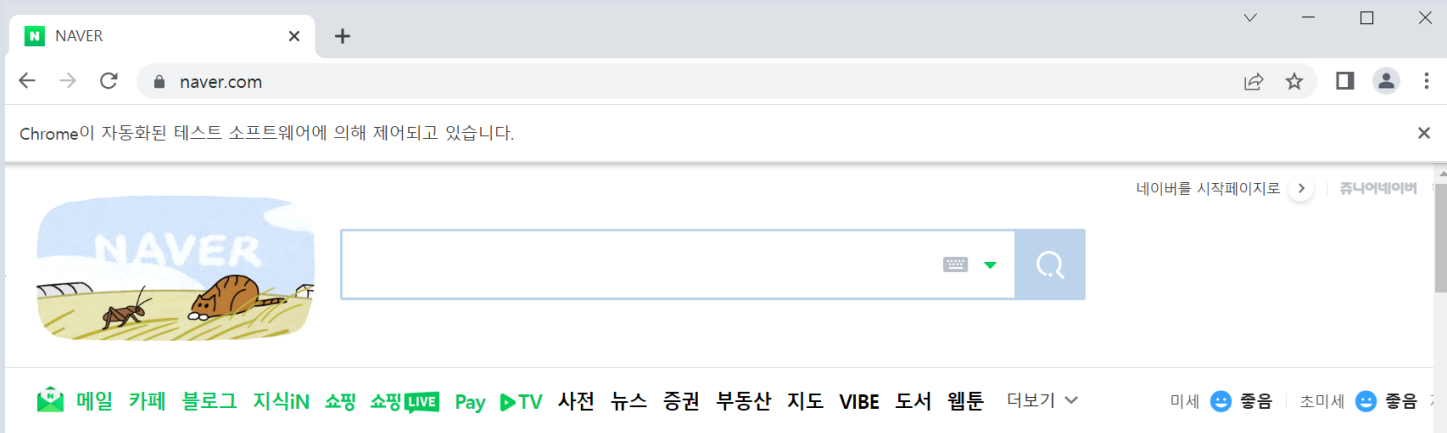


웹크롤링 심화

selenium

■ selenium

: 웹사이트 자동화 및 테스트를 위한 도구



- selenium 활용예시


- 웹사이트 로그인 자동화
- 메일 자동으로 보내기
- 좋아요 자동으로 누르기
- 포스팅 자동으로 작성하기
- 쪽지 자동으로 보내기
- 예약 매크로

- selenium 라이브러리 설치

```
pip install selenium
```

selenium

기본 템플릿



```
1 from selenium import webdriver
2 from selenium.webdriver.common.by import By
3 from selenium.webdriver.common.keys import Keys
4
5 # 크롬 드라이버 생성
6 driver = webdriver.Chrome()
7
8 # 페이지 이동
9 driver.get("https://www.naver.com")
10
11 # 태그 찾기
12 search = driver.find_element(By.CSS_SELECTOR, "#query")
13
14 # 1. 클릭
15 search.click()
16
17 # 2. 문자 입력
18 search.send_keys('스타트코딩')
19
20 # 3. 키 입력
21 search.send_keys(Keys.ENTER)
```

selenium

기본 사용법

- 태그 한개 선택하기

```
tag = driver.find_element(By.CSS_SELECTOR, "선택자")
```

* 찾았을 때 : Tag 객체 반환 * 못 찾았을 때 : NoSuchElementException 발생

- 태그 여러개 선택하기

```
tags = driver.find_elements(By.CSS_SELECTOR, "선택자")
```

* 찾았을 때 : Tag 객체 리스트 반환 * 못 찾았을 때 : 빈 리스트 반환

■ 자주 사용하는 Tag 객체의 속성과 메서드

명칭	기능
<code>tag.click()</code>	태그를 클릭
<code>tag.send_keys("message")</code>	태그에 메시지 입력
<code>tag.text</code>	태그안의 모든 텍스트 요소
<code>tag.get_attribute("속성명")</code>	태그의 속성값
<code>tag.find_element()</code>	해당 태그 안에서 한개 선택
<code>tag.find_elements()</code>	해당 태그 안에서 여러개 선택

