

# THÔNG TIN LUẬN ÁN

Tên luận án:

## NGHIÊN CỨU XÂY DỰNG MÔ HÌNH ĐỌC HIỂU TỰ ĐỘNG CHO VĂN BẢN TIẾNG VIỆT

Chuyên ngành: **Khoa học máy tính**

Mã số: **9480101**

Họ tên: **NGUYỄN VĂN KIỆT**

Cán bộ hướng dẫn: **PGS.TS. NGUYỄN LƯU THÙY NGÂN và TS. NGUYỄN GIA TUẤN ANH**

Cơ sở đào tạo: **TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN – ĐHQG TP.HCM**

### TÓM TẮT

Đọc hiểu tự động (Machine Reading Comprehension) là một lĩnh vực nghiên cứu được các nhà nghiên cứu quan tâm rất nhiều trong trí tuệ nhân tạo, đặc biệt là xử lý ngôn ngữ tự nhiên. Luận án tập trung vào xây dựng ngữ liệu và nghiên cứu các mô hình đọc hiểu, cùng với tích hợp mô hình đọc hiểu vào mô hình hỏi đáp trong văn bản tiếng Việt. Qua quá trình thực hiện các nghiên cứu trong luận án, nghiên cứu sinh (NCS) đã đóng góp ba nội dung chính:

- 1) Xây dựng ngữ liệu và đánh giá các mô hình đọc hiểu tự động trên ngữ liệu tiếng Việt: Trong đóng góp thứ nhất, NCS trọng tâm tập trung nghiên cứu, thiết kế và xây dựng các bộ ngữ liệu trên ngữ liệu tiếng Việt (Tiếng Việt là một ngôn ngữ có ít các bộ ngữ liệu cho phát triển và đánh giá các thuật toán học máy trong xử lý ngôn ngữ tự nhiên và trí tuệ nhân tạo). Các bộ ngữ liệu tiếng Việt trên miền mở Wikipedia và trên miền đóng tin tức sức khỏe được đề xuất nhằm để phát triển và đánh giá các mô hình MRC. Các mô hình MRC tiên tiến được đánh giá trên các bộ ngữ liệu đầu tiên trên văn bản tiếng Việt. Từ các kinh nghiệm nghiên cứu trong đọc hiểu và hỏi đáp tự động tiếng Việt, NCS đã đề xuất phát triển một bài toán đọc hiểu tự động tiếng Việt đầy thách thức tại VLSP 2021.

- 2) Đề xuất mô hình MRC tích hợp truy vấn minh chứng trên văn bản tiếng Việt: Trong đóng góp thứ hai, kế thừa từ các kết quả thử nghiệm đầu tiên đã đạt được trên các bộ ngữ liệu đã đề xuất trong đóng góp thứ nhất, luận án thiết kế và xây dựng các mô hình đọc hiểu tự động và hỏi đáp dựa trên các mô hình ngôn ngữ dựa trên kiến trúc transformer và truy xuất minh chứng. Thêm vào đó, NCS nghiên cứu và mở rộng đánh giá, so sánh với nhiều mô hình đọc hiểu tiên tiến trên thế giới để hiểu mô hình đọc hiểu đề xuất trên nhiều bộ ngữ liệu khác nhau.
- 3) Đề xuất mô hình hỏi đáp tích hợp mô hình đọc hiểu trên văn bản tiếng Việt: Trong đóng góp thứ ba, kế thừa từ các kết quả đầu tiên đã đạt được trên các bộ ngữ liệu đã đề xuất trong đóng góp thứ nhất và mô hình đọc hiểu tự động đề xuất tiếng Việt trong đóng góp thứ hai, luận án đề xuất, thiết kế và triển khai các mô hình hỏi đáp dựa trên mô hình ngôn ngữ và truy vấn minh chứng đạt hiệu quả cao trên văn bản tiếng Việt.

Trong thời gian hoàn thành luận án, NCS đã công bố bảy bài báo khoa học, trong đó: bốn bài báo đăng tại các tạp chí uy tín (ba tạp chí SCIE và một tạp chí quốc gia) và ba bài báo khoa học đăng tại các hội nghị quốc tế uy tín.

#### **CÁN BỘ HƯỚNG DẪN**

#### **NGHIÊN CỨU SINH**

NGUYỄN LƯU THÙY NGÂN    NGUYỄN GIA TUẤN ANH    NGUYỄN VĂN KIỆT