
CSC 526 Assignment 3

Q1 (Programming in **PERL**, pattern matching):

Choose your favorite song; store the song lyrics in a text file (test input data).

Write a short program that does the following:

- Open the text file
- Read all the lines into an array (or read line by line into a variable, it is up to you).
- Find all words that have ≥ 8 consecutive characters (such as the words “spinning” or “instructions”), use pattern matching technique (regular expression) to find the words.
- Write the words to an output file (the words can repeat / have duplicates).

Estimated time: 1.5 hours

Q2 (Programming in **PERL**, hashes):

Create a hash/associative array, containing (hardcode) the first 10 entries of SS_02.txt (from assignment 2), in to match PDB ID to protein sequence.

e.g.

```
`101M' =>
`MVLSEGEWQLVLVHVWAKVEADVAGHGQDILIRLFKSHPETLEKFDKRVKHLKTEAEMKASEDLKKGVTVLTAALGAILKKGHEA
ELKPLAQSHATKHKIKYKYLEFISEAIIHVLHSRHPGNFGADAQGAMNKALELFRKDIAAKYKELGYQG'
```

- Ask user to enter a PDB ID, the program should return the corresponding protein sequence.
- Provide three test input cases/entries.

Estimated time: 1.5 hours

Submission instructions:

Please submit a paper copy and an electronic copy.

Paper copy:

- Please submit the paper copy at the beginning of the class.
- Provide (create) **test input data** to all programming questions, and capture the related outputs as screen captures (or output files).
- Print the program source code files, test input data and **the output screen captures (or output files)**. If no output screen capture (or output file) is submitted, it would be assumed that the related program does not compile. If the print-out is not readable, **no mark will be awarded**.
- Identify each assignment question by writing the question number at the top of each page.
- Add the following statement to the first page of your submission: “I have abided by the UNCG Academic Integrity Policy on this assignment”. Please write your full name and sign next to the statement. If the statement or the signature is not found, **75% of the possible points will be deducted**.

Electronic copy:

- Please submit a **lastname_firstname_assignment03.zip** (or lastname_firstname_assignment01.rar) file through the Blackboard Digital Dropbox. This zip (or rar) file should contain all submission files.
- Put the answers of all written questions in a **lastname_firstname_assignment03.doc** file.
- For programming questions, you only need to submit the *.pl files.
- Name each *.pl file according to the question number (e.g. Q1_*.pl).

Grading guidelines (programming questions):

Your programs will be judged on several criteria, which are shown below.

- Correctness (50%): Does the program compile correctly? Does the program do what it's supposed to do?
- Design (20%): Are operations broken down into functions / procedures in a reasonable way?
- Style (10%): Is the program indented properly? Do variables have meaningful names?
- Robustness (10%): Does the program handle erroneous or unexpected input gracefully?
- Documentation (10%): Do all program files begin with a comment that identifies the author, the contents, and the compiler used for that particular file? Are all the functions, procedures and data fields clearly documented? Are unclear parts of code documented?

A program that does not compile will get at most **50% of the possible points**.