

(토픽모델링)윤석열 대통령 연설문에서 대통령의 핵심 가치 및 주제찾기

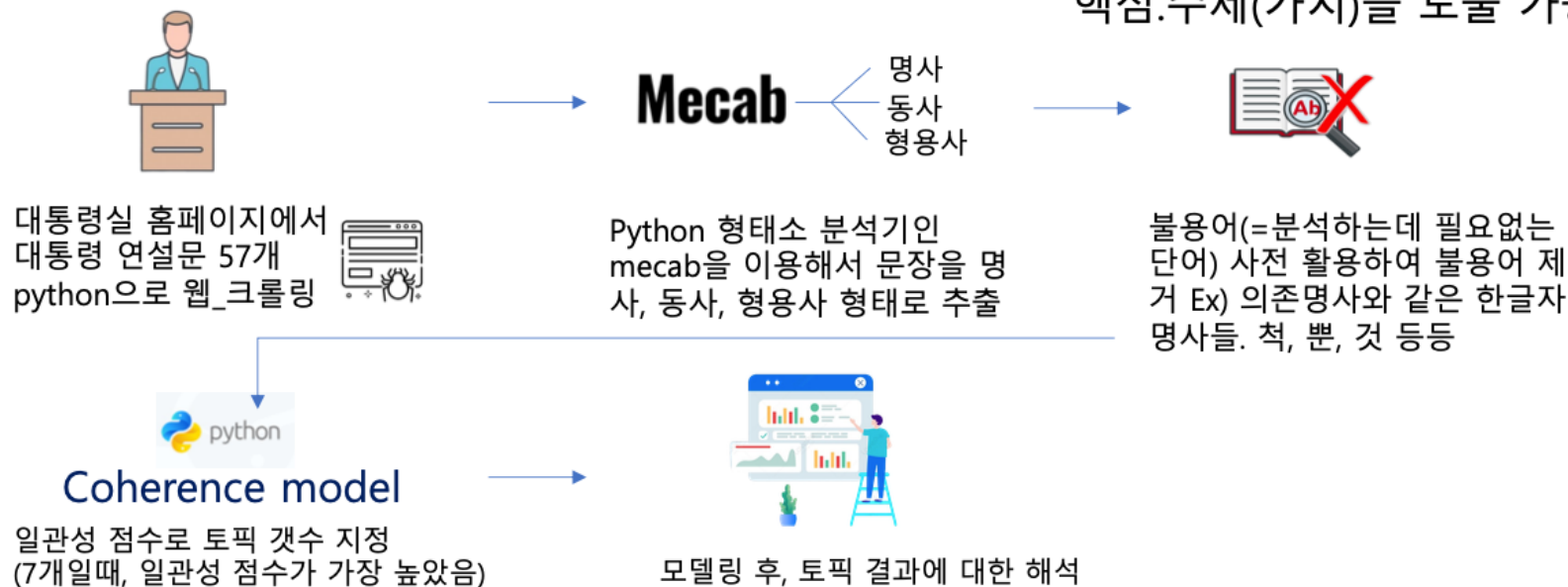
안동수



Lda 토픽모델링이란?

LDA(Latent Dirichlet Allocation) 토픽 모델링은 확률 기반의 모델링 기법을 통해 방대한 양의 문서 데이터를 분석함으로써 문서 내에 어떤 토픽이, 어떤 비율로 구성되어 있는지를 분석합니다.

→ 즉, 큰 방대한 텍스트 데이터에서 핵심토픽을 추출해줌. -> 57개 연설문에서 대통령이 말하고자 하는 핵심.주제(가치)들 도출 가능





대통령의 말과 글



2023.05.16
"3대 개혁은 더 이상 미룰 수도, 미뤄서도 안돼...비상한 각오로 임해야"
-제20회 국무회의 윤석열 대통령 모두 발언
제20회 국무회의를 시작하셨습니다. 정부 출범 1년 차 첫 국무회의입니다. 남다른 소회와 함께 새로운 각오를 다지겠습니다. 저는 지난 대선 당시 국민민주주의와 법치주의를 선포하며 새로운 국민의 나라를 만들겠다고 국민들께 약속드렸습니다. 지난 1년간 속 가빠져 달려왔습니다. 국민들께서 나라의 변화를 체감하실 수 있도록...

