Crizensolution 발표 대본

1. 안녕하세요 크리젠솔루션 인턴십 결과물 발표에 있어 전달성을 위해 대본을 작성하였고 대본을 토대로 발표를 시작하도록 하겠습니다.
2. 목차는 개발 준비 , python 개발, java개발 html, javascript 준비, 스프링부트 그리고 마무리로 이루어져 있고 또한 발표에 들어가기에 앞서, 기술적인면의 디테일은 과장님께서 저보다 훨씬 잘 아시기때문에 기술을 설명하기보다는 그동안 어떤 것을 해왔는지 위주로 발표를 진행하도록 하겠습니다.
3. 우선 처음에 네이버부동산 스크래핑을 진행하기 전 과장님께 필요한 사전 지식과 정보들을 전해들었고 이를 토대로 개발을 시작하게 되었습니다. 개발을 시작하기에 앞서 웹은 어떻게 작동하는지, 그리고 mvc에 대한 이해를 위해 스프링부트를 경험해 보았습니다.
4. 이후 제게 익숙한 언어인 python을 이용한다면 함수선택과, 코드작성을 자바보다 유연하게 할 수 있을 것 같다는 생각과, 원하는 결과물을 보다 빠르게 도출해 낼 수 있겠다라는 생각으로 파이썬코드로 우선 크롤링을 시작하게 되었습니다. 크롤링에 들어가기에 앞서 웹에 대해 전혀 경험해보지 못했고, 그로인해 html 이 뭔지, 어떻게 구성되고 코드가 읽혀지는지, 또한 어떤 태그들이 있는지 등을 공부하였습니다.
5. html에 대해 알게 된 후 본격적으로 크롤링의 구현능력을 높히기위해 태그명이 아닌 XPath를 활용하는 방법, http에 요청을 보내는 requests api, 그리고 규칙적인 상황에서의 크롤링을 원활하게 하기위해 re 라는 정규식 api에 대해 공부하였습니다.
6. 다음으로, 크롤링을 위한 모듈인 python 에 beautifulsoup의 기본 함수들을 경험해 보았고
7. 이후 html 소스코드를 보며 크롤링을 하는 연습을 시작하였습니다.

크롤링 함수사용에 익숙해 지기 위한 연습의 첫번째로 네이버웹툰 목록의 제목들을 모두 가져와보는 크롤링을 연습삼아 해보았고,

1. 다음으로 네이버웹툰의 평점도 크롤링 해보았습니다.
2. 그 이후에는 단지 텍스트를 긁어오는것이 아닌 논리함수를 활용하여 ,쿠팡의 노트북상품 중에 리뷰가 100개 이상이면서 광고상품이 아닌 제품들만 가져오는 조건으로 크롤링을 하는 코드를 짜보았습니다.
3. 하지만 제 목표는 정적인 페이지가 아닌 동적페이지를 크롤링 하는 것 이였기 때문에 이후 Selenium을 시작하였습니다. 다음과 같이 엔터키 탭키 텍스트입력과같은 것을 Selenium을 이용하여 페이지를 동적으로 작동시켜보았고,
4. 이후엔 Beautifulsoup과 Selenium을 동시에 활용하여 동적으로 페이지를 컨트롤하고 정적인 상태의 페이지 소스를 가져와 원하는 데이터를 추출하는 방식으로 크롤링해보았습니다. 또한 스크롤과같이 자바스크립트를 활용해해야만 하는 경우도 있었기 때문에 파이썬에서 자바스크립트 문법을 끌어와 사용해보았습니다.
5. 앞서 말씀드린 동적페이지, 정적페이지 들을 크롤링해본 경험을 바탕으로 본격적으로 네이버 부동산 크롤링을 시작하였습니다. 앞서 경험했던 사이트들보다 꽤나 복잡하였고 예외적인 일들이 일어날 수 있는 상황에서 대처를 할 수 있도록 코드를 짜는 것에 시간투자를 많이 했던 기억이 있습니다.
6. 결과물은 과장님께서 말씀하신대로 다음과 같이 데이터를 인덱스로 접근하는것이 아닌 key와 value값이 있는 json 형태로 구현을 하였고 또한 체계적인 JSON형태로 만들기 위해 과장님께 여러 번 자문을 얻었습니다.
7. 하지만 스프링부트를 이용하기위해서 자바가 필요하였고, 저는 파이썬 코드를 토대로 자바 언어로 코드를 짜기 시작하였습니다. 하지만 자바로는 꽤 많은 시간이 소비되었는데 그 이유로는 변수 선언의 방식과 문법이 파이썬과 상이 하여 적응하는데에 시간이 걸렸고, 자바의 장점이기도한 유지보수관리가 유연한 객체지향으로 해보려고 했지만 이미 절차지향으로 하드코딩 했던 파이썬 코드를 바꾸기엔 쉽지 않았습니다.
8. 자바 또한 개발과정에서 많은 오류와 예외 문제가 있었지만 결국 파이썬과 마찬가지로 Json 형태로 결과물을 완성하였습니다.

