Student Submission Integrity Diagnosis [SSID]: User Guide

NUS SCHOOL OF COMPUTING | NUS WEB IR/NLP GROUP

Table of Contents

1. Introduction

Student Submission Integrity Diagnosis (SSID) is a web application suite for managing courses, assignments, staff, teaching assistants, students, and student code submissions and most importantly**,** to detect and visualize plagiarism among these student code submissions.

1.1. About this User Guide

This user guide provides a quick start guide to use SSID. This user guide provides documentation of all the various features offered by SSID and frequently asked questions. To navigate between the different sections, you can use the table of contents above.

Additionally, throughout this user guide, there will be various icons used as described below:

	This indicates an important note. These are essential notes for you to take note when using SSID.
	This indicates an additional note. These are notes that are good for you to know but not essential to note when using SSID.

2. Features

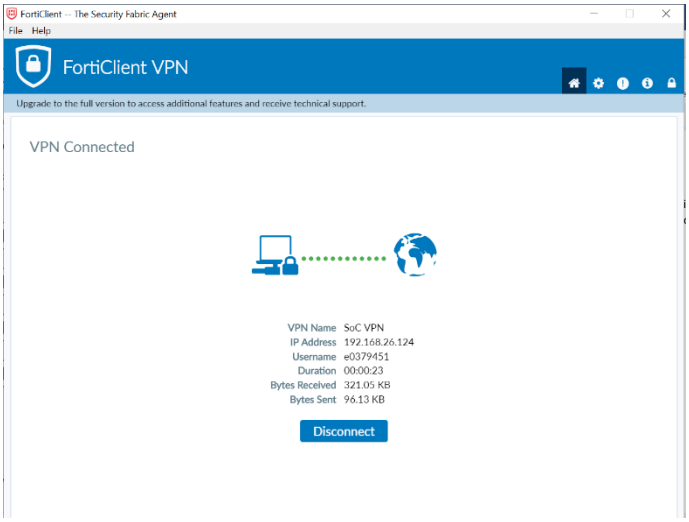
Some of the main features of SSID include:

- Pairwise plagiarism detection in submissions
- Clustering analysis
- Plagiarism history of students

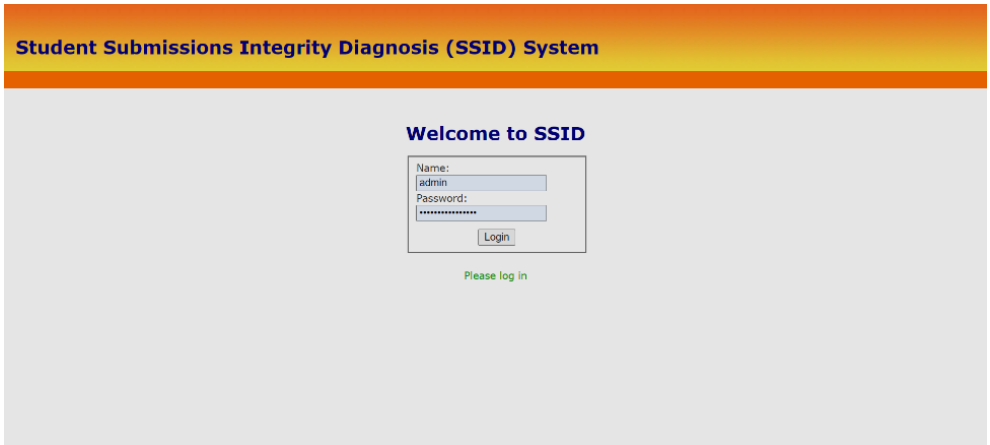
Below, you will get more information on how to use SSID to aid you in detecting and preventing plagiarism.

