

# ‘텍스트 마이닝’을 활용한 ‘금리예측’



팀명: 이거 끝나고 춤 한번 추조  
팀원: 김광희, 황호영, 강다영, 우지원, 최하은

# 목차

---

## 1 프로젝트 목적

## 2 접근방법

2-1) 데이터 수집

2-2) 텍스트 전처리

2-3) 텍스트 분석

## 3 분석결과



# 1. 프로젝트 목적

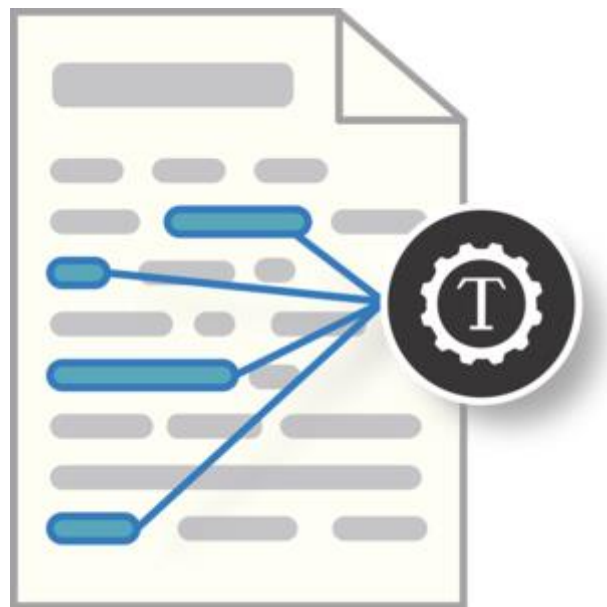
중앙은행의 커뮤니케이션은 '통화정책' 방향 결정

# 1. 프로젝트 목적

중앙은행의 커뮤니케이션은 **절제**된 표현 多

>> 커뮤니케이션에 내재된 **정보 추출 한계**

# 1. 프로젝트 목적



텍스트 마이닝을 통해 기준금리 변동에 대한 설명과 예측  
>> 통화정책 진단의 효율성을 높임

## 2. 접근방법

### 데이터 수집

1. 채권 분석 리포트
2. 네이버 뉴스기사(금리)
3. 의사록
4. 콜금리, 기준금리



### 텍스트전처리

1. eKoNLPy 활용
  - token
  - ngram
2. 콜금리 변동 확인



### 텍스트분석

1. 사전 구축
2. 어조 분석



## 2-1) 데이터 수집

\* 데이터 수집

### 채권분석 리포트



#### 채권분석

2008.04-2020.12 3,873개

- 채권 투자 전략
  - 금리전망
  - 채권시황
-

# 2-1) 데이터 수집

## 채권 분석 리포트 크롤링 코드

```
from bs4 import BeautifulSoup
import pandas as pd
import requests
import wget

def con(urls):
    headers = {"authority": "finance.naver.com",
, "method": "GET"
, "path": "/research/debenture_list.nhn?keyword=&brokerCode=&searchType=writeDate&writeFromDate=2005-01-01&writeToDate=2021-01-04&x=29&y=27"
, "scheme": "https",
"accept": "text/html,application/xhtml+xml,application/xml;q=0.9,image/webp,image/apng,*/*;q=0.8,application/signed-exchange;v=b3;q=0.9",
"accept-encoding": "gzip, deflate, br",
"accept-language": "ko,en;q=0.9,en-US;q=0.8",
"cookie": "NRTK=ag#all_gr#1_ma#-2_si#0_en#0_sp#0; NNB=V5Q30A3RX7JF6; nx_ssl=2; naver_stock_codeList=005930%7C252670%7C; page_uid=U/7UZwp0Jy0ss
"referer": "https://finance.naver.com/research/debenture_list.nhn?keyword=&brokerCode=&searchType=writeDate&writeFromDate=2005-01-01&writeToDa
"sec-fetch-dest": "document",
"sec-fetch-mode": "navigate",
"sec-fetch-site": "same-origin",
"sec-fetch-user": "?1",
"upgrade-insecure-requests": "1",
"user-agent": "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/87.0.4280.88 Safari/537.36 Edg/87.0.664
    response = requests.get(urls,headers=headers)
    soup = BeautifulSoup(response.text, "html.parser")
    links = soup.find_all('td',{'class':'file'})
    for link in links:
        news_urlbefore =link.find('a').get('href')
        wget.download(news_urlbefore)
for i in range(1,131):
    urls='https://finance.naver.com/research/debenture_list.nhn?keyword=&brokerCode=&searchType=writeDate&writeFromDate=2005-01-01&writeToDate
    con(urls)
```



