

Lecture 08

Selenium for Web Crawling

1. Web Crawling 원리

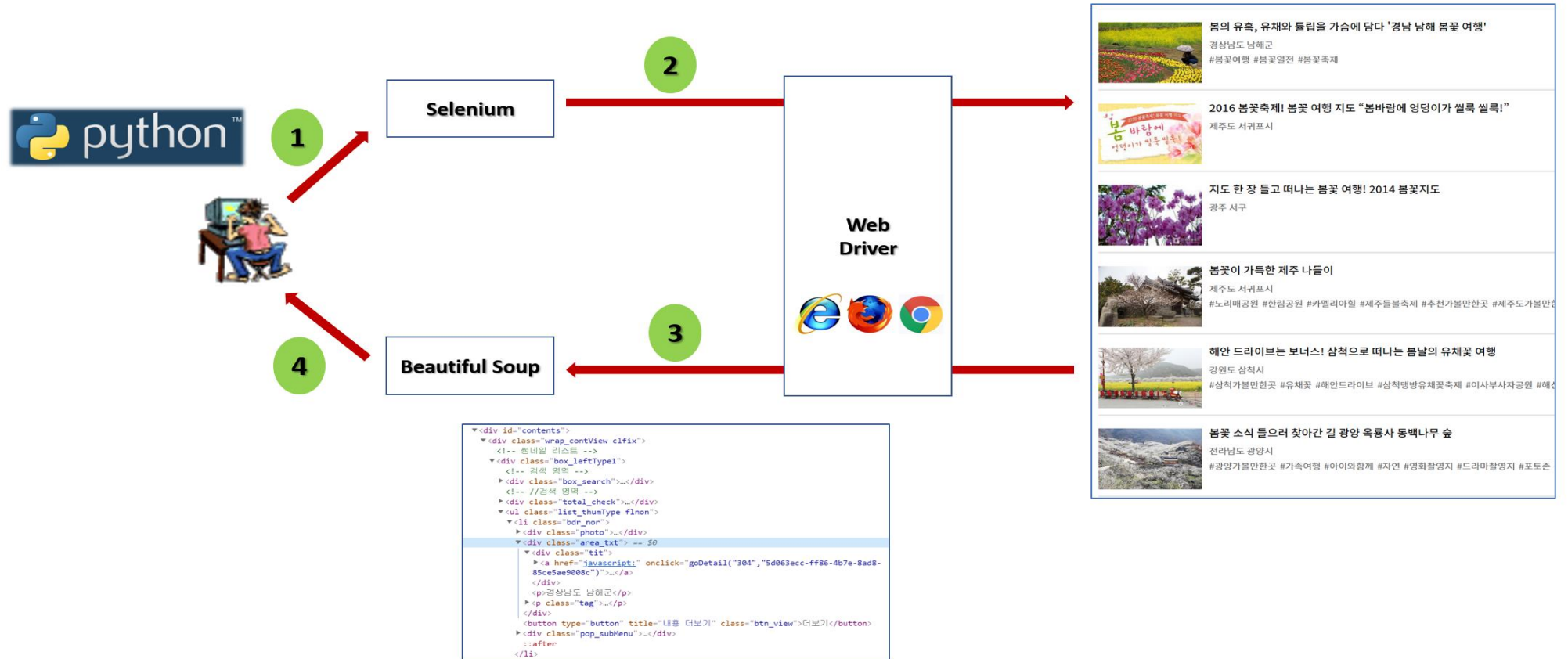
2. selenium 설치 및 확인

3. Chrome web driver 설치 및 확인

4. 한국관광공사 홈페이지 접속 크롤러 구현 실습

5. selenium에서 제공되는 함수

6. 네이버 접속 크롤러 구현 실습



1. Web Crawling 원리
2. selenium 설치 및 확인
3. Chrome web driver 설치 및 확인
4. 한국관광공사 홈페이지 접속 크롤러 구현 실습
5. selenium 에서 제공되는 함수
6. 네이버 접속 크롤러 구현 실습

■ Anaconda Navigator 활용

The screenshot shows the Anaconda Navigator application window. The left sidebar contains navigation options: Home, Environments, Learning, and Community. The main panel displays the 'base (root)' environment. A search bar at the top of the main panel is used to search for packages. The search results are displayed in a table with columns: Name, Description, and Version. The table shows three packages: pytest-selenium (2.0.1), r-seleniumpipes (0.3.7), and selenium (3.141.0). The selenium package is highlighted with a green checkmark in the Name column. A yellow callout bubble points to the selenium package with the text 'selenium 확인'.

