

| 자유 게시판 | >

[#1. 카톡크롤링응답] - 나도 할수 있다! 원하는 값 도출하기 1편: 환율조회

토끼는어흥
챗봇 중수
1:1 채팅

2020.02.20. 15:29
조회 163

댓글 6
URL 복사

[#1. 카톡크롤링응답] - 나도 할수 있다! 원하는 값 도출하기 1편

우선 이걸 뭐라고 정의해야 할까 하다가... 이 카페의 목적이 카카오톡봇이 목적이니 봇이 응답하는 것에 초점을 둔

크롤링 하는 방법을 시간날 때 하나씩 올려볼까 합니다.

저는 개발자도 아니며, Java, Nodejs, Python, C, C++, C# 등 소스를 볼줄만 알지 개발할 줄은 모릅니다.

그냥 관심갖다가 DarkTonardo 맞나요? 이분의 블로그의 영향을 많이 받고 이것저것 해보고 있습니다.

기회가 되면 DarkTonardo 님과 단기 프로젝트를 하나 진행하고 싶네요~

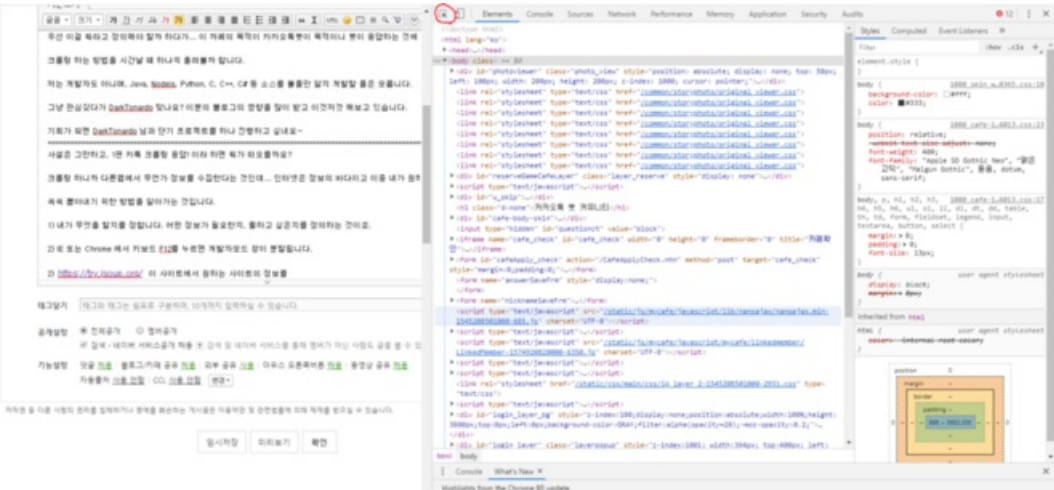
사설은 그만하고, 1편 카톡 크롤링 응답! 이라 하면 뭐가 떠오를까요?

크롤링 하니까 다른웹에서 무언가 정보를 수집한다는 것인데... 인터넷은 정보의 바다이고 이중 내가 원하는 데이터만

썩썩 뽑아내기 위한 방법을 알아가는 것입니다.

1) 내가 무엇을 할지를 정합니다. 어떤 정보가 필요한지, 뭘하고 싶은지를 정의하는 것이죠.

2) IE 또는 Chrome 에서 키보드 F12를 누르면 개발자모드 창이 분할됩니다. 아래 그림처럼 말이죠.



뭔가 굉장히 복잡해 보이지만, 네 자세히 알면 이것저것 기능들이 굉장히 많지요~

해당 페이지의 상태 코드값이나, 소스코드, 네트워크, Request/Response 알려고 하면 꽤나 많답니다.

일단, 여기서 필요한건 빨간색으로 동그라미 친 화살표가 보이지요? 해당 버튼을 클릭하고 내가 크롤링 하고 싶은 키워드를 클릭합니다.

