

Python을 활용한 웹 크롤링

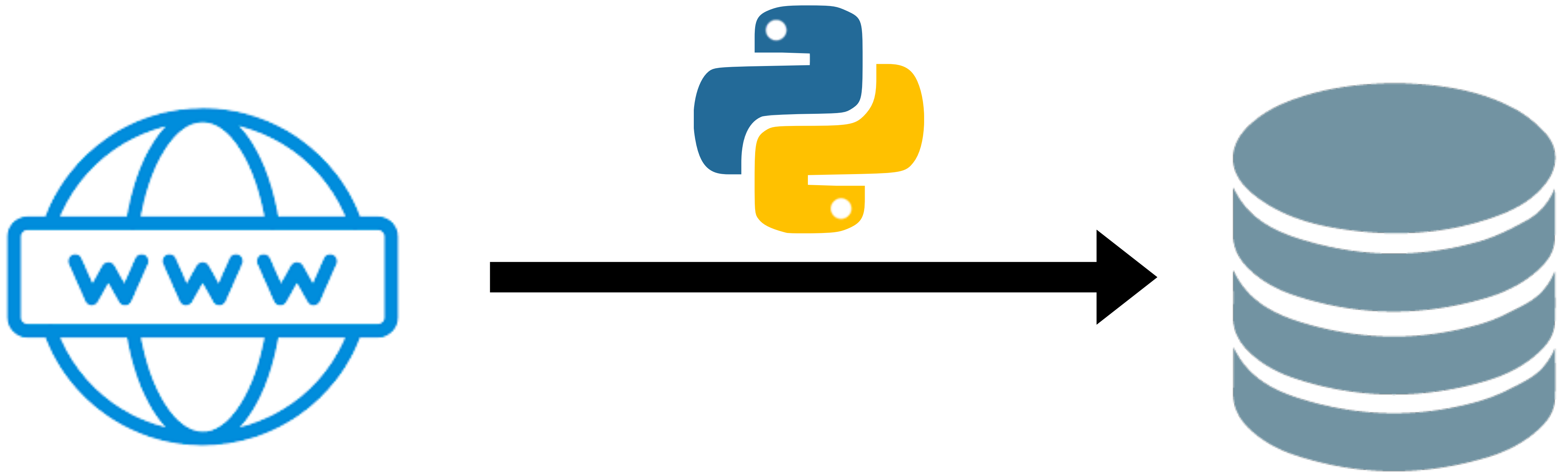
2018.04.03.
시범강의

크롤링, 스크래핑이란??

컴퓨터 소프트웨어 기술을 사용하여 웹에서 원하는 정보를 자동적으로 추출하는 것

크롤링, 스크래핑이란??