## 2-1) 데이터 수집

\* 데이터 수집

### 네이버 뉴스 기사



#### 뉴스기사

2005.01-2021.01 105,446개

- '금리' 키워드로 추출
  - 연합뉴스인포맥스
  - 연합뉴스
  - 이데일리
-

## 2-1) 데이터 수집

### 뉴스기사 크롤링 코드

```
from time import *
from datetime import *
from bs4 import BeautifulSoup
import urllib.request
from urllib.parse import quote
import pandas as pd
import datetime
import requests
title_set=[]
datetime_set=[]
article_set=[]
def cond(url):
    headers = {"authority": "search.naver.com",
               "method": "GET",
               "path": "/search.naver?where=news&query=%EA%B8%88%EB%A6%AC&sm=tab_opt&sort=0&photo=0&field=0&reporter_article=&
               "scheme": "https",
               "accept": "text/html,application/xhtml+xml,application/xml;q=0.9,image/avif,image/webp,image/apng,*/*;q=0.8,app
               "accept-enco 링크로 이동 (Ctrl+클릭) , br",
               "accept-lang 9,en;q=0.8",
               "cookie": "N 로드 중... _si#0_en#0_sp#0; NNB=7J5EOJKNQDZF6; nid_inf=809684172; NID_AUT=1PpROqf85py4ARS
               "referer": "https://search.naver.com/search.naver?where=news&query=%EA%B8%88%EB%A6%AC&sm=tab_opt&sort=0&photo=0
               "sec-fetch-dest": "document",
               "sec-fetch-mode": "navigate",
               "sec-fetch-site": "same-origin",
               "sec-fetch-user": "?1",
               "upgrade-insecure-requests": "1",
               "user-agent": "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/87.0.428
```

## 2-1) 데이터 수집

### 뉴스기사 크롤링 코드

```
response = requests.get(url,headers=headers)
soup = BeautifulSoup(response.text, "html.parser")
links = soup.find_all('div',{'class':'news_area'})
for link in links:
    #press = link.find('span',{'class':'_sp_each_source'}).get_text()
    news_urlbefore =link.find_all('a')#.get('href')
    news_url=news_urlbefore[1].attrs["href"]
    if ('einfomax.co.kr' in news_url):
        new_link = requests.get(news_url,headers={'User-Agent':'Mozilla/5.0'})
        news_html = BeautifulSoup(new_link.text,'html.parser')
        try:
            title = news_html.find('div',{'class':'article-head-title'}).get_text()
        except:
            title=""
        datetime = news_html.find('ul',{'class':'no-bullet auto-marbtm-0 line-height-6'})

        try:
            datetime2 = datetime.find_all('li')
            if '승인' in datetime2[1].get_text().strip():
                datetime2 = datetime2[1].get_text().strip().strip('승인').strip()
        except:
            datetime2=""
        try:
            article= news_html.find('div',{'id':'article-view-content-div'}).get_text().strip().replace("\t", "").replace("\n", "").replace("\v", "").replace("\f", "")
        except:
            article=""
        title_set.append(title)
        datetime_set.append(datetime2[:10])
        article_set.append(article)
```

## 2-1) 데이터 수집

### 뉴스기사 크롤링 코드

```
elif ('edaily.co.kr' in news_url):
    new_link = requests.get(news_url,headers={'User-Agent':'Mozilla/5.0'})
    news_html = BeautifulSoup(new_link.text,'html.parser')
    try:
        title = news_html.find('h2').get_text().strip("\t")
    except:
        title=""
    datetime = news_html.find('div',{'class':'dates'})
    datetime2 = datetime.find_all('p')
    try:
        datetime2=datetime2[0].get_text().strip()
    except:
        datetime2=""
    try:
        article= news_html.find('div',{'class':'news_body'}).get_text().strip().replace("\t", "").replace("\n", "")
    except:
        article=""
    title_set.append(title)
    datetime_set.append(datetime2[:10])
    article_set.append(article)
elif ('yna.co.kr' in news_url):
    new_link = requests.get(news_url,headers={'User-Agent':'Mozilla/5.0'})
    news_html = BeautifulSoup(new_link.text,'html.parser')
    try:
        title = news_html.find('h1',{'class':'tit'}).get_text().strip("\t")
    except:
        title=""
    try:
        datetime = news_html.find('p',{'class':'update-time'})
    except:
        datetime=""
    try:
        datetime = datetime.get_text().strip('송고시간')
    except:
        datetime=""
    article=""
```

