

**BỘ GIÁO DỤC VÀ ĐÀO TẠO  
ĐẠI HỌC ĐÀ NẴNG**



**VÕ ĐÌNH TÁ**

**HOÀN THIỆN HỆ THỐNG NGỮ VỤNG TIẾNG HRÊ  
ỨNG DỤNG XÂY DỰNG TỪ ĐIỂN HRÊ – VIỆT  
VÀ VIỆT – HRÊ**

**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.TS. Phan Huy Khánh**

Phản biện 1: **PGS. TS. Đoàn Văn Ban**

Phản biện 2: **PGS. TS. Võ Trung Hùng**

Luận văn được bảo vệ tại Hội đồng chấm Luận văn tốt nghiệp thạc sĩ kỹ thuật họp tại Đại học Đà Nẵng ngày 10 và 11 tháng 09 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

Cộng đồng dân tộc thiểu số tỉnh Quảng Ngãi sống tập trung chủ yếu ở 330 thôn của 63 xã thuộc 6 huyện miền núi và 16 xã miền núi thuộc 6 huyện đồng bằng. Dân số khoảng 284.770 người, dân tộc thiểu số chiếm 53,42% dân số trên địa bàn và chiếm 11,83% dân số toàn tỉnh; trong đó dân tộc H'rae khoảng 110.000 người; dân tộc Cor khoảng 28.000 người, dân tộc Cadong khoảng 16.000 người, dân tộc khác khoảng 300 người.

Tộc người thiểu số H'rae sử dụng ngữ hệ Nam Á, hiện sống chủ yếu ở các huyện Ba Tơ, Minh Long, Sơn Hà, Sơn Tây tỉnh Quảng Ngãi và huyện An Lão, tỉnh Bình Định.

Tiếng H'rae không có chữ viết truyền thống và chưa được latin hoá chính thức trong thời gian vừa qua. Trong thời gian kháng chiến chống Pháp và chống Mỹ, một số cán bộ người Kinh và người H'rae đã có một số cố gắng latin hóa văn bản tiếng H'rae để làm tài liệu tuyên truyền và dạy chữ nhưng quá trình này đến nay chưa có kết quả chính thức, hoàn chỉnh và có hệ thống.

Ở miền Nam dưới chế độ cũ, từ năm 1958 đến 1971, Viện ngôn Ngữ học Mùa hè (Summer Institute of Linguistics -SIL) của Mỹ đã cử một số nhà khoa học đến nghiên cứu các ngôn ngữ thiểu số phía Nam và những mục đích riêng, trong đó có tiếng H'rae. Trên cơ sở đó đã có một số sản phẩm về phương án chữ viết, về ngữ pháp và từ vựng được sử dụng nhưng cũng chưa chính thức.

Hiện nay, tiếng H'rae được phát trên sóng Đài Phát thanh Truyền hình Quảng Ngãi và đài truyền thanh các huyện miền núi. Văn bản tiếng H'rae dạng viết vẫn là phương án tạm thời.

Tóm lại, hiện trạng tiếng Hrê hiện nay là: chưa được tin học hóa; ngữ vựng chưa hoàn thiện; sử dụng không thống nhất; còn quá ít tài liệu nghiên cứu và hướng dẫn học tiếng Hrê,... Tuy nhiên, với sức sống mãnh liệt của một dân tộc đã trải qua lịch sử phát triển lâu đời và có những đóng góp to lớn trong công cuộc đấu tranh bảo vệ đất nước, tiếng Hrê cần được giữ gìn và phát triển lên một tầm cao mới nhằm góp phần bảo tồn những giá trị văn hóa tốt đẹp của dân tộc mình, đồng thời là phương tiện để đồng bào nâng cao đời sống vật chất cũng như tinh thần trong bối cảnh hội nhập trong nước cũng như quốc tế.

Xuất phát từ thực tế trên, một giải pháp nhằm góp phần nâng cao hiệu quả việc học tiếng Hrê, từ đó nâng cao hiệu quả tuyên truyền chủ trương của Đảng, chính sách của Nhà nước đến đồng bào, đồng thời góp phần xây dựng hoàn chỉnh hệ thống chữ viết Hrê phục vụ cho công tác dạy và học tiếng Hrê, tôi thực hiện đề tài: “Hoàn thiện hệ thống ngữ vựng tiếng Hrê, ứng dụng xây dựng từ điển Hrê – Việt và Việt – Hrê”.

## **2. Mục đích của đề tài**

Mục đích chính của đề tài là trên cơ sở những cái đã có xây dựng hoàn thiện kho ngữ vựng Hrê có cấu trúc mở, dễ kế thừa. Từ đó, ứng dụng xây dựng từ điển Hrê-Việt và Việt – Hrê để phục vụ cho công tác dạy và học tiếng Hrê.

## **3. Đối tượng phạm vi nghiên cứu**

Đối tượng nghiên cứu của đề tài là giới hạn trong phạm vi: tìm hiểu đặc điểm, cấu trúc tiếng Hrê; nghiên cứu tìm hiểu những tài liệu đã có về tiếng Hrê, từ đó hoàn thiện kho ngữ vựng, ứng dụng xây dựng từ điển Hrê – Việt và Việt – Hrê

#### **4. Phương pháp triển khai**

- Công cụ được xây dựng bởi ngôn ngữ lập trình ASP.NET/C# trên nền Dot Net 2005 truy cập dữ liệu từ XML. Quá trình thực hiện:

- Thu thập tài liệu từ sách, báo, internet và các nguồn khác về tiếng Hrê. Sau đó, tổng hợp các tài liệu liên quan.

- Nghiên cứu đặc trưng của tiếng Hrê, vấn đề từ điển, cơ sở dữ liệu đa ngữ.