컴퓨터 소프트웨어 기술을 사용하여 웹에서 원하는 정보를 자동적으로 추출하는 것



HOW TO CRAWLING

1. Python

1. Selenium
2. BeautifulSoup
3. Requests

2. WEB

1. HTML DOM
2. GET / POST

3. DATABASE

1. SQL
2. Python 연동

HOW TO CRAWLING

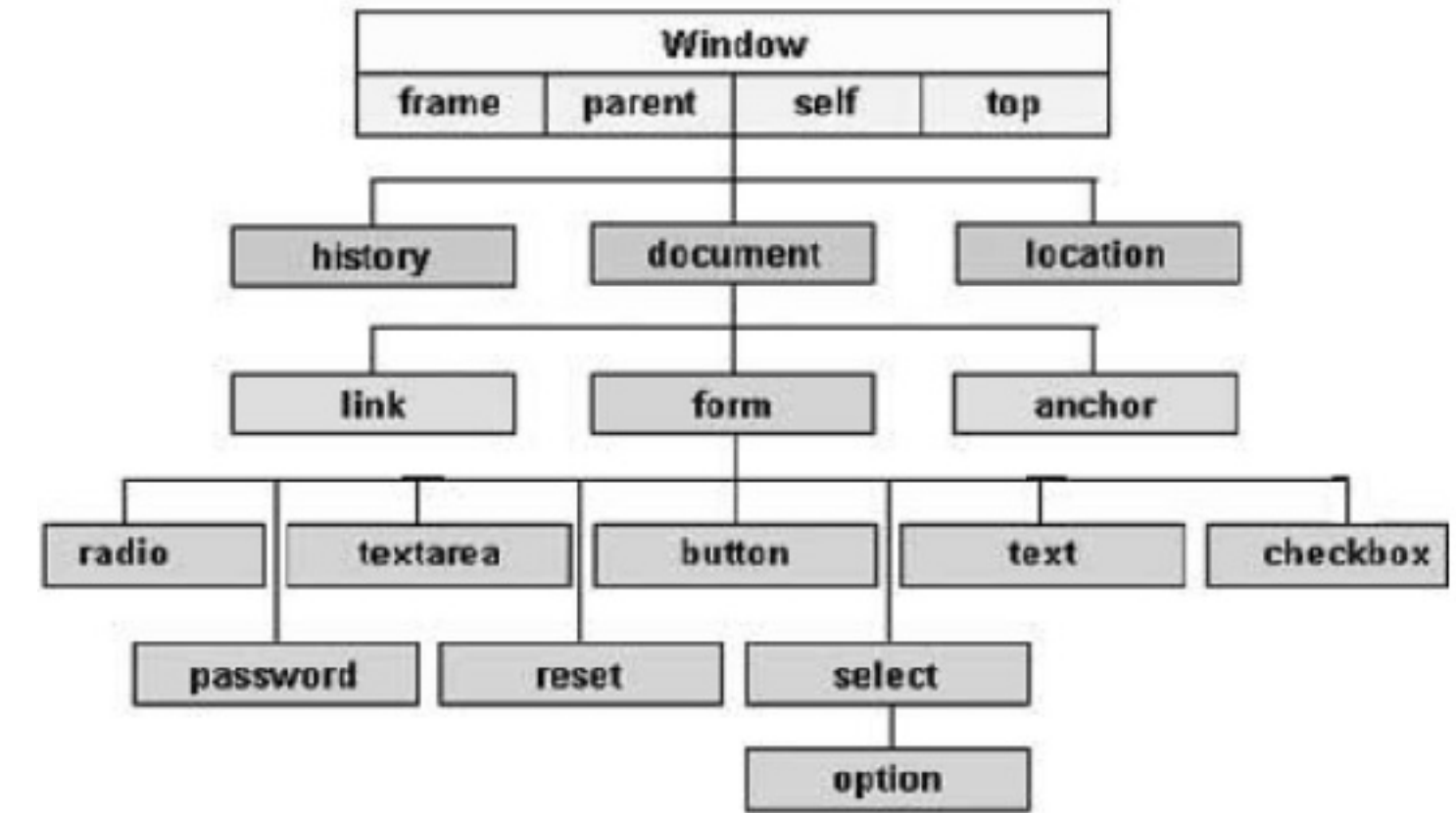
1. Python

1. Selenium
2. BeautifulSoup
3. Requests

Selenium



requests

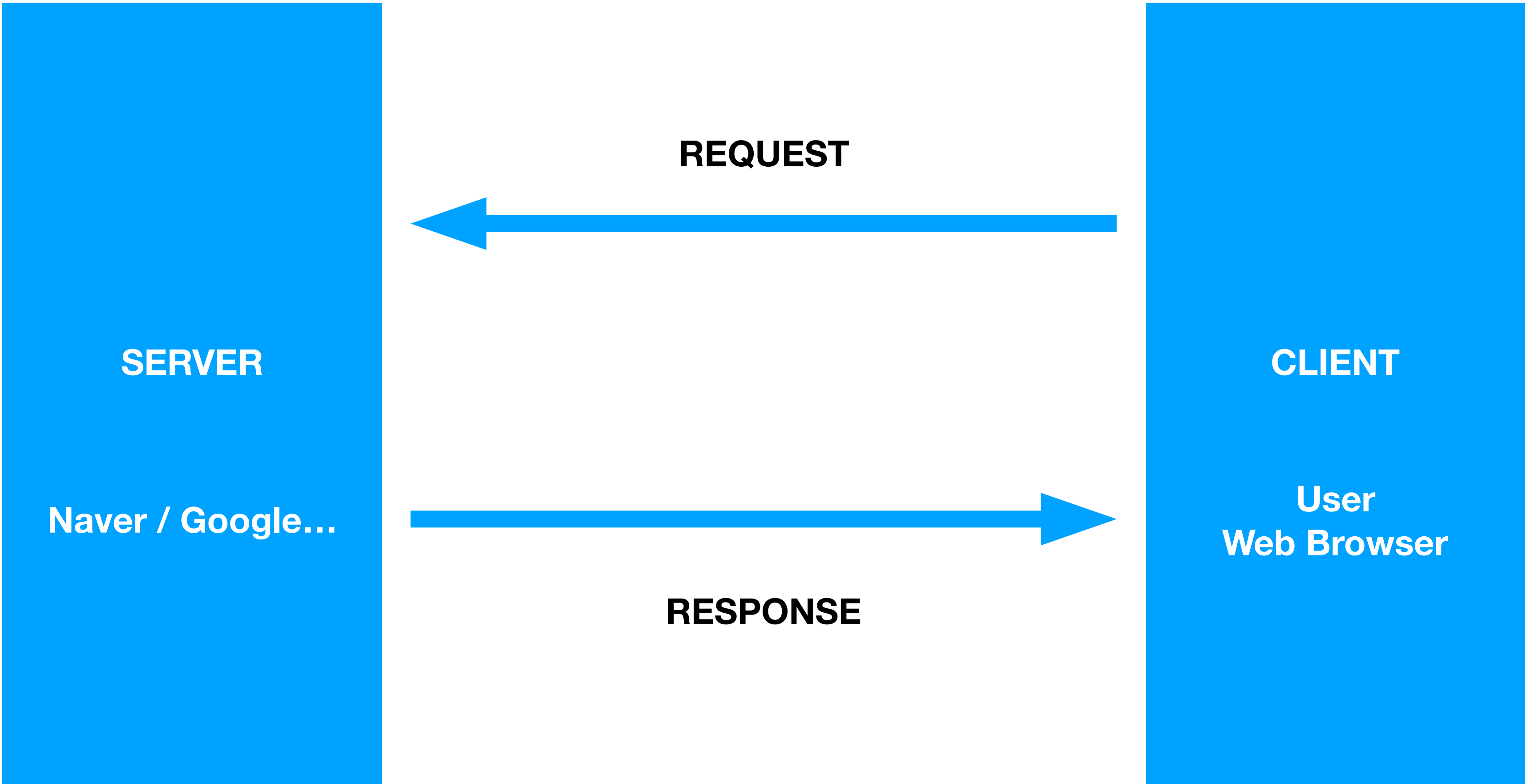


BeautifulSoup

```
10 <header role="banner">
11   <h1>HTML Essential Training</h1>
12   <nav role="navigation">
13     <ul>
14       <li><a href="intro.htm" title="Introduction to HTML5">Introduction</a>
15       <li><a href="syntax.htm" title="HTML Syntax">HTML Syntax</a>
16       <li><a href="structure.htm" title="HTML Structure">HTML Structure</a>
17       <li><a href="links.htm" title="HTML Links">HTML Links</a>
18       <li><a href="reference.htm" title="HTML Reference">HTML Reference</a>
19     </ul>
19   </nav>
20 </header>
```

