Crizensolution 발표 대본

1. 안녕하세요 크리젠솔루션 인턴십 결과물 발표에 있어 전달성을 위해 대본을 작성하였고 대본을 토대로 발표를 시작하도록 하겠습니다.
2. 목차는 개발준비 , python 개발, java개발 html, javascript 준비, 스프링부트 그리고 마무리로 진행하겠습니다.
3. 우선 처음에 네이버부동산 스크래핑을 진행하기 전 과장님께 필요한 사전 지식과 정보들을 전해들었고 이를 토대로 개발을 시작하게 되었습니다. 개발을 시작하기에 앞서 웹은 어떻게 작동하는지, mvc에 대한 이해를 위해 스프링부트를 경험해 보았습니다.
4. 이후 제게 익숙한 언어인 python을 이용한다면 함수선택과, 코드작성을 자바보다 유연하게 할 수 있을것 같다는 생각과, 원하는 결과물을 보다 빠르게 도출해 낼수있겠다 라는 생각으로 파이썬으로 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 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.
8. 크롤링 함수사용에 익숙해 지기 위한 연습의 첫번째로 네이버웹툰 목록의 제목들을 모두 가져와보는 크롤링을 연습삼아 해보았고,
9. 다음으로 네이버웹툰의 평점을 크롤링을 구현해 보았습니다.
10. 그 이후에 단지 텍스트를 긁어오는것이 아닌 논리함수를 활용하여 ,쿠팡의 노트북상품 중에 리뷰가 100개 이상이면서 광고상품이 아닌 제품들만 가져오는 방식으로 크롤링을 하는 코드를 짜보았습니ㄷ.
11. 하지만 제 목표는 정적인 페이지가 아닌 동적페이지를 크롤링 하는 것이였기 때문에 이후 Selenium을 시작하였습니다. 다음과 같이 엔터키 탭키 텍스트입력과같은 것을 Selenium을 이용하여 페이지를 동적으로 작동시켜보았고,
12. 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 사용해보았습니다.
13. 앞서 말씀드린 동적페이지, 정적페이지 들을 크롤링해본 경험을 바탕으로 본격적으로 네이버 부동산 크롤링을 시작하였습니다. 파이썬 코드로는 약 4일 정도 시간에 걸쳐 완성하였고,
14. 결과물은 과장님께서 말씀하신대로 다음과 같이 key와 value값이 있는 json 형태로 구현을 하였습니다.
15. 하지만 스프링부트를 이용하기위해서 자바가 필요하였고, 저는 파이썬 코드를 토대로 자바 언어로 구현을 시도하였습니다. 하지만 자바로는 꽤 많은 시간이 소비되었는데 그 첫번째 이유로, 자바로써의 구현 과정은 파이썬과 마찬가지로, 원하는 데이터를 추출하는 따라서 자바로 개발을 하면서는 누가 봐도 이 코드는 어떤것을 하고있는 코드인지 직관적으로 알 수 있게끔 주석을 달며 코드를 작성해 보았습니다.
16. 자바 또한 개발과정에서 많은 오류와 예외 문제가 있었지만 결국 파이썬과 비슷하게 Json 형태로 결과물을 완성하였습니다.
17. 이로써 이제 스프링부트를 이용하여 웹을 만들 준비를 시작 하였습니다. 숫자를 입력 받고 버튼을 누르면 그 숫자에 해당하는 부동산 정보를 크롤링 하는 것을 생각했고, 이를 방식은 동일하였으나, Jsoup 과 Beautifulsoup은 비슷하면서도 문법의 활용이 달랐기때문에 Jsoup을 적응하기위한 시간이 필요하였습니다. 또한, 파이썬으로 코드를 작성하면서 제가어떤 코드를 작성하고있고, 파이썬 코드를 작성할때 이 변수가 품고있는 데이터의 의미들을 따로 주석처리로 적어주지 않았었는데, 비록 제가 얼마전에 만든 코드임에도 불구하고 정확히 기억이 나지 않았습니다.

만약 이 코드가 나 혼자만 보는게 아니라 같이 협업하며 만들어야 했다면 코드를 해석하는 데에 있어 비효율적이라고 생각이 들었고,

1. 구현하기위해 테이블 작성법에 대해 알아야 했습니다.
2. 그 당시 과장님께서 HTML 과 javascript를 이용하라고 힌트를 주셨고, 매물마다 정보의 개수가 모두 다르기 때문에 저는 동적테이블을 활용해야 한다고 생각했습니다. 따라서, 설 기간에 자바스크립트라는 언어를 공부하게 되었습니다.
3. 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.
4. 이는 클래스변수로 Json Object를 받아서 사용하였고 이렇게 변수에 저장해서 사용한 이유는
5. 면적별로 정리하여 만든 새로운 테이블을 조회할때 시간 절약을 위해 이미 스크래핑된 Json 데이터를 재사용하여 구현하기 위함 이였습니다.

실 근무일은 얼마 되지 않았지만 약 한달여간 인턴쉽을 하면서 처음에는 자책을 많이 했던것 같습니다. 이전에는 프로그래밍을 시작한지 얼마 되지않았지만 나 정도면 빠르게 성장을 하는편이라고 생각했었습니다.

하지만 이렇게 인턴으로 실무를 조금이라도 경험을 해보니 전 아는게 많이 없다는걸 많이 느꼈습니다.

그렇게 인턴과정에서 자바와 웹 프로그래밍을 경험을 해보았고, 과장님께서 말씀하신대로 저는 html 자바스크립트 스프링을 할 줄 아는 사람이 아닌, 진입장벽이 그저 낮아진 상태 라는 것을 인지하였고 제 무지함에 실망을 느꼈습니다.

하지만 인턴을 하면서 프로그래밍을 보는 시야가 그래도 조금은 넓어짐을 느꼈고 프로그래밍에 점점 더 흥미가 붙어, 자바를 이용한 웹개발과 파이썬을이용한 데이터 분석이 가능한 개발자가 되고싶다 라는 생각을 하게 되었습니다.

앞으로는 스프링을계속 공부해보려고 하고있고 웹프로그래밍쪽으로 자기계발을 할 생각입니다.