Time and Space Optimal Algorithms for Text Processing and Data Compression

S. Cenk Sahinalp
University of Warwick, UK

Abstract

Among the main approaches in string algorithmics, the labeling paradigm of Karp, Miller and Rosenberg stands out as the simplest and the most generally applicable. Before we introduced the first optimal parallel algorithm for suffix tree construction (1), the labeling paradigm was considered not competitive with other approaches due to the lack of robust parsing schemes. In this talk, we show how to further improve the power of the labeling paradigm via a new Locally Consistent Parsing (LCP) technique which we apply to obtain the first time and space optimal algorithms for several fundamental problems in string processing (2). These include approximate string matching, dynamic text indexing, and dictionary matching. We then present time and space optimal algorithms for two very promising novel techniques for high performance data compression: the Hershkovits-Yokoo-Ziv (HYZ) context sensitive data compression method (3), and the Flexible Parsing (FP) scheme for optimal dictionary compression (4). Our algoritmic results complement the strong information theoretic properties satisfied by these compression methods.

Joint work with (1) U. Vishkin, (2) U. Vishkin, (3) Y. Matias, S. Muthukrishnan and J. Ziv, (4) Y. Matias.