HOW TO CRAWLING

- 2. WEB
 - 1. GET / POST
 - 2. HTML DOM



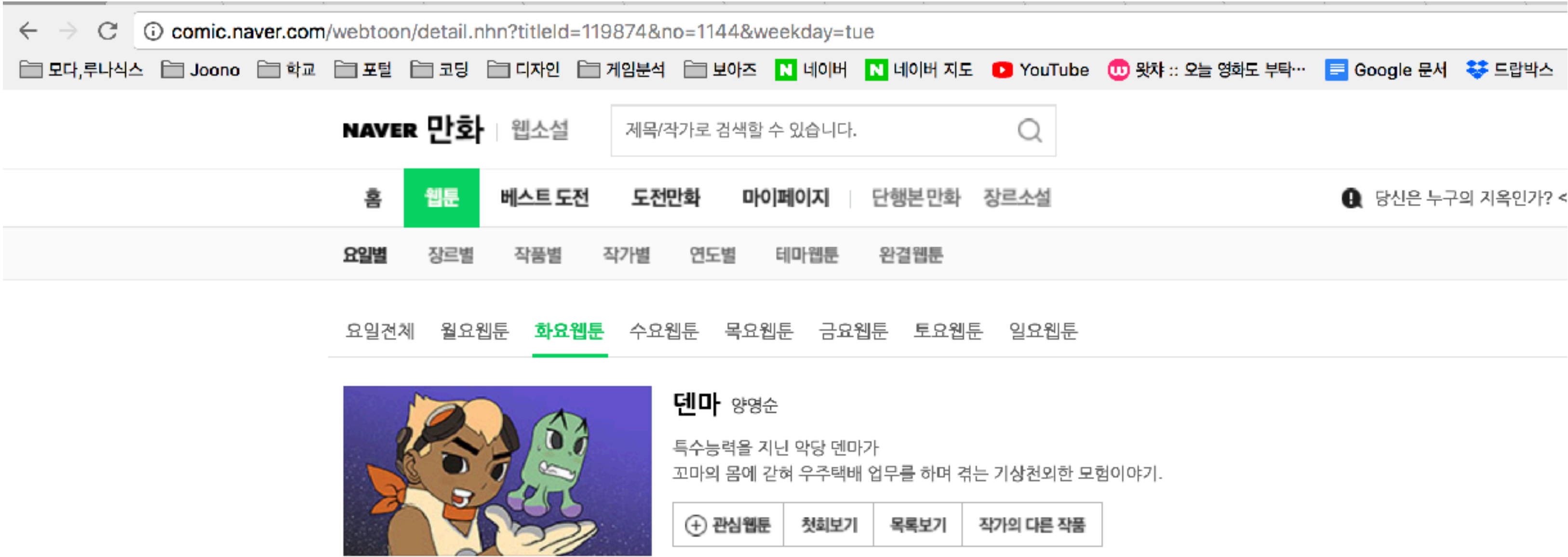
HOW TO CRAWLING

