

Übung 04

Holen von Daten mittels **Playwright** und **Pandas**

INFI-IS

5xHWII

Albert Greinöcker

October 20, 2025



CCA - COMPETENCE CENTRE

HTL Anichstraße

Ziel der Übung:

- Umgang mit Playwright, vor allem bezgl. Automatisierung von Browsereingaben
- Export der Daten mit Pandas

1 Installation und Einrichten von Playwright

Grundinstallation:

```
pip install playwright
```

Browser-Komponenten installieren:

```
playwright install
```

2 Veranstaltungen der Treibhaus-Seite

Die Beispiele liegen in den pandas-examples unter:

ex_05_pandas_playwright_demo und ex_05_pandas_playwright_demo_moodle

Bitte Daten zu den Veranstaltungen (es reicht Datum, Beginnzeit, Link und Beschreibungstext) vom Treibhaus als CSV speichern. Das erreicht man mit folgenden Schritten:

- a. Chromium-Browser Öffnen (aber mittels playwright)
- b. Navigieren zu <https://treibhaus.at/programm>
- c. Auslesen der einzelnen Ergebnisse (Text und Link) und diese Über Pandas als CSV-Datei speichern. Dafür muss der dynamisch erzeugte Quellcode analysiert werden. Diesen bekommt man über die Web-Entwickler-Werkzeuge [F12], dort unter Inspector (Firefox) bzw. Elements (Chrome)

Folgende Tags und Attribute könnten hier wichtig sein:

- `<div class_="event-date">`, danach ``
- `` also `page.locator('span[itemprop="name"]')`

3 Aufgabe mit Parameter-Übergabe

Die Beispiele liegen in den pandas-examples unter:

ex_05_pandas_playwright_demo und ex_05_pandas_playwright_demo_moodle

Folgende Schritte sind auszuführen:

- Chromium-Browser Öffnen (aber mittels playwright).
- Navigieren zu <https://duckduckgo.com>
- Herausfinden des Suchfeldes (über den Quellcode [STRG + U]) und Eingabe eines Suchbegriffes (automatisiert, nicht manuell).
- Warten, bis die Ergebnisseite geladen ist.
- Erzeugen eines Screenshots der Ergebnisseite
- Auslesen der einzelnen Ergebnisse (Text und Link) und diese Über Pandas als CSV-Datei speichern.

3.1 Abfragen von Daten aus einen Login geschützten Bereich

Hier ist ersichtlich wie man z.B. in Moodle die Kurse, in die man eingeschrieben ist, abfragt:

ex_05_pandas_playwright_demo_moodle. Es ist keine Benutzerinteraktion notwendig, es werden die Eingabefelder automatisch ermittelt und befüllt.

Bitte selbst überlegen von welchen Quellen man Daten holen möchte. Hier ein paar Vorschläge:

- Das bestehende Moodle-Beispiel erweitern
- von Github, z.B. die aktuellen Repositories abholen
- <https://books.toscrape.com/> Den Buchbestand so weit wie möglich beziehen (ist nur eine Demo-Seite)

Geht das auch bei großen Anbietern wie Amazon? (nicht ausprobieren)

Beschreibung `.env` (wie im Beispiel

ex_05_pandas_playwright_demo_moodle verwendet)

Beispiel einer beliebigen `.env`-Datei:

```
DEBUG=True
DATABASE_URL=postgres://user:password@localhost:5432/meinedatenbank
SECRET_KEY=abc123xyz
PLAYWRIGHT_HEADLESS=False
```

Installation: `pip install python-dotenv`

So wird dann zugegriffen:

```
1 from dotenv import load_dotenv
2 import os
3
4 # .env-Datei einlesen
5 load_dotenv()
6
7 # Zugriff auf Variablen
8 debug_mode = os.getenv("DEBUG")
9 db_url = os.getenv("DATABASE_URL")
10 print("Debug:", debug_mode)
11 print("DB:", db_url)
```