Department of Computer Science

The University of Hong Kong

Final Year Project

Project Plan

The next generation eLearning platform for computing education

Supervisor: Dr. Chui Chun Kit

Group Members: Cai Tung San 3035090470

Chan Wai Lun 3035089835

Cheung Siu Wai 2013588661

Wong Ching Quen 3035098161

Table of Content

1 Summary

This project plan aims to provide an overview about our final year project, titled "The Next Generation eLearning platform for computing education". The eLearning platform ("The platform") aims to provide students a collaborative programming environment with different features for them to be able to learn better on programming and computing. It can analyze data and deliver timely feedback to provide various aids to students. More explanation will be provided below.

2 Problem Statement

Programming is not an easy subject for learning. One of the challenges that code learners often encounter is the lack of learning support, so the platform aims to auto generate feedback and guidelines to students when they are learning programming. It also wants to facilitate the collaboration and idea exchange between students, thus communication tools like charts, chat rooms will be provided to students.

3 Background

Currently, there are many eLearning platforms available in the education field. Majority of them like Moodle and i-Class focus on learning resources (lecture slides, reference materials...) distribution and assignment submission. However, there is no specialized platform available for programming and computing education. The platform aims to fill this position.

By supporting collaborative programming and various coding support, student could be expected to learn programming in self-learning manner with the aid of the platform or from peers.

4 Scope

In this project, it is assumed that the platform will mainly be used by the students in the course COMP2123 Programming Technologies and Tools. As the course adopts C++ as the teaching language, the system will provide C++ as coding language for students and analysis will be centered around C++. The data analysis will focus on classifying student's performance on passing the test cases. Examples of evaluation metrics for student performance are time elapsed in development and number of test cases passed.

5 Prerequisites

Server and virtual machine are needed for compilation of codes submitted by students. As huge amount of data can be expected to be generated during the learning process of users, high computational power is essential for the platform. Hence, Cloud computing technologies and servers are needed for operation of the platform. PHP framework is used as the base of the platform. Data analytics tool and libraries are needed as machine learning process is involved in the platform. Students in the course COMP2123 Programming Technologies and Tools will be the target users of the platform to provide data to verify and improve our machine learning models.

6 Architecture

This project will create an online eLearning platform. There are two main features in the platform. First, we will provide collaborative tools to help students learning programming. Second, we will carry out data analysis on the data produced by students to generate auto feedback for their programming in editor. Details of the platform will be shown and explained below.
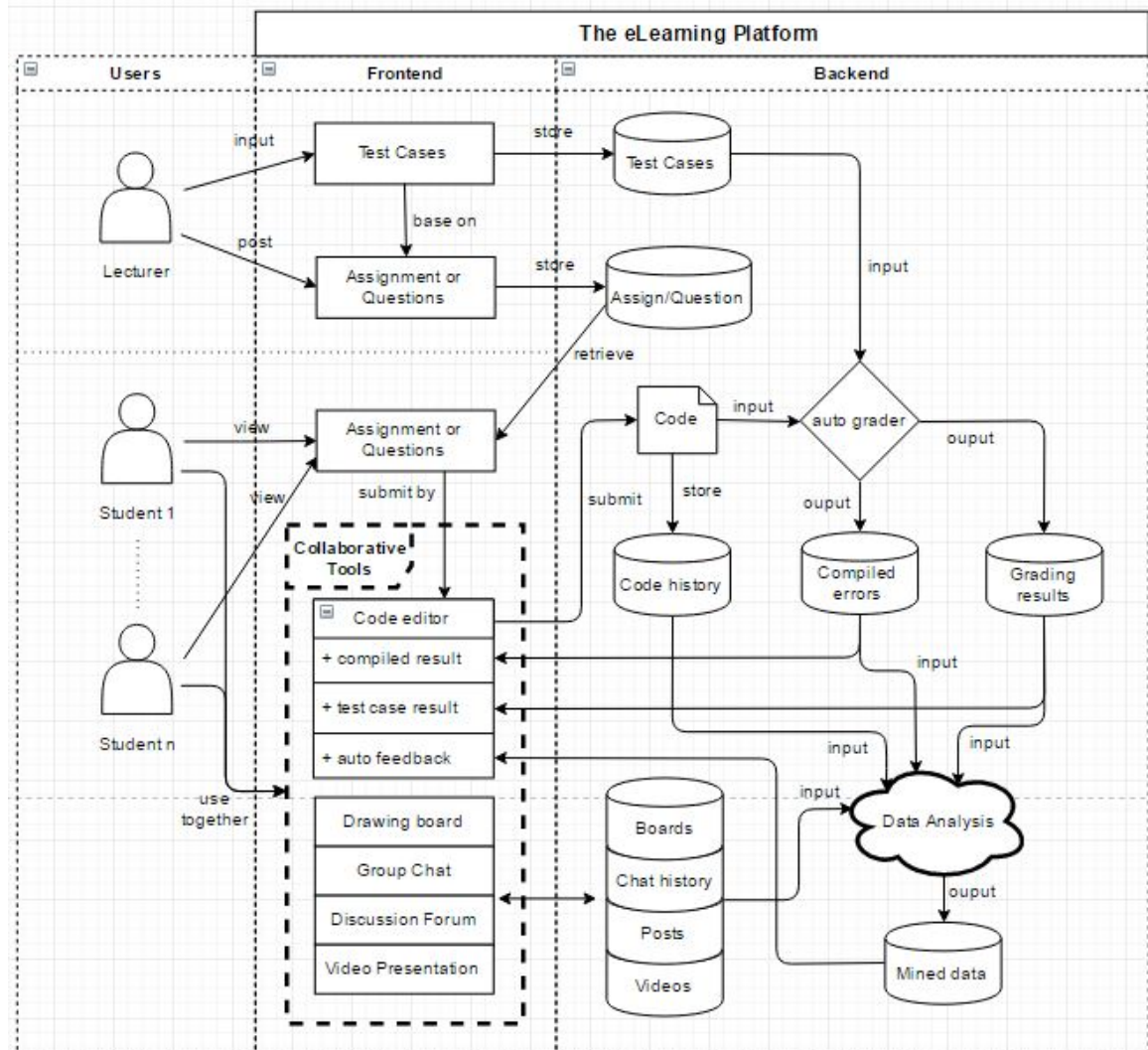
Figure 1. Structure of the online eLearning Platform

