

요구사항 분석서

[트위터 데이터를 활용한 감정분석 프로그램]

Capstone 디자인(01)

담당 교수 : 안용학, 양효식, 박기호

팀장 : 16010945 노수민 컴퓨터공학

팀원 : 16010573 이아현 컴퓨터공학

16011041 홍주희 컴퓨터공학

16011189 양승주 디지털콘텐츠

목 차

1. 시스템 개요	3
2. 기존 시스템	3
3. 요구사항 목록	4
3.1 기능별 요구사항	4
4. 요구사항 분석	7
4.1 요구사항 목록	7
4.2 요구사항 세부 명세	8
5. 기능 외 요구사항	12
6. 제약사항	12
7. 소프트웨어 개발 환경	13

1. 시스템 개요

- 트위터 API를 통해 유명인의 트윗 데이터를 수집하고, 게시글 속 문장의 감정을 분석하여 긍정, 중립, 부정 3가지의 감정으로 분류해 시각화해 보여주는 프로그램을 웹 애플리케이션으로 제작한다.
- 트위터 API 라이선스 문제로 인해 9월부터 11월까지의 데이터 수집을 통해 진행한다.
- 기존 연산 방식을 NZB 인덱싱으로 변형하여 모델의 용량과 연산량을 감소시키고, NZB 인덱싱이 적용된 모델을 이용하여 데이터 분석을 한다..

2. 기존 시스템

시중에 이미 존재하는 유사한 기존 시스템의 분석 내용 혹은 기능적인 면을 중점으로 분석하여 다음과 같은 애플리케이션을 찾았다.

