

**ĐẠI HỌC THÁI NGUYÊN
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG**

DƯƠNG VĂN LƯƠNG

**THUẬT TOÁN TỤ NĂNG LƯỢNG BEA (BOND ENERGY
ALGORITHM) TRONG THIẾT KẾ CƠ SỞ DỮ LIỆU PHÂN TÁN
TUYỂN SINH TRUNG HỌC PHỔ THÔNG
TỈNH THÁI NGUYÊN**

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

THÁI NGUYÊN - 2016

**ĐẠI HỌC THÁI NGUYÊN
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG**

DƯƠNG VĂN LƯƠNG

**THUẬT TOÁN TỤ NĂNG LƯỢNG BEA (BOND ENERGY
ALGORITHM) TRONG THIẾT KẾ CƠ SỞ DỮ LIỆU PHÂN TÁN
TUYỂN SINH TRUNG HỌC PHỔ THÔNG
TỈNH THÁI NGUYÊN**

Chuyên ngành: Khoa học máy tính

Mã số: 60 48 01 01

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

NGƯỜI HƯỚNG DẪN KHOA HỌC: PGS. TS. Đỗ Trung Tuấn

Thái Nguyên, 2016

Lời cam đoan

Tôi cam đoan những kết quả trong luận văn là của việc tìm hiểu, có trích dẫn và tham chiếu đến các nguồn tư liệu tin cậy. Nội dung luận văn không sao chép từ các kết quả của các luận văn, luận án khác.

Lời cảm ơn

Trước tiên, tôi xin gửi lời cảm ơn đến trường Đại học Công nghệ thông tin và truyền thông Thái Nguyên, đã tạo điều kiện và tổ chức khóa học này để tôi có thể có điều kiện tiếp thu những kiến thức mới, có thời gian học tập và hoàn thành luận văn cao học này.

Tôi xin chân thành cảm ơn các thầy cô đã truyền đạt cho chúng tôi những kiến thức quý báu trong quá trình học tập và làm luận văn.

Tôi chân thành cảm ơn ta bè cùng lớp đã giúp đỡ, động viên tôi trong quá trình học tập cũng như thực hiện luận văn.

Cuối cùng, tôi xin gửi thành quả này tới gia đình và người thân của tôi, những người đã hết lòng tạo điều kiện và động viên tôi để tôi có được kết quả ngày hôm nay.

Mục lục

Lời cam đoan.....	iii
Lời cảm ơn	iv
Mục lục	v
DANH MỤC CÁC TỪ VIẾT TẮT	viii
CÁC BẢNG, HÌNH VẼ.....	ix
MỞ ĐẦU	1
1. Lý do chọn đề tài	1
2. Cấu trúc luận văn.....	5
Chương 1.....	7
KHÁI NIỆM VỀ CƠ SỞ DỮ LIỆU PHÂN TÁN.....	7
1.1. Hệ thống cơ sở dữ liệu tập trung.....	7
1.2. Giới thiệu hệ thống cơ sở dữ liệu phân tán.....	8
1.3. Ưu điểm, nhược điểm hệ thống cơ sở dữ liệu phân tán.....	9
1.3.1. Ưu điểm.....	9
1.3.2. Khuyết điểm	10
1.4. Các đặc điểm của hệ thống xử lý phân tán	10
1.5. Kiến trúc của một hệ CSDL phân tán	11
1.6. Các hệ thống phân tán.....	13
1.6.1. Mô hình khách chủ	13
1.6.2. Mô hình phân tán ngang hàng	14
1.6.3. Mô hình nhân bản chính - phụ	15
1.6.4. So sánh các mô hình.....	16

1.7. Nhu cầu về cơ sở dữ liệu phân tán.....	16
1.7.1. Công tác tuyển sinh THPT tỉnh Thái Nguyên.....	16
1.7.2. Hạ tầng cơ sở dữ liệu về giáo dục Thái Nguyên.....	19
1.7.3. Hạ tầng dịch vụ Internet tại Thái Nguyên và ứng dụng cho giáo dục.....	20
1.7.4. Nhu cầu thiết kế cơ sở dữ liệu phân tán phục vụ công tác tuyển sinh THPT.....	22
1.8. Kết luận chương.....	24
Chương 2.....	25
CƠ SỞ LÝ THUYẾT LIÊN QUAN.....	25
2.1. Phân đoạn dữ liệu.....	25
2.1.1. Giới thiệu.....	25
2.1.2. Nhu cầu phân đoạn dữ liệu.....	26
2.1.3. Các điều kiện ràng buộc trong thiết kế phân đoạn.....	27
2.2. Phân đoạn ngang.....	28
2.3. Phân đoạn dọc.....	28
2.3.1. Giới thiệu.....	28
2.3.2. Quá trình phân đoạn.....	29
2.3.4. Thuật toán tụ năng lượng BEA và ứng dụng.....	30
2.3.5. Xác định ma trận sử dụng các thuộc tính.....	35
2.3.6. Xây dựng ma trận hấp dẫn xếp cụm.....	36
2.3.7. Phân đoạn ma trận hấp dẫn.....	43
2.4. Kết luận chương.....	50
Chương 3.....	51

