

Metody indeksowania atrybutów zawierających zbiory

Mikołaj MORZY, Zbyszko KRÓLIKOWSKI

Politechnika Poznańska, Instytut Informatyki
ul. Piotrowo 3a, 60-965 Poznań

Otrzymano 18 listopada 2002 roku

Streszczenie. Atrybuty zawierające zbiory są naturalnym i wygodnym sposobem modelowania złożonych obiektów świata rzeczywistego. Współczesne relacyjno-obiektowe systemy baz danych umożliwiają przechowywanie zbiorów w postaci pojedynczych atrybutów oraz wspierają wydawanie zapytań do takich atrybutów. Efektywne wykonywanie zapytań w bazie danych jest możliwe między innymi dzięki indeksom, które wydatnie przyspieszają dostęp do danych. W przeciwieństwie do atrybutów skalarnych, dla atrybutów zawierających zbiory trudno jest zaproponować satysfakcjonujący mechanizm indeksowania, który jednakowo efektywnie wspierałby różne rodzaje zapytań do takiego atrybutu, np. zapytania o zawieranie zbioru lub zapytania o niepuste przecięcie z danym zbiorem. Niniejsza praca stanowi przegląd aktualnie dostępnych technik indeksowania zbiorów. Przedstawiono w niej zarówno techniki wywodzące się z relacyjnych i obiektowych systemów baz danych (grupowy indeks bitmapowy, haszowy indeks bitmapowy), jak i wywodzące się z dziedzin pokrewnych, np. metody zaczerpnięte z przetwarzania tekstu (pliki odwrócone, indeks sygnaturowy, S-drzewo) czy systemów informacji geograficznej (RD-drzewo). Praca zawiera również wyniki eksperymentów obliczeniowych, których celem było porównanie efektywności wykorzystania wybranych indeksów w rozwiązywaniu problemów charakterystycznych dla eksploracji danych.

Słowa kluczowe: indeksowanie zbiorów, indeks bitmapowy, indeks haszowy, eksploracja danych