

텍스트 마이닝과 데이터 마이닝

Part 05. 토픽 모델링과 워드 클라우드

정 정 민

Chapter 11. 토픽 모델링

1. 토픽 모델링이란
2. 토픽 모델링 사례

토픽 모델링이란

토픽 모델링 (Topic Modeling)

- 토픽(Topic) : 문서 집합 안에서 논의되는 주제나 개념을 의미
- 모델링(Modeling) : 통계적인 방법으로 데이터의 패턴을 추출하는 과정을 의미
- 따라서 토픽 모델링이란,
 - 대규모 텍스트 데이터에 존재하는 다양한 주제를 자동으로 식별하고 분류하는 과정
 - 이를 통해 데이터 안에 숨겨진 주제 구조를 발견하고 이해하는 목적
- 텍스트 마이닝, 자연어 처리, 정보 검색 등 다양한 분야에서 활용



토픽 모델링의 주제

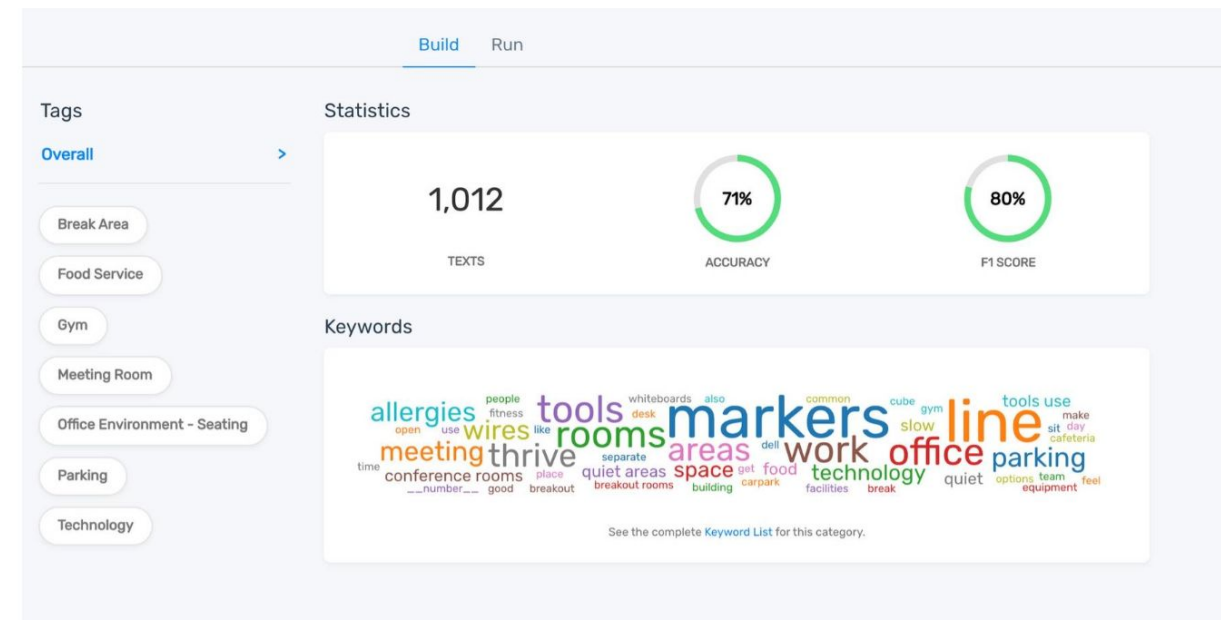
- 토픽 모델링의 주제는 일반적으로 단어의 집합으로 표현
- 텍스트 내의 특정 패턴이나 빈도를 기반으로 선택
- 데이터 소스에 따라 주제의 범위가 결정
 - 뉴스 기사, 소셜 미디어, 학술 논문 등
- 또한, 하나의 문서에 다수의 주제를 포함
- 데이터의 트렌드와 특정 이벤트를 반영하는 경향
- 선택된 주제는 다른 텍스트 분석에 긍정적인 도움이 되는 방향으로 사용될 수 있음
 - 특정 주제와 비슷한 텍스트만 추려내거나
 - 분석 결과를 해석하는 과정에서 토픽과 관련된 부분만을 취하거나
 - 등등

토픽 모델링 사례

직원 복지를 위해 설문 결과를 분석 - Dell 사례

- Dell 회사는 많은 직원들의 고민과 생각을 알고 싶어함
- 이를 위해 직원 만족도 설문 조사를 분석
- 약 10,000개가 넘는 텍스트 설문 조사 과정을 진행
- 과정에서 토픽 모델링을 활용
- **직원들의 의견을 빠르고 직관적으로 파악** 가능
- 운영 효율 증가

predetermined categories: *Break Area, Food Service, Gym, Meeting Room, etc.*:



검색어 트렌드 - NAVER

- 국내 검색 엔진 회사인 네이버에서 제공하는 인기 검색어
- 사람들의 관심도와 전반적인 트렌드를 확인할 수 있음
- 이를 통해 **사업 인사이트, 마케팅 아이디어** 등을 얻을 수 있음
 - 겨울이 다가오니 털신발을 팔아보자!
 - 여름이 다가오니 기능성 운동복이 좋겠다.
 - 요즘은 “도파민”이라는 단어를 많이 사용하니, 우리 제품에도 그런 단어를 녹여봐야지!
 - 등등

분야별 인기 검색어		인기분야	
패션의류		월간	
2023.09.	2023.10.	2023.11.	2023.12.
2023.09.01. ~ 2023.09.30.	2023.10.01. ~ 2023.10.31.	2023.11.01. ~ 2023.11.30.	2023.12.01. ~ 2023.12.31.
1 원피스	1 원피스	1 숏패딩	1 롱패딩
2 트위드자켓	2 트위드자켓	2 패딩	2 여성패딩
3 써스데이아일랜드원피스	3 맨투맨	3 롱패딩	3 숏패딩
4 가을원피스	4 경량패딩	4 여성패딩	4 여성롱패딩
5 가디건	5 후드티	5 여성롱패딩	5 패딩
6 블라우스	6 패딩조끼	6 패딩조끼	6 코트
7 후드집업	7 가디건	7 경량패딩	7 원피스
8 나이키바람막이	8 후드집업	8 코트	8 몽클레어패딩
9 지고트원피스	9 니트원피스	9 원피스	9 코듀로이팬츠
10 케네스레이디원피스	10 트랜치코트	10 코듀로이팬츠	10 남자패딩

E.O.D