IS602 Syllabus

Fall 2014

OVERVIEW

Description

This course serves as instruction and guidance for advanced topics in programming. Using Python students will be exposed to a variety of skills and techniques to program for data analysis. Emphasis is on learning basic concepts of a broad range of topics, specific topics are described below. Lectures and tutoring will be provided by the course instructor. Weekly homework and a final project program will be required.

Objectives

- Learn/review basic Python concepts
- Acquire and store data effectively
- Process and manipulate data efficiently
- Create library of reusable code
- Construct basic GUI

Learning Assessment

Students should gain understanding of the course topics. Assessment will be achieved by three methods. First, homework will be assigned on a weekly basis. Homework assignments build on one another and show the learning progression of students. Second, a semester project will be assigned. This project will incorporate the same topics first presenting the in the weekly homework, and will give students the opportunity to explore more completely topics of interest to them. Finally, discussion participation will be used to more fully explain concepts and resolve concerns.

Grading

- Homework 35%
- Project 40%
- Participation 25%

MATERIALS

Required Text

Dive Into Python by Mark Pilgrim (ISBN: 1590593561)

For Purchase: http://www.amazon.com/Dive-Into-Python-Mark-Pilgrim/dp/1590593561

The material in the book is also free at: http://www.diveintopython.net/

Optional Texts

Learning Python, Second Edition by Mark Lutz (ISBN: 0596002815)

This book is good for those who haven't programed or need more review

For purchase: http://www.amazon.com/Learning-Python-Second-Edition-Mark/dp/0596002815

Python for Data Analysis by Wes McKinney (ISBN: 1449319793)

Great book that covers much of what we'll be covering later in the semester.

For purchase: http://www.amazon.com/Python-Data-Analysis-Wes-McKinney/dp/1449319793

Book website: http://pandas.pydata.org/

Required Software

All software tools are freely available and open source. They include:

- Python 2.7.x (http://www.python.org/)
- NumPy and SciPy libraries (http://www.numpy.org/ and http://www.scipy.org/)
- IPython (http://ipython.org/)
- pandas library (http://pandas.pydata.org/)
- Matplotlib library (http://matplotlib.org/)

Optional Software

We won't be using a integrated development environment in class, but many programmers find them useful.

- PyCharm (http://www.jetbrains.com/pycharm/)
- Ninja-IDE (http://ninja-ide.org/)
- PyDev for Eclipse (http://pydev.org/and http://www.eclipse.org/)

Why Python?

Many programming languages and computational tools (Excel, MATLAB, R, C/C++, Java) can be used analyze data. For many new to data programming, Python is a good choice. It is a complete programming language, and it is simple to learn and easy to write and read. You can get started with it right away, and it comes with many great libraries that will simplify your analysis.

COMPONENTS

This course is presented mostly asynchronously online. Every other week we will meet online for instruction and discussion. Students may work at their own pace; however, homework and project deadlines must be observed.

Homework

Homework will be assigned on a weekly basis. Deadlines are marked on each assignment. Since the homework builds on each other, keeping up to date is essential. Therefore late homework will only be accepted with a loss of 5 points per day past the deadline.

Discussion

Some homework assignments contain portions that must be submitted via the course discussion group. This gives students the opportunity to exchange ideas and collaborate.

Project

The project is a solo effort. Final submission will be at the end of the semester. There will be three major milestones that each student complete, they are:

1. Project Proposal
2. Formal Project Specification
3. Completed Program

Topics

- Part 1 : Introduction and Review
    - Introduction/Review Part 1 : Syntax , datatypes, loops, search, sort
    - Introduction/Review Part 2 : functions, classes
- Part 2 : I/O, Parsing, etc.
    - I/O Part 1 : files, exceptions, regular expressions
    - I/O Part 2 : web pages, using an API
    - Basic Datamining
- Part 3: Numpy, Scipy, and pandas
    - Numpy
    - Scipy Part 1 : regressions using scipy
    - Scipy Part 2 : computer vision
    - Pandas
- Part 4 : Visualization
    - Matplotlib
    - IPython Part 1 : Monte Carlo simulation
    - IPython Part 2: parallel computing