- Nghiên cứu giải pháp kỹ thuật và công cụ cập nhật làm giàu kho ngữ vựng (sử dụng phương pháp cập nhật tự động và thủ công dựa trên việc kế thừa các nguồn dữ liệu có sẵn).

- Nghiên cứu khai thác kho ngữ vựng.

- Xây dựng ứng dụng web.

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

Về mặt ý nghĩa khoa học, đề tài là cơ sở tiền đề để phục vụ cho các bài toán xử lý ngôn ngữ tự nhiên (dịch, từ điển, phần mềm học tập tiếng Hrê...), về ý nghĩa thực tiễn thì kết quả của đề tài là kho ngữ vựng và từ điển Hrê - Việt - Hrê để phục vụ cho công tác dạy và học tiếng Hrê đồng thời phục vụ cho công tác truyền thông của đài phát thanh truyền hình tỉnh, đài truyền thanh các huyện miền núi và các đơn vị chức năng trong công tác tuyên truyền chủ trương của Đảng, chính sách của Nhà nước đến đồng bào cũng như truyền đạt những thông tin có ích về trồng trọt, chăn nuôi, sức khỏe, giáo dục cho đồng bào Hrê từ đó nâng cao đời sống vật chất lẫn tinh thần cho đồng bào. Qua đó, góp phần tăng cường khối đại đoàn kết dân tộc.

## **6. Bố cục luận văn**

Luận văn được tổ chức thành 3 chương:

### **Chương 1: Tìm hiểu tiếng Hrê**

Nêu thực trạng tình hình sử dụng tiếng Hrê hiện nay, đồng thời trình bày một số nội dung cơ bản của tiếng Hrê như: từ ngữ âm, nguyên âm, phụ âm, các phương án phiên âm.

### **Chương 2: Cơ sở lý thuyết**

Trong chương này, sẽ nêu ra cơ sở lý thuyết dùng để xây dựng ứng dụng; tìm hiểu về từ điển; nêu các công cụ và kỹ thuật cập nhật kho dữ liệu.

### **Chương 3: Xây dựng Từ điển Hrê – Việt – Hrê**

Trong chương cuối này sẽ nêu giải pháp cập nhật cơ sở dữ liệu; phân tích thiết kế hệ thống và cuối cùng là triển khai thí nghiệm đánh giá kết quả chương trình.

## CHƯƠNG 1: TÌM HIỂU TIẾNG HRÊ

### 1.1. Tình hình sử dụng tiếng Hrê

#### 1.1.1. *Vài nét về tiếng Hrê*

Ngôn ngữ của người Hrê thuộc ngữ hệ (họ) Nam Á. Đây là một ngữ hệ có phạm vi tồn tại rộng lớn ở Đông Nam Á, từ Indonesia đến một số khu vực thuộc vùng phía nam Trung Quốc.

Từ trong tiếng Hrê ngày xưa có cấu tạo đa âm tiết (polysyllable) nhưng ngày nay đã rơi rụng dần và gần như một ngôn ngữ đơn âm (nói từng tiếng rời). Một số tiền tố, hậu tố trong từ chuyển thành dấu hiệu căng chùng khi phát âm như đặc trưng thanh hầu, họng ở đầu và đặc trưng căng cao ở cuối như trong 'mau (lúa), hnoiq (nói).

Hiện nay, tiếng Hrê được phát trên sóng Đài phát thanh truyền hình Quảng Ngãi và các đài truyền thanh huyện miền núi. Văn bản tiếng Hrê dạng viết vẫn là phương án tạm thời. Hiện nay có rất nhiều tài liệu biên soạn chữ viết về tiếng Hrê nhưng tất cả vẫn còn sơ khai, chưa thống nhất và chưa được Nhà nước công nhận chính thức, vì thế gây nhiều khó khăn trong việc dạy và học tiếng Hrê.

Chữ viết Hrê đã có là loại chữ ghi âm tự dạng la – tinh. Hệ thống chữ đầu tiên được các nhà khoa học thuộc Viện Ngữ học mùa hè (SIL) xây dựng vào khoảng những năm 70 của thế kỷ 20 (trước giải phóng miền Nam). Nó đã được dùng để ghi tiếng Hrê ở các tỉnh Quảng Ngãi và Bình Định, sử dụng trong một số sách dạy và học tiếng Hrê, nhưng chưa thực sự phổ biến trong cộng đồng Hrê.

Sau giải phóng, tập thể các ông Đinh Văn Bay, Đinh Xuân Trâm và Đinh Văn Lâm (trí thức của dân tộc Hrê) đã soạn thảo ra một bộ chữ khác. Hệ thống chữ này cho đến nay rất ít người được biết. Căn cứ trên bộ chữ của các ông Đinh Văn Bay, Đinh Xuân Trâm..., ông Đinh Văn

Thành – trí thức Hrê, công tác tại Sở Giáo dục Bình Định – đã hiệu chỉnh và chế tác một hệ thống chữ để ghi tiếng Hrê. Trên cơ sở chữ này, ông Đinh Văn Thành đã biên soạn một số tài liệu phục vụ cho dạy và học tiếng Hrê.

### **1.1.2. Hệ thống ngữ âm**

#### **1.1.2.1. Từ ngữ âm**

Tiếng Hrê có hai dạng từ ngữ âm: từ đơn tiết và từ đa tiết

Từ ngữ âm đơn tiết: chỉ gồm một âm (một tiếng), ví dụ:

aw (tôi, tao)

hla (lá)

am (đi)

hnim (nhà)

maw (lúa)

khe (trăng)...