2.1. Accessing SSID

a) First, ensure that you are connected to the SOC VPN through FortiClient. If you need help in setting up your VPN, please click this link. Once you have logged into SOC VPN successfully, you should see this in your FortiClient console:



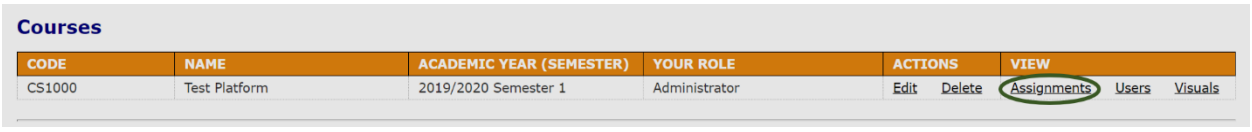
b) After which browse to ssid-i.comp.nus.edu.sg and you will see the below webpage. Please log in to SSID with your given username and password.



2.2. Viewing assignments

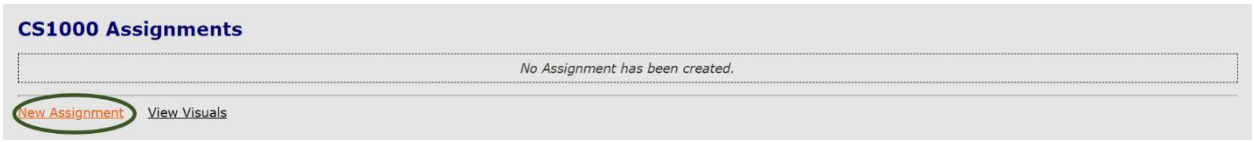
a) To view assignments, click Courses from the top menu.

b) Under the View tab, click Assignments to view the assignments.

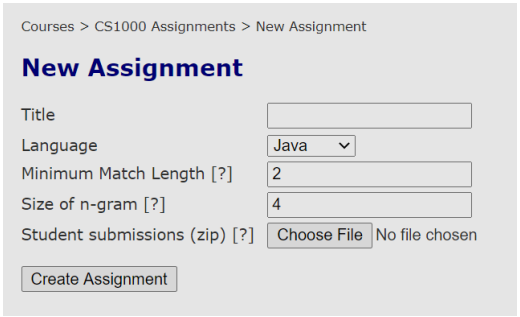


2.3. Creating assignments

a) To create an assignment, under the assignments listing, click New Assignment to continue.



b) Fill in the respective fills and select the zipped submission file (.zip) which contains the student submissions to upload. Then, click Create Assignment to continue.



	Please ensure that the zipped submission contain **1 folder** for **each student**'s code files. Ensure that different students' files are **separated into different folders** which are named according to the student's ID. If you have any codes or files that you wish to exclude from the system (such as sample code that you gave to the students), put those codes or files under a folder called **"skeleton"**.

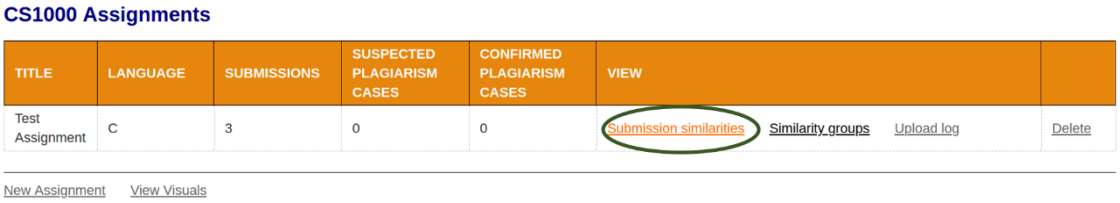
	*Minimum Match Length* refers to the number of words or lines that match between two or more entries. *Size of n-gram* refers to the size of the consecutive token tiles that are used to match.

	If there are multiple files within a student's folder, do take note that all these files will **be merged into a single file** by SSID to process it. Therefore, if the assignment contains several code files for different questions which are not necessary to be compared with each other, please **divide it into different question submission zips and upload separately**. You would be then able to see this merged file after the processing is done under the Submission Similarities tab which will be explained in the next section.

