[특별 부록] 1권 핵심 라이브러리 총정리

> 이 페이지는 우리가 1권에서 사용한 핵심 라이브러리(`Selenium`, `Beautiful Soup`, `pandas`)의 주요 메서드들을 정리한 '치트 시트(Cheat Sheet)'입니다. 앞으로 크롤러를 만들 때마다 자주 찾아보게 될 내용이니, 확실하게 복습하고 넘어가시길 바랍니다.

1. `Selenium` - 웹 브라우저 조종사

- `driver = webdriver.Chrome(service=service)`
  - Chrome 브라우저를 실행하고, 조종할 수 있는 `driver` 객체를 생성합니다.
- `driver.get(URL)`
  - `URL` 주소의 웹페이지로 이동시킵니다.
- `driver.quit()`
  - 실행했던 브라우저를 완전히 종료합니다.
- `driver.find_element(By.TYPE, "VALUE")`
  - 페이지에서 조건에 맞는 **첫 번째 요소 하나**를 찾습니다. (`By.CSS_SELECTOR`를 가장 많이 사용합니다.)
- `element.click()`
  - 찾아낸 요소(`element`)를 마우스로 클릭합니다.
- `element.send_keys("TEXT")`
  - 찾아낸 요소(주로 입력창)에 `TEXT`를 키보드로 입력합니다.
- `driver.execute_script("JAVASCRIPT_CODE")`
  - 현재 페이지에서 `JAVASCRIPT_CODE`를 직접 실행합니다. (주로 스크롤링에 사용합니다.)
- `driver.page_source`
  - 현재 브라우저에 렌더링된 **페이지의 전체 HTML 소스 코드를 문자열로** 반환합니다.

2. `Beautiful Soup` - HTML 요리사

- `soup = BeautifulSoup(html, "lxml")`
  - HTML 문자열(`html`)을 `lxml` 파서를 이용해, 탐색하기 쉬운 `soup` 객체로 변환합니다.
- `soup.select("CSS_SELECTOR")`
  - `CSS_SELECTOR` 조건에 맞는 **모든 요소를 리스트** 형태로 반환합니다.
- `soup.select_one("CSS_SELECTOR")`
  - `CSS_SELECTOR` 조건에 맞는 **첫 번째 요소 하나**를 반환합니다.
- `element.text`
  - 찾아낸 요소(`element`)의 HTML 태그를 모두 제거하고, **안에 있는 순수한 텍스트**만 추출합니다.
- `element['attribute_name']`
  - 찾아낸 요소(`element`)의 특정 속성(`attribute_name`) 값을 가져옵니다. (예: `card['href']`)

3. `pandas` - 데이터 정리 전문가

- `df = pd.DataFrame(list_of_dictionaries)`
  - 딕셔너리들이 담긴 리스트를 엑셀 시트와 같은 `DataFrame` 객체로 변환합니다.
- `df.to_csv("filename.csv", index=False, encoding='utf-8-sig')`

- `DataFrame` 객체를 `filename.csv` 라는 이름의 CSV 파일로 저장합니다.
  - `index=False`: 불필요한 행 번호를 저장하지 않습니다.
  - `encoding='utf-8-sig'`: 한글 깨짐을 방지합니다.