THỬ NGHIỆM VÀ ỨNG DỤNG THỰC TẾ.....	51
3.1. Hệ thống cơ sở dữ liệu phục vụ tuyển sinh THPT tại Sở	51
3.1.1. Sơ đồ cơ sở dữ liệu phân tán	51
3.1.2. Các truy cập dữ liệu từ các trường	52
3.2. Phân đoạn dọc	52
3.2.1. Đặt vấn đề.....	52
3.2.2. Áp dụng thuật toán tụ năng lượng BEA.....	53
3.2.3. Thực hiện với dữ liệu cụ thể.....	56
3.3. Các sơ đồ thuật toán	58
3.3.1. Chương trình chính.....	58
3.3.2. Thuật toán phân đoạn ma trận hấp dẫn phân cụm.....	58
3.3.3. Thuật toán tìm phần trên, phần dưới trong ma trận hấp dẫn.....	59
3.4. Kết luận chương.....	61
KẾT LUẬN.....	62
Kết quả luận văn đạt được.....	62
Phương hướng tìm hiểu và thực nghiệm tiếp theo của luận văn...	63
TÀI LIỆU THAM KHẢO	64
Tiếng Việt.....	64
Tiếng Anh.....	64
PHỤ LỤC.....	65

DANH MỤC CÁC TỪ VIẾT TẮT

CLR	Server Common Language Runtime
CNTT	Công nghệ Thông tin
CSDL	Cơ sở dữ liệu
CSDLPT	Cơ sở dữ liệu phân tán
DB	Database, cơ sở dữ liệu
DB2	Tên cơ sở dữ liệu của IBM
DD	Data Dictionary
DDB	Distributed DataBase
DDBMS	Distributed Data Base Management System
GDDT	Giáo dục và đào tạo
GV	Giáo viên
HS	Học sinh
ICT	Công nghệ thông tin và truyền thông
IDE	Integrated Development Environment
ISO	International Organization for Standardization
Khách/ chủ	Mô hình hệ thống máy chính - máy dự phòng
Multitier Client – Server	Mô hình đa khách - chủ
Peer - to - Peer	Mô hình nhóm làm việc
RCS	Revision Control System
SQL	Structured Query Language
THPT	Trung học phổ thông
VNPT	Vietnam Posts and Telecommunications Group
XML	eXtensible Markup Language

CÁC BẢNG, HÌNH VẼ

Hình 1.1: Môi trường CSDL tập trung.....	7
Hình 1.2. Hệ thống cơ sở dữ liệu phân tán	8
Hình 1.3: Môi trường CSDL phân tán	9
Hình 1.4: Sơ đồ tạo CSDL phân tán	11
Hình 1.5: Sơ đồ quá trình tạo CSDL phân tán	12
Hình 1.6. Hệ thống khách/ chủ.....	14
Hình 1.7. Mô phỏng hệ thống Khách/ chủ.....	15
Bảng 1.1: So sánh các mô hình phân tán	16
Hình 1.8. Trang Web của Sở GDĐT Thái Nguyên	16
Hình 1.8. Một số trang Web của các trường	17
Bảng 1.2. Các trường THPT trong mạng lưới phân tán tại tỉnh Thái Nguyên	17
Hình 1.10. Cổng thông tin điện tử liên quan đến cơ sở dữ liệu văn bản	19
Hình 1.11. Cơ sở dữ liệu tập trung.....	23
Hình 1.12. Hình dạng thí dụ về cơ sở dữ liệu phân tán đối với công tác tuyển sinh	24
Hình 2.1. Minh họa về sự phân đoạn trên ổ đĩa	25
Hình 2.2. Phân đoạn dữ liệu.....	26
Hình 2.3. Minh họa tách bảng.....	29
Hình 2.4. Bảng dữ liệu biết được khi thiết kế cơ sở dữ liệu	30
Bảng 2.1. Dữ liệu từ hình trên.....	30
Hình 2.5. Thí dụ về mức tụ năng lượng ME.....	32

Hình 2.6. Chương trình SQL để khai báo lược đồ.....	35
Bảng 2.2. Quá trình P_k truy cập C_i	35
Hình 2.7. Ma trận sử dụng các thuộc tính.....	35
Hình 2.8. Các tần suất truy cập của các quá trình P_j	36
Bảng 2.3. Các AFF tính được dựa trên các ACC, như kết quả trong hình 5..	36
Hình 2.9. Ma trận hấp dẫn xếp cụm.....	37
Hình 2.10. Ma trận hấp dẫn xếp cụm với đường chéo được tính lại	37
Hình 2.11. Thí dụ tính BOND (C_1, C_3).....	38
Hình 2.12. Xét C_3 đối với cặp C_1, C_p đã hình thành	38
Hình 2.13. Thuộc tính C_3 được đặt giữa C_1, C_p do $CONT(C_1, C_3, C_p)$ đạt max	39
Hình 2.14. Vị trí tương đối trước khi xét C_4	40
Hình 2.15. Kết quả xếp cột C_4	41
Hình 2.16. Cột C_5 cần xếp đối với bốn cột đã có.....	41
Hình 2.17. Ma trận hấp dẫn cuối cùng.....	43
Hình 2.18. Ma trận hấp dẫn cuối cùng đã sắp xếp.....	43
Hình 2.19. Ma trận sử dụng các thuộc tính, như hình 4, nhưng cho T truy cập nhiều A_i	44
Hình 2.20. Chọn điểm x	45
Hình 2.21. Ma trận đầu vào cho việc tìm x lần đầu	46
Hình 2.22. Ma trận đầu vào đối với C_5, C_5	47
Hình 2.23. Chuyển vị sau một lần tính giá trị cực đại	47