

캡스톤 디자인(1) 계획서

1. 프로젝트 제목

- 자연어 처리를 활용한 한국어 유튜브 댓글 종류 분류

2. 프로젝트 구성원

-

3. 프로젝트 개요

SNS, 개인 블로그, 동영상 채널 등 다양한 방식의 콘텐츠 생산이 보편화된 오늘날, 방대한 양의 데이터가 인터넷에 저장되고 있습니다. 이를 마케팅 등 원하는 분야에서 적절히 사용하기 위해서는 필요한 데이터를 수집하고 분석하여 원하는 정보를 얻어내는 능력이 필요합니다.

본 프로젝트에서는 Youtube 사이트에 작성된 한국어 댓글들을 수집하고 목적에 맞는 다양한 기준으로 분류해 보는 것이 목표입니다.

4. 진행 계획

1) 적용 기술

- Python 사용 예정, github로 소스코드 관리
- 웹 크롤링 : BeautifulSoup , Selenium, Google Youtube api
- 전처리 : KoNLPy, kss(Korean Sentence Splitter)
- 모델 학습 : scikit-learn, PyTorch

2) 진행 방법 & 일정 계획

- 3월 : 웹 크롤링 공부 및 데이터 수집
- 4월 : 분류 모델 설계, 데이터 전처리
- 5월 : 모델 학습, 분류 , 결과 개선.