Efficient Cross-Trees for External Memory

Roberto Grossi *
Dipartimento di Sistemi e Informatica
Università di Firenze
via Lombroso 6/17, 50134 Firenze, Italy

Giuseppe F. Italiano †
Department of Computer Science
Hong Kong University of Science & Technology
Clear Water Bay, Kowloon, Hong Kong

Abstract

We describe efficient methods for organizing and maintaining large multidimensional data sets in external memory. This is particular important as accesses to external memory are currently several order of magnitudes slower than accesses to main memory, and current technology advances are likely to make this gap even wider. We focus particularly on multidimensional data sets which must be kept simultaneously sorted under several total orderings: these orderings may be defined by the user, and may also be changed dynamically by the user throughout the lifetime of the data structures, according to the application at hand. Besides standard insertions and deletions of data, our proposed solution can perform efficiently *split* and *concatenate* operations on the whole data sets according to any ordering. This allows the user: (1) to dynamically rearrange any ordering of a segment of data, in a time that is faster than recomputing the new ordering from scratch; (2) to efficiently answer queries related to the data contained in a particular range of the current orderings. Our solution fully generalizes the notion of B-trees to higher dimensions by carefully combining space-driven and data-driven partitions. As a result, our methods are competitive with other popular data structures used as multidimensional indexing tools (including k-d trees, quad-trees, grid files, R-trees).