CS 4630/5630 Spring 2020 Kresman: Homework 1

Learning Outcome: I can use language libraries to solve basic computational problems in STEM domains

Bioinformatics integrates biology and informatics in the collection and analysis of complex data such as genomics data. Messenger RNA, or mRNA, is a family of RNA molecules that convey genetic information from DNA to the ribosome. We will learn about this topic a bit later in the semester, but for now a simple exercise that, hopefully, perks up your interest.

Messenger RNA: Find all occurrences of the motif CAG in an mRNA sequence - a string containing some combination of the letters A, C, T, G, etc. An example of an mRNA sequence is CCACTGCACTCACCGCACCCGGCCAATTTTTGTGTTTTTAGTAGAGACTAAATACCATATAGTGAACACCTAAGA

Write a python GUI application, with an input text box, output text box and a ‘find’ button. The user enters the mRNA sequence in the input box, and hits find button. The program then populates the output box, for example: 3, 7 (meaning CAG occurs at position 3 and 7). Make sure the input box is a wraparound box so it can hold long strings like the one above, though the physical box is not big. Reasonable documentation please. Non-GUI application, such as command line python code, incurs a 40% penalty.

Later this semester, we hope to learn couple of fast methods Bioinformaticians use to search such genomics data, but for this homework use ONLY the stuff covered through Unit 4.

Use Anaconda iPython. Name for the file prefix is lastNameHW1 (for example, my file prefix is kresmanHW1.)

CS 5250 ONLY: Add a 2nd input box that lets the user to also enter the motif to look for.

Resource:

* Sample programs on Canvas (Files/Programs)
* Unit 4 Slides 11-12

Canvas turn in: lastNameHW1.ipynb and a 2nd file (lastNameHW1RunSnapshot.doc) – run snapshot file includes at least two runs – images that show the input and output.