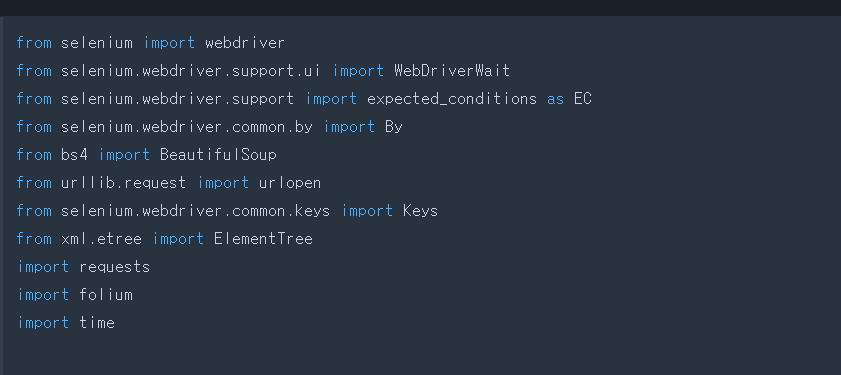
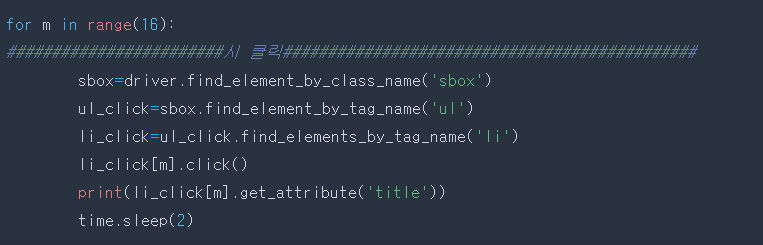
터미널

사용 라이브러리

URL: <https://www.bustago.or.kr/newweb/kr/terminalInfo/terminallist.do#>

16개 광역시/도 에 대해서 한 개씩 클릭하며 위치정보 수집

광역시도 데이터는 sbox 클래스안에 List 형태로 존재한다.

LIst인덱스를 0에서부터 1씩 증가 시켜가며 모든 광역시도에 대한 데이터 수집

광역시/도를 선택한 후 정보를 가져온다.

문제점

1.10번째 페이지에서 다음 페이지로 넘어가는 버튼과 1~10번쨰 페이지 넘어가는 버튼이 다르다.

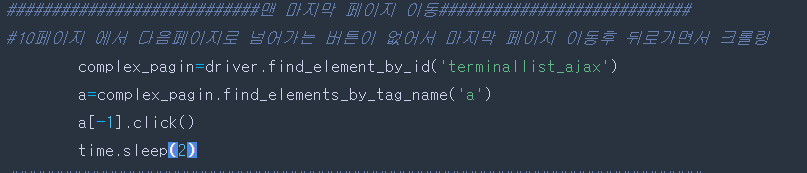


2. 오른쪽의 버튼을 클릭하면 다음 페이지가 아닌 마지막 페이지로 넘어간다.

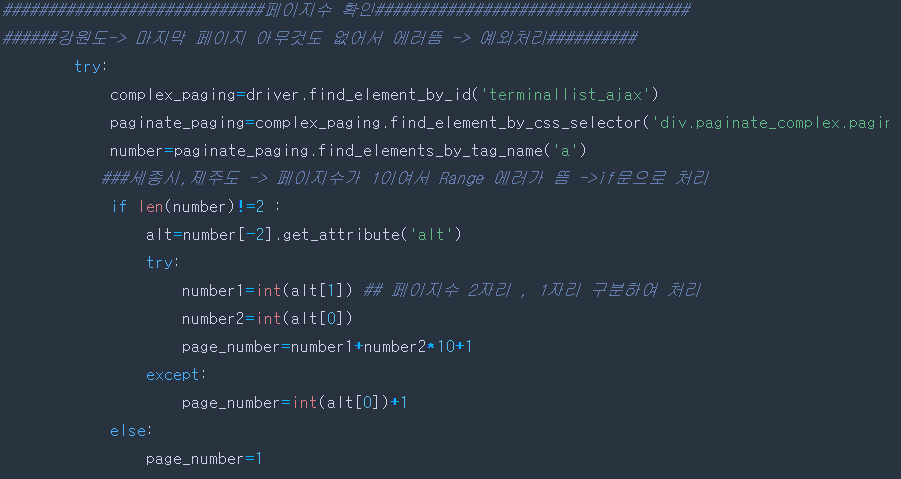
->해결방안 : 처음부터 맨 마지막 페이지로 이동 후 앞으로 가면서 데이터 수집



버튼들은 a 태그에 존재한다. 이를 list로 받아와서 맨 마지막 페이지 버튼 인덱스를 이용하여 맨 마지막 페이지로 이동한다.



맨 마지막 페이지로 이동 후 페이지의 수를 확인하여 반복 회수를 결정한다.



