**Tên đề tài luận văn: Khai thác mô hình ngôn ngữ lớn nâng cao hiệu năng đo lường độ tương tự ngữ nghĩa giữa văn bản tiếng Việt**

Xây dựng đề cương nghiên cứu

Chương 1. Tổng quan về bài toán STS

* **Giới thiệu, Motivation**
  + Định nghĩa và ý nghĩa của Semantic Textual Similarity.
  + Ứng dụng của STS trong các lĩnh vực như tìm kiếm thông tin, dịch máy, tổng hợp văn bản, và phân loại văn bản.
* **Lịch sử và phát triển của STS (Các nghiên cứu liên quan)**
  + Sự phát triển của STS từ những năm đầu và các bước ngoặt quan trọng.
  + Các phương pháp cổ điển và hiện đại.
* **Các thách thức trong STS**
  + Sự mơ hồ ngữ nghĩa và sự phụ thuộc ngữ cảnh.
  + Khó tạo nên được bộ dữ liệu ngữ nghĩa chuẩn.
  + Đánh giá độ tương tự ngữ nghĩa.
* **Mục tiêu và phạm vi của nghiên cứu**
  + Mục tiêu chính của nghiên cứu.
  + Giới hạn và phạm vi nghiên cứu.

Chương 2. STS theo tiếp cận Độ đo (ghi chú tiếng anh)

Đưa ra một độ đo khoảng cách tuân thủ theo 3 tiên đề: đối xứng, phản xạ, bắc cầu

Gợi ý: Đo thông qua 1 xâu chung dài nhất có thể (normalize), biểu đồ ven, đọ tương quan Spearman.

* **Định nghĩa Measure trong STS**
  + Khái niệm và vai trò của cách tiếp cận measure
* **Các phương pháp measure được nghiên cứu**
  + Longest common substring
  + Trọng số được tính toán từ vị trí trong câu
* **Phân tích chi tiết các triển khai và đánh giá các phương pháp prediction**
  + Ưu và nhược điểm của từng phương pháp.
  + Các tiêu chí đánh giá hiệu quả của prediction, tổng quan kết quả.

Chương 3. STS theo tiếp cận prediction

* **Định nghĩa và phân loại prediction**
  + Khái niệm và vai trò của prediction trong STS.
  + Phân loại prediction: supervised, unsupervised, semi-supervised.
* **Các phương pháp prediction được nghiên cứu**
  + MLP, RNN
* **Phân tích chi tiết các triển khai và đánh giá các phương pháp prediction**
  + Ưu và nhược điểm của từng phương pháp.
  + Các tiêu chí đánh giá hiệu quả của prediction, tổng quan kết quả.

Chương 4. Thực nghiệm và phân tích.

* **Thiết kế thực nghiệm**
  + Mục tiêu thực nghiệm.
  + Dữ liệu thực nghiệm: nguồn gốc, đặc điểm và xử lý dữ liệu, quy trình chuẩn hóa.
  + Các bước tiến hành thực nghiệm.
* **Thực hiện các phương pháp measure**
  + Triển khai và đánh giá các phương pháp measure.
  + So sánh kết quả giữa các phương pháp cũ.
* **Thực hiện các phương pháp prediction**
  + Triển khai và đánh giá các phương pháp prediction.
  + So sánh kết quả giữa các phương pháp cũ.
* **Phân tích kết quả**
  + Đánh giá tổng thể kết quả thực nghiệm.
  + So sánh giữa measure và prediction.
  + Các yếu tố ảnh hưởng đến hiệu quả của STS.

Chương 5: Kết luận

* **Tóm tắt kết quả nghiên cứu**
  + Tóm tắt các phát hiện chính từ nghiên cứu.
* **Đóng góp của nghiên cứu**
  + Những đóng góp lý thuyết và thực tiễn.
* **Hạn chế và hướng nghiên cứu tương lai**
  + Các hạn chế của nghiên cứu hiện tại.
  + Đề xuất các hướng nghiên cứu tiếp theo.

Xây dựng bộ dữ liệu và thực nghiệm với một số tiếp cận cơ bản

Nghiên cứu về đánh giá mô hình ngôn ngữ lớn