A. Watson Tone Analyzer

- ① 트위터, 온라인 리뷰, 이메일, 텍스트를 7 가지의 톤으로 분석
- ② 영어, 프랑스어 2 가지 언어를 제공

B. DeepMoji

- ① 영어 문장의 감정을 분석해 여러 이모지로 표현

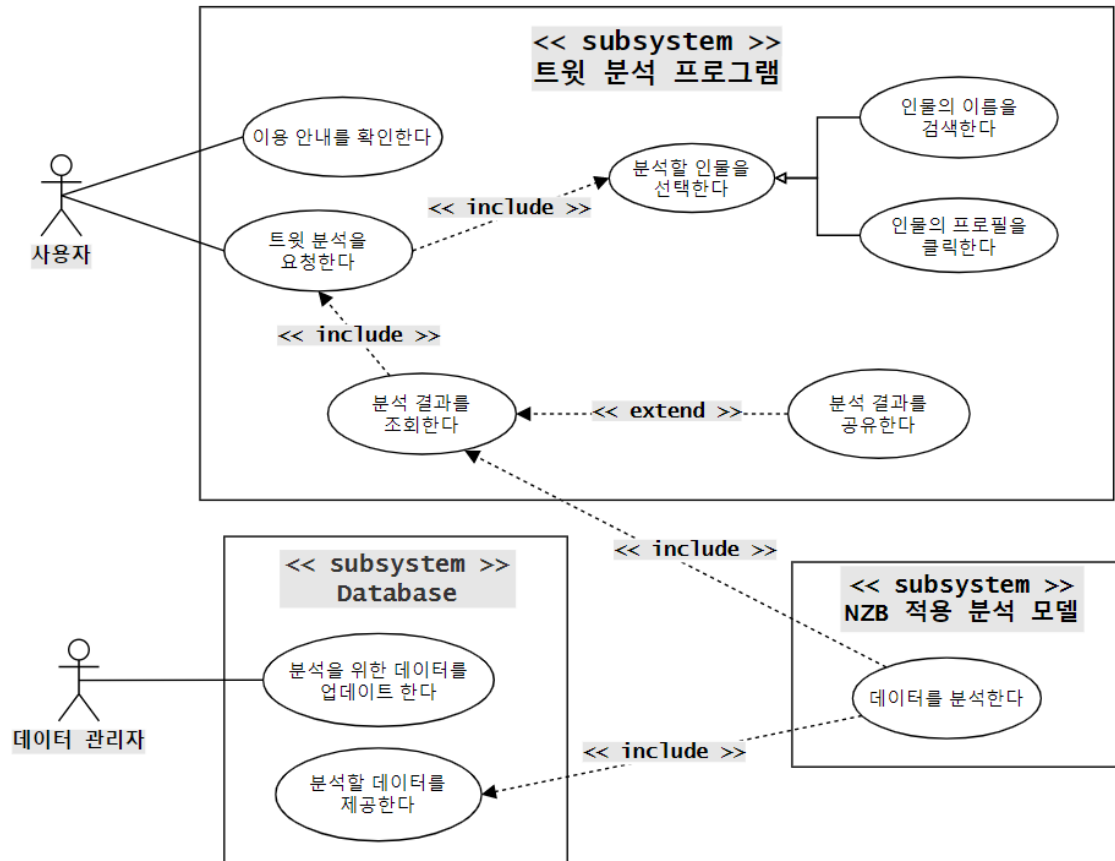
Watson Tone Analyzer와의 차이점

- 하나의 게시글을 가지고 분석을 진행하지만, 본 프로젝트에서 만들 웹 애플리케이션은 여러 게시글을 분석하여 긍정, 중립, 부정 문장으로 나누어 통계를 낸다.

DeepMoji와의 차이점

- 한 문장을 대상으로 여러가지 감정으로 표현하지만, 본 프로젝트에서 만들 웹 애플리케이션은 여러 문장을 분석하여 긍정, 중립, 부정 문장으로 나누어 통계를 낸다.

3. 요구사항 분석



요구사항 분석서

트위터를 활용한 감정분석 프로그램

작성일 2020-10-03

3.1. 기능별 요구사항

A. 트윗 분석 시스템

요구사항	유명인 검색			
	ID	R001	우선순위	상
요구사항 내역	<ul style="list-style-type: none"> - 사용자는 이름 검색을 통해 분석 대상을 선택할 수 있다. - 검색어 자동완성 기능을 제공한다. 			

요구사항	DB내 유명인 목록 표시			
	ID	R002	우선순위	상
요구사항 내역	<ul style="list-style-type: none"> - 검색 창 하단에 현재 분석 가능한 유명인들의 프로필을 한다. 			

요구사항	분석 결과 조회			
	ID	R003	우선순위	상
요구사항 내역	<ul style="list-style-type: none"> - 사용자가 분석하고자 하는 인물의 프로필을 클릭 시, 저장된 트윗 데이터를 통해 분석 후 결과를 시각화 해준다. 			

요구사항	이용 안내 조회			
	ID	R004	우선순위	중
요구사항 내역	<ul style="list-style-type: none"> - 프로그램의 서비스 설명(분석 방법)을 안내해준다. - 프로그램의 개발자 소개 및 개발 의도를 제공한다. 			

요구사항	분석 결과 공유			
	ID	R005	우선순위	하
요구사항 내역	<ul style="list-style-type: none"> - 유명인의 언어습관 분석 결과를 타 사이트에 공유 가능하다. 			

요구사항 분석서

트위터를 활용한 감정분석 프로그램

작성일 2020-10-03

B. NZB 적용 분석 모델

요구사항	데이터 분석			
	ID	R006	우선순위	상
요구사항 내역	<ul style="list-style-type: none"> - 미리 훈련된 모델이 사용자가 원하는 유명인의 트윗 데이터를 분석한다. 			

C. 데이터베이스

요구사항	데이터 저장			
	ID	R007	우선순위	상
요구사항 내역	<ul style="list-style-type: none"> - 데이터베이스에 분석에 사용할 트윗 데이터를 txt 파일 형식으로 저장한다. - 분석 인물 이름, 트위터 계정, 트윗, 해당 트윗 날짜를 저장한다. - 1주일마다 트윗 데이터를 업데이트 한다. 			

