



BÁO CÁO ĐỒ ÁN

Đề tài: NGHIÊN CỨU BÀI TOÁN KIỂM TRA LỖI VÀ XÂY DỰNG ỨNG DỤNG SỬA LỖI TRONG BÀI VIẾT TIẾNG ANH



Môn: Phương pháp luận nghiên cứu khoa học – CS519.L11

GVHD: PGS.TS. Lê Đình Duy

Thành viên nhóm:

Nguyễn Đỗ Tường Phát	18520329
Lê Huỳnh Đức	17520357
Lường Thị Thu Nguyệt	18520414
Trần Ngọc Yến Nhi	18520328

17/1/2021

Nội dung

01 Giới thiệu

02 Mục tiêu

03 Nội dung và phương pháp thực hiện

04 Kết quả dự kiến





Giới thiệu



Giới thiệu

Bài toán: Nghiên cứu và cải thiện các bài toán, phương pháp kiểm tra lỗi và sửa lỗi bài viết tiếng Anh, đặc biệt là phần sửa lỗi ngữ pháp.

Input: Đoạn văn bản lấy từ các nguồn khác nhau (người dùng nhập, đoạn văn có sẵn).

Output: Đoạn văn bản đó, nhưng những từ hoặc đoạn sai ngữ pháp sẽ được đánh dấu và có gợi ý để thay thế từ/ngữ pháp.





Giới thiệu

Khả năng ứng dụng:

- Phát hiện lỗi sai theo thứ tự ưu tiên là: chính tả, ngữ pháp, ngữ nghĩa.
- Áp dụng cho:
 - + Người đang học tiếng Anh
 - + Những người biên soạn tài liệu tiếng Anh cần kiểm tra lại ngữ pháp,...





Mục tiêu

Mục tiêu



- Nghiên cứu và cải thiện những mô hình, thuật toán, các phương pháp máy học liên quan đến xử lý ngôn ngữ tự nhiên mà cụ thể hơn là sửa lỗi ngữ pháp.
- Xây dựng một hệ thống kiểm tra, sửa lỗi bài viết tiếng Anh dựa trên những phương pháp đã được cải thiện thành công.



Nội dung và phương pháp thực hiện

Đối tượng nghiên cứu

Các mô hình máy học

Mô hình ngôn ngữ
(Language model)

Mô hình n grams

Mô hình thống kê
(Statistical model)

Các thuật toán, phương pháp máy học

Hồi quy tuyến tính
(Linear Regression)

Cây quyết định
(Decision tree)

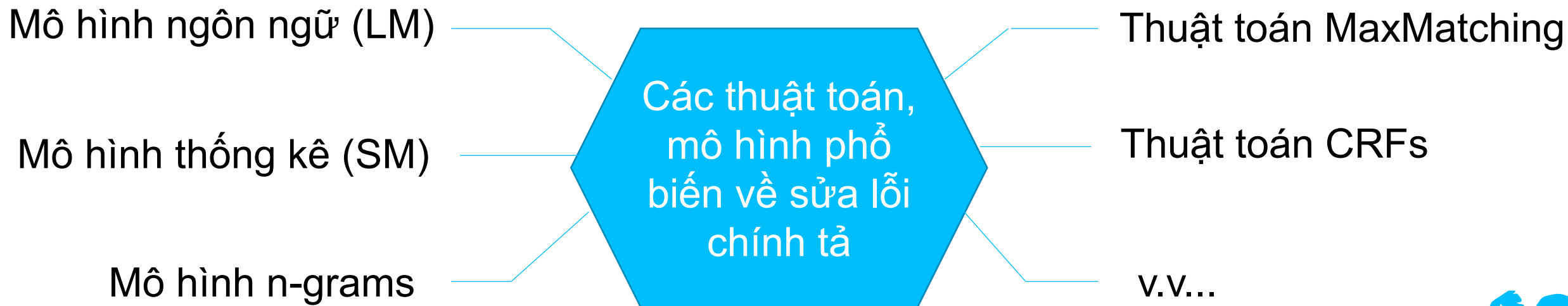
Mạng thần kinh
(Neural Networks)

Mạng Bayes
(Bayesian Network)



Phương pháp

- o Nghiên cứu về các thuật toán, mô hình phổ biến nhất hiện nay về sửa lỗi chính tả.



Phương pháp

- o Tiến hành cài đặt thử nghiệm các thuật toán, các mô hình và tiến hành nghiệm thu, đánh giá kết quả.






Phương pháp

- o Cải tiến các phương pháp bằng cách cài đặt thêm/ bớt tính năng (feature), xoay chuyển các bộ dữ liệu để tiến hành so sánh và đánh giá kết quả với các phiên bản cũ.
- o Chọn ra phương pháp tối ưu nhất và tiến hành xây dựng hệ thống phát hiện lỗi chính tả.



Kết quả dự kiến



Kết quả dự kiến

- Xây dựng thành công một hệ thống kiểm và sửa lỗi văn bản. Sửa lỗi chính tả (chính xác 80%), sửa lỗi ngữ pháp (chính xác 80%), ngữ nghĩa (chính xác 40%).
- Cải thiện được những thiếu sót từ các dự án tham khảo.
- Xây dựng được ứng dụng kiểm tra và sửa lỗi trong bài viết tiếng Anh.



Thank You



Any question