

BỘ GIÁO DỤC VÀ ĐÀO TẠO
ĐẠI HỌC ĐÀ NẴNG

NGUYỄN THỊ PHƯƠNG DUNG

**XÂY DỰNG HỆ THỐNG HỖ TRỢ TƯ VẤN
TUYỂN SINH VÀ ĐÀO TẠO TÍN CHỈ TẠI
TRƯỜNG ĐẠI HỌC QUẢNG NAM**

**Chuyên ngành: KHOA HỌC MÁY TÍNH
Mã số: 60.48.01**

TÓM TẮT LUẬN VĂN THẠC SĨ KỸ THUẬT

Đà Nẵng - Năm 2011

Công trình được hoàn thành tại
ĐẠI HỌC ĐÀ NẴNG

Người hướng dẫn khoa học: **PGS.TSKH. Trần Quốc Chiến**

Phản biện 1: **PGS.TS Võ Trung Hùng**

Phản biện 2: **TS. Nguyễn Mậu Hân**

Luận văn được bảo vệ trước Hội đồng chấm Luận văn tốt nghiệp thạc sĩ kỹ thuật họp tại Đại học Đà Nẵng vào ngày 15 tháng 10 năm 2011

Có thể tìm hiểu luận văn tại:

- Trung tâm Thông tin-Học liệu, Đại học Đà Nẵng
- Trung tâm Học liệu, Đại học Đà Nẵng

MỞ ĐẦU

1. Lý do chọn đề tài

Trong giai đoạn đất nước phát triển hiện nay, giáo dục luôn là vấn đề quan tâm hàng đầu của xã hội. Hiện nay hai vấn đề được quan tâm nhiều nhất là công tác tuyển sinh đầu vào và chất lượng đào tạo ở các trường đại học.

Đối với công tác tuyển sinh đã có rất nhiều chương trình tư vấn tuyển sinh rộng khắp trên tất cả các tỉnh, thành trong cả nước với nhiều hình thức phong phú. Tuy nhiên thực tế thí sinh còn cảm thấy thiếu thông tin về các trường, các ngành mình quan tâm. Với mục đích trên, luận văn đi vào Xây dựng hệ thống hỗ trợ tư vấn tuyển sinh cho Trường Đại học Quảng Nam.

Hình thức đào tạo tín chỉ là hình thức mới, gây ra nhiều khó khăn cho nhiều sinh viên. Vì thế luận văn cũng đi vào Xây dựng hệ thống hỗ trợ tư vấn đào tạo tín chỉ cho sinh viên mà đặc biệt tư vấn hướng dẫn sinh viên chọn môn học phù hợp.

2. Mục đích nghiên cứu

Bước đầu nghiên cứu một số cơ sở lý thuyết khai phá dữ liệu để hỗ trợ phần nào giúp các thí sinh có thể chọn đúng ngành, nghề và cấp học phù hợp khi đăng kí thi vào trường, giúp các bạn sinh viên hiểu rõ hơn về học chế tín chỉ, từ đó chọn được môn học phù hợp, chủ động trong xây dựng kế hoạch học... Đề tài cũng đưa ra chương trình demo nhỏ minh họa việc xây dựng hệ thống hỗ trợ tư vấn.

3. Đối tượng và phạm vi nghiên cứu

Đối tượng nghiên cứu:

- Kho dữ liệu, phương thức quản lý và vận hành kho dữ liệu. Nghiên cứu kỹ thuật khám phá tri thức và khai phá dữ liệu.

Phạm vi nghiên cứu

- Nghiên cứu xây dựng, quản lý kho dữ liệu, khai phá dữ liệu trong công tác tư vấn tuyển sinh. Nghiên cứu xây dựng hệ thống đào tạo tín chỉ tại trường Đại học Quảng Nam.

4. Phương pháp nghiên cứu

Nghiên cứu lý thuyết :

- Dựa vào tri thức về khai phá dữ liệu như cây quyết định và luật kết hợp và các thuật toán suy diễn để xây dựng hệ thống tư vấn tuyển sinh trực tuyến trên mô hình khai phá dữ liệu Business Intelligence Development Studio. Sử dụng các công cụ của hệ quản trị SQL server trong data warehouse

Nghiên cứu thực nghiệm

- Dựa trên các nghiên cứu về lý thuyết để xây dựng ứng dụng “Tư vấn tuyển sinh và đào tạo tín chỉ tại trường Đại học Quảng Nam”. Chạy ứng dụng thử nghiệm trên máy đơn.

5. Ý nghĩa khoa học và thực tiễn của đề tài

Về mặt khoa học

Đề tài sẽ đưa ra một hệ thống tư vấn hỗ trợ công tác tư vấn tuyển sinh và đào tạo tín chỉ

Về mặt thực tiễn

Đề tài tạo ra được kho dữ liệu hỗ trợ tư vấn, tra cứu nhằm nắm được những thông tin về tư vấn tuyển sinh. Bên cạnh đề tài còn đưa ra hệ thống tư vấn chọn các môn học trong đào tạo tín chỉ.

6. Bố cục của luận văn

Báo cáo của luận văn được được tổ chức thành 3 chương

CHƯƠNG 1. TÌM HIỂU TƯ VẤN TUYỂN SINH

VÀ ĐÀO TẠO TÍN CHỈ TẠI TRƯỜNG ĐẠI HỌC QUẢNG NAM

1.1. Giới thiệu về công tác tư vấn tuyển sinh và đào tạo tín chỉ tại trường Đại học Quảng Nam

1.1.1. Về công tác tư vấn tuyển sinh

Trường Đại học Quảng Nam trước đây là Trường Cao đẳng Sư phạm Quảng Nam, được thành lập ngày 08/6/2007. Nhà trường có chức năng, nhiệm vụ đào tạo đa cấp, đa ngành, đa hệ từ trung cấp chuyên nghiệp; cao đẳng cho đến bậc đại học với các hình thức đào tạo: chính qui, liên thông, vừa làm vừa học; bồi dưỡng chuẩn hoá và bồi dưỡng thường xuyên cho giáo viên các cấp;

Hàng năm cứ đến khoảng tháng 3, trường Đại học Quảng Nam bắt đầu thông báo tuyển sinh các ngành học với hình thức thông báo đa dạng từ thông tin trên cuốn Cẩm nang tuyển sinh đại học – cao đẳng, các báo, đài, website chính thức của trường, phối hợp với báo Thanh Niên, báo Tuổi Trẻ tham gia các buổi tư vấn tuyển sinh trên địa bàn tỉnh cũng như các tỉnh lân cận, nhằm thu hút nhiều hơn sự quan tâm của thí sinh đến với trường.