Name	Description	Version
<input type="checkbox"/> pytest-selenium		2.0.1
<input type="checkbox"/> r-seleniumpipes		0.3.7
<input checked="" type="checkbox"/> selenium		3.141.0

3 packages available matching "selenium"

selenium 확인

1. Web Crawling
원리

2. selenium 설치
및 확인

3. Chrome web
driver 설치 및
확인

4. 한국관광공사
홈페이지 접속
크롤러 구현
실습

5. selenium
에서 제공되는
함수

6. 네이버 접속
크롤러 구현
실습

- conda 또는 pip 활용하여 BeautifulSoup4 패키지 설치

```
(base) C:\Users\tina> conda install selenium
```

```
(base) C:\Users\tina> pip install selenium
```

1. Web Crawling 원리

2. selenium 설치 및 확인

3. Chrome web driver 설치 및 확인

4. 한국관광공사 홈페이지 접속 크롤러 구현 실습

5. selenium 에서 제공되는 함수

6. 네이버 접속 크롤러 구현 실습

https://www.selenium.dev/documentation/ko/introduction/the_selenium_project_and_tools/

- 셀레니움(Selenium)은 웹 브라우저를 자동화할 수 있게 돕는 일련의 도구 모음인 엄브렐라 프로젝트(umbrella project) 중 일부입니다.
- Selenium 은 여러 프로젝트를 포함하지만, **웹 브라우저 자동화를 위한 도구가 핵심**입니다.
- 브라우저 인스턴스를 **원격으로 제어**하고, 사용자와 브라우저와의 상호작용을 에뮬레이트 하기 위한 최고의 기술들이 적용되어 있습니다.
- 최종 사용자가 수행하는 일반적인 활동들을 시뮬레이션 할 수 있습니다.
- 필드에 텍스트 입력, 드롭 다운 및 체크 박스 선택, 링크 클릭, 마우스 이동, 임의의 JavaScript 실행 등의 기능이 제공됩니다.
- 웹 사이트의 **프론트엔드(front-end) 테스트**에 주로 사용되지만, Selenium은 브라우저 사용자 에이전트(agent) 라이브러리의 핵심입니다.
- 인터페이스는 응용 프로그램에 맞게 어디서나 사용될 수 있도록 최적화 되어있고, **목적에 맞게 다른 라이브러리와의 구성을 장려**합니다.

1. Web Crawling
원리

2. selenium 설치
및 확인

3. Chrome web
driver 설치 및
확인

4. 한국관광공사
홈페이지 접속
크롤러 구현
실습

5. selenium
에서 제공되는
함수

6. 네이버 접속
크롤러 구현
실습

<https://www.selenium.dev/>

공식 홈페이지

<https://www.selenium.dev/history/>

Selenium 역사

이야기는 2004년에 시작됩니다.

이야기는 2004년 시카고의 ThoughtWorks에서 시작됩니다. Jason Huggins는 내부 시간 및 비용 응용 프로그램(Python, Plone) 테스트를 위해 "JavaScriptTestRunner"로 핵심 모드를 구축했습니다. 이 컨설팅의 Agile 성향을 고려할 때 모든 애플리케이션의 자동 테스트는 ThoughtWork 스타일의 핵심입니다. 그는 Paul Gross와 Jie Tina Wang의 도움을 받았습니다. 그들에게 이것은 하루 일과였습니다.

Jason은 여러 동료에게 테스트 도구를 시연하기 시작했습니다. 많은 사람들이 즉각적이고 직관적인 시각적 피드백과 다른 웹 애플리케이션을 위한 재사용 가능한 테스트 프레임워크로 성장할 수 있는 잠재력에 흥분했습니다.

<https://support.google.com/chrome/answer/95414?co=GENIE.Platform%3DDesktop&hl=ko>

1. Web Crawling
원리

2. selenium 설치
및 확인

3. Chrome web
driver 설치 및
확인

4. 한국관광공사
홈페이지 접속
크롤러 구현
실습

5. selenium
에서 제공되는
함수

6. 네이버 접속
크롤러 구현
실습

Chrome 업데이트

최신 보안 업데이트로 사용자를 보호하기 위해 Chrome은 기기에서 사용 가능한 브라우저의 새 버전이 출시될 때마다 자동으로 업데이트됩니다. 업데이트로 인해 브라우저가 다르게 보이는 경우도 있습니다.

컴퓨터

Android

iPhone/iPad

새 버전이 출시될 때마다 Chrome 업데이트

일반적으로 컴퓨터의 브라우저를 닫았다가 다시 열면 백그라운드에서 업데이트가 진행됩니다. 한동안 브라우저를 닫지 않은 경우 대기 중인 업데이트가 표시될 수도 있습니다.

1. 컴퓨터에서 Chrome을 엽니다.
2. 오른쪽 상단에서 더보기 : 를 찾습니다.
3. 업데이트가 대기 중인 경우 아이콘이 다음 색상으로 표시됩니다.
 - **녹색**: 업데이트가 출시된 지 2일이 지나지 않았습니다.
 - **주황색**: 업데이트가 출시된 지 4일 정도 지났습니다.
 - **빨간색**: 업데이트가 출시된 지 1주일 이상 지났습니다.