## 2-1) 데이터 수집

### 뉴스기사 크롤링 코드

```
try:
    article= news_html.find('div',{'class':'story-news article'}).get_text().strip().replace("\t", "").replace("\n", "")
except:
    article=""
title_set.append(title)
datetime_set.append(datetime[:10])
article_set.append(article)
elif ('news.naver.com' in news_url):
    new_link = requests.get(news_url,headers={'User-Agent':'Mozilla/5.0'})
    news_html = BeautifulSoup(new_link.text,'html.parser')
    try:
        title = news_html.find('h3',{'id':'articleTitle'}).get_text()
    except:
        title=""
    try:
        datetime = news_html.find('span',{'class':'t11'}).get_text()
    except:
        datetime=""
    try:
        article= news_html.find('div',{'id':'articleBodyContents'}).get_text().strip().replace("\t", "").replace("\n", "")
    except:
        article=""
    title_set.append(title)
    datetime_set.append(datetime[:10])
    article_set.append(article)
else:
    new_link = requests.get(news_url,headers={'User-Agent':'Mozilla/5.0'})
    news_html = BeautifulSoup(new_link.text,'html.parser')
    title = news_html.find('h3',{'id':'articleTitle'}).get_text()
    datetime = news_html.find('span',{'class':'t11'}).get_text()
    article= news_html.find('div',{'id':'articleBodyContents'}).replace("\t", "").replace("\n", "")
    title_set.append(title)
    datetime_set.append(datetime[:10])
    article_set.append(article)
```

# 2-1) 데이터 수집

## 뉴스기사 크롤링 코드

```
url_query = quote('금리')
thirty=5100
now = datetime.datetime.now()
for i in range(5101,5401):#5848-일 반복
    now_minus= now - datetime.timedelta(days=i)
    x=now_minus.strftime('%Y.%m.%d')
    y=now_minus.strftime('%Y%m%d')
    url = 'https://search.naver.com/search.naver?where=news&query='+url_query+'&sm=tab_opt&sort=0&photo=0&field=0&reporter_article=&pd=3&ds={}&de={}&docid={}'
    cond(url)
    print("여기까지"+str(i))
    try:
        headers = {"authority": "search.naver.com",
                    "method": "GET",
                    "path": "/search.naver?where=news&query=%EA%B8%88%EB%A6%AC&sm=tab_opt&sort=0&photo=0&field=0&reporter_article=&pd=3&ds=2021.01.04&de=2021.01.04&docid={}"
                    "scheme": "https",
                    "accept": "text/html,application/xhtml+xml,application/xml;q=0.9,image/avif,image/webp,image/apng,*/*;q=0.8,application/signed-exchange;v=b3;q=0.9",
                    "accept-encoding": "gzip, deflate, br",
                    "accept-language": "ko,en-US;q=0.9,en;q=0.8",
                    "cookie": "NRTK=ag#all_gr#1_ma#-2_si#0_en#0_sp#0; NNB=7J5EOJKNQDZF6; nid_inf=809684172; NID_AUT=1PpROqf85py4ARSAzT7KkM82jnuLvdfEmRI1bbOBfp1GKeFu",
                    "referer": "https://search.naver.com/search.naver?where=news&query=%EA%B8%88%EB%A6%AC&sm=tab_opt&sort=0&photo=0&field=0&reporter_article=&pd=3&ds=2021.01.04&de=2021.01.04&docid={}"
                    "sec-fetch-dest": "document",
                    "sec-fetch-mode": "navigate",
                    "sec-fetch-site": "same-origin",
                    "sec-fetch-user": "?1",
                    "upgrade-insecure-requests": "1",
                    "user-agent": "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/87.0.4280.88 Safari/537.36"}
        response = requests.get(url,headers=headers)
        soup = BeautifulSoup(response.text, "html.parser")
        next =soup.find('a',{'class':'btn_next'}).get('href')
        url2 = 'https://search.naver.com/search.naver' + next
        cond(url2)
    except:
        pass
    if i%30==0:
        newdata_set={'date':datetime_set,'title':title_set,'content':article_set}
        newdata_set=pd.DataFrame(newdata_set)
        newdata_set.to_csv('newdata_set'+str(thirty)+'.csv',encoding='utf-8')
        thirty+=1
```