1.1.2. Về đào tạo tín chỉ

Bắt đầu từ khóa tuyển sinh 2010-2011, trường sẽ đào tạo theo hệ thống tín chỉ đối với sinh viên hệ đại học.

Cho đến nay, trường ĐH Quảng Nam đã áp dụng từng bước việc đào tạo tín chỉ thay cho đào tạo thường niên, quá trình chuyển giao này chắc chắn không tránh khỏi những khó khăn, thách thức từ cả

người dạy và người học. Vì thế rất cần một hệ thống hỗ trợ, cung cấp thông tin, giải đáp thắc mắc cho mọi người tham gia.

1.2. Yêu cầu của hệ thống

a. Đối với tư vấn tuyển sinh

Thao tác dễ dàng, đơn giản, chỉ cần thí sinh có máy vi tính và đường truyền internet là có thể sử dụng được.

Đảm bảo cung cấp đầy đủ thông tin về công tác tuyển sinh của trường

Xây dựng cơ chế tư vấn, giúp thí sinh có thể chọn được đúng ngành nghề phù hợp.

b. Đối với tư vấn đào tạo tín chỉ

Sử dụng đơn giản, dễ dàng, thích hợp với mọi người.

Hệ thống có thể đưa ra những lựa chọn phù hợp cho sinh viên trong các vấn đề về việc học như : đăng ký môn học, lựa chọn số môn/học kì...

1.3. Mô tả hoạt động của hệ thống

Hệ thống giao tiếp được thông qua giao diện website, dễ sử dụng và thân thiện với hầu hết mọi người. Đối với tư vấn tuyển sinh, sau khi cung cấp một số thông tin cơ bản cho hệ thống (trả lời các câu hỏi mà hệ thống đưa ra), người sử dụng sẽ nhận được các tư vấn về chọn ngành, chọn cấp bậc thi phù hợp với bản thân.

Đối với tư vấn đào tạo tín chỉ, sinh viên mỗi ngành học sẽ nhận được các tư vấn về chọn môn học cho mỗi học kì, môn nào cần học trước, để học rút ngắn thì cần kế hoạch học ra sao, nếu thi lại môn đó thì có thể đăng kí vào thời gian nào.

1.4. Tìm hiểu mô hình tuyển sinh và đào tạo tín chỉ

1.4.1. Tư vấn tuyển sinh

1.4.1.1. Hình thức tư vấn tuyển sinh

Tư vấn tuyển sinh trước các kì thi Đại học – Cao đẳng luôn là công việc hết sức quan trọng và đã được tổ chức thường xuyên, rộng khắp hàng năm với rất nhiều các hình thức và nội dung như :

- Cẩm nang tuyển sinh đại học – cao đẳng
- Tư vấn trực tiếp – ngày hội tư vấn tuyển sinh hướng nghiệp

Ngoài ra còn có tư vấn qua website của trường, tư vấn trực tuyến, tư vấn qua chat, tư vấn qua đài phát thanh – truyền hình, tư vấn qua điện thoại...

1.4.1.2. Đặc điểm tuyển sinh tại trường Đại học Quảng Nam

Trong công tác tuyển sinh, trường cũng đã chủ động xây dựng trang web tuyển sinh riêng cho mình nhằm cung cấp thông tin đầy đủ nhất về tuyển sinh ĐH-CĐ hàng năm. Ngoài ra trường cũng tích cực phổ biến giới thiệu các ngành nghề đào tạo, chỉ tiêu thông qua báo chí, tham gia công tác tư vấn tại các trường THPT, giải đáp trực tiếp qua điện thoại...

1.4.2. Tư vấn đào tạo tín chỉ

1.4.2.1. Hình thức đào tạo tín chỉ

Đào tạo theo Hệ thống tín chỉ cho phép sinh viên có thể chủ động học theo điều kiện và năng lực của mình. Những học chế tín chỉ là mỗi môn học được lượng hóa bằng một tín chỉ. Sinh viên (SV) tích lũy dần, hoàn thành chương trình học của mình theo số tín chỉ chứ không phải lên lớp theo từng học kì, từng năm học như ở phổ thông.

Để hiểu rõ hơn về đào tạo tín chỉ, cần phải tìm hiểu một số định nghĩa:

Tín chỉ (credit)

Một tín chỉ (credit unit)

Giờ tín chỉ (credit hour)

1.4.2.2. Đào tạo tín chỉ tại trường Đại học Quảng Nam

Theo lộ trình, trường Đại học Quảng Nam sẽ chính thức triển khai áp dụng từ năm học 2010-2011 cho tất cả các ngành bậc đại học hệ chính quy khóa K10 (tuyển sinh vào năm 2010).

1.5. Kết luận

Trong nội dung chương này, tôi đã trình bày cơ sở về công tác tư vấn tuyển sinh và đào tạo tín chỉ tại trường Đại học Quảng Nam và mô tả hoạt động của hệ thống tư vấn.

Phần tiếp theo của luận văn này, tôi xin trình bày cơ sở lý thuyết về kho dữ liệu, khai phá dữ liệu bằng cây quyết định và tìm hiểu SQL Server 2005 và khai phá dữ liệu với Business Intelligence Development Studio (BIDS) trong SQL Server. Từ đó xây dựng các hệ thống tư vấn dựa trên khai phá dữ liệu bằng cây quyết định trong chương 3 được mô tả rõ ràng.

CHƯƠNG 2. KHO DỮ LIỆU VÀ TỔNG QUAN VỀ KHAI PHÁ DỮ LIỆU

2.1. Kho dữ liệu (Data Warehouse – DW)

2.1.1. Tổng quan về kho dữ liệu

Kho dữ liệu là tuyển tập các cơ sở dữ liệu tích hợp, hướng chủ đề, được thiết kế để hỗ trợ cho chức năng trợ giúp quyết định.

Theo John Ladley, Công nghệ kho dữ liệu (Data Warehouse Technology) là tập các phương pháp, kỹ thuật và các công cụ có thể kết hợp, hỗ trợ nhau để cung cấp thông tin cho người sử dụng trên cơ sở tích hợp từ nhiều nguồn dữ liệu, nhiều môi trường khác nhau.

Định nghĩa do W.H. Inman đề xướng: DW được hiểu là một tập hợp các dữ liệu tương đối ổn định (không hay thay đổi), cập nhật theo thời gian, được tích hợp theo hướng chủ đề nhằm hỗ trợ quá trình tạo quyết định về mặt quản lý.

