NightCore Mech SimiLabs User Manual

First Edition

Group:

Compiled by: The NightCore Mech Teams

Table of Contents

1. Introduction

Thank you for choosing SimiLabs, a system designed to identify academic misconduct committed by students at North-West University.

1.1. Background of the application

The NWU Registrar must address plagiarism by evaluating each case individually and appointing experts to prepare technical reports. External subject matter experts (SMEs) are requested to examine the technical reports with an additional report that provides a deeper insight into the alleged plagiarism if the technical reports do not self-evidently emphasize the severity of the plagiarism. Manually comparing the allegedly plagiarized text in issue with the original text as evidence text is a requirement for the technical report, which can become difficult and lead to certain similarities being overlooked.

1.2. Purpose of the application

Through the use of a similarity metric, the software should reduce the time spent manually comparing two texts and generalise the assessment of how severe the conjectured copying is. The software must combine text-matching skills with stylometric analytics to provide more accurate reports, better explain academic misbehaviour, and enable improved decision-making.

1.3. System Capabilities

The system developed and shown in this user manual has the following capabilities:

- The system is capable of being used by multiple users at any given time.
- The user can upload documents to do a quick text analysis, extensive text analysis, and a stylometric analysis.
- The user can download reports of the results of the various analyses.

The following sections will explain how each of these functionalities works and will also assist with accessing the application on a web browser.

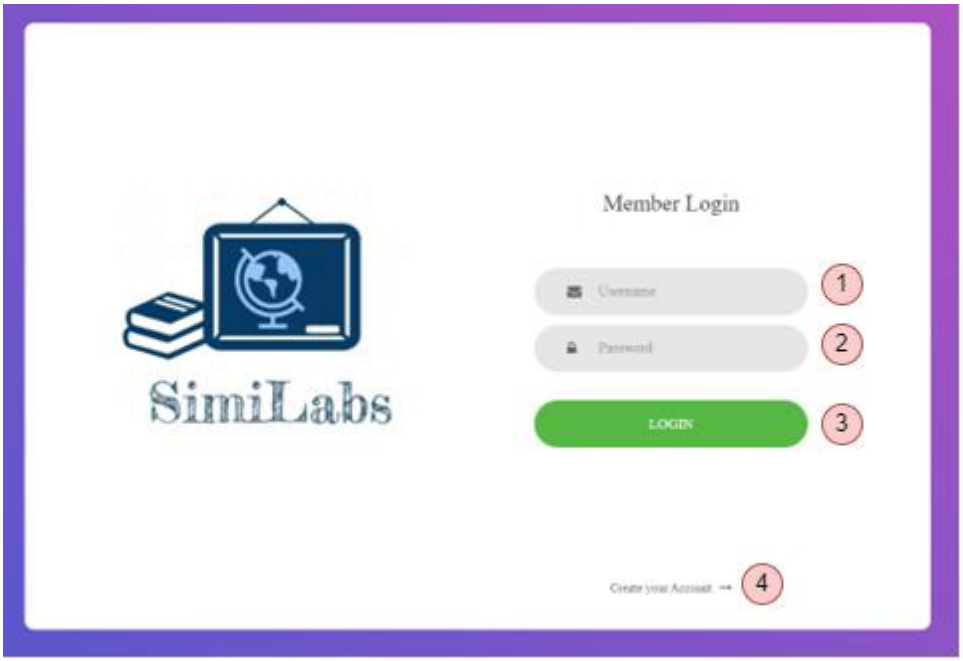
2. Using the system

2.1. Accessing the application via a web browser

The application can be accessed through any web browser on a desktop device, including a desktop PC and a laptop, by entering the following URL in the search bar:

2.2. Logging in

The first page the user will see is the login page:



1) The user inputs their username in this box.
2) The user inputs their password in this box.
3) The user clicks on this button to log in with their credentials.
4) The user clicks on this to be redirected to the registration page if they do not have an account.

The user must log in before the application can be used.