selenium을 이용한 자동화 프로그램 만들기

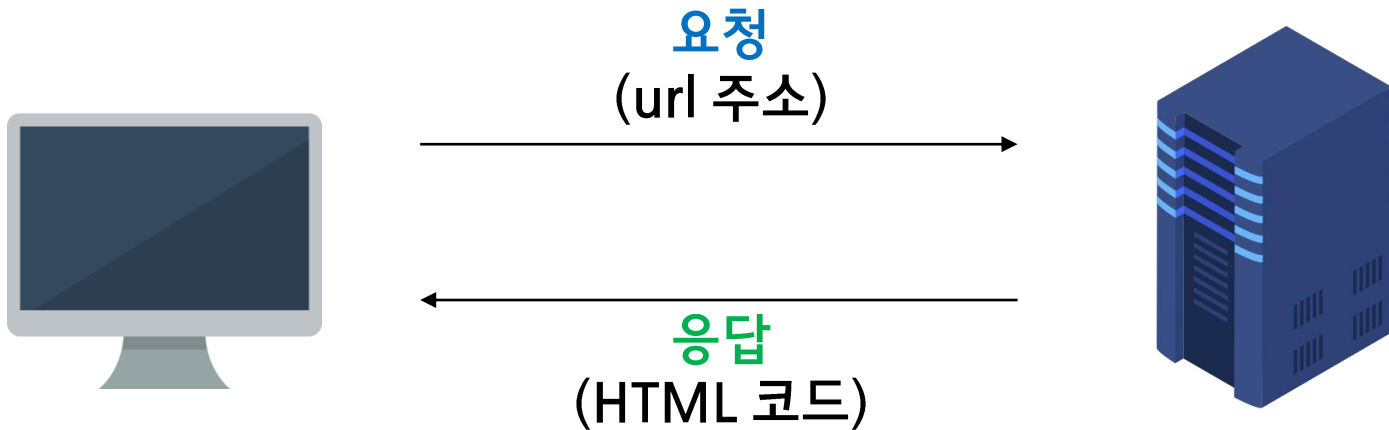
정적 페이지
동적 페이지

정적 페이지



정적 페이지 예시

동적 페이지



동적 페이지



동적 페이지 예시

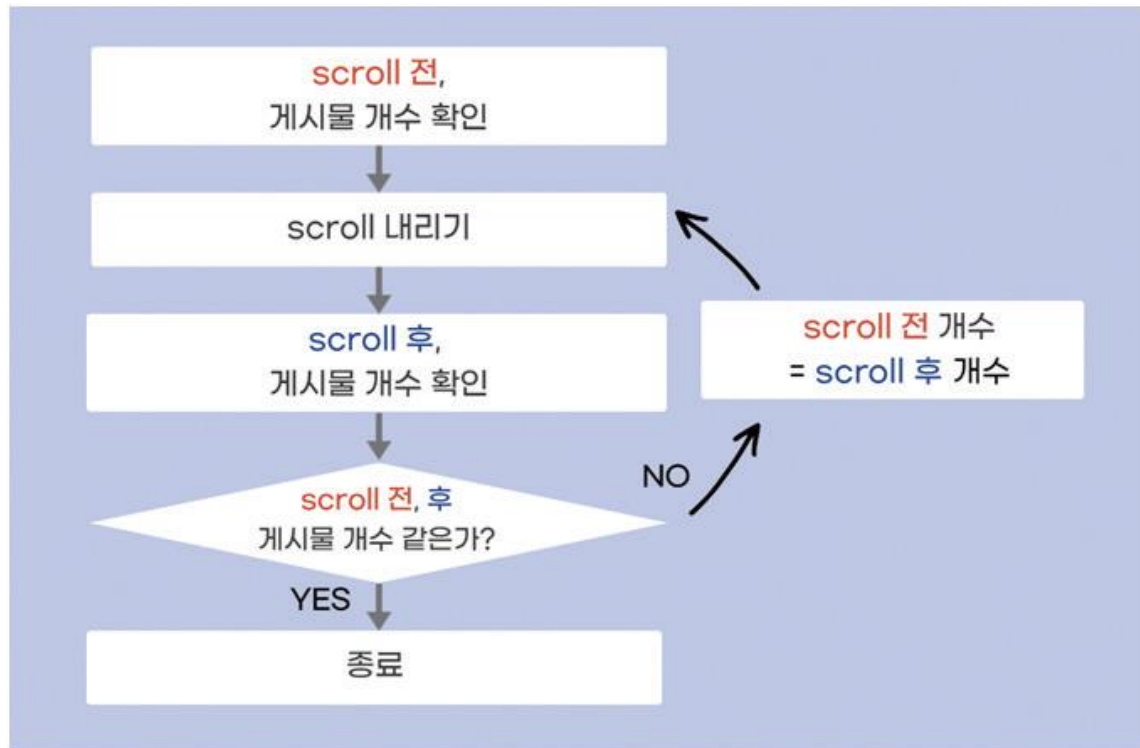
selenium

사용하는 이유

- 동적인 페이지 크롤링 할 때
- 로그인이 필요할 때
- 웹사이트 자동화를 할 때

동적인 페이지 크롤링 실습

무한 스크롤 처리방법



selenium 을 이용한 크롤링 개인 실습 과제

Ajax를 이용한 크롤링

■ Ajax

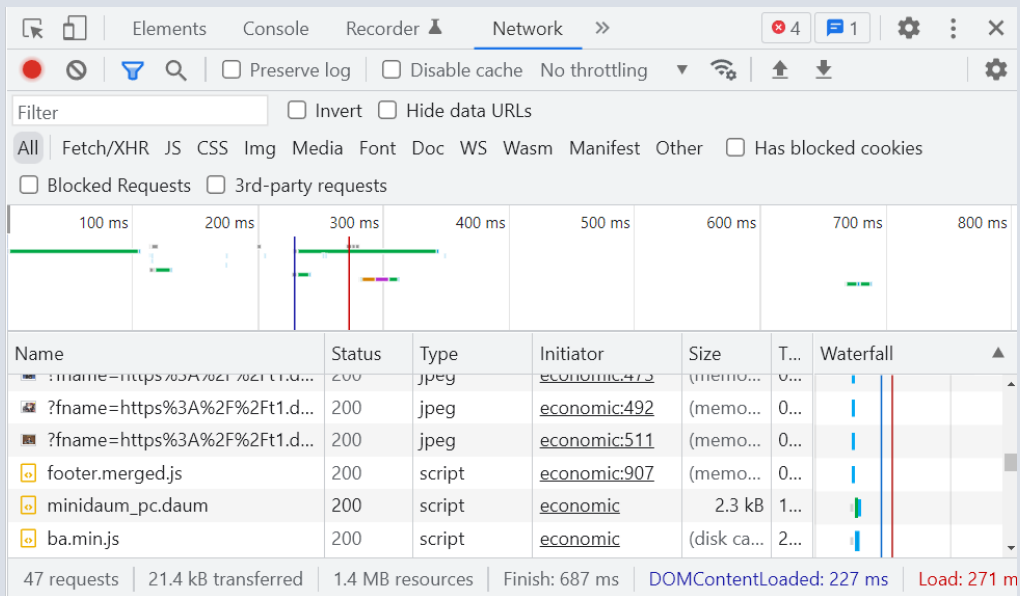
: 자바스크립트를 이용해서 비동기식으로 서버와 통신하는 방식

실시간 경제 뉴스

전체기사보기

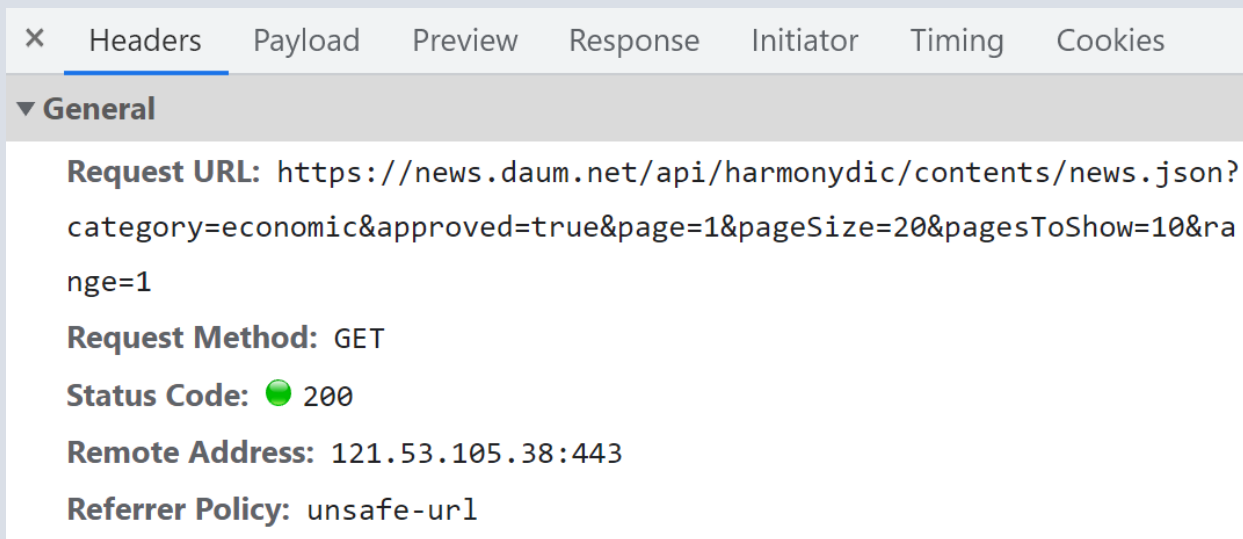
- 22:32 ○ 이재용-최태원-정의선 등 재계총수, 빈 살만과 협력강화 모색(종합) 연합뉴스
