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

MA NGỌC KHÁNH

NGHIÊN CỨU PHƯƠNG PHÁP RÚT GỌN VĂN BẢN VÀ CHUYỂN ĐỔI CÚ PHÁP NGÔN NGỮ KÝ HIỆU VIỆT NAM

Chuyên ngành: Khoa học máy tính
Mã số: 8480101

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

Người hướng dẫn khoa học: PGS.TS. PHÙNG TRUNG NGHĨA

THÁI NGUYÊN, 2018

MỤC LỤC

DANH MỤC HÌNH

DANH MỤC BẢNG BIỂU

LỜI CAM ĐOAN

Tôi là: **Ma Ngọc Khánh**

Lớp: CK15

Khoá học: 2016 - 2018

Chuyên ngành: Khoa học máy tính

Mã số chuyên ngành: 8480101

Cơ sở đào tạo: Trường Đại học Công nghệ thông tin và Truyền thông Thái Nguyên.

Giáo viên hướng dẫn: **PGS.TS. Phùng Trung Nghĩa**

Tôi xin cam đoan luận văn ***"Nghiên cứu phương pháp rút gọn văn bản và chuyển đổi cú pháp ngôn ngữ ký hiệu Việt Nam"*** này là công trình nghiên cứu của riêng tôi dưới sự hướng dẫn của PGS.TS. Phùng Trung Nghĩa. Các số liệu sử dụng trong luận văn là trung thực. Các kết quả nghiên cứu được trình bày trong luận văn chưa từng được công bố tại bất kỳ công trình nào khác.

Thái Nguyên, ngày 30 tháng 5 năm 2018

HỌC VIÊN

Ma Ngọc Khánh

LỜI CẢM ƠN

Học viên xin gửi lời cảm ơn chân thành tới Thầy hướng dẫn PGS.TS. Phùng Trung Nghĩa, Trường Đại học Công nghệ Thông tin và Truyền thông - Đại học Thái Nguyên, người đã tận tình hướng dẫn giúp học viên hoàn thành luận văn tốt nghiệp.

Học viên cũng xin gửi lời cảm ơn sâu sắc đến các thầy cô giáo của Trường Đại học Công nghệ Thông tin và Truyền thông - Đại học Thái Nguyên, cùng các thầy cô giáo của Viện Công nghệ thông tin - Viện khoa học Việt Nam đã nhiệt tình giảng dạy, truyền đạt kiến thức cho học viên trong suốt 2 năm học để học viên có thể hoàn thành được luận văn của mình.

Ma Ngọc Khánh

MỞ ĐẦU

Hiện nay theo thống kê, Việt Nam có khoảng trên 2.5 triệu người khiếm thính [1]. Do khả năng nghe bị suy giảm nên khả năng giao tiếp bằng lời ở cộng đồng người khiếm thính bị hạn chế rất nhiều. Để thay thế cho khả năng giao tiếp bằng tiếng nói, ngôn ngữ ký hiệu, là ngôn ngữ tượng hình sử dụng biểu diễn, chuyển động của bàn tay, cơ thể, và sắc thái biểu cảm của khuôn mặt đã ra đời Tuy nhiên, việc sử dụng ngôn ngữ ký hiệu chưa phát huy được hiệu quả giúp người khiếm thính hòa nhập được với xã hội do việc giao tiếp giữa người khiếm thính và người nghe tốt còn gặp nhiều khó khăn.

Trên thế giới hiện nay đã và đang nghiên cứu phát triển và đưa ra nhiều dịch vụ thông dịch và sản phẩm công nghệ nhằm hỗ trợ người khiếm thính trong giao tiếp xã hội như máy trợ thính dành cho người nghe kém, găng tay chuyển đổi ngôn ngữ ký hiệu thành giọng nói [9], các phần mềm dịch từ văn bản/giọng nói sang ngôn ngữ ký hiệu hay các từ điển tra cứu ngôn ngữ ký hiệu online [12], v.v… Tuy nhiên mỗi một nghiên cứu hay sản phẩm đều có những hạn chế và chưa đáp ứng được việc hỗ trợ trong giao tiếp hai chiều giữa người khiếm thính và người nghe tốt trong thực tế.

Việc nghiên cứu xử lý ngôn ngữ ký hiệu trên máy tính ở nước ta còn rất mới mẻ. Chúng ta chưa thực sự có một hệ thống ngôn ngữ đồng nhất cho ngôn ngữ ký hiệu tiếng Việt [6]. Bên cạnh vấn đề ngôn ngữ học, việc phát triển sản phẩm ứng dụng công nghệ để phát huy ngôn ngữ ký hiệu nhằm nâng cao trình độ, tiếp nhận thông tin, khả năng giao tiếp cho người khiếm thính lại càng ít và kém hiệu quả.

