Lab 2-2: GC-Content with Attitude

Due date: October 7/October 8.

About the Lab

This is the second part of **Lab 2**. In Part 1 you needed to build and deliver in a course of one hour and thirty minutes a working program to compute the GC content for a given DNA sequence. Because you had very limited time to work on the program, the usability, functionality and the overall complexity of the program was reduced to a bare minimum.

We have discussed in class (and so did your BIO 441 partners) how knowledge of GC-content of different regions can be used to learn more about properties of specific parts of the DNA molecule. In Part 2 of the lab you will extend your GC content computation software to accommodate the evolving needs of your BIO 441 partners.

The lab is designed to span two lab periods: October 2 and October 7. The tentative outline of the software engineering process is as follows:

Date(s)	Stages
October 2 lab	Requirements and Design
October 2-7	Implementation (and redesign)
October 7 lab	Testing and delivery
October 8	Final delivery, submission

Lab Assignment

Your BIO 441 partners want to use the ability to compute GC content of DNA fragments for discovery of interesting locations in the genome. Your goal for this lab is to modify your Lab 2-1 program to make their discovery process easier. At the start of your October 2 lab discuss with your BIO 441 partners their needs, and work with them on finalizing their requirements document.

While specific requirements are left to individual teams, *and specific versions of the program each team will develop may be somewhat different from each other*, here are some overall changes and improvements over the **Lab 2-1** version of your program, that each team will need to account for.

Extended Functionality. Your Lab 2-1 version of the program computes the GC content of a single range from the given DNA sequence. In class we have discussed different ways to enhance DNA analysis through repeated computation of GC content over parts of DNA sequences. The requirements from your BIO 441 partners may contain request for functionality implementing such enhancements (unlike this document, BIO 441 requirements will have to specify what is needed precisely).

Expanded Input. Your new program may need to handle larger data inputs and still produce the results fast. It may also need to handle some test cases that you have not seen yet, and that either may force your **Lab 2-1** program to choke and die, throw exceptions, or simply produce incorrect output. If this happens, the requirements you receive will outline precisely the extended input, and possibly, output needs.

New questions. At the moment, your program computes GC content of a given DNA region. In the new version, the information need may change from an outright computation to a discovery task.

Note: When planning your work, please take into account that the intended use of the Tuesday,October 7 lab period is to *test* your program both in terms of accuracy and satisfaction of requirements, but also in terms of usability. Please make sure that your team comes to class with software that your BIO 441 partners can help you test. *Do NOT* rely on the October 7 lab period for significant implementation activities - the implementation stage should effectively be completed by then.

Submission Instructions

You need to submit the following items:

- source code of your implementation
- `README` file with usage instructions/other documentation
- initial version of the requirements document (before you saw it)
- final version of the requirements document
- the binary/executable version of your program
- test cases

Submission via handin. Submit your source and `README` via `handin` using the following commmand:

```
$handin dekhtyar 448-lab2-2 <files>
```

The submitted `README` file must identify all team members (and contain the team name if you have it).

Submission via Piazza. Submit your binary/executable file, both versions of requirments documents and a suite of test cases used on Piazza. Select your team as the "target" of the Piazza posts, and tag them with your team's folder tag.

Please note that part of the deliverables, namely requirements and test cases may be the responsibility of your BIO 441 partners to submit on Piazza. Make certain it happens before the submission deadline.

Deadlines. Each team will perform initial software delivery to the BIO 441 students by the end of the October 7 lab. Please make sure the initial Piazza deliverables are up by then.

However, we are cognizant of the fact that your BIO 441 partners may not have enough time to work with your final version of the software. Because of this, there is a 24-hour *grace period* in effect. During this grace period, you may work on any bugs/issues noticed by the BIO 441 partners and replace the old deliverables on Piazza (and the matching source code you submit via `handin`) with the new ones.

The official **hard deadline** to have the ultimate Piazza and `handin` submissions is *October 8, 11:59pm*.

Good Luck!