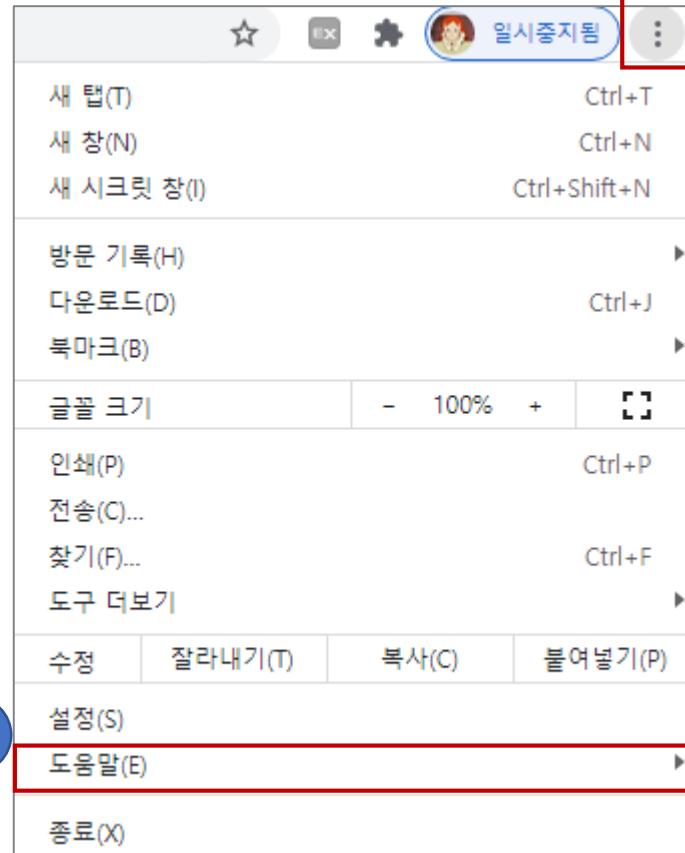
Chrome을 업데이트하려면 다음 단계를 따르세요.

1. 컴퓨터에서 Chrome을 엽니다.
2. 오른쪽 상단에서 더보기 : 를 클릭합니다.
3. **Chrome 업데이트**를 클릭합니다.
 - **중요**: 최신 버전을 사용하고 있다면 이 버튼이 표시되지 않습니다.
4. **다시 실행**을 클릭합니다.

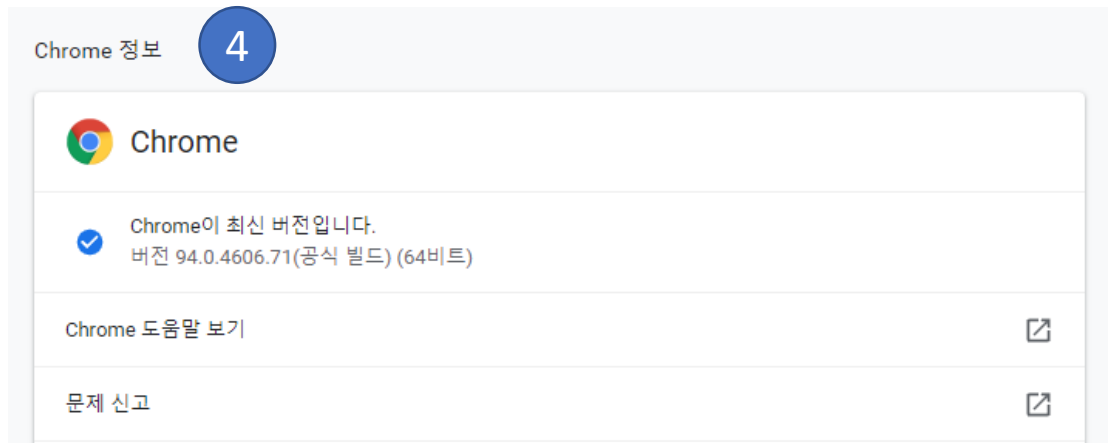
열려 있던 탭과 창이 브라우저에 저장되었다가, 브라우저가 다시 시작할 때 자동으로 다시 열립니다. Chrome을 다시 시작해도 시크릿 창은 다시 열리지 않습니다. 바로 다시 시작하고 싶지 않은 경우 **나중에 하기**를 클릭합니다. 그러면 다음에 브라우저를 다시 시작할 때 업데이트가 적용됩니다.

1. Web Crawling 원리
2. selenium 설치 및 확인
3. Chrome web driver 설치 및 확인
4. 한국관광공사 홈페이지 접속 크롤러 구현 실습
5. selenium 에서 제공되는 함수
6. 네이버 접속 크롤러 구현 실습

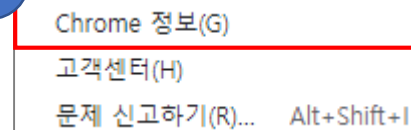
chrome web browser 버전 확인



1



3



1. Web Crawling
원리
2. selenium 설치
및 확인
3. Chrome web
driver 설치 및
확인
4. 한국관광공사
홈페이지 접속
크롤러 구현
실습
5. selenium
에서 제공되는
함수
6. 네이버 접속
크롤러 구현
실습

참고

selenium이 사용할 컴퓨터에 chrome driver와 더불어
우리가 직접 사용하는 chrome web browser가 설치되어 있어야 합니다.