6.1 Web system and user interface

Web system and user interface which is the frontend part of the platform are provided for users to interact with the platform.

Laravel[1] is the framework chosen for the platform since it provides many library functions such as user access control and security support such as CSRF protection. This can quicken the development of the platform.

---

[1] Otwell, T. (n.d.). Official Laravel Website. Retrieved September 27, 2016, from https://laravel.com/

Another advantage of adopting Laravel is the huge amount of online support documents. Stemming from the popularity of Laravel, there are many discussions and communities concerning the development of Laravel. Solutions to problems arose in the development process can be expected to be found easily. Thus, development time and difficulties could be reduced by the aid of Laravel.

Laravel use traditional MVC (Model-View-Controller) structure for development to maintain high modularity and easiness.

The Model part handles the storage and query of data which we will use MySQL as relational database and Eloquent ORM included in Laravel to perform database model creation and database query.

The View part which displays result to users, HTML will be the selected language and CSS will be supporting tools for aesthetics purpose and visual effects.

The Controller part which handles the interaction between users and the platform, PHP is the proposed candidate.

6.2 Collaborative tools

One of the objectives of the platform is providing a collaborative environment for computing students. So collaborative tools for idea exchange and elaboration. Drawing board, group chat, discussion forum and video presentation will be provided for them to communicate and share their ideas. Also Code editor is provided for them to code together and submit any assignments or questions posted by the Lecturer.

6.2.1 Code editor

The platform will embed external code editor for the advantage of code completeness and high degree of functionalities support. The proposed candidates for code editor are Firepad[2] and Ace[3].

---

[2] Official Firepad Website. (n.d.). Retrieved September 27, 2016, from https://firepad.io/
[3] Official Ace Website. (n.d.). Retrieved September 27, 2016, from https://ace.c9.io/

Firepad serves as real-time collaborative code editing tool; while Ace serves as a common code editor by providing general features like syntax highlighting, code completion and multiple cursors. Detailed integration tutorial between these two tools is available in Firepad website. Thus, it is believed that low amount of work and time is required for integrating the tools into the platform.

Besides the embed code editor, the platform will integrate auto feedback in it to help students programming. The generation of auto feedback is explained in the data analysis part below.

The platform will also provide area next to the code editor to display the test case result upon the submission of code by students as well as the compiled errors if the code contains any compile errors.

6.3 Compiler

The compiler in the platform collects code input and returns result to user, it is an important role in the platform. The project's supervisor, Dr. Chui, suggests that there is existing compilation environment in the Moodle system. The compiler is hosted by HKU CS server and use VM (virtual machine) for compilation to prevent memory issues, which commonly arise when user commits mistake in learning. Considered the convenience from this existing solution, it is proposed that we bridge the compiler hosted by HKU CS server with the platform.

