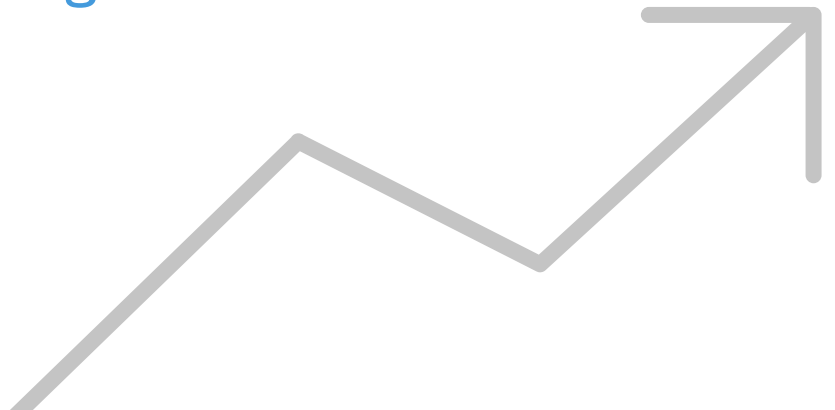
DESTATIS
Statistisches Bundesamt

Web scraping for beginners – Simple automation of online data collections

Conference on New Techniques and Technologies for Statistics

9 March 2021 - 11 March 2021



The growing informational content of the internet

- More and more information on economy, society i.a. is freely available
- Regarding e-commerce:
  - Dynamic universe in terms of product availability and prices
  - Broader variety of products
  - Price setting often nontransparent
  - New phenomena on the online market (dynamic and personalized pricing)

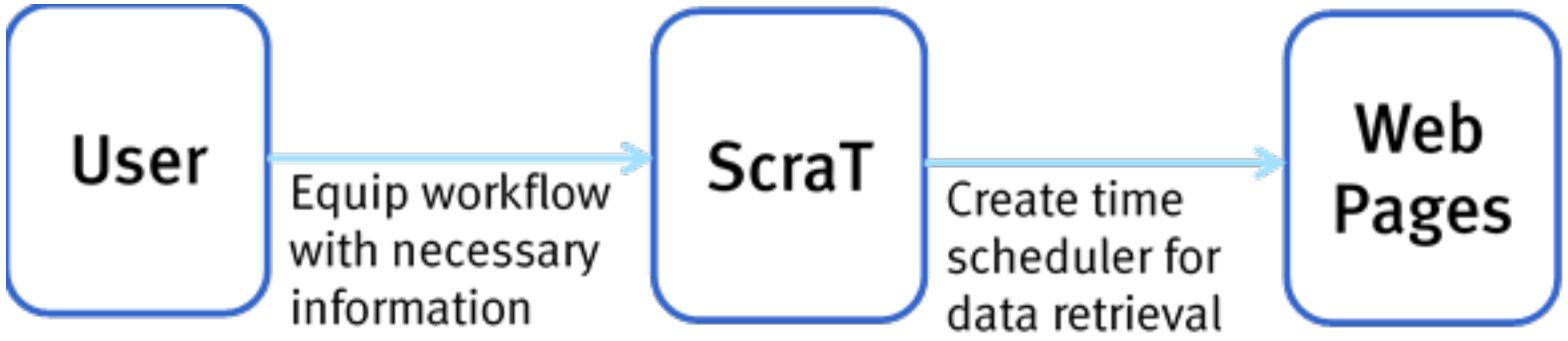
➢ Especially statistical institutes achieve added value utilizing online data sources for the production or validation of their statistics

What‘s web scraping???

- Stands for the automated collection of data on the internet
- Ability to exploit huge amount of data at any given time
- Increase of efficiency and productivity: synergy effects by collecting data on the same web page
- Shift of tasks: less resources needed for the collection of data, more resources needed for the maintenance of programs and analysis of retrieved data
- Profound programming skills necessary

Development of a generic web-based application

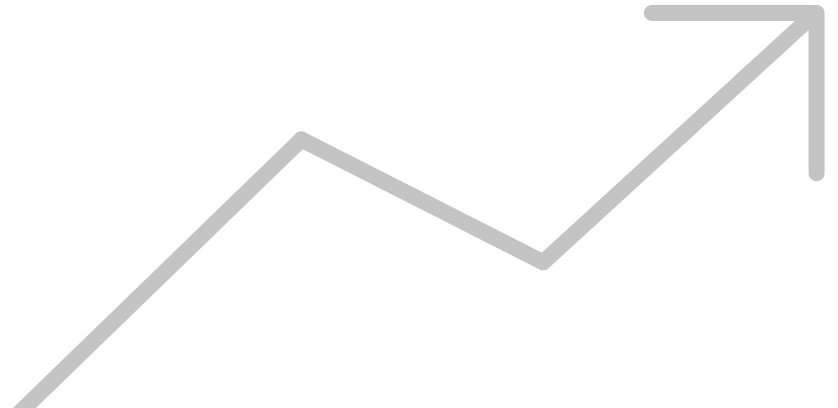
- Name: “ScraT” (compound of the words “scraping” and “tool”)
- Program code is hidden behind user masks
- Motto: “clicking” instead of “coding”
- Process of data collection:



Contact

Statistisches Bundesamt
65180 Wiesbaden
Germany

Contact Person
Christian Blaudow
christian.blaudow@destatis.de
Phone +49 611 75-2178

www.destatis.de

www.destatis.de/kontakt