캡스톤디자인 프로젝트 중간보고

인공지능 기반 Windows 악성코드 탐지 및 PE 파일분석 웹 서비스

아메바 (우건희, 김동영)

목차

개발 목적

- 코로나 19라는 Pandemic을 겪으며, 인프라의 디지털/비대면화가 가속되고 있습니다.

- 이 과정에서 악성코드를 비롯한, 보안 위협들이 증가하고 있습니다.
(국가정보원, 한국인터넷진흥원 등, “2021 국가정보보호백서”. 국가정보원, 한국인터넷진흥원 등. https://www.kisa.or.kr/20303/form?postSeq=0241&page=1)

- 악성코드 분석가가 모든 악성코드에 대한 세부 분석(eg., 리버싱)을 실시 할 수 없습니다.

- 인공지능을 통해 악성파일 여부를 알려주고, 분석하는데 유용한 PE값을 제공하여,
악성코드 분석 시간을 줄여주는 웹 서비스를 개발하고자 합니다.

담당 분야

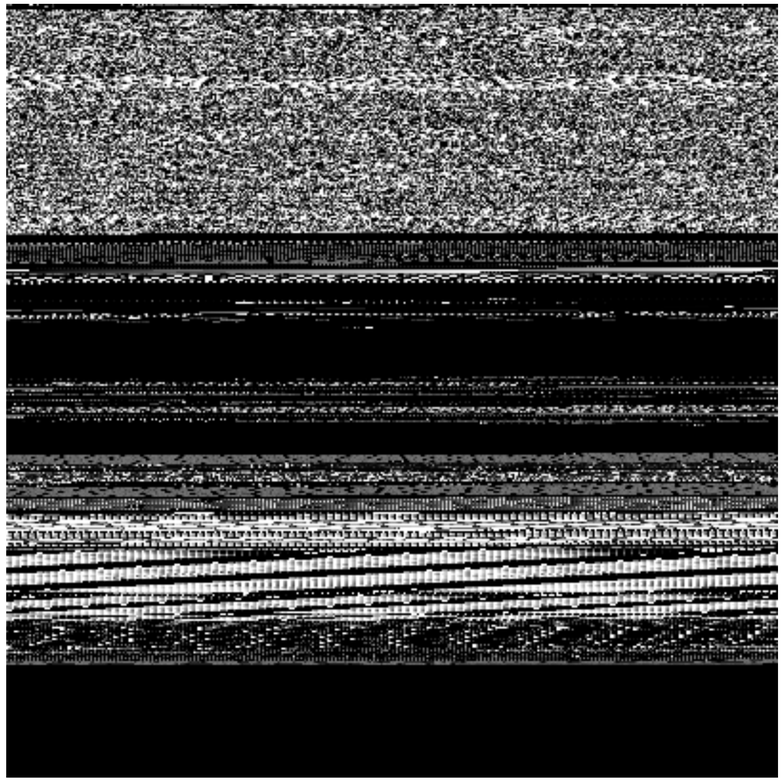
- 크게 2가지 부분으로 나뉘어 개발이 진행됩니다:
  - 악성코드 분석 인공지능 부문 (김동영 담당)
    - 악성, 일반 프로그램 (데이터) 수집 및 특징 공학
    - 인공지능 모델 개발
    - 악성코드 관련 코드 및 악성코드 분석 서버 개발 담당
  - 웹 서비스 부분 (우건희 담당)
    - 웹 서버 (Full-Stack) 개발
    - DB 서버 (스키마 작성, DB 구축) 담당
    - 클라우드 서버 담당

악성코드 분석 인공지능 부문

악성코드 분석 인공지능 부문

진행 내역

- 악성코드 크롤러 개발 및 악성코드 수집 완료
  - 악성코드 약 3만 9천개, 일반 프로그램 800개
- SHA-256 해시값 추출 및 파일 이름 변경 코드 개발 완료
- PE 파일만을 골라내는 코드 개발 완료
- PE값 중 일부를 추출하는 코드 개발 완료
- CNN 분석을 위한 악성코드 -> 이미지 변환 코드 개발 완료
- 악성코드 분석을 위한 CNN 개발 중
  - Loss Function을 Binary Cross Entropy로 했으나,
    Loss 값이 소수 값으로 가는 Overfitting 현상 등을 보임
  - Confusion Matrix를 통해 분류 여부 확인, 샘플수 조절로 문제를 해결해 나갈 예정



