

## VII-O-4

### VECTƠ BIT ĐỘNG: CÁCH TIẾP CẬN MỚI ĐỂ KHAI THÁC TẬP PHỔ BIẾN

*Võ Đình Bầy<sup>1</sup>, Lê Hoài Bắc<sup>2</sup>*

<sup>1</sup>Khoa Công nghệ Thông tin, Trường ĐH Kỹ thuật Công nghệ Tp.HCM

<sup>2</sup>Khoa Công nghệ Thông tin, Trường ĐH Khoa học Tự nhiên – ĐHQG Tp.HCM

#### Tóm tắt

Khai thác tập phổ biến là một tiến trình quan trọng trong khai thác luật kết hợp. Mục tiêu của nó là tìm tất cả các tập phổ biến có độ phổ biến (support) thỏa ngưỡng cho trước. Đã có nhiều phương pháp được đề xuất để giải quyết vấn đề này. Bài báo nhằm đề nghị một cách tiếp cận mới cho khai thác tập phổ biến bao gồm: i) Đưa ra khái niệm DBV (Dynamic Bit Vector – vectơ bit động) và một số thao tác trên các DBV; ii) Đề nghị thuật toán khai thác tập phổ biến dựa trên DBV. Kết quả thực nghiệm chứng tỏ DBV hiệu quả hơn các tiếp cận dựa trên định dạng dữ liệu dọc là Tidset và BitTable

Từ khoá: Tập phổ biến, vectơ bit động, BitTable, luật kết hợp

### DYNAMIC BIT VECTOR A NEW APPROACH FOR MINING FREQUENT ITEMSETS

*Vo Dinh Bay<sup>1</sup>, Le Hoai Bac<sup>2</sup>*

<sup>1</sup>Faculty of Information Technology, University of Technology HCMC

<sup>2</sup>Faculty of Information Technology, University of Science - VNU HCMC

#### Abstract

Mining frequent itemsets is an important task in mining association rules. Its purpose is finding all itemsets having supports satisfy minimum support threshold. Up to now, there are a lot of approaches to solve this problem. This paper proposes a new approach for mining frequent itemsets: i) propose the concept of Dynamic Bit Vector (DBV), and operations in DBV; ii) propose an algorithm for mining frequent itemsets based on DBV. Experimental results show that our approach is more efficient than that of Tidset and BitTable.

Key words: frequent itemset, Dynamic Bit Vector, BitTable, association rules