

주식 시세 현황 분석



소프트웨어융합대학원
진혜진



1. 프로젝트 개요

2. 웹 기본 지식 이해하기

3. 웹 스크래핑

4. 데이터 수집 및 가공

5. 데이터 시각화



프로젝트 개요

1. 프로젝트의 목표

■ 웹 페이지의 데이터 추출하기

- 웹 페이지에서 확인한 데이터를 옮길 때 문제점
 - 옮겨 적다가 오타가 발생하거나 일부 데이터를 누락할 수도 있음
 - 데이터 분량이 많은 경우 일일이 모든 페이지를 열어 옮겨 적는 것은 매우 비효율적임
- 웹 스크래핑(Web Scraping)
 - 웹 페이지를 구성하는 실제 소스 코드를 읽으면서 원하는 데이터를 추출하는 기술
 - 웹 페이지에서 확인한 데이터의 정보를 가져올 때 사용함
- 이번 프로젝트의 목표
 - 웹 페이지를 구성하는 기본 기술(HTML, HTTP 등)에 대해 이해하기
 - 웹 스크래핑 기술을 사용해 웹 페이지 데이터를 추출하는 방법 학습함

1. 프로젝트의 목표

✓여기서 잠깐! 웹 스크래핑과 웹 크롤링(Web Crawling)의 차이점

- 웹 스크래핑
 - 특정 웹 페이지의 데이터를 추출하는 데 초점을 맞추고 있음
- 웹 크롤링
 - 웹 페이지에 있는 웹 주소들을 추출해 웹 스크래핑을 할 수 있도록 대상 페이지를 찾는 개념

2. 문제 정의

■ 주식 시세 현황 분석하기

- 웹 사이트 데이터를 추출 → 원하는 방식으로 분석 → 웹 스크래핑 사용

이직을 하면서 퇴직금을 수령했는데, 꽤 목돈이라서 수익이 좋은 곳에 투자하고 싶습니다. 최근 주식 시장이 많이 하락했다고 들어서 여러 주식 지표에 관심을 가지게 되었고, 전망이 좋은 종목에 투자하기로 했습니다.

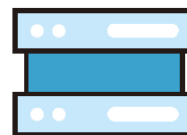
주로 방문하는 주식 시세 확인 웹 사이트에서 괜찮아 보이는 종목을 선택한 뒤 해당 값들을 제가 원하는 대로 분석하고 그 결과를 보관하고 싶은데, 이 웹 사이트에서는 데이터를 다운로드하는 기능을 제공하고 있지 않아요. 그렇다고 데이터를 눈으로 보면서 수작업으로 옮기자니 시간이 너무 오래 걸릴 것 같습니다. 좋은 방법이 없을까요?



HTML 웹 사이트



웹 스크래핑



데이터

웹 페이지 데이터를 추출할 수 있는 웹 스크래핑



웹 기본 지식 이해하기

1. 웹의 탄생

■ 웹(Web)

- WWW(월드 와이드 웹 World Wide Web)의 줄임말
 - 인터넷에 연결된 컴퓨터를 통해 사람들이 정보를 공유할 수 있는 전 세계적인 정보 공간
 - 그림이나 문자와 같은 데이터를 교환하는 전송 방식을 뜻하기도 함
- 웹 브라우저
 - 웹은 세계의 여러 대학과 연구기관에서 일하는 물리학자들의 신속한 정보 교환과 공동 연구를 지원하기 위해 탄생함
 - 이러한 정보를 열람할 수 있도록 웹 브라우저가 제작됨
 - ✓ 윈도우 엣지, 크롬, 사파리와 같은 웹브라우저가 정보를 주고 받기 위해 고안된 도구임

2. HTML로 웹 페이지 만들기

■ HTML(Hyper Text Markup Language)

