BB1000 Programming in Python
Computer Exercise 9-10:
File handling in Python
(2017-04-28, RB33, 9:00-12:00)
(2017-05-03, RB33, 10:00-13:00)

Update: questions 4, 5, 6, and 7 are optional for the current Computer Exercise 9-10 and will not be graded.

1.(a) Write a program which asks the user for a number, make the sum with '42', and add the original number as well as the result of the sum to an external file of numbers (one number per line). [E,C,A]

1.(b) Give the name of the output file of 1.(a) an extension like ".e1". Calculate the average of all numbers in this file and write the output to a file with extension ".e2". [E,C,A]

2. A file is called ordered if there is an order defined on the data given in the file and if these data are ordered in an ascending order. Imagine that on the two ordered files *f* and *g* each number is given on one line (a number only appears once in this exercise; *f* and *g* may contain different amount of numbers). Write a program which merges *f* and *g* to one newly ordered file *h*. [C,A]

3. Write a program which writes a text on the screen (name: 3_given_file_story), taking into account the line-partition and a screen width of 70 characters. [C,A]

4. Write a script that repeatingly asks for any number until the user enters the word "quit". The program should print the average of the entered numbers, as well as the standard deviation. The results should be given up to 3 digits of accuracy. [E,C,A]

Note: a standard deviation is defined as

$$\sigma = \sqrt{\frac{\sum_i (n_i - ave)^2}{N}},$$

with N being the amount of numbers n_i and ave representing the average.

5. Write a script that interactively creates an address book. The program should ask for your first name and email address, and then save this data in a dictionary, as long as you do not finish the program by typing '0'. At the output, the program is to list the names created in alphabetical order and the corresponding email addresses. [A]

6. Write a script that will return as many numbers in the Fibonacci sequence (on different lines on screen) as the user will ask. The number should be given as a parameter to the program at the moment the job is started. [A]

7. Write a script that asks for any sentence. The script analyzes typographic typographical errors. If the third or last letter of a word in a sentence is large, eg '' Felix eaTs micE and Two pigeonS.'', the program marks the word in capital letters, ie ''Felix EATS MICE and Two PIGEONS '' and prints the modified sentence again at the output. [C,A]

8. Write a program that generates random amino acid sequences and writes them in the form of fasta files (the first line is a title, the rest contains a sequence of max. 70 characters in each line). The amino acids are represented by one letter:

amin = {"A": "Ala" , "C": "Cys", "D": "Asp", "E": "Glu", "F": "Phe" , "G": "Gly", "H": "His", "I": "Ile" , "K": "Lys" , "L": "Leu", "M": "Met","N": "Asn", "P": "Pro", "Q": "Gln", "R": "Arg", "S": "Ser" , "T": "Thr" , "V": "Val" , "W": "Trp", "Y": "Tyr"}

The user specifies how many program files are to be generated and which random sequence length is to be stored in them, eg 3 random sequences = 3 files: sek1.fasta, sek2.fasta, sek3.fasta. The user also specifies the length of the random sequence. The title in each fasta file is 'Sequence: (number of the fasta file)'; for e.g. sek2.fasta, the title is then 'Sequence: 2'. To work with random numbers, use `import random` and `random.choice()`. [A]

9. Write a program which analyses files in pdb format (file: 3IE9.pdb). The program loads the coordinates of the atoms of the protein molecule and writes them in xyz format to the protein.xyz file (the lines starting with '' ATOM ''). The coordinates of all other atoms in the structure (lines starting with ''HETATM'') are written to ''other_residues.xyz ''. [E,C,A]

10. Write a program which analyses files in pdb format (file: 3IE9.pdb). Knowing that the lines in the file beginning with the keyword '' HETATM '' contain the coordinates of the solvent or ion molecules, check that the structure contains copper, magnesium, zinc, sodium or calcium ions. If so, the program should print out the coordinates of those ions shifted by vector [3, 4, 2]. The output file should have the name of the pdb-file, without the .pdb extension and with "_shift.xyz". [A]