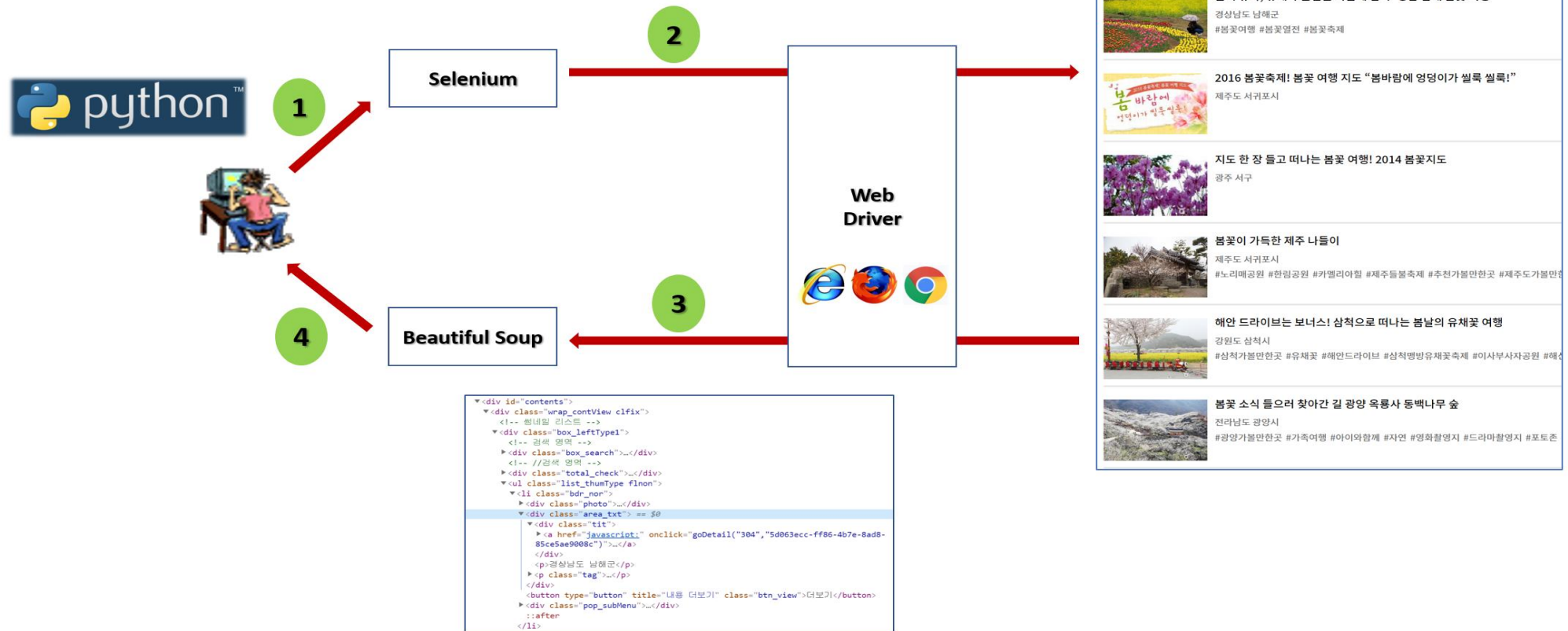


# Selenium for Web Crawling

## • Web Crawling 원리

- selenium 설치 및 확인
- Web driver 설치 및 확인 (Chrome Driver 활용)
- 자동 검색 기능 구현 (한국관광공사 홈페이지를 활용한 실습)
- Exercise



- Web Crawling 원리
- **selenium 설치 및 확인**
- Web driver 설치 및 확인  
(Chrome Driver 활용)
- 자동 검색 기능 구현  
(한국관광공사 홈페이지를  
활용한 실습)
- Exercise

## ■ 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 a search for 'selenium' under the 'base (root)' environment. The search results table lists three packages: 'pytest-selenium' (version 2.0.1), 'r-seleniumpipes' (version 0.3.7), and 'selenium' (version 3.141.0). The 'selenium' package is selected with a green checkmark. A status bar at the bottom indicates '3 packages available matching "selenium"'. A 'Sign in' button is visible in the top right corner.

