



Student Submission Integrity Diagnosis [SSID]: User Guide

Table of Contents

1. Introduction.....	Page 2
1.1. About this User Guide.....	Page 2
2. Features.....	Page 2
2.1. Accessing SSID.....	Page 3
2.2. Viewing assignments.....	Page 4
2.3. Creating assignments.....	Page 4
2.4. Viewing pairwise submissions comparison results.....	Page 6
2.5. Reporting a suspicious case.....	Page 7
2.6. Removing a suspicious case.....	Page 7
2.7. Confirming a plagiarism case.....	Page 8
2.8. View student log.....	Page 9
2.9. Creating and viewing similarity cluster groups.....	Page 10
2.10. Mapping between directory name & student roster.....	Page 12
3. Frequently Asked Questions & Answers.....	Page 16



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

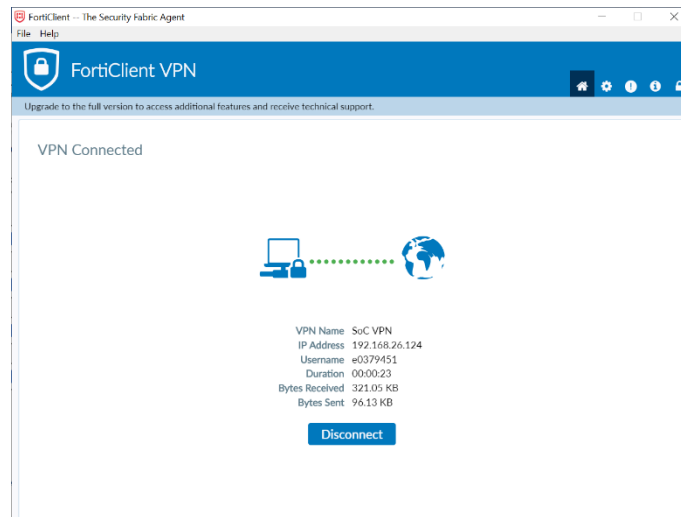
Some of the main features of SSID include:

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

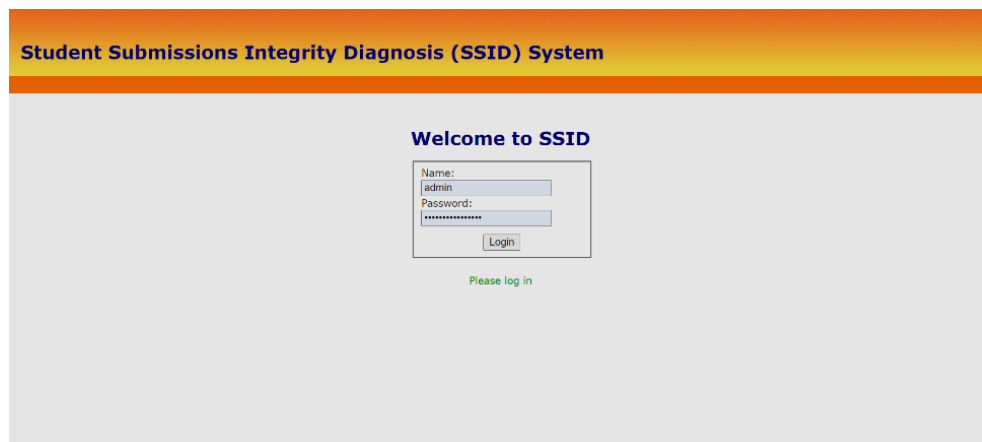
Below, you will get more information on how to use SSID to aid you in detecting and preventing plagiarism.

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.

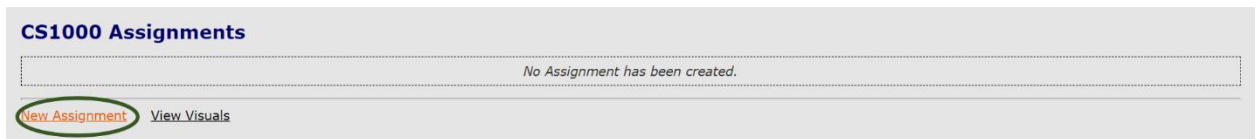


- b) Under the View tab, click [Assignments](#) to view the assignments.

Courses						
CODE	NAME	ACADEMIC YEAR (SEMESTER)	YOUR ROLE	ACTIONS		VIEW
CS1000	Test Platform	2019/2020 Semester 1	Administrator	Edit	Delete	Assignments Users Visuals

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.

Courses > CS1000 Assignments > New Assignment

New Assignment

Title

Language


Minimum Match Length [?]


Size of n-gram [?]

Student submissions (zip) [?] No file chosen



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"**.

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

	<p>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.</p> <p>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.</p> <p>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.</p>
-----------------------------------------------------------------------------------	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

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.

