

회의록(3주차)

#캡스톤/회의록

21.03.19

각자 한 일 공유

용훈

- 유튜브 데이터 크롤링 방법 공부
 - 방법1) 특정 채널 아이디 입력 → 해당 채널에 대한 조회수, 좋아요, 기타 정보 출력
 - 방법2) 인기 동영상 순위 데이터를 출력

지석

- 빅데이터 키워드 추출
 - 방법1) Konlpy 를 이용한 최다 키워드 순위 처리

논의

방법2만 이용해서 크롤링 할 것인지, 방법1도 이용해서 크롤링 할 것인지

방법1은 특정 채널의 데이터만 가져오기 때문에 종합적인 데이터를 가져올 수 없음.

방법2만 이용해서 하자고 합의

크롤링 데이터를 주간별로 사이트에 반영할지, 일간별로 사이트에 반영할지

매일매일 특정 시간에 데이터를 크롤링하여 사이트를 업데이트하되,

구독자 뉴스레터는 주간별로 데이터를 모아 종합하여 키워드 추출한 값을 전송하기로 한다

주간별 데이터를 독립적으로 추출해도 될지, 아니면 지난 주와 비교하여 순위를 매길지

인기동영상 선정 방식에서 이미 지난 주와 데이터를 비교하여 선정함. 이에 따라 독립적으로 순위를 매겨도 괜찮다는 의견으로 합의.

유튜브 데이터만 뽑을지, 다른 사이트 데이터도 뽑을 지

지석 의견) 프로젝트를 기획할 때, 나는 문화 & 취미 를 위주로 데이터를 뽑고 싶었다. 이에 따라 유튜브 데이터만 사용하는 것이 괜찮다고 생각

용훈 의견) 유튜브만 하면 데이터가 특정 분야에 치우칠 것이라 생각한다. 이에 따라 다른 사이트 데이터도 취합하여 포괄적 사용자층을 확보하고자 한다.

합의 내용 : 일단 빅데이터 처리를 위해선 데이터 형식이 일치해야 한다. 일단 유튜브 데이터만을 이용하여

처리하고, 추후 남는 시간에 데이터를 처리하는 방식으로 하자

일정 계획 합의

- 중간 발표 전까지 구현 목표 설정 (03/29)

- ☐ 중간 발표 자료 만들기
- ☐ 크롤링 데이터 만들기
- ☐ 적합한 키워드 추출
- ☐ 서버 탑재 → 서버 문의 확인
- ☐ 뉴스레터 프로세스 확인
- ☐ 서버 DB 구현
- ☐ 크롤링 데이터 출력 형식에 대한 논의 완료

다음 회의 준비사항

- ☐ 용훈 : 사이트 와이어 프레임
- ☐ 지석 : csv 돌려서 결과 확인
- ☐ 지석 : 서버 문의사항 확인