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

**selenium 확인**

- Web Crawling 원리
- **selenium 설치 및 확인**
- Web driver 설치 및 확인  
(Chrome Driver 활용)
- 자동 검색 기능 구현  
(한국관광공사 홈페이지를  
활용한 실습)
- Exercise

## ■ conda 활용하여 인스톨하기

```
(base) C:\WINDOWS\system32>conda install selenium
```

## ■ pip 활용하여 인스톨하기

```
(base) C:\WINDOWS\system32>pip install selenium
```

- Web Crawling 원리
- **selenium 설치 및 확인**
- Web driver 설치 및 확인  
(Chrome Driver 활용)
- 자동 검색 기능 구현  
(한국관광공사 홈페이지를  
활용한 실습)
- Exercise

[https://www.selenium.dev/documentation/ko/introduction/the\\_selenium\\_project\\_and\\_tools/](https://www.selenium.dev/documentation/ko/introduction/the_selenium_project_and_tools/)

- Selenium 은 웹 브라우저를 제어합니다.
- **Selenium** 은 여러 프로젝트를 포함하지만, 웹 브라우저 자동화를 위한 도구가 핵심입니다.  
브라우저 인스턴스를 원격으로 제어하고, 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 프론트-엔드 테스트에 주로 사용되지만, **Selenium**은 브라우저 사용자  
에이전트 *라이브러리*의 핵심입니다. 인터페이스는 응용 프로그램에 맞게 어디서나 사용될 수  
있도록 최적화 되어있고, 당신의 목적에 맞게 다른 라이브러리와 구성을 장려합니다.

<https://www.selenium.dev/> 공식 홈페이지

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

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

- Web Crawling 원리
- selenium 설치 및 확인
- **Web driver 설치 및 확인 (Chrome Driver활용)**
- 자동 검색 기능 구현 (한국관광공사 홈페이지를 활용한 실습)
- Exercise

## 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.

- Web Crawling 원리
- selenium 설치 및 확인
- **Web driver 설치 및 확인 (Chrome Driver 활용)**
- 자동 검색 기능 구현 (한국관광공사 홈페이지를 활용한 실습)
- Exercise

## chrome web browser 버전 확인

The image shows a Chrome browser window with several steps highlighted by numbered blue circles and red boxes:

- 1**: A red box highlights the three-dot menu icon in the top right corner of the browser window.
- 2**: A red box highlights the '도움말(E)' (Help) option in the dropdown menu.
- 3**: A red box highlights the 'Chrome 정보(G)' (About Chrome) option in the submenu.
- 4**: A screenshot of the 'Chrome 정보' (About Chrome) page, which displays the current version: 'Chrome이 최신 버전입니다. 버전 89.0.4389.114(공식 빌드) (64비트)'.

- Web Crawling 원리
- selenium 설치 및 확인
- **Web driver 설치 및 확인 (Chrome Driver활용)**
- 자동 검색 기능 구현  
(한국관광공사 홈페이지를  
활용한 실습)
- Exercise

## 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.common.exceptions.WebDriverException: Message: unknown error: cannot find Chrome binary"](#)

라는 오류가 나오면서 chrome driver가 실행이 안될 수 있습니다.



- Web Crawling 원리
- selenium 설치 및 확인
- Web driver 설치 및 확인 (Chrome Driver활용)
- 자동 검색 기능 구현 (한국관광공사 홈페이지를 활용한 실습)
- Exercise

## selenium 패키지가 사용할 Web driver 설치

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

접속하기

### Downloads

#### Current Releases

- If you are using Chrome version 90, please download [ChromeDriver 90.0.4430.24](#)
- If you are using Chrome version 89, please download [ChromeDriver 89.0.4389.23](#)
- If you are using Chrome version 88, please download [ChromeDriver 88.0.4324.96](#)
- If you are using Chrome version 87, please download [ChromeDriver 87.0.4280.88](#)
- 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 90.0.4430.24

Supports Chrome version 90

- Resolved issue 3708: GetElementLocation's inView fails for elements in shadow DOM
- Resolved issue 3721: Collect performance log for service worker when enabled
- Resolved issue 3731: AdbImpl::GetPidByName does not work correctly on non-standard Android OS's like on OnePlus
- Resolved issue 3733: Catastrophic failure with domain on localhost
- Resolved issue 3734: Driver should Error Attempting to set Cookie on Wrong Domain
- Resolved issue 3743: Refactor call function and fix remove element issue

For more details, please see the [release notes](#).

