**연구계획서**

부서명 : 통로이미지㈜ 디지털 연구소

작성자 : 김근수

최초 작성일 : 2020.07.03  
최종 수정일: 2020.12.28

|  |  |
| --- | --- |
| 연구명 | 머신러닝\_이미지\_텍스트\_추출\_연구개발 |
| 개요 | 이미지 콘텐츠를 구매하려는 고객중 구매한 콘텐츠의 텍스트를 변경하여 사용하기 위해 텍스트가 삽입되기 전의 원본 콘텐츠를 원하는 경우가 있음    이를 위해 텍스트가 없는 콘텐츠를 추가로 입고받고 있으나  입고 받는 과정에서 사람이 수동으로 확인해야 하는 과정이 필요하게 되고 이 과정에서 누락이 발생할 수 있음  이미지 콘텐츠에 텍스트 포함 여부를 자동으로 검사하게 하여  콘텐츠 선별 업무에 효율성 증대시키기 위함 |
| 목표 | 1. 이미지에서 한글 및 영문 텍스트 추출 2. 텍스트 추출 정확도 향상 3. 한글, 영문을 제외한 언어 인식 4. 언어 인식률 향상을 위한 머신러닝 학습 |
| 활용기술 | Python Anaconda OpenCV  Tesseract  PyTesseract. |