2.1.2. Mục đích của kho dữ liệu

Mục tiêu chính của kho dữ liệu tổng quát là nhằm đáp ứng các tiêu chuẩn cơ bản sau:

- Phải có khả năng đáp ứng mọi yêu cầu về thông tin của NSD
- Hỗ trợ để các nhân viên của tổ chức thực hiện tốt, hiệu quả công việc của mình, như có những quyết định hợp lý, nhanh và bán được nhiều hàng hơn v.v.
- Giúp cho tổ chức, xác định, quản lý và điều hành các dự án, các nghiệp vụ một cách hiệu quả và chính xác.
- Tích hợp dữ liệu và các siêu dữ liệu từ nhiều nguồn khác nhau

2.1.3. Đặc tính của kho dữ liệu

- Tính tích hợp (Integration)
- Dữ liệu gắn thời gian và có tính lịch sử

- Dữ liệu có tính ổn định (nonvolatility)
- Dữ liệu không biến động
- Dữ liệu tổng hợp

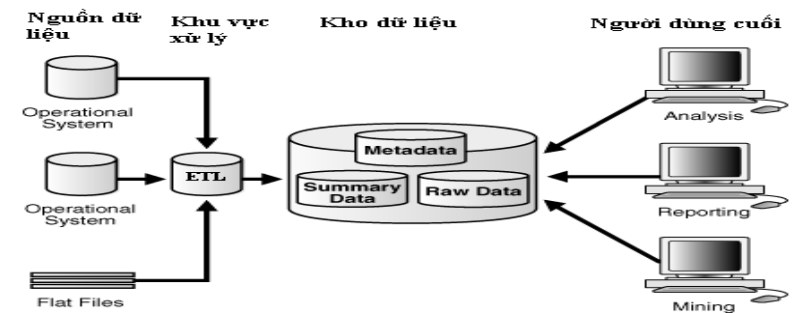
2.1.4. Quy trình xây dựng kho dữ liệu và các vấn đề liên quan

2.1.4.1 Kho dữ liệu và cơ sở dữ liệu

- Trước tiên DW là database rất lớn
- Database hướng về xử lý thời gian thực, DW hướng về tính ổn định.
- Phục vụ xử lý transaction, cập nhật. Datawarehouse thường chỉ đọc, phục vụ cho những nhu cầu báo cáo.
- DW sẽ lấy thông tin có thể từ nhiều nguồn khác nhau: DB2, Oracle, SQLserver thậm chí cả File thông thường rồi làm sạch chúng và đưa vào cấu trúc của nó-đó là VLDB (very large database).
- Một điểm quan trọng là Database thường được chuẩn hóa (Dạng chuẩn 1, 2, 3, BCK) để khai thác.

2.1.4.2. Kiến trúc kho dữ liệu

Mô hình kiến trúc của kho dữ liệu cơ bản gồm có ba thành phần : Dữ liệu nguồn, khu vực xử lý và kho dữ liệu.



Hình 2.1 Kiến trúc kho dữ liệu

2.2. Khám phá tri thức và khai phá dữ liệu

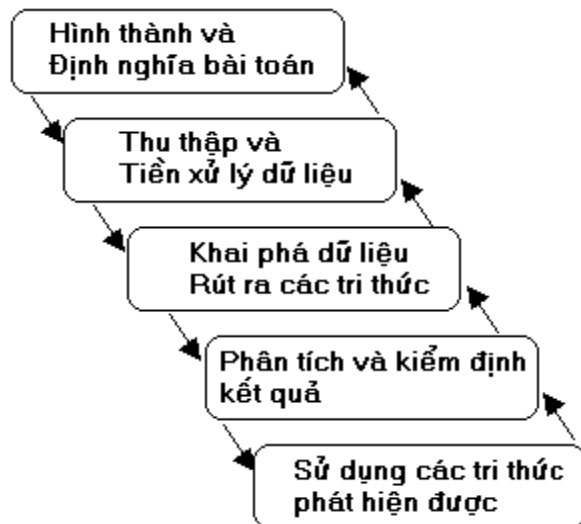
2.2.1. Tổng quan về khám phá tri thức và khai phá dữ liệu

Phát hiện tri thức (Knowledge Discovery) trong các cơ sở dữ liệu là một qui trình nhận biết các mẫu hoặc các mô hình trong dữ liệu với các tính năng: hợp thức, mới, khả ích, và có thể hiểu được.

Khai phá dữ liệu là một bước trong quy trình phát hiện tri thức gồm có các thuật toán khai thác dữ liệu chuyên dùng dưới một số quy định về hiệu quả tính toán chấp nhận được để tìm ra các mẫu hoặc các mô hình trong dữ liệu.

2.2.2. Quá trình phát hiện tri thức

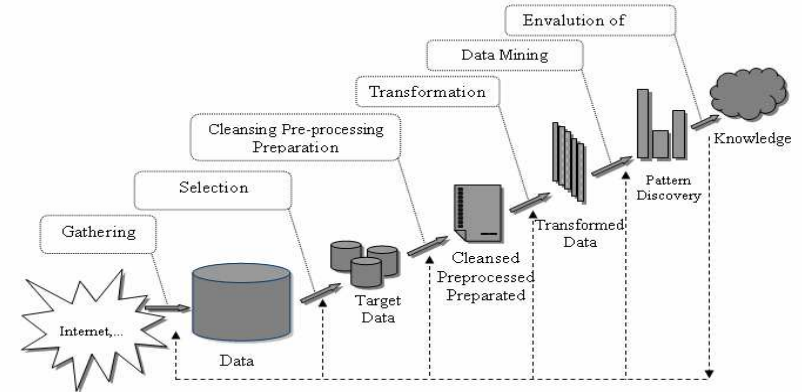
Quá trình khám phá tri thức được tiến hành qua 5 bước sau:



Hình 2.5. Quá trình khám phá tri thức

2.2.3. Quá trình khai phá dữ liệu

Quá trình này gồm có 6 bước:



Hình 2.6. Quá trình khai phá dữ liệu

2.2.4. Các kỹ thuật khai phá dữ liệu

Trong thực tế có nhiều kỹ thuật khai phá dữ liệu khác nhau nhằm thực hiện hai chức năng mô tả và dự đoán.

- Kỹ thuật khai phá dữ liệu mô tả
- Kỹ thuật khai phá dữ liệu dự đoán

Một số kỹ thuật phổ biến thường được sử dụng để khai phá dữ liệu hiện nay là: Phân lớp dữ liệu, phân cụm dữ liệu, khai phá luật kết hợp, hồi quy, giải thuật di truyền, mạng *noron*, cây quyết định

2.2.5. Khai phá dữ liệu bằng cây quyết định

2.2.5.1. Định nghĩa cây quyết định

Cây quyết định là một mô tả tri thức dạng đơn giản nhằm phân các đối tượng dữ liệu thành một số lớp nhất định. Các nút của cây

được gán nhãn là tên các thuộc tính, các cạnh được gán các giá trị có thể của các thuộc tính, các lá miêu tả các lớp khác nhau. Các đối tượng được phân lớp theo các đường đi trên cây, qua các cạnh tương ứng với giá trị của thuộc tính của đối tượng tới lá.