- 다양한 문자와 그림들을 웹 브라우저에서 볼 수 있도록 표현하기 위해 사용하는 기술
- 웹 페이지 표시를 위해 개발된 표준 마크업 언어

■ HTML로 단순한 웹페이지 제작하기

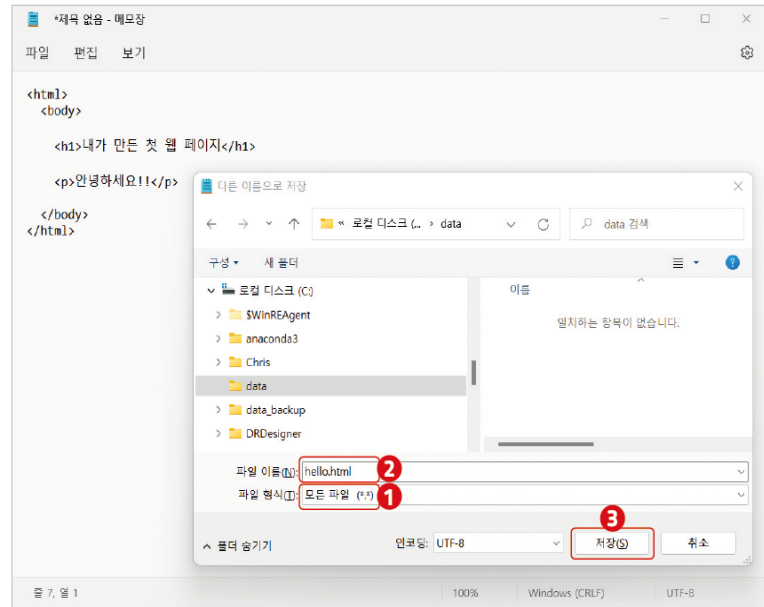
- 기본 텍스트 편집기인 메모장 사용
- 메모장 열기 → 다음 코드 입력 → <Ctrl> + <S>로 저장

```
<html>
  <body>
    <h1>내가 만든 첫 웹 페이지</h1>
    <p>안녕하세요!!</p>
  </body>
</html>
```

2. HTML로 웹 페이지 만들기

■ 웹페이지 저장 후 실행하기

- 저장 위치는 원하는 폴더로 설정



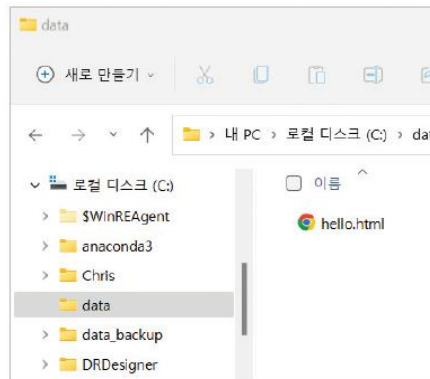
hello.html 파일 생성

- ① 파일 형식은 '모든 파일 (*.*)'로 선택합니다.
- ② 파일 이름은 hello.html로 설정합니다.
- ③ <저장> 버튼을 클릭합니다.

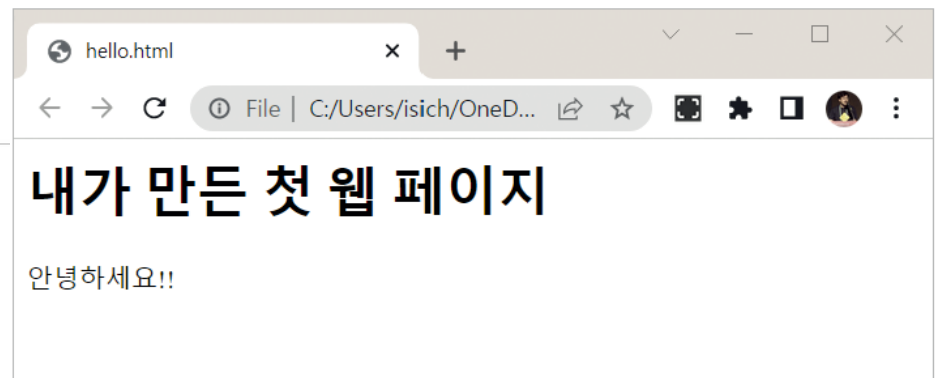
2. HTML로 웹 페이지 만들기

■ 웹페이지 저장 후 실행하기

- 저장한 파일을 윈도우 탐색기에서 확인
 - 웹 브라우저가 인식하는 파일을 의미 : 파일 아이콘이 웹 브라우저 모양으로 지정되어 있음
 - 해당 파일 더블클릭 → 웹 브라우저가 열리면서 HTML 파일에 입력한 문장이 보임



윈도 탐색기에서 hello.html 파일 확인하기



hello.html 파일을 웹 브라우저로 열기

3. HTML 분석하기

■ hello.html의 내용

```
<html>
  <body>
    <h1>내가 만든 첫 웹 페이지</h1>
    <p>안녕하세요!!</p>
  </body>
</html>
```