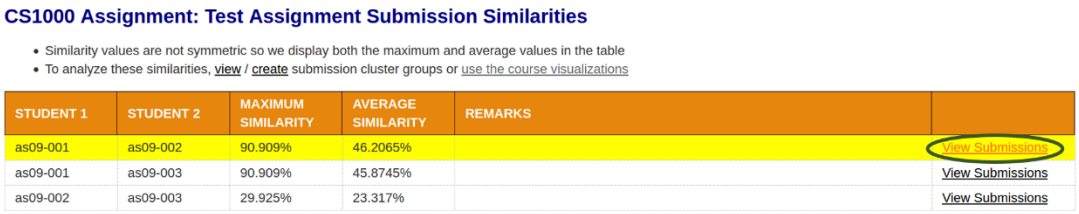
2.4. Viewing pairwise submissions comparison results

	You must have at least one assignment to view pairwise submissions comparison. To create an assignment, refer to section 2.3

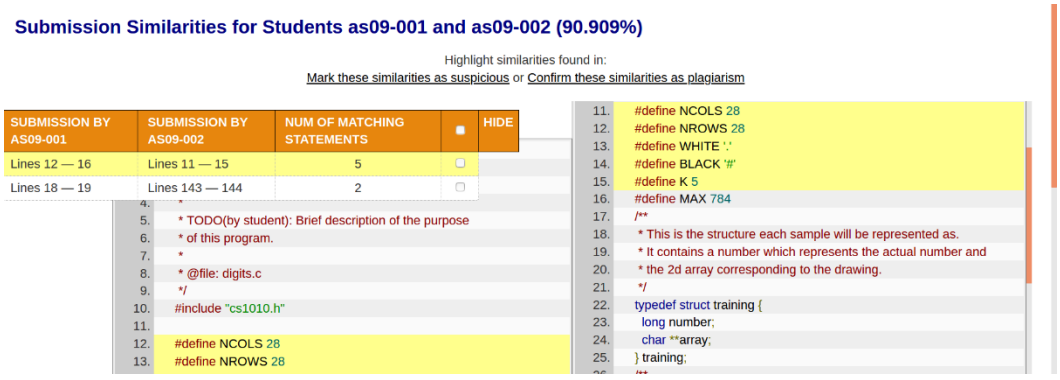
a) To view pairwise submissions comparison for an assignment, firstly, go to the assignments listings. Under your desired assignment, click Submission Similarities under the view tab.



b) From the rightmost column, click View Submissions of the desired pair of student submissions to view the pairwise submissions comparison result.

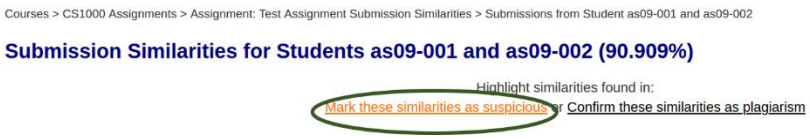


c) Click on the lines to view the similarities.

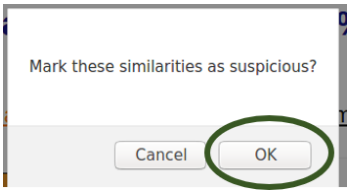


2.5. Reporting a suspicious case

a) Under the pairwise submissions comparison listing (as seen above), click Mark these similarities as suspicious on the top.



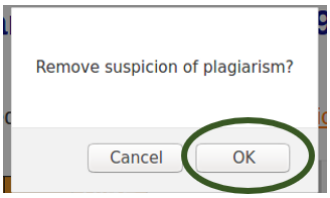
b) In the confirmation popup box, click OK to report this pair of students as suspicious.



2.6. Removing a suspicious Case

a) Under pairwise submissions comparison listing, click Remove suspicion on the top.

b) In the confirmation popup box, click OK to remove this pair of students as suspicious.

