
TÉCNICAS PARA SELECCIÓN DINÁMICA DE PERMUTANTES

ALEJANDRO SEBASTIÁN MEDINA NEGRETE
INGENIERO CIVIL EN COMPUTACIÓN

RESUMEN

El almacenamiento en bases de datos cambia progresivamente. Hoy en día no sólo es necesario almacenar números o palabras, también las bases de datos están siendo exigidas a almacenar documentos, archivos multimedia, muestras biológicas y huellas digitales entre otros tipos de datos, con muchos de los cuales ya no es posible realizar búsquedas de manera tradicional, debido a que no es posible agrupar o seleccionar por alguna característica de comparación campo a campo. Las consultas por similitud o búsqueda aproximada se presentan como una solución para realizar este tipo de búsquedas, para lo cual es necesario un índice, el cual se construye utilizando una función de distancia que determina que tan similares son dos elementos. Una técnica para la construcción de índices es la basada en permutantes y establece que dado un conjunto de datos se pueden seleccionar elementos como puntos de referencia, donde el grado de cercanía de dos elementos está dado por la similitud de sus permutaciones. Una desventaja que esta técnica tiene es que el índice sólo sirve para un conjunto estático de datos, o sea, no puede recibir operaciones de inserción, extracción ni edición, ya que no es posible adaptar el índice para que mantenga su eficiencia y, por lo tanto, pierde su propósito. El objetivo de este trabajo es tener índices que se construyan de forma dinámica y que puedan recibir operaciones, pero que no vean afectada de manera significativa su eficiencia. Para ello, se exponen técnicas dinámicas que permiten la construcción del índice de manera incremental y la efectividad del índice construido se determina comparando búsquedas sobre el índice construido con búsquedas sobre el conjunto total de datos. De este trabajo se concluye que las técnicas para manipular los índices implementados tienen distinto rendimiento según el tipo de dato. El parámetro importante a considerar es el valor de α , el cual regula la selección de permutantes y tiene directa relación en el rendimiento del índice. Además, se ha determinado que los índices no pierden rendimiento cuando se extraen elementos. Lo anterior indica que los resultados y conclusiones obtenidas son un indicio a considerar en otros espacios métricos.