* STREM이라는 화학물질 판매 업체의 웹사이트에 등록되어 있는 물질명과 가격 등의 정보를 크롤링하여 엑셀에 StremCrawlingResult.xlsm 이라는 이름의 파일로 저장하는 프로그램입니다. 크롤링을 위해 Beautifulsoup와 selenium을 같이 사용했으며, 엑셀을 조작하기 위해 openpyxl 모듈 또한 사용했습니다. 크롤링은 구글 크롬 브라우저로 진행되며, 작동하기 위해서는 프로그램 폴더 내부에 chromedriver가 존재해야 합니다.
* CatalogCrawler 모듈은 엑셀 파일의 Catalog 시트를 구성하는 모듈입니다. 각각 대분류, 중분류, Catalog Number, Description, CAS number 5개의 열을 만들고, 간격을 설정한 후, 웹사이트에서 내용을 크롤링해 Catalog 시트에 크롤링한 텍스트를 입력하는 역할을 합니다.
* PriceCrawler 모듈은 엑셀 파일의 Price 시트를 구성하는 모듈입니다. 이 모듈은 Catalog 시트의 Catalog Number 열의 내용들을 하나씩 웹사이트에 입력하여 사이트에서 출력되는 무게 단위, 가격, 재고 유무를 크롤링하여 가져오고, 이 크롤링이 진행된 시각도 함께 Price 시트에 입력합니다. 같은 물질이라도 무게 단위가 여러 개 있기 때문에 한번에 전부 진행할 경우 작업 완료까지 시간이 매우 길게 소요되기 때문에, 우선 100개의 Catalog Number만 진행하게끔 작성해 두었습니다.
* RestPriceCrawler 모듈은 폴더에 이미 StremCrawlingResult.xlsm 파일이 존재해야만 작동하는 모듈로, PriceCrawler 모듈에서 이미 진행된 Catalog Number 이외의 물질들의 정보들을 크롤링해 엑셀로 저장하는 모듈입니다. 크롤링하는 정보는 PriceCrawler 모듈과 같습니다.
* StremCrawler 모듈은 가장 먼저 실행되는 모듈로, 현재 폴더에StremCrawlerResult.xlsm 파일이 존재하지 않으면 CatalogCrawler 모듈과 PriceCrawler 모듈을 차례대로 실행하고, 파일이 이미 존재하면 RestPriceCrawler 모듈을 실행하도록 선택하는 역할을 합니다.
* StremPandasProcess는 데이터의 정렬을 위해 만들어 둔 모듈로, CatalogCrawler 모듈의 마지막에 데이터들을 대분류와 중분류를 기준으로 알파벳 순 정렬을 하는 모듈입니다.
* 코드의 맨 마지막에는 엑셀 시트를 보호된 시트로 전환하는 기능이 있으며, 시작 부분에는 이 보호를 해제하는 기능이 있어 부분적으로 시트의 임의 수정을 방지할 수 있도록 해 두었습니다.