[python] Selenium으로 스크래핑하기

​

1. Selenium으로 DOM요소 선택 - 요소를 찾지 못하면 NoSuchElementException 발생

|  |  |
| --- | --- |
| 이름 | 설명 |
| 처음요소를 추출 | |
| find\_element\_by\_id(id) | id속성으로 요소를 하나 추출 |
| find\_element\_by\_name(name) | name 속성으로 요소를 하나 추출 |
| find\_element\_by\_css\_selector(query) | css 선택자로 요소를 하나 추출 |
| find\_element\_xpath(query) | xpath를 지정해 요소를 하나 추출 |
| find\_element\_by\_tag\_name(name) | 태그 이름이 name에 해당하는 요소를 하나 추출 |
| find\_element\_by\_link\_text(text) | 링크 텍스트로 요소를 추출 |
| find\_element\_by\_partial\_link\_text(text) | 링크의 자식 요소에 포함되 있는 텍스트로 요소를 하나 추출 |
| find\_element\_by\_class\_name(name) | 클래스 이름이 name에 해당하는 요소를 하나 추출 |
| 모든 요소를 추출 | |
| find\_element\_by\_css\_selector(query) | css 선택자로 요소를 여러개 추출 |
| find\_element\_by\_xpath(query) | xpath를 지정해 요소를 여러개 추출 |
| find\_element\_by\_tag\_name(name) | 태그이름이 name에 해당하는 요소를 여러개 추출 |
| find\_element\_by\_class\_name(name) | 클래스 이름이 name에 해당하는 요소를 여러개 추출 |
| find\_element\_by\_partial\_link\_text(text) | 링크의 자식 요소에 포함돼 있는 텍스트로 요소를 여러개 추출 |

​

2. Selenium으로 요소르르 조작하기

|  |  |
| --- | --- |
| 메서드/ 속성 | 설명 |
| clear() | 글자를 지운다 |
| click() | 요소를 클릭 |
| get\_attribute(name) | 요소 속성중 name에 해당하는 속성 값을 추출 |
| is\_displayed() | 요소가 화면에 출력되는지 확인 |
| is\_enabled() | 요소가 활성화돼 있는지 확인 |
| is\_selected() | 체크박스 등의 요소가 선택된 상태인지 확인 |
| screenshot(filename) | 스크린샷 |
| send\_keys(value) | 키를 입력 |
| submit() | 입력 양식을 전송 |
| value\_of\_css\_property(name) | name에 해당하는 css속성 값을 추출 |
| id | id |
| location | 요소의 위치 |
| parent | 부모요소 |
| rect | 크기와 위치 정보를 가진 사전자료형 리턴 |
| screenshot\_as\_base64 | 스크린샷을 base64로 추출 |
| screenshot\_as\_png | 스크린샷을 png형식의 바이너리로 추출 |
| size | 요소의 크리 |
| tag\_name | 태그 이름 |
| text | 요소의 내부 글자 |

​

3. send\_key()에서 특수키 입력

from selenium.Webdriver.common.keys import Keys

ARROW\_DOWN / ARROW\_LEFT / ARROW\_RIGHT / ARROW\_UP

BACKSPACE / DELETE / HOME / END /INSERT /

ALT / COMMAND / CONTROL / SHIFT

ENTER / ESCAPE /SPACE / TAB

F1 / F2 / F3 ............./ F12

​

​

4. Selenium 드라이버 조작

|  |  |
| --- | --- |
| add\_cookie( cookie\_dict) | 쿠키값을 사전 형식으로 지정 |
| back() / forward() | 이전 페이지/ 다음페이지 |
| close() | 브라우저 닫기 |
| current\_url | 현재 url |
| delete\_all\_cookies() | 모든 쿠키 제거 |
| delete\_cookie(name) | name에 해당하는 쿠키 제거 |
| execute( command, params) | 브라우저 고유의 명령어 실행 |
| execute\_async\_script( script, \*args) | 비동기 처리하는 자바스크립트를 실행 |
| execute\_script( script, \*args) | 동기 처리하는 자바스크립트를 실행 |
| get(url) | 웹 페이지를 읽어들임 |
| get\_cookie( name) | 특정 쿠키 값을 추출 |
| get\_cookies() | 모든 쿠키값을 사전 형식으로 추출 |
| get\_log(type) | 로그를 추출(type: browser/driver/client/server) |
| get\_screenshot\_as\_base64() | base64형식으로 스크린샷을 추출 |
| get\_screenshot\_as\_file(filename) | 스크린샷을 파일로 저장 |
| get\_screenshot\_as\_png() | png형식으로 스키란샷의 바이너리를 추출 |
| get\_window\_position(windowHandle='current') | 브라우저의 위치를 추출 |
| get\_window\_size( windowHandle='current') | 브라우저의 크기를 추출 |
| implicitly\_wait(sec) | 최대 대기 시간을 초 단위로 지정해서 처리가 끈날때 까지 대기 |
| quit() | 드라이버를 종료 시켜 브라우저 닫기 |
| save\_screenshot(filename) | 스크린샷 저장 |
| set\_page\_load\_timeout( time\_to\_wait) | 페이지르르 읽는 타임아웃 시간을 지정 |
| set\_script\_timeout(time\_to\_wait) | 스크립트의 타임아웃 시간을 지정 |
| set\_window\_position(x,y,windowHandle='current') | 브라우저 위치를 지정 |
| set\_window\_size(width, height, windowHandle='current') | 브라우저 크기를 지정 |
| title | 현재 타이틀을 추출 |

​

출처: <https://fenderist.tistory.com/168> [Devman]