Bioinformatics Group
Prof. Dr. Rolf Backofen
Florian Eggenhofer
Michael Uhl
Stefan Mautner

Algorithms and Datastructures
WS 2017 / 2018

http://www.bioinf.uni-freiburg.de/Lehre/

UNI FREIBURG

Exercise sheet 4

Deadline: Tuesday, 21.11.2017 11:00 AM

Exercise 1 (15 points)
Implement a *geo_names_analyzer.py* file with the following functionalities:

1. A function *read_info_from_file* to get the described information from the *GeoNames.org* tab-separated table file *allCountries.txt* which you can find on our website. All lines in the file with a *P* in column 7 are localities (cities, villages ..). You can find the locality names in column 2. Only consider localities with > 0 inhabitants (column 15). You can find the country code in column 9.

2. A function *compute_names_by_sorting* which calculates the most-frequent world-wide locality names via sorting (using lists only).

3. A function *compute_names_by_map* which does the same, but instead of sorting lists via the usage of an associative array (use Python's *dictionary* class). Write some unit tests for both functions. Do not use the *allCountries.txt* file for testing, but instead generate a small table yourself (see lecture) or use the table for Austria (*AT.zip*) from the website.

   ATTENTION: Do not upload the *allCountries.txt* file into the SVN directory!

Hints:

- Read in the file and process it line by line, do not store the file in a data structure
- If you want to you can also read the file line by line without unpacking it to save space (using the *zipfile* module)

Exercise 2 (5 points)
Write a function *compare_runtimes* which calculates the 3 most-frequent locality names world-wide using your two functions (map versus sorting lists). Compare their runtime. Write down the results as well as a short discussion comparing the two methods in your *erfahrungen.txt* file.

Exercise 2 (5 points)
Change one of the two counting functions (or both) such that it again computes the 3 most-frequent locality names world-wide, but this time with the restriction that the locality name should be present at least once in Germany too (country code DE).

Commit
Commit your code into the SVN in a new subdirectory **uebungsblatt_04**. Commit your feedback in a text file *erfahrungen.txt* as usual. There specify which tasks have been difficult for you and where did you have problems? How much time did you spend to solve the exercises?