

KẾT LUẬN

Xử lý ngôn ngữ tự nhiên bao giờ cũng rất phức tạp và còn nhiều vấn đề mở cần tiếp tục nghiên cứu. Riêng về xử lý tiếng Việt thì phức tạp hơn vì cú pháp tiếng Việt tồn tại hiện tượng nhập nhằng (nhập nhằng cú pháp), vai trò của từ và cụm từ trong câu phụ thuộc vào ngữ cảnh rất nhiều. Vì thế, việc xác định cho được các từ và cụm từ trong câu để trợ giúp bắt lỗi chính tả không phải là đơn giản và hầu như không thể bắt lỗi với tỉ lệ 100%. Qua quá trình thực hiện, khoá luận đã đạt được một số kết quả như sau:

- Thực hiện nghiên cứu và cài đặt minh hoạ bắt lỗi chính tả tiếng Việt cả ở hai mức: mức âm tiết và mức cú pháp. Khẳng định khả năng có thể bắt lỗi lên đến 95% tỷ lệ bắt lỗi. Đồng thời, khẳng định phương pháp bắt lỗi dựa trên phân tích cú pháp là triệt để hơn.
- Thực hiện nghiên cứu ngữ pháp tiếng Việt để hỗ trợ cho phân tích cú pháp.
- Thực hiện nghiên cứu giải thuật Earley và cải tiến giải thuật cho phù hợp hơn với thực tế phân tích cú pháp. Giúp cho giải thuật Earley hoàn thiện hơn với xử lý ngôn ngữ tự nhiên.
- Từ điển được thiết kế nhỏ gọn và tối ưu trong tìm kiếm. Từ điển đã giải quyết được vấn đề lưu trữ Unicode.
- Xây dựng thuật toán phân tích từ mới hiệu quả hơn, nhưng vẫn còn hạn chế ở giá trị ngưỡng là 3.
- Xây dựng chương trình bắt lỗi chính tả tiếng Việt với phương pháp gợi ý khá hoàn chỉnh. Xây dựng một hệ thống các chương trình hỗ trợ cho chương trình bắt lỗi chính tả.

Bắt lỗi chính tả tiếng Việt là một vấn đề phức tạp. Khoá luận cũng chỉ mới là bước đầu đi vào giải quyết vấn đề mà thôi, cho nên còn nhiều hạn chế và chưa thể đáp ứng nhu cầu của người sử dụng. Trong thời gian tới, chương trình sẽ được phát triển theo hướng:

- Xây dựng chương trình phân tích văn bản ra các thành phần như: đoạn, câu, âm tiết, từ, kí hiệu, số... Chương trình có khả năng độc lập nền.
- Tối ưu hoá từ điển và tiếp tục cải tiến giải thuật Earley.
- Cải tiến nâng cao độ thông minh của phương pháp tạo gợi ý sửa lỗi.
- Tiếp tục nghiên cứu ngữ pháp tiếng Việt, đề ra các tiêu chuẩn cho phân chia từ loại, hoàn thiện hệ thống luật sinh cho câu tiếng Việt.
- Xây dựng hệ thống chương trình tiện ích hỗ trợ quá trình nghiên cứu. Phát triển khả năng học của chương trình.
- Nghiên cứu khả năng bắt lỗi mức ngữ nghĩa.

- Tích hợp vào các hệ thống soạn thảo văn bản hiện có trên thị trường.

Trong tương lai xa hơn, có thể sẽ phát triển:

- Hệ thống dịch tự động.
- Xây dựng các chương trình ứng dụng có khả năng hiểu dữ liệu nhập vào.

Như trong quản lý thư viện, người sử dụng có thể nhập một từ nào đó liên quan đến sách cần tham khảo thì chương trình sẽ tự động liệt kê tất cả các sách cần thiết. Hay người sử dụng có thể nhập từ sai nhưng chương trình vẫn có thể hiểu được người sử dụng muốn gì.