

6. R

6.1 텍스트 마이닝(Text mining)

- 문자로 이루어진 데이터에서 가치 있는 정보를 얻어내는 분석 기법.
- SNS나 웹사이트에 올라온 글들을 분석하여 이야기 주제 및 관심도를 파악하기 위해 활용
- 형태소 분석(Morphology Analysis) : 문장을 구성하는 어절들이 어떤 품사로 되어 있는지 분석

6.2 분석절차

1) 형태소 분석

- 데이터를 형태소에 따라 추출한다.

2) 명사, 동사, 형용사 등의 의미가 있는 단어들을 추출

- 무의미한 단어들(특수문자, 의미 없는 숫자 등)을 전처리를 통해 걸러낸다.

3) 빈도표 만들기

- 추출된 형태소 단어들 중 사용된 빈도를 확인하여 순위를 결정한다.

4) 시각화

- 순위가 결정된 단어들을 확인하기 쉽게 그래프 형태 및 워드 클라우드 형태로도 표현이 가능 하도록 데이터를 가공한다.