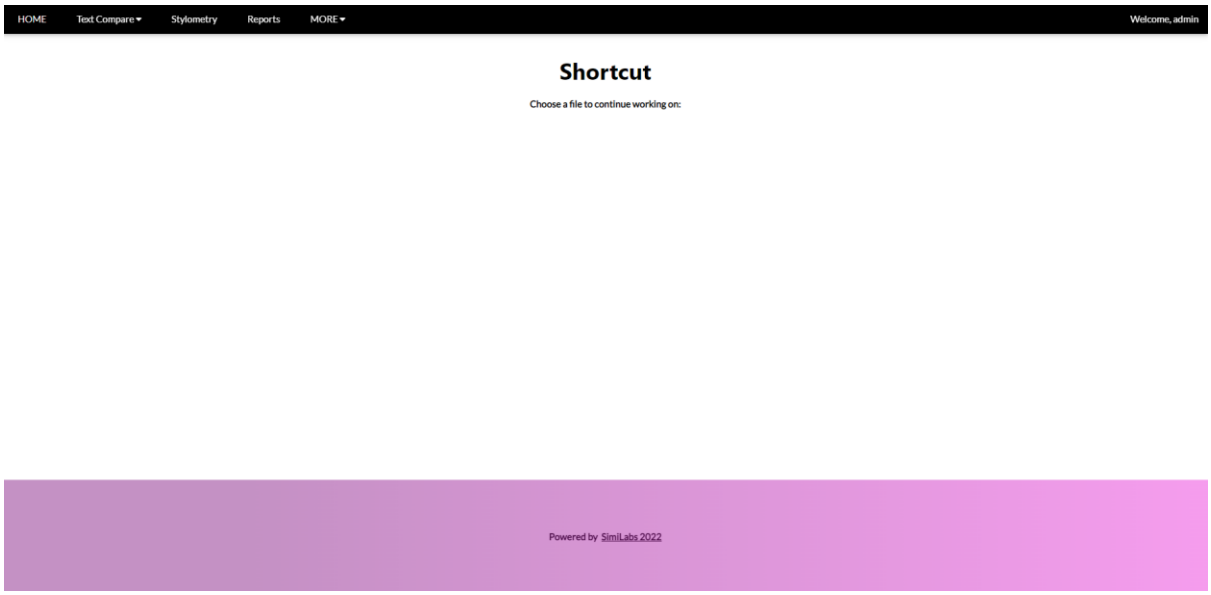
2.3. Creating an Account

If the user does not already have an account, they can create one by clicking on Create your Account at the bottom of the screen. This will redirect the user to the registration page:

1) The user chooses a username. The username must be a valid email address.
2) The user chooses a strong password.
3) The user enters the password again to make sure they did make any errors.
4) The user clicks on this button to register a new account with their credentials.
5) The user clicks on this to go back to the login page to log into the application.

6) Navigating the Home page

Upon successful log-in, the user will be redirected to the home page:



The toolbar at the top of the screen shows the different options available to choose from:

1) This redirects the user to the home page.
2) The user has an option of a quick text comparison or an extensive text comparison.
3) Redirects the user to the stylometry page.
4) Reports
5) The user can choose between a help option and logging out of the application.