만약 chrome web browser가 설치되어 있지 않을 경우에는
selenium 라이브러리를 사용할 때

[“selenium.common.exceptions.WebDriverException: Message: unknown error: cannot find Chrome binary”](#)

라는 오류가 나오면서 chrome driver가 실행이 안될 수 있습니다.

1. Web Crawling
원리
2. selenium 설치
및 확인
3. Chrome web
driver 설치 및
확인
4. 한국관광공사
홈페이지 접속
크롤러 구현
실습
5. selenium
에서 제공되는
함수
6. 네이버 접속
크롤러 구현
실습

selenium 패키지가 사용할 Web driver 설치

Step 1. <https://sites.google.com/a/chromium.org/chromedriver/downloads>

접속하기

Downloads

**Please note that we have migrated to a new
ChromeDriver site.
Current site will be deprecated soon.**

Current Releases

- If you are using Chrome version 94, please download [ChromeDriver 94.0.4606.41](#)
- If you are using Chrome version 93, please download [ChromeDriver 93.0.4577.63](#)
- If you are using Chrome version 92, please download [ChromeDriver 92.0.4515.107](#)
- For older version of Chrome, please see below for the version of ChromeDriver that supports it.

If you are using Chrome from Dev or Canary channel, please following instructions on the [ChromeDriver Canary](#) page.

For more information on selecting the right version of ChromeDriver, please see the [Version Selection](#) page.

ChromeDriver 94.0.4606.41







Supports Chrome version 94

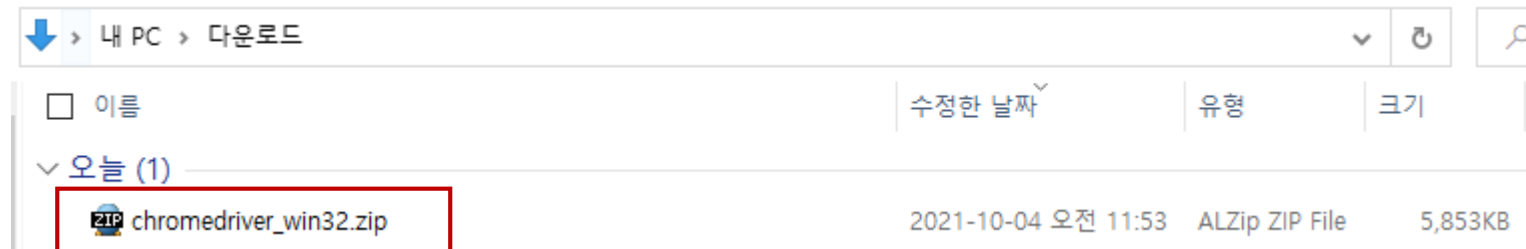
For more details, please see the [release not](#)

1. Web Crawling
원리
2. selenium 설치
및 확인
3. Chrome web
driver 설치 및
확인
4. 한국관광공사
홈페이지 접속
크롤러 구현
실습
5. selenium
에서 제공되는
함수
6. 네이버 접속
크롤러 구현
실습

Step 2. 94 버전의 화면에서 OS 환경에 적합한 크롬 드라이버를 선택하여 다운로드하기

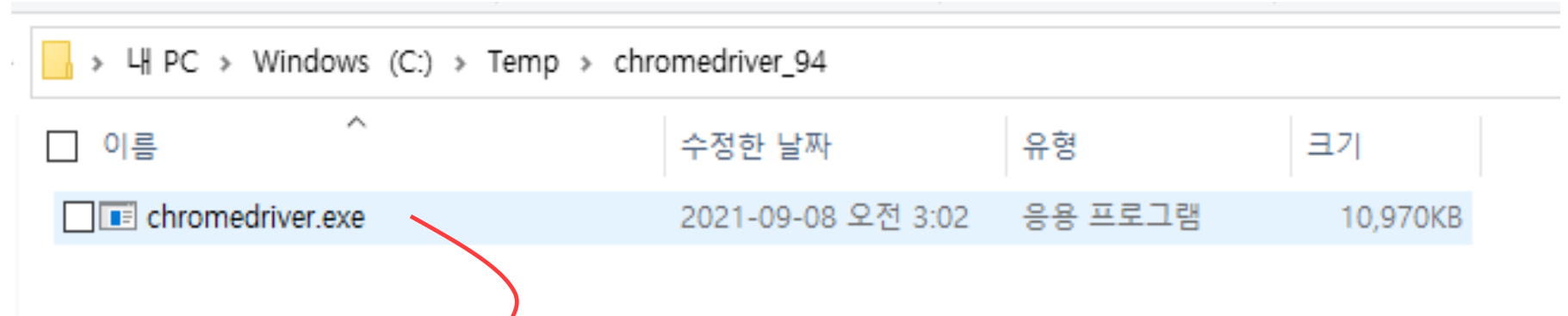
Index of /94.0.4606.41/

	Name	Last modified	Size	ETag
	Parent Directory	-	-	-
	chromedriver linux64.zip	2021-09-09 09:57:33	9.42MB	9a00e20a5e5646c009c0ad74a446df10
	chromedriver mac64.zip	2021-09-09 09:57:36	7.80MB	f89f291ccc799a9de32e832807ce7e42
	chromedriver mac64_m1.zip	2021-09-09 09:57:38	7.15MB	69b92932297c2ce2764aa4c7744e1c17
	chromedriver win32.zip	2021-09-09 09:57:41	5.72MB	c853398ecdb35290e60e57a833e61752
	notes.txt	2021-09-09 09:57:45	0.00MB	b88b7fe40b1af6891a9f57da13e08914