페이지 수 1인 경우 :

A 태그의 개수가 적어서 indexing 오류 발생 -> if else로 처리

페이지 수가 10페이지 이상인 경우

다음과 같이 “42페이지 입니다” 에서 앞에 숫자만 가져온다.

(get\_attribute(‘alt’) 이용 -> alt 값 가져옴)

10페이지 이하인 경우 : [(페이지수), ‘페’]

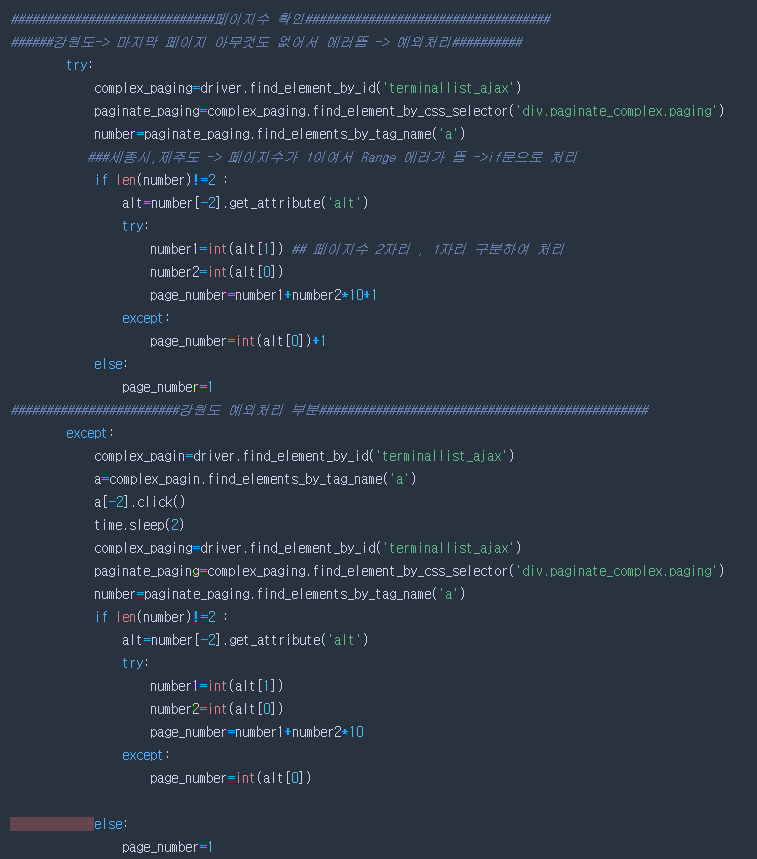
10페이지 이상인 경우:[(페이지수 10의자리),(페이지수 1의자리)]

다음과 같은 형식으로 데이터를 가져오기 때문에 2번째 데이터가 정수 변환 가능하면 10페이지 이상, 정수 변환 불가능 하면 10페이지 미만이다. ->try except 로 구분

10페이지 이상 이하 구분 후 , 페이지 수를 계산한다.

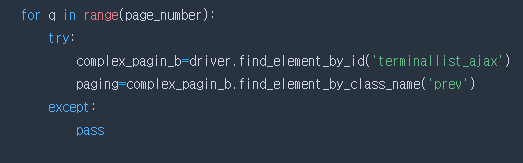
마지막 페이지에 아무 내용도 없는 경우에 에러가 뜬다

Try,except 를 이용해서 페이지에 아무 내용이 없어도 페이지 수를 가져오게 만들었다.



페이지수 가져오는 부분 전체 코드

* 강원도 예외 처리부분 좀더 짧게 만들 수 있는데 귀찮아서 복붙했습니다.

페이지수 만큼 반복하여 데이터 가져오기

변수 q를 이용하여 반복 제어

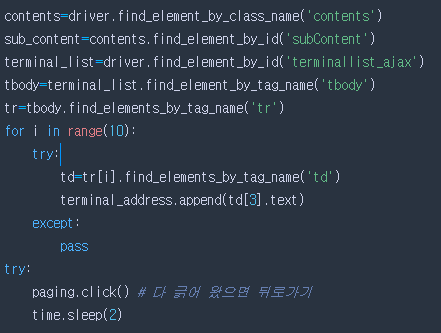
데이터 가져오기 전에 paging 변수를 이용해서 이 버튼을 먼저 찾아 놓음

->try except 사용한 이유:

역순으로 페이지 인덱스를 이용하며 데이터를 긁어오는 과정에서 맨 첫번째 페이지에 도착하면 이 버튼은 존재하지 않는다 -> 오류발생



데이터는 테이블 형식으로 들어있다 table->tbody->tr->td 순으로 데이터를 읽어온다



for i in range(10):

위에 표를 보면 데이터는 한 페이지에 10개씩 들어있다

하지만 맨 마지막 페이지에는 10개씩 들어있지 않은 경우가 존재한다

정확하게 하는 건 td의 개수로 반복회수를 결정하는 것이지만

10번 돌리고 try except를 써서 에러 발생을 방지했다.