

사례연구#3 텍스트 데이터 분석

A3팀(문태웅(팀장), 남원식, 이순규)

B3팀(유준(팀장), 문현진, 임성현)

C3팀(오준서(팀장), 천성한, 한호중, 황윤수)

팀별로 다음 과제를 수행하고 결과와 R code를 포함한 사례연구보고서를 제출하시오.

제공된 데이터는 1863년 11월 19일 게티스버그에서의 링컨 대통령 연설문의 국문 번역문이다.

1. 제공된 데이터를 이용하여 토픽 분석을 실시하여 단어구름으로 시각화 하고 단어 출현 빈도수를 기반으로 어떤 단어들이 주요 단어인지 설명하시오
2. 제공된 데이터를 이용하여 연관어 분석을 실시하여 연관어를 시각화 하고 시각화 결과에 대해 설명하시오
3. 다음 포털사이트의 실시간 뉴스(<https://news.daum.net/>)를 수집하고 실시간 토픽분석을 실행하여 단어구름으로 시각화 하고, 분석 시점에서 주요 이슈가 무엇인지 설명하시오.

공동수행사항:

- 1) R code에 #로 comment를 추가하여 단계별로 실행한 모듈 설명을 추가 하시오.
- 2) 본인의 컴퓨터에 관련 패키지가 설치되어 있더라도 필요한 패키지의 설치 명령어도 포함하고 comment 처리(#)하시오
- 3) 자율적으로 선택한 parameter값(예, 최소 단어 출현 빈도수, 지지도, 신뢰도 등)을 명시하고 선택한 이유에 대해 설명하시오.
- 4) 단어구름 형성 시 적용할 최소 단어 출현 빈도수는 적절한 수준(최소 2이상으로 A4 한 페이지 안에 들어가고 단어들이 구분될 정도의 수준)으로 팀에서 결정)

제출기한: 2021년 11월 26일(금) 09:30

제출처: 카페 내 과제제출 게시판

제출물: 사례연구보고서, R코드

제출양식: 자유(워드, 한글 등)

발표는 없음