- 22:22 ○ 남다른 스케일...객실 400개 빌리고 호텔은 철통 경호 YTN
- 22:06 ○ 여의도에 감원 한파 IB 출신 '별벌'...증권가 1년 새 천당에서 지옥으로 매경이코노미
- 22:06 ○ 집값 단기 조정 중...장기 폭락 없다 매경이코노미
- 22:03 ○ "플렉스가 뭔데요...?" 고물가에 '짬물 소비' 늘어난다 매일경제
- 22:01 ○ '금투세' 유예는 부자감세 vs 개미를 죽여난다 [어떻게 생각하십니까] 세계일보
- 21:59 ○ 고병원성 AI 수도권까지 뚫렸다...전국 확산 조짐 '비상' KBS
- 21:57 ○ 대선공약 '미래 모빌리티 선도도시'...유치 분수령 KBS
- 21:54 ○ 수능 끝...이제 수험표는 '상품권'이다 경향신문
- 21:53 ○ '탄소산업 중심' 전주형 일자리, 관건은? KBS
- 21:53 ○ 구인난 속 구직난...군산지역 일자리 불균형 '심각' KBS
- 21:52 ○ 동해선 철도 건설 과제 산적..."정치권 등 적극 나서야" KBS
- 21:50 ○ 원주 동부순환도로 단절 14년...내년 착공 KBS

