INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI
SHORT ABSTRACT OF THESIS

Name of the Student	:	SHIBSANKAR DAS
Roll Number	:	10612301
Programme of Study	:	Ph.D.

Thesis Title: On Approximate Parameterized String Matching and Related Problems

Name of Thesis Supervisor(s)	:	Dr. Kalpesh Kapoor
Thesis Submitted to the Department/ ~~Center~~	:	Mathematics
Date of completion of Thesis Viva-Voce Exam	:	14-August-2016
Key words for description of Thesis Work	:	Graph Theory, Approximate and Parameterized String Matching

SHORT ABSTRACT

This thesis investigates problems concerned with Approximate Parameterized String Matching (APSM) under Hamming distance error model and related problems in graph theory. We introduce a term called error class in APSM problem and explore various combinatorial properties of the error classes. We also provide a tight lower bound on the weights of Maximum Weight Bipartite Matching (MWBM) of graphs which is correlated with the counting of number of error classes in APSM problem. The problem of APSM for a pair of equal length strings under Hamming distance is computationally equivalent to MWBM problem in graph theory. Let $G = (V, E, Wt)$ be an undirected, weighted bipartite graph where V be the vertex set and E be the edge set of G with positive integer weights on the edges which are given by the weight function Wt: $E \rightarrow N$. We fine-tune the existing decomposition theorem, originally proposed by Kao et al., for computing a MWBM of G. It is used to design a modified deterministic algorithm to compute weight of a MWBM of G. This modified decomposition technique is used as a subroutine in the following three proposed solutions. (*i*) An $O(n\_P\, n\_T\, m)$-time algorithm for all pairs approximate parameterized string matching problem with error threshold k, among two sets P and T of m-length strings over two distinct alphabets, where n_P and n_T are the cardinality of P and T, respectively. We introduce Parikh vector based filtering technique in order to preprocess the given sets of strings to avoid the comparison of non-candidate pairs. (*ii*) An $O(nm)$-time algorithm for APSM problem under weighted Hamming distance for a pattern p in a text t. And (*iii*) an $O(m+k)$-time solution for parameterized string matching problem with k mismatches between a pair of m-length strings over two distinct alphabets under weighted Hamming distance. Experimental evaluation of some of the proposed techniques is also done and it is observed that the suggested techniques are efficient.