

Webautomation mit Selenium

Wie man einen Platz im Hochschulsport bekommt

Julian

Linx AG Blitztalk

October 25, 2019

Layout

- 1 Motivation
- 2 Die Buchungsseite
- 3 HSP
- 4 Selenium

warum?

Die Hochschulsport Buchung ist ein Krampf:

- Überlastete Server (weil mehrere Buchungstermine zur gleichen Zeit)
- First come first serve Prinzip ...

Wie kann man sich helfen?

- Credentials speichern hilft! (siehe spätere Abbildung)
- Maus / Keyboard Makros
- EIN SKRIPT!

Wie läuft die Buchung ab?

- 1 Kurs der Wahl finden auf der Auswahlseite
- 2 Auf der Seite mit Kursbeschreibung den Buchungs-Link finden
- 3 Persönliche Daten eingeben
- 4 Ein mal bestätigen
- 5 Pdf mit Buchungsbestätigung runterladen

Kursauswahl

Badminton						
141	Anf/Di	14.03-16.00	Spo 1	14.10.09.02	Flabbecca Dreier	40/50/60 € zum Angebot
Badminton Po						
157	Mi	16.03-18.00	Spo 2	14.10.09.02	Kerstin Zitz	20/30/40 € zum Angebot
	Mo	16.03-18.00	Spo 2			
	Do	16.03-17.00	Spo 2			
158	Di	16.03-17.00	Spo 2	14.10.09.02	Paula Oberdörstel-Spahr	20/30/40 € zum Angebot
1606						
	Mi	16.03-18.00	Spo 3			
	Mo	16.03-18.00	Spo 2			
	Do	16.03-17.00	Spo 2	14.10.09.02	Kerstin Zitz, Paul Oberdörstel-Eppler	35/55/75 € zum Angebot
	Di	16.03-17.00	Spo 2			
Badminton						
1701	Anf/Mo	16.03-17.00	OH	14.10.09.02	Conchita Caserio de Schumann	35/45/55 € zum Angebot

(a) Kursliste

The screenshot shows the course details for 'Badminton Training' on the University of Tübingen website. The page includes a header with the university logo and navigation links. The main content area is divided into several sections: 'Angebotsummer', 'Anmeldung', 'Anmeldungstermin', 'Anmeldungsort', 'Anmeldungskosten', 'Anmeldungshinweise', 'Anmeldungsbogen', 'Anmeldungstermin', 'Anmeldungsort', 'Anmeldungskosten', 'Anmeldungshinweise', 'Anmeldungsbogen'. The 'Anmeldung' section is highlighted in red. Below this, there is a table with columns for 'Anmeldungstermin', 'Anmeldungsort', 'Anmeldungskosten', and 'Anmeldungshinweise'. The table contains several rows of data, including dates and times, and a 'zum Angebot' link.

(b) Kursdetailseite

Kursbuchung: HTML Formulare

```
▼<form name="bsform" method="post" action="https://buchung.hsp.uni-tuebingen.de/cgi/Anmeldung.fcgi" data-onsubmit="chkForm"> event
  <input type="hidden" name="fid" value="d7aa912b21c1c638b73f7781774337ba7e20519c4c0a5e41f7b4692e">
  ▼<div id="bs_form_content">
    ▼<div id="bs_form_head">
      ▼<div id="bs_uni">
        <div class="bs_space"></div>
        <div id="bs_uni_text">Hochschulsport der Universität Tübingen</div>
      </div>
      ▼<div id="bs_ag">
        ▼<div class="bs_form_row">
          ▼<div class="bs_form_spl">
            Angebot:
            <div class="bs_form_extext">Offer</div>
          </div>
          ▼<div class="bs_form_sp2">
            <div class="bs_ag_text">l41</div>
            <div id="bs_ag_name">Badminton Anf.</div>
          </div>
        </div>
        ▼<div class="bs_form_row">
          ▼<div class="bs_form_spl">
            Zeitraum:
            <div class="bs_form_extext">Time period</div>
          </div>
        > <div class="bs_form_sp2">⌵</div>
      </div>
      ▼<div class="bs_form_row">
        > <div class="bs_form_spl">⌵</div>
        > <div class="bs_form_sp2">⌵</div>
      </div>
        > <div class="bs_form_row">⌵</div>
        > <div class="bs_form_row">⌵</div>
        > <div class="bs_form_row">⌵</div>
      </div>
```

(a) Ein Teil vom Buchungs HTML-Formular (blau markiert eine Session ID)

Security by Obstruction

▼ \level

```
<input type="checkbox" name="tnbed" value="1">
```

Die

(a) .. Im Buchungsformular

▼ \div class="hs_form_row hs_exspace"

```
<input type="hidden" name="tnbed" value="1 ">
```

▼ <div class="hs_form_row hs_exspace">

(b) .. Im Bestätigungsformular

Das Skript: Erste Version (Sommer 2018)

Herangehensweise:

- Http Formulare analysiert
- Kursseiten, Formularnamen und Session-IDs mit BeautifulSoup4 finden
- Mit dem Python Modul requests Session managen, Formulare abschicken (POST Requests ...)

Das Skript: Erste Version (Sommer 2018)

Herangehensweise:

- Http Formulare analysiert
- Kursseiten, Formularnamen und Session-IDs mit BeautifulSoup4 finden
- Mit dem Python Modul requests session managen, Formulare abschicken (POST Requests ...)

Problem: JavaScript

Das Skript: Zweite Version (Sommer 2019)

Nach langer Pause eine neues Vorgehen:

- Kontrolliere Headless Session eines Mozilla Firefox oder Chrome Browsers mit Selenium
- Imitiere Nutzer Eingaben
- JavaScript, Cookies, Sessions, ... Kein Problem mehr!

DEMO

`https://github.com/JulianFlesch/hsp`

Andere Lösungen

- Browser Plugin in JavaScript, das die gleichen Eingaben simuliert
- ~~Wie beim Ansatz von 2018, aber mit der JavaScript Engine phantomjs~~ **(Support discontinued)**

Selenium

Hier kommt bald ein kleines Intro ;)

Danke fürs Zuhören!