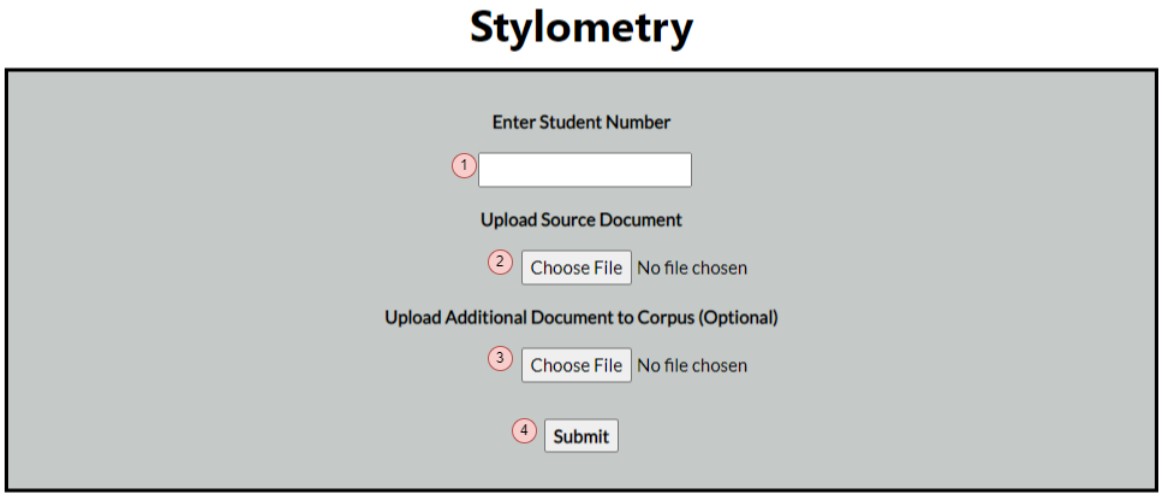
Using the Text Comparison Feature

Quick text comparison

Extensive text comparison

Using the Stylometry Feature

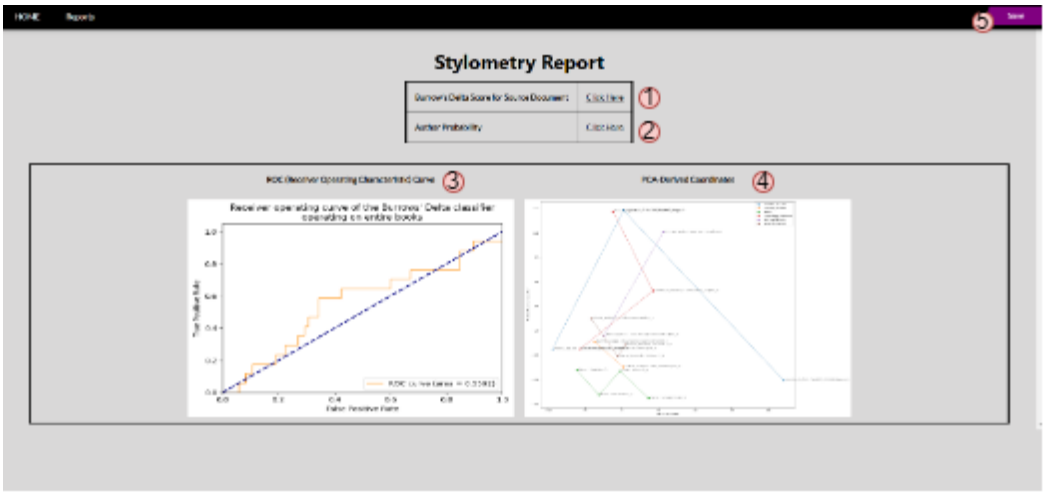
If the user chooses the stylometry option in the toolbar, they will be redirected to the stylometry page:



1) The user enters the student number of the student that submitted the document.
2) The user uploads the suspected document.
3) The user can choose to add additional documents to the corpus belonging to the student.
4) The user clicks on the submit button to receive the results of the stylometric analysis.

Stylometry results

After the analysis has been completed the user will be redirected to the Stylometry Report page:



1) Contains the Burrows' data value of the source document compared to all of the students in the corpus. The user can view these values by clicking on "Click Here". The Burrows' delta is a value between 0 and 3. The higher the value, the less likely it is that the student that submitted the document is the author of the document. The lower the value, the more likely that the student is the author.

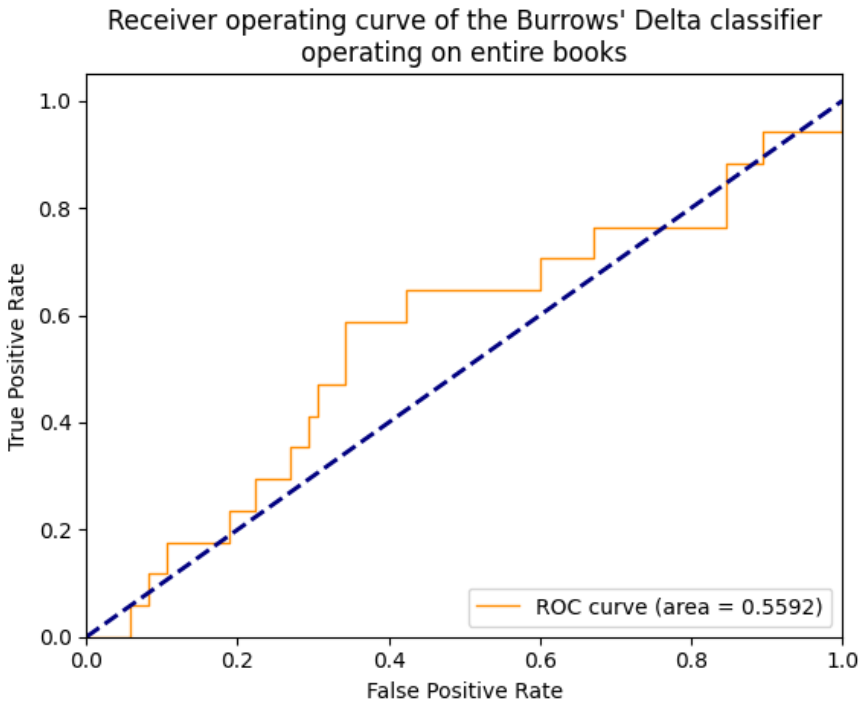
31775357 - ISE2022

author	
Annika_du_Toit	2.048234
Hanno_Visagie	1.576989
Hano	1.174867
Llewellyn_Anthony	2.093773
Michael_Rosin	1.771622
Shené_Boshoff	1.414503

2) Contains the probability that the student that submitted the document is the author of that document compared to all other students in the corpus. The higher the probability of another student, the less likely it is that the student that submitted the document is not the sole author. The higher the probability next to the student that submitted the document, the more likely it is that the student is the author. The probability next to that student should be the highest on the list. The user can view these values by clicking on "Click Here".