Tạo luật: Các luật được tạo ra nhằm suy diễn một số mẫu dữ liệu có ý nghĩa về mặt thống kê.

2.2.5.2. Vấn đề xây dựng cây quyết định

Quá trình xây dựng cây quyết định đều được chia ra làm 3 giai đoạn cơ bản: Xây dựng cây, cắt tía cây, đánh giá cây.

2.2.5.3. Rút ra các luật từ cây quyết định

Có thể chuyển đổi qua lại giữa mô hình cây quyết định và mô hình dạng luật (IF ...THEN...). Hai mô hình này là tương đương nhau.

2.2.5.4. Các thuật toán khai phá dữ liệu bằng cây quyết định

a) Thuật toán CLS:

Thuật toán CLS được thiết kế theo chiến lược chia để trị từ trên xuống.

b) Thuật toán ID3

ID3 xây dựng cây quyết định từ trên- xuống (top -down).

c) Thuật toán C4.5

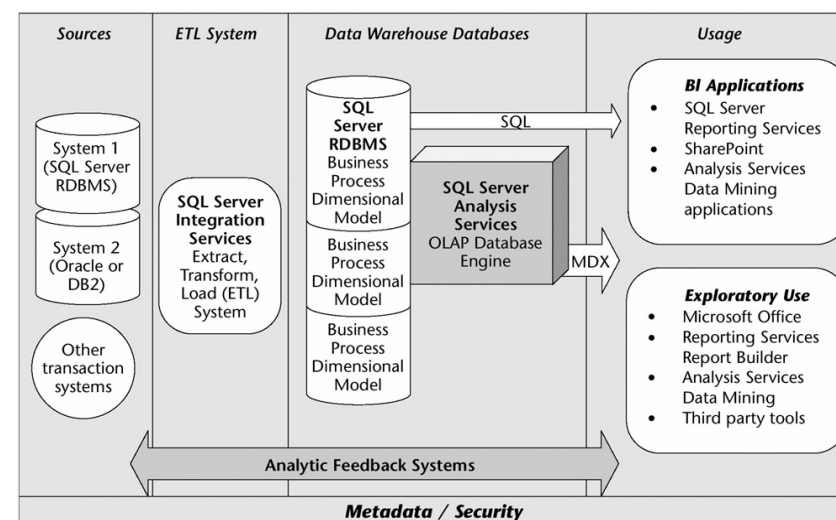
Thuật toán C4.5 là một thuật toán được cải tiến từ thuật toán ID3 với việc cho phép xử lý trên tập dữ liệu có các thuộc tính số (numeric attributes) và và làm việc được với tập dữ liệu bị thiếu và bị nhiễu. Nó thực hiện phân lớp tập mẫu dữ liệu theo chiến lược ưu tiên theo chiều sâu (Depth - First). Giới thiệu SQL server và công cụ xây dựng mô hình khai phá dữ liệu Business Intelligence Development Studio (BIDS)

2.3. Giới thiệu về ngôn ngữ SQL Server

Hệ quản trị cơ sở dữ liệu Microsoft SQL Server (MSSQL) là một trong những hệ quản trị cơ sở dữ liệu thông dụng hiện nay với ưu điểm có các công cụ quản lý mạnh mẽ giúp cho việc quản lý và bảo trì hệ thống dễ dàng, hỗ trợ nhiều phương pháp lưu trữ, phân vùng và đánh chỉ mục phục vụ cho việc tối ưu hóa hiệu năng. MSSQL 2005 có 4 dịch vụ lớn : Database Engine, Intergration Service, Reporting service, Analysis Services.

2.3.1. Xây dựng kho dữ liệu dựa trên các công cụ của Microsoft SQL Server

Các công cụ kho dữ liệu



Hình 2.8 Các công cụ của SQL server 2005

2.3.2. Giới thiệu công cụ xây dựng mô hình khai phá dữ liệu Business Intelligence Development Studio (BIDS)

BIDS là công cụ cho phép tổ chức quản lý và khai thác kho dữ liệu (Xử lý phân tích trực tuyến) cũng như xây dựng các mô hình khai phá dữ liệu rất dễ sử dụng và hiệu quả của Microsoft.

Quy trình Xây dựng mô hình khai phá dữ liệu với BIDS như sau :

- Tạo mới 1 project (Analysis Services Project)
- Tạo một Data Source
- Tạo một Data Source View
- Tạo một Mining model structure
- Tạo các Mining models
- Khai thác Mining models
- Kiểm tra độ chính xác của Mining Models
- Sử dụng Mining Models để dự đoán.

2.4. Kết luận

Chương này đã trình bày phân lý thuyết cơ bản về kho dữ liệu, khai phá dữ liệu, đồng thời nêu rõ việc sử dụng SQL Server và công cụ BIDS để

khai phá dữ liệu đưa ra cây quyết định, từ đó suy diễn được các luật.

CHƯƠNG 3. KHAI PHÁ DỮ LIỆU CHO HỆ THỐNG TƯ VẤN TUYỂN SINH VÀ PHÂN TÍCH HỆ THỐNG ĐÀO TẠO TÍN CHỈ

3.1. Tư vấn tuyển sinh

3.1.1. Kho dữ liệu trong tư vấn tuyển sinh

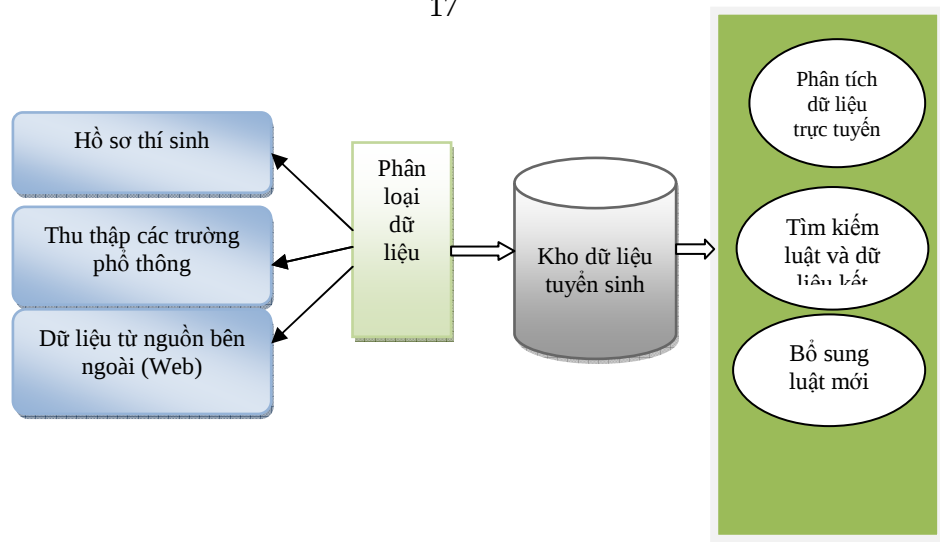
3.1.1.1. Mục đích xây dựng kho dữ liệu cho hệ thống tư vấn tuyển sinh

Các học sinh muốn tìm kiếm các thông tin bổ ích về các trường đại học cao đẳng trong cả nước để tham gia thi tuyển. Nhu cầu thông tin để các học sinh tham khảo thật sự cần thiết nhằm đảm bảo phù hợp với nhu cầu năng lực, sở thích, điều kiện kinh tế gia đình, điều kiện khoảng cách địa lý, giá trị bằng cấp của ngôi trường mình chọn...đây là một nhu cầu rất thiết thực và để đảm bảo thông tin đầy đủ và chính xác cần có một lượng dữ liệu lớn về thông tin tuyển sinh hằng năm được thu thập và phân loại chính xác. Từ đó đưa ra các quyết định nhằm hỗ trợ cho sinh viên có hướng chọn đúng ngành nghề phù hợp và đầy đủ các khía cạnh cho những sinh viên cụ thể. Từ đó tôi quyết định đưa ra giải pháp hình thành một kho dữ liệu nguồn được cập nhật và khai phá tốt để cung cấp cho việc định hướng và tư vấn cho thí sinh hằng năm muốn có thông tin tuyển sinh vào đại học Quảng Nam.

3.1.1.2. Thu thập dữ liệu và phân loại dữ liệu

Dữ liệu nguồn cho việc khai phá và đánh giá đưa ra quyết định cho thông tin tư vấn sẽ được lấy từ nhiều nguồn dữ liệu khác nhau. Dữ liệu sẽ được tập hợp từ các hồ sơ mà mỗi học sinh đã nộp vào trường để dự tuyển ta sẽ phân loại theo các thông tin

Việc thu thập dữ liệu được đưa ra như sơ đồ bên dưới:



Hình 3.1. Quá trình thu thập dữ liệu, phân lớp để giải quyết bài toán

Việc lưu trữ thông tin vào kho dữ liệu có dạng như dưới:

- (1) Bảng dữ liệu ngành
- (2) Bảng dữ liệu thông tin học
- (3) Dữ liệu lưu trữ điểm chuẩn theo từng ngành
- (4) Bảng dữ liệu lưu trữ kết quả thi

3.1.2. Khai thác và phân tích quy luật và lựa chọn giải pháp cho bài toán

Hệ thống cần phải đáp ứng làm sao đủ thông tin mà trường cung cấp trước mỗi đợt tuyển sinh. Có một hệ thống các câu hỏi được tạo ra một cách tự động và giải quyết được số lượng lớn các thắc mắc của từng học sinh, phụ huynh và những người quan tâm.

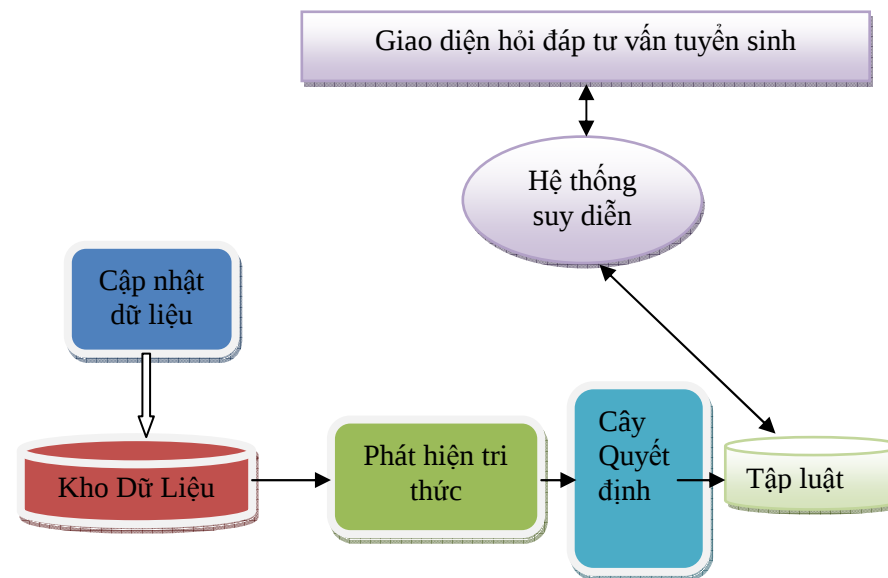
Việc đưa ra những quyết định sẽ được xác định từ những dữ liệu đầu vào là những câu trả lời từ những người sử dụng đã nhập vào. Hệ thống sẽ xem xét dựa trên cây quyết định mà đã được xây dựng để

đưa ra những tư vấn chính xác. Như vậy yêu cầu đặt ra ở đây là cây quyết định được phát sinh từ đâu. Đó chính là quá trình khai phá dữ liệu để tìm ra tri thức phục vụ nhu cầu và mục đích bài toán.

Kết quả mong muốn là xây dựng nên cây quyết định và tập các luật đưa ra cho bài toán tư vấn tuyển sinh dựa trên cây quyết định đó. Trong phạm vi đề tài sẽ tìm hiểu đến thuật toán mà microsoft đã sử dụng để phát sinh cây quyết định được tích hợp sẵn trong bộ Microsoft SQL server.

3.1.3. Khai phá tri thức và đưa ra tập luật dựa trên cây quyết định và ứng dụng suy diễn cho bài toán tư vấn tuyển sinh

3.1.3.1. Mô hình bài toán tư vấn tuyển sinh



Hình 3.2. Mô hình thực hiện bài toán tư vấn tuyển sinh

Đề tài chỉ tìm hiểu và ứng dụng công cụ phát sinh cây quyết định đã được tích hợp sẵn trong bộ Microsoft SQL Server và sử dụng tập luật được phát sinh từ suy diễn cây quyết định đó để đưa ra quyết định phân loại cho tập các câu hỏi mà sẽ hỗ trợ cho việc tư vấn tuyển sinh. Tập luật này sẽ được làm mới qua thời gian vì nó phụ thuộc vào kho dữ liệu nguồn.

3.1.3.2. Huấn luyện mô hình

Lựa chọn các thuộc tính sau :

Bảng 3.5. Bảng dữ liệu thống kê kết quả thi vào Đại học Quảng Nam

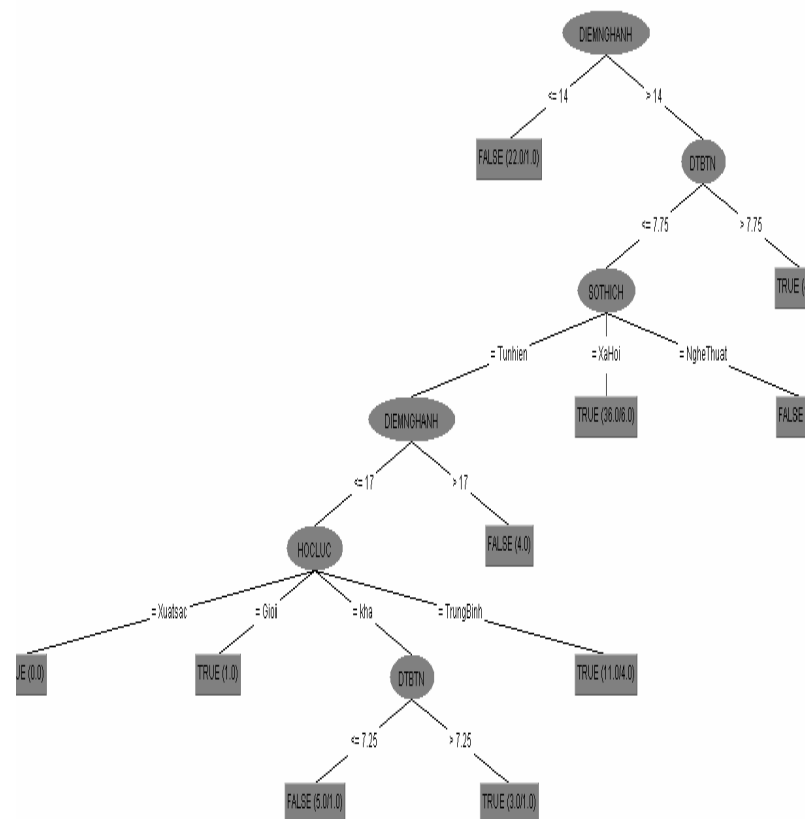
Các Thuộc Tính									Phân Lớp
Mã HS	Điểm TB Mon TN	Điểm TB Mon XH	HocLuc	SoThich	Khoi Chon	Diem thi	Ma Nghanh	DiemNghanh	Kếtquả Thi
AA01	9	9	Xuat sac	Tunhien	A	22	100	19	Yes
AA02	9	8	Gioi	Tunhien	A	20	100	19	yes
AA03	8.5	8	Gioi	Tunhien	A	19	101	18	Yes

3.1.3.3. Khai phá dữ liệu bằng SQL Server Business Intelligence

Development Studio

CSDL dùng để khai phá là bảng Data với các thuộc tính và một số dữ liệu mẫu như trên. Để tìm ra mối liên hệ giữa các thuộc tính để từ đó rút ra được các quy luật tư vấn, ta có thể xét sự liên quan giữa một số thuộc tính input và thuộc tính suy đoán

Sau khi thực thi ta có Cây quyết định



Hình 3.9. Cây quyết định

3.1.3.4. Phân tích Tập luật trong tư vấn tuyển sinh

Bộ luật sinh ra từ cây quyết định ta có tập 4370 luật. Từ tập dữ liệu đầu vào gồm 1200 bản ghi.

Với việc hệ thống đưa ra giao diện hỏi đáp để lấy các thông tin đầu vào của người sử dụng và motor suy diễn từ tập các luật được

phát sinh từ đó ta đưa ra các kết quả là các ngành phù hợp mà thí sinh đó dự thi có khả năng đậu cao nhất như dự đoán điểm số có thể đạt được của thí sinh đó.

Từ giải thuật cây quyết định, dựa trên CSDL có được, chúng ta tiến hành phân lớp và có được các luật liên quan đến câu hỏi học sinh thi đậu hay không khi đã biết học lực, sở thích và khối ngành muốn thi.

3.2. Phân tích hệ thống đào tạo tín chỉ

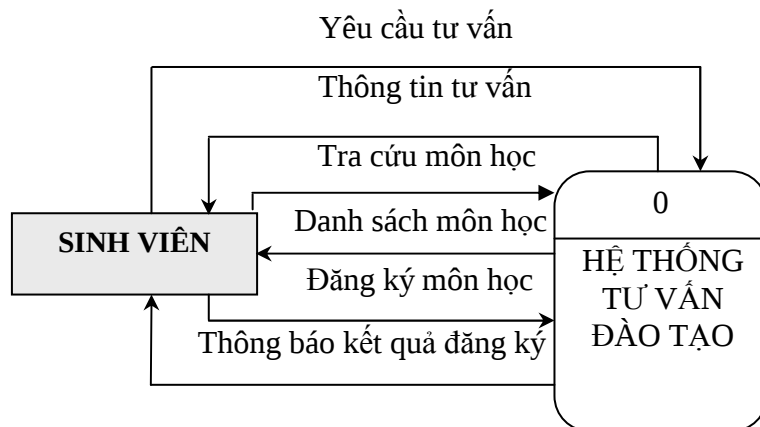
3.2.1. Tổng quan bài toán tư vấn đào tạo tín chỉ

