

2018-07-22 세미나 정리

무지개팀 - 화빈 언니

- BAR 그래프 그리는 함수:

개선할 점: 4년치 데이터인데 x축 날짜는 초반 10개 정도만 뜸.

원하는 데이터가 뚜렷해져야 세부 사항 수정을 할 수 있을 듯.

크롤링팀 - 준호 오빠

- GIT BASH나 DOCKER에서 사용할 기본적인 명령어 소개:

cd + enter 키 치면 본인 홈으로 들어가짐

pwd : 현재 작업 중인 디렉토리 반환

mv : 수정

cp/scp : cp는 복사 scp는 서버에서 로컬, 로컬에서 서버로 복사 시킬 때

- scp 복사 파일 이름 yusso724@154.8.160.186:/home/yusso724

mkdir : 폴더 생성

ssh 서버 접근 명령

:set nu git bash에서 행 번호 보임

vi 파일 이름 파일 열어서 볼 수 있음

htop cpu 작업상황 다 보임

- 현재 크롤링 된 키워드: Sangziteng, Kesou, Feiyan, Biantaotiyuan, Yumen, Twitter
- 개선할 점:
 - 바이두 키워드 크롤링 시, 필요한 수치를 캡처하고 찍힌 스크린샷을 인식하는 방식이어
서 3과 9가 모두 9로 읽히는 문제 발생, 수동으로 일일이 대조하고 수정해야 함.
 - 모든 NET 멤버에게 GITHUB 관리자 권한을 부여했는데, 파일 업로드 시 충돌 문제 발
생 가능성 있음.

돌고래탐 유소진

- 기본 SQL문 설명, 데이터 베이스 수정 사항
- 향후 개선할 점:
 - 크롤링 데이터에서 빠진 날짜 있는지 확인
 - Hourly 데이터 DATETIME 한 시간씩 앞으로 당겨보기
 - 미세먼지 데이터랑 키워드 DATE_를 키로 외부조인으로 묶기.

돌고래팀 재근

- Pymysql을 사용한 Mysql과 Python 연동 방법 소개
- 하나의 데이터 베이스에서 모든 테이블을 가져오는 함수 (load_all_tables.py) 설명
- 처음에 테이블 읽어 오면 튜플 형식으로 저장, Join(i)은 튜플 형식을 string으로 변환
- 논의할 사항: 클라우드 서버 인터넷 속도 업그레이드 필요한지?