1. Web Crawling 원리
2. selenium 설치 및 확인
3. Chrome web driver 설치 및 확인
4. 한국관광공사 홈페이지 접속 크롤러 구현 실습
5. selenium 에서 제공되는 함수
6. 네이버 접속 크롤러 구현 실습

Step 3. 원하는 디렉토리에 압축 풀고, 기억하기 편한 디렉토리명으로 변경하기



크롬 드라이버 버전을 기억할 수 있도록 위와 같이 디렉토리 이름에 버전을 기록하면 편리함

chromedriver_94 디렉토리 안에 chromedriver.exe 파일이 있음

chromedriver.exe

1. Web Crawling
원리

2. selenium 설치
및 확인

3. Chrome web
driver 설치 및
확인

4. 한국관광공사
홈페이지 접속
크롤러 구현
실습

5. selenium
에서 제공되는
함수

6. 네이버 접속
크롤러 구현
실습

■ 실습 노트 참고

- 08_(한국관광공사) 강의용 Selenium for Web Crawling.ipynb

<https://korean.visitkorea.or.kr/main/main.html>

한국 관광 공사

1. Web Crawling
원리

2. selenium 설치
및 확인

3. Chrome web
driver 설치 및
확인

4. 한국관광공사
홈페이지 접속
크롤러 구현
실습

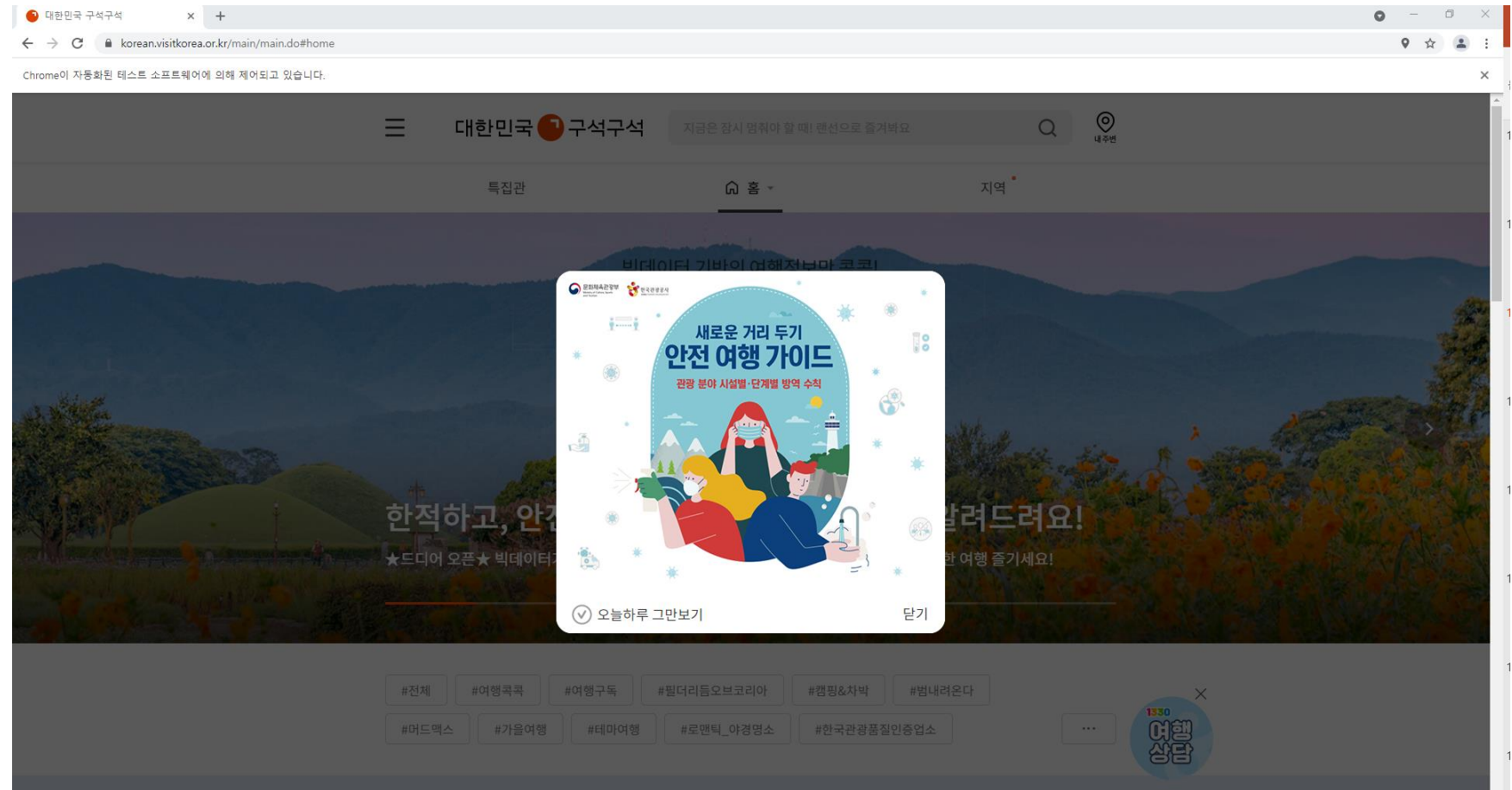
5. selenium
에서 제공되는
함수

6. 네이버 접속
크롤러 구현
실습

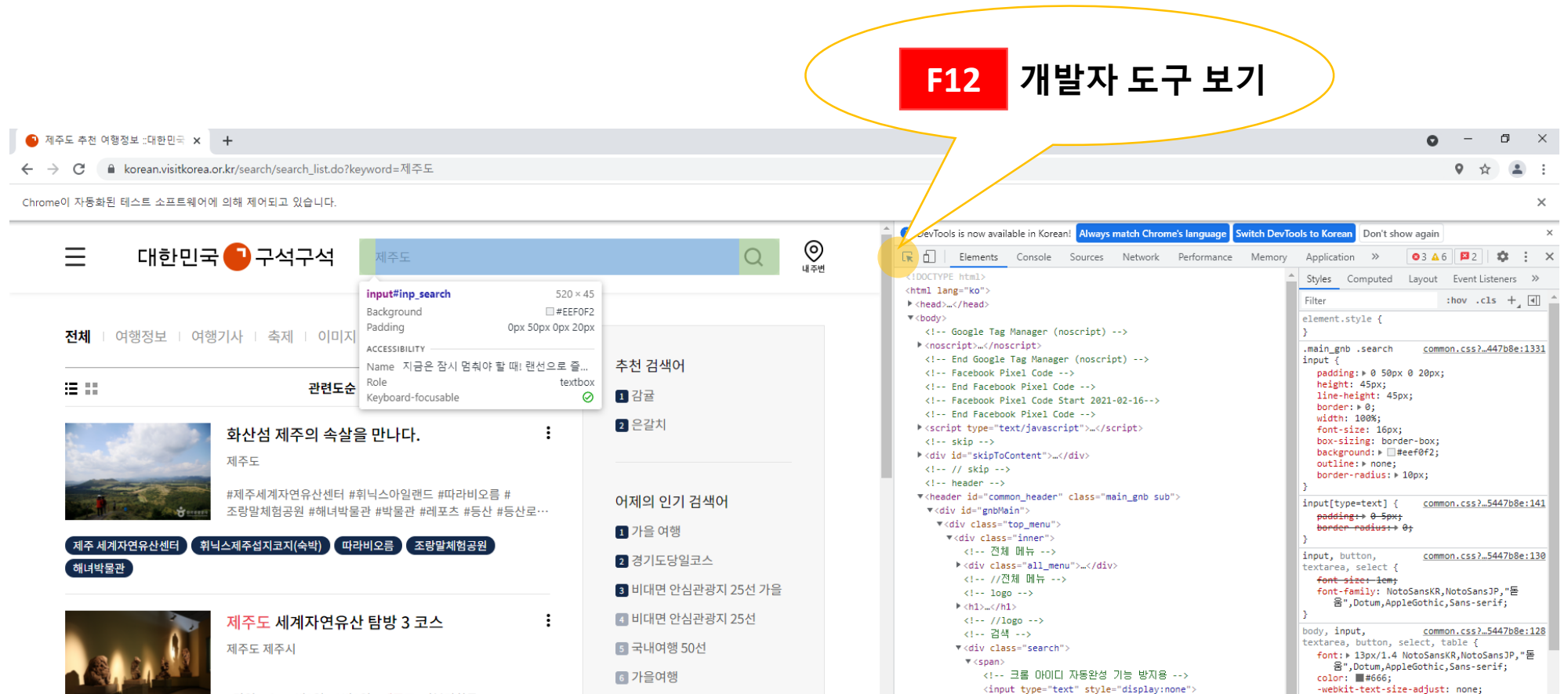
한국 관광 공사

<https://korean.visitkorea.or.kr/main/main.html>