#### ChromeDriver 89.0.4389.23

Supports Chrome version 89

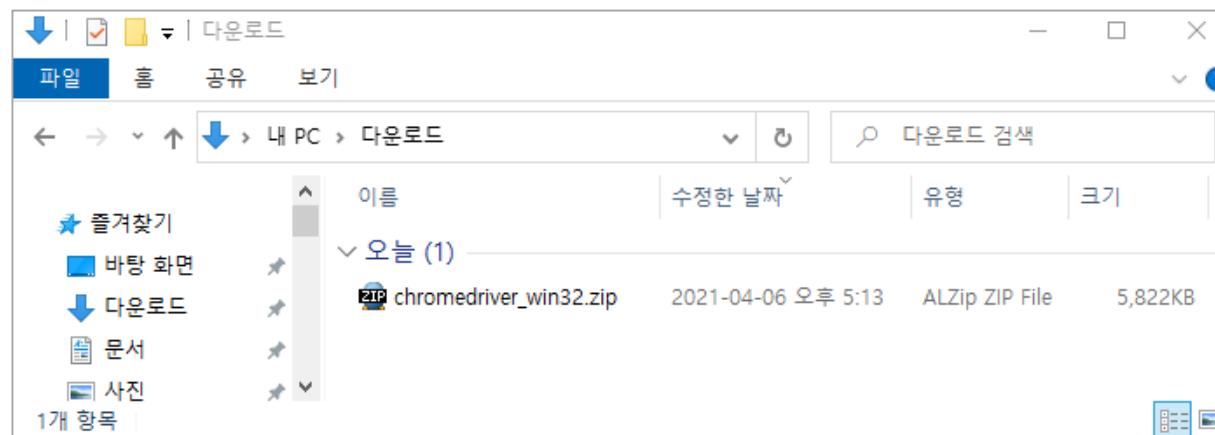
- Resolved issue 3667: Timed out receiving a message from renderer
- Resolved issue 3675: GetElementLocation uses inView method which mishandles boolean as array
- Resolved issue 3682: GetElementRegion returns incorrect size for elements which are partially out of viewport
- Resolved issue 3690: ChromeDriver fails if extensions use chrome.windows API

- Web Crawling 원리
- selenium 설치 및 확인
- Web driver 설치 및 확인 (Chrome Driver활용)
- 자동 검색 기능 구현 (한국관광공사 홈페이지를 활용한 실습)
- Exercise