악성코드 분석 인공지능 부문

추후 계획

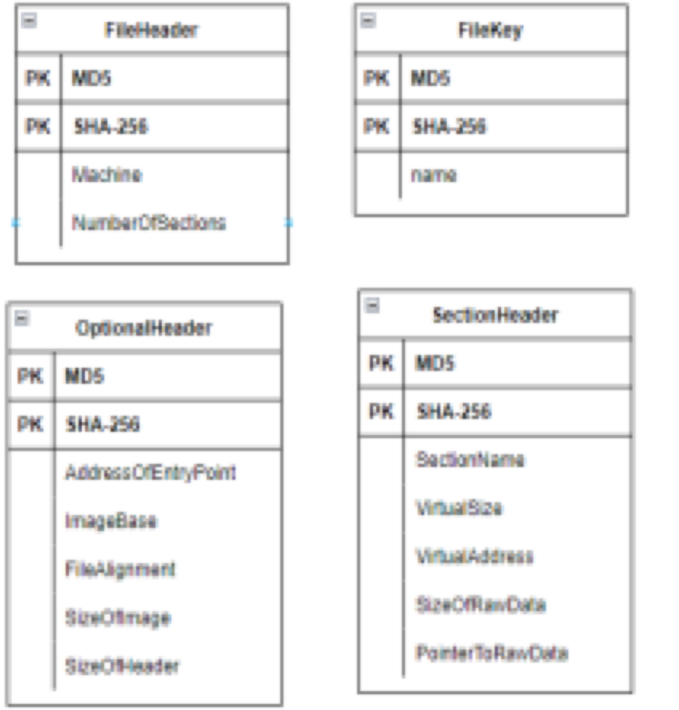
- DNN 모델을 완성할 계획입니다.
  - Imported API 정보를 이용해, 임베딩(word2vec, N-Gram)을 하여 학습데이터로 사용할 예정입니다.
  - CNN과 DNN을 개발한 후에는 Softvoting을 통해, 모델의 신뢰성을 높이려 합니다.
    * Softvoting: CNN과 DNN의 예측값의 평균값을 구하는 과정
- 추가 기능을 개발할 예정입니다.
  - Packing 파일에 대한 YARA Ruleset을 이용하여, 패킹여부와 분류하는 기능을 개발할 예정입니다.
  - CNN용 데이터 이미지와 분석한 String 데이터를 CSV를 출력하는 기능을 개발할 예정입니다.
- DB(MySQL)와 연동하는 코드를 개발할 예정입니다. 이미 접속 시험은 완료하였습니다.
- 작성한 코드들을 통합하여, 악성코드 분석 서버가 작동할 수 있도록 할 예정입니다.

웹 서비스 부문

웹 서비스 부문

진행 내역

- DB는 AWS의 RDS를 이용하여, 운영 및 관리 되고 있습니다.
  - DB에 대한 테이블 설계는 완료하였습니다.
- 웹서버는 AWS의 EC2를 이용하여, nodejs, React를 설치하였습니다.
  - 웹서버에 대한 Front-End 개발은 70% 정도 완성된 상태입니다.
  - 우측 하단에 있는 사진은 실제 개발 중인, 웹페이지의 사진입니다.

