

Webscraping

Profile mit Photos der Mitarbeitenden IS1

Richard Conrardy

1. Downloaden von [anki_webscraping.py](#)

```
import requests
from bs4 import BeautifulSoup
import csv

# Base URL
BASE_URL = "https://www.phbern.ch"
START_URL = "https://www.phbern.ch/ueber-die-phbern/personen?f%5B0%5D=institut%3A895"

def scrape_phbern_profiles():
    page = 0
    results = []
    visited_urls = set()

    while True:
        print(f"Scraping page {page + 1}...")
        url = f"{START_URL}&page={page}"
        response = requests.get(url)

        if response.status_code != 200:
            print("Error fetching page", response.status_code)
            break

        soup = BeautifulSoup(response.text, 'html.parser')

        # Find profile links and names
        profiles = soup.select('a[href^="/ueber-die-phbern/personen/"]')
        if not profiles:
            print("No more profiles found. Exiting.")
            break

        for profile in profiles:
            # Clean up the name to remove excessive spaces
            name = " ".join(profile.text.strip().split()).strip()
            href = profile.get('href')
            if href and name:
                full_url = f"{BASE_URL}{href}"
                if full_url not in visited_urls:
                    visited_urls.add(full_url)

            # Check for an image on the profile page
```

```

        image_url = None
        profile_response = requests.get(full_url)
        if profile_response.status_code == 200:
            profile_soup = BeautifulSoup(profile_response.text, 'html.parser')
            image = profile_soup.select_one('img[src^="/sites/default/files/"]')
            if image:
                image_url = f"<img src=\"{BASE_URL}{image['src']}\">"

        # Only add profiles that have an image
        if image_url:
            results.append({"name": name, "image_url": image_url})

    page += 1

    return results

if __name__ == "__main__":
    profiles = scrape_phbern_profiles()

    # Save or print the results
    with open("phbern_profiles.csv", "w", encoding="utf-8", newline="") as csvfile:
        csvwriter = csv.writer(csvfile)
        for profile in profiles:
            csvwriter.writerow([profile['image_url'], profile['name']])

    print(f"Scraped {len(profiles)} profiles with images and saved to phbern_profiles.csv.")

```