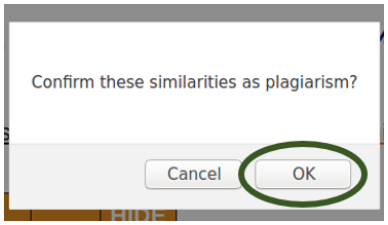


2.7. Confirming a plagiarism case

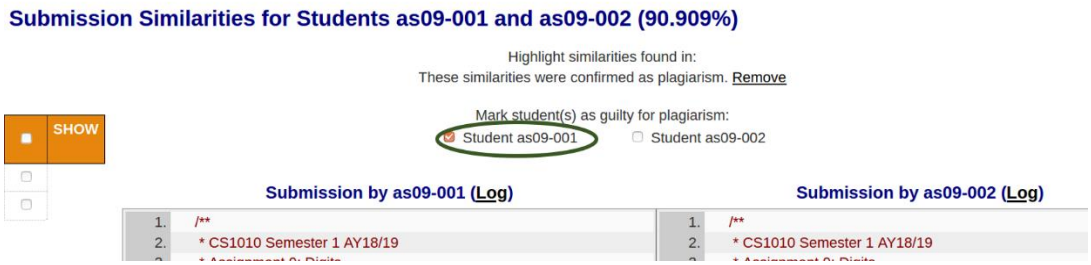
!	You must have been assigned with the teaching staff role for the module to confirm plagiarism cases.

a) Under pairwise submissions comparison listing, click Confirm as plagiarism on the top.

b) In the confirmation popup box, click OK to report this pair of students as suspicious.

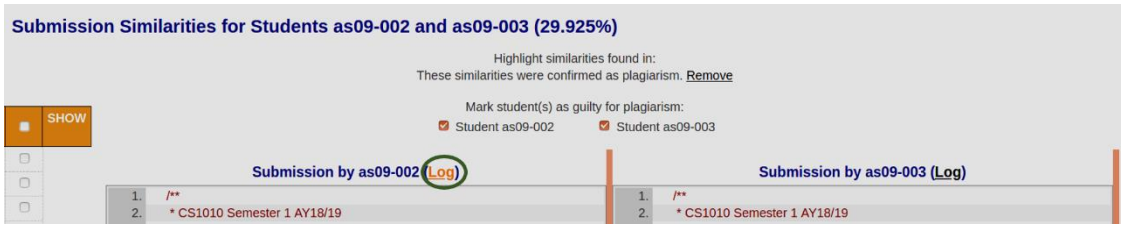


c) Mark the box belonging to the student ID whom is found guilty.



2.8. Viewing student log

a) Under pairwise submissions comparison listing, click Log beside the student ID of the student.



b) You would be able to see the past history of the student recorded in SSID.

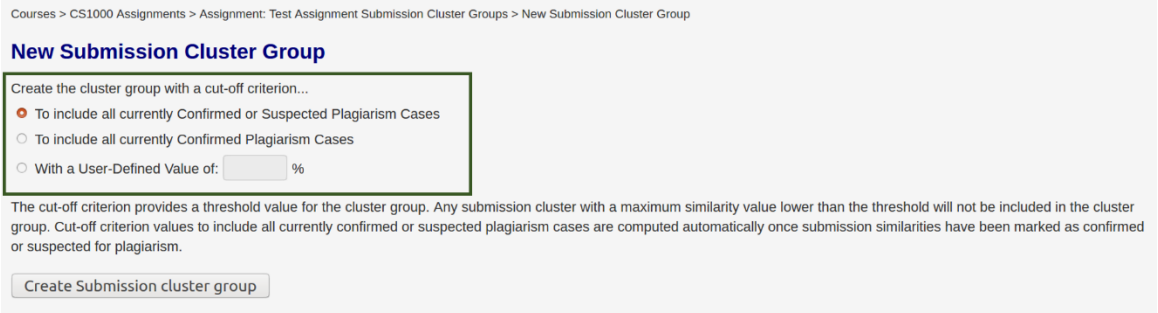
2.9. Creating and viewing similarity cluster groups

!	You must have at least one assignment to create grouping. To create an assignment, refer to section 2.3

a) Under the assignments listing, click Similarity groups.

b) Click New Cluster Group to create a new similarity cluster group.

