



LUND
UNIVERSITY

Faculty of Social Sciences

MKVA50, Data Collection from the Internet: An Introduction to Web Scraping, 3 credits

Datainsamling från internet: En introduktion till web scraping, 3 högskolepoäng
First Cycle / Grundnivå

Details of approval

The syllabus was approved by the board of the Department of Communication and Media on 2023-09-28 to be valid from 2024-01-15, spring semester 2024.

General Information

The course is offered as a freestanding course.

Language of instruction: English

Main field of studies

Media and Communication Studies

Depth of study relative to the degree requirements

G1N, First cycle, has only upper-secondary level entry requirements

Learning outcomes

Knowledge and understanding

On completion of the course, the student shall:

- Demonstrate an understanding of the differences between web scraping methods

Competence and skills

On completion of the course, the student shall:

- Demonstrate the ability to scrape simple web pages and transform the resulting data into a spreadsheet

Course content

The Internet offers a trove of data for both academic research and industry applications. This course introduces the student to the basics of collecting data from Internet web pages. After an introduction to the differences between approaches to web scraping, the student will learn to apply basic web scraping methods for collecting data. Limitations to what types of data are obtainable through these methods are also discussed. Data from web scraping can be transformed into a spreadsheet, which can be analyzed through any number of techniques (note: data analysis is not covered in the course). To pass the course, the student will need to demonstrate the ability to create a spreadsheet from web scraped data.

Course design

The mode of instruction is blended learning, involving lectures and seminars. The lectures provide overviews and a comparison of the methods taught, as well as coding workshops that the student can follow at their own pace. During the seminars, the student have the opportunity to clarify questions regarding their coding and general questions regarding the methodologies taught.

Assessment

Assessment is based on the following:

- The submission of two spreadsheets that demonstrate the student scraped data from a website
- A lab report outlining the student's experience of collecting data using the methods taught in the course

The course includes opportunities for assessment at a first examination, a re-examination close to the first examination, and a second re-sit for courses completed in the past year (i.e., a catch-up exam). Within one year after the course has undergone a major change or ceased, at least two additional opportunities are offered for the same course content. After this, further re-examination opportunities are offered but in accordance with the current course syllabus.

The examiner, in consultation with Disability Support Services, may deviate from the regular form of examination in order to provide a permanently disabled student with a form of examination equivalent to that of a student without a disability.

Subcourses that are part of this course can be found in an appendix at the end of this document.

Grades

Marking scale: Fail, Pass.

For the grade of Pass, the student must fulfil the learning outcomes specified for the course. At the start of the course, students are informed about the learning outcomes stated in the syllabus and about the grading scale and how it is applied on the course.

Entry requirements

General requirements for university studies in Sweden

Further information

Applicants are exempted from the general entry requirements of proficiency in Swedish.

Subcourses in MKVA50, Data Collection from the Internet: An Introduction to Web Scraping

Applies from V24

2401 Data Collection from the Internet, 7,5 hp
Grading scale: Fail, Pass