Nêu tổng quan bài toán tư vấn đào tạo tín chỉ

3.2.2. Phân tích thiết kế hệ thống

3.2.2.1. Mô hình chức năng

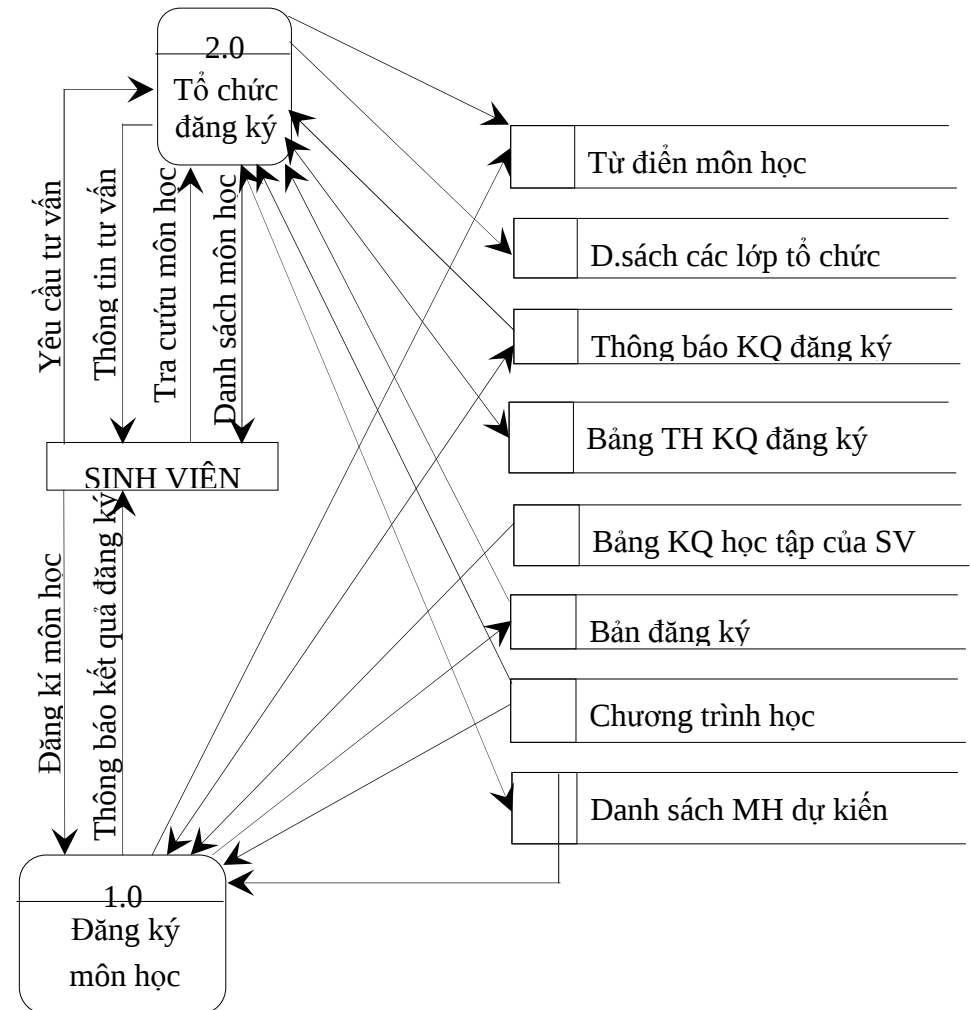
Biểu đồ ngữ cảnh



Hình 3.12 Biểu đồ ngữ cảnh hệ thống tư vấn đào tạo tín chỉ

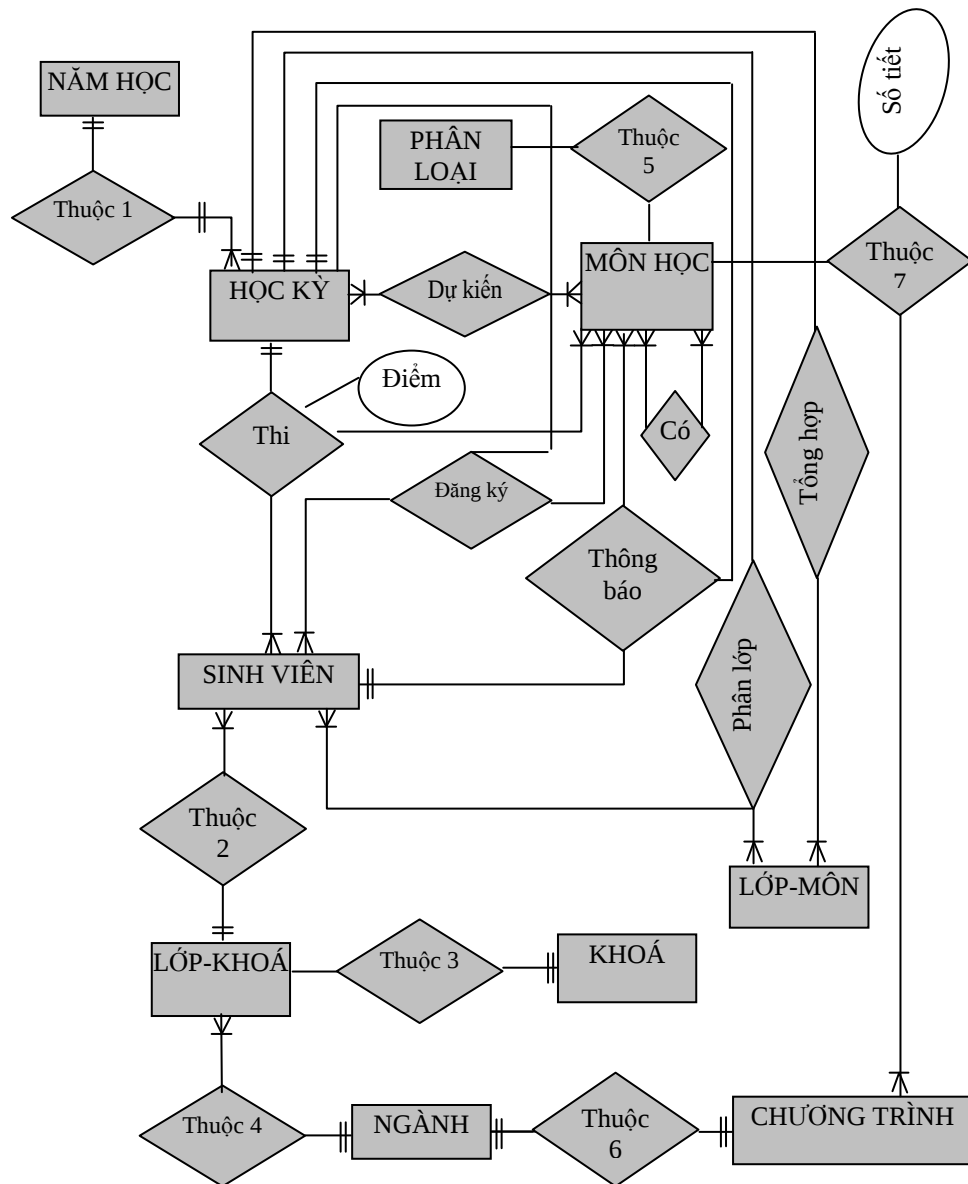
3.2.2.2. Biểu đồ luồng dữ liệu

Biểu đồ luồng dữ liệu mức 0



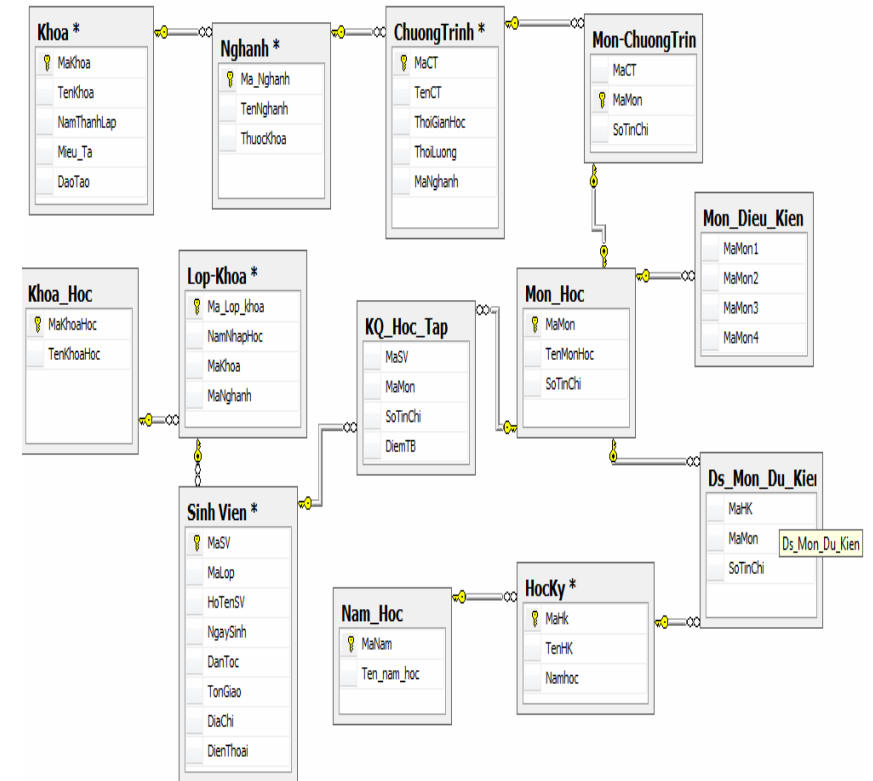
Hình 3.14 Biểu đồ luồng dữ liệu mức 0 của đăng ký môn học

3.2.2.3. Mô hình dữ liệu hệ thống - Mô hình E-R



Hình 3.16 Mô hình E-R

3.2.2.4. Thiết kế cơ sở dữ liệu



Hình 3.17 Lược đồ quan hệ cho mô hình tư vấn đào tạo tín chỉ

3.2.2.5. Quy trình hiện thực bài toán tư vấn đào tạo

Bài toán tư vấn sẽ được xây dựng mà không dùng hệ chuyên gia hay kho dữ liệu, vì có quá nhiều ràng buộc. Thay vào đó vấn đề này cần được xây dựng như 1 hệ thống thông tin mang tính tư vấn riêng.

3.3. Xây dựng giao diện

3.3.1. Màn hình tư vấn tuyển sinh

Để Được Tư Vấn Thông Tin Ngành Học Phù Hợp Với Bạn Nhất Xin Cung Cấp Chính Xác Các Thông Tin Bên Dưới

Thông Tin Cá Nhân Họ Và Tên : Dung * Nơi Sinh : Quảng Nam * Ngày Sinh : * Điện Thoại : 0902754636 * Email : Dung@duat.edu.vn * Trường THPT Bạn học : Lê Quý Đôn - Quảng Nam *		Thông Tin Tư Vấn Học Lực Của Bạn: Giỏi * Điểm Trung Bình Tự Nhiên : 9 * Điểm Trung Bình Xã Hội: 7.5 * Sở Thích : ☒ Tự Nhiên ☐ Xã Hội ☐ Nghệ Thuật * Khối Chọn Thi : A * Bạn Muốn Nhận Thông Tin Tư Vấn Thông Qua ? ☒ Email ☒ Điện Thoại ☐ Chỉ Xem	
--	--	--	--

Hình 3.18 Màn hình tư vấn tuyển sinh

3.3.2. Màn hình tư vấn đào tạo tín chỉ

TRƯỜNG ĐẠI HỌC QUẢNG NAM
QUANG NAM UNIVERSITY

Thông Tin Tuyển Sinh | Tư Vấn Tuyển Sinh | Tư Vấn Đào Tạo

Tên Sinh Viên : Nguyễn Thùy Linh