Từ ngữ đa âm tiết: gồm một (hoặc hai) âm tiết đứng trước (gọi là “tiền âm tiết” và âm tiết đứng sau được phát âm nhấn mạnh hơn (gọi là “âm tiết chính”, ví dụ

mangai (người)

kani (chuột)

tanih (đất)

tamui (khách)

alah (lười)

pahaceh (xẻ thành tấm), ...

#### **1.1.2.2. Hệ thống phụ âm**

Hệ thống phụ âm tiếng Hrê bao gồm các phụ âm đơn và các phụ âm kép (còn gọi là “tổ hợp phụ âm”). Chúng ở vị trí phần đầu và phần cuối của âm tiết



❖ Các phụ âm đơn

**Bảng 1.1. Các phụ âm đơn**

| Vị trí cấu âm<br>Phương thức phát âm |           | môi         | đầu lưỡi | mặt lưỡi | gốc lưỡi | hầu |
|--------------------------------------|-----------|-------------|----------|----------|----------|-----|
|                                      |           | T<br>Ă<br>C | vô thanh | p        | t        | c   |
| vô thanh bật hơi                     | (ph)      |             | (th)     |          | (kh)     |     |
| hữu thanh                            | (b)       |             | (d)      |          |          |     |
| hữu thanh thở                        | (bh)      |             | (dh)     | (jh)     | (gh)     |     |
| mũi                                  | m         |             | n        |          | ɳ        |     |
| KHÔNG<br>TẮC                         | vô thanh  |             | (s)      |          |          | h   |
|                                      | hữu thanh | w           |          | j        |          |     |
|                                      | bên       |             | l        | r        |          |     |
|                                      | rung      |             |          |          |          |     |

1.1.2.3. Hệ thống nguyên âm

Hệ thống nguyên âm Hrê gồm các loại sau:

Xét về số lượng các yếu tố cấu thành, có thể phân biệt nguyên âm đơn (chỉ gồm một yếu tố, ví dụ: i, ɛ, a, ...) với nguyên âm đôi (gồm hai yếu tố, ví dụ: ua,...)

Xét về cách phát âm tạo nên những “giọng” khác nhau, có thể phân biệt nguyên âm căng (hay còn gọi là “cứng”, được phát âm với “giọng” cao và trong, ví dụ: i, ea,...) với nguyên âm chùng (hay còn gọi là “mềm”, được phát âm với “giọng” trầm đục, có tiếng thở, ví dụ: ì, ε`a,...)

❖ *Hệ thống các nguyên âm đơn (căng và chùng) được trình bày qua bảng sau:*

**Bảng 1.9. Nguyên âm đơn**

| Dòng<br>Độ nâng | Trước |       | Sau            |       |          |         |
|-----------------|-------|-------|----------------|-------|----------|---------|
|                 | Căng  | Chùng | Không tròn môi |       | Tròn môi |         |
|                 |       |       | Căng           | Chùng | Căng     | Chùng   |
| Hẹp             | i     | ì     |                |       | u        | ù       |
| Trung bình      | e     |       |                |       | o        |         |
| Rộng            | ε`    | ε`    | a              | à     | ɔ        | ɔ`<br>, |

❖ *Hệ thống các nguyên âm đôi (căng và chùng) như sau:*

**Bảng 1.10. Các nguyên âm đôi**

|       |    |    |     |     |
|-------|----|----|-----|-----|
| Căng  | ia | ua | εa  | ɔa  |
| Chùng | ìa | ùa | ε`a | ɔ`a |

## 1.2. Phương án đề nghị phiên âm tiếng Hrê

### 1.2.1. *Những yêu cầu đối với chữ Hrê*

### 1.2.2. *Phương án đề nghị phiên âm tiếng Hrê*

#### 1.2.2.1. *Các chữ cái và dấu*

Các chữ cái (viết hoa và viết thường):

A a, B b, C c, D d, Đ đ, E e, Ê ê, G g, H h, I i, J j, K k, L l, M m, ‘M’ m, Nn, ‘N’ n, O o, Ô ô, P p, Q q, R r, ‘R’ r, S s, T t, U u, V v, W w, ‘W’ w, Y y, Y’ y.

Các dấu:

Dấu “ ’ ” (dùng để ghi yếu tố? trong phụ âm kép ở đầu âm tiết) và dấu “ ` ” (ghi tính chất “chùng” của nguyên âm): Dấu “ ~ ” và “ ” được dùng để ghi “giọng mũi” trong một số từ ngữ (ít gặp).

Ngoài ra, trong chính tả Hrê còn có các dấu:., !,, ?... (như chính tả các ngôn ngữ khác).

#### 1.2.2.2. *Cách ghi từ ngữ âm*

❖ *Đối với từ ngữ âm đơn tiết:*

Ghi bằng các chữ cái ghép lại từ trái sang phải

❖ *Đối với từ ngữ âm đa tiết:*

Từng âm tiết được ghi như đối với từ ngữ âm đơn tiết. Các âm tiết (tiền âm tiết và âm tiết chính) được ghi kế tiếp liền nhau (không tách rời; tiền âm tiết đứng trước âm tiết chính). Khi âm tiết chính mở đầu bằng phụ âm đơn ?, thì âm này được ghi bằng dấu “-”.