## 2-1) 데이터 수집

\* 데이터 수집

### 금융통화위원회 의사록



#### 의사록

2005.07-2020.11 145개

- 한국은행 금융통화위원회  
의사록 섹션 2,3 추출

## 2-1) 데이터 수집

### 의사록 크롤링 코드

```
from bs4 import BeautifulSoup
import pandas as pd
import requests
import wget
import os
import fitz
import re
name_list=[]
dates_list=[]
text_list=[]
def con(urls):
    headers = {"authority": "finance.naver.com"
, "method": "GET"
, "path": "/research/debenture_list.nhn?keyword=&brokerCode=&searchType=writeDate&writeFromDate=2005-01-01&wri
, "scheme": "https",
"accept": "text/html,application/xhtml+xml,application/xml;q=0.9,image/webp,image/apng,*/*;q=0.8,application/
"accept-encoding": "gzip, deflate, br",
"accept-language": "ko,en;q=0.9,en-US;q=0.8",
"cookie": "NRTK=ag#all_gr#1_ma#-2_si#0_en#0_sp#0; NNB=V5Q30A3RX7JF6; nx_ssl=2; naver_stock_codeList=005930%7C
"referer": "https://finance.naver.com/research/debenture_list.nhn?keyword=&brokerCode=&searchType=writeDate&w
"sec-fetch-dest": "document",
"sec-fetch-mode": "navigate",
"sec-fetch-site": "same-origin",
"sec-fetch-user": "?1",
"upgrade-insecure-requests": "1",
"user-agent": "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/87.0.4
```



## 2-1) 데이터 수집

### 의사록 크롤링 코드

```
dates = soup.find_all('td' ,{'class':'date'})
for i in range(0,int(len(names)/2)):
    name_list.append(names[2*i].text)
    dates_list.append(dates[2*i].text)
for i in range(1,3874):#3874
    filename='C:/Users/691/kdigital/채권분석리포트/{i}.pdf'.format(i)
    doc=fitz.open(filename)
    page_count=doc.pageCount
    page=0
    a=""
    while (page<page_count):
        p=doc.loadPage(page)
        #print(p.getText())
        a=a+p.getText().strip('\n').strip('\r').strip().replace('\n',"")
        page+=1
    text_list.append(a)
for i in range(1,131):
    urls='https://finance.naver.com/research/debenture_list.nhn?keyword=&brokerCode=&se
con(urls)
# print(len(dates_list))
# print(len(name_list))
# print(len(text_list))
bonds_data2 = {"date":dates_list,"title":name_list,"content":text_list}
bonds_data2 = pd.DataFrame(bonds_data2)
bonds_data2.to_csv('bonds_data2_save.csv', encoding='utf-8')
```

## 2-1) 데이터 수집

\* 데이터 수집

### 콜금리, 기준금리



#### 콜금리, 기준금리

기준금리

2005.01-2020.12 5,659개

콜금리

2005.01-2020.12 2,829개

---

## 2-1) 데이터 수집

### 콜금리 크롤링 코드

```
import csv
from bs4 import BeautifulSoup
import requests
import pandas as pd
call_rate=[]
call_date=[]
def print_stock_price(page_num):
    result = [[], []]
    for n in range(page_num):
        url = 'https://finance.naver.com/marketindex/interestDailyQuote.nhn?marketindexCd=IRR_CALL&page='+str(n+1)
        r = requests.get(url)
        html = r.content
        soup = BeautifulSoup(html, 'html.parser')
        tr = soup.select('table > tbody > tr')
        for i in range(1, len(tr)-1):
            if tr[i].select('td')[0].text.strip():
                a=tr[i].select('td')[0].text.strip()
                b=tr[i].select('td')[1].text.strip()
                call_rate.append(b)
                call_date.append(a)
        call_data={'날짜':call_date,'콜금리':call_rate}
        call_data=pd.DataFrame(call_data)
        call_data.to_csv('call_rate1.csv',encoding='utf-8')
    print_stock_price(573)
```

## 2-1) 데이터 수집

### 기준금리 크롤링 코드

```
date=[]
nom_rate=[]
url1 = 'https://www.bok.or.kr/portal/singl/baseRate/list.do?dataSeCd=01&menuNo=200643'
r = requests.get(url1)
html = r.content
soup = BeautifulSoup(html, 'html.parser')
tr = soup.select('.fixed tbody tr')
for i in range(1, len(tr)-1):
    if tr[i].select('td')[0].text.strip():
        a=tr[i].select('td')[0].text.strip()+'. '
        b=tr[i].select('td')[1].text.strip().replace('월 ', '.').replace('일', '')
        c=tr[i].select('td')[2].text.strip()
        date.append(a+b)
        nom_rate.append(c)
nom_rate={'날짜':date, '기준금리':nom_rate}
nom_rate=pd.DataFrame(nom_rate)
nom_rate.to_csv('nom_rate.csv',encoding='utf-8')
print("끝")
```

