Compact and queryable representation of raster datasets

Susana Ladra, José R. Paramá, and Fernando Silva-Coira

Universidade da Coruña, Facultade de Informática,
Campus de Elviña s/n, 15071 A Coruña, Spain
{susana.ladra,jose.parama,fernando.silva}@udc.es

Palabras clave

Geographic information systems, data compression, query processing, raster datasets

Publicación original

Procs. of the 28th International Conference on Scientific and Statistical Database Management (SSDBM'16), ACM, Article No. 15, 2016.
(DOI: 10.1145/2949689.2949710)

Resumen

Compact data structures join in a unique data structure a compressed representation of the data and the access mechanisms to retrieve such data. The objective is to keep that data structure always in main memory, and thus, data should be managed directly in that compressed representation. This approach produces two benefits: we can manage larger datasets in main memory and we make a better usage of the memory hierarchy.

This work presents a compact data structure, called k^2-*raster*, designed to represent raster data, which usually represents attributes of the space (i.e., temperatures, elevation measures, etc.). It uses a quadtree-based strategy combined with compressed representations of trees and integer sequences. This obtains very little spaces, requiring less than 5% of the original space for some datasets. In addition, it is capable of speeding up queries, such as efficiently returning the value stored at any cell or retrieving all cells containing values within a specific range of values inside any subregion of the raster.

There are previous compact data structures designed to store raster data, which work well when the raster dataset has few different values. Nevertheless, when the dataset has a large number of different values, the performance in space consumption and query response time degrades. The presented data structure competes with previous approaches in the first scenario, but scales much better when the number of different values and the size of the dataset increase, which is critical when applying over real datasets.