Abstract:

In this talk I will discuss two related data structure problems. The first is to preprocess two strings $S$ and $T$, each of length $n$, into a data structure that can report the optimal alignment score (Longest common subsequence / edit distance) between any substring of $S$ and any substring of $T$. I will describe a data structure with query time $(\log n)^{(2+o(1))}$, and construction time and space $n^{(2+o(1))}$. This is optimal up to polylogarithmic/subpolynomial factors since computing just the alignment between $S$ and $T$ cannot be done in subquadratic time assuming the strong exponential time hypothesis.

The second problem is to preprocess a directed weighted planar graph into a data structure that reports the distance between any two vertices. Using the concept of voronoi diagrams, dramatic progress was made on such distance oracles in the past few years. We now have oracles of almost linear $n^{(1+o(1))}$ size that answer queries in $(\log n)^{(2+o(1))}$ time. However, the construction time of these oracles is not nearly linear, but $O(n^{(3/2)})$, which is not known to be optimal.

I will describe how ideas developed for planar distance oracles lead to the data structure for the string alignment problem. The structure of the alignment graph allows for a simpler presentation of some of the techniques involved, and is further exploited to obtain a faster construction time. It is not uncommon that techniques originally developed for pattern matching problems are later extended and generalized to planar graphs. Here ideas propagate in the opposite direction; techniques originally developed for planar graphs are specialized to a pattern matching problem.

Based on joint works with Panos Charalampopoulos, Pawel Gawrychowski and Oren Weimann (SODA