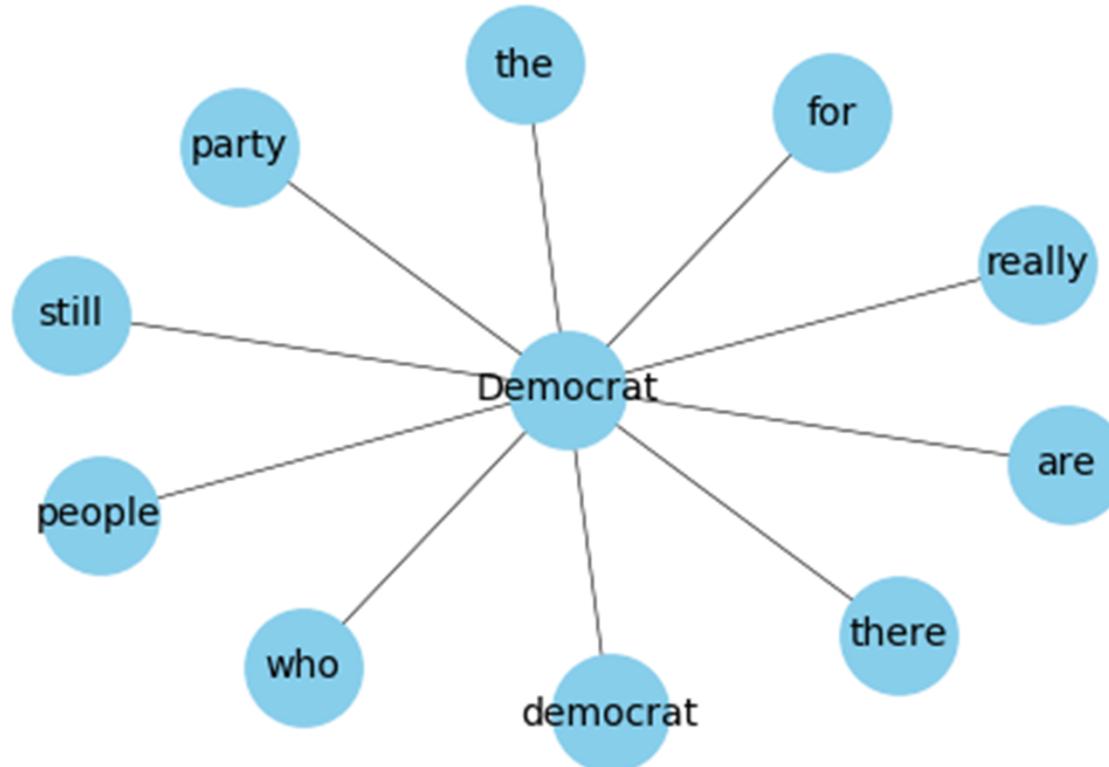


ව්‍යුත්පන ආකෘතිය

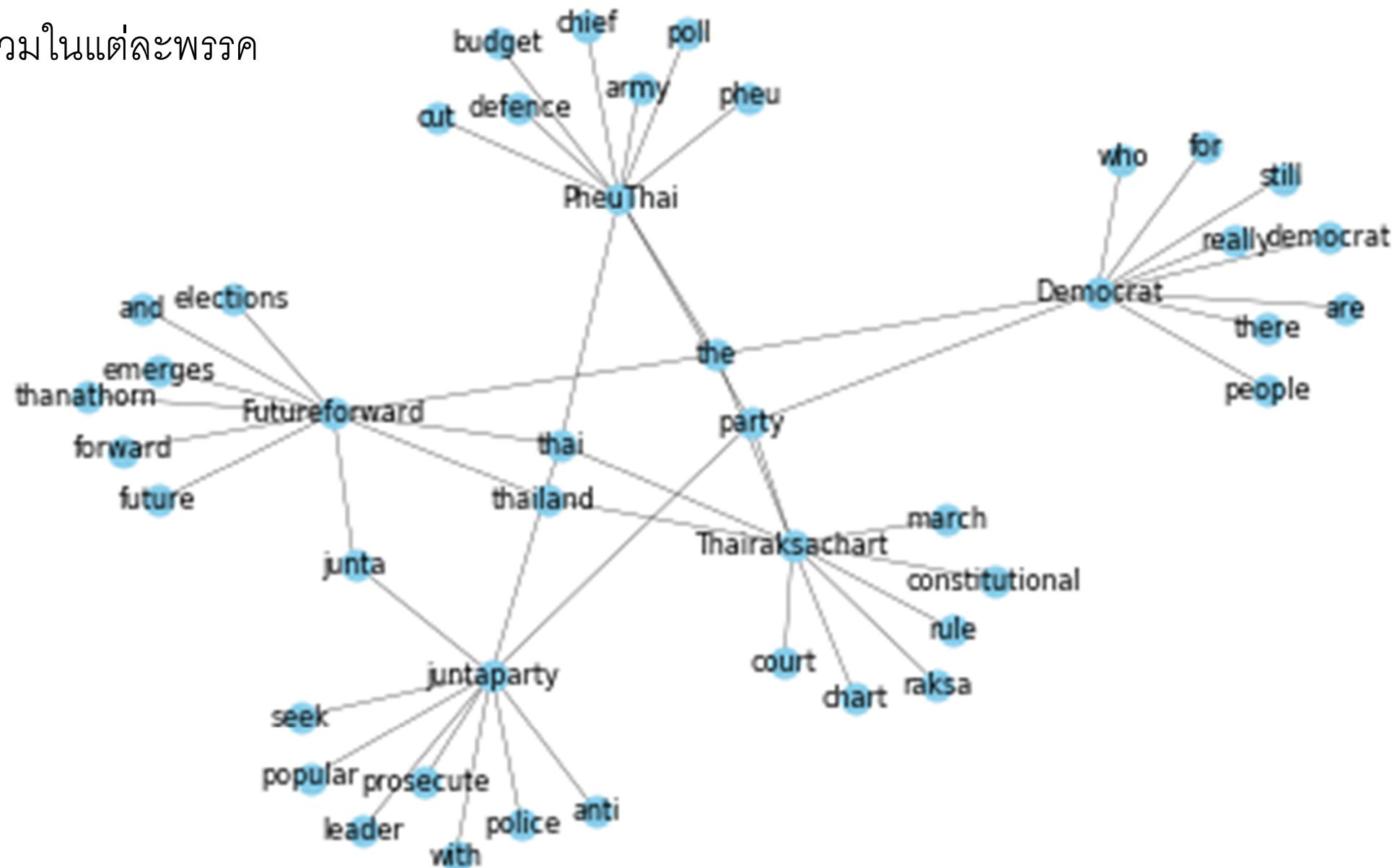
RegEx : (pheu