요구사항 분석서

트위터를 활용한 감정분석 프로그램

작성일 2020-10-03

4. 요구사항 분석

4.1. 요구사항 목록

번호	분류	요구사항명	상세내용	난이도
R001	트윗 분석 시스템	유명인 검색	찾고자 하는 유명인의 이름을 검색	상
R002	트윗 분석 시스템	DB내 유명인 목록 표시	DB에 존재하는 유명인을 화면에 표시	상
R003	트윗 분석 시스템	분석 결과 조회	사용자가 선택한 유명인의 트윗 데이터 분석 조회	상
R004	트윗 분석 시스템	이용 안내 조회	웹사이트 이용 방법을 설명	중
R005	트윗 분석 시스템	분석 결과 공유	타 사이트에 분석 결과를 공유	하
R006	NZB 적용 분석 모델	데이터 분석	유명인의 트윗 데이터를 분석	상
R007	데이터베이스	데이터 저장	데이터베이스에 분석에 사용할 트윗 데이터를 저장	상

요구사항 분석서

트위터를 활용한 감정분석 프로그램

작성일 2020-10-03

4.2. 요구사항 세부 명세

● R001

요구사항명	유명인 검색	요구사항구분	트윗 분석 시스템
요구사항설명	사용자가 프로그램에서 분석하고 싶은 유명인을 검색한다.		
해결방안	검색창에 유명인의 이름을 입력할 경우, 자동완성 기능으로 검색을 돕는다.		
위험요소	DB에 존재하지 않는 경우, 분석이 불가능하다고 안내한다.		
설계 시 고려사항	유명인의 이름, 예명을 정확하게 사용해야 한다.		
관련요구사항	DB내 유명인 목록 표시		
시나리오	1. 사용자가 웹페이지 중앙의 검색창에 유명인의 이름을 입력한다. 2. 해당 유명인이 존재할 경우 자동완성 기능으로 입력을 돕는다.		

● R002

요구사항명	DB내 유명인 목록 표시	요구사항구분	트윗 분석 시스템
요구사항설명	사용자가 분석 가능한 유명인들을 웹에서 확인한다.		
해결방안	분석 가능한 유명인들의 프로필을 표시하고, 클릭하면 분석이 가능하다. 프로필에는 사진, 이름, 트위터 아이디를 표기한다.		
위험요소	사용자가 프로필을 실수로 클릭할 경우를 위해, 선택된 인물의 분석을 진행할지 확인하는 안내창을 보인다.		
설계 시 고려사항	50명 이상의 프로필을 화면에 나타낼 경우, 가독성에 지장이 생길 수 있다.		
관련요구사항	분석 결과 조회		
시나리오	1. 사용자가 웹페이지 하단에 존재하는 분석 가능한 유명인의 프로필을 확인한다. 2. 분석하고 싶은 유명인의 프로필을 클릭한다. 3. 해당 유명인의 트윗 분석을 진행할지 확인하는 안내창이 뜬다.		

요구사항 분석서

트위터를 활용한 감정분석 프로그램

작성일 2020-10-03

● R003

요구사항명	분석 결과 조회	요구사항구분	트윗 분석 시스템
요구사항설명	사용자가 확인하고 싶은 유명인의 트윗 데이터 분석 결과를 화면에 표시한다.		
해결방안	데이터 분석 결과를 통계 제출 후, 시각화를 통해 화면에 표시한다. 1. 긍정 / 중립 / 부정 문장의 비율 제공 2. 자주 사용하는 단어 제공		
위험요소	사용자가 긍정 / 중립 / 부정 의 판단 기준을 궁금해 할 수 있다.		
설계 시 고려사항	긍정 / 중립 / 부정 의 기준을 잘 설정해야 한다.		
관련요구사항	DB내 유명인 목록 표시		
시나리오	1. 사용자가 분석을 요청해, 유명인의 트윗 데이터 분석을 한다. 2. 분석 결과를 화면에 표시한다. 3. 사용자는 유명인의 트윗 게시물 내 문장의 감정 비율과 자주 사용하는 단어를 확인한다.		

● R004

요구사항명	이용 안내 조회	요구사항구분	트윗 분석 시스템
요구사항설명	웹 애플리케이션의 이용 안내를 확인한다.		
해결방안	웹 상단에 about 카테고리를 통해 이용 안내를 제공한다.		
설계 시 고려사항	웹의 어느 페이지에서도 접근 가능하도록 한다. 사용자가 사용하기 쉬운 UI로 구성한다.		
시나리오	1. 사용자가 상단의 about 버튼을 클릭한다. 2. 사용자는 개발 의도, 서비스 설명(분석 방법), 개발자 소개를 확인한다.		

