

Lab Exercise 07

IT1007: Introduction to Programming with Python and C

Instructions

1. There are two parts in this lab. For Part A, you have to submit within the same day of your lab session. For your Part B, you have six days to work on it. (E.g. if your TLab is on Monday, then your deadline is the coming Sunday midnight.)
2. Complete your code using the skeleton files provided, then **test your code on your computer first** before submitting to Coursemology.
3. To submit your code on Coursemology, click on “Labs” in the sidebar followed by the appropriate “Attempt” button.
4. **Copy ONLY the required function** from your completed skeleton file into the Coursemology code window.
5. Click “Run Code” to test that your function works on Coursemology.
6. Click “Finalise Submission” to submit your code for the **ENTIRE Part A/B. You will not be able to amend your code after you have finalised your submission.**
7. You must name your functions exactly as the questions state.

Failure to follow each of the instruction will result in **10% deduction** of your marks.

Part A (Arrays) [30 marks] (Deadline: TLab 2359H)

Question 1: Code Translation (10 marks)

Translate the following Python snippet into its equivalent C code.

(**Note:** Your solution should: (a) store the values in arrays, (b) produces the **exact same** output as the Python program)

Code Snippet

```
size = input("Start from? ")
intSize = int(size)
myArray = [i for i in range(intSize, intSize + 10)]
reversal = []

for num in myArray[::-1]:
    reversal.append(num)

print("Generated array of 10 elements.")
print(myArray)
print(reversal)
```

Question 2: Debugging (10 marks)

Bob is trying to store 50 evenly-spaced values of $\sin x$ in the range $[0, 2\pi]$ in an array and print every even-indexed value item in said array. However, he is unable to execute his program and needs help debugging it. Correct his code and submit the corrected code on Coursemology.

Code Snippet

```
#include <stdio.h>
#include <math.h> # for sin()

int main() {
    int pi = 3.14159265358979323846;
    int[] array;
    for (i == 0, i <= 50, i++) {
        array[i] = sin(((2 * pi) / 49) * i);
    }

    for (i == 0, i <= len(array), (i++)++) {
        printf("a[" , i, "]" == ", array[i]);
    }
}
```

Question 3: Multiply and Sum (10 marks)

You are given two integer arrays, **a1** and **a2**, of size 10, containing random values.

Calculate the sum of the pairwise products of each element in both arrays. Mathematically, you will need to calculate the following:

$$\text{Sum} = \sum_{i=0}^{\text{length}} \mathbf{a1}[i] \times \mathbf{a2}[i]$$

Question 4: Ends with (Just for Fun) (0 marks)

Write a program that will take in name of a person (at most 20 characters) and prints out the last 3 characters of the name.

Sample Output

```
>>> Micheal
eal
>>> Bharadwaj
waj
>>> Cheng Ho-lun
eng
```

Extra: Modify the program such that it will accept input with spaces (e.g. **Cheng Ho-lun** should return **lun**)

Question 5: Character Frequency Analysis (Just for Fun) (0 marks)

Write a program that will accept, in a single line, a character and word (at most 20 characters) as input and prints out the number of times the character appears in the word.

Sample Output

```
>>> l Basically
Count of l: 2
>>> a Bharadwaj
Count of a: 3
>>> L Alan
Count of L: 0
>>> r error
Count of r: 3
```

Part B (DNA Sequence Analysis)

[65 marks + 5 marks for meaningful code comments]
(Deadline: TLab + 6)

Question 6: Deoxyribonucleic acid (65 marks)

Deoxyribonucleic acid (commonly known as DNA) is a molecule that carries genetic information in a living organism. DNA is made out of **4 nucleotides**, which can be represented using the following symbols: A, G, C and T. Please submit **one program** that incorporates **ALL QUESTIONS** in this part. **ALL NECESSARY LIBRARIES HAVE BEEN IMPORTED!**

- (a) (10 marks) Write a program that will generate an arbitrary DNA sequence comprising of **100 nucleotides**. The generated DNA sequence should be printed out and on top of that, stored in an array. The program should use a **switch** statement and a control loop of your choice. (**Hint:** Refer to Tutorial 2 solutions)

Sample Output

```
AGCTAGGTACAGGATTAAA ... TTCGTACCA
```

- (b) (20 marks) Print out the count and the probability of occurrence for **each type** of nucleotide in the generated DNA sequence.

Sample Output

```
Nucleotide Frequencies:
A: Occurences = 36, Probability = 0.36
G: Occurences = 21, Probability = 0.21
T: Occurences = 25, Probability = 0.25
C: Occurences = 18, Probability = 0.18
```

Counting single nucleotides is not useful for the purposes of analysing DNA sequences. Thus, we need to be able to search for **two nucleotide patterns** in a given set of patterns to be matched. In this exercise, we will be searching for the following patterns in the generated DNA sequence:

{AG, CT, TA, GC, AA}

- (c) (25 marks) For each two nucleotide pattern, identify the location (indices) where the pairs occur in the DNA sequence. For example, given a sequence:

AGCTAGGTACAGGATTAAA ... TTCGTACCA

AG pair occurs at the location 0,4,10 and so on. You need to display the locations for each pair with sufficient clarity. Refer to the sample output given below:

Sample Output

```
Base Pair: AG
Number of Occurences: 11
Indices: (0, 4, 10, 19, 21, ...)

Base Pair: AA
Number of Occurences: 19
Indices: (7, 23, 31, ...)
```

- (d) (10 marks) From the values calculated in part (c), print out the pair with the highest count (frequency).

Sample Output

```
<Previous Question's Output>

Base pair with highest count: GC
```