## 2-2) 텍스트 전처리

### 뉴스기사

Unnamed: 0	Unnamed: 0.1	date	title	content	ngrams	
2	2	19	2021-01-05	"현금 쥐도 소비 대신 저축만...포스트 코로나 규제완화에 방점"	[전미경제학회 2021]재무장관, 하버드대 총장 지낸 래리 서머스명저 '테일러 경제...	fed/NNG;양적완화/NNG;축소/NNG,마이너스/NNG;떨어/VV,침체/NNG;...
3	3	18	2021-01-05	"지금 주식시장 과열인가"...증권사 리서치센터장들이 답했다(종합)	(서울=연합뉴스) 증권팀 = 코스피 지수가 5일 2,990선에 오르며 3,000 시...	회복/NNG;이익/NNG;개선/NNG,어렵/VA;유동성/NNG;위험/NNG,인플레이...
4	4	17	2021-01-05	씨티은행, 3개월 만기 2.8% 정기예금 가입 이벤트	[이데일리 김유성 기자] 한국씨티은행은 이달 29일 혹은 한도소진시까지 개인고객을 ...	통화/NNG;가격/NNG;하락/NNG,가격/NNG;하락/NNG;우려/NNG
8	8	12	2021-01-05	12월 신용대출 11개월만에 줄었지 만...전세대출 증가폭 커져	지난 한해 5대은행 가계대출 10%↑...신용대출 22% 전세대출 31% 급증(서울=연...	신용/NNG;대출/NNG;급증/NNG,가계/NNG;대출/NNG;증가/NNG,가계/N...
12	12	4	2021-01-05	美 국채 10년물 금리 상승...조지아 선거 주시	(서울=연합인포맥스) 신윤우 기자 = 5일 아시아 시장에서 10년 만기 미국 국채 ...	국채/NNG;가격/NNG;하락/NNG,회복/NNG;물가/NNG;상승/NNG
...	...	...	...	...	...	
105440	105440	7347	2004-12-31	LG카드 증자 극적 타결(종합)	[edaily 김기성 박호식 최한나기자] LG카드 증자를 위한 채권단과 LG그룹의 ...	채권/NNG;증가/NNG,채권/NNG;인하/NNG
105441	105441	7346	2004-12-31	한은총재 신년사 "내년 경제 불확실성 여전히 높다"	[edaily 김현동기자] 박승 한국은행 총재는 31일 내년 통화신용정책 방향과 관...	외환시장/NNG;모니터링/NNG;강화/NNG,정책/NNG;효과/NNG;제약/NNG;...
105444	105444	7343	2004-12-31	<채권> LG카드 해결에 회사채 금리 하락	(서울=연합뉴스) 김문성기자= LG카드 사태의 해결로 회사채 금리가 하락했다. ...	국고채/NNG;내리/VV
105445	105445	7342	2004-12-31	지표금리 3.28% 마감..올 평균금리 4.09%(마감)	[edaily 강종구기자] 채권금리가 올해 마지막 거래를 보험권으로 마감했다. 장중...	유료뉴스/NNG;시장/NNG;플러스/NNG,물량/NNG;부담/NNG;우려/NNG,통...
105446	105446	7361	2004-12-31	(월가시각)예민해진 시장	[뉴욕=edaily 안근모특파원] 중국 쇼크가 뉴욕시장을 이틀째 지배했다. 어제는 ...	이익/NNG;모멘텀/NNG;악화/NNG,가격/NNG;부담/NNG,물가/NNG;오르/VV

