기차역

url :

<http://info.korail.com/mbs/www/jsp/station/station_list.jsp?id=www_020110020000&category=1>



텍스트이(가) 표시된 사진

자동 생성된 설명url 맨 뒤의 숫자가 데이터 페이지와 연결됨



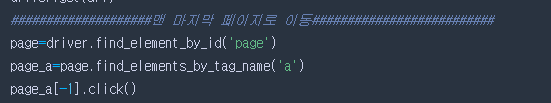
url 뒤의 숫자로 페이지 넘어감



다음 페이지 버튼:

1페이지에서 누르면 2로 넘어가는게 아니라 10페이지로 넘어감

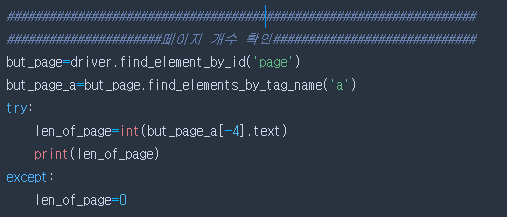
해결하기 위해 맨 마지막 페이지부터 앞으로 넘어오면서 크롤링



맨 마지막 페이지 이동 후 페이지 수 계산 -> 반복 회수 결정



맨 마지막 페이지로 이동 후 맨 마지막 페이지 수는 가져올 수 없다(13페이지) -> 그 다음 페이지 기준으로 페이지 수 계산 (12페이지)

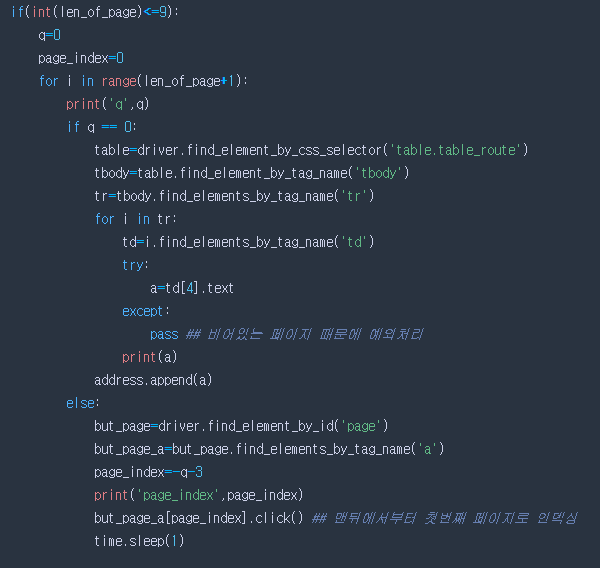


페이지 수가 0일 때 a 태그의 개수가 안 맞아서 인덱스 오류 발생-> try except로 에러 발생 방지

페이지 수가 10페이지 이상인 경우와 아닌 경우로 나누어서 코딩-> 왜?