그러면, 해당 키워드의 우측창 소스코드위치로 이동한답니다. 자, 일단 여기까지 하고,,,

카페에 글을 적으면서 캡처를 뜬 부분이고. 실제 예를 들어서 가입하면서 만든 환율정보를 표기하는 방법을 예로

들어보겠습니다.

2) 우리에게 친숙한 네이버로 가서 환율을 검색합니다.

https://finance.naver.com/marketindex/?tabSel=exchange#tab_section

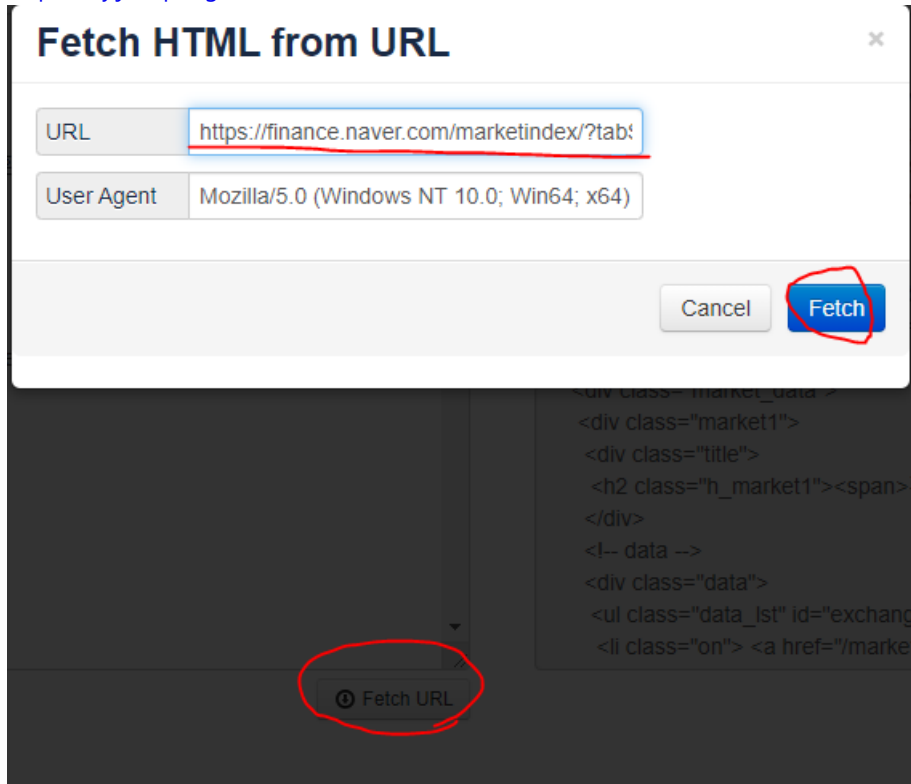
위와 같은 사이트에서 자세한 환율정보를 공시하고 있네요.

음.. 난 원달러환율 정보를 카톡에서 나오게 할테야~

F12 버튼을 누르고, 화살표 표기 아이콘 클릭 > 원달러환율 금액이 적힌 곳을 클릭해줍니다. 자, 일단 여기까지 하고~

3) Jsoup을 테스트 할 수 있는 사이트에 접속합니다.

<https://try.jsoup.org/> 이 사이트에서 원하는 사이트의 URL 을 집어넣어 사이트의 소스코드를 불러올 수 있습니다.



