

프로젝트 제안서

영화 캐릭터 분석 서비스 4팀

타이틀

1. 프로젝트 주제
“영화 대본 파일 분석을 통해 영화 제작자를 위한 요약본 제공 서비스”
2. 엔드유저에게 보이는 웹서비스 타이틀
TOY SCRIPT
3. 팀 구성원의 전체 이름과 역할
 - 김유나 : PM, Frontend
 - 김수연 : 서기, Frontend
 - 김수람 : Backend
 - 노하람 : Backend
 - 이보람 : 코드 총괄, Backend

데이터세트

1. 사용할 데이터셋과 수집 계획
 - Toy Story 3 (2010) PDF 대본 분석
 - 이 외에 <https://imsdb.com/scripts> 과 같은 사이트 이용 및 검색 수집 예정
2. 데이터 소스에 대한 링크, 분석을 위해 데이터를 준비하는데 예상되는 어려움
 - 대본의 형식 불일치로 인해 동일한 전처리가 어려울 수 있음
 - 행동 묘사를 캐릭터의 성격이나 태도로 정보 조합하는 부분
 - 분석결과로 나온 파이썬 데이터를 웹에서 사용자에게 PDF 데이터로 변환하는 부분

기획 방법

1. 문제 정의와 가설 설정 방법
 - 문제 정의: 많은 작가로부터 대본을 받는 영화 제작자는 모든 작품을 읽고 캐릭터를 이해하기에 시간적 어려움이 있다.
 - 가설 설정:

- 1) 대본 내 장소별 중요성과 장소별 등장인물 빈출도를 시각화, 등장인물 성격과 관계도를 시각화하면 대본을 자세히 읽지 않아도 대략적인 영화의 정보를 파악할 수 있다.
- 2) 밤에 자주 나타나는 캐릭터의 대사를 감정 분석했을 때, 부정적인 단어가 더 많이 나올 것이다. (기본 분석 데이터로 설정한 토이스토리 대본 한정)
- 3) 출연 빈도가 높은 인물 일수록, 많은 장소에 나올 것이다.

2. 엘리스에서 배운 데이터 분석 개념, 알고리즘 및 라이브러리 응용 방법

0. 분석 툴 : jupyter notebook

1. 시각화 :

- 그래프 : Matplotlib, Seaborn
- 이미지 : PIL

2. 데이터 처리 : Numpy, Pandas

- PDF 파일 변환 : PyPDF2, pdfminer
- 자연어처리(영어) : spaCy or nltk
- 워드 임베딩 등이 필요할 경우는 차후에 생각

3. 데이터 분석 종류 (시각화, 통계 대시보드)

- 캐릭터 수 : 통계
- 캐릭터 성격 : 통계 대시보드
- 캐릭터 관계 : 시각화, word association
- 장소별 자주 등장하는 캐릭터 : 통계 대시보드, 시각화
- 시간별 자주 등장하는 캐릭터 : 통계 대시보드, 시각화
- 캐릭터 및 배경 소개 서비스 : 시각화

서비스 설명

1. 데이터분석 웹 서비스의 최종적인 기능과 형식

- 사용자가 영화명을 입력하여 'GO' 버튼을 누르면, 해당 영화를 분석한 캐릭터별 성격과 관계구도와 같은 대본 요약 결과를 웹 상에 표시하고 PDF 파일 다운로드 기능을 지원한다.

2. 웹 서비스의 사용자가 데이터 분석 시각화 자료를 통해 얻는 인사이트

- 총 등장 인물의 수, 가장 빈도가 높게 나타나는 장소
- 등장 인물의 성격 및 관계도
- 장소별 등장인물 빈출도

1. 메인 페이지에서 영화 이름을 검색한다.
 - 검색 결과가 있는 경우, 2번으로 넘어간다.
 - 검색 결과가 없는 경우, 화면에 에러 메시지를 띄운다.
2. 대본을 분석한 전체 결과를 화면에 띄운다.
 - 캐릭터 소개, 장소, 시간대 탭 메뉴가 나타난다.
 - 캐릭터 소개 탭이 첫번째로 보여진다.
 - 다른 탭을 클릭하면 각각의 세부 결과가 보이게 한다.(또는 이동)
3. 분석 전체 결과를 웹에 보여주고, PDF 파일로 다운로드 할 수 있다.