SyntaxSentinels

Leveraging advanced technology to improve code plagiarism detection

By Dennis Fong, Lucas Chen, Luigi Quattrociocchi, Mohammad Mohsin Khan, and Julian Cecchini
With help from Dr. Hassan Ashtiani

Motivation

Plagiarism detection for coding in academia and online competitions is becoming increasingly challenging as students develop sophisticated ways to evade traditional syntax-based analysis tools like MOSS (Measure of Software Similarity) which analyze the structure of the code. These systems struggle with advanced "semantic" plagiarism techniques, in which the code is modified structurally while retaining the same logic. As a result, many instances of code plagiarism go undetected leading to false negatives and compromising academic integrity.
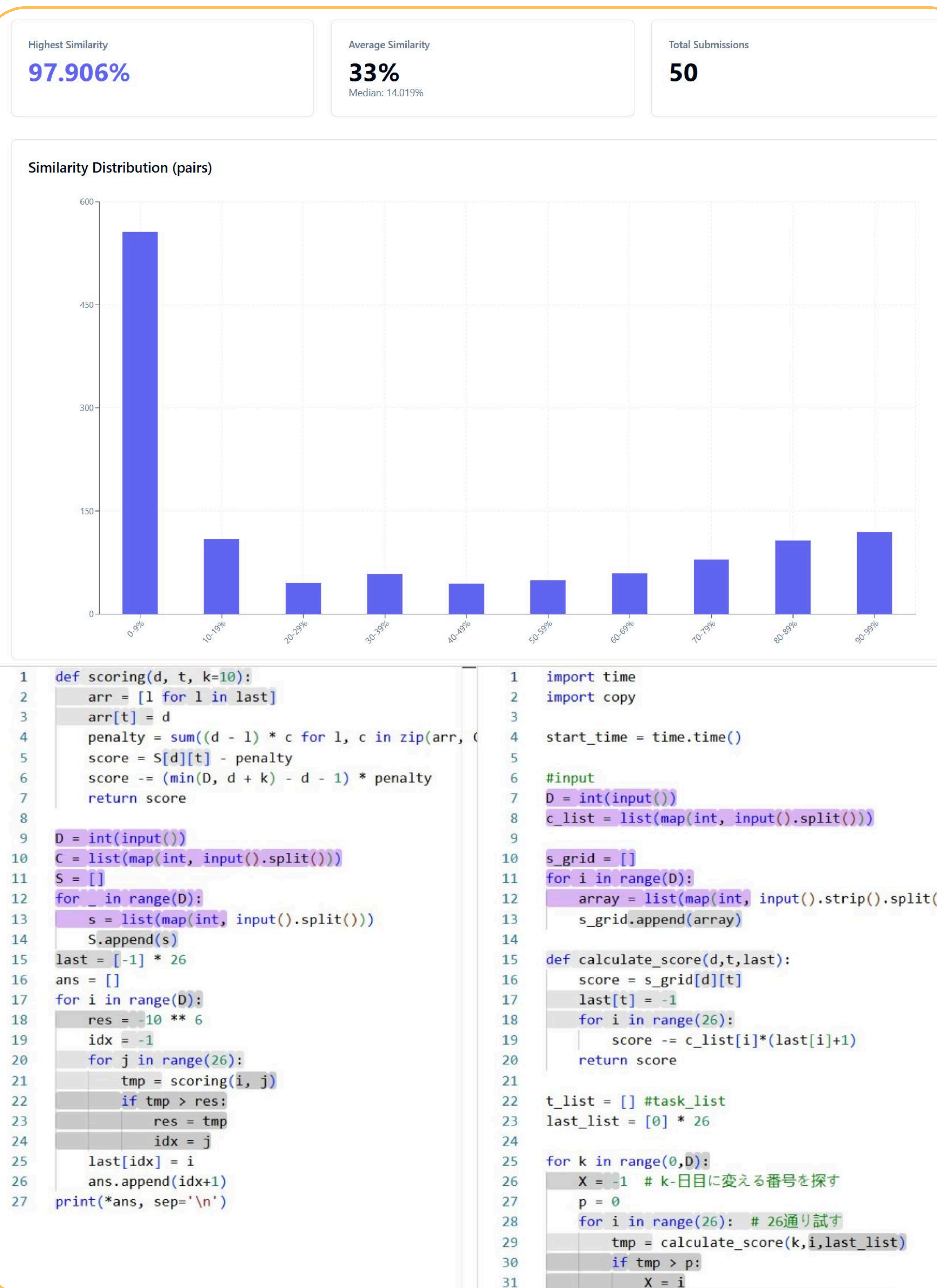
Features