 10페이지 이상인 경우는 이전 페이지 버튼을 한번 눌러야 하기 때문에

페이지수가 10보다 작을 때

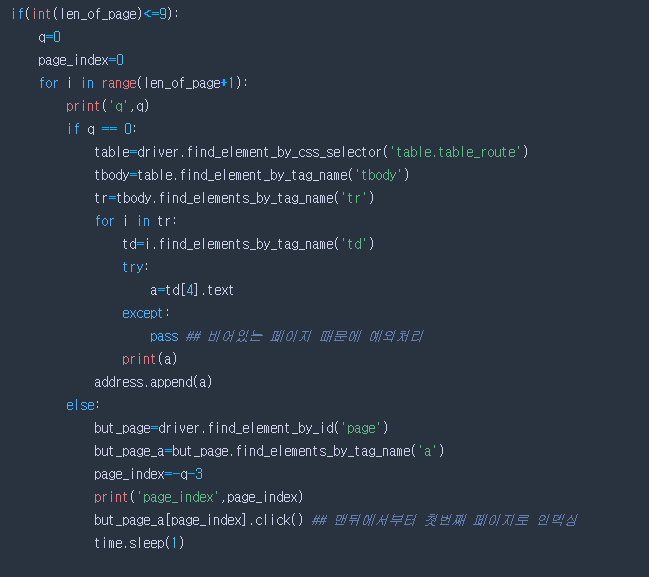


제어 변수 q: 변수 i와 함께 0부터 페이지 수 만큼 증가하면서 반복 회수 제어

왜 i 와 별도로 설정? ->10페이지 넘어가는 경우일 때 1~10 페이지 읽을 때와 11~13 페이지 읽을 때의 경우를 나누기 위하여

Q=0 즉 맨 마지막 페이지를 if 로 나누었음 이걸 나눠주지 않으면 맨 첫 페이지 데이터 읽기 전에 다음 페이지로 넘어감.

밑에 설명 계속



Page index 변수

이 사이트는 뒤로 가기 버튼을 누르면 한 페이지 전으로 가는게 아니라 맨 마지막으로 간다 그래서 index를 직접 이용했다

이때 page index는 변수 q를 이용하여 제어한다.

.데이터 가져오기

테이블 안에 들어있다

Table->tbody->tr->td 순으로 접근한다.



-3

-4

-5

-6

-7

-8

-9

-10

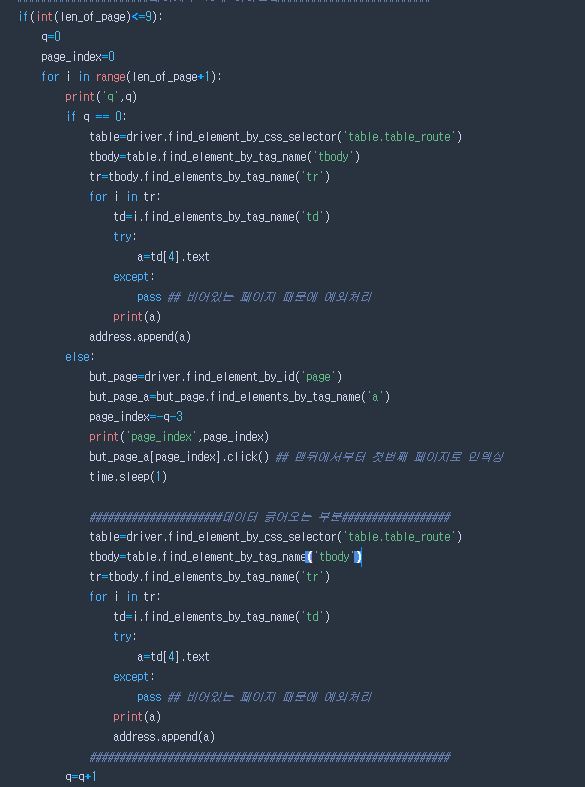
-11

-12



버튼 array indexing

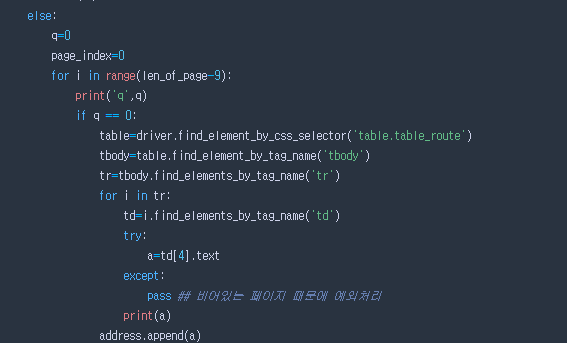
페이지 10보다 작을 경우 전체 코드



페이지수 10보다 클 때->

이전 페이지 버튼 누르기 전과 후로 코딩

1. 이전 페이지 버튼 누르기 전



1. 페이지 인덱싱



-3

-4

-5

위에서 설명한대로 마지막 페이지 다음 페이지 수 즉 12 페이지를 읽어온다.

len\_of\_page 변수에는 12 가 저장된다.

