Call for the
**Interdisciplinary project (IDP)**

„Creating and expanding a data crawler for the automated collection of tabular online content“

**Description**

The goal of this interdisciplinary project (IDP) is to create and expand a user friendly concept for automatically collecting structured data from the internet. The resulting program should be able to scan websites for certain types of structures, extract data and prepare it for use in common statistics software packages. Maintaining full usability for users with no coding knowledge is a requirement.

The envisioned method for tackling the project involves the assessment of existing open solutions regarding provided features, then making adaptations and adding functionality where necessary. Finally, an integration of the system into the commonplace workflows of empirical economic research is to be enabled.

The age of the internet is nothing less than a game-changer with respect to empirical economic research. A plethora of questions that previously could only be addressed theoretically due to a lack of real data can today be answered empirically using information publicly available on the internet.

Great amounts of quantitative data relevant to scientific research are publicized on hierarchically structured websites, oftentimes in tabular formatting. Manually copying data into programs such as Excel for further processing and analysis is an unnecessarily tedious process, particularly for large sets of data. Automating these repetitive tasks constitutes a highly utilile solution to this problem.

Concerning the IDP’s organization, the project’s proposed activities include drafting a detailed project roadmap and conducting a requirement analysis. The latter includes an analysis of the gap between the status quo and the target state, which is conducted in close collaboration with the chair. Moreover, an analysis examining the project’s goals regarding feasibility is envisioned.

Following the completion of these steps, the student(s) participating in the project have the opportunity to further familiarize themselves with relevant programming languages and the programming environment. In the following the project is to be completed as jointly agreed upon in coordination with members of the chair.
Upon successful project completion the preparation of a project report is mandatory. The project’s contents and development as well as its results are to be presented to the chair’s staff subsequently.

Pertaining to the mandatory lecture visit during the IDP, the following courses are offered by the chair and aid the prevention of errors (concept-related or otherwise) during the project: "Grundzüge der Volkswirtschaftslehre" ("Principles of Economics") (WI001056(_E)), "Volkswirtschaftslehre I respectively II" ("Economics I" or “Economics II”) (WI000021(_E) respectively WI000023(_E)), as well as "Volkswirtschaftslehre III" (WI000100), alternatively "Economics III" (WI000100).

**Requirements**

- Knowledge in Python
- Experience in using and handling datasets and databanks
- Independent style of work and willingness to take the initiative

**Organization**

Number of participants: 1-2 students
Registration: beginning immediately
Duration: 1 semester (part time, including lecture-free period)

Distribution of ECTS credits:
- 6 ECTS credits: lecture as mentioned above (any of the following: WI001056(_E), WI000021(_E), WI000023(_E) or WI000100(_E))
- 7 ECTS credits: practical project work
- 2 ECTS credits: project documentation
- 1 ECTS credit: final presentation

Please direct applications including a brief cover letter, CV, and transcript of records to:

Julian Hackinger, Lehrstuhl für Volkswirtschaftslehre
julian.hackinger@tum.de