* STREM 크롤러보다 나중에 만든 크롤러이기 때문에 STREM 크롤러에 비해 몇 가지 업데이트가 있습니다. 이름이 FirefoxEnamineCrawler인 이유는 원래 chromedriver와 크롬 브라우저를 이용하는 크롬 버전과 geckodriver와 파이어폭스 브라우저를 사용하는 2가지 버전으로 만들었으나, 크롬 버전은 실행할 때마다 빈 콘솔창이 같이 실행되고, 실행 속도 또한 파이어폭스 버전에 비해 더 느렸기 때문에 이후에 크롬 버전을 삭제했기 때문입니다.
* 작동에는 Tor 브라우저를 설치하면 생기는 Tor Browser 폴더 전체와 geckodriver, 파이어폭스 브라우저가 필요합니다.
* 개발한지 시간이 좀 지났기 때문에 이 크롤러는 약간의 추가 수정 작업을 제가 개인적으로 진행했습니다. 드라이버와 브라우저를 최신 상태에서 작동하도록 수정했고, 크롤링 대상 사이트가 개발 당시와 조금 구성이 달라져 있어 변경 이후의 사이트에서도 정상적으로 크롤링이 가능하도록 코드를 조금 수정하였습니다.
* Enamine 사는 크롤링 방지를 위해 특정 UserAgent가 지속적으로 사이트를 탐색할 경우 그 유저의 ip를 일시적으로 차단하는 기능을 운영하고 있었습니다. ip 차단을 방지하기 위해, 약 10분 내외의 간격으로 ip를 무작위로 변경해 주는 브라우저인 Tor 브라우저와 프로그램을 실행할 때마다 무작위로 UserAgent를 생성해 주는 라이브러리인 fake\_useragent를 사용했습니다.
* 하지만 Tor 브라우저 자체가 생소하기도 하고, 혹시나 사용할 때 생길 수 있는 다양한 문제들을 생각해 Tor 브라우저를 사용하는 부분만을 제외한 Only UserAgent Version을 따로 만들어 두었습니다.
* 이 이전에 제작했던 STREM 크롤러는 STREM 사가 취급하는 상품의 가짓수가 적은 편이었으므로 전체 상품의 정보를 모두 크롤링하는 방식으로 제작할 수 있었지만, Enamine 사는 취급 상품이 매우 많기 때문에 같은 방식으로 제작이 불가능했습니다. 그래서, 일종의 물질 구분 코드인 CAS\_Number를 이용하여 검색하고, 검색 결과로 나온 물질 1가지의 정보들만 가져오는 방식을 선택하게 되었습니다.
* 아무래도 검색창에 직접 검색을 해서 정보를 가져오는 방식인 만큼, STREM 크롤러 때와는 다르게 검색 결과가 존재하지 않을 때의 대책이 필요했습니다. 그래서 검색 결과가 없을 때 발생하는 에러를 이용해 그 에러가 발생하면 검색 결과가 없으니 다른 물질을 검색하라는 메시지가 출력되도록 구성했습니다.
* 정보를 엑셀 파일에 작성할 때, 그 엑셀 파일이 열려 있으면 크롤링한 정보를 파일에 쓸 수 없습니다. 그렇기 때문에 파일이 열려 있으면 파일을 닫고 다시 시도하라는 메시지가 출력되게끔 해 두었습니다.
* 엑셀 파일이 열려 있거나 찾고자 하는 물질이 취급되지 않는 물건이라 정보가 없는 등 오류가 발생할 만한 일이 생겼을 때, 프로그램이 종료되어 버리면 다시 실행해야 하는 번거로움이 발생하기 때문에 반복문을 활용해 오류가 생기면 다시 CAS\_Number 입력부터 진행할 수 있도록 처리했습니다.