CS1000 Assignments

TITLE	LANGUAGE	SUBMISSIONS	SUSPECTED PLAGIARISM CASES	CONFIRMED PLAGIARISM CASES	VIEW	
Test Assignment	C	3	0	0	Submission similarities	Similarity groups Upload log Delete

[New Assignment](#) [View Visuals](#)

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

CS1000 Assignment: Test Assignment Submission Similarities

- Similarity values are not symmetric so we display both the maximum and average values in the table
- To analyze these similarities, [view](#) / [create](#) submission cluster groups or [use the course visualizations](#)

STUDENT 1	STUDENT 2	MAXIMUM SIMILARITY	AVERAGE SIMILARITY	REMARKS	
as09-001	as09-002	90.909%	46.2065%		View Submissions
as09-001	as09-003	90.909%	45.8745%		View Submissions
as09-002	as09-003	29.925%	23.317%		View Submissions

- c) Click on the lines to view the similarities.

Submission Similarities for Students as09-001 and as09-002 (90.909%)

Highlight similarities found in:

[Mark these similarities as suspicious](#) or [Confirm these similarities as plagiarism](#)

SUBMISSION BY AS09-001	SUBMISSION BY AS09-002	NUM OF MATCHING STATEMENTS		HIDE
Lines 12 — 16	Lines 11 — 15	5	<input type="checkbox"/>	
Lines 18 — 19	Lines 143 — 144	2	<input type="checkbox"/>	

4. *	11. #define NCOLS 28
5. * TODO(by student): Brief description of the purpose	12. #define NROWS 28
6. * of this program.	13. #define WHITE ' '
7. *	14. #define BLACK '#'
8. * @file: digits.c	15. #define K 5
9. */	16. #define MAX 784
10. #include "cs1010.h"	17. /**
11.	18. * This is the structure each sample will be represented as.
12. #define NCOLS 28	19. * It contains a number which represents the actual number and
13. #define NROWS 28	20. * the 2d array corresponding to the drawing.
	21. */
	22. typedef struct training {
	23. long number;
	24. char **array;
	25. } training;
	26. **

2.5. Reporting a suspicious case

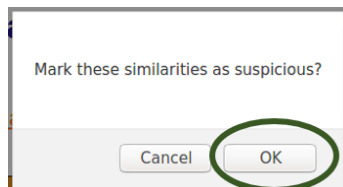
- a) Under the pairwise submissions comparison listing (as seen above), click [Mark these similarities as suspicious](#) on the top.

Courses > CS1000 Assignments > Assignment: Test Assignment Submission Similarities > Submissions from Student as09-001 and as09-002

Submission Similarities for Students as09-001 and as09-002 (90.909%)

Highlight similarities found in:
[Mark these similarities as suspicious](#) or [Confirm these similarities as plagiarism](#)

- b) In the confirmation popup box, click [OK](#) to report this pair of students as suspicious.



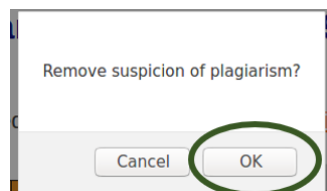
2.6. Removing a suspicious Case

- a) Under pairwise submissions comparison listing, click [Remove suspicion](#) on the top.


Submission Similarities for Students as09-001 and as09-002 (90.909%)

Highlight similarities found in:
These similarities were marked as suspicious. [remove suspicion](#) or [Confirm as plagiarism](#)

- 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.

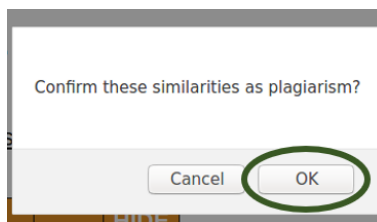
- a) Under pairwise submissions comparison listing, click [Confirm as plagiarism](#) on the top.

Courses > CS1000 Assignments > Assignment: Test Assignment Submission Similarities > Submissions from Student as09-001 and as09-002

Submission Similarities for Students as09-001 and as09-002 (90.909%)

Highlight similarities found in:
These similarities were marked as suspicious. [Remove suspicion](#) or [Confirm as plagiarism](#)

- 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.

