

# 프로젝트 기술서 : 사용언어 (Python) / 2020. 07.

<b>주요 내용</b>	<ul style="list-style-type: none"> <li>- 2017년 한국언론의 정당정파성 탐구</li> <li>&gt; 19대 대선 정책공약집 빈출단어로 정당 간 상관관계 파악</li> <li>&gt; 2017년 신문사 사설 자연어처리로 사설별 경향성(주제) 파악</li> </ul>
<b>처리 기법</b>	<ul style="list-style-type: none"> <li>- 자연어처리 (konlpy, word2vec, doc2vec, t-SNE, pyLDAvis)</li> <li>- 모델 테스트 (sklearn - LogisticRegression, MLPClassifier)</li> </ul>
<b>참고</b>	- rowena22323.github.io

**서론**

- 도입**
  - 19대 대통령 정책공약집은 각 정당별 정치이념을 반영하고 있으며, 신문사 사설은 해당 신문사가 그날의 중요 사건에 대해 신문사의 시각을 명백히 밝히는 컨텐트이다.
  - 정당과 언론의 관계성은 대통령의 정책수행방향과 신문사 사설의 논조에서 유사하게 드러날 것이다.
- 방법**
  - 자연어처리기법을 이용하여 정당별 정책공약집과 신문사별 사설의 관계 파악
- 기대**
  - 언론 연구분야 중 하나인 한국형 언론-정당 병행관계 탐구

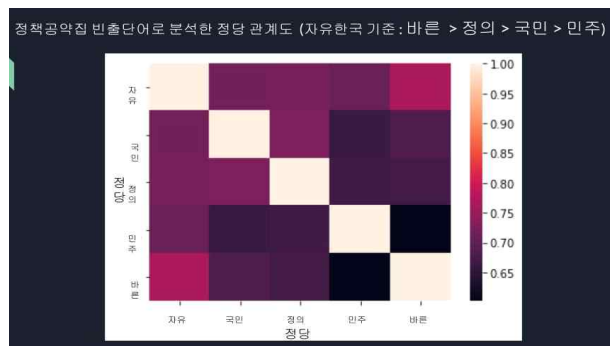
> 참고  
한국언론의 정당정파성 형성과정에 대한 연구(재미화, 2015) / 주요신문의 정권초기 대통령관련 보도태도 (백주연, 2014)

**19대 대선공약집 자연어처리**

- 19대 대선공약집 자연어처리 : 정당별 최다빈출단어 1000개 수집

- 1) 최다빈출단어로 정당간 관계성 분석
- 2) word2vec > 정당별 중점 공약 파악

> 참고 : 빈도수로 보는 정당 / 정책공약집 word2vec  
<https://programmers.co.kr/learn/courses/21/lessons/1698> word2vec - tSNE  
<https://wikidocs.net/book/2155> 딥러닝을 이용한 자연어처리



▲ 빈출단어로 분석한 정당 관계도 (자유한국당기준)  
자유 > 바른 > 정의 > 국민 > 민주

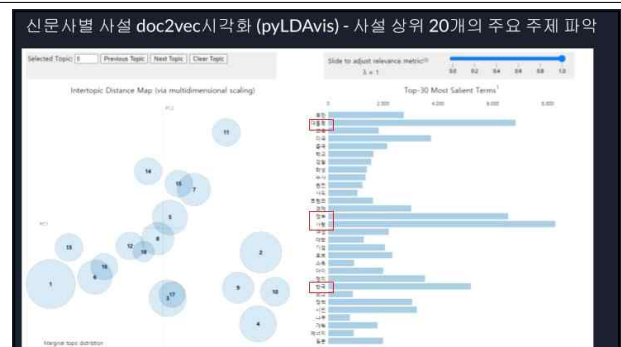


▲ 정당별 정책 공약집 word2vec시각화  
10대정책(좌)과 시각화 결과(우) 간 유사성 확인

**2017년 신문사별 사설 자연어처리**

- 토픽모델링 : 신문사 사설건별로 주제단어 추출
- 전체 코퍼스에서 30000 이상을 차지하는 신문사의 사설 20건에 대한 시각화
- Doc2Vec으로 신문사 - 신문사 코퍼스 간 경향성 파악

> 참고 : 신문사별 사설 토픽모델링 (경향, 한겨레, 한국경제, 한국일보, 매일신문, 서울신문)  
<https://sainteller98.tistory.com/167?category=731441> 토픽모델링  
[http://herodotus.com/blog/projects/2018/01/21/maver\\_movie\\_review\\_Doc2Vec](http://herodotus.com/blog/projects/2018/01/21/maver_movie_review_Doc2Vec)



▶ 신문사별 사설 doc2vec시각화 (경향/한겨레/동아/한국/매일/서울)  
- 신문사별 빈출 주제 20건에 대한 주제별 유사도 및 빈출단어 30개 분석

<b>결과</b>	<ul style="list-style-type: none"> <li>- 정당별 단어유사도 추출로 정당 정책 공약집 주제파악</li> <li>- 신문사 사설별 주제 분류</li> <li>- sklearn 다중레이블 분류 / 로지스틱 선형회귀 테스트 실시</li> </ul>
<b>이후 연구과제</b>	<ul style="list-style-type: none"> <li>- 신문사 사설과 정당 간 유사도 분석</li> <li>- 유사 경향을 보이는 정당과 신문사 탐구</li> </ul>