1.2.2.3. *Cách ghi các phụ âm*

❖ *Đối với các phụ âm đơn:*

Ghi bằng các chữ cái riêng lẻ hoặc ghép lại

❖ *Đối với các phụ âm kép (hai yếu tố và ba yếu tố):*

Ghi bằng cách ghép các chữ cái để thể hiện các yếu tố, sát liền nhau, theo thứ tự từ trái sang phải, thể hiện ở bảng sau:

1.2.2.4. *Cách ghi các nguyên âm*

❖ *Đối với các nguyên âm đơn:*

Ghi bằng một chữ cái; các nguyên âm đơn cũng được ghi bằng chữ cái không có dấu, các nguyên âm chùng được ghi bằng chữ có dấu “ ` ”

❖ *Đối với các nguyên âm đôi:*

Ghi bằng hai chữ cái kế tiếp nhau; các nguyên âm đôi cũng được ghi bằng hai chữ cái không dấu, các nguyên âm đôi chùng được ghi bằng hai chữ cái có dấu “ ” trên chữ cái thứ nhất

### **1.3. Ý nghĩa của việc xây dựng kho ngữ vựng Hrê–Việt–Hrê**

Việc xây dựng một kho ngữ vựng Hrê có tính mở, dễ khai thác, cập nhật và ứng dụng xây dựng từ điển Hrê – Việt và Việt – Hrê giúp cho việc xây dựng các ứng dụng phục vụ cho các mục đích học tập, nghiên cứu chữ viết Hrê. Qua đó, giúp cho việc xây dựng một bộ chữ viết Hrê thống nhất được Nhà nước công nhận chính thức, từ đó phục vụ cho công tác tuyên truyền các chủ trương của Đảng, chính sách của Nhà nước đến đồng bào dân tộc Hrê được tốt hơn.

## CHƯƠNG 2: CƠ SỞ LÝ THUYẾT

### 2.1. Cơ sở công nghệ

#### 2.1.1. Công nghệ dotNET

2.1.1.1. Giới thiệu Microsoft.NET

2.1.1.2. ASP.net

2.1.1.3. Ngôn ngữ lập trình C#.Net

2.1.1.4. Visual Studio.Net

2.1.1.5. Visual C# 2008

#### 2.1.2. Mô hình ADO.NET

Thành phần của ADO.NET được thiết kế nhằm tăng tốc độ truy cập và thao tác dữ liệu trong môi trường đa lớp, gồm 2 thành phần chính là đối tượng DataSet và .NET Provider.

.NET Provider là một tập các đối tượng trong thành phần .NET bao gồm Connection (Kết nối cơ sở dữ liệu), Command (Lệnh thi hành), DataReader (Bộ đọc dữ liệu) và DataAdapter (Kết nối dữ liệu). Khi sử dụng DataSet chúng ta có thể kết nối nguồn dữ liệu, thực hiện truy vấn một hay nhiều bảng, sau đó yêu cầu hệ thống đóng kết nối.

Provider được xem như một cầu nối giữa ứng dụng với cơ sở dữ liệu, chúng dùng để kết nối nguồn dữ liệu, thực hiện câu lệnh và nhận dữ liệu trả về. Những dữ liệu này có thể được xử lý trực tiếp, hay lưu trữ trên đối tượng DataSet của ADO.NET.

ADO.NET bao gồm hai Provider như sau:

- ❖ *SQL Server.NET data provider*: Sử dụng cho SQL Server 7.0 và các phiên bản sau.

- ❖ *OLE DB.NET data provider*: Sử dụng cho dữ liệu kết xuất thông qua OLE DB.

### **2.1.3. Hệ quản trị cơ sở dữ liệu SQL Server 2005**

SQL Server 2005 là một hệ thống quản lý CSDL (Relational Database Management System - RDBMS) sử dụng Transact – SQL để trao đổi dữ liệu giữa máy Client và SQL Server. Một RDBMS bao gồm Databases, database engine và các ứng dụng dùng để quản lý dữ liệu và các bộ phận khác nhau trong RDBMS.

## **2.2. Tìm hiểu từ điển**

### **2.2.1. Khái niệm từ điển**

#### **2.2.2. Một số loại từ điển**

2.2.2.1. Từ điển giấy

2.2.2.2. Từ điển máy tính

2.2.2.3. Từ điển Internet

### **2.2.3. Nhận xét**

## **2.3. Công cụ và kỹ thuật cập nhật kho ngữ vựng**

### **2.3.1. Cơ sở dữ liệu từ vựng đa ngữ**

2.3.1.1. Tổng quan về CSDL

2.3.1.2. CSDL từ vựng đa ngữ

2.3.1.3. Những vấn đề cần xử lý khi xây dựng CSDL

### **2.3.2. Các công cụ trợ giúp xây dựng CSDL**

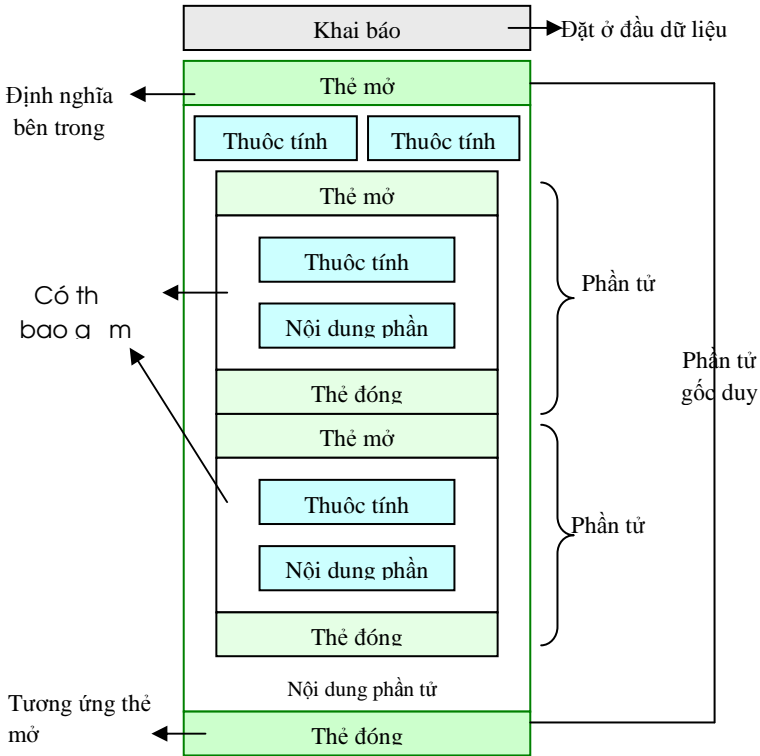
2.3.2.1. Microsoft Access

2.3.2.2. Microsoft SQL Server

2.3.2.3. Microsoft Word

2.3.2.4. XML (eXtensible Markup Language)

Một trang XML cần phải theo đúng các quy luật sau:



**Hình 2.4. Cấu trúc của tài liệu XML**

### 2.3.3. Các kỹ thuật cập nhật dữ liệu

2.3.3.1. Công cụ cập nhật tài liệu bằng RTF của Microsoft Word

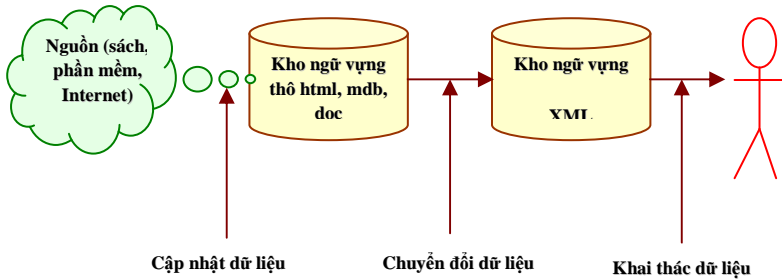
2.3.3.2. Kỹ thuật cập nhật sử dụng các macro

## CHƯƠNG 3: XÂY DỰNG TỪ ĐIỂN HRÊ-VIỆT-HRÊ

### 3.1. Ý tưởng xây dựng

### 3.2. Hoàn thiện kho ngữ vựng Hrê – Việt – Hrê

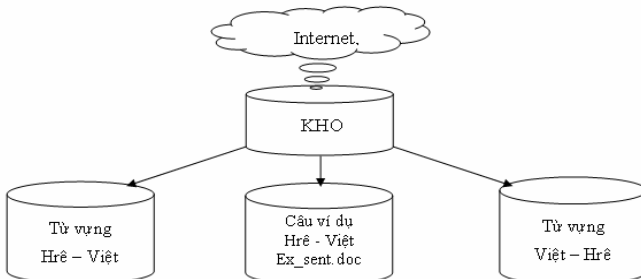
#### 3.2.1. Tổng quan về quá trình xây dựng kho ngữ vựng



Hình 3.1. Mô hình tổng quát xây dựng kho ngữ liệu

#### 3.2.2. Xây dựng cấu trúc kho ngữ vựng

##### 3.2.2.1. Cấu trúc kho dữ liệu thô



Hình 3.2. Cấu trúc kho dữ liệu thô



Nguồn CSDL Hrê – Việt trong tệp HV.doc gồm các thành phần:

Mục từ tiếng Hrê

Từ loại

Nghĩa tiếng Việt tương ứng

Nguồn CSDL Việt – Hrê trong tệp VH.doc gồm các thành phần:

Mục từ tiếng Việt

Từ loại

Nghĩa tiếng Hrê tương ứng

Nguồn CSDL câu ví dụ trong tệp Ex\_sent.doc gồm các thành phần:

Câu ví dụ tiếng Hrê

Câu ví dụ tiếng Việt tương ứng

3.2.2.2. *Cấu trúc kho ngữ vựng Hrê – Việt – Hrê dưới dạng XML*

Kho ngữ vựng Hrê – Việt – Hrê được xây dựng dưới dạng 2 tệp tin XML là HV.xml và VH.xml có chung cấu trúc như sau [5]:

<Dictionary>

<Word>

<Entry\_name>[từ tiếng Hrê]</Entry\_name>

<Type\_word>[từ loại]</Type\_word>

<Word\_similar>[từ            có            nghĩa            tương  
ứng]</Word\_similar>

<Viet\_equa>[nghĩa tiếng Việt tương ứng]</Viet\_equa>

<Sentence>

<Hre\_sent>[Câu ví dụ tiếng Hrê]</Hre\_sent>

<Viet\_sent>[Câu ví dụ tiếng Việt]</Viet\_sent>

</Sentence>

</Word>

</Dictionary>

### **3.2.3. Chuyển đổi dữ liệu thô sang dữ liệu XML**

- Sử dụng phần mềm HTMLtoRTF Converter Pro để chuyển đổi các trang HTML thành các trang văn bản Word.

- Sử dụng đoạn mã Macro VBA Winword để hợp nhất dữ liệu, tạo thành một tệp \*.rtf hoàn chỉnh.

- Từ tệp \*.rtf, sử dụng macro chuyển đổi để xuất dữ liệu cơ sở dữ liệu Access \*.mdb. Từ đây, sử dụng công cụ Export để chuyển sang tệp cơ sở dữ liệu \*.xml.

### **3.2.4. Giải pháp cập nhập dữ liệu**

#### **3.2.4.1. Cập nhật tự động**

Trước hết chuyển tất cả các nguồn dữ liệu thô từ internet hay các tập tin dữ liệu trên máy tính về dạng Microsoft Word và tổng hợp thành hai tập tin là HV.doc chứa từ vựng Hrê – Việt và VH.doc chứa từ vựng Việt – Hrê. Tùy vào định dạng của nguồn dữ liệu mà sử dụng các công cụ cập nhật khác nhau để chuyển đổi sang dạng Word.