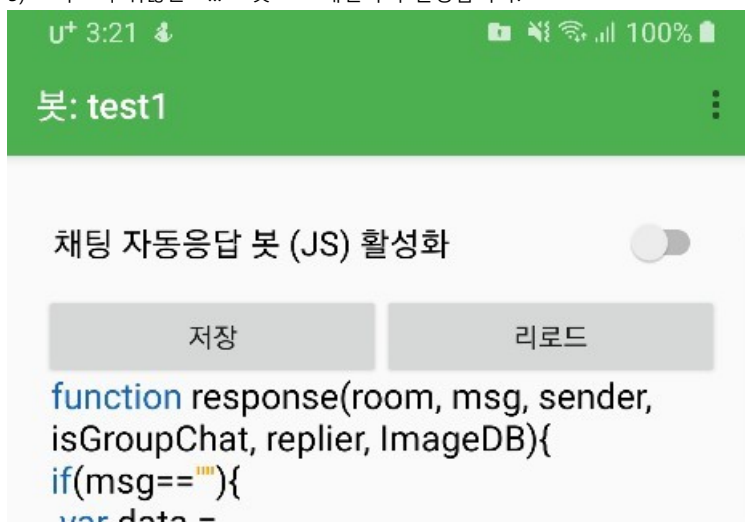
화면 가운데 Fetch URL 을 클릭하여, 2) 에서 언급한 환율공시 사이트 URL 을 넣어주고 Fetch 버튼을 클릭합니다. 그러면 소스코드를 불러오게 됩니다.

4) Select CSS query 상자에 원하는 값을 빼낼 Class 를 언급합니다.

가격을 나타내는 Class 는 공통 span 을 사용했군요. 해당 공간에 span 입력을 하니 아래로 가격 공시 span class 를 사용한 값들만 추려서 나오는군요~ 곳~

한 1/3 정도 왔습니다. 힘내세요~~

5) 쓰다보니 귀찮군요... 스샷으로 대신하여 설명합니다.



```
org.jsoup.Jsoup.connect("https://
finance.naver.com/marketindex/
?tabSel=exchange#tab_section").get().se
lect("span")+"";
//원하는 사이트 url 에서 span class 선택
data= data.replace(/<[^>]+>/g,"");
//각종 잡다한태그 제거
data = data.split("환전 고시 환율")[1];
data = data.split("KEB하나은행 기준")[0];
//어디부터 어디까지 자를지 정함
data = data.trim();
//공백제거
data = data.split("\n");
//엔터 자르기
data = data[1].trim();

replier.reply("[원달러환율]\n"+data+"원
입니다.");
//응답메시지 꾸미기
}
}
```

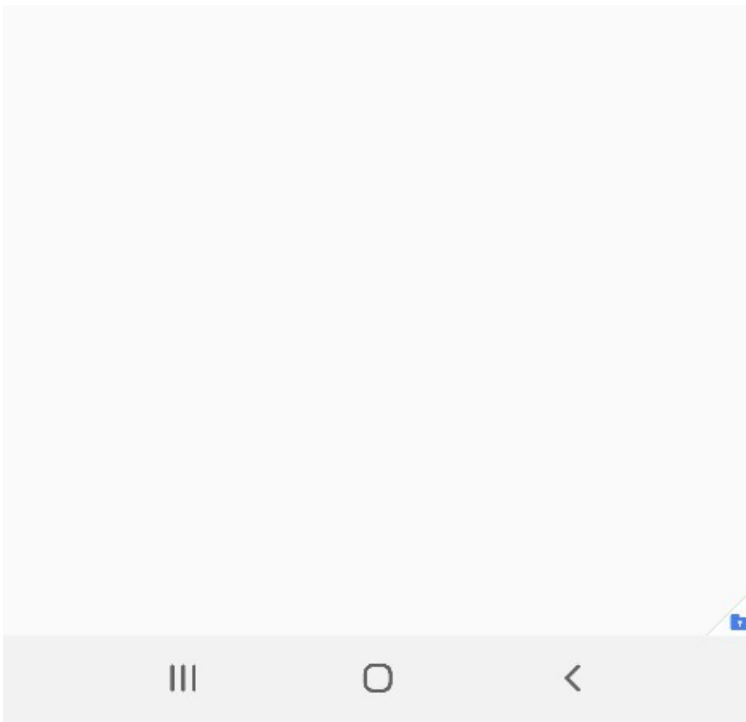
- Dark Tonardo 님의 초록봇입니다. 초록봇에서 Jsoup 으로 최대한 간결하게 코드를 작성해 봅니다.
get().select() 를 사용하여, span class 를 끄집어 냅니다.