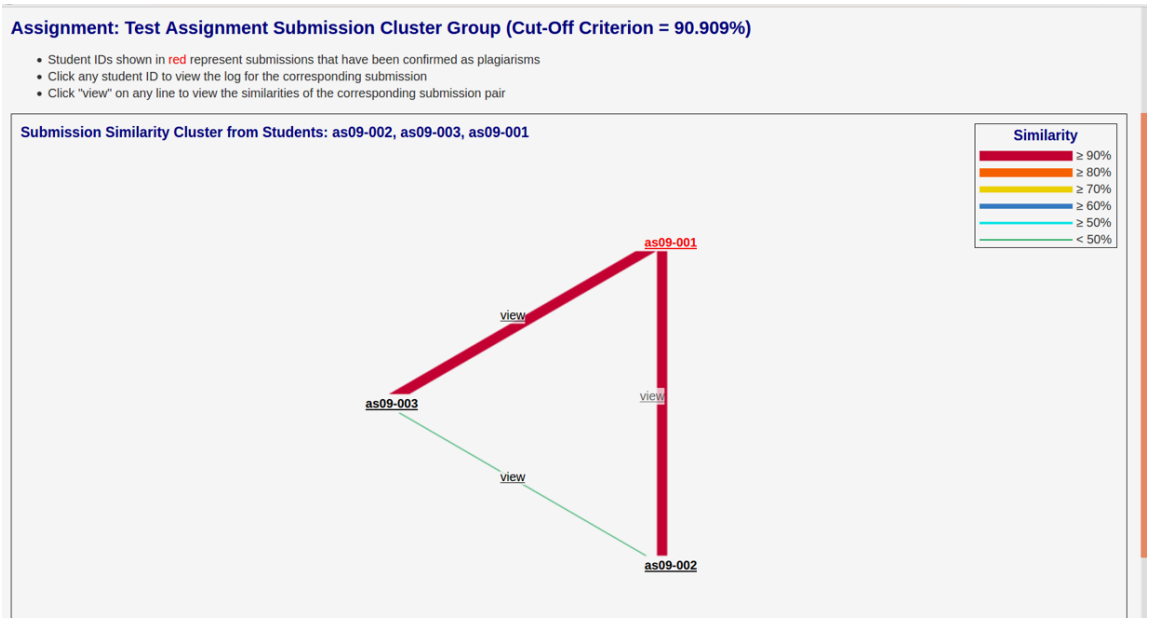
c) Create a cluster group based on your requirements. Then, click Create Submission Cluster group to continue



!	Note that the options "Confirmed plagiarism cases" and "Suspicious or confirmed plagiarism cases" are only available if there exists at least one confirmed plagiarism case and/or reported suspicious case respectively

d) Click View Clusters to view the similarity cluster group created.

e) View the cluster group to make observations.



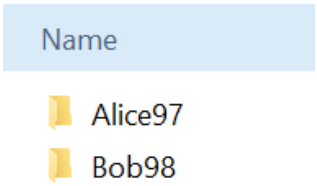
2.10. Mapping between directory name & student roster

This feature allows users to upload a mapping file that maps between a directory name (in the uploaded zip file) and the student roster that you might be using for your modules. To better illustrate this feature, we will use an example.

