


<p style="writing-mode: vertical-rl; transform: rotate(180deg);">NGUYỄN ĐỨC THỌ</p>	<p style="text-align: center;"><b>ĐẠI HỌC THÁI NGUYÊN</b></p> <p style="text-align: center;"><b>TRƯỜNG ĐẠI HỌC KỸ THUẬT CÔNG NGHIỆP</b></p> 
<p style="writing-mode: vertical-rl; transform: rotate(180deg);">LUẬN VĂN THẠC SĨ KỸ THUẬT</p> <p style="writing-mode: vertical-rl; transform: rotate(180deg);">KỸ THUẬT ĐIỆN TỬ</p> <p style="writing-mode: vertical-rl; transform: rotate(180deg);">N 2013</p>	<p style="text-align: center;"><b>LUẬN VĂN THẠC SĨ KỸ THUẬT</b></p> <p style="text-align: center;">NGÀNH: KỸ THUẬT ĐIỆN TỬ</p> <p style="text-align: center;"><b>NGHIÊN CỨU MỘT SỐ PHƯƠNG PHÁP TỔNG HỢP TIẾNG NÓI</b></p> <p style="text-align: center;">Ngành: KỸ THUẬT ĐIỆN TỬ  Học Viên: NGUYỄN ĐỨC THỌ  Người HD Khoa học: PGS.TS. LƯƠNG CHI MAI</p> <p style="text-align: center;">THÁI NGUYÊN 2013</p>



## LUẬN VĂN THẠC SĨ

**Họ và tên học viên** : Nguyễn Đức Thọ  
**Ngày tháng năm sinh** : Ngày 22 tháng 03 năm 1980  
**Nơi sinh** : Chi Lăng - Quế võ - Tỉnh Bắc Ninh  
**Nơi công tác** : Trường Cao đẳng nghề Cơ điện và Xây dựng Bắc Ninh  
**Cơ sở đào tạo** : Trường Đại học Kỹ thuật Công nghiệp Thái Nguyên  
**Chuyên ngành** : Kỹ thuật điện tử  
**Khóa học** : K13- KTĐT  
**Ngày giao đề tài** : Ngày.....Tháng.....Năm.....  
**Ngày hoàn thành đề tài** : Ngày.....Tháng.....Năm.....

### TÊN ĐỀ TÀI:

NGHIÊN CỨU MỘT SỐ PHƯƠNG PHÁP TỔNG HỢP TIẾNG NÓI

**Người hướng dẫn khoa học:** PGS.TS. *Lương Chi Mai*  
Viện công nghệ thông tin, viện khoa học và công nghệ Việt Nam

**GIÁO VIÊN HƯỚNG DẪN**

**HỌC VIÊN**

PGS.TS. *Lương Chi Mai*  
**BAN GIÁM HIỆU**

**Nguyễn Đức Thọ**  
**KHOA SAU ĐẠI HỌC**

**ĐẠI HỌC THÁI NGUYÊN  
TRƯỜNG ĐẠI HỌC KỸ THUẬT CÔNG NGHIỆP**

**LUẬN VĂN THẠC SĨ**

**Tên đề tài: NGHIÊN CỨU MỘT SỐ PHƯƠNG PHÁP TỔNG HỢP TIẾNG NÓI**

**Người hướng dẫn: PGS .TS Lương Chi Mai**

**Học viên: Nguyễn Đức Thọ**

**Chuyên ngành: Kỹ thuật điện tử**

**Thái Nguyên 2013**

## LỜI CAM ĐOAN

Tên tôi là: **Nguyễn Đức Thọ**

Học viên lớp Cao học khoá 13- Kỹ thuật điện tử - Trường Đại học Kỹ thuật Công nghiệp Thái Nguyên

Tôi xin cam đoan luận văn là kết quả nghiên cứu của riêng tôi, các số liệu, kết quả nêu trong luận văn này là trung thực và không sao chép của ai, luận văn này không giống hoàn toàn bất cứ luận văn hoặc các công trình đã có trước đó.

Thái Nguyên, Ngày ..... tháng ..... năm

2013

*Nguyễn Đức Thọ*

## *Lời cảm ơn*

Trước hết, em xin được gửi lời biết ơn sâu sắc tới các thầy cô giáo trong trường Đại học kỹ thuật Công nghiệp Thái Nguyên, các thầy cô giáo trực tiếp giảng dạy, tận tình hướng dẫn em những kiến thức và kinh nghiệm quý báu trong suốt thời gian học tập và rèn luyện.

Em xin bày tỏ lời cảm ơn chân thành tới các thầy cô giáo trong Khoa Điện điện tử Trường Đại học kỹ thuật Công nghiệp Thái Nguyên, khoa sau Đại học Trường Đại học kỹ thuật Công nghiệp Thái Nguyên đã tạo điều kiện thuận lợi cho em trong thời gian học tập.