#### **3.2.4.2. Cập nhập thủ công**

Nhập trực tiếp dữ liệu từ nguồn sách, từ điển giấy, dữ liệu do các chuyên gia về ngôn ngữ cung cấp.

Đối với CSDL kho ngữ vựng Hrê – Việt – Hrê thì hầu như là phải nhập bằng phương pháp thủ công, vì tài liệu về tiếng Hrê trên Internet và trên máy tính là hầu như chưa có gì

### **3.2.5. Thống kê kho dữ liệu**

Đã xây dựng được kho ngữ vựng 1500 từ và 200 câu ví dụ

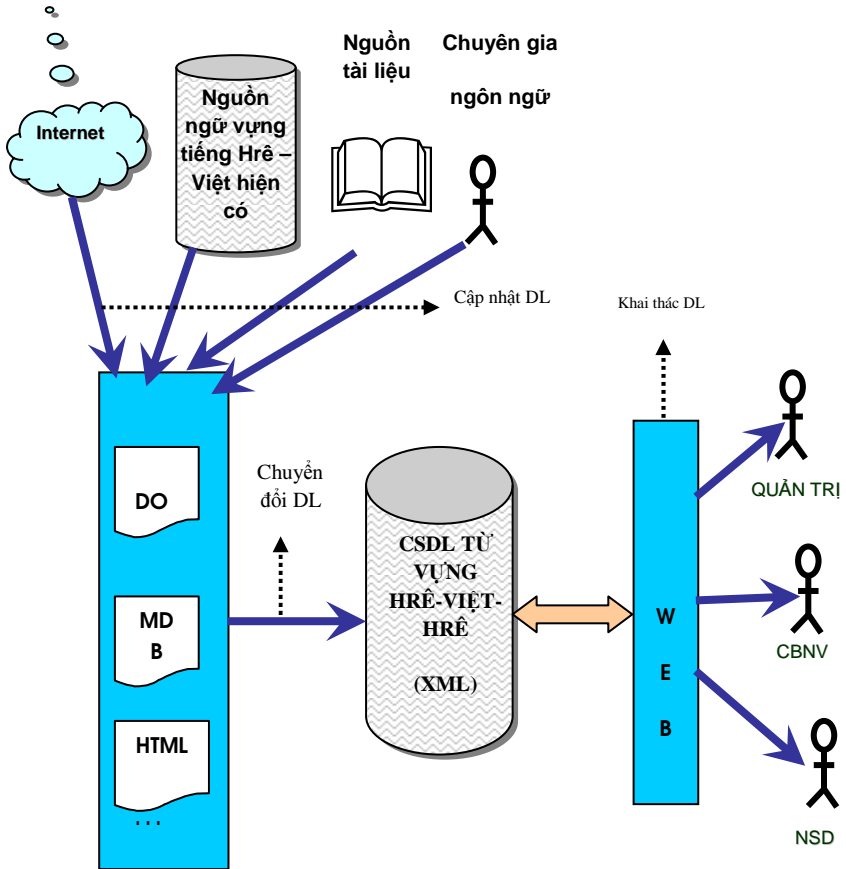
## **3.3. Phân tích thiết kế hệ thống**

### **3.3.1. Mô tả hệ thống**

#### **3.3.1.1. Yêu cầu đối với hệ thống**

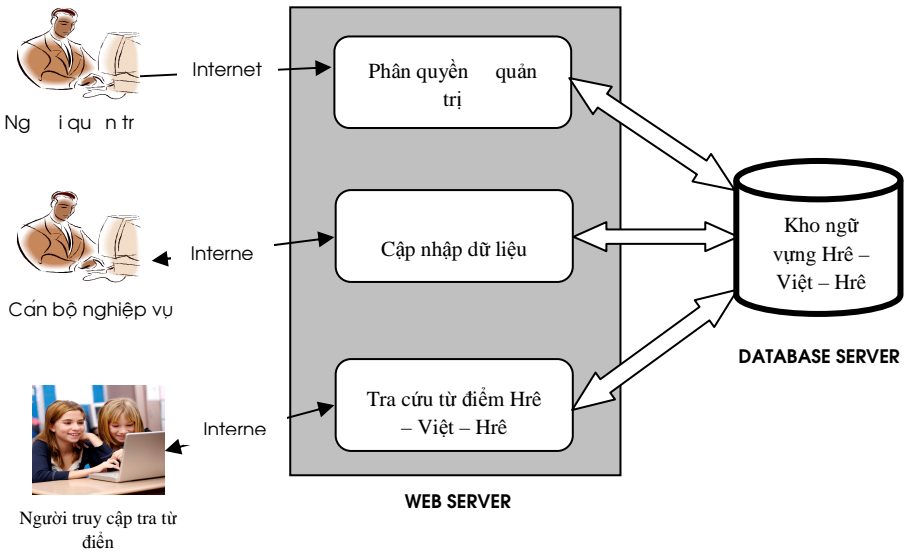
#### **3.3.1.2. Kiến trúc tổng thể**

Kiến trúc tổng thể của hệ thống bao gồm những thành phần sau:



*Hình 3.3. Kiến trúc tổng thể hệ thống*

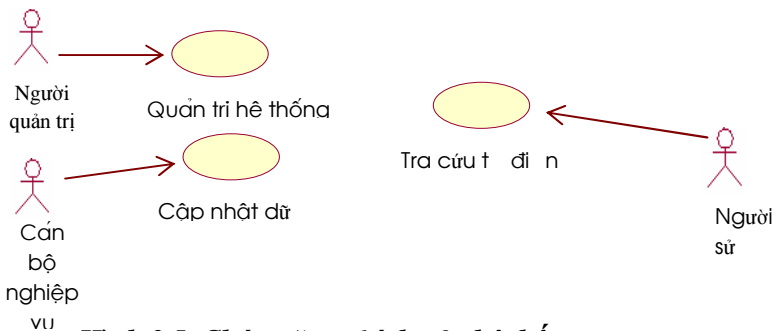
### 3.3.1.3. Mô hình hoạt động



**Hình 3.4. Mô hình hoạt động của hệ thống**

### 3.3.2. Thiết kế chức năng

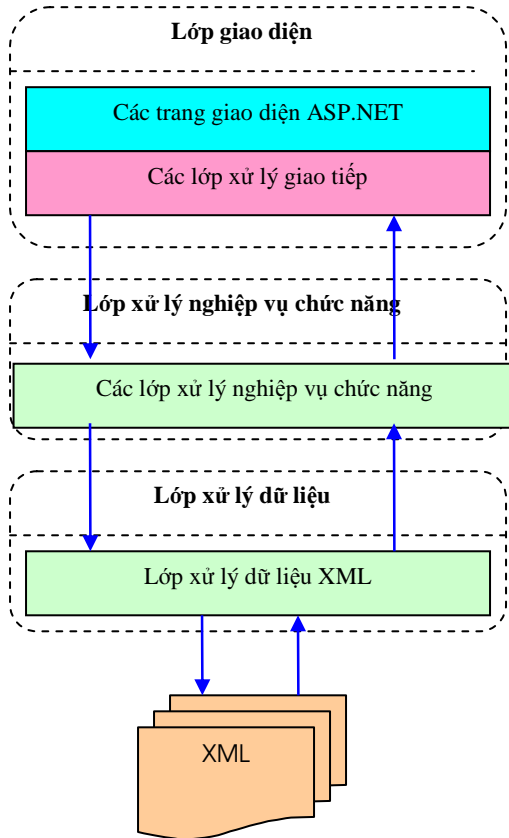
Hệ thống cần có những chức năng cơ bản sau:



**Hình 3.5. Chức năng chính của hệ thống**

### 3.4. Triển khai thí nghiệm và đánh giá kết quả

#### 3.4.1. Mô hình cài đặt



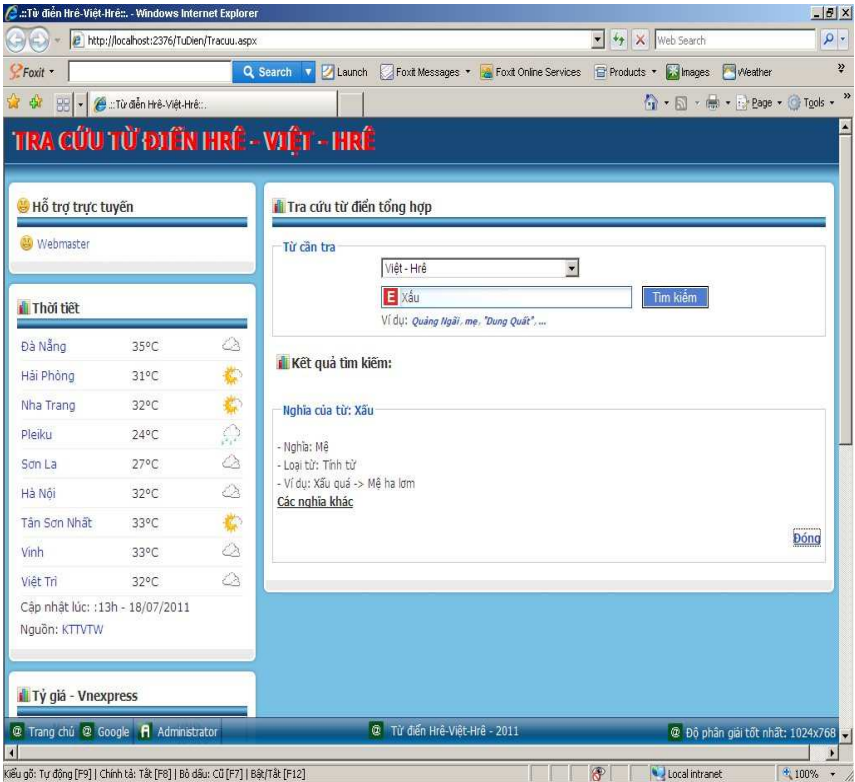
**Hình 3.6. Mô hình cài đặt**

3.4.1.1. Thành phần lớp giao diện

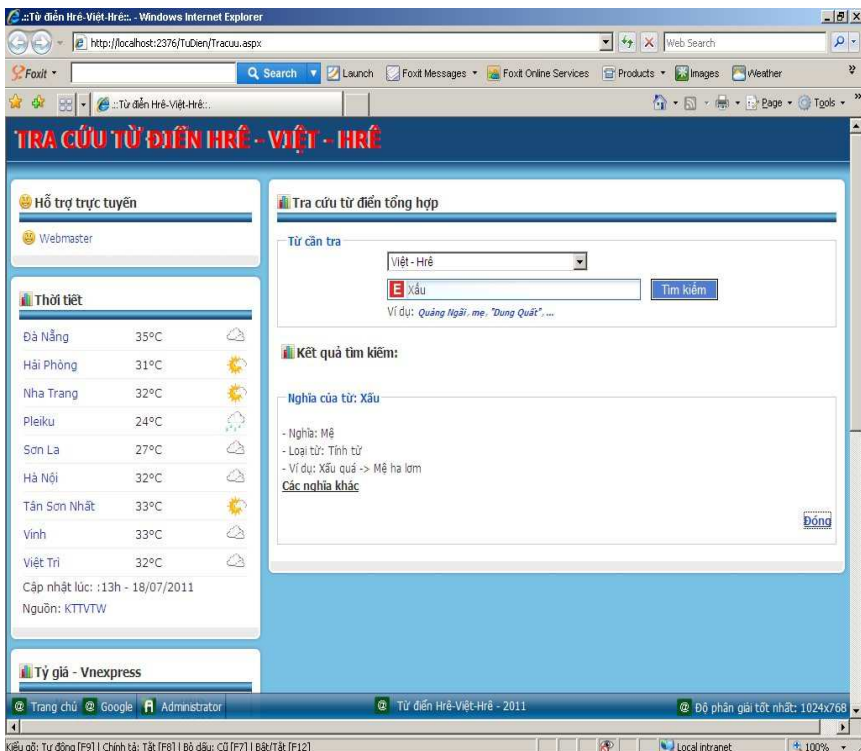
3.4.1.2. Thành phần lớp xử lý các nghiệp vụ chức năng

