Universität Konstanz

Introduction to Computation for the Social Sciences

Assignment 5

Prof. Dr. Karsten Donnay, Marius Giebenhain, Stefan Scholz
Winter Term 2019 / 2020

Please solve the exercises below and commit your solutions to our GitHub Classroom until Dec, 3rd midnight. Submit all your code in one executable file (*py / ipynb*). You can score up to 10 points in this assignment. You will get individual feedback in your repository.

Exercise 1: HTTP Connection (4 Points)

Working with HTTP connections is essential for many data gathering tasks. The Python module *urllib* provides all functionality we need.

Write a Python function `open_url(url)` that:

- uses *urllib* to establish a HTTP connection to an arbitrary website
- retrieves and prints the first 200 characters of the html resource, i.e. the html source code, of the chosen website
- handles the exceptions thrown by the `urllib.request` function

The basic syntax for exception handling in Python is as follows.

```
try:
    ...
    return ...
except SomeError1 as e:
    # error-specific exception handling
except SomeError2 as e:
    # error-specific exception handling
except
    # general exception handling
```

Exercise 2: Logging Module (3 Points)

The module *logging* in Python provides functionality for logging and debugging purposes. Use the module *logging* to extend the error handling for the function that you implemented to establish a HTTP connection. All exceptions thrown by your function shall be logged as errors.

To accomplish the task:

- write a Python function `init_log(file_name, file_mode, level, format, date_format)` that initializes a custom log file to which all debugging

information and errors are appended using a format that includes the
date, time, level and the message of the logging event
- log occurring errors by calling `logging.error(…)`
- close the log after completing your task by calling `logging.shutdown()`

If you choose not to complete *Exercise 1*, test the logging functionality with a few
examples of your own.

Exercise 3: Download a File (3 Points)

In *Exercise 1*, you used the module *urllib* to establish a HTTP connection. You
can also use the module *urllib* to perform simple file downloads.

Write a Python function `download_file(url, path)` that:
- checks whether the input URL points to a .txt file
- if the input URL points to a .txt file, uses the module *urllib* to download
  and write the text file to the given path on your machine
- logs an error *"No text file found at given URL, download aborted!"* to the
  log file created in *Exercise 2* if the input URL does not point to a .txt file.
- properly handles exceptions

Use the function `download_file()` to download William Shakespeare's drama
Macbeth as a plain text file *here[1]*.

[1] https://ia802707.us.archive.org/1/items/macbeth02264gut/0ws3410.txt