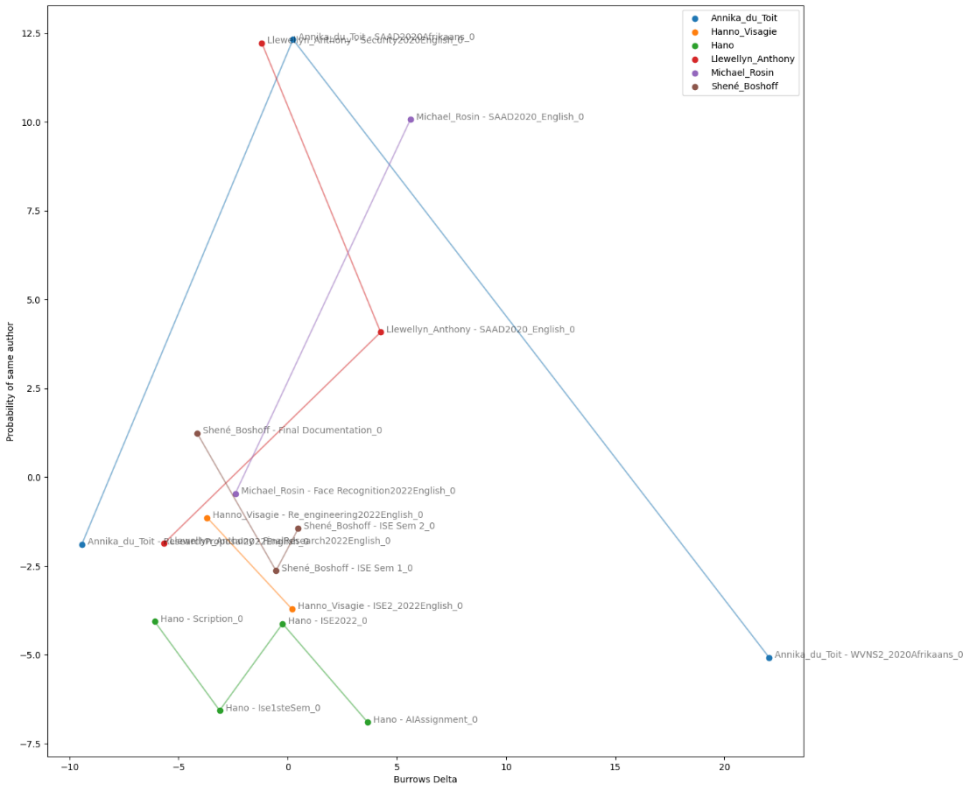
author	
Annika_du_Toit	0.482595
Hanno_Visagie	0.497711
Hano	0.510615
Llewellyn_Anthony	0.481135
Michael_Rosin	0.491466
Shené_Boshoff	0.502926

3) The Receiver Operating Characteristic (ROC) Curve is an indication of how well the analysis performed. The ROC evaluation is performed using cross-validation. Every document in the corpus is taken out and a Burrows' model is trained on the remainder of the document, which is then tested against the document taken out. The probability scores of the results are used to calculate the ROC curve. The Area under the ROC Curve (AUC) measures the area under the ROC curve. AUC provided an aggregate measure of performance across all possible classification thresholds. It represents the probability that a random positive is positioned to the right of a random negative. The orange line represents the AUC. The closer it is to 1 the better the model performs and the more accurate it is. The model will improve as the corpus grows. The ideal AUC will have a y value of 1 and an x value of 1.



4) The stylistic similarities between the documents in the corpus are visualized by calculating their differences and using Principle Component Analysis (PCA). Z-scores are calculated for the top 50 most common words used in every document. The z-

score represents the fingerprint of each document. Ideally, the same authors should be clustered together.



5) The user can generate and save a PDF report of the stylometric results. The user will be able to choose the download location of the report.

Possible errors that can be encountered

The list of all potential known faults that the user could run into when utilising the application is provided in the next part, along with some extra suggestions and guidance on how to resolve these issues.

Error Message	Reason for Error	Possible Fix
Valid username is required.	The username entered is incorrect or doesn't exist.	If the user has an account, enter the username correctly or create an account.
Password is required.	The password was not entered.	Enter the correct password.
Invalid credentials.	The username or password was entered incorrectly.	Enter the correct username and password.
Confirmation password is required.	The confirmation password was not entered.	Enter the confirmation password.

Passwords do not match.	The password and confirmation password is not the same.	Make sure both passwords are correct and the same.
Please select a file.	A file was not selected for upload.	Select a file to upload.
Something went wrong. 400 Bad Request: missing length.	The length of the substring to search for was not entered when choosing the substring algorithm.	Enter the length of the substring when the substring algorithm is chosen.
Please fill out this field.	The student number was not entered.	Enter the student number.
Something went wrong.		

If the user encounters any errors that are not listed above, the support team can be contacted.