Making personalized education scalable using sequence alignment algorithm
Lakhan Bhojwani (lb9316@rit.edu)
Advisor: Dr. Carlos Rivero
Rochester Institute of Technology

Introduction
• Nowadays the number of students enrolling in computing science are increasing with a tremendous speed and thus providing personalized feedback to all students for all assignments is impossible.
• The idea of this project is to develop an application which will take all the submitted assignments from students and then identify the response patterns and make clusters of similar responses into one group.
• The proposed algorithm is the modified version of Smith and Waterman Algorithm [1].

Smith – Waterman Algorithm
• Let us consider two sequence S1 = “HELLOWORK” and S2 = “HELLOWORLD”.
• Smith and Waterman algorithm is divided into five major steps: Creating a table, Table Initializing, Calculating the Score for each cell in the matrix, Trackbacking the matched sequence and Identify matched subsequence [1].

Proposed Method

Results
• The data was taken from the codechef [2] website and the total number of Java test code samples are 625. Assignment name was Life, Universe and Everything and the total number of users submitted solution for this assignment were 197.

Conclusions and Future Work
• Proposed algorithm gives more accurate result as compared to techniques with Longest Common Sequence Algorithm and metrics based Algorithms. For example Smith and Waterman algorithm.
• Proposed algorithm works faster than Abstract Syntax Tree based techniques and the Program Dependencies Graph based techniques because it doesn’t have to create any graph or tree before real implementation.
• Future work would be to create a visual graphic of matched lines instead of just showing line numbers that are been matched. This can be done by making an Eclipse plugin so that it will be easy for professors to check plagiarism.

References