- 사진에서와 같이 친절하게 귀차니즘을 무릅쓰고(표준어가 맞는지 모르겠네요) 주석을 달아주었습니다.

- 네... 주석과 같습니다. 하나하나 주석을 쳐가며 어떻게 변화가 되는지 직접 해보시기를 추천드립니다.

왜 그렇게 되고 왜 그렇게 되고 하는지...

- 정규표현식에 관해서는 직접 많이 해보시는게 외우는것보다 깨달음이 큽니다. 직접 부딪혀서 몇번이고 해보십시오.



디버그 룸으로 와서 if(msg=="") 안에 내가 원하는 input 값을 집어넣습니다.
당연히.. output 은 replier.reply() 에서 작동합니다.

시작은 잘 쓰려다가 점점 귀차니즘과 일을 해야하기에 여기까지 할게요~ 끝~~ 1편 끝~~

등업 좀 해주세요~

* 2편은 여러가지 기능을 하나의 스크립트로 만드는 방법? 을 설명해 드리지요~

제소스도 더러워서 최적화해야 하는데 귀차니즘으로 강 방치상태네요... -ㅅ-



토끼는어흥님의 게시글 더보기 >

❤ 좋아요 3 💬 댓글 6

🔗 공유 | 신고

댓글 등록순 최신순 C

댓글알림 ☐



토끼는어흥 작성자

소스코드는 다른 글에 올려두었으니, 참고하세요~

강 퍼다가 쓰시면 응용해서 다른것들을 못하니 직접 하나하나 따라해가며 해보시기를 추천드립니다.

2020.02.20. 15:30 답글 쓰기



토끼는어흥 작성자

이전 글이 중급소스로 이동되었네요~ 카페 규칙이 있나 봅니다. 죄송합니다. 스샷을 참고하시길~

2020.02.20. 15:31 답글 쓰기



주in

좋은 글 감사합니다~

2020.02.20. 15:37 답글 쓰기



하루나리이

오 저도 마침 공부중이었는데 도움이 많이 될 것 같습니다!
퇴근하고 정독하겠습니다!

2020.02.20. 16:02 답글 쓰기



AlphaDo

닥토상은 고였어요

2020.02.20. 23:18 답글 쓰기



시스티나8

다들 너무 고이신듯...항상 감사합니다

2020.02.27. 16:37 답글 쓰기

Hibot

댓글을 남겨보세요



등록

글쓰기

답글

목록

▲ TOP

'자유 게시판' 게시판 글

이 게시판 새글 구독하기 ☐

루팅없이 카톡 DB획득 성공! 🤖 [12]

Semicolon

2020.02.20.

이 코로나 사이트 파싱하면 매우좋을듯 [5]

bangjiwoo1

2020.02.20.

[#1. 카톡크롤링응답] - 나도 할수 있다! 원하는 값 도출하기 1편: 환율조회 🤖 [6]

토끼는어흥

2020.02.20.

질문있어용.. [4]

손부장

2020.02.20.

가입했습니다. 반가워요~ [6]

토끼는어흥

2020.02.20.

1 2 3

전체보기

이 카페 인기글

[치**, 한**] 님 강제 탈퇴

AlphaDo

♡1 💬2



틱택토 (Player vs Player)



틱택토 (Player vs Bot)

반가워요.

천방지축하연

♡0 💬4

kakao.py에 대해서

undefined

♡0 💬13

오픈이발 괴롭히기(메봇R)

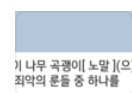
ERROR

♡0 💬7

가르치기 리로드..

Milk2

♡0 💬9



비 나무 국경에(노말)(의
최악의 문들 중 하나를

5000분의 1을 얻은 사람

북채/그림책 구분

시초는중국발

♡0 💬8

