산학 캡스톤 과제 미팅 정리

# [2019-01-31]

1. 코드 공유 및 의견 공유는 bitbucket을 이용하여 한다

( <https://bitbucket.org/hanyang_cooperation/hanyang_cooperation/src> )

1. 향 후 일정

* 2월 중반 온라인 미팅 (WBS를 토대로 기획서 준비 필)
* 2월 후반 오프라인 미팅
* 3월 리뷰 미팅 (결과 안나오면 연속적인 3월 미팅 필요 예상)

1. DB 관련

* DB 포맷은 bitbucket에 올려주기로 하였고 필요한 attribute만 가져다 쓰면 된다
* Raw data table, after analysis table, category table 등등이 필요로 예상

1. Restful API interface는 회사측은 server, 팀측은 client의 역할만 존재하고, 분석 요청과 분석한 데이터를 가져가는 요청 두 가지 인터페이스가 존재
2. 감성분석

* 감성분석시 긍정, 부정, 중립 세 가지로 나누고, 긍/부정이 나타나지 않는 경우를 중립으로 한다.
* 추가적으로 딥한 감성분석을 원한다면 강한 긍정, 강한 부정 단계를 만드는 것을 추천해주심
* 단순한 긍/부정 말고도 기능과 부분을 나누어 구체적인 긍/부정 케이스를 나누는 팁을 주심
* 핸드폰을 예로 들면, 디스플레이 부분, 버튼 부분, 배터리 부분, 소프트웨어 부분의 긍/부정을 따로따로 산정한다던가 하는 식

1. 쓰레기 데이터 처리 관련

* 쓰레기 데이터에 대한 정의를 실제 댓글들을 보면서 다시 정해보라고함
* 크롤링 단계에서 기본적인 전처리가 필요

(가볍게 걸러내는 작업 필요 : 영상/댓글 길이가 너무 짧은 것들, url, 이모티콘 등)

* 의미가 없어보인다고 해도 쓰레기 데이터인지 한번 더 생각해 볼 것

1. AWS를 설 이후 빠른 시일 내에 주시겠다고함

* 받으면 바로 환경구축해야함

1. 제품군 선정 관련

* 제품 선정 시에 경쟁 브랜드를 넣는 것을 원하시는 듯함

( 핸드폰 2가지 이외에도 LG 회사이니 LG 핸드폰 분석도 해보라하심)

* 현재 선정한 제품들 중에서 보면 쓸만한 댓글 내용 많이 없기 때문에 좀 더 자세히 리뷰한 댓글이 많은 유튜브 동영상/주제를 찾아야함
* 현재 선정 되어있는 제품들 중에서 그래픽카드가 와닿지 않는다고 하심. 대중적인 다른 제품으로 바꿔야할 듯
* 새로운 제품 선정 시 기준이 되야할 것
  + 공감대, 관심 많은 분야 (캡스톤 발표의 청중은 교수님들과 학생들)
  + 영상에 대해 어떤 식으로 달리는지 확인
  + 영상 내에 좋은 댓글이 달리는 걸 찾아서
  + 사람 눈으로 데이터를 보았을 때 최종 결과가 예상되는 영상들로

(결과를 미리 예상해두고 코딩하는 방법을 추천해주심)

* + 예시 : 30대 이상 여자들 관심분야 제품들, 스마트스피커 등등

1. 이번 프로젝트에서 Hadoop과 spark를 안 쓸 계획
2. 서승우 사원님의 코딩 팁

* 형태소 분석이 잘 안되는 단어가 있다면 의미 없는 영어로 바꾸어놓고, 의미 없는 영어 단어를 강제로 어떤 분류로 해석하도록 코딩. 형태소 의미 분석 성공!

이후 다시 원래 단어로 바꿔주는 작업이 필요

* 정확도를 올릴 때 단어 간의 거리를 따져서 의미를 엮어서 분석하는 아이디어
* 분석시 타이틀과 댓글을 묶어서 분석

(example: 타이틀이 부정이고 댓글이 “찬성해요” 라고 했을 때 찬성해요는 긍정이지만 부정으로 해석할 필요가 있음)

* 감정 사전을 체계화하여 분야와 기능들로 쪼개서 체계화하여 사용

1. 어느 정도가 의미있는 데이터량이 모였다고 판단할 수 있나?

* 통계 모수 이론에 따른다고함

1. 웹

* 누가 데이터를 보는가에 초점 (end user!)
* 먼저 캡스톤 발표에 초점을 맞춘다면 발표를 보는 교수님들, 대학생들이 위주

교수님과 대학생들의 나이대를 생각하여 어떤 view를 제공할지 생각해볼 것

* 유튜브 데이터라는 속성을 생각해보고, flow 분석이 가능한다거나 조사한 것이 경쟁작이면 비교하는 view를 만든다거나, 시간 데이터를 이용할 수 있는 그래프를 쓴다던가 하는 식으로 생각해볼 것
* React라는 new web 기술을 쓴다는 것에 호의적.
* 빠른 시일 내에 해야할 것 (경수가 카톡에 올린 것들이 대부분)
* Wbs바탕으로 다다음주까지 기획서 작성
* 현재 선정한 제품에서 댓글 내용이 목적과 다른부분이 크기때문에 좀더 제품을 자세히 리뷰한 댓글이 많은 유튜브 동영상 찾기
* 제품 선정에 대해 다시 이야기 나눠보기
* 리눅스나 윈도우에서 형태소분석기(mecab)환경구축
* Aws계정받으면 환경구축
* 테이블구조 예시 받으면 가장 필요한 속성선택후 우리 웹에 사용할 정보들 추가 (테이블 구조는 이야기가 많이 필요할 듯)