5기 스터디 네이버DB 생성 프로젝트 권기남 백승렬

**프로젝트 소개:**

Growth Hackers 학회 자체 데이터를 생성하기 위해 네이버의 나이/ 요일/ 시간별 트렌딩 검색어를 축적하고자 함. 프로젝트의 중점은 이 데이터를 활용한 데이터분석이 아닌 데이터 축적에 있음. 본래 프로젝트의 목적은 상시로 돌아가는 GH 서버에서 실시간으로 GH 전용 DB를 업데이트하는 프로그램의 제작이었음. 최종적으로 만든 결과물은 실행시킬 시 DB를 최신으로 업데이트해주는 프로그램/ 코드의 제작이 됨.

**배우고자 한 점:**

1. Database 구축 방법
2. Bigquery 사용법 (database 생성 측면)
3. 크롤러 학습
4. 자동화
5. AWS 서버 사용법
6. 서버를 이용한 24시간 작동되는 서비스/ 툴의 제작

**배우지 못한 점:**

1. Database 구축 방법 – 의도한 database는 구축하였지만 일반적인 경로인지는 모르겠음
2. AWS 서버 사용법 – 정상작동하던 chromedriver가 안되고, 이유없이 서버가 터지고 새로운 서버를 파게 되면서 시간문제로 AWS의 사용을 포기함. AWS를 많이 다루어 볼 수 있었지만 최종 결과물에는 아쉽게도 사용하지 못함
3. 24시간 작동되는 서비스의 제작 – AWS에 문제가 생기며 실행이 힘들어짐. 이론은 구축했으나 chromedriver를 서버에서 읽지 못하고, 24시간 가동 시 서버가 감당 못하고 중간에 터지며 데이터가 안정적이게 구축되지 못하게 됐을 것을 감안하여 안정적인 방법을 택하기로 함

**프로젝트 결과물:**

1. 시작부터 끝까지 실행 시 GH DB에 NaverCrawler.Trending20 테이블을 업데이트 하는 ipynb 코드
2. [edu@ghsnu.com](mailto:edu@ghsnu.com) 의 빅쿼리 계정에 저장된 2019년 1월 1일부터 최신까지 저장된 일자/시간/나이대별 검색 키워드 top 20

**코드 설명:**

Ipynb 자체에 주석으로 설명되어 있음. 전체적으로는

1. DB에 연결한 후
2. 네이버 실시간데이터 정보의 특정 날짜/ 시간을 크롤러로 긁어온 후
3. 그 시간의 모든 나이대의 키워드를 긁어와서 하루치 dataframe을 제작한 후
4. 하루단위로 DB에 정보를 업데이트함

**추후 활용 가능 방향**

1. SQL 교육에 자체 DB의 활용
2. 다양한 프로젝트에서 과거 데이터를 볼 때 참고할 수 있는 DB의 활용

예) 2019년 5월 중순에 일본 비행기표가 매출이 떨어졌는데, 기업의 잘못인지, 혹은 사회적 문제 때문이었는지 분석

예2) 브로앤팁스의 매출이 2019년 10월에 급증했는데, 특정 제품의 성과 때문인지, 특정 방송/언론의 노출 때문인지, 혹은 사회적 이슈가 있었기 때문인지

1. 자연어처리 학습
2. 기타 다양한 분야에서의 적극적인 데이터 활용