■ 태그

- HTML 문서를 이루는 문법적인 표시 : 문자를 꺾쇠 기호(<>)로 감싼 형태로 작성
- 시작 태그와 종료 태그로 구성되고 그 사이에 내용이 삽입됨
 - 시작 태그 : (<>)
 - 종료 태그 : 문자 앞에 슬래시 (</>)
 - 쌍으로 구성된 태그의 경우 시작 태그를 사용했으면 반드시 종료 태그로 닫아야 함

```
<html>  # 시작 태그
...
</html> # 종료 태그
```

3. HTML 분석하기

■ <body> 태그

- HTML 문서의 내용 중 실제로 웹 브라우저에 표기할 내용을 담음
- <html> 태그가 열린 다음에 등장하기 때문에 </html> 태그로 닫히기 전에 </body> 태그로 먼저 닫아야 함

```
<body>  
...  
</body>
```

3. HTML 분석하기

■ <h1>, <p> 태그

- <h1> 태그 : 가장 큰 제목을 표현→ '내가 만든 첫 웹페이지'
- <p> 태그 : 일반 단락을 표기→ '안녕하세요!!'

→ 출력 결과 글자체 모양과 크기가 다른 것을 알 수 있음

✓ <h1> 태그

```
<h1>내가 만든 첫 웹 페이지</h1>
```

✓ <p> 태그

```
<p>안녕하세요!!</p>
```

→ HTML 태그는 각각 웹 브라우저에 어떻게 데이터를 보여줄지 표현하는 용도로 사용함

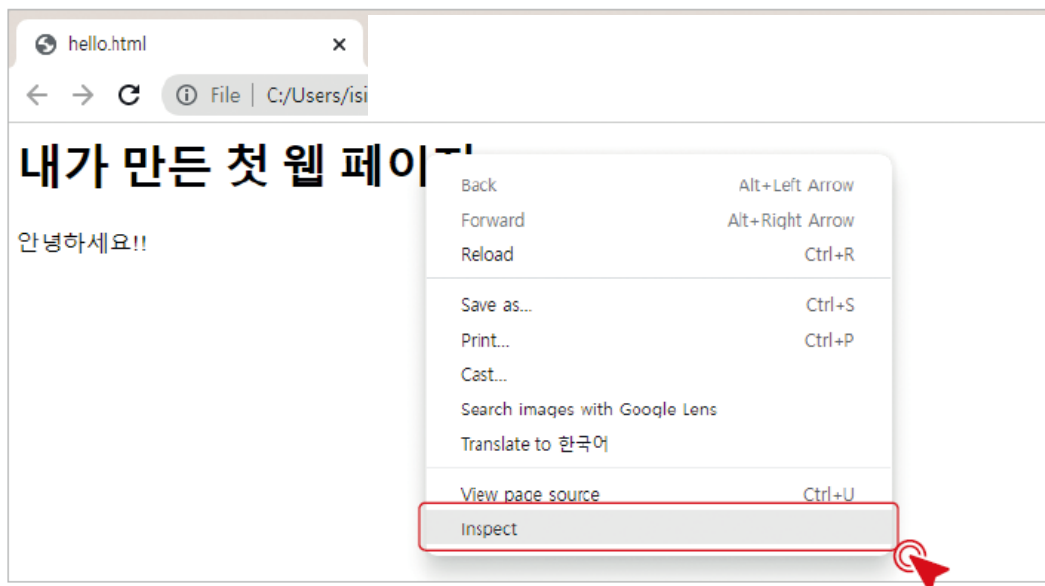
→ 웹 브라우저→ HTML 태그로 구성된 파일을 서버에서 다운로드해 사용자에게 보여줌