**Step 2.** 89 버전의 화면에서 OS 환경에 적합한 크롬 드라이버를 선택하여 다운로드하기

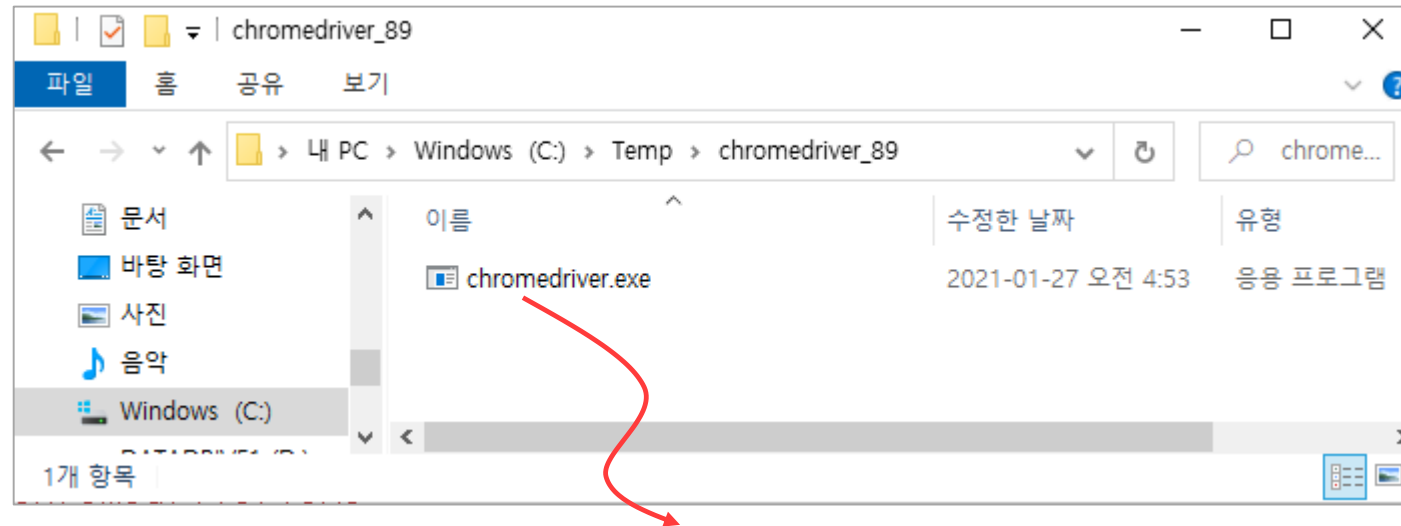
Index of /89.0.4389.23/

Name	Last modified	Size	ETag
<a href="#">Parent Directory</a>	-	-	-
<a href="#">chromedriver_linux64.zip</a>	2021-01-28 17:30:52	5.57MB	24686a3cc3ccf8cbc60cf744baa47692
<a href="#">chromedriver_mac64.zip</a>	2021-01-28 17:30:53	7.97MB	a6620c6a6804fa08365dfc6e8c8724e6
<a href="#">chromedriver_mac64_m1.zip</a>	2021-01-28 17:30:55	7.17MB	1544d2a1b1a6fbd55ff5dac0b48d89f
<a href="#">chromedriver_win32.zip</a>	2021-01-28 17:30:57	5.68MB	0bf4bc39f34cee67f5f95afd8a24c191
<a href="#">notes.txt</a>	2021-01-28 17:31:00	0.00MB	4e3ff6354a7462edfe5d95ab8690f7c8



- Web Crawling 원리
- selenium 설치 및 확인
- **Web driver 설치 및 확인 (Chrome Driver활용)**
- 자동 검색 기능 구현 (한국관광공사 홈페이지를 활용한 실습)
- Exercise

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



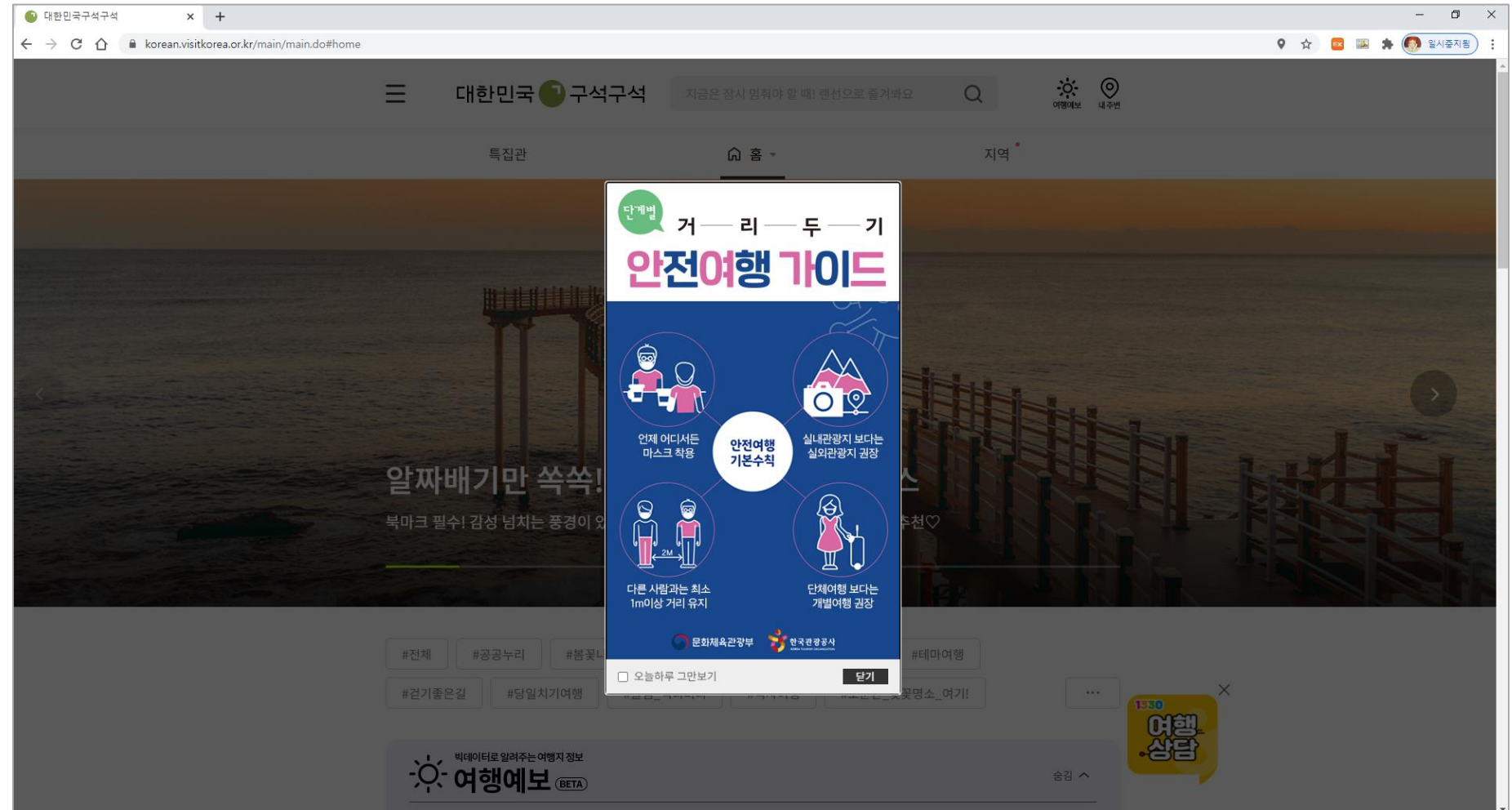
크롬 드라이버 버전을 기억하기 편하게 디렉토리 이름을 변경하여 코딩에 활용하면 편리함.  
**Chromedriver\_89 디렉토리** 안에 chromedriver.exe 파일이 있음

