

팀 미팅 회의록

팀명	Do it !	차수	5 차
일 시	2019년 4월 01일 월요일 15시 00 분 – 17시 00분 (2시간 00분)		
장 소	7호관 자율주행스튜디오		
참석자	전원		
불참자	없음		
안 건	수행계획서 피드백 반영 토의		
회의내용	<ol style="list-style-type: none"> 문서 파일에서 악성코드에 해당하는 특성(매크로 등)을 어떻게 추출할지 → 매크로 등 악성코드에 해당하는 특성을 태그 정보로 구분 및 판별 학습 모델 결정 → 실험했던 결과들을 나열. 딥러닝은 얇은 신경망인 CNN을 사용하되 딥러닝 뿐만 아니라 여러 머신러닝 모델을 사용. 학습 데이터의 규모 및 데이터의 불균형 → PDF 파일의 경우 약 20만개의 데이터 수집. → 정상 파일과 악성 파일의 양을 비슷한 수로 맞춘다. → 다양한 데이터 공유 사이트에서 크롤링을 통한 수집 플랫폼에 대한 구체적인 계획 → 웹서버와 엔진을 어떻게 연동시키는지에 대한 구체적인 설명 추가 정량적 평가 다양화 → Precision, Recall, F1 – score 평가 방법을 사용 기존의 악성파일 탐지 기법과 비교하여 새로운 기술적 시도를 담고 있는지 → 기존 연구에서는 시그니처 기반으로 탐지했지만 이에 비해 문서파일의 구조적 특징을 피쳐 해싱 기법을 통해 벡터화한다. 또한 여러 분야에 접목시킬 수 있도록 확장성이 높은 엔진을 개발하고자 한다. 어떠한 feature 들을 분석하려고 하는지에 대한 구체적인 설명 추가 → 문서 구조를 특징으로 피쳐로 삼고자 한다. 구조에 대한 피쳐는 문서의 유형마다 다르므로 연구가 필요하다. 		

결과물	<ol style="list-style-type: none"> 1. 지속적인 데이터 수집 2. 주간 팀별 구체적인 일정 조율 3. 웹 페이지에 파일 업로드 기능 구현 완료