파이썬과 방식은 동일하였으나, Jsoup 과 Beautifulsoup은 비슷하면서도 문법의 활용이 달랐기때문에 Jsoup을 적응하기위해 구글링을 하여 지식을 습득하였고 또한, 파이썬으로 코드를 작성하면서 제가 어떤 코드를 작성하고있고, 파이썬 코드를 작성할 때 이 변수가 품고있는 데이터의 의미들을 따로 주석처리로 적어주지 않아, 제가 얼마전에 만든 코드임에도 불구하고 이 부분이 무엇을 하는 부분인지 정확히 기억이 나지 않았습니다.

만약 이 코드가 나 혼자만 보는게 아니라 같이 협업하며 만들어야 했다면 코드를 해석하는 데에 있어 비효율적인 코드를 짰다고 생각했습니다.

1. 이후엔 이제 스프링부트를 이용하여 웹을 만들 준비를 시작 하였습니다. 숫자를 입력 받고 버튼을 누르면 그 숫자에 해당하는 부동산 정보를 크롤링 하는 모양에 대해 과장님께서 아이디어를 주셨고, 그래서 틀부터 우선적으로 만들어보았습니다.
2. 하지만 사용자입장에서는 json형태의 데이터를 보고싶어 하는것이 아니기때문에, 이 문제를 해결하기위해 테이블 작성법에 대해 알아야 했습니다.

그 당시 과장님께서 HTML 과 javascript를 이용하라고 힌트를 주셨고, 매물마다 정보의 개수가 모두 다른 경우이면서 HTML, Javascript를 이용을 해야한다면 동적테이블을 활용해야 한다고 생각했습니다. 따라서, 설 기간에 자바스크립트라는 언어를 공부하게 되었습니다.

1. Css로 보기 좋게 꾸미진 못했지만, HTML 과 Javascript를 이용하여 테이블을 만드는데 성공하였고, 다음과 같이 숫자를 입력하고 버튼을 누르게 되면 매물정보가 테이블에 정리되어 나오는 식으로 구현하였습니다.
2. 또한 과장님께서는 사용자 입장에서 면적별로 비교를 하고싶을수도 있지만 특정 면적에 해당하는 정보만 얻고 싶을 수 있다라는 말씀을 주셨고 새로운 페이지에 다른형태의 동적테이블을 만들기 위해 클래스 변수에 Json 객체 정보를 저장하였고

1. 면적별로 정리한 새로운 테이블을 조회할 때 이미 스크래핑된 Json 데이터를 클래스 변수에 저장한후 JavaScript로 끌어와 사용하는 방법으로 구현하였습니다.
2. 인턴십을 하면서 입문수준이지만 웹프로그래밍을 경험해 보고 이렇게 인턴으로 실무를 옆에서 지켜본 것 만으로도 전 아는 것이 정말 많이 없다고 느꼈습니다. 스프링, html, javascript 등을 해보았지만 과장님께서 말씀하신대로, 이것들을 할 줄 아는 사람이 아닌 진입장벽이 조금 낮아진 사람이지만 이렇게 경험한 것 만으로도 만족하고 있습니다. 저는 이후에 웹프로그래밍에 대해 더 많이 공부해보고 싶다는 생각이 들었고 저는 자바를 이용하여 웹 개발도 할 수 있고, 파이썬을 이용하여 데이터분석도 가능한 개발자가 되고 싶은 목표가 생기게 되었습니다. 인턴 과정에 있어서 많이 모르지만 항상 친절하게 알려주셔서 감사드리고 앞으로 더 열심히 해서 나중엔 개발자 대 개발자로써 다시 과장님을 뵐 수 있도록 노력하겠습니다. 기술적인면 뿐만아니라 사회생활도 배울 수 있어서 저에게는 뜻 깊은 시간 이였습니다. 감사합니다.