

Python Programming

Exercise 7: The Pandas library

Prof. Dr. Thomas Kopinski

November 17, 2021

Abstract

This exercise will focus on working with data - importing, reading, rearranging, cleaning etc.

Task 1: Tables

- In the repository, you'll find the BBC data set under data/bbc – Import both the raw and the clean version into separate data frames
- Inspect the data sets: What is the difference and can you see it directly by using basic methods such as `head()` or `tail()`
- What kind of results do you get when using the `describe()` function? Does it make sense for all the entries?
- Extract single series from the data frame.
- Sometimes you don't want to work with all the data: Remove columns by choice using the `drop()` function. What test the various parameters - what influence does the `axis` parameter have and what happens when using the `inplace = True` setting?
- Try to access single elements from the data frame - what is the best way to do this?
- clean the data set as good as possible - remove unnecessary items, use imputing to fill gaps, count the NaN values and fill the void with reasonable values, convert the data into required data types.
- Use string cleaning where required: Inspect the data and write functions which can be used to remove unwanted elements from string elements.
- Revert the index of the data frame!