**멘토링 회의록**

| **프로젝트명** | 안숨은그림찾기 | | |
| --- | --- | --- | --- |
| **일 시** | 2022년 09월 17일  09:00∼17:50 | **참 석 자** | 데이터에듀 : 정대석 이사, 김현호 연구원 |
| **장 소** | 한국품질재단  한가람IT 교육센터  강의실(202호) | 1조 : 김동주, 안성현,  이정빈, 이한빈,  허아현 |

| **멘토링 내용 (요약)** |
| --- |
| **【 멘토링 받은 내용 】**  tip  프롬프트를 이용하여 서버에 접속하는 방법 익혀두면 도움됨  우분투(리눅스 버전의 한 종류)를 이용하면 좋다  CPU, GPU, batch\_size, epochs 에 대한 전반적인 설명   * epochs을 늘리게 되면 과대적합이 나타날 수 있고 batch\_size를 조정하여 테스트를 해보면 좋은 데이터가 될 것이다   가상환경의 종류 및 장단점   * 아나콘다는 편리하게 가상환경을 설치 할 수 있고 Python venv는 버전에 대한 이동이 자유롭다. Docker가 있는데 내컴퓨터안에 컴퓨터를 하나 더 넣는 느낌이다, 다른 OS들을 올려서 쓸 수 있고 GPU세팅을 잡을 수 있고 내가 쓰던 환경을 다른 사람에게 그대로 옮길 수 도 있다, 도커를 잘쓰면 장점이 될 수 있다.   함수 표기법   * VideoInfo(url) : 카멜표기법 * video\_info : 스네이크(파이썬) 표기법 * video-info : 웹표기법   GET : 서버쪽으로 전송을 할 때 주소창에 텍스트가 뜸 (보안의 문제가 있음) 장점 : 간단함  POST : 내용을 감추거나 내용이 많을 때 쓰게 됨  포스트맨을 쓰는 것을 권장   * 포스트맨 : 웹개발에 있어 형식들을 자동으로 만들어주는것   영상처리를 다루는 것이 큰 메리트가 될 수 있다.  BERT모델을 이용한 질의 응답시스템, 단어에 대한 빈도수를 이용해 주제나 해시태그를 뽑아낼 수 있는 기능을 만들 수 있다  이미지 유사도 검출방법   * PySceneDetect : 영상에서 어떤 변화가 일어났을 때 해당 프레임 및 시간을 뽑아주는 역할을 함     **【 요청 사항 】**  ○ 질의사항  1) 두런 사이트 소스 확인 = Youtube 사이트 연결방법 확인   * 유튜브 사이트 크롤링을 해서 유튜브에 대한 url을 통해 그 유튜브의 데이터를 가져오게됨 영상을 가져온다는 것 자체는 불법이다   2) 텍스트 검색시 해당 타임스탬프로 이동하는 기능 구현 가능여부(검색기능)   * jQuery를 이용해 간단하게 해결 할 수 있고 간략한 페이지 만드는 소스를 줄 수 있다.   3) 회사 서버 사용가능 여부   * 데이터 에듀 쪽 서버제공, 학습데이터의 크기가 크기 때문에 학습데이터들을 서버에서 생성을 시켜도 되고 서버를 업로드시켜도 됨, 메모리 용량 24기가, 그래픽카드의 3배의 서버를 할당   4) 두런에 동영상들이 어떻게 들어가있는지   * DB에 저장되어 있음   5) IT 이슈를 어디서 보는지   * SOTA(STATE OF THE ART) : 어떤 모델에 대해 순위를 매겨주는 사이트, SOTA를 쓸줄알면 상당히 큰 메리트가 될 수 있다   6) STT에 대해 배울 수 있는지   * 진입장벽 순위 : CV << TTS <<<STT = NLP, AI는 블랙박스이기 때문에 안의 구조를 보지 못하기 때문에 진입장벽이 생김   6) 추천하는 자격증이 있는지   * 최근에 포트폴리오에 지금까지 뭘 했는지 기록해둘 수 있기 때문에 자격증은 의미가 없다. 그리고 포트폴리오를 만드는 툴로 notion을 추천함   **# check 사항**  10월 8일 최종점검 |
|