Plagiarism Checker for Assignments in Computer Science Courses

Project Abstract
In some curricula, students are required to complete a project or answer some open-ended questions. However, some students may cheat by copying former students' assignments. The goal of the project is to develop a tool to automatically check duplicate assignments. It can help teachers to check whether there is plagiarism in students' Assignments.

Project Description
In some curricula, students are required to complete a project or answer some open-ended questions. However, some students may cheat by copying former students' assignments. To deal with this problem, this project will build a tool to automatically check duplicate assignments. It will compare the received assignment with historical files in the database and find the possible plagiarized part.
The core features of this tool are the following:
(1) It can extract the text from files in different formats i.e. word, pdf, jpg, etc. You can use some techniques like OCR(Optical Character Recognition) to achieve this goal.
(2) It contains two kinds of users, teachers and students. Users can access it online.
(3) It contains two duplication detection algorithms for text and source code respectively. For text duplication detection, it can set different detection granularity like sentence or paragraph. For source code duplication detection, it can tackle some simple transformation like variable name substitution. Teachers can set a threshold to control the allowed complexity of transformations.
(4) Teachers can save the historical data in an online database and create different databases for different assignments.
(5) It can present the detected duplication parts and the corresponding historical file.
The project must include a user-friendly interface. Only registered users are allowed to use this tool.

Project Scope
Your project should implement at least the features described above. Optional features like better results visualization could be added to the application at the team's discretion. Teams must also provide documentation that describes how to get started with the tool.

Process Requirements
There are no specific requirements on the process to be followed.

Environmental Constraints
No constraints.

Project Restrictions
None.

Project License
MIT License.

Level of Sponsor Involvement
Sponsor contact:
Teams are free to contact the sponsor at any time, but answers may be delayed
depending on sponsor availability. Brief Skype meetings can be arranged on demand if
necessary.