3.4.1.3. Thành phần lớp xử lý dữ liệu

### 3.4.2. Một số giao diện của chương trình



Hình 3.7. Màn hình giao diện tra cứu từ điển Hrê – Việt



**Hình 3.8. Màn hình giao diện tra cứu từ điển Việt – Hrê**

### **3.4.3. Kịch bản sử dụng**

#### **3.4.3.1. Người quản trị:**

- Khởi động chương trình
- Đăng nhập quyền admin
- Phân quyền người sử dụng
- Cập nhật dữ liệu vào kho ngữ vựng

#### **3.4.3.2. Cán bộ nghiệp vụ:**

- Đăng nhập hệ thống
- Cập nhật dữ liệu vào kho ngữ vựng

### 3.4.3.3. Người truy cập tra từ điển:

Đối với người dùng tra từ điển Hrê – Việt – Hrê cần phải có các trình duyệt Web như Internet Explorer, Firefox, Google Chrome...; phần mềm hỗ trợ gõ các ngôn ngữ như Vietkey2000 hay Unikey; máy tính của người dùng phải kết nối đến máy chủ Localhost chứa CSDL và chương trình từ điển). Các bước tra từ điển như sau:

- Khởi động chương trình bằng cách truy cập vào trang <http://localhost:2376/tudien/tracuu.aspx>.

- Chọn từ điển cần tra (Việt – Hrê, Hrê – Việt)

- Nhập hoặc tìm từ cần tra, xem kết quả

#### **3.4.4. Đánh giá kết quả:**

- Chương trình cho kết quả chính xác đối với các từ vựng có sẵn trong CSDL. Những từ không tìm thấy, chương trình thông báo không tìm thấy.

- Khi có lỗi hay sự cố, chương trình luôn thông báo thông tin cho người sử dụng về lỗi hay sự cố và đưa ra hướng dẫn cách khắc phục hoặc gửi thông báo lỗi đến admin.

- Chương trình cho phép sử dụng đồng thời 50 người dùng cuối.



## KẾT LUẬN

### 1. Kết quả đạt được của đề tài

Với mục tiêu là hoàn thiện hệ thống ngữ vựng tiếng Hrê và ứng dụng xây dựng từ điển Hrê – Việt – Hrê, Luận văn đã làm được những việc như: đã nêu được tình hình sử dụng tiếng Hrê hiện nay; giới thiệu những nét chính về tiếng Hrê, hệ thống ngữ âm, phương án đề nghị phiên âm tiếng Hrê.

Về mặt lý thuyết luận văn đã tìm hiểu về cơ sở công nghệ để xây dựng ứng dụng: công nghệ dotNET, mô hình ADO.NET, hệ quản trị cơ sở dữ liệu SQL Server. Tìm hiểu về cơ sở dữ liệu từ vựng đa ngữ, các công cụ trợ giúp xây dựng CSDL, các kỹ thuật cập nhật dữ liệu. Luận văn đã đi sâu tìm hiểu về XML và đã chọn XML để xây dựng CSDL kho ngữ vựng.

Luận văn đã xây dựng được kho ngữ vựng Hrê – Việt với 1.500 từ, từ đó đã ứng dụng xây dựng từ điển Hrê – Việt – Hrê chạy trên môi trường website. Điều này cho phép người sử dụng có thể tra cứu từ điển Hrê – Việt – Hrê một cách nhanh chóng, thuận tiện, phục vụ cho nhu cầu học tập, nghiên cứu tiếng Hrê.

Tuy nhiên, do thời gian và trình độ của tác giả còn hạn chế nên luận văn chỉ mới dừng lại ở đó mà chưa làm được một số việc còn tồn tại như: chưa đưa phần phiên âm vào từ điển; chưa xây dựng được những câu tiếng Hrê thông dụng trong đời sống xã hội vào kho ngữ vựng.

### 2. Hướng phát triển đề tài

Tuy kết quả đạt được của đề tài còn khiêm tốn nhưng đó là cơ sở tiền đề để phát triển, mở rộng đề tài theo hướng:

- Tiếp tục hoàn thiện cơ sở dữ liệu ngữ vựng Hrê – Việt bằng cách nghiên cứu bổ sung đầy đủ phần phiên âm tiếng Hrê, các câu tiếng Hrê thông dụng trong đời sống xã hội.

- Tiếp tục phát triển cơ sở dữ liệu, đặc biệt là các câu ví dụ, thành ngữ tiếng Hrê tạo điều kiện thuận lợi cho việc xây dựng ứng dụng dịch tự động.

- Phát triển kho ngữ vựng Hrê – Việt thành kho ngữ vựng Hrê – Việt – Anh, ứng dụng xây dựng từ điển Hrê – Việt – Anh; Hrê – Việt – Pháp.

- Kế thừa kho ngữ vựng xây dựng phần mềm học tiếng Hrê.