

Improving Clustered Quad-trees Performance through Local Reorganization

Vicente Yriarte, Antonio Vivas and Walter Cunto

Universidad Simón Bolívar

Departamento de Computación y Tecnología de la Información

Apartado 89000, Caracas, Venezuela

e-mail: yriarte@ldc.usb.ve, avivasl@yahoo.com

e-mail: Walter_Cunto@mckinsey.com

ABSTRACT

Traditionally, analysis of quad-trees performance unrealistically assume that keys are uniformly distributed. Actually keys tend to cluster. In order to model clustered data, we propose that insertion in quad-trees may occur in any of the tree's leaves with equal probability. This leads to more unbalanced trees and suggests that suitable tree balance techniques could be useful. For this reason, we apply local reorganizations to fringe sub-trees of constant size. The impact of such a technique is a 26% reduction in the average depth of a quad-tree with n keys, from $\frac{4}{3}H_n + O(1)$ to $\frac{2.9539}{3}H_n + O(1)$, while the variance improves by 71%. This improvement in the average depth is 27% away from the optimal.

Keywords: quad-tree, clustered data, local reorganization, locally balanced quad-trees, fringe analysis, multi-dimensional data, data structure, algorithms analysis.

RESUMEN

Tradicionalmente, en los análisis del comportamiento de los quad-trees se ha asumido, de manera irreal, que las claves están uniformemente distribuidas. Pero en realidad las claves tienden a formar *clusters*. Con el fin de modelar "clustered data," proponemos que la inserción en los quad-trees puede ocurrir en cualquiera de las hojas del árbol con igual probabilidad. Esto conduce a árboles más desbalanceados y sugiere que se usen técnicas adecuadas de balanceo. Por esta razón, aplicamos reorganización local a sub-árboles de tamaño constante en la copa del árbol. El resultado de dicha reorganización es una reducción de 26% en la profundidad promedio de un quad-trees con n claves, de $\frac{4}{3}H_n + O(1)$ a $\frac{2.9539}{3}H_n + O(1)$, y una mejora de 71% en la varianza. Esta mejora en la profundidad promedio coloca a la estructura a sólo 27% del valor óptimo.

Palabras Claves: *quad-tree*, datos agomerados (*clustered data*), reorganización local, *quad-trees* localmente balanceados, análisis de la copa (*del fringe*), datos multi-dimensionales, estructura de datos, análisis de algoritmos.