## 2-2) 텍스트 전처리

### 채권분석리포트

Unnamed: 0	Unnamed: 0.1	date	title	content	ngrams
0	0	0	20.12.30 원점회귀(原點回歸)	원점회귀(原點回歸) Issue Comment Credit Credit 김상만, ...	전망/NNG;상향/NNG;전망/NNG;하향/NNG하위등급/NNG;스프레드/NNG;축...
1	1	1	20.12.23 포트폴리오 다변화로 가는 여정	8 글로벌 채권   2020.12.222020.12.222020.12.222020....	국고채/NNG;스프레드/NNG;축소/NNG;확대/NNG국채/NNG;금리/NNG;변동...
2	2	2	20.12.23 2021년 전망 업데이트	2021년 전망 업데이트 Issue Comment Credit Credit 김상...	전망/NNG;상향/NNG;전망/NNG;하향/NNG전망/NNG;하향/NNG;경기/NN...
3	3	3	20.12.18 일드 픽업 이제는 장기물이 답이다	그림 1. P/F 및 벤치마크 수익률 추이 주1: 매매수수료 및 이자소득세는 반영...	회사채/NNG;신용등급/NNG;상향/NNG기업/NNG;펀더멘털/NNG;회복/NNG채...
4	4	4	20.12.18 10년 발행 감소, 초장기 공급부담 완화	Compliance Notice 본 조사자료는 고객의 투자에 정보를 제공할 목적으로...	경기/NNG;불확실성/NNG;우려/NNG자산/NNG;듀레이션/NNG;확대/NNG부채...
...	...	...	...	...	...
3868	3868	3868	08.04.21 항상 열리는 있는 가능성	FOMC 이후 원자재 가격 상승세 둔화 여부에 관심 대내외 요인이 금리에 상반된 영...	fed/NNG;원자재/NNG;가격/NNG;상승/NNG원자재/NNG;가격/NNG;상승...
3869	3869	3869	08.04.14 총재의 변신은 무죄다?	2008_04 월간채권투자채 권 시 장전 망금 용 시 장차 트북對應과 限界월간채...	미/NNG;국채/NNG;금리/NNG;안전자산/NNG;선호/NNG경기/NNG;개선/N...
3870	3870	3870	08.04.07 시간차 공격	Fixed Income Weekly 2008. 4. 7. #918 Fixed In...	글로벌/NNG;신용경색/NNG;완화/NNG단기/NNG;인플레이션/NNG;우려/NNG...
3871	3871	3871	08.04.07 돌아온 박스권	4월 금통위는 정부-한은간 policy mix 합의의 반영일 수도 예상대로, 올해...	예상/NNG;경제/NNG;성장률/NNG;하향/NNG금리/NNG;인하/NNG;기대/N...
3872	3872	3872	08.04.01 대우증권 월간 채권투자 (4월)	경기 전망은 비교적 분명... 인플레이션 명분이 시점 결정할 것 현 금리 레벨이 5월 인하...	미/NNG;국채/NNG;금리/NNG;안전자산/NNG;선호/NNG금리/NNG;인하/N...

## 2-2) 텍스트 전처리

### 의사록

Unnamed: 0		title	date	content	ngrams
0	0	금융통화위원회 의사록(2005년도 제 14차)(2005.7.7)	2005-08-23	일부 위원은 지난달 연평도 포격 사태·「」에도 불구하고 스왑레이트 상승으로 ...	소비자/NNG;물가/NNG;농산물/NNG;가격/NNG;안정/NNG세계/NNG;경제/...
1	1	금융통화위원회 의사록(2005년도 제 21차)(2005.10.11)	2005-11-29	일부 위원은 미국의 재정지출 확대, 보호무역주의 강화 등으로 미 연준의 금리인상 ...	국제/NNG;금융시장/NNG;fed/NNG;금리/NNG;인상/NNG불확실성/NNG;...
2	2	금융통화위원회 의사록(2006년도 제 2차)(2006.1.12)	2006-02-28	일부 위원은 최근 시중자금에 제 금융권의, 2「」 실적배당 상품에서 은행의 정기...	소비자/NNG;물가/NNG;국제/NNG;유가/NNG;하락/NNG소비자/NNG;물가/...
3	3	금융통화위원회 의사록(2006년도 제 3차)(2006.1.26)	2006-03-14	일부 위원은 프리미엄이 금융기관에 대·CDS「」한 신용평가를 정확하게 나타...	소비자/NNG;물가/NNG;농축산물/NNG;가격/NNG;상승/NNG물가/NNG;상...
4	4	금융통화위원회 의사록(2006년도 제 4차)(2006.2.9)	2006-03-28	일부 위원은 일본이 과거 20여년 동안 저성장 하에서도 경상수지가 흑자를 보이면서...	소비자/NNG;물가/NNG;상승률/NNG;물가/NNG;안정/NNG물가/NNG;상승률...
...	...	...	...	...	...
140	140	금융통화위원회 의사록(2019년도 제 22차)(2019.11.29.)	2019-12-17	일부 위원은 자본유입에 대해서는 관대 한 반면 자본유출에 대해서는 비교적 엄격한 경...	fed/NNG;금리/NNG;인상/NNG;기대/NNG;상승/NNGGecb/NNG;금리/...
141	141	금융통화위원회 의사록(2019년도 제 24차)(2019.12.26.)	2020-01-10	일부 위원은 2002년부터 거의 예외 없 이 연초에는 원화의 강세 현상이 나타나 왔...	근원/NNG;인플레이션/NNG;소비자/NNG;물가/NNG;안정/NNG국채/NNG;수...
142	142	금융통화위원회 의사록(2020년도 제 1차)(2020.1.17.)	2020-02-04	일부 위원은 외화유동성 사정을 분석, 「」 함에 있어서 유동외채비율이 중요한...	소비자/NNG;물가/NNG;국제/NNG;유가/NNG;하락/NNG금융시장/NNG;국제...
143	143	금융통화위원회 의사록(2020년도 제 20차)(2020.09.10)	2020-09-29	일부 위원은 최근 국제금융시장에서 국제신·「」용평가사들이 선진국의 국가신용등...	인플레이션/NNG;기대/NNG;물가/NNG;상승률/NNG;하락/NNG경제/NNG;부...
144	144	금융통화위원회 의사록(2020년도 제 25차)(2020.11.26)	2020-12-15	일부 위원은 최근 국제금융시장에서 금융기, 「」 간간 거래상대방 위험이 증대되면...	물가/NNG;국제/NNG;원자재/NNG;가격/NNG;하락/NNG기업/NNG;자금/N...