4. 웹 브라우저 개발자 도구

■ 웹 브라우저에서 HTML 파일을 확인하는 방법

① 개발자 도구 이용 : 크롬 웹 브라우저 기준으로 확인해 보기

→ HTML 태그를 확인하고 싶은 곳에 마우스 오른쪽 버튼을 클릭 → [Inspect]를 클릭

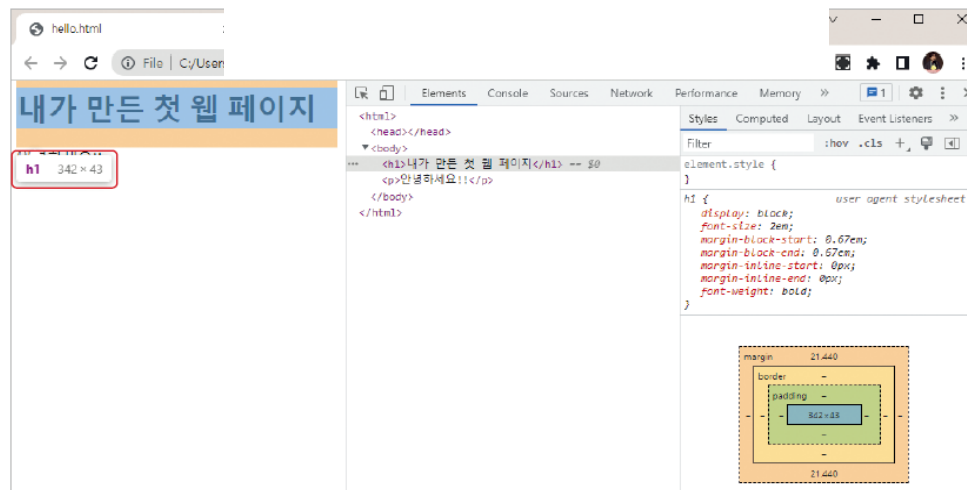


[Inspect(검사)] 메뉴 실행하기

4. 웹 브라우저 개발자 도구

■ 웹 브라우저에서 HTML 파일을 확인하는 방법

- 웹 페이지 오른쪽 : HTML 코드를 확인할 수 있는 개발자 도구 창의 [Elements] 탭
- 왼쪽의 웹 페이지 : 검사를 실행한 영역이 박스 모양으로 표시
- 풍선 도움말 형태로 HTML 태그가 보임
 - 보라색으로 표시된 h1이 이 영역에 적용된 태그임
 - 개발자 도구 창을 보면 이 영역에 해당하는 HTML 코드가 선택되어 있음



개발자 도구로 HTML 확인하기

4. 웹 브라우저 개발자 도구

■ 웹 브라우저에서 HTML 파일을 확인하는 방법

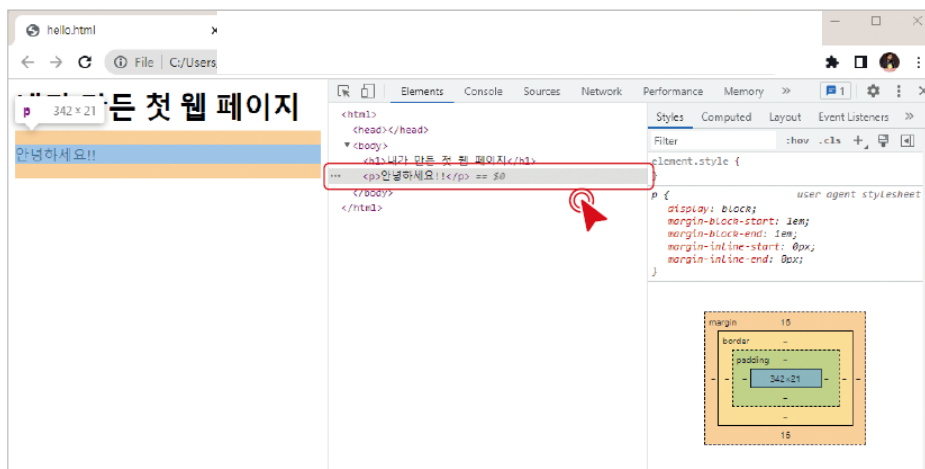
② 개발자 도구 창에서 커서 옮기기

- <h1>태그 영역의 커서를 <p> 태그로 옮기기

→ 왼쪽 웹 페이지 영역 : 박스 위치 이동, 적용된 태그를 알려주는 보라색 글자도 p로 변경됨

→ [Elements] 탭 : 웹 페이지를 구성하고 있는 HTML 태그 정보를 쉽게 확인할 수 있음

✓이 정보는 웹 스크래핑을 수행할 때 수집할 정보를 가리키는 지표가 됨



마우스 커서 옮기기

4. 웹 브라우저 개발자 도구

✓여기서 잠깐! 직접 작성하지 않은 HTML 태그가 보이는 경우

- HTML 문서 속 직접 작성하지 않은 <head>...</head> 태그
 - 기본적으로 작성해야 하는 태그이지만 누락되었기 때문에 내용이 없는 채로 추가되어서 보이는 것입니다. 크게 신경 쓰지 않아도 됩니다.

✓여기서 잠깐! 웹 페이지 스타일을 지정하는 CSS

- HTML : 웹 페이지 구성을 표현하는 주체
- CSS(Cascading Style Sheets) : 웹 페이지의 보이는 스타일을 지정
 - 여기에서는 다루지 않는 영역이니 신경 쓰지 않아도 됩니다..