- Comprehensive Detection via an NLP approach.
- Detailed reports to show similarity scores and line by line analysis of file pairs.
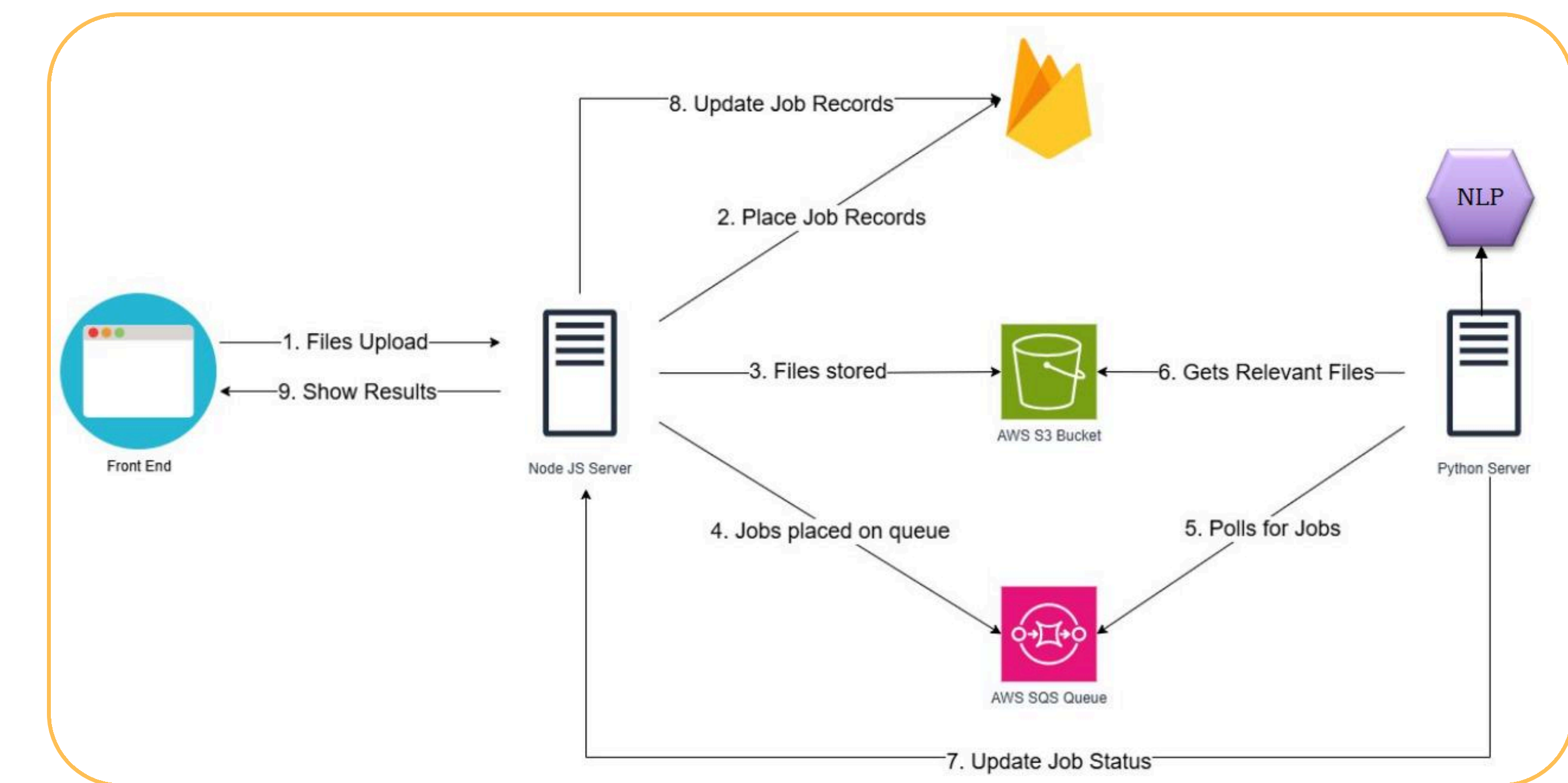
Key Benefits