thai)|(pheuthai)|(pheuthaiparty)



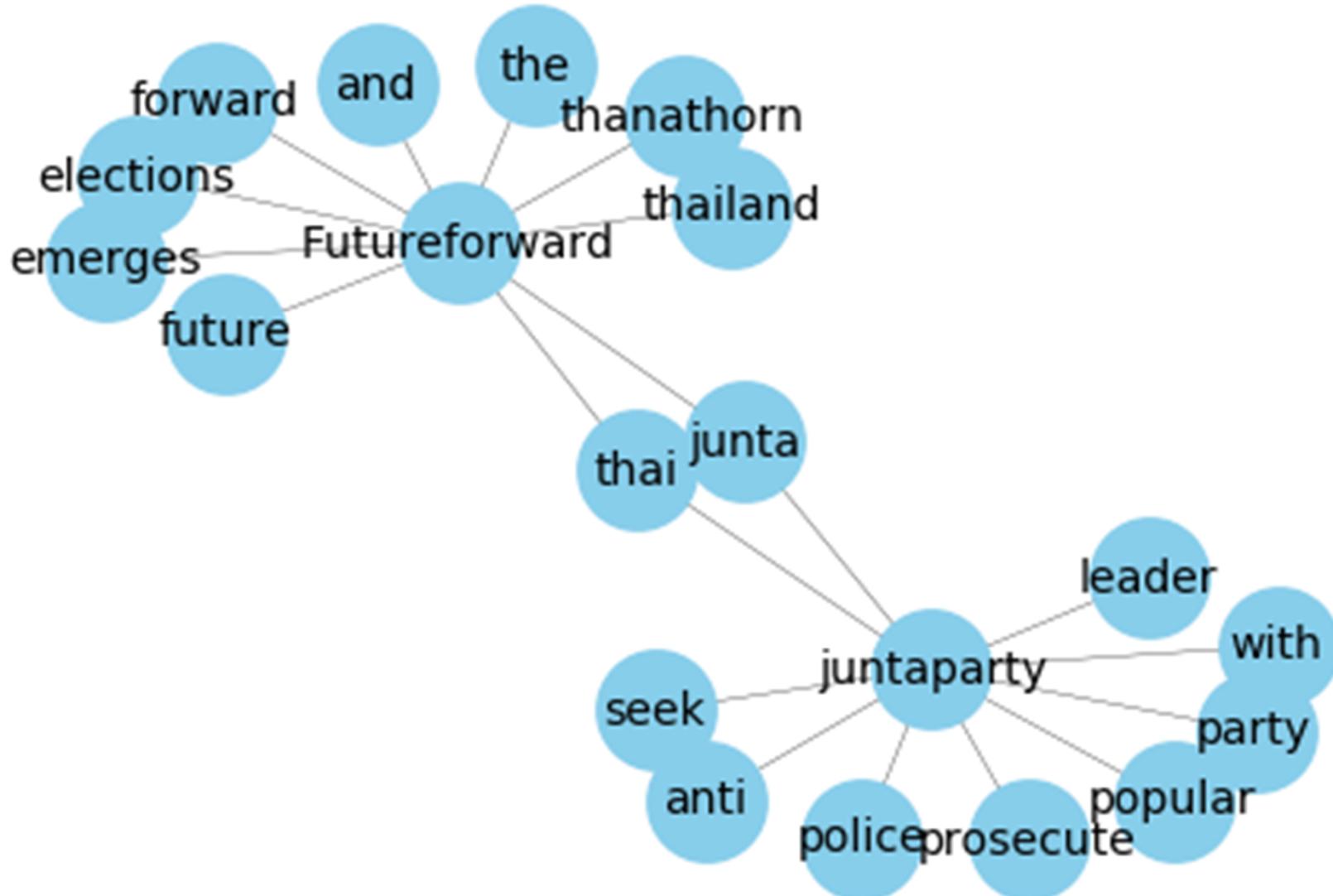
RegEx : (\s+democrat\s+)|(\s+democratparty\s+)

