

1. 경제 기사 데이터를 마침표가 있는 문장별로 분리

```
In [4]: import os
print(os.getcwd())
```

/Users/heosangbeom

```
In [5]: os.chdir("/Users/heosangbeom/Desktop/data_final")
os.getcwd()
```

Out[5]: '/Users/heosangbeom/Desktop/data_final'

```
In [6]: import sys
import matplotlib.pyplot as plt
```

```
In [7]: import pandas as pd
import numpy as np
df = pd.read_csv("total_articles_100.csv", encoding='cp949', )
df.head(10)
```

Out[7]:

	Unnamed: 0	content_new
0	0	8개월 앞으로 다가온 베이징 올림픽이후 중국증시에 대한 전망이 크게 엇갈리고 있다고...
1	4	모간스탠리의 이코노미스트 출신인 앤디 시에는 올 한해 아시아 증시가 상승세를 보일 ...
2	14	주식시장은 늘 '낙관'과 '비관'간 대립의 연속이다.최근 주식시장 강세가 지속되면서...
3	38	지난해 우리나라 수출이 5년 연속 두자릿수 증가율을 기록하며 3700억달러를 넘어섰...
4	54	프로그램 매물 충격으로 코스피가 2008년 첫 거래일에 2% 이상 하락하며 장을 마...
5	70	국제유가가 사상 처음으로 배럴당 100달러를 돌파하며 세자리수 유가시대가 시작됐다....
6	88	[한겨레] 국제유가가 한때 배럴당 100달러를 넘어섬에 따라 새 정부의 새해 경제정...
7	109	지난해 최고의 투자처는 역시 중국시장이었는데요. 올해는 중국시장에 대한 전망이 조금...
8	128	뉴욕=연합인포맥스) 김홍규 특파원 = 1월3일로 끝난 주간의 미국의 30년만기 ...
9	136	오늘 약보합으로 출발한 국제 유가는 미국의 원유 재고가 월가의 예상치 보다 많은 4...

```
In [8]: content = df[["content_new"]]
content.head()
```

Out[8]:

	content_new
0	8개월 앞으로 다가온 베이징 올림픽이후 중국증시에 대한 전망이 크게 엇갈리고 있다고...
1	모간스탠리의 이코노미스트 출신인 앤디 시에는 올 한해 아시아 증시가 상승세를 보일 ...
2	주식시장은 늘 '낙관'과 '비관'간 대립의 연속이다.최근 주식시장 강세가 지속되면서...
3	지난해 우리나라 수출이 5년 연속 두자릿수 증가율을 기록하며 3700억달러를 넘어섰...
4	프로그램 매물 충격으로 코스피가 2008년 첫 거래일에 2% 이상 하락하며 장을 마...

```
In [9]: df['split'] = df.content_new.str.split('.').str[0]
df['split']
```

```
Out[9]: 0      8개월 앞으로 다가온 베이징 올림픽이후 중국증시에 대한 전망이 크게 엇갈리고 있다고...
1      모간스탠리의 이코노미스트 출신인 앤디 시에는 올 한해 아시아 증시가 상승세를 보일 ...
2      주식시장은 늘 '낙관'과 '비관'간 대립의 연속이다
3      지난해 우리나라 수출이 5년 연속 두자릿수 증가율을 기록하며 3700억달러를 넘어섰다
4      프로그램 매물 충격으로 코스피가 2008년 첫 거래일에 2% 이상 하락하며 장을 마감했다
5      국제유가가 사상 처음으로 배럴당 100달러를 돌파하며 셰리리수 유가시대가 시작됐다
6      [한겨레] 국제유가가 한때 배럴당 100달러를 넘어섬에 따라 새 정부의 새해 경제정...
7      지난해 최고의 투자처는 역시 중국시장이었는데요
8      뉴욕=연합인포맥스) 김홍규 특파원 = 1월3일로 끝난 주간의 미국의 30년만기 ...
9      오늘 약보합으로 출발한 국제 유가는 미국의 원유 재고가 월가의 예상치 보다 많은 4...
10     (서울=연합인포맥스) 한창현 기자= NH투자증권은 미국 경기침체 우려와 물가 국...
11     고유가 및 국제금융시장 불안등 부정적인 대외여건 여파로 가계와 기업 심리지표가 악화...
12     하반기 이후엔 ""완화"" vs ""긴축"" 전망 팽팽- 1월 콜금리 ""동결 확실시""
13     (서울=연합인포맥스) 이효지 기자= 상하이증시는 위안화 강세와 기업들의 실적 전망에...
14     ■ 파이낸셜 타임스▲中 국부펀드 비판에 방어 나서-선진국들이 운용투명성과 정당성에 대...
15     미국 경제에 부정적인 조짐들이 가중되고 있는 것으로 보인다
16     유가 고공행진을 업고 석유 수출로 막대한 달러를 벌어들인 걸프만 연안 국가들이 해외...
17     올해 외국계 기업의 취업문이 지난해보다 좁아질 것으로 보입니다
18     - 다우 소폭 반등 성공-나스닥 7일째 하락 - ""美경기 이미 후퇴 국면"" 진단...
19     건설업 구직자들은 올해 취업난 해소에 대한 기대감이 높은 것으로 나타났다
20     (서울=연합뉴스) 박대환 기자 = 임종룡 재정경제부 경제정책국장은 9일 과천청사에...
21     미국의 서브프라임 모기지 관련 부실이 올 상반기까지 계속 확대될 것이라는 전망이 나왔다
22     중동지역 국가들이 올해 기업들의 인수합병(M&A) 시장에서 '태풍의 눈'으로 떠오를...
23     국은행 금융통화위원회가 새해 첫 회의를 갖고 1월 콜금리 목표치를 현 수준인 연 5
24     [애그플레이션 비상]농산물발 물가 상승인 '애그플레이션'(Agflation)에 대한...
25     서울=연합인포맥스) 이규창 기자= 10일 한국은행 금융통화위원회는 1월 정기회의를 ...
26     올해 코스닥 지수가 불안한 대외변수에도 불구하고 코스피에 비해 선방하고 있다
27     국은행 금융통화위원회는 10일 1월 콜금리 운용목표를 연 5
28     황건호 증권업협회 회장은 11일 ""자본시장의 의견 수렴을 위해 금융통화위원을 증권...
29     ""인위적인 경기부양책은 쓰지 않겠다

...

70     김석동 재정경제부 제1차관은 ""국내 성장과 물가 고용 등 대내적 요인들은 양호한 ...
71     우리나라 고용률이 최근 6년간 63%대를 벗어나지 못하며 제자리걸음을 하고 있는 것...
72     굿모닝신한증권은 23일 미국 연준의 금리인하 단행으로 패닉현상은 단절되면서 증시가 ...
73     도미니크 스트로스 칸 국제통화기금(IMF) 총재가 ""미국의 경기 침체가 전 세계로 ...
74     - 금융시장 안정대책으로 국민연금 주식자금 조기 집행을 발표한 재정경제부가 24일...
75     국제 유가가 90달러를 밀돌며 하락 마감했습니다
76     긴급 수입 제한 조치인 세이프가드가 엄격해진 발동 요건이나 세계무역기구 분쟁시 조...
77     장중 코스피지수 ****선이 무너지며 주식시장이 공황상태에 빠졌다
78     미 연방준비제도이사회(FRB)의 기준금리 긴급 인하와 정부의 경기부양책에 대한 의지...
79     외국인 주식매도로 달러 수요↑...15개월만에 최고엔캐리 거래청산 엔화도 강세원·달러 ...
80     미국의 정책금리 인하로 한국은행 역시 콜금리 목표치를 내릴 수밖에 없을 것이라는 기...
81     증권예탁결제원이 지난해 금융기관간 시스템 연계를 통해 실시간(real-time)으...
82     5% 상승했다고 블룸버그 통신이 24일 보도했다
83     세계경제포럼(WEF) 주최로 지난 23일(현지시간)부터 스위스에서 열리고 있는 다보...
84     뉴욕 증시가 장 막판 반등에 성공했습니다
85     지난 24일 중국 국가통계국이 2007년 주요 경제지표를 발표한 가운데 전문가들은 ...
86     다보스 포럼 폐막...ECB 트리세 총재 금리인하 여부 묵묵부답미국의 서브프라임 모기지...
87     [클라크 쉐 HSBC 대체투자그룹 미국 연구소장]올해 시장 변동성이 확대되면서 헤지...
88     환율이 나홀만에 반등 950원선에 바짝 다가섰다
89     글로벌 증시의 악재가 아시아를 강타하며 국내 증시도 큰폭으로 하락하고 있다
90     올 영업이익 2兆원대 회복 예상최근 현대차의 주가가 분기실적 개선에도 불구하고 부진한...
91     29일 뉴욕 증시 개장을 앞두고 지수 선물은 일제히 상승하고 있다
92     - ""금융시장 동요로 인해 더 악화될 수도"" 전미경제조사연구소(NBER) 소장인...
```

```

93  美 의회에 한미 FTA 조속한 비준동의 촉구      (워싱턴=연합뉴스) 김병수 특파원...
94  사공일 대통령직 인수위원회 국가경쟁력강화특위 공동위원장은 지난주 스위스에서 열린 ...
95  한은 中 주가·부동산 분석..."장기불황 가능성 낮아" 최근 미국의 경기침체 우려와...
96  [미국증시 반등...FRB 50bp추가인하 기대감] 미국증시가 전일 낙폭을 만회하며 상승했다
97  지난해 인도와 두바이 간 무역 거래가 크게 늘어났다고 인도 경제지 이코노믹 타임스가...
98  (서울=연합인포맥스) 도쿄증시는 미 연방준비제도이사회(FRB)가 기준금리를 인하할...
99                                30일 아시아 증시는 오전장에서 혼조세가 뚜렷하다
Name: split, Length: 100, dtype: object