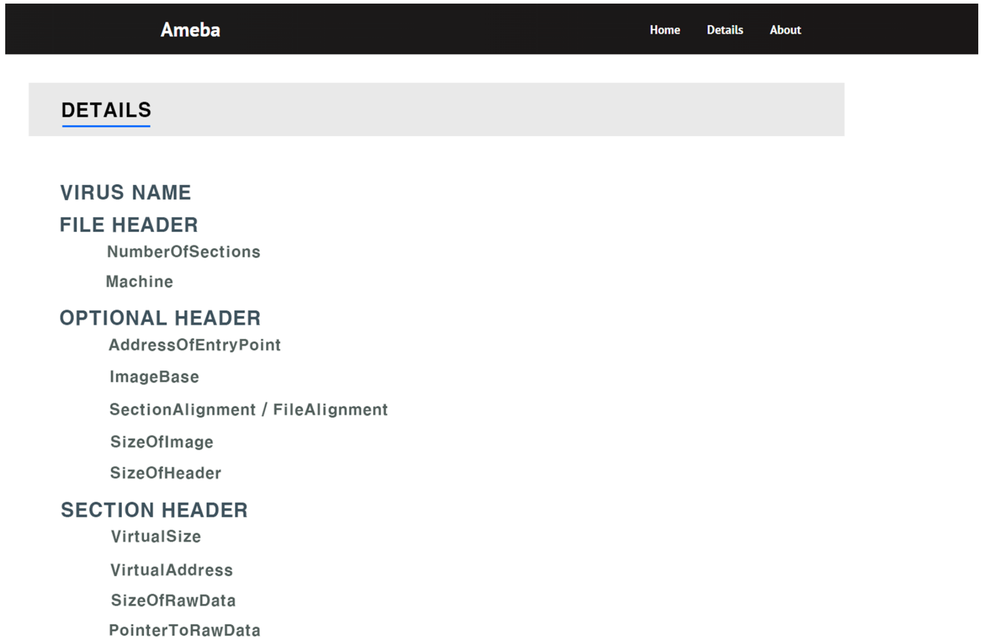


웹 서비스 부문

추후 계획

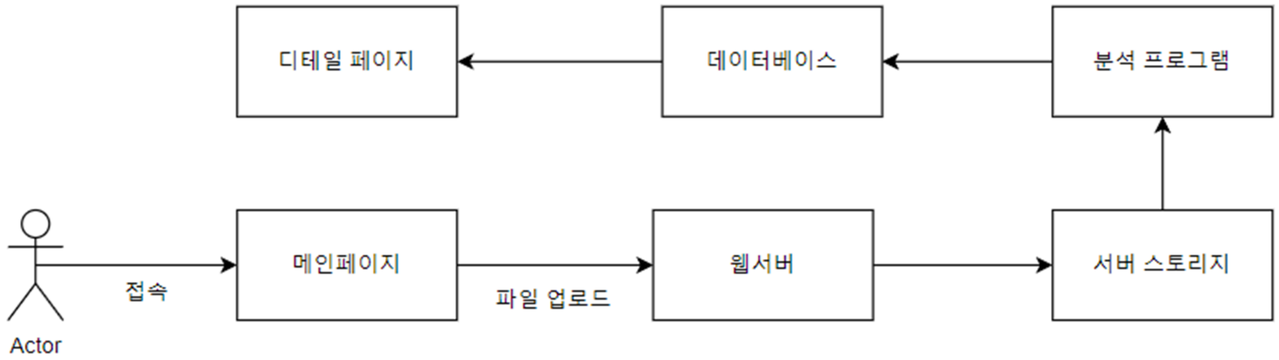
- Front-End의 남은 부분을 완성할 예정입니다.
    - AboutPage 구성, Detail Page 구성, Footer 링크 연동.
- Back-End를 개발할 예정입니다.
    - 파일이 업로드 되는 부분을 구현할 예정입니다.
    - 파일의 SHA-256 값을 구해, DB에 저장된 정보가 있는지 확인할 예정입니다.
        - (저장된 정보가 있을 경우) DB에 저장된 정보를 출력하도록 수정할 예정입니다.
    - Front/Back-End의 개발 완료시, AWS를 이용해 웹서비스를 운영할 예정입니다.
    - 정상 작동 시, 시연동영상 촬영과, 보고서 작성, 통합 테스트 등을 진행할 예정입니다.

웹 서비스 부분 예시 #1



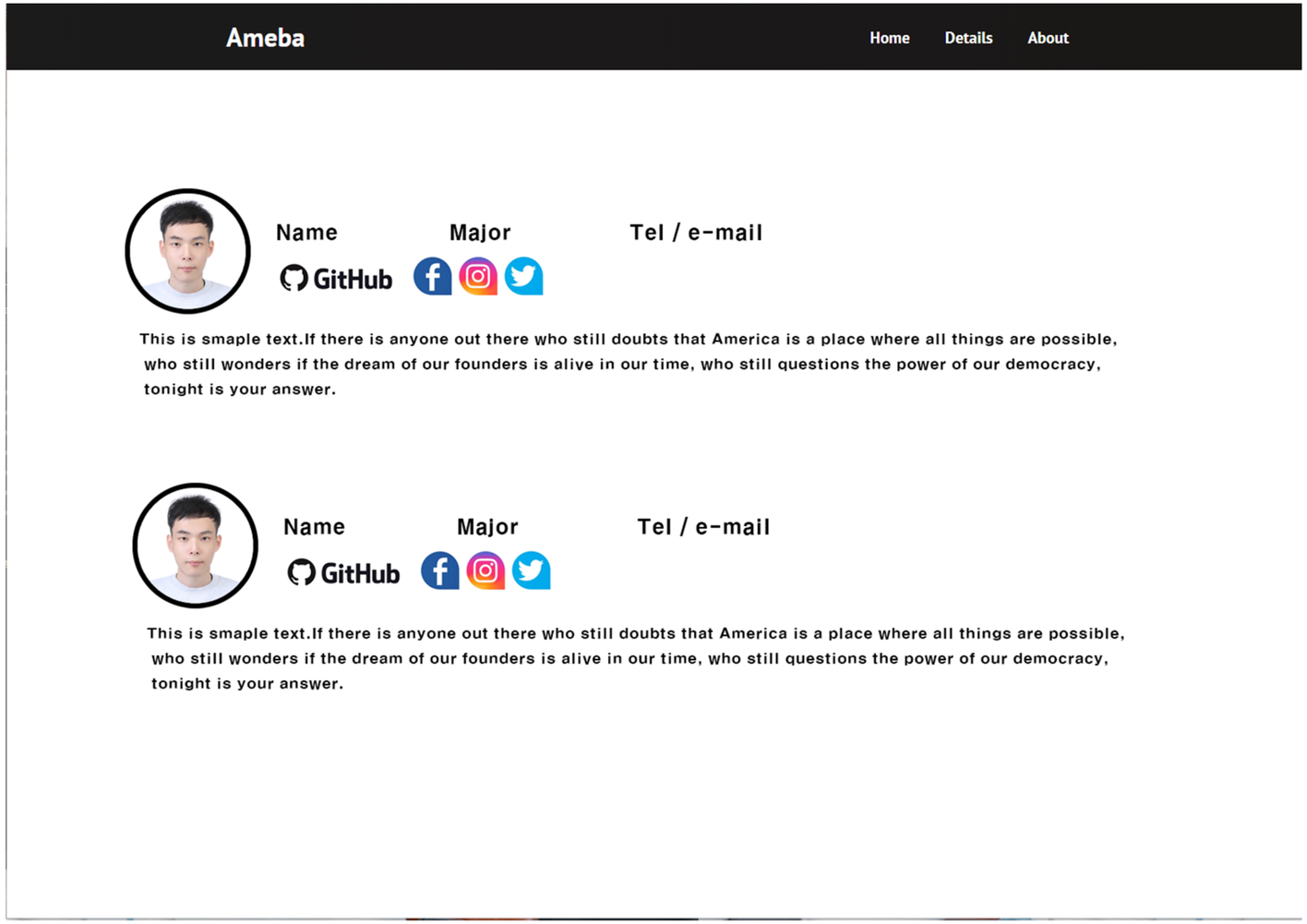
디테일 페이지 예시입니다.