6.4 Data analytics

The system will use data analytics for classifying test cases into different categories, which corresponds to different programming concepts. Having the ability to categorize test cases allows pinpointing the weakness of student and giving corresponding suggestions to student. Below are the details of it.
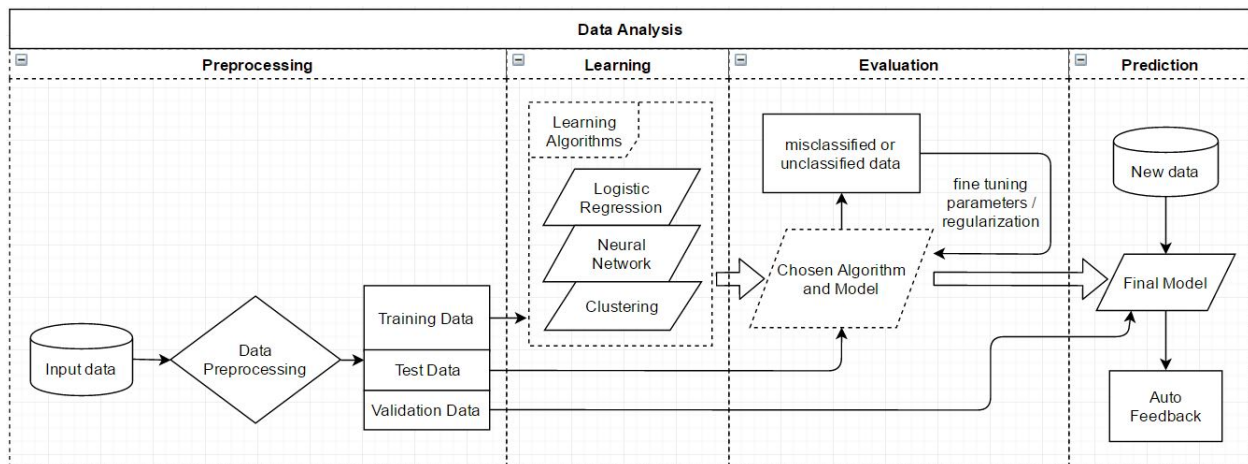
Figure 2. Data analytics workflow

6.4.1 Preprocessing

Data passing in for data analysis will undergo preprocessing before being used. For example, DIFF function will be used to spot any code changes from the code history. The DIFF function's behavior is similar to the counterpart in Bash[4], which is returning the difference between two files. All this will also undergo normalization or standardization for better comparison or visualization of output.

These data serves as the input for the data analysis. They will be split into training data, validation data and test data in about 50-25-25 proportion. The training data is responsible for setting parameters in the classification algorithm. The validation data is responsible for comparing performance between algorithm candidates. The test data is responsible for verification and evaluating the accuracy of the algorithm.

6.4.2 Learning

After having the data for input, learning algorithms need to be decided for classification. There are multiple algorithm candidates, logistic regression[5], artificial neural network(ANN), clustering are the possible

---

[4] Diff Man Page. (n.d.). Retrieved September 29, 2016, from http://ss64.com/bash/diff.html
[5] Hosmer Jr, D. W., & Lemeshow, S. (2004). *Applied logistic regression*. John Wiley & Sons.

choices. The algorithm with highest prediction accuracy will be chosen for the platform's data analytics.

(a) Logistic regression

Logistic regression is a supervised learning algorithm used that have categorized input and output. By minimizing a specific cost function with respect to the parameters, optimal set of parameters which has the least prediction error is obtained. Thus, the obtained parameters can be used to classify new data and the probability of wrong prediction can be expected to be low.

The advantage of it is simple and easy to implement. Moreover, due to its simplicity, its computation is faster. However, this algorithm requires training data to be labelled, extra manpower for labelling may be needed as a result. In addition, the model only supports binary classification in its simplest form. One-vs-all technique can be used for equipping the model with the ability of multi-class categorization.

(b) Artificial neural network (ANN)

ANN is a learning algorithm that is widely used nowadays. Classification is one of ANN's abilities. By feeding labelled data into ANN, optimal parameters can be obtained and classification can be achieved, similar to logistic regression. ANN is good for modeling non-linear regression with large number of input features (e.g. pictures). However, ANN is complex and hard to implement since extra work is needed to understand the algorithm. What is more, ANN is computationally expensive.

(c) Clustering

Clustering is an unsupervised learning algorithm. Under the potential situation that most input data is not labelled, the algorithm can group

similar data entries and allows manual tagging of the group later. Work is reduced compared to supervised learning method. However, there is technical difficulty in the partition process since the K-means algorithm used is NP-hard and required high computational power.