Submission Similarities for Students as09-001 and as09-002 (90.909%)

Highlight similarities found in:
These similarities were confirmed as plagiarism. [Remove](#)

Mark student(s) as guilty for plagiarism:

☒ Student as09-001 ☐ Student as09-002



Submission by as09-001 (Log)

1.	/**
2.	* CS1010 Semester 1 AY18/19
3.	* Assignment 0: Rights

Submission by as09-002 (Log)

1.	/**
2.	* CS1010 Semester 1 AY18/19
3.	* Assignment 0: Rights

2.8. Viewing student log

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

Submission Similarities for Students as09-002 and as09-003 (29.925%)

Highlight similarities found in:
These similarities were confirmed as plagiarism. [Remove](#)

Mark student(s) as guilty for plagiarism:
☒ Student as09-002 ☒ Student as09-003

[SHOW](#)

[Submission by as09-002 \(Log\)](#) [Submission by as09-003 \(Log\)](#)

1.	/**	1.	/**
2.	* CS1010 Semester 1 AY18/19	2.	* CS1010 Semester 1 AY18/19

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

Courses > CS1000 Assignments > Assignment: Test Assignment 2 Submission Log for Student as09-002

Student: as09-002

DATE/TIME	COURSE	ASSIGNMENT	GRADER	REMARKS
2020-09-23 01:21:46	CS1000 Testing Platform	Test Assignment	SSID Administrator	Confirmed as plagiarism with submission by as09-001
2020-09-23 01:21:32	CS1000 Testing Platform	Test Assignment	SSID Administrator	Unmarked as plagiarism with submission by as09-001

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](#).

CS1000 Assignments

TITLE	LANGUAGE	SUBMISSIONS	SUSPECTED PLAGIARISM CASES	CONFIRMED PLAGIARISM CASES	VIEW	
Test Assignment	C	3	0	1	Submission similarities Similarity groups Upload log	Delete

- b) Click [New Cluster Group](#) to create a new similarity cluster group.

Assignment: Test Assignment Submission Cluster Groups

No cluster group has been created.
Cluster groups allow you to identify clusters of students with submission similarity values above a defined threshold.

[New Cluster Group](#)

- c) Create a cluster group based on your requirements. Then, click [Create Submission Cluster group](#) to continue

Courses > CS1000 Assignments > Assignment: Test Assignment Submission Cluster Groups > New Submission Cluster Group

New Submission Cluster Group

Create the cluster group with a cut-off criterion...

- ☒ To include all currently Confirmed or Suspected Plagiarism Cases
- ☐ To include all currently Confirmed Plagiarism Cases
- ☐ With a User-Defined Value of: %

The cut-off criterion provides a threshold value for the cluster group. Any submission cluster with a maximum similarity value lower than the threshold will not be included in the cluster group. Cut-off criterion values to include all currently confirmed or suspected plagiarism cases are computed automatically once submission similarities have been marked as confirmed or suspected for plagiarism.

[Create Submission cluster group](#)



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.

Courses > CS1000 Assignments > Assignment: Test Assignment Submission Cluster Groups

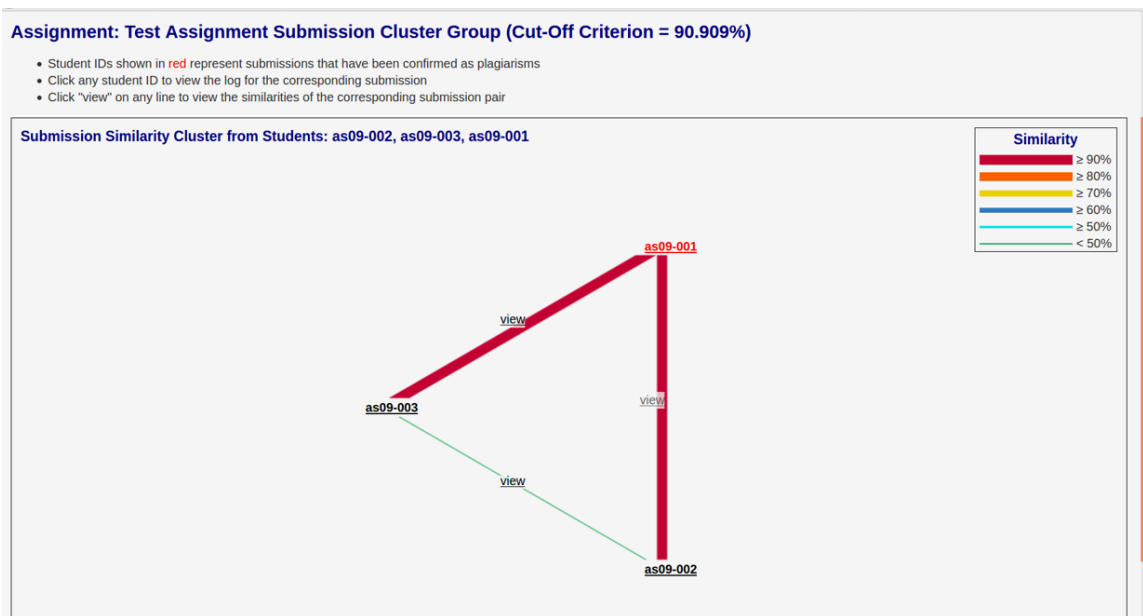
Assignment: Test Assignment Submission Cluster Groups

Submission cluster group was successfully created.