GET

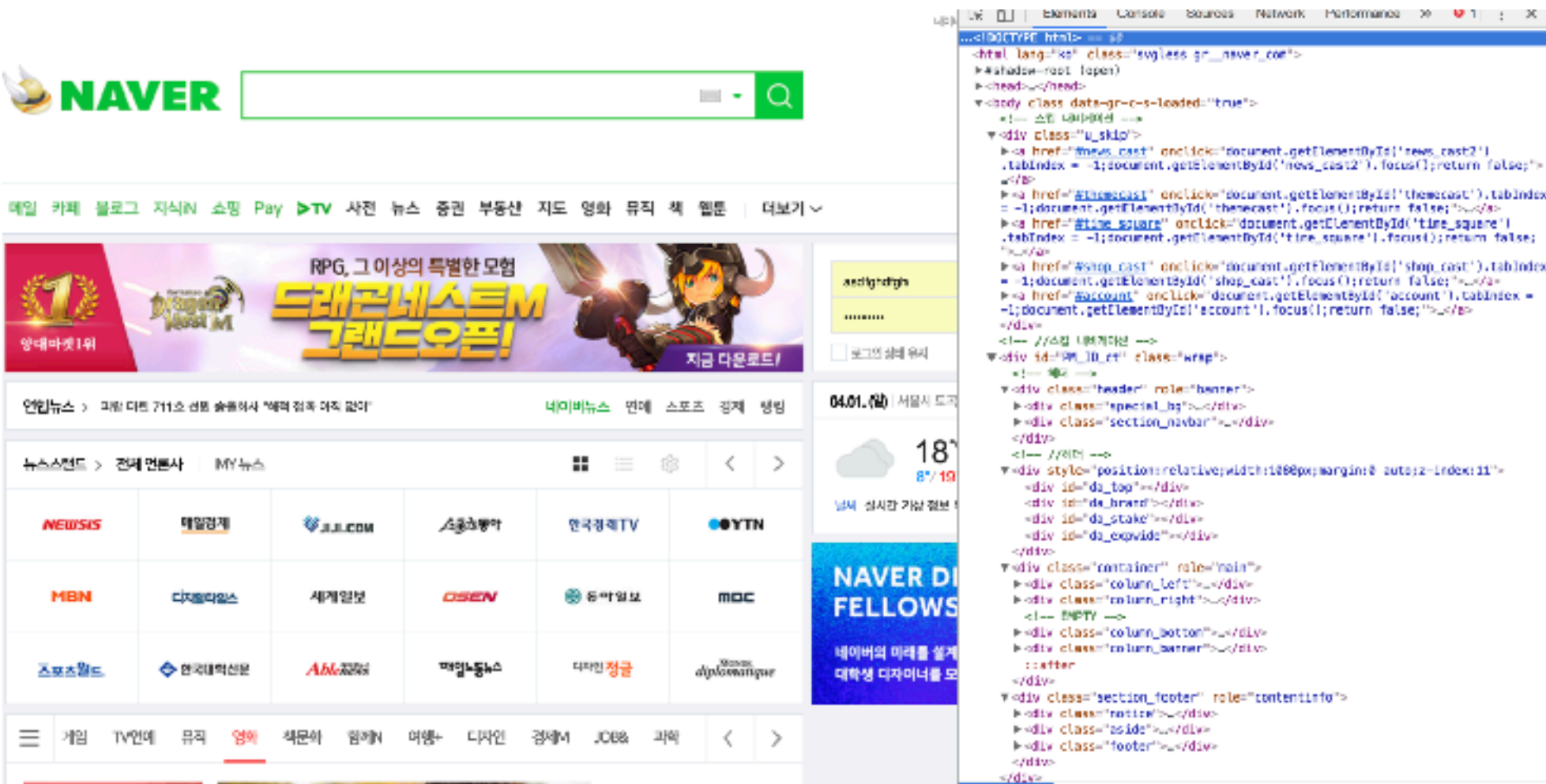
http://comic.naver.com/webtoon/detail.nhn?titleId=119874&no=1144&weekday=tue

