BỘ GIÁO DỤC VÀ ĐÀO TẠO

ĐẠI HỌC THÁI NGUYÊN

TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN & TRUYỀN THÔNG

NINH QUANG TRUNG

THUẬT TOÁN XÁC ĐỊNH CHA CHUNG GẦN NHẤT CỦA HAI NÚT TRONG CÂY ỨNG DỤNG PHÂN TÍCH ĐA DẠNG LOÀI VI SINH VẬT

***Chuyên ngành*: Khoa học máy tính**

***Mã số*: 60.48.01**

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

Thái Nguyên, năm 2014

BỘ GIÁO DỤC VÀ ĐÀO TẠO

ĐẠI HỌC THÁI NGUYÊN

TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN & TRUYỀN THÔNG

NINH QUANG TRUNG

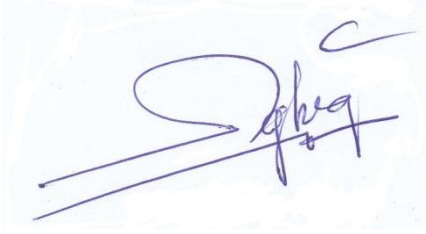
CÁC PHƯƠNG PHÁP XÁC ĐỊNH CHA CHUNG GẦN NHẤT CỦA HAI NÚT TRONG CÂY, ỨNG DỤNG PHÂN TÍCH ĐA DẠNG LOÀI VI SINH VẬT

***Chuyên ngành*: Khoa học máy tính**

***Mã số*: 60.48.01**

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

Người hướng dẫn khoa học: TS. Nguyễn Cường



LỜI CAM ĐOAN

Tôi xin cam đoan: Luận văn này là công trình nghiên cứu thực sự của cá nhân, được thực hiện dưới sự hướng dẫn khoa học của *Tiến sĩ Nguyễn Cường*. Các số liệu, những kết luận nghiên cứu được trình bày trong luận văn này trung thực và chưa từng được công bố dưới bất cứ hình thức nào.

Tôi xin chịu trách nhiệm về nghiên cứu của mình.

Học viên

Ninh Quang Trung

LỜI CẢM ƠN

Lời đầu tiên, tôi xin chân thành cảm ơn ***Tiến sĩ Nguyễn Cường*** người đã trực tiếp hướng dẫn tôi hoàn thành luận văn. Với những lời chỉ dẫn, những tài liệu, sự tận tình hướng dẫn và những lời động viên của Thầy đã giúp tôi vượt qua nhiều khó khăn trong quá trình thực hiện luận văn này.

Tôi cũng xin cảm ơn quý Thầy (Cô) giảng dạy chương trình cao học "Khoa học máy tính" đã truyền dạy những kiến thức quý báu, những kiến thức này rất hữu ích và giúp tôi nhiều khi thực hiện nghiên cứu.

Xin cảm ơn các quý Thầy (Cô) công tác tại Trường Đại học Công nghệ thông tin và truyền thông – Đại học Thái Nguyên đã tạo điều kiện cho tôi được tham gia và hoàn thành khóa học.

Tôi xin chân thành cảm ơn.

Học viên

Ninh Quang Trung

MỤC LỤC

DANH MỤC CÁC HÌNH ẢNH

DANH MỤC CÁC BẢNG BIỂU

DANH MỤC CÁC TỪ VIẾT TẮT-THUẬT NGỮ

Cụm từ viết tắt	Cụm từ chi tiết
ASCII	American Standard Code for Information Interchange
BFS	Breadth First Search
Bp	Basepair
DFS	Depth – First – Search
DNA	Deoxyribo Nucleic Acid
E	Edges
G	Graph
GA	Genome Analyzer
HGP	Human Genome Project
LCA	Lowest Common Ancestor
MEGAN	MetaGenomeANalyzer
MGA	MetaGeneAnnotator
NCBI	National Center for Biotechnology Information
NGS	Next Genration Sequencing
PCR	Polymerase chain reaction
RMQ	Range-Minimum Query

MỞ ĐẦU

Lý thuyết đồ thị là ngành khoa học xuất hiện từ rất lâu nhưng lại có nhiều ứng dụng hiện đại. Những ý tưởng cơ bản của nó được đưa ra từ thế kỷ thứ 18 bởi nhà toán học Thụy Sĩ Leonhard Euler. Ông đã dùng đồ thị để giải quyết bài toán cây cầu Konigsberg nổi tiếng. Từ đó lý thuyết đồ thị ngày càng khẳng định được vị trí quan trọng của mình trong việc áp dụng để giải quyết bài toán thực tế nhờ vào việc tìm ra ngày càng nhiều của các định lý, công thức và thuật toán.

Đặc biệt trong những năm trở lại đây, cùng với sự ra đời của máy tính điện tử và sựphát triển nhanh chóng của Tin học, Lý thuyết đồ thị càng được quan tâm đến nhiều hơn. Đặc biệtlà các thuật toán trên đồ thị đã có nhiều ứng dụng trong nhiều lĩnh vực khác nhau như: Mạng máy tính, Lý thuyết mã, Tối ưu hoá, Kinh tế học, tìm đường đi ngắn nhất, bài toán luồng cực đại, bài toán vận chuyển, bài toán luồng tổng quát và bài toán xác định cha chung gần nhất của hai nút trong cây cũng là một ứng dụng trong lý thuyết đồ thị. Hiện nay, lý thuyết đồ thị là một trong những kiến thức cơ sở của bộ môn khoa học máy tính.

Trong phạm vi một đề tài không thể nói kỹ và nói hết những vấn đề của lý thuyết đồ thị. Luận văn này trình bày lý thuyết đồ thị dưới góc độ khảo sát những thuật toán cơ bản nhất có thể cài đặt được trên máy tính một số ứng dụng của nó.

Trong khuôn khổ luận văn học viên ứng dụng lý thuyết đồ thị để giải bài toán xác định cha chung gần nhất của hai nút trong cây nhằm mục đích phân tích đa dạng loài vi sinh sử dụng công nghệ đọc trình tự thế hệ mới.

Ngày nay thông tin về trình tự gen rất hữu ích trong những nghiên cứu về sinh học phân tử và trong nhiều lĩnh vực ứng dụng như chuẩn đoán, sinh học pháp y, hệ thống sinh học…Quá trình đọc trình tự hay giải trình tự