

작심두달 웹크롤링

파이썬 활용 + 웹구조 학습

1. 초보자를 위한 파이썬 300제 소개
2. Pyautogui 라이브러리 사용법 정리
3. 웹 구조(HTML, CSS selector) 학습
4. BeautifulSoup, requests 학습

초보자를 위한 파이썬 300제



초보자를 위한 파이썬 강의/기초 300 ...



PYSTACK

나중에 시

공유

초보자를 위한 파이썬 300제 Introduction

파이스탁

PyautoGUI 사용법 정리

1. 마우스 자동화
2. 키보드 자동화
3. 메시지 박스

반복적인 업무(노가다)를 컴퓨터에게
시킬 수 있다.

웹 구조 학습

HTML?

웹사이트의 구조를 표시하기 위한 언어
태그로 구성되어 있다.

태그(요소)?

〈태그이름〉 내용 〈/태그이름〉

〈h1〉 스타트코딩 〈/h1〉

〈태그이름 속성= "속성값"〉 내용 〈/태그이름〉

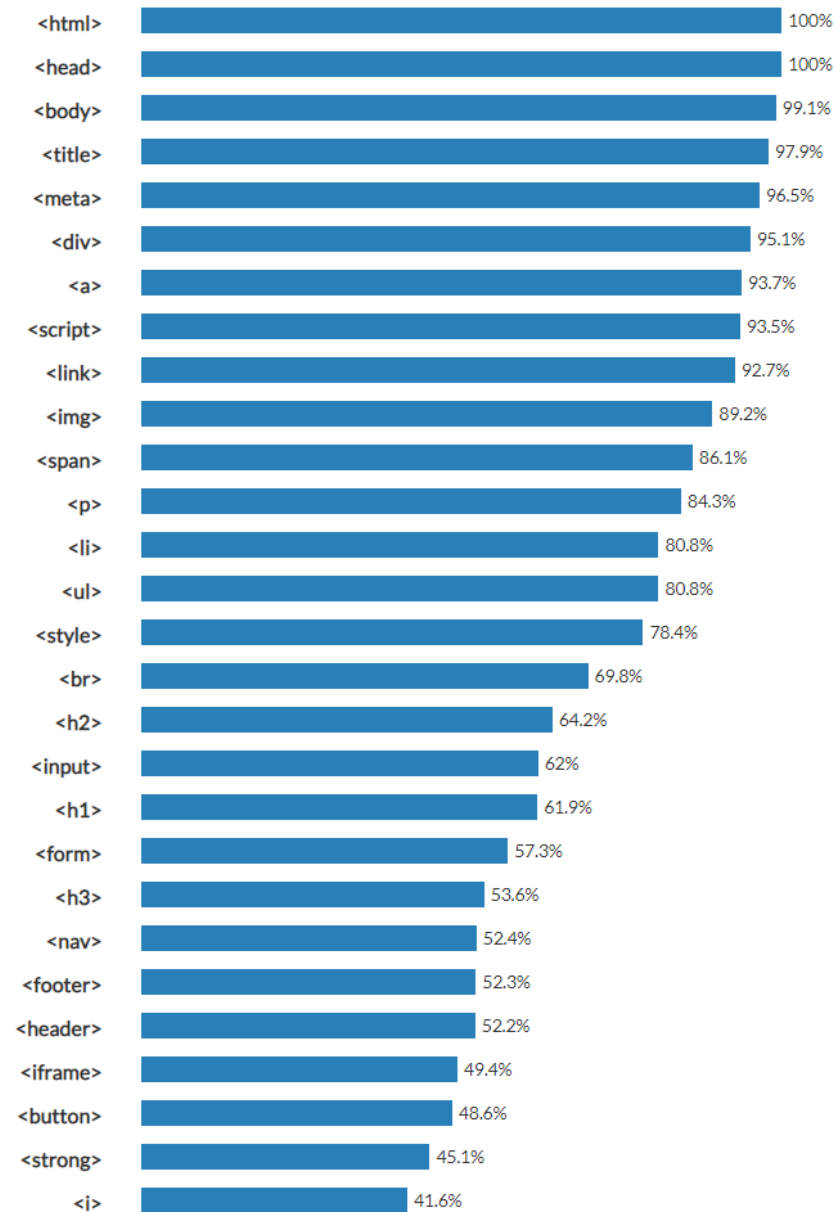
〈a href="https://www.naver.com"〉
네이버로이동 〈/a〉

HTML 문서구조

```
<!DOCTYPE html>  
<html>  
  <head>  
    </head>  
  <body>  
    </body>  
</html>
```

태그의 종류

The **thirty-two elements** used on most pages, ordered by appearance frequency:



문제!

`<strong class="title">전망 좋은 30평 대 아파트 추천`

1. 태그이름?
2. 속성명?
3. 속성값?
4. 내용?

문제!

```
<div id="themecast" class="sc_themecast">  
  <h2 class="blind">주제별 캐스트</h2>  
</div>
```

1. 태그이름?
2. 속성명?
3. 속성값?
4. 내용?

문제!

```
<div id="themecast" class="sc_themecast">  
  <h2 class="blind">주제별 캐스트</h2>  
</div>
```

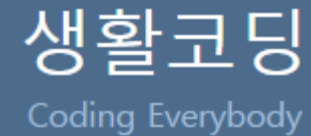
1. h2의 부모
2. div의 자식

문제!

```
<div id="themecast" class="sc_themecast">  
  <h2 class="blind">주제별 캐스트</h2>  
  <div class="theme_head">  
    <strong class="group_title">오늘 읽을 글</strong>  
  </div>  
</div>
```

1. strong의 조부모
2. div.sc_themecast의 후손

HTML 학습 사이트

The logo for w3schools.com, featuring the text "w3schools.com" in a sans-serif font. The "w3" is in dark blue, "schools" is in dark blue, and ".com" is in green.The logo for 생활코딩 (Saenghwal Coding), featuring the text "생활코딩" in white on a blue background, with "Coding Everybody" in a smaller font below it.

<https://www.w3schools.com/html/default.asp> <https://opentutorials.org/course/3084>

Beautifulsoup 라이브러리?

설치

```
pip install beautifulsoup4
```


Beautifulsoup 사용법 및 실습

CSS 선택자

문제!

1. `div#header`
2. `a.title`
3. `div#header > span.text`
4. `div.title.red > ul.container > strong.text`

requests 라이브러리

설치

```
pip install requests
```

HTTP 통신

