HORT 53000

Introduction to Computing for Biologists

Spring semester
Credit Hours: 3

Instructor: Kranthi Varala
Contact Information:
Office: HORT Rm #307, Phone: 765-494-0152, E-mail: kvarala@purdue.edu

Lecture: **Tuesday 3:00-4:15 PM Rm#BRNG B282**
Lab: **Thursday 3:30-5:20 PM Rm#Hicks G959**

Course Website: https://www.purdue.edu/hla/sites/varalalab/**idab**/

Objectives

The aim of this course is to train graduate students and seniors in the basic data analysis skills and methods required for most of current and future biological data analysis. This course will teach the basic skills and knowledge required to operate in a UNIX environment and program in Python. The course is designed to teach students in Biology, lacking any computational skills, the skill set required to analyze and interpret their own datasets. The lectures will cover the concepts and knowledge of data handling and programming. The associated lab section will provide the students hands-on experience in working with large data sets in a high-performance computing environment. The students are encouraged to bring their own biological dataset that they will work on throughout the semester. Generic data sets will be provided for students who are yet to generate their own large data sets.

Details

The course is divided into two roughly equal parts. The first half of the course deals with the knowledge needed to operate in a UNIX environment, since most large servers and supercomputers use some version of UNIX as their operating system. During the first six weeks of the course, we will go over concepts in computer architecture, basic UNIX commands, command line tools for text manipulation, shell scripting and finally how to work with supercomputers. The second half of the course teaches the fundamentals of programming using the programming language Python. A full, up to date list of the lecture topics and this syllabus can be found at the course website listed above. The last two lectures will go over special topics that are relevant to the specific projects chosen by the students (see below for details).

Learning outcomes

This course is intended as a way to ease Biology students, with no computational training, into using programming and other computational tools to further their research. The lectures, lab and final project are built so that the students gain the following skills:

1. Familiarity with a command-line interface
2. Use high performance computing (aka supercomputers)
3. Framing data analysis problems as algorithms
4. Basics of programming in Python

Course Books

This course does not use a textbook. All lecture materials will be provided on the course website. However, the students might find the following two books useful during the course as well as good reference books for continued learning of UNIX and programming skills:

1. Unix, in a nutshell. 4th Edition, 2014. Author: Arnold Robbins. ISBN-13: 978-0596100292
2. Learning Python. 5th Edition, 2013. Author: Mark Lutz. ISBN-13: 978-1449355739

Final Project

This course is aimed to teach you practical knowledge in how to analyze your own biological dataset. Therefore, the students in this course are required to identify a programming project that they will initiate before the end of March. This project will serve as the final exam for the course. The students will submit a short description (One paragraph) describing the input data and the desired output for the project. The instructor will communicate with the students individually to adjust the scope and complexity of the project. The student and instructor will then agree on the project description and goals, no later than March 30th. Once defined, this description will be used to grade the quality of the final project submission. The program for this project MUST be written in the Python language and has to run on the Purdue High-Performance computing cluster (Access to the cluster will be provided to students). The project submission HAS to include the input data set, the student's code and all instructions required to execute the code. Project submission deadline will be due 12 PM on the Thursday of the final's week.

Grades

Quizzes (7 in total, Given at the start of the lab section): 70%
Final Project: 30%
The letter grade scale is:
A = 85%-100%, B = 75% -84%, C = 60%-74%, D = 50%-59%, F = <50%.