지체의 보기

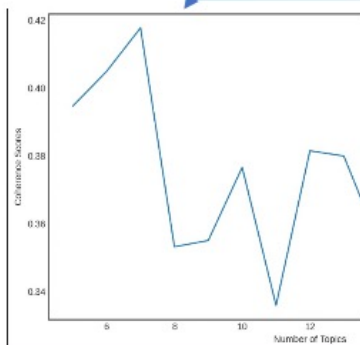
연설문57개에서 모든 문장들 수집

명사 : 개혁, 비상, 국무회의
동사 : 지키다, 만들다
형용사 : 어렵다

'척', '뿐', '때문', '저는'과 같은 단어 등 제거

문장 형태소 분석

토픽 모델링 분석에 유의미하지 않은 단어 제거



토픽 모델링 하기 위해 토픽 갯수 지정 필요
- 몇개의 주제(토픽)이 모델링 하는데 적합한지,
일관성 점수 Coherence model 라이브러리를
이용해 토픽 개수 7개 지정

LDA 토픽모델링



DATA MODELING

모델링 실행



토픽 모델링 결과 분석

Ⅲ 토픽 모델링(1)

토픽의 수 : 7개 선정

| 토픽 1 | |
|------|-------|
| 자유 | 1.226 |
| 평화 | 0.492 |
| 가치 | 0.427 |
| 시민 | 0.363 |
| 세계 | 0.306 |
| 보편 | 0.305 |
| 인권 | 0.255 |
| 연대 | 0.248 |
| 지키다 | 0.242 |
| 번영 | 0.225 |

자유_보편적 인권 가치

| 토픽 2 | |
|------|-------|
| 혁신 | 0.876 |
| 규제 | 0.786 |
| 민간 | 0.374 |
| 성장 | 0.315 |
| 도약 | 0.273 |
| 추진 | 0.183 |
| 과학 | 0.167 |
| 개선 | 0.160 |
| 필요 | 0.152 |
| 기업 | 0.150 |

규제혁신_민간성장
(기업 중심 도약 추진 필요)

| 토픽 3 | |
|------|-------|
| 산업 | 0.813 |
| 국가 | 0.482 |
| 기술 | 0.482 |
| 정부 | 0.301 |
| 우주 | 0.300 |
| 안보 | 0.292 |
| 미래 | 0.275 |
| 전략 | 0.273 |
| 육성 | 0.264 |
| 핵심 | 0.253 |

미래산업_전략육성

| 토픽 4 | |
|------|-------|
| 국민 | 2.238 |
| 지키다 | 0.354 |
| 나라 | 0.354 |
| 책임 | 0.330 |
| 안전 | 0.317 |
| 생명 | 0.303 |
| 만들다 | 0.256 |
| 정부 | 0.252 |
| 재산 | 0.241 |
| 서다 | 0.239 |

국민안전(생명 및 재산)_국가책임

| 토픽 5 | |
|------|-------|
| 피해 | 0.964 |
| 지원 | 0.543 |
| 복구 | 0.432 |
| 신속 | 0.407 |
| 입다 | 0.308 |
| 크다 | 0.273 |
| 집중호우 | 0.270 |
| 이번 | 0.250 |
| 일상 | 0.243 |
| 최선 | 0.227 |

집중호우_피해복구

| 토픽 6 | |
|------|-------|
| 반갑다 | 1.661 |
| 뵙다 | 0.142 |
| 위원 | 0.109 |
| 이렇다 | 0.093 |
| 존경 | 0.090 |
| 충남 | 0.075 |
| 도민 | 0.071 |
| 임원 | 0.043 |
| 연합회 | 0.042 |
| 선거 | 0.041 |

첫 인사

| 토픽 7 | |
|------|-------|
| 감사 | 2.237 |
| 진심 | 0.075 |
| 자리 | 0.071 |
| 현신 | 0.037 |
| 의료진 | 0.035 |
| 역사 | 0.029 |
| 빌리다 | 0.026 |
| 노력 | 0.023 |
| 일깨우다 | 0.021 |
| 교직원 | 0.020 |

감사 인사



57개 연설문 -> 토픽 7개로 압축
토픽별 구성하고 있는 단어 분포 모습 및 결과해석

감사합니다