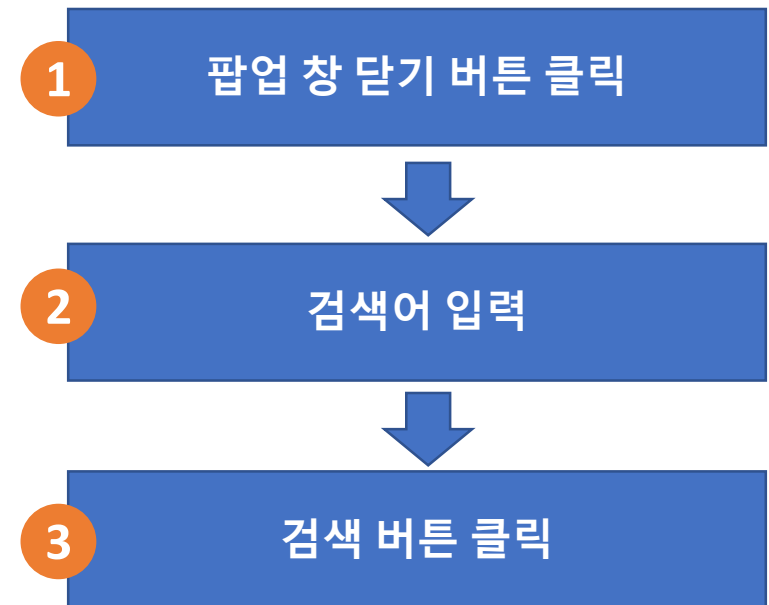
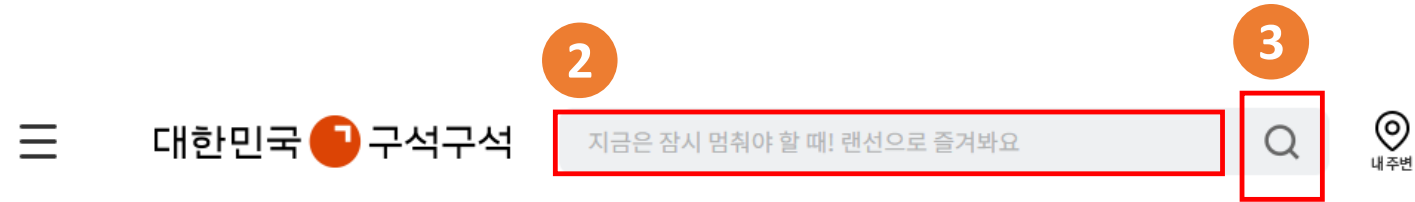
2021. 10. 04.



1. Web Crawling 원리
2. selenium 설치 및 확인
3. Chrome web driver 설치 및 확인
4. 한국관광공사 홈페이지 접속 크롤러 구현 실습
5. selenium 에서 제공되는 함수
6. 네이버 접속 크롤러 구현 실습



1. Web Crawling 원리
2. selenium 설치 및 확인
3. Chrome web driver 설치 및 확인
4. 한국관광공사 홈페이지 접속 크롤러 구현 실습
5. selenium 에서 제공되는 함수
6. 네이버 접속 크롤러 구현 실습



1. Web Crawling
원리

2. selenium 설치
및 확인

3. Chrome web
driver 설치 및
확인

4. 한국관광공사
홈페이지 접속
크롤러 구현
실습

5. selenium
에서 제공되는
함수

6. 네이버 접속
크롤러 구현
실습

예제) 한국관광공사 홈페이지 접속 크롤러 구현

Step 1. 크롤링에 필요한 모듈과 함수를 임포트하고 검색어 입력받기

```
1 from bs4 import BeautifulSoup
2 from selenium import webdriver
3 import time
4
5 query_txt = input('크롤링할 키워드는 무엇입니까?: ')
```

크롤링할 키워드는 무엇입니까?: 제주도

Step 2. 크롬 드라이버를 사용해서 웹 브라우저 실행하기

```
1 path = "c:/temp/chromedriver_94/chromedriver.exe"