2. WEB
1. GET / POST

2. HTML DOM



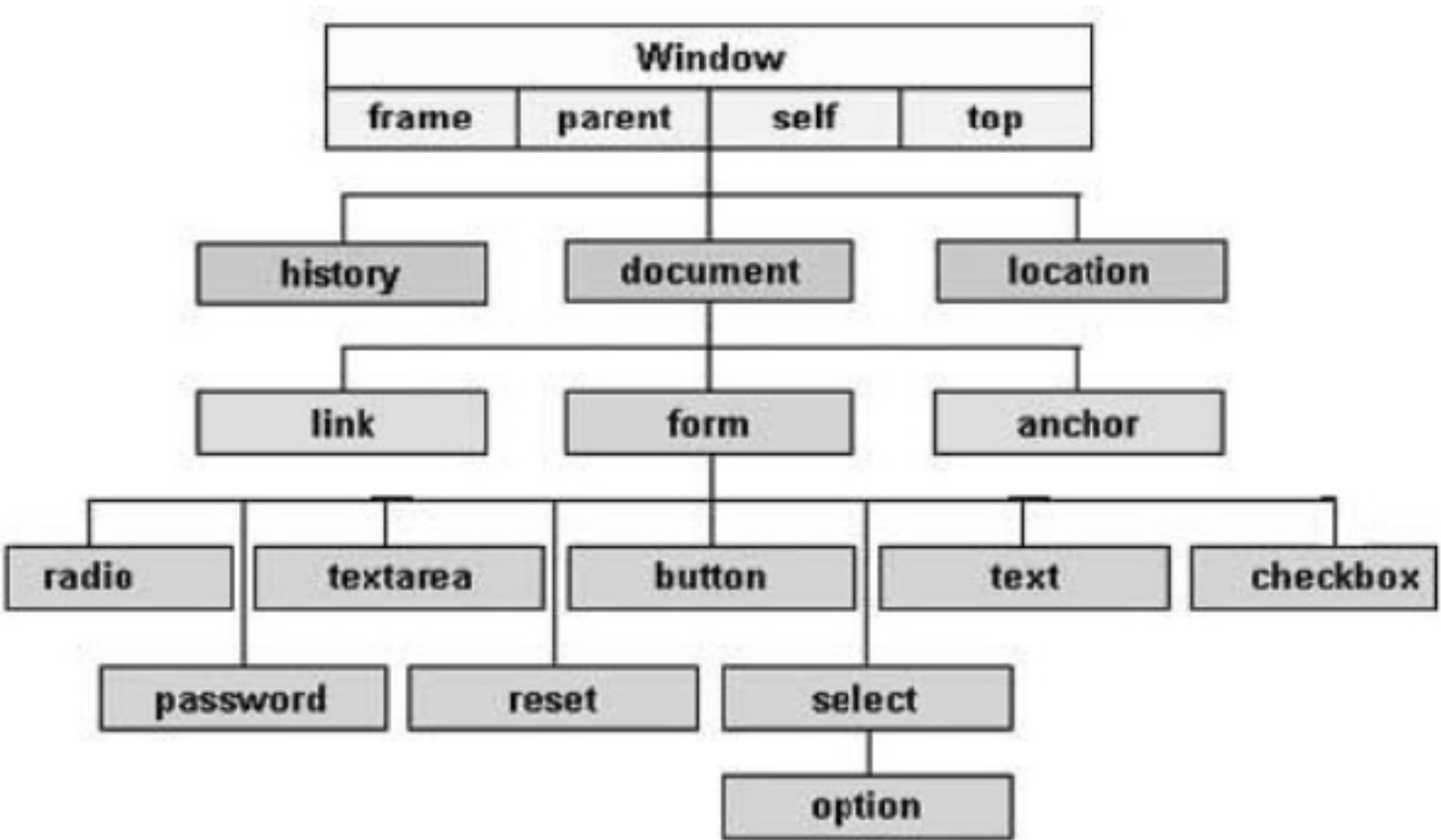
HOW TO CRAWLING



2. WEB

1. GET / POST

2. HTML DOM



HOW TO CRAWLING

SQL이란?

- RDBMS(관계형 데이터베이스 관리 시스템)의 데이터를 관리하기 위해 설계된 특수 목적의 프로그래밍 언어

왜 필요한가?

- 크롤링한 데이터가 방대한 경우, 혹은 관계형 데이터베이스에 관리해야 할 경우

Python에서는 어떻게 사용할까?

- Pymysql 패키지를 사용하여, Python script안에서 SQL을 사용

3. DATABASE

1. SQL

2. Python 연동

PRACTICE

**[https://github.com/jjjjoonno/fc_wc](https://github.com/jjjjoonno/fc_wc/blob/master/코드/test_case/exam1.py)
[/blob/master/코드/test_case/exam1.py](https://github.com/jjjjoonno/fc_wc/blob/master/코드/test_case/exam1.py)**