 chromedriver.exe

# 자동 검색 기능 구현 (한국관광공사 홈페이지 활용)

한국 관광 공사

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



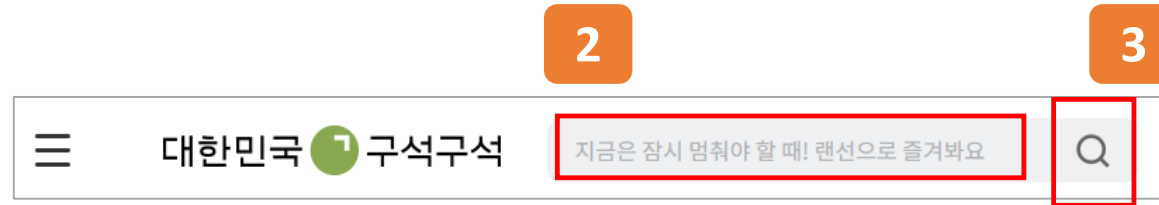
- Web Crawling 원리
- selenium 설치 및 확인
- Web driver 설치 및 확인 (Chrome Driver 활용)
- 자동 검색 기능 구현 (한국관광공사 홈페이지를 활용한 실습)
- Exercise

# 자동 검색 기능 구현 (한국관광공사 홈페이지 활용)

- Web Crawling 원리
- selenium 설치 및 확인
- Web driver 설치 및 확인 (Chrome Driver 활용)

- 자동 검색 기능 구현 (한국관광공사 홈페이지를 활용한 실습)

- Exercise



- Web Crawling 원리
- selenium 설치 및 확인
- Web driver 설치 및 확인 (Chrome Driver 활용)
- 자동 검색 기능 구현 (한국관광공사 홈페이지를 활용한 실습)
- Exercise

## 한국관광공사 홈페이지 접속 크롤러 구현

Step 1. 크롤링에 필요한 모듈과 함수를 로딩하고 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_89/chromedriver.exe"
2 driver = webdriver.Chrome(path)
```

Step 3. 한국 관광 공사 홈페이지 들어가기

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

## 자동 검색 기능 구현 (한국관광공사 홈페이지 활용)

- Web Crawling 원리
- selenium 설치 및 확인
- Web driver 설치 및 확인 (Chrome Driver 활용)
- 자동 검색 기능 구현 (한국관광공사 홈페이지를 활용한 실습)
- Exercise

### Step 4. 팝업 창 닫기

```
1 driver.find_element_by_id("chkForm01").click()
```

### Step 5. 입력 창 클릭하여 검색어 전달하기

```
1 element = driver.find_element_by_id("inp_search")
2 element.send_keys(query_txt)
```

### Step 6. 검색 버튼 눌러 실행하기

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



# 자동 검색 기능 구현 (한국관광공사 홈페이지 활용)

- Web Crawling 원리
- selenium 설치 및 확인
- Web driver 설치 및 확인 (Chrome Driver 활용)
- 자동 검색 기능 구현 (한국관광공사 홈페이지를 활용한 실습)
- Exercise

유용하게 활용됨

**F12 개발자 도구 보기**

The screenshot shows the Korean Tourism website (korean.visitkorea.or.kr) with the developer tools (F12) open. The left pane shows the DOM tree with a selected link. The right pane shows the CSS styles for the selected element.

**DOM Tree:**