2 driver = webdriver.Chrome(path)
```

Step 3. 한국 관광 공사 홈페이지 접속하기

```
1 driver.get("https://korean.visitkorea.or.kr")
2 time.sleep(2) # 위 페이지가 모두 열릴 때 까지 2초 기다립니다.
```

1. Web Crawling
원리

2. selenium 설치
및 확인

3. Chrome web
driver 설치 및
확인

4. 한국관광공사
홈페이지 접속
크롤러 구현
실습

5. selenium
에서 제공되는
함수

6. 네이버 접속
크롤러 구현
실습

Step 4. 팝업 창 닫기

```
1  #xpath 사용
2  #닫기 부분 클릭
3  driver.find_element_by_xpath('//*[@id="safetyStay1"]/div[1]/div/div[2]/button').click()
4
5  #오늘하루 그만보기 부분 클릭
6  #driver.find_element_by_xpath('//*[@id="safetyStay1"]/div[1]/div/div[2]/label').click()
```

Step 5. 입력 창 클릭하여 검색어 전달하기

```
1  #xpath 사용
2  #element = driver.find_element_by_xpath('//*[@id="inp_search"]')
3  #element.send_keys(query_txt)
4
5  #id 사용
6  element = driver.find_element_by_id("inp_search")
7  element.send_keys(query_txt)
```