사용자가 메인페이지에서 파일을 업로드하면 웹서버를 통해 파일이 서버 스토리지로 이동합니다. 들어온 파일은 분석 프로그램을 통해 분석되고 데이터베이스에 있는지 확인을 합니다. 없으면 데이터베이스에 새로 분석한 자료를 Insert하고 디테일 페이지에선 데이터베이스에 있는 해당 자료를 사용자에게 보여줍니다.

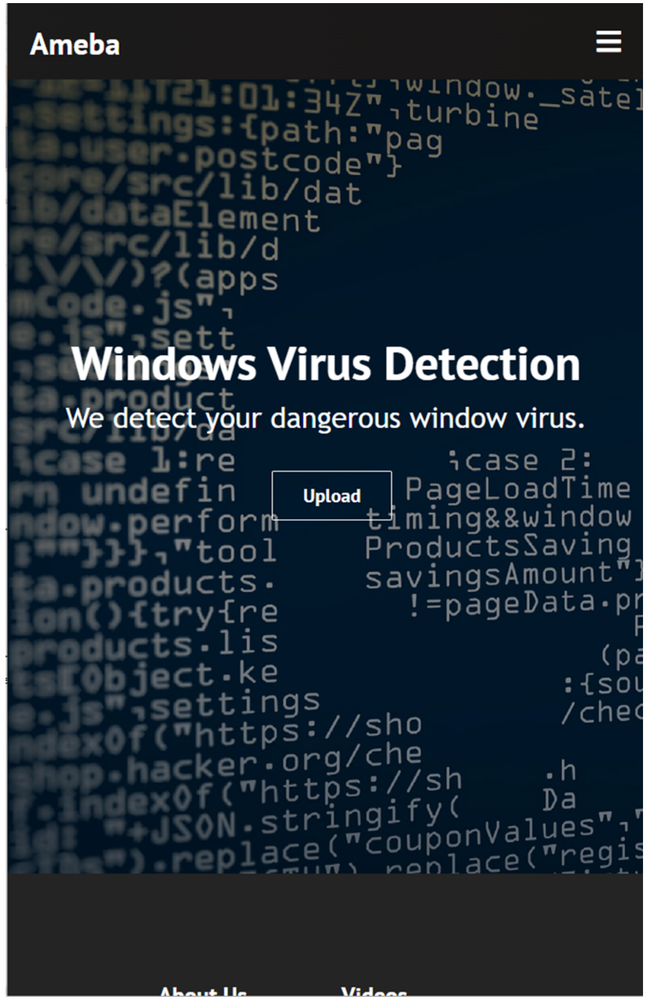


사용자가 서비스를 사용하는 흐름입니다.

웹 서비스 부분 예시 #2



어바웃 페이지 예시입니다.



모바일 페이지 예시입니다.

감사합니다.