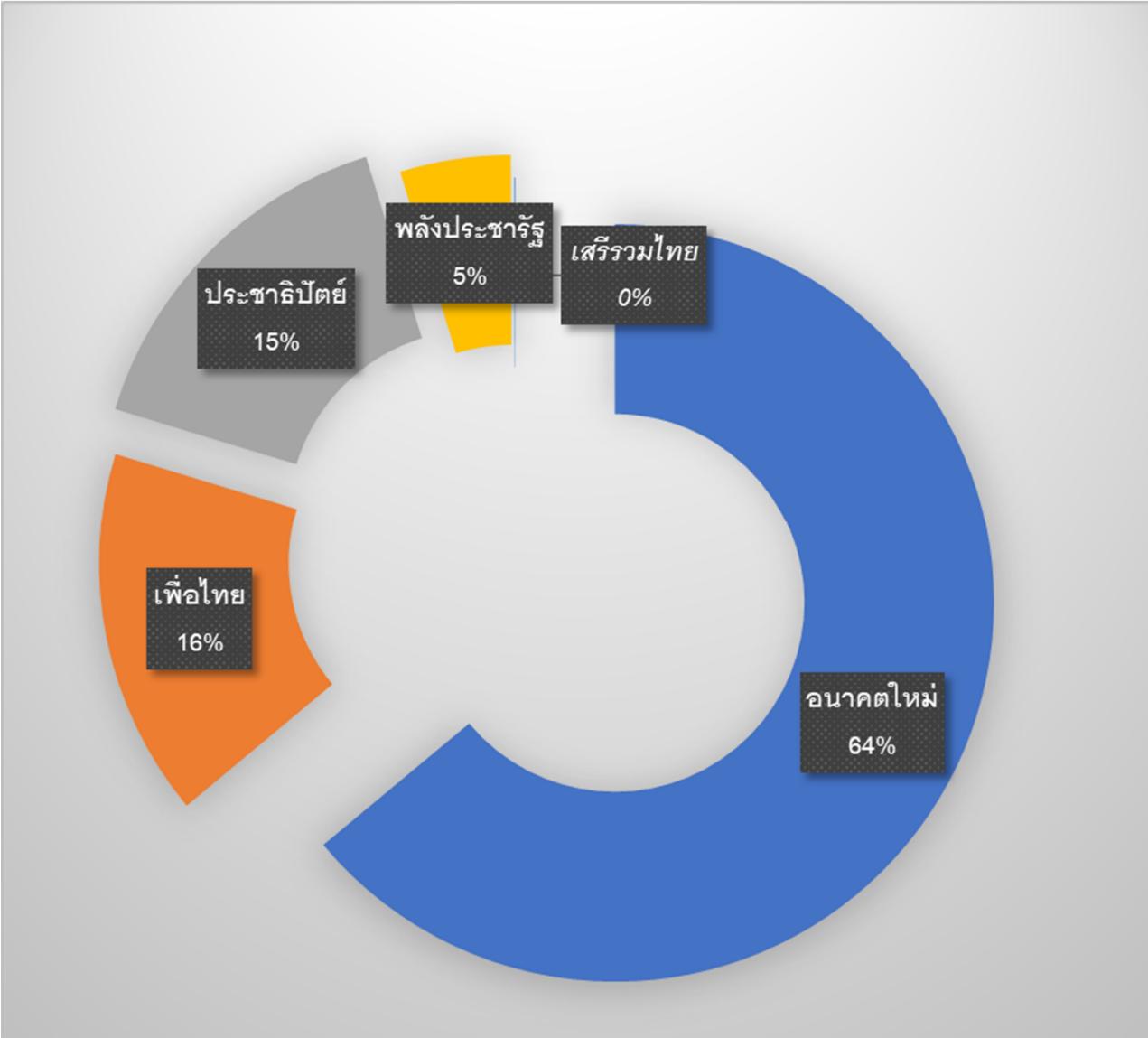


## คำที่ถูกพูดร่วมในแต่ละพรรคร่วม

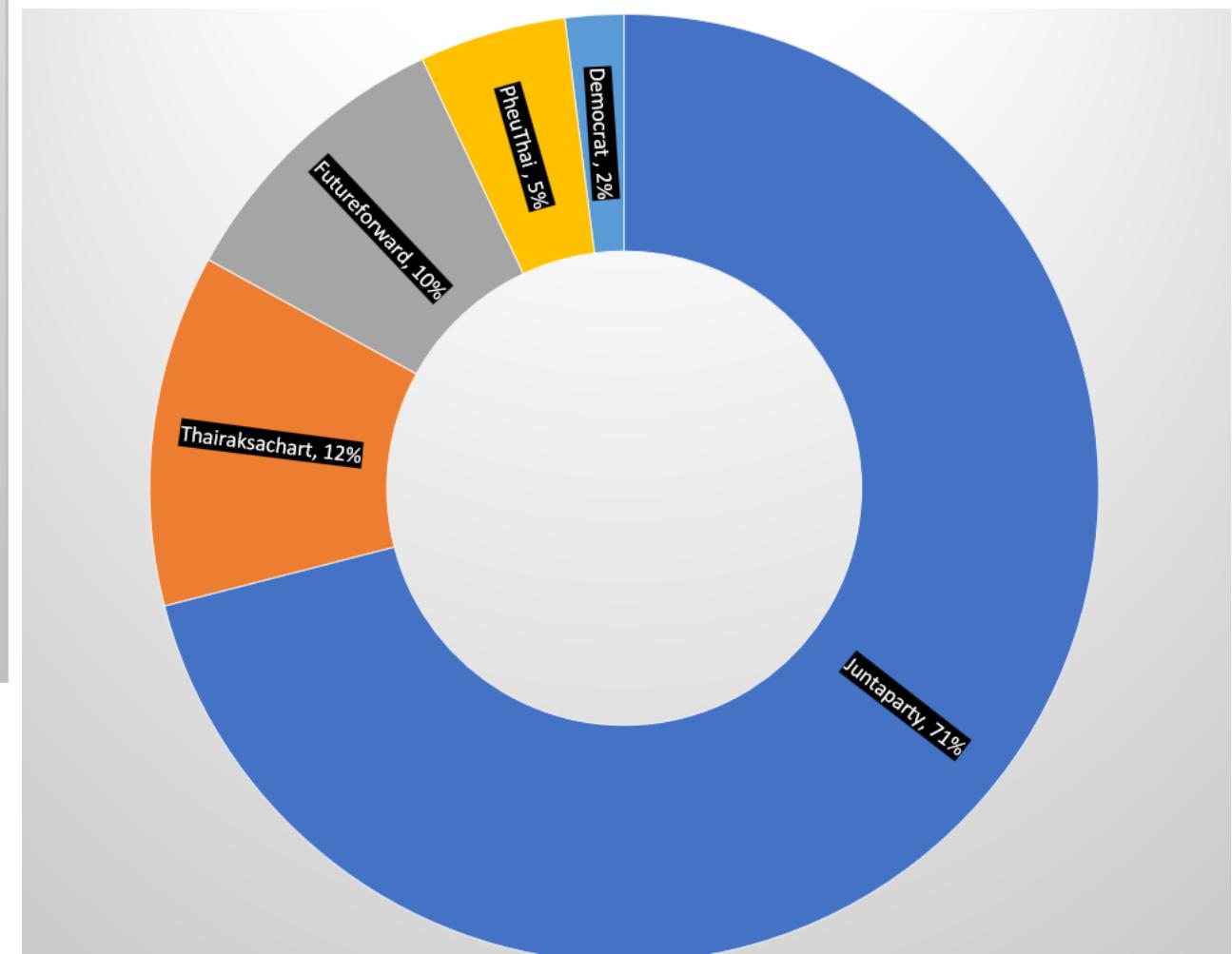


# "junta" ถูกพูดมากที่สุดใน 2 พรรคร่วม





## พิจารณาเมือง ใน twitter ภาษาไทย กับ ต่างประเทศ





# ឧបគណ