

전체 문서 토픽별 단어 분포

Overall Coherence Score (TF-IDF): 0.413619858849582

file:///Users/kimsuyeon/Desktop/Project2/lda_visualization_tfidf.html#topic=6&lambda=1&term=

1) 주제 간 거리 그래프(왼쪽 그래프)

서로 다른 주제들이 어떻게 관계되어 있는지를 2차원으로 표현한 그래프로 각 주제는 원으로 나타나며, 원들 사이의 거리는 주제 간 유사성을 의미한다.

- 원들이 가까이 있는 경우: 해당 주제들이 서로 유사한 단어를 많이 공유한다.
- 원들이 멀리 있는 경우: 주제들이 서로 상당히 다르다.

각 원의 크기는 해당 주제가 문서에서 차지하는 비중을 의미한다.

- 큰 원: 문서에서 많이 다뤄진 주제
- 작은 원: 문서에서 적게 다뤄진 주제

2) 주제 단어에 대한 막대 그래프 (오른쪽 그래프)

왼쪽 그래프에서 주제를 클릭하면 오른쪽 그래프에 해당 주제를 정의하는 주요 단어들이 막대 그래프로 표시된다.

- 파란색 막대: 문서 전체에서 해당 단어가 얼마나 자주 등장하는지
- 빨간색 막대: 해당 주제 내에서 그 단어가 얼마나 중요한지

단어가 얼마나 중요한지는 λ 값에 따라 달라진다.

- 낮은 λ 값: 주제에 특화된 단어들 표시
- 높은 λ 값: 전체 코퍼스에서 빈도가 높은 단어들 표시

문서별 토픽 분포

위의 결과에서 키워드를 보고 각각의 토픽은 다음과 같다고 생각하였다.

Topic0 : 국방, Topic1: 외교, Topic2: 교육문화, Topic3:사회, Topic4: 경제, Topic5: 정치, Topic6: 과학기술

2023년도 말, 2023년도 초, 2022년도 초, 2020년도 초, 2019년도 초, 2017년도 초, 2013년도 초, 2012년도 초에는 7가지의 토픽이 동일하게 분포하였다. 2021년도 초에는 정치의 비중이 가장 높았고 그 외의 주제들은 동일한 분포를 가지고 있다. 2014년도, 2015년도, 2016년도, 2108년도 초에는 외교의 비중이 가장 높았고 그 외의 주제들은 동일한 분포를 가지고 있다. 2014년도 초에 비해 2015년도 초에 외교의 비중이 더 컸고 그 이후부터는

줄어들다가 2018년도 초에 급격히 증가하였다가 다시 감소하였다.