**ĐẠI HỌC QUỐC GIA TP.HCM**

**TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN**

Họ tên HVCH: NGUYỄN DUY CHINH

**ĐỀ CƯƠNG NGHIÊN CỨU ĐỒ ÁN**

**TỐT NGHIỆP THẠC SĨ**

Tên đề tài: KHAI THÁC TOP-K SỰ KIỆN ĐỒNG XUẤT HIỆN VỚI BITTABLE

Chuyên ngành: HỆ THỐNG THÔNG TIN

Mã số chuyên ngành: 60 48 01 04

Xác nhận của cán bộ hướng dẫn

(Ký tên và ghi rõ họ tên)

Họ tên:…………………………

**Tp. HCM, tháng 08 năm 2017**

1. **Mục tiêu đồ án**

Bài toán khai thác top-k sự kiện đồng xuất hiện được đề xuất đầu tiên bởi Zhi-Hong Deng từ năm 2015. Trong bài báo của mình ông đã trình bày 4 thuật toán cơ bản và 2 thuật toán dựa vào cấu trúc Pi-Tree để giải quyết bài toán. Mục tiêu của đồ án là áp dụng một cấu trúc dữ liệu khác đó là BitTable để giải quyết bài toán cho thời gian xử lý nhanh hơn.

1. **Tóm tắt nội dung**

Khai thác tập phổ biến được đề xuất đầu tiên bởi Agrawal et al. nhằm phân tích giỏ hàng để rút ra tất cả các luật liên kết. Kể từ khi việc khai thác dữ liệu này được đề xuất và những thuật toán khai thác hữu ích liên quan, đã có hàng nghìn công trình nghiên cứu theo sau.

Trong khai thác tập phổ biến, một tập là phổ biến nếu tuần suất xuất hiện của nó trong cơ sở dữ liệu không nhỏ hơn một ngưỡng cho trước. Có nghĩa là, tập sự kiện phổ biến là một khái niệm bao trùm trong phạm vi tất cả cơ sở dữ liệu nhưng chúng không có sự liên quan đến nhau . Tuy nhiên, trong một số ứng dụng như các hệ thống gợi ý hoặc mạng xã hội, người ta sẽ quan tâm đến những sự kiện có liên quan với nhau nhiều hơn. Ví dụ trong hệ thống đưa ra gợi ý của một cửa hàng trực tuyến, bất cứ khi nào một người dùng thanh toán một vài sản phẩm, hệ thống phải đưa ra những gợi ý những sản phẩm khác mà thường mua cùng với những sản phẩm đó để bán thêm nhiều sản phẩm.

Đề tài nghiên cứu các thuật toán cơ bản để giải quyết bài toán, đồng thời áp dụng cấu trúc dữ liệu BitTable để giải quyết bài toán cho thời gian xử lý nhanh hơn.

1. **Yêu cầu**

* Nắm bắt được các thuật toán cơ bản.
* Cài đặt cấu trúc bittable và áp dụng để giải quyết bài toán.

1. **Thời gian thực hiện đồ án**

Từ tháng 8/2017 đến tháng 10/2017

1. **Phương pháp nghiên cứu**

* Thu thập tài liệu về khai thác top-k
* Tìm hiểu các thuật toán cơ bản
* Tìm hiểu cấu trúc BitTable
* Tìm hiểu môi trường phát triển trên ngôn ngữ lập trình C#

1. **Tài liệu tham khảo**

* Mining Top-K Co-Occurrence Items - Zhi-Hong Deng – 2015
* BitTableFI: An efficient mining frequent itemsets algorithm – Jie Dong,Min Han

Thời gian làm đồ án dự kiến từ tháng 8 đến tháng 10 năm 2017