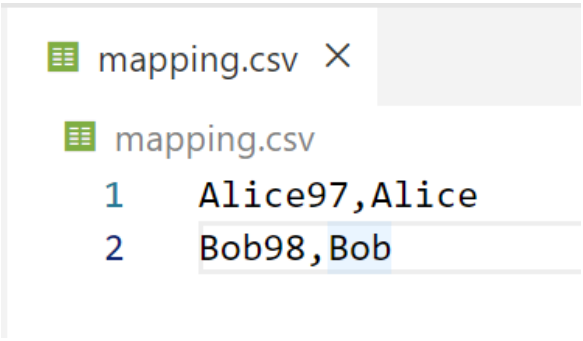
Assume your class has two students whose name and GitHub ID are given below:

	Name	**GitHub ID**
Student 1	Alice	Alice97
Student 2	Bob	Bob98

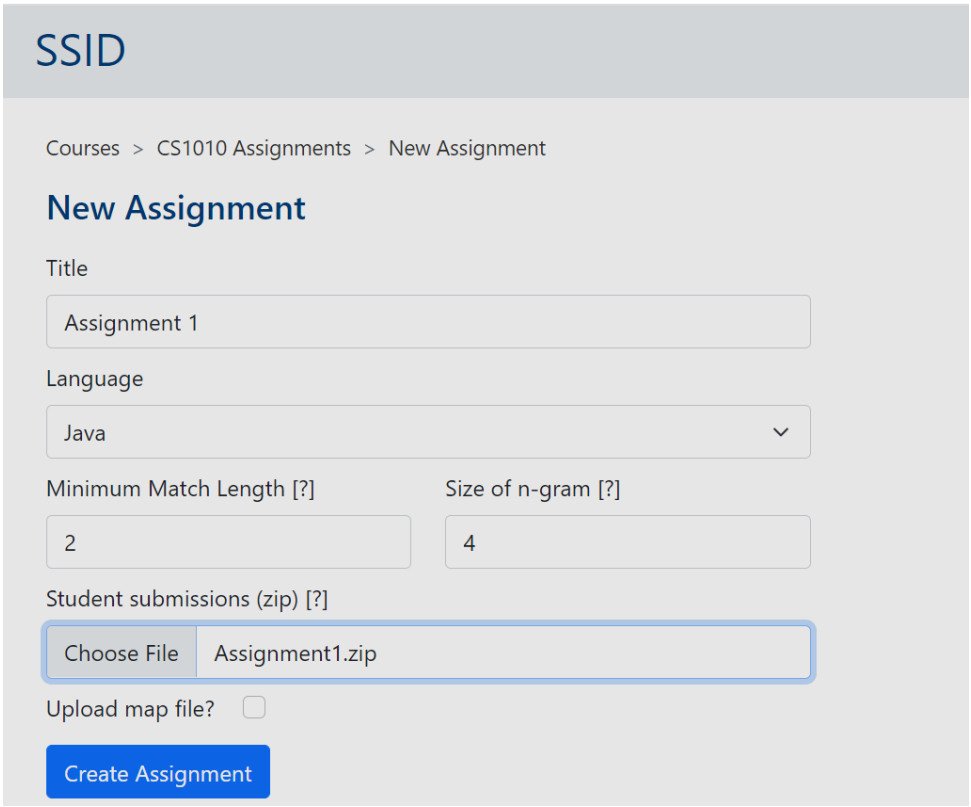
Now, you might encounter a situation where the files that they submitted are named after their GitHub ID (as shown below) while you want the file names to contain their actual name when the plagiarism results are displayed in the SSID page. How can you accomplish this?