왜 len\_of\_page – 9 ? : 11페이지 찍고 이전 페이지 버튼을 눌러야 에러가 안생긴다. 즉 상단의 페이지 array를 예로 들면 13 ,12,11 총 3번의 array를 누르고 뒤로기 버튼을 누르도록 코딩하였다.

데이터 가져오는 부분은 모두 동일하다.

1. 이전 페이지 누른 후

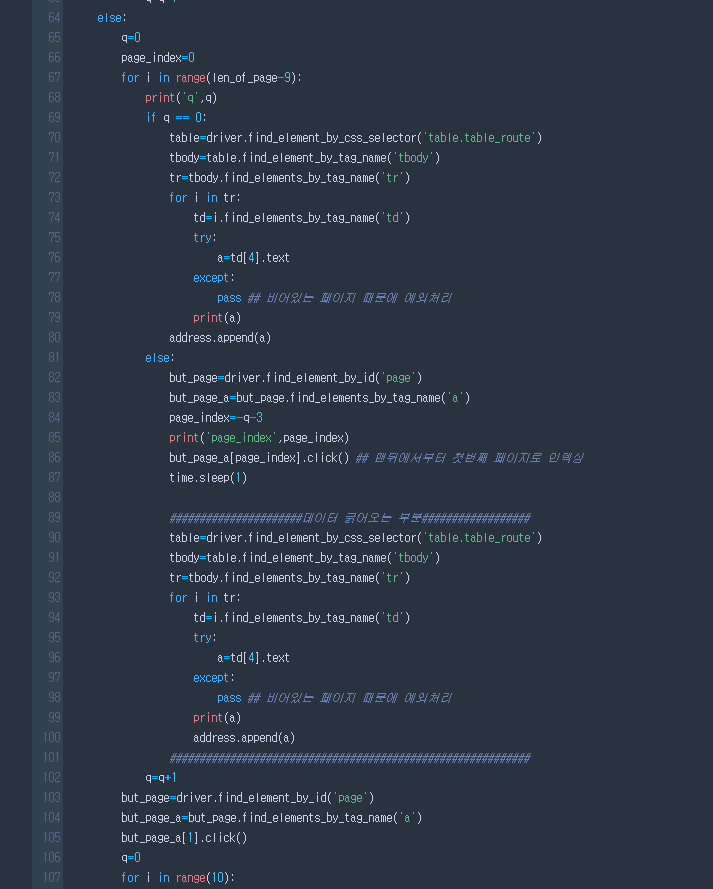
텍스트이(가) 표시된 사진

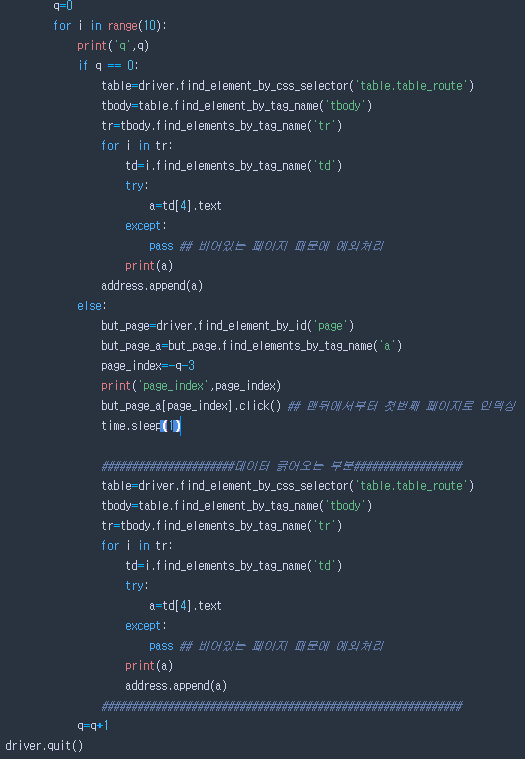
자동 생성된 설명

이전 페이지를 누르는 부분이다



이 상태에서는 10페이지 이하인 경우와 동일하게 진행한다.

페이지 11~1N 읽어오는 부분 전체 코드

페이지 1~10 읽어오는 부분 전체 코드