# Python Programming
# Exercise 12: Build a web crawler

Prof. Dr. Thomas Kopinski

December 15, 2021

**Abstract**

In this exercise you are going to inspect the beautifulsoup package to build a web crawler.

## Task 1: Plotting and explaining

Use the package to implement a scraping tool to extract information from this wikipedia page. Open the website with a browser and make sure you understand the content and the source code.

- In the chapter 'History and relationships to other fields' there a various subchapters, i.e. Artificial intelligence, Data mining etc. Create a pandas data frame carrying the relevant information from these chapters. Create a column for chapter content (the text), the links from this chapter (hrefs), and the footnotes. Each entry (row) should be a subchapter, find an own unique id for it.
- When the first task is done, answer the questions: How often do the word combinations 'machine learning', 'artificial intelligence' and 'data' appear.
- Make sure you use other packages such as regex to achieve the task.
- Remove irrelevant words such as stop words from the text - use additional packages as necessary.