

Part. 01

실전 데이터 분석 프로젝트

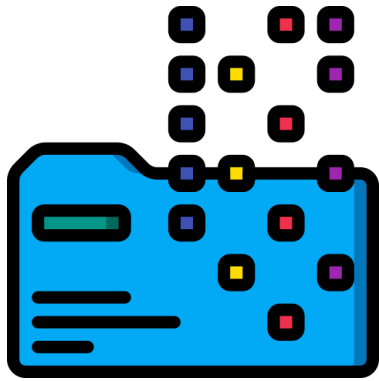
| 텍스트마이닝을 활용한 카카오톡 토픽분석

FASTCAMPUS
ONLINE

머신러닝과 데이터분석A-Z

강사. 김용담

I 텍스트 마이닝 과정



텍스트 수집



텍스트 전처리



텍스트 분석



텍스트 시각화

I 텍스트 시각화



- 텍스트 분석 결과를 도표나 그래프로 시각화 하는 과정
- 기존 데이터 분석과의 차이점은 **텍스트**로 시각화가 가능하다는 점

예) wordcloud, semantic network analysis

I 대표 예시 1. 워드클라우드(WordCloud)



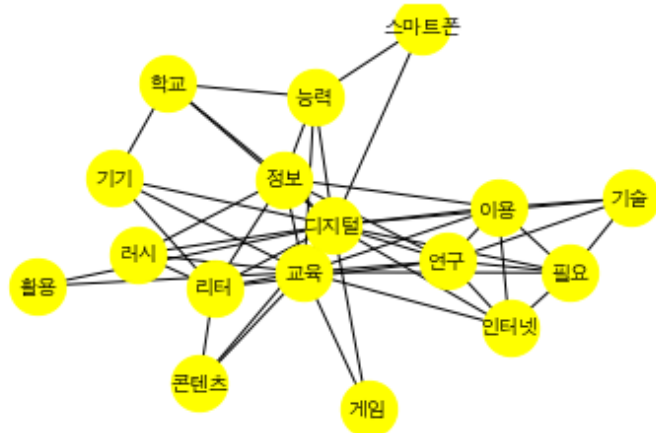
Made In Wordcloud.kr

워드클라우드 예시

- 텍스트에 있는 키워드들을 다양한 크기와 색깔로 표현해주는 기법
- 중요도(보통은 빈도) 별로 색깔과 크기가 다름
- 생성기
 - <http://wordcloud.kr/>
 - <https://www.wordclouds.com/>
 - <https://pypi.org/project/wordcloud/>

Reference : <http://wordcloud.kr/56>

I 대표 예시 2. 의미 연결망 분석(Semantic Network Analysis)



네이버 뉴스 기사 연결망 분석 결과

- 그래프의 노드에 단어를, 노드 간 연결에 단어 사이의 관계를 정의하여 연결
- 문서에서 등장하는 키워드 간의 관계를 분석하기 위해 사용
- 노드에 단어가 아닌 문장, 문서도 가능하고 노드 간 연결을 단어들의 동시 등장 횟수나, 문서 벡터 사이의 거리로 정의하는 응용이 가능

Part. 01

실전 데이터 분석 프로젝트

I 감사합니다.

FASTCAMPUS
ONLINE

머신러닝과 데이터분석A-ZI

강사. 김용담