

캡스톤 회의록 14

2020-09-27-일~2020-09-28-월

16010573 이아현 컴퓨터공학

(팀장) 16010945 노수민 컴퓨터공학

16011041 홍주희 컴퓨터공학

16011189 양승주 디지털콘텐츠

I. 회의 주제

1. 주제 변경 의견 - 0927
2. 주제 유지 및 데이터 관련 변동사항- 0928

II. 회의 내용

1. 주제 변경 의견 - 0927

A. Twitter 외 웹 크롤링 후 감정 분석

- ① 웹 크롤러 관련 안정성 없다 생각되어 제외

B. 문서 요약

- ① 두 개의 문서를 요약 후 비교

- 인풋 : 사용자가 텍스트 입력 or 파일 2개 업로드
- 아웃풋 : 문서 별 요약문, 문서 별 키워드(공통 키워드, 문서 키워드), 문서 내 단어 연관관계 가시화 등
- 문서 요약 알고리즘 선택 필요
 - TextRank 등
- NZB 적용 가능 여부 확인 필요
- 기능 구체화(차별성 위함) 필요

- 제안서 PPT 재 작성
- 안용학 교수님과의 미팅 전

C. 안용학 교수님과의 미팅(9/28 PM 3:00) 전 Word2Vec, TextRank 등 공부

2. 주제 유지 및 데이터 관련 변동사항 - 0928

A. 안용학 교수님과 미팅, 9/28(월) PM 3:00

- ① 감정분석->문서요약으로의 주제 변경 추천하지 않음
- ② 실시간 분석이 아닌 서비스에서 제공된 인물만 분석해도 괜찮음
 - Twitter API 로 인한 서비스 제약 조건 정확히 명시할 것
 - 팀원 한 명당 n 명씩 맡아 수동 수집 후 누적 데이터로 분석하는 것 추천
 - 미국 대통령 등 민감한 인물 데이터 사용 안 할 것 추천
- ③ 주제 유지 및 데이터 관련 변경 양효식 교수님께 확인 받을 것
- ④ 주제 변경 없더라도 세부 변경 사항 수정한 제안서 제출 필요

B. 변동 사항 정리

- ① 주제 : 유명인 Tweet 을 통한 언어습관 분석 그대로 유지
 - 단, Twitter API 무료 버전 사용으로 인한 서비스 사용 제약 조건 분명히 명시
 - Twitter 유료화 버전 진행 시 프로그램 기능 확장 가능
- ② 모델 : Word2Vec 그대로 사용
 - 현재 구현 예정인 기능에는 오버 스펙
 - 그러나 추후 기능 확장 가능성 있도록 유지
 - 팀원 모두 모델에 대한 정확한 숙지 필요
- ③ 데이터 수집 - 매일 수동 수집
 - 수집 인원 수 : 100 명 (최근 활동이 있는 계정으로 선별)

- 수집 기간 : 09/29(화) 시작, 11/24(화) 종료
- 수집 담당 : 주희, 수민
- 수집 규칙 : 1 주일에 한 번 수집
 - 파일명 통일
 - 기간 별, 계정 별 rawData.txt 파일 별도 저장
 - 데이터 github 저장 여부는 개인정보 문제로 반려

④ DB 사용 : MongoDB

- 저장할 내용 : Twitter 사용자 계정 정보, 해당 사용자의 tweet 을 저장한 txt 파일
- 용량 증대로 인해 사용료 나올 경우, 팀원 분할 계산

⑤ 인풋 : User 가 입력한 유명인 Twitter 계정

⑥ 아웃풋

- 5 가지 유형 분류 제거
- 긍정적 단어 비율, 부정적 단어 비율 그래프 표현
- Tweet 내 단어를 기반으로 분석한 결과라는 점 안내 표시
- 추가적인 시각화 필요(아웃풋 부족)
- 자주 사용하는 단어 태그 표시(최대 5 개)
- Tweet 출처 계정 안내
- Twitter API 무료 버전 사용으로 인한 검색 인물 제한적임 안내
- 분석에 사용된 Tweet 이 작성된 날짜 안내

⑦ 역할 분담, 개발 일정

- 팀장이 수정 후 공지 예정

C. 웹 Input & Output 이미지 제작 후 피드백 – 승주

- 세부 웹 UI 진행 예정

D. 요구사항 분석서 초안 진행 - 승주

E. 유스케이스 초안 진행 - 아현

F. 데이터 수집 - 수민, 주희

- ① 이전 수집한 유명인 300 명 계정 목록서 정치인물들 삭제
- ② 유명인 250 명 계정의 활동 여부 확인
- ③ 주기적 활동하는 계정으로 각각 50 명씩 담당
- ④ 총 100 명의 유명인 Twitter 계정서 매주 데이터 수집 예정

III. 공지 사항

- 1. 양효식 교수님께 변동사항 확인 받기, 9/28(화) 메일 전송 완료
- 2. 박기호 교수님께 변동사항 전달하기, 9/29(화) 까지
- 3. 변동사항 정확하게 숙지하기
- 4. Word2Vec 모델 이해하기
- 5. 데이터 수동 수집, 9/29(화) 시작
- 6. 수정 제안서 제출, 9/29(화) 까지
- 7. 요구사항 분석서, 유스케이스 완성, 10/3(토) 까지
- 8. 발표 영상 완성, 10/4(일) 까지
- 9. 요구사항 분석서, 유스케이스, 발표영상 제출, 10/5(월) 까지