

## **Module: IS 930 Web Scraping for Researchers**

### **Contents:**

In today's data-driven world, the ability to obtain web data is a game-changer for empirical research. Top-tier research papers in marketing, management, accounting, finance, and information systems increasingly rely on data obtained from scraping websites or querying APIs. Examples of such data include customer reviews from Amazon, search trends on Google, Instagram posts, job postings, investment disclosures, and much more. The skill to obtain novel, rich, and unique web data has become critical for business researchers.

This hands-on course equips doctoral students and postdocs in business with the essential skills to collect, extract, and process web data for research purposes. You will follow a structured, step-by-step approach to:

- Identify requirements for effective data extraction
- Automate data collection from websites or APIs
- Parse and clean data
- Navigate ethical and legal considerations

Through practical exercises in small groups, you will develop a fully functional web scraper using the Python programming language, which you can easily adapt for your own research projects later on.

You do not need to be familiar with Python to attend the course. If you have not worked with Python yet, we provide setup instructions before the course starts.

Dates HWS 2025:

23.10.2025, 09:00-17:30: Day 1 (Fundamentals, HTML, Crawling, Fetching, Parsing)

24.10.2025, 09:00-17:30: Day 2 (Advanced Scraping, Methodological and Ethical Issues)

31.10.2025, 11:00-15:00: Q&A and consultation on group project

### **Learning outcomes:**

By the end of the course, students ...:

- ..can evaluate the costs and benefits of automating data extraction from websites and APIs,
- ..can implement web scraping techniques in Python to collect, parse, and clean data from websites and APIs,
- ..can address common technical challenges in web scraping, including dynamic content and scalability,
- ..can explain ethical and methodological considerations related to web scraping.

<b>Prerequisites: -</b> <b>Formal: -</b> <b>Recommended: -</b>		
<b>Obligatory registration:</b> Yes, you must register for the course and the number of participants is limited.	<b>Further Information on the registration:</b>	
<b>Courses</b>	<b>Hours per week</b>	<b>Self-study</b>
Lecture	2	6
<b>ECTS in total</b>	6	
<b>Form of assessment</b>	Group assignment (80%), Class participation (20%) Participants will work in pairs to develop a Python-based web scraper. Detailed instructions and requirements will be provided during the course.	
<b>Preliminary course work</b>	-	
<b>Lecturer/Person in charge</b>	Prof. Dr. Jens Förderer	
<b>Duration of module</b>		
<b>Offering</b>	HWS and FSS	
<b>Language</b>	English	
<b>Learning Goals / Learning Objectives</b>	-	
<b>Range of application</b>	Doctoral program, MMBR	