

USING MOSS FOR PLAGIARISM DETECTION IN LARGE-SCALE DATA STRUCTURES

This document presents a step-by-step guide on setting up and using MOSS (Measure of Software Similarity) to detect plagiarism in large-scale data structure assignments.

1. Folder Setup:

- Start by creating a new folder on your local system and name it "submissions". This will be your central location for housing all student submissions and relevant MOSS files.

2. Registration for a MOSS Account:

- Compose a new email and address it to moss@moss.stanford.edu
- In the body of the email, type the following:

```
'''
```

```
registeruser
```

```
mail username@domain
```

```
'''
```

Note: Please replace "username@domain" with your actual email address.

3. Obtaining and Saving the MOSS Script:

- Once you've sent the email, wait for a reply from MOSS. This reply will contain the necessary Perl script.
- Save this script as "moss.pl". Make sure to grant it execute permissions with the command `chmod ug+x moss.pl`
- Move the "moss.pl" file into the "submissions" folder you created earlier

4. Downloading and Organizing Student Submissions:

- Download all student submissions in zip file format from Canvas.
- Move the downloaded zip files into the "submissions" folder.

5. Running the Python Script:

- The Python script "extract_and_process_files.py" should be placed in the "submissions" folder.
- From within the "submissions" folder, run this script via your terminal or command prompt.
- The script will then ask you to choose the operation you want to carry out: extract the submissions and individual files, run the MOSS plagiarism check, or perform both tasks..

6. Understanding Script Operations:

- The Python script unzips each file into its respective directory in and flattens the directory structure.
- All zipped code files are extracted and placed into their respective directories within the "extracted_code_folders".
- The script takes care of any filename conflicts and removes empty directories.
- If selected by the user, the script will automatically initiate the MOSS plagiarism check.

7. Reviewing MOSS Results:

- After the MOSS analysis is complete, the results will be saved in a file named "output.txt" within the "submissions" folder.
- Review the contents of the "output.txt" file to identify potential plagiarism. Any instances where the plagiarism score is above 700% should be closely examined.
- If the script encounters any errors during its execution, these will be recorded in a file named "error.txt" in the "submissions" folder. This file provides detailed information about the errors for troubleshooting purposes.

If you have any questions or require further assistance, please feel free to reach out.

Best wishes,
Goutam Kalburgi
gk325@nau.edu