

Chapter. 01

데이터 크롤링 개요

I Chrome에 대해서...

FASTCAMPUS
ONLINE

금융공학/퀀트 I

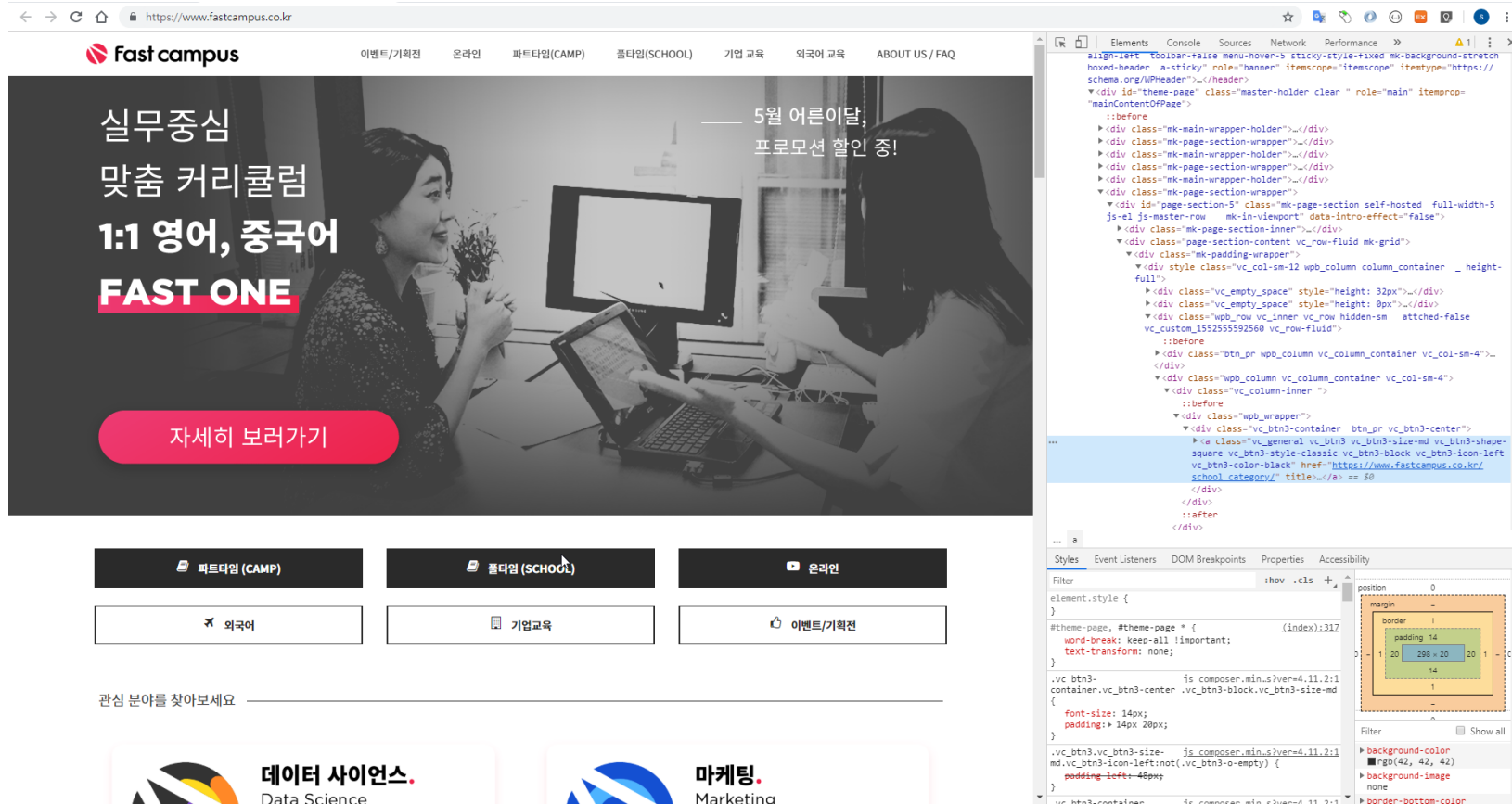
강사. 서찬웅

I 이번 시간에 알아볼 내용

1. chrome를 활용하는 방법
2. 개발자 모드
 1. elements
 2. network

I 크롬에서 키보드 F12 버튼을 눌러 개발자 도구를 실행합니다.

- 크롬을 실행하여 키보드 F12 키를 누르면 개발자 도구가 나타납니다. 개발자 도구가 중요한 이유는 우리가 원하는 데이터가 html 안에서 어느 위치에 존재하는지 알려줍니다. 개발자 도구안에서 중요한 몇 가지 내용에 대해서 알아보겠습니다.
- 아래 화면은 <https://www.fastcampus.co.kr/> 페이지에서 개발자 도구를 실행한 화면입니다.



I Elements 탭의 기능을 살펴보겠습니다.

1. 해당 기능을 클릭합니다.

2. 가져오고 싶은 데이터 부분을 클릭합니다. 클릭을 하게 되면 해당 부분이 선택 됩니다.