6.4.3 Evaluation

The selected model will be evaluated using test data. Overfitting and underfitting are common problems encountered in machine learning. By using test data to find any misclassified or unclassified data, the parameters of the model will be frequently fine tuning using regularization when evaluation to produce a better and much accurate model.

6.4.4 Prediction
The final model could then be used for prediction. New data by students could be input in it and the model will try to classified the test cases and provide feedback back to the code editor.

7 Feasibility assessment

To auto generate feedback for students, we rely on machine learning algorithm to analyze the performance of students on passing provided test cases on exercises. We believe that we can classify test cases into programming concepts and different difficulties. For classification, we will use hierarchical clustering, which is an unsupervised machine learning techniques for classifying data[6]. Depending on situation, if manual label of data is available, we may use supervised learning model (more specifically, logistic regression or neural network).

Regarding implementation, Python will be chosen as programming language due to the abundance of machine learning libraries. Scikit-learn is the proposed library for providing learning algorithm[7] and Numpy is the proposed library for

---

[6] A., Ng. (n.d.). Machine Learning. Retrieved September 17, 2016, from https://www.coursera.org/learn/machine-learning#

[7] Scikit-learn. (n.d.). Retrieved September 17, 2016, from http://scikit-learn.org/stable/

providing numerical computation[8], which appears frequently in machine learning process.

8 Project management info

Responsibilities	Person(s) in charge
User Interface design	Cai Tung San
Collaborative tools development	Cheung Siu Wai Wong Ching Quen
Data analytics development	Cai Tung San Chan Wai Lun
Testing and Debugging	Cai Tung San Chan Wai Lun Cheung Siu Wai Wong Ching Quen
Documentation	Cai Tung San Chan Wai Lun Cheung Siu Wai Wong Ching Quen

9 Risks, challenges & Mitigation

Risks / Challenges	Description	Mitigation Strategy	Magnitude
Timeliness issue	Project time lag	To prioritize functionalities: Core functions like data analytics will be ranked and processed according	High

---

[8] NumPy. (n.d.). Retrieved September 17, 2016, from http://www.numpy.org/

		to its importance. Effective prioritizing can ensure the timeliness of functionalities' delivery.	
Steep learning curve of programming technologies and tools	There are many programming technologies and tools could be used in the development of the system. Although the team has experience with some of them (e.g. Python, Web programming…), the team has to learn new tools like Numpy, Laravel framework etc.	To continually develop the core functions of the application	High
Difficulty in choosing between Machine Learning algorithms	A suitable candidate has to be determined in order to obtain the best output for the system.	A part of data serves as validation set to compare the performance between different algorithms; To seek advice from experts.	Moderate
Porting to existing compilation server	Technical difficulties to connect the platform with the compilation server provided by HKU CS.	To contact HKU CS Technical staff for assistance	Low

10 Schedule

Date	Progress
August, 2016	Project Planning and literature review.

September	Build user interface and basic functions of the eLearning platform including user login system, place for viewing assignment, place for compiling and executing codes, display basic cases results upon execution of codes and storage of result into the database.
October	
November	Implement more features to the platform including group allocation and group chat, design diagram and discussion forum. Create training and test dataset for preprocessing of machine learning. Design and implementation of the predicative models and algorithms decided.
December	
January, 2017	Fine tune and testing to make sure the system is bug-free before being used by students.
February	The system should be ready. Students of COMP2123 can use the system in their class and data should be collected. Data collected should be used for fine tuning the system.
March	
April	Prepare for final report and presentation

11 Mini-conclusion

It is quite demanding to quickly build the eLearning platform before January. Therefore, we should devote more time on this project and have a better time management so that we can finish the project on schedule.

Also, advices given by the supervisor are also important for the success of the project. We should meet the supervisor at least once in a month so that we can listen to his advice and work through the right direction.

12 Appendices/References

1. Otwell, T. (n.d.). Official Laravel Website. Retrieved September 27, 2016, from https://laravel.com/

2. Official Firepad Website. (n.d.). Retrieved September 27, 2016, from https://firepad.io/

3. Official Ace Website. (n.d.). Retrieved September 27, 2016, from
   https://ace.c9.io/

4. Diff Man Page. (n.d.). Retrieved September 29, 2016, from
   http://ss64.com/bash/diff.html

5. Hosmer Jr, D. W., & Lemeshow, S. (2004). *Applied logistic regression*.
   John Wiley & Sons.
6. A., Ng. (n.d.). Machine Learning. Retrieved September 17, 2016, from
   https://www.coursera.org/learn/machine-learning#

7. Scikit-learn. (n.d.). Retrieved September 17, 2016, from
   http://scikit-learn.org/stable/

8. NumPy. (n.d.). Retrieved September 17, 2016, from
   http://www.numpy.org/