## 2-2) 텍스트 전처리

### 전체

Unnamed: 0_x	title	date	content	ngrams	Unnamed: 0.1_x	Unnamed: 0.1.1	Unnamed: 0_y	Unnamed: 0.1_y	콜금리	date2	call2	up_down
0	금융통화위원회의사록(2005년도 제14차)(2005.7.7)	2005-08-23	일부 위원은 지난 달 연평도 포격 사태 · 「 J 」 에도 불구하고 스왑레이트 상승으로 ...	소비자/NNG;물가/NNG;농산물/NNG;가격/NNG;안정/NNG 세계/NNG;경제/...	NaN	NaN	2724.0	2724.0	3.24	2005-07-26	3.27	down
1	금융통화위원회의사록(2005년도 제21차)(2005.10.11)	2005-11-29	일부 위원은 미국의 재정지출 확대, 보호무역주의 강화 등으로 미 연준의 금리인상 ...	국제/NNG;금융시장/NNG;fed/NNG;금리/NNG;인상/NNG를 확실성/NNG;...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	금융통화위원회의사록(2006년도 제2차)(2006.1.12)	2006-02-28	일부 위원은 최근 시중자금이 제 금융권의 , 2 「 J 」 실적배당상품에서 은행의 정기...	소비자/NNG;물가/NNG;국제/NNG;유가/NNG;하락/NNG소비자/NNG;물가/...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	금융통화위원회의사록(2006년도 제3차)(2006.1.26)	2006-03-14	일부 위원은 프리미엄이 금융기관에 대 · CDS 「 J 」 한 신용평가를 정확하게 나타...	소비자/NNG;물가/NNG;농축산물/NNG;가격/NNG;상승/NNG물가/NNG;상...	NaN	NaN	2623.0	2623.0	3.98	2006-02-14	3.73	up
4	금융통화위원회의사록(2006년도 제4차)(2006.2.9)	2006-03-28	일부 위원은 일본이 과거 20여년 동안 저성장 하에 서도 경상수지가 흑자를 보이면서...	소비자/NNG;물가/NNG;상승률/NNG;물가/NNG;안정/NNG물가/NNG;상승률...	NaN	NaN	2615.0	2615.0	3.97	2006-02-28	3.75	up
...	...	...	...	...	...	...	...	...	...	...	...	...
84851	105440 LG카드 중자 극적 타결(종합)	2004-12-31	[edaily 김기성 박호식 최한나기자] LG카드 중자를 위한 채권단과 LG그룹의 ...	채권/NNG;중가/NNG, 채권/NNG;인하/NNG	105440.0	7347.0	NaN	NaN	NaN	NaN	NaN	NaN
84852	105441 한은총재 신년사 "내년 경제 부활" 신년 기자회견	2004-12-31	[edaily 김현등기자] 박승 한국은행 총재는 31일 내년	외환시장/NNG;모니터링/NNG;강화/NNG, 정책/NNG; 증가/NNG; 제	105441.0	7346.0	NaN	NaN	NaN	NaN	NaN	NaN



## 2-3) 텍스트 분석

### 1 사전구축

	level_0	index	up_count	down_count	total_count
0	0	가격/NNG;상승/NNG	581.0	427.0	1008.0
1	1	경제/NNG;성장/NNG	543.0	482.0	1025.0
2	2	물가/NNG;안정/NNG	527.0	357.0	884.0
3	3	가격/NNG;하락/NNG	384.0	386.0	770.0
4	4	경제/NNG;회복/NNG	378.0	392.0	770.0
...	...	...	...	...	...
2159	5786	생산/NNG;감축/NNG	3.0	14.0	17.0
2160	5790	담배가격/NNG;인상/NNG	3.0	17.0	20.0
2161	6817	대출/NNG;촉진/NNG	3.0	14.0	17.0
2162	6915	심각/NNG;훼손/NNG	3.0	14.0	17.0
2163	10296	유동성함정/NNG;빠지/VV	2.0	14.0	16.0