Em xin bày tỏ lòng biết ơn chân thành, lời cảm ơn sâu sắc đối với cô giáo **PGS.TS. Lương Chi Mai** đã trực tiếp hướng dẫn, định hướng cho em giải quyết nhiều vấn đề trong luận văn.

Em cũng muốn gửi lời cảm ơn tới tập thể lớp thạc sĩ kỹ thuật điện tử k13 đã tạo một môi trường thi đua học tập lành mạnh, tạo điều kiện cho sự phát triển của các thành viên trong lớp.

Cuối cùng em xin được cảm ơn những người thân, bạn bè đã quan tâm, động viên giúp đỡ em trong quá trình học tập, nghiên cứu và hoàn thành luận văn.

Tác giả luận văn

*Nguyễn Đức Thọ*

## MỤC LỤC

Lời cam đoan.....	i
Lời cảm ơn .....	ii
Mục lục.....	1
Danh mục các chữ viết tắt.....	5
Danh mục bảng các thuật ngữ Anh – Việt .....	5
Danh mục hình vẽ và đồ thị.....	6
Danh mục các bảng biểu.....	7

<b>NỘI DUNG</b>	Trang
<b>LỜI MỞ ĐẦU</b> .....	8
<b>Chương 1. Tổng quan về tổng hợp tiếng nói</b> .....	12
1.1 Giới thiệu về tổng hợp tiếng nói.....	12
1.1.1 khái niệm về tổng hợp tiếng nói.....	12
1.1.2 Mô hình của một bộ tổng hợp tiếng nói.....	12
1.1.2.1. Thành phần xử lý ngôn ngữ tự nhiên.....	15
1.1.2.2. Thành phần xử tín hiệu số.....	18
1.2. Lịch sử phát triển và ứng dụng tổng hợp tiếng nói.....	20
<b>Chương 2. Các phương pháp tổng hợp tiếng nói</b> .....	25
2.1 Phương pháp tổng hợp theo cấu âm.....	25
2.2 Phương pháp tổng hợp theo formant.....	28
2.3 Phương pháp Tổng hợp ghép nối.....	30
2.4 Phương pháp tổng hợp theo ghép chuỗi.....	31
2.5 Đánh giá chung các phương pháp tổng hợp tiếng nói.....	34
<b>Chương 3. Chuẩn hoá văn bản và thuật toán phân tích văn bản</b> .....	38

3.1 Tổng quan.....	38
3.2.Các vấn đề đối với bài toán Chuẩn hóa văn bản tiếng Việt.....	38
3.2.1.Bài toán chuẩn hóa văn bản tiếng Anh.....	38
3.2.2 Đặc điểm văn bản tiếng Việt.....	39
3.2.3.Sự tổ chức không có quy chuẩn của một token.....	39
3.2.3.1 Sự phức tạp của NSW.....	40
3.2.3.2 Xử lý các NSW khác nhau.....	41
3.2.3.3 Vấn đề phân loại NSW.....	41
3.2.3.4.Vấn đề mở rộng NSW.....	41
3.3.Giải pháp đề xuất.....	42
3.3.1 Phân loại các từ chưa chuẩn hóa cho tiếng Việt.....	42
3.3.2 Tổ chức gán thẻ cho NSW.....	47



3.3.3.Chuẩn phân tách.....	49
3.3.4.Phát hiện NSW.....	50
3.3.5.Phân loại NSW.....	50
3.3.6.Mở rộng cách đọc.....	52
3.4.Các kỹ thuật và giải thuật .....	52
3.4.1.Tổng quan về cây quyết định.....	53
3.4.2. Các bước xây dựng cây quyết định.....	54
3.4.2.1.Cắt tỉa cây.....	55
3.4.2.2.Độ đo thuộc tính.....	55
3.4.2.3.Rút luật phân lớp từ cây quyết định.....	56
3.4.2.4.Hiện tượng “quá vừa” (overfitting) dữ liệu huấn luyện.....	56
3.4.2.5.Xử lý quá vừa dữ liệu.....	57
3.4.2.6.Áp dụng cây quyết định.....	58
3.4.3.Mô hình ngôn ngữ chữ cái (Letter Language Modeling).....	60
3.4.4.Kỹ thuật làm trơn mô hình Kneser-Ney cải tiến.....	60
3.5.Kết chương.....	61
<b>Chương 4. Cài đặt thử nghiệm modul chuẩn hóa văn bản cho tiếng Việt</b>	<b>62</b>
4.1. Cài đặt và thử nghiệm.....	62
4.1.1.Cài đặt.....	62
4.1.2.Thử nghiệm chương trình.....	69

4.1.2.1. Phương pháp đánh giá kết quả.....	69
4.1.2.2. Bộ dữ liệu thử nghiệm Training.....	70
4.2. Kết quả thử nghiệm.....	72
<b>KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN.....</b>	<b>74</b>
<b>TÀI LIỆU THAM KHẢO.....</b>	<b>75</b>