

# 딥러닝 기반 핵심 산업별 빅데이터 분석

## <키워드 분석 프로젝트 - 2팀>

일 자	2018년 11월 19일	주 제	정치 키워드 분석
팀 장	유영재 <yyj3053@naver.com>	팀 원	장은경, 공정배, 민유진

### 주제 : 네이버 상위권 뉴스와 정부 지지율 관계성 파악

- 정치 / 경제 / 사회 3가지 카테고리를 이용
- 기사 제목의 키워드 빈도수를 이용
- 기사 랭크 top5 와 top30 나눠서 분석
- 문재인 정부의 1년 간의 지지율 데이터 추출 및 관계도 파악 (2017년11월1일~2018년11월15일)

#### 1. 기사 크롤링 함수

- 대통령 지지율 크롤링 함수
- 취임부터 2018년 11월까지 데이터 추출
- 출처: 위키피디아, 한국갤럽

#### 2. 키워드 추출하는 함수

- 빈도수가 높은 순으로 정렬
- 원하는 카테고리를 정해서 추출

#### 3. 단어 필터

- 버전을 2가지로 형성 (많은 키워드 수집을 위한 약한 필터 / 임팩트 있는 키워드 수집을 위한 강한 필터)
  - 정치
  - 경제
  - 사회
  - 종합 (정치 경제 사회 모두를 한꺼번에 고려한 것)

#### 4. main 함수

- extract\_keyword 내부에서 호출 함수 이용
- 키워드를 얻기 위한 자잘한 과정들을 내부적으로 모듈화

#### 5. wordCloud 생성함수

- 생성한 파일을 월별로 워드클라우드 생성
- mask 이미지를 적용할 수 있도록 생성

## 6. 전체적인 동작

- 기사 크롤링 함수 동작
- 정부 지지율 크롤링 함수 동작
- main 함수로 키워드 추출
- 데이터 시각화

## 7. 데이터 시각화

- 정부 지지율 월 단위 그래프
- 정치/경제/사회/ 카테고리 총합 월 단위 워드클라우드
- 기사 top5 or top30 월 단위 워드클라우드

**감사합니다**