- Streamlines the user experience of educators in maintaining academic integrity.
- Improves precision and reduces false positives.
- Visualizations provide a high-level overview of plagiarism within a corpus.
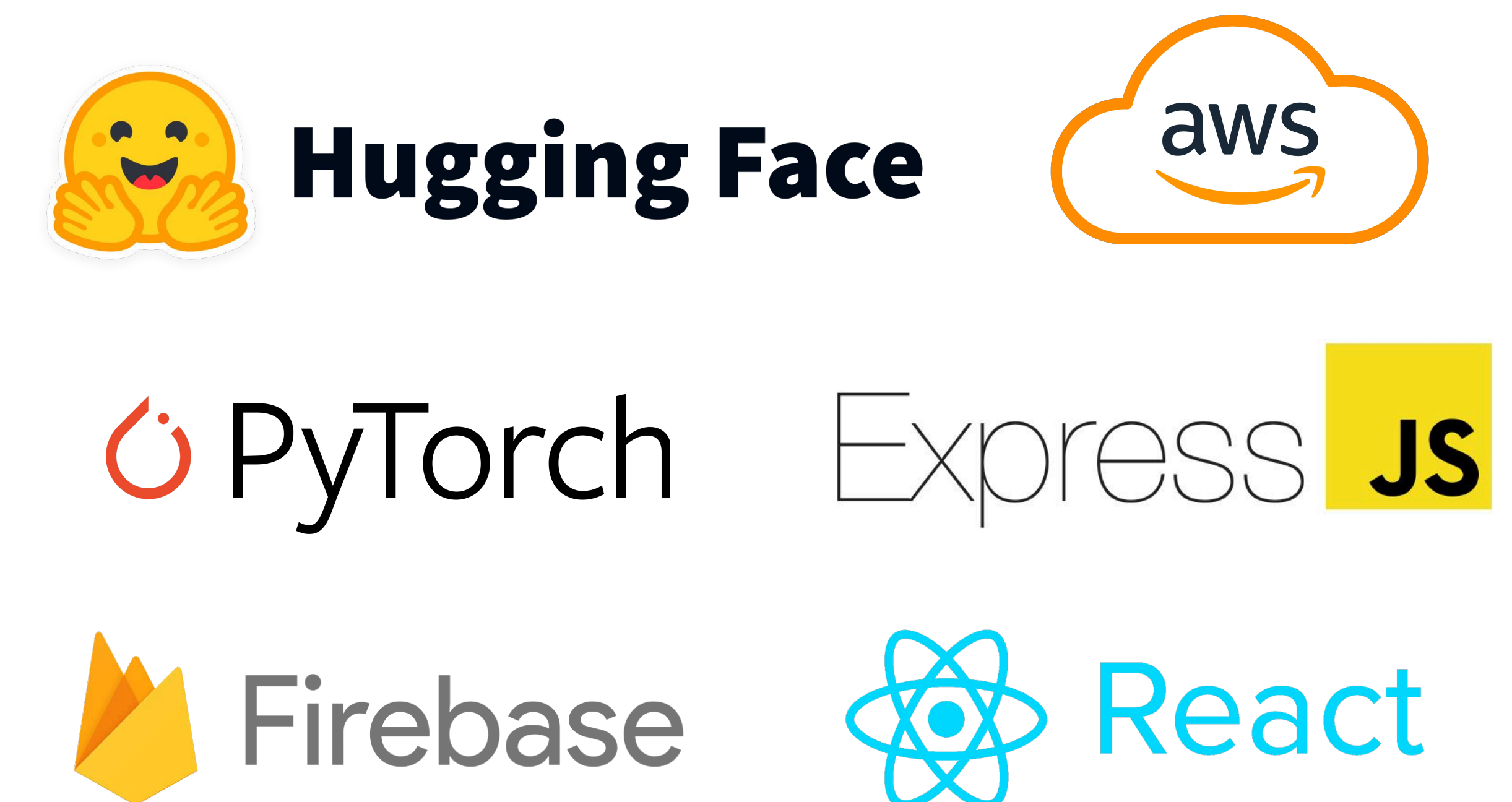
Overview

A code plagiarism detector that uses Natural Language Processing (NLP) among other traditional techniques to gain deeper insights on potential plagiarism within a batch of code submissions.

System Architecture

Tools Used

Hugging Face · aws · PyTorch · ExpressJS · Firebase · React

Meet the Team

Dennis Fong · Lucas Chen · Luigi Quattrociocchi · Mohammad Mohsin Khan · Julian Cecchini