**5월 프로젝트 데이터 수집 처리**

2022대구AI스쿨 심화과정 지상준

**1. 실행 방법**

chromedriver.exe, crawling.py, dataset.py, main.py

위 파일을 같은 경로에 두고 main.py를 실행합니다.

**2. 코드 설명**

|  |
| --- |
| crawling.py |
| selenium을 이용하여 웹 페이지를 검색 및 검색결과를 크롤링합니다.  검색결과를 경우에 따라 base64 인코딩 된 값이나, 웹 url 결과에 맞게 cv2에서 쓰는 이미지로 변환합니다.  cv2 이미지를 255, 255 비율에 맞게 resize하며, 정사각형이 아닌 이미지는 padding 됩니다.  지정된 경로에 이미지가 저장됩니다.  with ~ as ~ 구문을 이용할 수 있도록 하여 자원 반환에 용이합니다. |

|  |
| --- |
| dataset.py |
| 생성자에서 제공받은 torchvision.transforms을 이용하여 이미지를 변환합니다.  transform 처리에 걸린 시간을 기록합니다.  라벨에 부여된 숫자를 반환합니다. |

|  |
| --- |
| main.py |
| 크롤링 후 자원 반환을 위해 with ~ as ~ 구문을 적용하였습니다.  torchvision.transforms의 인스턴스를 만들고, 이미지 처리를 위한 Resize, RandomCrop을 매개변수로 넘겨주었습니다. |