Với sự quan tâm đặc biệt của Đảng và Nhà nước, đã có nhiều trường học, trung tâm hỗ trợ dạy học và việc làm riêng cho người khiếm thính. Vì vậy việc nghiên cứu về các thuật toán và xây dựng phần mềm rút gọn văn

bản, chuyển đổi cú pháp đối với ngôn ngữ ký hiệu Việt Nam là cần thiết [2]. Do đó tôi chọn đề tài ***"Nghiên cứu phương pháp rút gọn văn bản và chuyển đổi cú pháp ngôn ngữ ký hiệu Việt Nam"***

Mục tiêu của luân văn là Nghiên cứu các lý thuyết đã có để phân tích, đánh giá về các tính chất, các luật rút gọn, chuyển đổi cú pháp đối với ngôn ngữ ký hiệu Việt Nam. Dựa trên các cơ sở lý thuyết và các phân tích, đánh giá sẽ nghiên cứu cài đặt thực nghiệm các thuật toán rút gọn, chuyển đổi cú pháp này và xây dựng phần mềm hỗ trợ rút gọn và chuyển đổi cú pháp ngôn ngữ ký hiệu Việt Nam.

Nội dung chính của luận văn bao gồm 3 chương:

Chương 1. Tổng quan về ngôn ngữ ký hiệu Việt Nam

Chương 2. Luật rút gọn văn bản và chuyển đổi cú pháp đối với ngôn ngữ ký hiệu Việt Nam.

Chương 3. Xây dựng hệ thống rút gọn văn bản và chuyển đổi cú pháp ngôn ngữ ký hiệu Việt Nam.

Khi viết báo cáo này học viên đã cố gắng để đạt được những mục tiêu và định hướng nghiên cứu đề ra ban đầu, song điều kiện thời gian và năng lực còn hạn chế nên không tránh khỏi thiếu sót. Học viên mong nhận được sự góp ý của thầy giáo hướng dẫn, thầy cô giáo để học viên có được những kinh nghiệm thực tế và bổ ích để sau này có thể xây dựng được một chương trình hoàn thiện hơn.

CHƯƠNG 1
TỔNG QUAN VỀ NGÔN NGỮ KÝ HIỆU VIỆT NAM

1.1. Tổng quan về ngôn ngữ ký hiệu

1.1.1. Khái niệm ngôn ngữ ký hiệu

Ngôn ngữ ký hiệu (hay ngôn ngữ dấu hiệu, thủ ngữ) là ngôn ngữ dùng những biểu hiện của bàn tay thay cho âm thanh của tiếng nói. Ngôn ngữ ký hiệu do người điếc tạo ra nhằm giúp họ có thể giao tiếp với nhau trong cộng đồng của mình và tiếp thu tri thức của xã hội. Việc thay thế âm thanh của tiếng nói có thể liên quan đến đồng thời sự kết hợp các hình dạng tay, hướng và chuyển động của bàn tay, cánh tay hoặc cơ thể, và nét mặt để thể hiện trôi chảy những suy nghĩ của người nói. Ngôn ngữ kí hiệu có nhiều điểm tương đồng với ngôn ngữ nói (đôi khi được gọi là "ngôn ngữ bằng miệng" - mà phụ thuộc chủ yếu vào âm thanh), đó là lý do tại sao ngôn ngữ học xem xét cả hai dạng ngôn ngữ là ngôn ngữ tự nhiên. Tuy nhiên cũng có một số khác biệt đáng kể giữa các ngôn ngữ ký hiệu và ngôn ngữ nói. Đặc biệt không nên nhầm lẫn ngôn ngữ kí hiệu với ngôn ngữ cơ thể, là một loại giao tiếp phi ngôn ngữ.

Bất cứ đâu trong cộng đồng người khiếm thính trên thế giới, ngôn ngữ ký hiệu đều được phát triển. Ngôn ngữ kí hiệu không chỉ được sử dụng bởi người điếc mà nó cũng được sử dụng bởi những người có thể nghe thấy, nhưng thể chất bị hạn chế để có thể nói chuyện bình thường. Ngôn ngữ kí hiệu có những thuộc tính ngôn ngữ riêng biệt. Hiện nay, hàng trăm ngôn ngữ ký hiệu được sử dụng trên thế giới và phát triển trong cộng đồng người khiếm thính ở tất cả các quốc gia. Một số ngôn ngữ ký hiệu có được công nhận pháp lý, trong khi một số khác thì chỉ mang tính cục bộ, địa phương.

Một quan niệm sai lầm phổ biến là tất cả các ngôn ngữ ký hiệu là trên toàn thế giới là hoàn toàn giống nhau hoặc ngôn ngữ ký hiệu là một ngôn ngữ