```

2. 문장에서 동시 출현 빈도가 5번째로 높은 단어 토큰 쌍을 찾자

2-1. 단어 토큰은 okt 형태소 분석기 함수인 nouns()로 얻어진 단어 토큰을 이용한다.

```

In [10]: from konlpy.tag import Okt
         okt = Okt()

```

```

In [11]: text = "30일 아시아 증시는 오전장에서 혼조세가 뚜렷하다"
         print(okt.nouns(text))

['아시아', '증시', '오전', '혼', '조세']

```

```

In [12]: # dataframe을 list로 변환 후 텍스트화
         text_list = np.array(df['split'].tolist())
         text_list
         text = ",".join(text_list)

```

```

In [22]: #단어 단위로 토큰화
         nouns = okt.nouns(text)
         print(nouns)

'정당성', '대해', '제기', '비판', '반박', '중', '국부펀드', '세계', '경제', '위협',
'것', '편협', '인식', '금융', '보호', '주의', '금지', '유러', '존', '경제', '폭',
'위축', '미국', '경제', '부정', '조짐', '가중', '것', '유가', '행진', '석유', '수출',
'막대', '달러', '걸프만', '연안', '국가', '해외', '기업', '인수', '박차', '올해', '외
국', '기업', '취업', '문', '지난해', '것', '다우', '폭', '반등', '성공', '나스닥',
'하락', '경기', '후퇴', '국면', '진단', '달', '유가', '대로', '급락', '현지', '시간',
'뉴욕', '주식시장', '혼', '세로', '건설업', '구직', '자', '올해', '취업', '해소', '대
한', '대감', '것', '서울', '연합뉴스', '박대', '기자', '임종룡', '재정경제부', '경제정
책', '국장', '과천', '청사', '올해', '경제', '운용', '방향', '브리핑', '올해', '우
리', '경제', '위협', '요인', '미국', '경기', '둔화', '중국', '인플레이션', '고유', '국
제', '금융시장', '변동성', '등', '것', '라며', '반면', '수가', '점', '새', '정부',
'출범', '대감', '결집', '등', '기회', '요인', '수', '고', '말', '미국', '서브프라임',
'모기지', '관련', '부실', '상반기', '계속', '확대', '것', '전망', '중동', '지역', '국
가', '올해', '기업', '인수합병', '시장', '태풍', '눈', '것', '전망', '국', '은행',
'금융통화위원회', '새해', '첫', '회의', '콜', '금리', '목표', '현', '수준', '연', '애
그플레이션', '비상', '농산물', '발', '물가', '상승', '애그플레이션', '대한', '우려', '고
조', '가운데', '남아시아', '지역', '식량난', '것', '서울', '연합인포맥스', '이규창', '기
자', '한국은행', '금융통화위원회', '정기', '회의', '개최', '콜', '목표', '금리', '연',
'올해', '나스닥', '지수', '대인', '변수', '부각', '크시하', '선박', '그', '은행', '그

```

```
In [25]: from collections import Counter
         from konlpy.tag import Twitter
         import pytagcloud

         count = Counter(nouns)
         count.most_common(10)
```

```
Out[25]: [('미국', 33),
          ('것', 27),
          ('금리', 18),
          ('증시', 17),
          ('경제', 17),
          ('경기', 16),
          ('지난해', 14),
          ('등', 13),
          ('침체', 12),
          ('수', 12)]
```

따라서, 5번째로 높은 순서를 가지고 있는 단어는 "경제"이다.

```
In [ ]:
```