CUT-OFF CRITERION	NUMBER OF CLUSTERS	DESCRIPTION	ACTION
90.909	1	[Created on 2020-09-23 01:25:00 +0800] Based on cases confirmed or suspected as plagiarism	View Clusters Delete

[New Cluster Group](#)

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?

Name

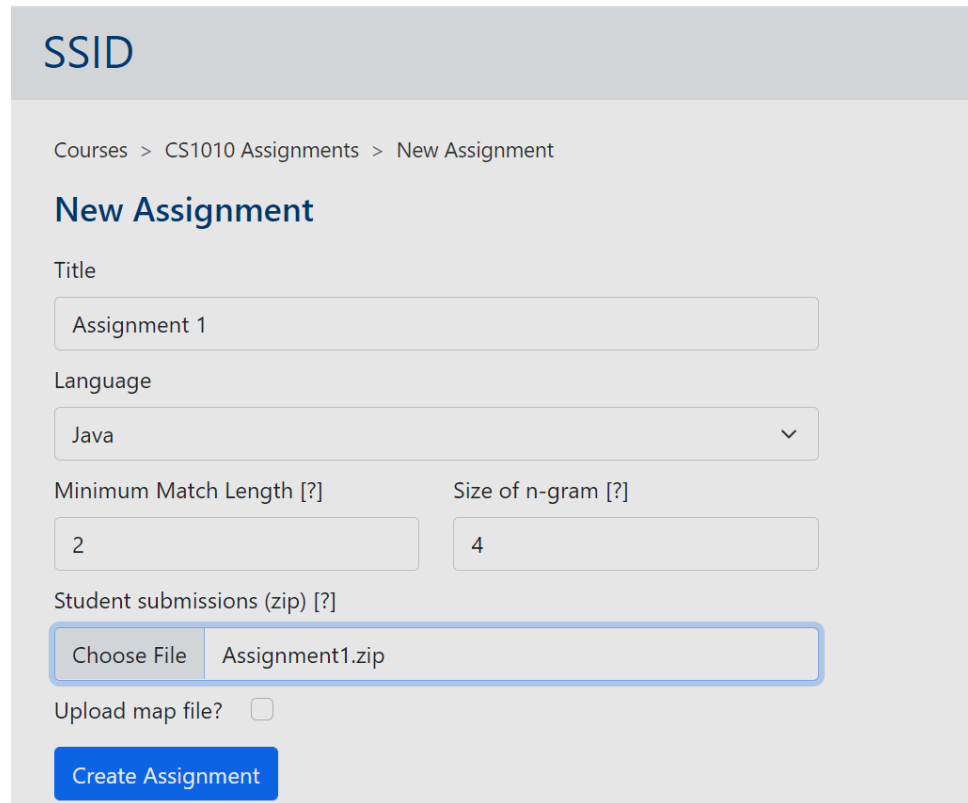
📁 Alice97

📁 Bob98

- 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.

```
mapping.csv ×
mapping.csv
1 Alice97,Alice
2 Bob98,Bob
```

- 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.



The screenshot shows the 'New Assignment' form in the SSID interface. The breadcrumb trail is 'Courses > CS1010 Assignments > New Assignment'. The form fields are as follows:

- Title:** A text input field containing 'Assignment 1'.
- Language:** A dropdown menu with 'Java' selected.
- Minimum Match Length [?]:** A text input field containing '2'.
- Size of n-gram [?]:** A text input field containing '4'.
- Student submissions (zip) [?]:** A file upload area with a 'Choose File' button and the filename 'Assignment1.zip' displayed.
- Upload map file?** A checkbox that is currently unchecked.
- Create Assignment:** A blue button at the bottom of the form.

- 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.

Courses > CS1010 Assignments > New Assignment

New Assignment

Title

Assignment 1

Language

Java

Minimum Match Length [?]

2

Size of n-gram [?]

4

Student submissions (zip) [?]

Choose File Assignment1.zip

Upload map file? ☒

Mapping file (csv) [?]

Choose File mapping.csv

Create Assignment

- 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.

SSID

HomeCoursesUsersSettingsUser GuideLog Out

Courses > CS1010 Assignments > Assignment: Assignment 1 Submission Similarities

CS1010 Assignment: Assignment 1 Submission Similarities

- Similarity values are not symmetric so we display both the maximum and average values in the table
- To analyze these similarities, [view](#) / [create](#) submission cluster groups or [use the course visualizations](#)
- To allow non-SSID users to view and analyse the results, [create](#) a sharable link

Student 1	Student 2	Maximum Similarity	Average Similarity	Remarks	Actions
Bob (Bob98)	Alice (Alice97)	100.0%	87.235%		⋮

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.