

## Spring 2016 Seminar Series

APPROXIMATE SEQUENCE MATCHING: ALGORITHMS AND APPLICATIONS

FRIDAY FEBRUARY 19, 2016

10:00 AM – HEC 438

Approximate sequence matching algorithms are at the heart of many applications in computational biology. For example, given a large collection of short sequences, we may be interested in finding all pairs of sequences that share a common sub-string under some length and mismatch constraints. Clustering reads, which is one of the key step in meta-genome assembly as well as in various reads compression algorithms can be modelled as the above problem. The bounded mismatch average common sub-string is another approximate sequence matching problem with applications in phylogeny inference. While seed-and-extend type heuristics are widely used in practice to solve various such problems, provably efficient solutions are still evasive for many. In this talk, I present the recent efforts by our group on designing algorithms with worst/average case run time guarantees for some of these problems and efficient strategies to parallelize them. It is worth noting that the earlier solutions to those problems takes near quadratic time, where as the new algorithms are strictly sub-quadratic.

Part of this work appeared in RECOMB 2015 Conference.

**DR. SHARMA V. THANKACHAN**  
Georgia Institute of Technology



Sharma Thankachan is a Research Scientist in the School of CSE, Georgia Institute of Technology, Atlanta. His research interests are in the area of algorithms and data structures, primarily for string searching/indexing/compression problems. Sharma holds a PhD in Computer Science from Louisiana State University.

*Hosted by: Dr. Mark Heinrich*