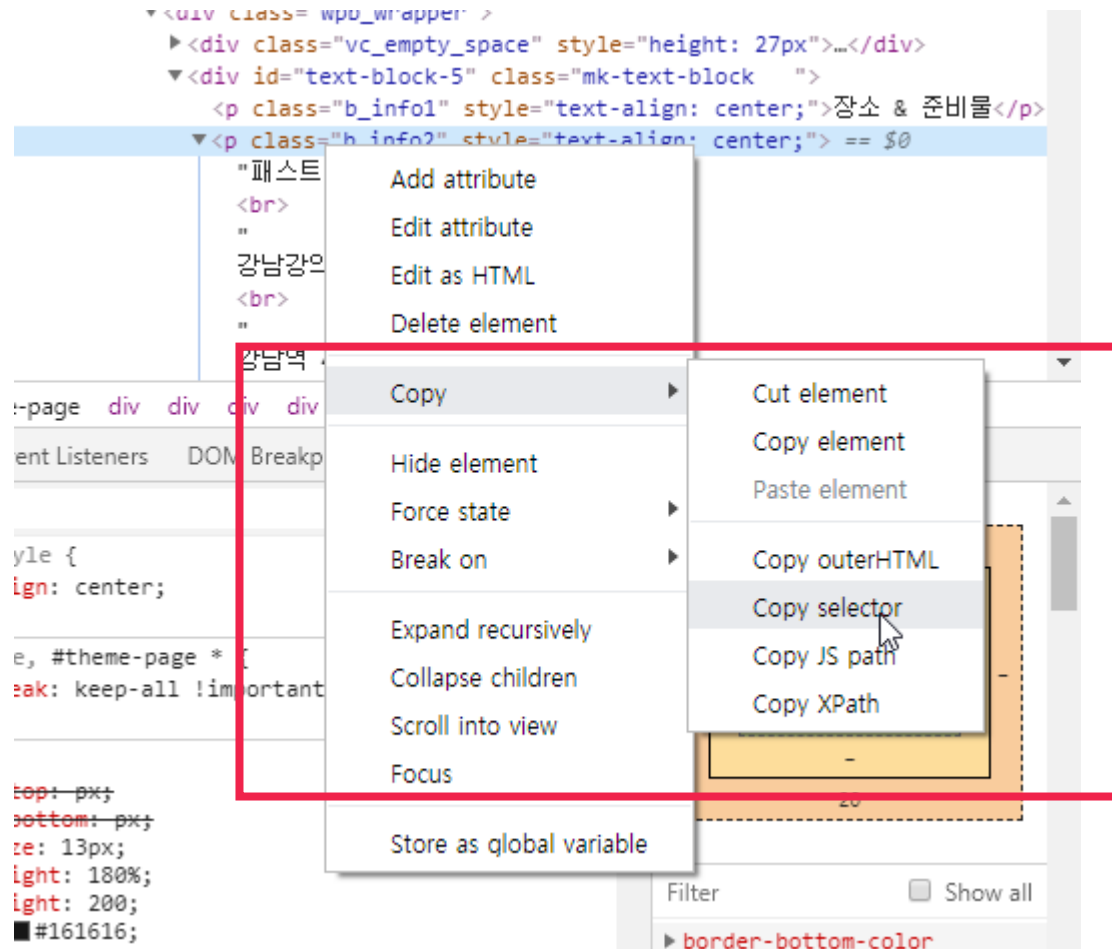
3. 클릭한 부분이 현재 보이는 페이지의 html에서 어느 위치에 해당하는지 알려줍니다.

```

<div id="theme-page" class="master-holder clear" role="main" itemprop="mainContentOfPage">
  ::before
  <div class="mk-main-wrapper-holder">...</div>
  <div class="wpb_row vc_row mk-fullwidth-true attached-false vc_custom_1552962152353 vc_row-fluid js-master-row mk-in-viewport">...</div>
  <div class="mk-main-wrapper-holder">...</div>
  <div class="wpb_row vc_row mk-fullwidth-true attached-false 0 vc_custom_1533792781628 vc_row-fluid js-master-row mk-in-viewport">
    ::before
    <div class="mk-grid">
      <div style class="vc_col-sm-12 wpb_column column_container _ height-full">
        <div class="wpb_row vc_inner vc_row box0 attached-false vc_custom_1533876857794 vc_row-fluid">
          ::before
          <div class="wpb_wrapper">
            <div class="vc_empty_space" style="height: 20px;">...</div>
            <div id="text-block-5" class="mk-text-block">
              <p class="b_info1" style="text-align: center;">장소 & 준비물</p>
              <p class="b_info2" style="text-align: center;">
                "패스트캠퍼스"
                <br>
                "강남강의장"
                <br>
                "강남역 4번 출구, 미왕빌딩"
              </p>
            </div>
          </div>
        </div>
      </div>
    </div>
  </div>
</div>
  
```

I Elements 탭의 기능을 살펴보겠습니다.

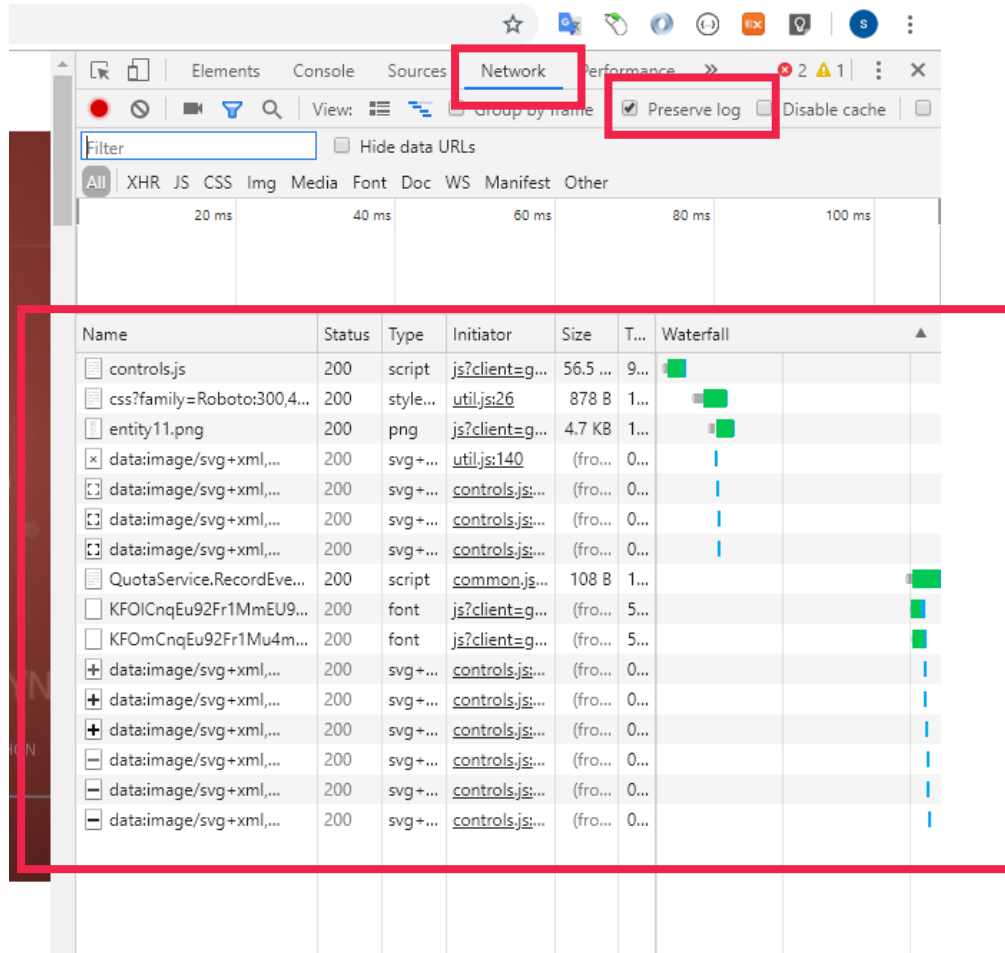
- 클릭한 부분의 html 태그 정보를 가지고 크롤링을 할 재료로 사용합니다.
 - tag, class, id 값을 찾고 그 정보를 바탕으로 정보를 가져옵니다.
- 3번 부분에서 해당하는 소스 위에서 마우스 오른쪽 버튼을 누르고 아래 화면처럼 클릭해보겠습니다.



copy selector 클릭을 하면 해당 정보의 select 정보가 클립보드에 복사가 됩니다. 이 정보도 크롤링에 활용됩니다.

Network 탭의 기능을 살펴보겠습니다.

- network 탭의 기능은 내가 접속한 페이지와 현재 크롬하고는 서로 통신을 하고 있습니다.
- 페이지에서 로그인이나 혹은 페이지 안의 버튼을 클릭 한다는건 서버에게 이런 정보를 요청을 한다는 거고, 그 정보를 다시 제게 보내줘야 크롬에서 정상적으로 페이지가 출력됩니다. network 탭은 이 정보를 출력해줍니다.
- 곧 서버-클라이언트 간의 통신 내용을 알수 있습니다.



옆에 목록에 있는 내용은 서버-클라이언트(내컴퓨터) 간에 데이터가 왔다갔다한 흔적입니다. 이 흔적을 조사하면 로그인 정보나, 검색어, 데이터가 어떻게 전달되는지 조사할수 있습니다. 결국 우리의 목적은 데이터를 가져오는거니 network 탭도 중요합니다.

※ Preserve Log 체크박스를 꼭 체크해주세요. 해당 기능은 체크를 하면 이전 기록도 지우지 않고 보관되어 계속 누적으로 출력됩니다. (체크하지 않으면 새로고침하면 초기화됩니다.)

I firefox에 대해서...

- firefox도 동일한 기능을 제공하고 있습니다. firefox 사용자들도 똑같은 기능을 사용할 수 있습니다.
- 거의 비슷하니 쉽게 접근할 수 있습니다. (개발자 도구 F12 도 동일)

감사합니다