

ĐÁNH GIÁ KHẢ NĂNG GÁN NHÃN NGỮ NGHĨA TỰ ĐỘNG CỦA MÔ HÌNH NGÔN NGỮ LỚN CHO MỤC TỪ DANH TỪ TIẾNG VIỆT

Tóm tắt:

Mô hình ngôn ngữ lớn (LLM) ngày càng được ứng dụng trong nhiều lĩnh vực, đặc biệt trong xử lý ngôn ngữ tự nhiên. Nghiên cứu này khai thác LLM để gán nhãn ngữ nghĩa tự động trực tiếp cho các mục từ danh từ tiếng Việt dựa trên hệ thống WordNet. Hiệu quả gán nhãn được đánh giá thông qua bộ ngữ liệu vàng gồm 880 mục từ, được xây dựng từ ba nguồn: (1) Open English WordNet, (2) Từ điển OALD song ngữ Anh-Việt và (3) Từ điển tiếng Việt. Kết quả cho thấy LLM đạt độ chính xác 82,39% khi gán nhãn, qua đó khẳng định tiềm năng ứng dụng LLM trong việc xây dựng và mở rộng hệ thống nhãn ngữ nghĩa phù hợp với đặc trưng của tiếng Việt.

Từ khóa

Gán nhãn ngữ nghĩa; mô hình ngôn ngữ lớn; WordNet; danh từ tiếng Việt; bộ ngữ liệu vàng.

Thông tin các tác giả

- (1) Phan Thị Mỹ Trang: NCS., Trường Đại học Khoa học xã hội và Nhân văn, ĐHQG-HCM, số 10-12 Đinh Tiên Hoàng, Phường Sài Gòn, TP. HCM; ThS., Trường Đại học Công nghệ Sài Gòn, 180 Cao Lỗ, Phường Chánh Hưng, TP. HCM; số điện thoại: 090.305.7986; e-mail: mytrang779@gmail.com
- (2) Đỗ Quốc Trí: Sinh viên ngành Công nghệ thông tin, Trường Đại học Khoa học Tự nhiên, ĐHQG-HCM, 227 Nguyễn Văn Cừ, Phường Chợ Quán, TP. HCM; số điện thoại: 091.986.2313; e-mail: doquoctri2003@gmail.com
- (3) Phan Văn Bá Hải: Sinh viên ngành Công nghệ thông tin, Trường Đại học Khoa học Tự nhiên, ĐHQG-HCM, 227 Nguyễn Văn Cừ, Phường Chợ Quán, TP. HCM; số điện thoại: 076.776.7200; e-mail: phanvanba-hai@gmail.com
- (4) Đặng Ngọc Lê: PGS.TS., Trường Đại học Nguyễn Tất Thành, Phường Xóm Chiếu, TP. HCM; số điện thoại: 091.446.4660; e-mail: ngocle.dang1309@gmail.com
- (5) Đinh Điền: PGS. TS., Trường Đại học Khoa học Tự nhiên, ĐHQG-HCM, 227 Nguyễn Văn Cừ, Phường Chợ Quán, TP. HCM; số điện thoại: 090.827.8207; e-mail: ddiem@fit.hcmus.edu.vn
- (6) Trần Thị Minh Phượng: TS., Trường Đại học Khoa học xã hội và Nhân văn, ĐHQG-HCM, số 10-12 Đinh Tiên Hoàng, Phường Sài Gòn, TP. HCM; số điện thoại: 091.834.2789; e-mail: minhphuongtrn@hcmussh.edu.vn

Author: TRANG, Phan Thị Mỹ (Trường Đại học Khoa học xã hội và Nhân văn - ĐHQG TP.HCM; Trường Đại học Công nghệ Sài Gòn)

Co-authors: Mr TRÍ, Đỗ Quốc (Trường Đại học Khoa học Tự nhiên, ĐHQG-HCM); Mr HẢI, Phan Văn Bá (Trường Đại học Khoa học Tự nhiên - ĐHQG HCM); LÊ, Đặng Ngọc; ĐIỀN, Đinh (Trường Đại học Khoa học Tự nhiên, ĐHQG TP.HCM); PHƯỢNG, Trần Thị Minh (ĐHKHXHN, ĐHQG HCM)

Track Classification: Tiểu ban 1: Những tiến bộ và thành tựu mới trong lĩnh vực Ngôn ngữ học Tính toán