

# THÔNG TIN LUẬN ÁN

Tên luận án:	<b>Tóm tắt đoạn văn bản tiếng Việt dựa trên cách tiếp cận tạo sinh</b>
Chuyên ngành:	Khoa học máy tính
Mã số:	62.48.01.01
Họ tên NCS:	<b>Trần Trung</b>
Hướng dẫn khoa học:	<b>PGS. TS. Nguyễn Tuấn Đăng, PGS. TS. Phạm Hữu Đức</b>
Cơ sở đào tạo:	Trường Đại học Công nghệ Thông tin, ĐHQG-HCM

---

## 1. TÓM TẮT:

Mục tiêu của luận án là đề xuất giải pháp tạo sinh đoạn văn bản tóm tắt nhằm tóm tắt nội dung thông tin của đoạn văn bản tiếng Việt cho trước. Mô hình tóm tắt văn bản được đề xuất trong Luận án gồm hai thành phần:

- Thành phần “Phân tích và Biểu diễn ngữ nghĩa văn bản tiếng Việt”. Thành phần này nhận đầu vào là đoạn văn bản tiếng Việt. Đầu ra của thành phần này là một cấu trúc biểu diễn ngữ nghĩa trừu tượng của văn bản nguồn.
- Thành phần “Tạo sinh văn bản tiếng Việt”. Đầu vào của thành phần này là một cấu trúc biểu diễn ngữ nghĩa trừu tượng. Đầu ra của thành phần này, cũng là kết quả cuối cùng của mô hình, là một đoạn văn bản tóm tắt hoàn chỉnh.

Luận án tập trung nghiên cứu mô hình tóm tắt cho các đoạn văn bản tiếng Việt ngắn, trong đó có thể xuất hiện các yếu tố hỏi chỉ.

## 2. NHỮNG ĐÓNG GÓP CHÍNH CỦA LUẬN ÁN:

### ○ Những đóng góp khoa học:

- Đề xuất mô hình giải pháp tạo sinh đoạn văn bản tóm tắt.

- Đề xuất những phương pháp tạo dựng cấu trúc biểu diễn ngữ nghĩa trừu tượng Discourse Representation Structure – DRS (theo Discourse Representation Theory).
  - Phương pháp sử dụng cấu trúc ngữ đoạn bề mặt gồm hai giai đoạn. Giai đoạn thứ nhất, luận án đề xuất phương pháp tạo dựng một cấu trúc biểu diễn cấp độ bề mặt được gọi là cấu trúc ngữ đoạn bề mặt. Cũng trong giai đoạn này, luận án đề xuất phương pháp dựa trên quy tắc để chuyển đổi về đoạn văn bản bao gồm các câu tiếng Việt có cấu trúc đơn giản. Giai đoạn thứ hai, luận án đề xuất phương pháp xác định mối liên hệ giữa từng yếu tố hồi chỉ với yếu tố tiền ngữ tương ứng đồng thời tạo dựng DRS của đoạn văn bản kết quả của Giai đoạn thứ nhất.
  - Phương pháp sử dụng cấu trúc đồ thị ngữ đoạn được gán nhãn gồm hai giai đoạn. Giai đoạn thứ nhất, luận án đề xuất phương pháp tạo dựng một cấu trúc biểu diễn cấp độ bề mặt được gọi là cấu trúc đồ thị ngữ đoạn được gán nhãn. Trong giai đoạn này, luận án đồng thời đề xuất phương pháp xác định mối liên hệ giữa từng yếu tố hồi chỉ với yếu tố tiền ngữ tương ứng. Giai đoạn thứ hai, luận án đề xuất phương pháp chuyển đổi mô hình biểu diễn trung gian thành DRS.
- Đề xuất những phương pháp xác định yếu tố tiền ngữ cho những dạng yếu tố hồi chỉ khác nhau trong đoạn văn bản tiếng Việt nguồn.
- Đề xuất những phương pháp tạo sinh câu và đoạn văn bản tóm tắt tiếng Việt dựa trên một dạng cấu trúc biểu diễn ngữ nghĩa trừu tượng.
- **Những đóng góp thực tiễn:**
  - Kết quả nghiên cứu của Luận án có thể ứng dụng vào việc xây dựng các hệ thống tóm tắt thông tin văn bản có chất lượng và hiệu quả cao.
  - Kết quả nghiên cứu của Luận án có thể ứng dụng trong những hướng nghiên cứu khác của chuyên ngành Xử lý ngôn ngữ tự nhiên như: phân tích cú pháp, khai phá văn bản, tạo sinh ngôn ngữ tự nhiên,...

- Kết quả nghiên cứu của Luận án có thể ứng dụng vào việc xây dựng các hệ thống tương tác giữa người và máy như: các hệ thống robot, trợ lý ảo,...

Các kết quả nghiên cứu và thực nghiệm của luận án được công bố trong 16 bài báo khoa học: 02 bài báo tạp chí quốc tế, 02 bài báo tạp chí trong nước, 08 bài báo ở hội nghị quốc tế chuyên ngành có phản biện độc lập mà các kỷ yếu hội nghị này được phát hành bởi các nhà xuất bản uy tín Springer, IEEE; 04 bài báo công bố trong kỷ yếu của các hội nghị chuyên ngành quốc gia.

**Cán bộ hướng dẫn 1**

**Cán bộ hướng dẫn 2**

**Nghiên cứu sinh**

**PGS. TS. Nguyễn Tuấn Đăng**

**PGS. TS. Phạm Hữu Đức**

**Trần Trung**