Lab 2-2: GC-Content with Attitude

**Due date:** April 16/April 17.

## 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: April 11 and April 16. The tentative outline of the software engineering process is as follows:

| Date(s) | Stages |
|---|---|
| April 11 lab | Requirements and Design |
| April 12-16 | Implementation (and redesign) |
| April 16 lab | Testing and delivery |
| April 17 | Final delivery, submission |

**Note.** It is the task of your BIO 441 partners to provide you with the correct and complete requirements for your software. Because of this, this document is intentionally vague when talking about the functionality and the functional requirements for the Lab 2-2 software. This does not mean that you will have to implement something un- or under-specified: the requirements you get from your BIO 441 partners are expected to be precise and comprehensive. Whenever you discover issues with the requirements, you need to attempt to resolve them with your partners. Failing that, please do not hesitate to contact myself, Dr. Goodman or one of the TAs (Aldrin, Ryan, Jan) and request further clarification. By the time you are doing *program design* and *implementation* the CS side of the team must have a clear picture of the expected software.

Also of note is the fact, that starting this assignment, different teams may get requirements that are somewhat different, and therefore may develop software which will differ from each other. As long as (a) your final software deliverable satisfies the overall goals of the lab as set by Dr. Goodman to BIO 441 students, and (b) satisfis the specific requirements provided to you by your BIO 441 partners, the programs submitted for grading will be allowed to diverge from each other. Because we are fixing the UI in this lab, it is less on issue here, but this is also a warning for future lab assignments.

In a nutshell, your software will be graded against your team's requirements, if your team's requirements satisfy the actual goals of the lab for BIO 441.

## Lab Assignment

Your assignment for this lab is design (or redesign) a program to compute GC content of a DNA sequence in a variety of ways as specified in the new requirements document you will receive from your BIO 441 partners at the beginning of April 11 lab.

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.

**Graphical UI.** Prior to the start of the April 11 lab, we will release code for Java UI in support of your new GC content computation program. **It is a requirement of this lab that your team uses the exact UI provided to you without alterations**.

Your BIO 441 partners are aware of this and should be familiar with how the UI looks. *Please*, make certain that their requirements document *reflects their familiarity* with the mandated UI. Whenever you find significant discrepancies between the requirements and the UI capabilities, please notify myself, or Dr. Goodman immediately.

**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.

**Note:** When planning your work, please take into account that the intended use of the Tuesday, April 16 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 April 16 lab period for significant implementation activities - the implementation stage should effectively be completed by then.

## Submission Instructions

There are two CSC 448 deliverables that will be graded: the program that you develop and deliver to your partners and the program documentation/usage instructions. In order to properly evaluate your program though, we need you to submit two additional deliverables:

- The initial requirements document (as you see it at the beginning of the class);
- The final requirements document (as you modify it throughout the class);

(if your requirements document did not change, submit it twice and indicate that no changes were made).

**Submission via handin.** The full set of deliverables (source code, compilation/running instructions (README), user documentation, and the two requirements documents) needs to be submitted via the following `handin` command.

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

The submitted `README` file must identify all team members (and contain the team name if you have it). The deliverables submitted via `handin` are the ones that will be graded.

**Submission via Piazza.** You must deliver the final (working!) version of your software to your BIO 441 partners via Piazza. While your BIO 441 partners are responsible for posting the two (initial, final) versions of your team's requirements documents there (as well as any test cases/input data on which the program needs to run), the CS side of the team is responsible for posting the final executable/runnable program, and user instructions.

**Deadlines.** It is expected that each team will perform initial software delivery to the BIO 441 students by the end of the April 16 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 *April 17, 11:59pm*.

**Grading vs. Maintenance.** The final `handin` submission you make prior to *April 17, 11:59pm* for each team **will be graded**. However, because your software will continue being in *active use* by your BIO 441 colleagues throughout the quarter, you may be asked to (a) fix discovered bugs, and (b) add some features to the software, that were part of the original assignment, but were missing from the requirements documents (e.g., due to an oversight on the part of your BIO 441 partners). Each team is expected to handle such request graciously and without regard for the fact that the software you are asked to modify has already been graded for the course. (that is - you are working on real programs used by real people for real needs. Your partners **need** your software to complete their coursework.)

**Good Luck!**