## 2-3) 텍스트 분석

〈Polarity Score 공식〉

$$Polarity\ Score = \frac{P(Ngram | Hawkish)}{P(Ngram | Dovish)} = \frac{P(Ngram | Hawkish) / P(Hawkish)}{P(Ngram | Dovish) / P(Dovish)}$$

## 2-3) 텍스트 분석

# 1 사전구축

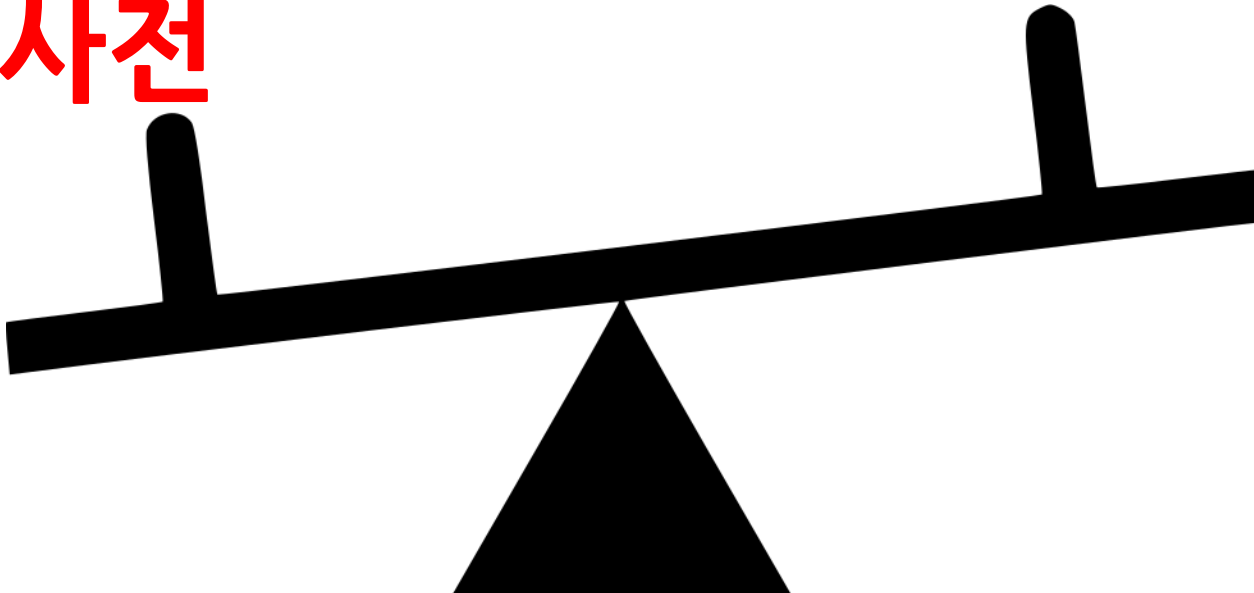
	level_0	index	up_count	down_count	total_count	polar_score	polarity
0	0	가격/NNG;상승/NNG	581.0	427.0	1008.0	1.265972	nothing
1	1	경제/NNG;성장/NNG	543.0	482.0	1025.0	1.048162	nothing
2	2	물가/NNG;안정/NNG	527.0	357.0	884.0	1.373467	Hawkish
3	3	가격/NNG;하락/NNG	384.0	386.0	770.0	0.925592	nothing
4	4	경제/NNG;회복/NNG	378.0	392.0	770.0	0.897184	nothing
...	...	...	...	...	...	...	...
2159	5786	생산/NNG;감축/NNG	3.0	14.0	17.0	0.199374	Dovish
2160	5790	담배가격/NNG;인상/NNG	3.0	17.0	20.0	0.164190	Dovish
2161	6817	대출/NNG;촉진/NNG	3.0	14.0	17.0	0.199374	Dovish
2162	6915	심각/NNG;훼손/NNG	3.0	14.0	17.0	0.199374	Dovish
2163	10296	유동성함정/NNG;빠지/VV	2.0	14.0	16.0	0.132916	Dovish

## 2-3) 텍스트 분석

### 1 사전구축

부정  
사전

긍정  
사전





# Thank you

---