

```

import requests
from bs4 import BeautifulSoup as bs
import pandas as pd

page = requests.get("https://library.gabia.com/")
soup = bs(page.text, "html.parser")

elements = soup.select('div.esg-entry-content a.eg-grant-element-0')

titles = []
links = []
for index, element in enumerate(elements, 1):
    titles.append(element.text)
    links.append(element.attrs['href'])

df = pd.DataFrame()
df['titles'] = titles
df['links'] = links

df.to_excel('./library_gabia.xlsx', sheet_name='Sheet1')

```

requests 라이브러리 import
BeautifulSoup 라이브러리 import
표로 만들기 위한 pandas 라이브러리 import

https://library.gabia.com/ 주소로 get 요청을 보내고 응답을 받습니다. 상태 코드와 HTML 내용을 응답받을 수 있습니다.
응답받은 HTML 내용을 BeautifulSoup 클래스의 객체 형태로 생성/반환

BeautifulSoup 가 제공하는 기능 중 CSS 셀렉터를 이용하여 원하는 정보를 찾을 수 있는 기능

titles 배열과 links 배열의 값으로 Pandas의 DataFrame 을 생성합니다.

DataFrame 의 to_excel() 함수를 이용하여 엑셀파일을 작성합니다.

colab.research.google.com/drive/1T174QMPVHizNfa8OMhQX8zgZvtU--N1c#scrollTo=ZrkiOhhOwB7Y

웹파싱.ipynb

파일 수정 보기 삽입 런타임 도구 도움말

명령어 코드 텍스트

파일

- Gemini가 작성한 코드로 파일 분석 업로드
- sample_data
- library_gabia.xlsx

400 번째 게시글: '도메인'이란 무엇일까요? (https://library.gabia.com/contents/domain/688/)

401 번째 게시글: 홈페이지, 비워두지 마세요: 도메인 파일 활용하기, (https://library.gabia.com/contents/domain/689/)

402 번째 게시글: 포털 메일 대신 도메인 메일을 쓰는 것이 좋은 세 가지 이유! (https://library.gabia.com/contents/domain/690/)

403 번째 게시글: '도메인 포워딩 서비스'를 이용하여 내 홈페이지의 접속 통로 넓히기, (https://library.gabia.com/contents/domain/691/)

404 번째 게시글: 서버와 일반 컴퓨터, 무엇이 다를까?, (https://library.gabia.com/contents/domain/692/)

405 번째 게시글: KR로 대표되는 'ccTLD'의 등록 주기는 어떻게 될까?, (https://library.gabia.com/contents/domain/693/)

406 번째 게시글: COM, NET으로 대표되는 'gTLD'의 등록 주기는 어떻게 될까?, (https://library.gabia.com/contents/domain/694/)

407 번째 게시글: 도메인, 영원한 소유는 없다! '도메인 라이프 사이클' 알아보기, (https://library.gabia.com/contents/domain/695/)

408 번째 게시글: 도메인에도 '단계'가 있다!, (https://library.gabia.com/contents/domain/696/)

409 번째 게시글: 익숙한 듯, 익숙하지 않은 'TLD'란 무엇일까?, (https://library.gabia.com/contents/domain/697/)

410 번째 게시글: 도메인이란?, (https://library.gabia.com/contents/domain/698/)

```

import requests
from bs4 import BeautifulSoup as bs
import pandas as pd

page = requests.get("https://library.gabia.com/")
soup = bs(page.text, "html.parser")

elements = soup.select('div.esg-entry-content a.eg-grant-element-0')

titles = []
links = []
for index, element in enumerate(elements, 1):
    titles.append(element.text)
    links.append(element.attrs['href'])

df = pd.DataFrame()
df['titles'] = titles
df['links'] = links

df.to_excel('./library_gabia.xlsx', sheet_name='Sheet1')

```