Programming task for CNI/NFDI4BIOIMAGE job offer

March 02, 2023

Scientists acquire lots of data for their research projects stored in files, directories, databases, notebooks or just their head. It is easy to lose track of what data is where and how it relates to one another, let alone sharing and explaining data with other scientists. We want to make this much easier by proposing standards, give advice and guidance, as well as providing software tools.

In this task you are given a directory full of data from different measurements for an artificial research project. It consists of 100 small photos of objects taken on a certain date and measurements (histogram, statistics, segmentation) on these photos stored in various files, one photo and its measurements belong to one run of an experiment. Unfortunately the data is not very well organized, just by date for one experiment run. The following table offers an overview of the files and their content:

Data type	Extensions
photo	jpg, png, bmp
object	txt
histogram	tsv
segmentation	gif, tif
statistics	csv
date	in file names

There are multiple problems with this organization: 1) file naming is not too informative, 2) all files are stored flat in one directory, which makes them difficult to browse and find, 3) file formats have been used inconsistently, 4) the data is missing a high-level description and overview for sharing with other scientists.

Your tasks are:

0. Extract the attached ZIP file to a directory such as `unorg_data/` and familiarize yourself with the contained data.
1. Design an improved data organization concept, addressing the above problems.
2. Implement a Python program for organizing the data according to your concept. It should run from the command line with two arguments, the input path of the data and the output path for the reorganized result, e.g., to run it like this: `python organize.py unorg_data/ org_data/` where `unorg_data/` and `org_data/` are the input and output directories.
3. Hand in your Python program and your organized result as ZIP file, along with a very brief description of your concept (this can be part of your program or result).

These tasks shall give you the opportunity to gain more insight into challenges we may be facing and showcase your programming capabilities and creativity.

Please send your solution to torsten.stoeter@lin-magdeburg.de by Friday, March 10th. In case you have any question, just email me.

Looking forward to seeing your solution!