요구사항 분석서

트위터를 활용한 감정분석 프로그램

작성일 2020-10-03

● R005

요구사항명	분석 결과 공유	요구사항구분	트윗 분석 시스템
요구사항설명	유명인의 트윗 분석 결과를 외부로 공유한다.		
해결방안	분석 결과를 조회할 때, 공유하기 버튼을 추가한다. 페이스북, 트위터, 카카오톡 공유를 가능하게 한다.		
위험요소	공유 가능한 범위를 지정한다.		
설계 시 고려사항	많은 사람들이 사용하는 사이트의 공유 기능은 필수로 제공한다.		
관련요구사항	분석 결과 조회		
시나리오	1. 사용자가 트윗 분석 결과를 조회한다. 2. 사용자가 웹 외부로 분석 결과를 공유한다.		

● R006

요구사항명	데이터 분석	요구사항구분	NZB 적용 분석 모델
요구사항설명	사용자가 분석하고 싶은 유명인의 트윗 데이터를 분석한다.		
해결방안	DB에 저장된 유명인이 작성한 트윗을 대상으로 분석을 시작한다.		
관련요구사항	데이터 저장		
시나리오	1. DB에서 유명인의 트윗 데이터를 불러온다. 2. 분석 모델이 데이터를 분석하여 문장의 긍정, 중립, 부정의 통계를 낸다. 3. 자주 사용하는 단어의 통계를 낸다.		

요구사항 분석서

트위터를 활용한 감정분석 프로그램

작성일 2020-10-03

● R007

요구사항명	데이터 저장	요구사항구분	데이터베이스
요구사항설명	분석을 위해 유명인의 트윗 데이터가 필요하다.		
해결방안	트위터 API 라이선스 문제로 인해 1주일마다 데이터 관리자가 데이터를 수집한 후, DB에 저장한다.		
위험요소	중복되는 데이터가 생길 수 있어서, 확인이 필요하다.		
설계 시 고려사항	가져온 트윗 게시글의 데이터 업로드 날짜를 구분하여 저장한다.		
관련요구사항	데이터 분석		
시나리오	1. 사용자가 분석하고 싶은 유명인을 선택한다. 2. 해당 유명인의 언어습관 파악을 위해 분석할 트윗 데이터가 필요하다. 3. DB에서 미리 저장해둔 데이터를 불러온다.		

요구사항 분석서

트위터를 활용한 감정분석 프로그램

작성일 2020-10-03

5. 기능 외 요구사항

번호	요구사항 내용
R01	희소행렬 처리 방식인 NZB 인덱싱을 적용한 모델을 개발해야 한다.
R02	기존 모델과 NZB 인덱싱을 적용한 모델의 성능 (연산 속도, 메모리 사용량, 정확성)을 측정하고 비교한다.
R03	유명인의 트윗을 1주일마다 수집과 전처리를 통해 DB 내용을 업데이트 한다.
R04	실시간 처리를 통해 신속하고 정확하게 화면에 분석 결과 출력한다. -> 분석 시간 최장 5초 이내 처리한다. -> 10초 이상 지연 시 오류 메시지 사용자에게 제시한다.
R05	PC/모바일에 따라 별도의 화면 제공 브라우저(Chrome, Internet Explorer, Microsoft Edge)에 제한받지 않는다.

6. 제약 사항

번호	제약 조건 내용
C001	트위터 API 라이선스 문제로 실시간 게시글 분석이 불가능하다.
C002	DB에 저장된 약 50명 정도의 유명인만 트윗 분석이 가능하다.
C003	모델 훈련 시 영어 데이터를 사용할 예정으로, 영어 외엔 다른 언어의 분석은 불가능하다.

요구사항 분석서	트위터를 활용한 감정분석 프로그램
	작성일 2020-10-03

7. 소프트웨어 개발 환경

소프트웨어 환경		하드웨어 환경		네트워크 환경	
Language	Javascript, python, C++	CPU	Intel core i5	Cloud computing service	AWS
DBMS	Mongodb	RAM	8GB	VM	EC2
WAS	Node.js	O/S	Windows10	AMI	Ubuntu Server
Editor	Jupyter notebook, VS Code			Server region	Seoul
Tech. Used	ejs				