

## devlog - webcrawler 만들어 보기

소방서에서 하는 업무 중에는 의소대라는 업무가 있다. 소방인력 중 하나인 의용소방대에 관한 업무이다. 허나 농촌 지역인 만큼 의소대의 정보를 기록하는 데에 있어서 애로사항들이 있다. 담당 반장님께서서는 이제 도로명으로 바뀐 새주소와 기존의 번지 주소와의 비교를 통해 의소대 대원들의 주소를 확인하는 작업을 하시는데 기존의 번지주소를 일일이 인터넷에 검색하여 변환하는 작업을 하고 계셨다. 간단한 작업이지만 이를 일일이 검색하여 엑셀에 작성하기란 쉽지 않고 시간 역시 많이 걸리는 작업이었다.

마침 페이스북에서 카이스트 사회복무요원이 우편물 분류 작업을 자동화작업을 해내면서 뉴스를 보고 이에 자극을 받아 파이썬을 통해 나 역시 간단한 웹크롤러 코드를 짜보려 했으며 딱 마침 실제 업무에서 이를 적용해 볼 수 있는 기회였다.

기본 로직은 엑셀 시트에서 주소를 불러 온 다음 : 오픈파이엑셀, 로드 시트, 변수 할당  
주소 변환 사이트를 BS로 펴오고 검색창을 불러온다 : BS4  
불러온 검색창에 주소를 넣고 검색버튼을 클릭하여 변환된 주소를 받는다 : 셀레니움,  
웹드라이버  
변환된 주소를 엑셀에 넣고 저장한다. : 세이브시트

파이썬의 외부모듈을 불러오기 위해서는 커맨드 창에서 파이썬 패스 지정후 모듈 다운로드 받아야했다.

pip install module를 통해 필요한 모듈들을 다운받고 실행에 나섰다.

하지만 어떤 문제에서인지는 모르겠으나 불행하게도 오픈파이엑셀에 동작하지 않았다..  
다행인 점은 그렇게 고난도의 작업이 아니라 다행이도 그냥 간단한 엑셀 리드와 엑셀 라이트 라이브러리로 대신할 수 있었다. 중간중간에 콘솔로그를 대신할 프린트 구문을 넣어 콘솔에 작업 현황을 알 수 있도록 하였다.

브라우저에서 html을 긁어오는 것은 문제가 안되나 클릭이라는 작업을 수행하기 위해서 크롬 드라이버를 설치해야했다. 셀레니움에서 웹드라이버로 크롬드라이버를 선택해주고 드라이버로 변수할당 한다.

하지만 계속하여 실패가 생겼다. 크롬드라이버가 돌아가는 모습을 눈으로 보면서도 왜 실패하는가 전혀 이해할 수 없었는데 한순간 커다란 깨달음이 다가왔다. 바로 크롬드라이버 속도가 느리기 때문에 타임 인풋을 넣어서 브라우저가 검색 결과를 로딩하는 시간을 기다려 줘야한다는 점이다. 평소에 기계적인 코드를 생각하다가 컴퓨터가 작동하는 현장에서의 코드에 타임 모듈을 넣어 기다리는 시간을 주는 행위는 정말 '현실적'이었다.

하지만 무엇보다도 가장 크게 와닿았던 현실은 바로 데이터 관리의 중요성이다. 아무래도 소방서이다 보니 다른 곳처럼 이곳 또한 데이터를 엑셀에 단순 저장하는데 교대근무에다가 각 센터별로 담당자도 모두 다르다 보니 서식이 모두 제각각이었다. 코드는 그저 단순하게 지정된 열만 순행하는데 주소 데이터는 각기 다른 열에 있거나 혹은 띄어쓰기나 빈칸이 멋대로 있는 등 문제가 많았다. 데이터의 통일성을 훼손하지 않기 위해 지정된 서식을 공유하고 엑셀이 아닌 데이터베이스를 사용해보는 연습이 필요했다.