

A High-Efficiency Querying Algorithm for Multi-Attribute Text String Identifiers

馬恆, 鄭弘裕

Industrial Engineering and System Management
Management

hengma@chu.edu.tw

Abstract

This research focuses on developing a high-speed query algorithm for multi-attribute large-volume data, in which data compression and fast querying can be simultaneously achieved. In the past decade, this type of problem is usually resolved by storing the data in a database, and querying is carried out through database-embedded methods. As the data volume is growing in a tenfold fashion, these methods are encountering difficulties in data manipulation as far as efficiency is concerned. A number of approaches, e.g., the Bloom Filters, have been proposed for querying purposes in large-volume data using a compact structure requiring less memory space. These approaches often limit themselves to single-attribute querying, and the multi counterpart can only be achieved by layered implementations where errors, e.g., the false positives, are prone to increase. We propose an improved CMAC mechanism seeking to achieve fast multi-attribute querying with less memory space and errors. Experimental results show favorable performances for our proposed mechanism.

Keyword : Multi-attribute, CMAC, Bloom Filter