

How Fast is my code?

Project 1 Re-Scope (V2)

Please read below for the eventual goal of this project ([Eventual Goal – V1](#)). In order to fulfil the requirements of project, I have had to re-scope this project. Here is the updated problem statement, solution, and the classes required to solve this problem.

Problem Statement:

In order to solve [V1 Problem Statement](#), I need to implement a scaled down version of this problem. The scaled down problem reads as follows: Given a homework implemented in a Jupyter notebook, can we find how long it takes to execute each python cell in this notebook? If so, can we find some form of analysis to showcase these differences between execution times and plot them?

Solution:

To solve the scaled down version of this problem, I will be implementing a Jupyter extension that will allow students to run my solution. The solution allows the student to visualize how long each cell takes compared to the other cells in the notebook. This solution also provides a couple of graphing solutions (Histogram, and BoxPlot), and it also allows the ability to use multiple external sources for visualizing these plots.

Solution Details:

Controller

This class is the driver of all the functionalities provided in this implementation. It has a single method `execute` which takes the arguments provided from the notebook i.e., the notebook location, notebook url, and notebook auth details. Using this information, this class executes a [Processor](#) call which processes the notebook, this class then makes an [Executor](#) call which executes all the python cells, then this class creates as many [IPlotters](#) as needed to plot graphs, and finally the controller sends these plots to the [IDisplayers](#) for displaying purposes.

Processor

This class is responsible for processing all the information provided in the jupyter notebook. It creates a dictionary with a mapping between the cell number and the data contained in the notebook.

Executor

This class is responsible for making sure all valid python cells are executed. It then creates a dictionary between the cell number and the length it took to execute this solution.

IPlotter

This is an interface which provides the functionality of plotting. There are two known implementations as of now: HistogramPlotter, and BoxPlotter. As the names suggest HistogramPlotter plots a histogram and BoxPlotter plots a boxplot.

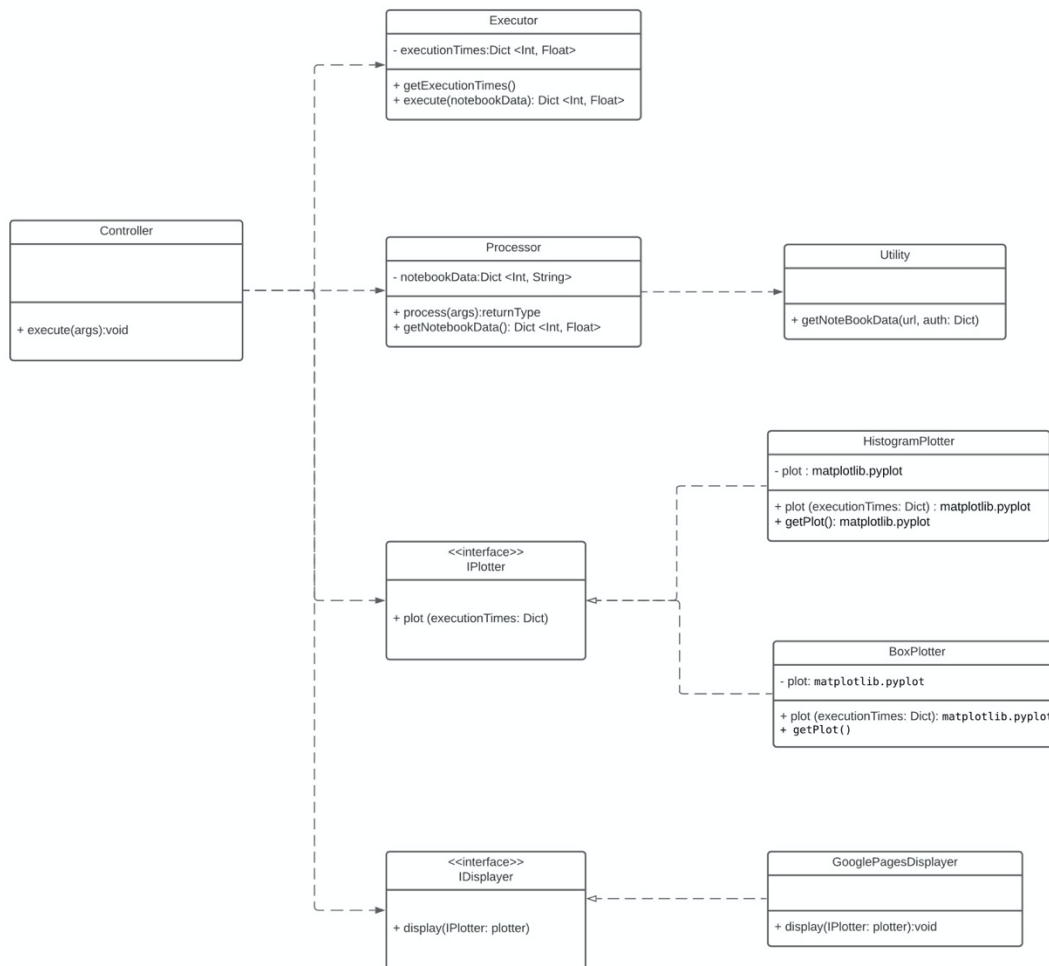
IDisplayer

This is an interface which provides the functionality of displaying. There is one known implementation as of now: GithubPagesDisplayer.

Utility

This class will be used to facilitate different functionalities that other classes would like to outsource and/or functionalities commonly required between different classes.

UML Diagram of Solution:



Eventual Goal – V1

Problem Statement:

Plenty of times, as a student, I write some beautiful code. My code runs really well but then I start doing some experiments to determine how good my code is. In order to determine this, I will ideally do some metric collection. The biggest factor that comes to mind is to see how the performance of my code is. I run my code and my code takes about 0.2 seconds. Wow look at that, I wrote a very complex algorithm that takes about 0.2 seconds. Feeling so proud, I post this to W200-python group then I find out that other students took about 0.1 seconds. So, I guess my code wasn't as good as others. Then I start optimizing and get my code to 0.09 seconds. Then another student posts 0.08 seconds. I get frustrated because I don't have much time now to optimize further since the students took 1 day each to get back with their results, which totaled about 3

days to get the performance of other students code. And the homework is due in 30 mins.

Solution:

What if there was a way to have this information handy? What if I was able to have the students' scores handy (without their name and their details obviously)? What if I could just get the percentile of my code? What if I could get the average it took for all the students code to run? What if I could get the most frequent number of this information? All these are details that are worth investigating and will be useful during my execution of my code.

Motivation:

I ran into this problem in Week 6. I wrote some beautiful code that used recursive backtracking, some form of Permutations, and several other functionalities. Although I knew my code would be slow, I didn't mind it since there was a 7-character limit but then I didn't know how slow my code was compared to other students. I discovered this when I went through the w200-python length of time it took for other students to run their algorithms. Then I realized my code was very slow and needed improvement. It would have been very helpful if I was able to know this information beforehand.

Dependencies:

Since this involves some form of getting students code runtime, I will need to work with the Professors and TAs of W200. I will be needing the following information for each Homework:

1. Time it took for each homework of each student.
 - a. This could be a tag mechanism. For example, we would have a column called homework number and then another column could hold the time it took for that homework.
2. Apis exposed so that I could retrieve these details. I could potentially be willing to write these apis.
3. Changes in the Auto Grader. Maybe one cell could be specifically for performance metric collection.