

420-SN1-RE Programming in Science - Lab Exercise 14

November 7, 2024

Introduction

Goals for this lab:

- Working with files
- Reading files, iterating through data contained in files

Submission

For this lab you will submit your work in 1 .zip folder. Create a folder named Lab14 and do all your work in this folder. It is really important to be organized and keep all .py and .txt files needed and created for this lab in this same folder.

Exercise 1: Calculating AT content from DNA strings in a file

Your program calculates the AT content of a given DNA string. The AT content is the percentage of (the total number of A's and T's) to the length of a DNA string.

```
#For a given dna string (dna): "AT_count" is the total number of A's and T's
AT_content = (AT_count/len(dna))*100
```

Provided with this lab is the file `dna.txt`. This file contains 12 DNA strings of various lengths (ranging from 700 to 3000). Each string is stored on an individual "line" of the file. (In a text file, a line contains 0 or more characters followed by the newline character "\n" which denotes the end of a line. Refer to the example `fp_iterate.py` in the notes.

Create a new file `Lab14-p1.py`. Write a short program which opens and reads the provided `dna.txt` file. Your program should then analyze the AT content of each DNA string provided in the file.

The output of your program should resemble the output below, with each truncated DNA string (the first 10 characters) followed by the AT content of that string, rounded to two decimals.

```
AT content analysis:
tcataggaag...: 51.80
aagcgtagga...: 52.86
cgacgttctg...: 51.71
cacatacata...: 54.00
acgcggttct...: 52.00
taggtgatga...: 48.43
taggcgcaaa...: 49.14
gtccctaacc...: 49.71
aatgcgccgg...: 50.86
attttagacg...: 51.57
tatcggtcac...: 49.71
atacgctgcg...: 48.50
```

Exercise 2: Counting letter frequencies

Enclosed in this lab is a file, `alice.txt`, that contains the entire text of the book “Alice’s Adventures in Wonderland”¹. I have also enclosed a program named `Lab14-p2.py` that helps you to get started. Your program computes the frequencies of the letters of the alphabet, based on the text in the book. Represent the letter frequencies as a `dict` named `frequencies`, where each key will be a lower case letter, and the value will be the number of times that letter was found in the text. Look at Lab 13, Exercise 1, # 10 for a hint about creating a dictionary with lowercase letters as the keys.

The program should do the following:

1. Use `open()` to access the file `alice.txt`
2. Read all the contents of the file in a str variable `txt` (e.g. `read-all.py`) and close the file.
3. Use a for loop over `txt` to count all of the letters 'a' through 'z' in the file contents. Be sure to convert all letters to lower case, so all letters are counted together. **Do not count non-letters.**
4. Print all of the letters and their frequencies in alphabetical order (e.g. from 'a' to 'z').

Exercise 3: Formatted Printing:

The following 4 lists contain the information about the 8 employees of ABC Inc. Write the program code required to print all employee information in a tabular format. Make sure that your code respects all formatted and tabular printing rules as seen in the notes. Choose appropriate formatting for each field. Write your code in a new file `Lab14-P3.py`

```
employee=['Bob', 'Helen', 'Joanne', 'Robert', 'Sam', 'Stephanie', 'Steve', 'Helen']
salary=[80000, 56000, 67300, 105000, 69100, 70230, 81500, 40000]
status=['full-time', 'temporary', 'part-time', 'executive', 'study-leave',
        'maternity-leave', 'full-time', 'part-time']
vac_pay=[4.0, 4.25, 5.5, 4.75, 3.6, 4.9, 4.5, 5.25]
#Code should print
Bob          80000 4.00    full-time
Helen        56000 4.25    temporary
Joanne       67300 5.50    part-time
Robert       105000 4.75    executive
Sam          69100 3.60    study-leave
Stephanie    70230 4.90    maternity-leave
Steve        81500 4.50    full-time
Helen        40000 5.25    part-time
```

What to hand in

Submission requirements

Submit one .zip folder `Lab14.zip` that contains all the following files:
`Lab14-p1.py`, `Lab14-p2.py`, `Lab14-p3.py`, `alice.txt`, `dna.txt`

¹Courtesy of Project Gutenberg, www.gutenberg.org