Mã Số : SV012011

Hệ Thống Đào Tạo Tín Chỉ Trường Đại Học Quảng Nam

Tra Cứu Môn Học :

Đăng Ký Môn Học :

Thông Tin Tư Vấn :

Hình 3.22. Màn hình giao diện chính tư vấn đào tạo tín chỉ

3.3.3. Đánh giá kết quả

Các chức năng hệ thống hoạt động theo đúng yêu cầu đặt ra, bao gồm tư vấn ngành nghề cho hệ tư vấn tuyển sinh, đồng thời hỗ trợ tư vấn cho việc lựa chọn môn học phù hợp cho sinh viên.

KẾT LUẬN

1. Các kết quả đã đạt được trong luận văn

Luận văn đã trình bày tổng quan các vấn đề lý thuyết như kho dữ liệu, khai phá dữ liệu, cây quyết định và các luật. Dựa vào motor suy diễn để suy diễn ra kết quả nhằm tư vấn lựa chọn ngành phù hợp với trình độ năng lực của từng học sinh.

Trong quá trình thực hiện cũng đã xây dựng được hệ thống tư vấn phục vụ cho công tác tư vấn tuyển sinh và đào tạo tín chỉ của nhà trường.

2. Kiến nghị hướng phát triển :

Nguồn dữ liệu khai thác cho bài toán tuyển sinh nên được khai thác từ nhiều nguồn khai phá từ tất cả các thể loại dữ liệu được lưu trữ ở trong mạng lan của trường. cũng như thông qua internet của các cơ quan giáo dục có tham gia tuyển sinh khác. Để từ đó có nguồn dữ liệu lớn để có thể đưa ra các luật có tính khách quan và sát với thực tế.

Hệ thống phải cải tiến bằng cách phát hiện và phân loại các tập luật mới và phân loại mức hữu dụng của các tập luật này.

Bài toán tư vấn tuyển sinh và đào tạo tín chỉ mới dừng lại ở việc đưa ra các quyết định tư vấn theo hệ thống cây quyết định và các luật xây dựng sẵn . Đây mới là một phần của hệ chuyên gia vì vậy mong muốn phát triển hệ thống này lên một hệ chuyên gia. Có khả năng tương tác với người sử dụng cao và hệ thống tập luật và motor suy diễn tốt .