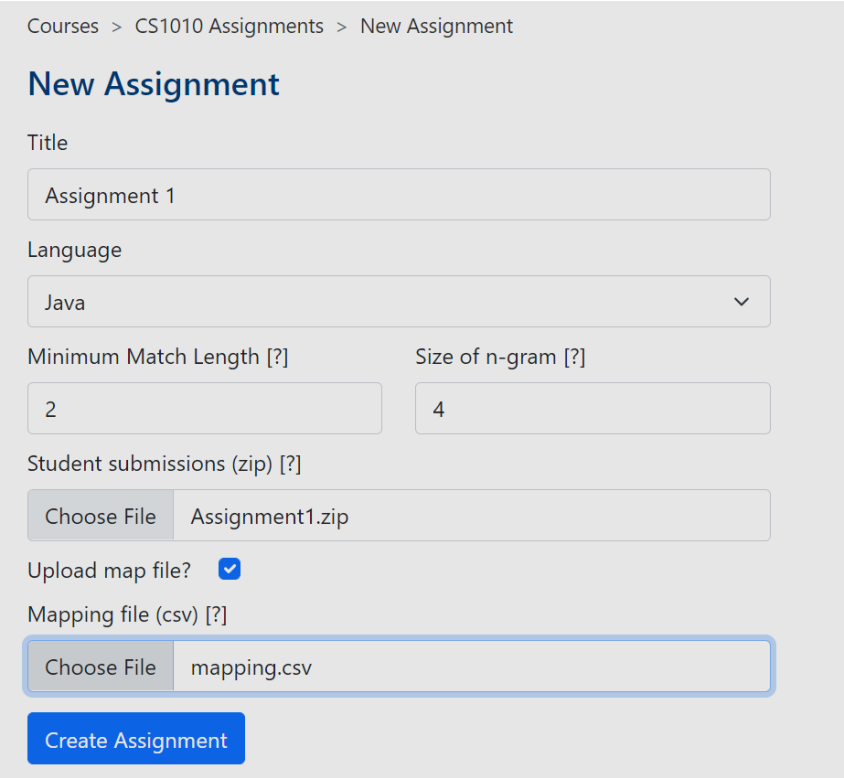
a) First create a csv file that contains the mapping between the GitHub ID and the actual names, as seen below. We will refer this as the mapping file from here on.



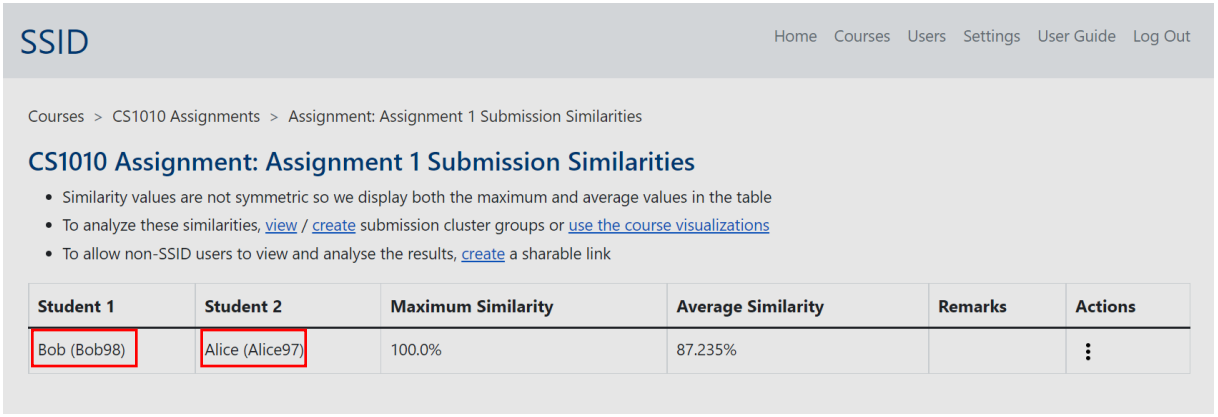
b) Now, create a new assignment in SSID. Fill in the respective fields and upload the submission zip folder as shown below. You may want to refer to section 2.3 if you are unsure.



c) Now, tick the box under Upload map file. Then, upload the mapping file that you created in step a. Then, click Create Assignment to start the plagiarism detection.



d) Once the plagiarism detection has been completed, you can look at the results and you will now see that the students' name and their respective GitHub ID are displayed.



3. Frequently Asked Questions & Answers

- What is the recommended percentage value to define a pair of plagiarized submissions?

  There exists no static lower-bound percentage of matching. However, based on our experiment, the lowest similarity between a pair of plagiarized submission is 57%, with the minimum-match-length be 2 and size of *N*-gram be 4.