- 구글 개발자 도구의 Network 탭
: 서버와 주고 받은 통신 내역을 확인가능



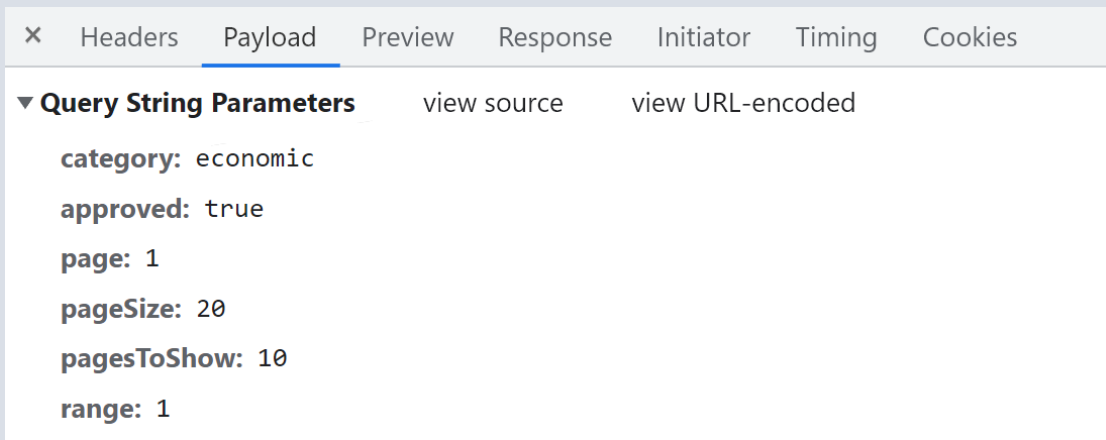
- Headers

: Request URL, Request Header 등을 확인할 수 있다.



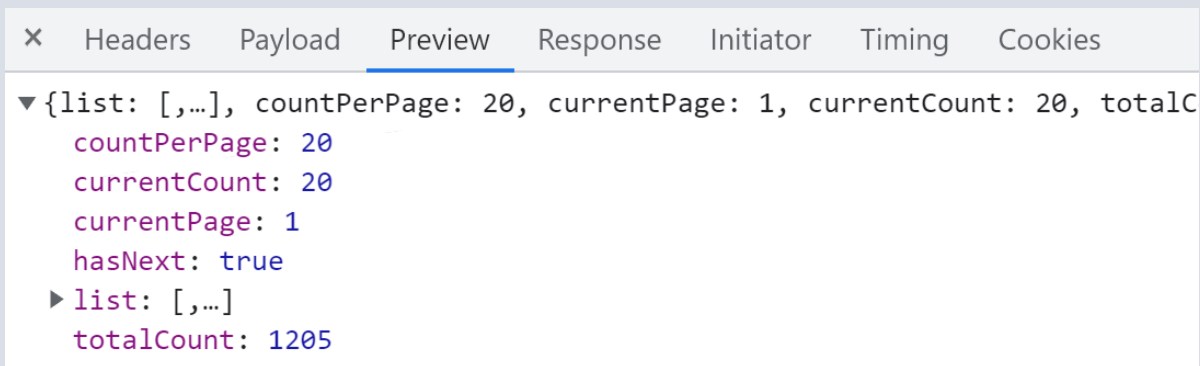
■ Payload

: Request Parameter를 보기 좋게 표시



■ Preview

: Response data를 보기 좋게 표시



Ajax 크롤링 실습

프로젝트

사후테스트