1. Web Crawling
원리
2. selenium 설치
및 확인
3. Chrome web
driver 설치 및
확인
4. 한국관광공사
홈페이지 접속
크롤러 구현
실습
5. selenium
에서 제공되는
함수
6. 네이버 접속
크롤러 구현
실습

Step 6. 검색 버튼 눌러 실행하기

```
1  #텍스트 사용
2  driver.find_element_by_link_text("검색").click()
3
4  #class name 사용 (결과는 위와 동일)
5  #driver.find_element_by_class_name("btn_search").click()
6
7  #xpath 사용 (결과는 위와 동일)
8  #driver.find_element_by_xpath('//*[@id="gnbMain"]/div[1]/div/div[2]/span/a').click()
```

selenium 에서 제공되는 find 함수

<https://selenium-python.readthedocs.io/locating-elements.html>

1. Web Crawling
원리
2. selenium 설치
및 확인
3. Chrome web
driver 설치 및
확인
4. 한국관광공사
홈페이지 접속
크롤러 구현
실습
5. selenium
에서 제공되는
함수
6. 네이버 접속
크롤러 구현
실습



Navigation

1. [Installation](#)
2. [Getting Started](#)
3. [Navigating](#)
4. [Locating Elements](#)
 - [4.1. Locating by Id](#)
 - [4.2. Locating by Name](#)
 - [4.3. Locating by XPath](#)
 - [4.4. Locating Hyperlinks by Link Text](#)
 - [4.5. Locating Elements by Tag Name](#)
 - [4.6. Locating Elements by Class Name](#)
 - [4.7. Locating Elements by CSS Selectors](#)
5. [Waits](#)
6. [Page Objects](#)
7. [WebDriver API](#)
8. [Appendix: Frequently Asked Questions](#)

4. Locating Elements

There are various strategies to locate elements in a page. You can use the most appropriate one for your case. Selenium provides the following methods to locate elements in a page:

- `find_element_by_id`
- `find_element_by_name`
- `find_element_by_xpath`
- `find_element_by_link_text`
- `find_element_by_partial_link_text`
- `find_element_by_tag_name`
- `find_element_by_class_name`
- `find_element_by_css_selector`

To find multiple elements (these methods will return a list):

- `find_elements_by_name`
- `find_elements_by_xpath`
- `find_elements_by_link_text`
- `find_elements_by_partial_link_text`
- `find_elements_by_tag_name`
- `find_elements_by_class_name`
- `find_elements_by_css_selector`

Apart from the public methods given above, there are two private methods which might be useful for locating page elements:

- `find_element`
- `find_elements`

Example usage:

<https://selenium-python.readthedocs.io/locating-elements.html>

1. Web Crawling
원리

2. selenium 설치
및 확인

3. Chrome web
driver 설치 및
확인

4. 한국관광공사
홈페이지 접속
크롤러 구현
실습

5. selenium
에서 제공되는
함수

6. 네이버 접속
크롤러 구현
실습

4. Locating Elements

There are various strategies to locate elements in a page. You can use the most appropriate one for your case. Selenium provides the following methods to locate elements in a page:

- *find_element_by_id*
- *find_element_by_name*
- *find_element_by_xpath*
- *find_element_by_link_text*
- *find_element_by_partial_link_text*
- *find_element_by_tag_name*
- *find_element_by_class_name*
- *find_element_by_css_selector*

To find multiple elements (these methods will return a list):

- *find_elements_by_name*
- *find_elements_by_xpath*
- *find_elements_by_link_text*
- *find_elements_by_partial_link_text*
- *find_elements_by_tag_name*
- *find_elements_by_class_name*
- *find_elements_by_css_selector*

1. Web Crawling
원리
2. selenium 설치
및 확인
3. Chrome web
driver 설치 및
확인
4. 한국관광공사
홈페이지 접속
크롤러 구현
실습
5. selenium
에서 제공되는
함수
6. 네이버 접속
크롤러 구현
실습

■ 실습 노트 참고

- 08_(네이버) 강의용 Selenium for Web Crawling.ipynb

1. Web Crawling
원리

2. selenium 설치
및 확인

3. Chrome web
driver 설치 및
확인

4. 한국관광공사
홈페이지 접속
크롤러 구현
실습

5. selenium
에서 제공되는
함수

6. 네이버 접속
크롤러 구현
실습

지시 사항

한국관광공사 홈페이지 접속 크롤러를 참고로 네이버 접속 크롤러를 완성하시오.

네이버 접속 크롤러 구현

- Naver 검색창에서 검색어 입력하여 다음 페이지로 이동하는 프로그램 구현하기

Step 1. 크롤링에 필요한 모듈과 함수를 임포트하고 검색어 입력받기

1
2
3
4
5

크롤링할 키워드는 무엇입니까?:

Step 2. 크롬 드라이버를 사용해서 웹 브라우저 실행하기

1
2

Step 3. 네이버 홈페이지 들어가기

1
2

Step 4. 입력 창 클릭하여 검색어 전달하기

1
2

Step 5. 검색 버튼 눌러 실행하기

1
2