```

<li>
  <a href="javascript:goAlllist('73662aeb-4f78-488e-a126-d34226d779d5','건기좋은길');">
    <span>#건기좋은길</span> == $0
  </a>
</li>

```

**Styles:**

```

element.style {
}

@media screen and (max-width: 1023px)
.main_taglist ul li a span {
  display: block;
  height: 28px;
  padding: 0 10px;
  font-size: 13px;
  line-height: 26px;
  border: 1px solid #d8d7d7;
}

```



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

- Web Crawling 원리
- selenium 설치 및 확인
- Web driver 설치 및 확인  
(Chrome Driver 활용)
- 자동 검색 기능 구현  
(한국관광공사 홈페이지를  
활용한 실습)
- Exercise

■ 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*

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

- Web Crawling 원리
- selenium 설치 및 확인
- Web driver 설치 및 확인  
(Chrome Driver 활용)
- 자동 검색 기능 구현  
(한국관광공사 홈페이지를  
활용한 실습)
- Exercise

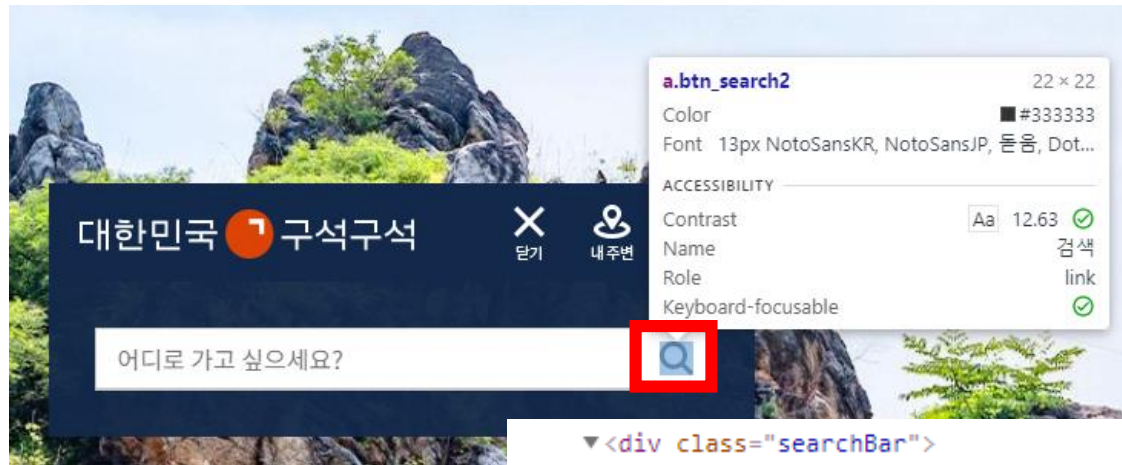
■ Selenium provides the following **methods** to locate elements in a page:

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*

# 자동 검색 기능 구현 (한국관광공사 홈페이지 활용)

- Web Crawling 원리
- selenium 설치 및 확인
- Web driver 설치 및 확인 (Chrome Driver 활용)
- 자동 검색 기능 구현 (한국관광공사 홈페이지를 활용한 실습)
- Exercise



2020년도  
홈페이지

```
<div class="searchBar">
  <input type="text" placeholder="어디로 가고 싶으세요?" class="inp_search" id="inp_search" name="inp_search" title="검색" onkeyup="completechk();" onkeydown="wordsListDelete();" onfocus="headerReadonly();" readonly>
  <a href="javascript:" class="btn_search2" onclick="btnGnbSearch();" >검색</a> == $0
  <!-- 0604 클래스 수정 -->
  <button type="button" class="btn_del" onclick="wordsListDelete();" >삭제</button>
```

- `find_element_by_link_text` → 검색
- `find_element_by_class_name` → btn\_search2
- `find_element_by_xpath` → `//*[@id="gnbMain"]/div/div/div[1]/div[1]/a`

# 자동 검색 기능 구현 (Naver 홈페이지 활용)

- Web Crawling 원리
- selenium 설치 및 확인
- Web driver 설치 및 확인 (Chrome Driver 활용)
- 자동 검색 기능 구현 (한국관광공사 홈페이지를 활용한 실습)

## • Exercise

### 네이버 접속 크롤러 구현

- Naver 검색창에서 검색어 입력하여 다음 페이지로 이동하는 프로그램 구현하기

Step 1. 필요한 모듈과 함수를 로딩하고 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	