BS Thesis 1 ------------------------

Background:
Principles and practices of software engineering have been significantly improved with the introduction of agile methods instead of the traditional waterfall model, leading to customer and business focus as well as improved software quality. One of the major advantages of agile methods is the faster release of a product. Continuous Integration (CI) is the practice of automatically compiling, building and testing software. It is part of the agile method and has been widely adapted. For example, Google, Amazon and over 900k open source projects are using Travis CI. The adaption of CI provides several benefits such as faster and continuous feedback [1][2], more frequent releases [3][1][2], improved product quality and developer productivity [1][3], increased customer satisfaction [2] and early bug detection [3]. Regression testing, one of the challenges in CI [4][5] must be cost-effective to provide faster feedback to developers. To improve productivity, developers should get immediate feedback, after they submit their code to main base. A high frequency of system builds and tests requires a large amount of time and resources [3]. For example, the feedback time, after testing, to developers in Google is between 45 minutes and 9 hours [6]. It is not an easy task to achieve faster feedback during regression testing. Many researchers have investigated this challenge such as 'regression feedback time' in [4], 'long running tests' in [1][4] and 'time consuming testing' in [5][7]. RTS (Regression test selection) involves relevant and important selection of test cases, from a large repository, and TCP (test case prioritization) involves an efficient re-order of test cases to be executed to provide faster feedback. Similarity based test case selection has been reported as an effective approach for RTS and TCP [8].

Project Description:
This BSc thesis project is a part of research project named AAT (Aspects of Automated Testing), under the umbrella of Software Center (https://www.software-center.se/). At this moment, we have successfully implemented a technique that can find similarities between test cases [8] but the limitation of our implementation is that it can only be applied to specific datasets (i.e. test cases in Python), received from our partners in Software Center. The first aim of this thesis is to integrate this technique with standard input and output, so it can be applied to any other data set. Standard input/output means that we need to convert required information into the proper and general format such as XML or other, due to the fact that each company follows specific language for writing test cases or definitions. This may require an investigation of common standards or frameworks for representing test case’s information. Another important contribution of this thesis is to run the selected test case in CI environment (i.e. Jenkins). This can be achieved by writing or extending a plugin that does not only run selected test cases, but also validate the resulting reduced test suit.

Steps:
1. Find the standard input/output that can be used for input to similarity-based test case selection technique, that we developed as part of AAT3 project. (your contribution)
2. Integrate above methods with technique, that is already developed. (your contribution)
3. Run the technique and identify the subset of test cases (Already developed)
4. Run selected test cases on source code in Jenkins (i.e. we may use open source software for source code) *(your contribution)*

5. Evaluate the test execution data to find the efficiency of selected test cases *(your contribution)*

Outcomes:
The final outcomes of this thesis can be a plugin that can be integrated within CI environment that accommodate the steps provided above.

References:


