

**ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**



**ỨNG DỤNG XỬ LÝ NGÔN NGỮ TỰ NHIÊN
CHO CÂU TRUY VẤN LIÊN QUAN TỚI KIẾN
THỨC LUẬT VỀ BẢO HIỂM THẤT NGHIỆP**

Sinh viên thực hiện: Nguyễn Trùng Dương		
STT	Họ tên	MSSV
1	Nguyễn Trùng Dương	18520030
2		

TP. HỒ CHÍ MINH – 5/2022

1. GIỚI THIỆU

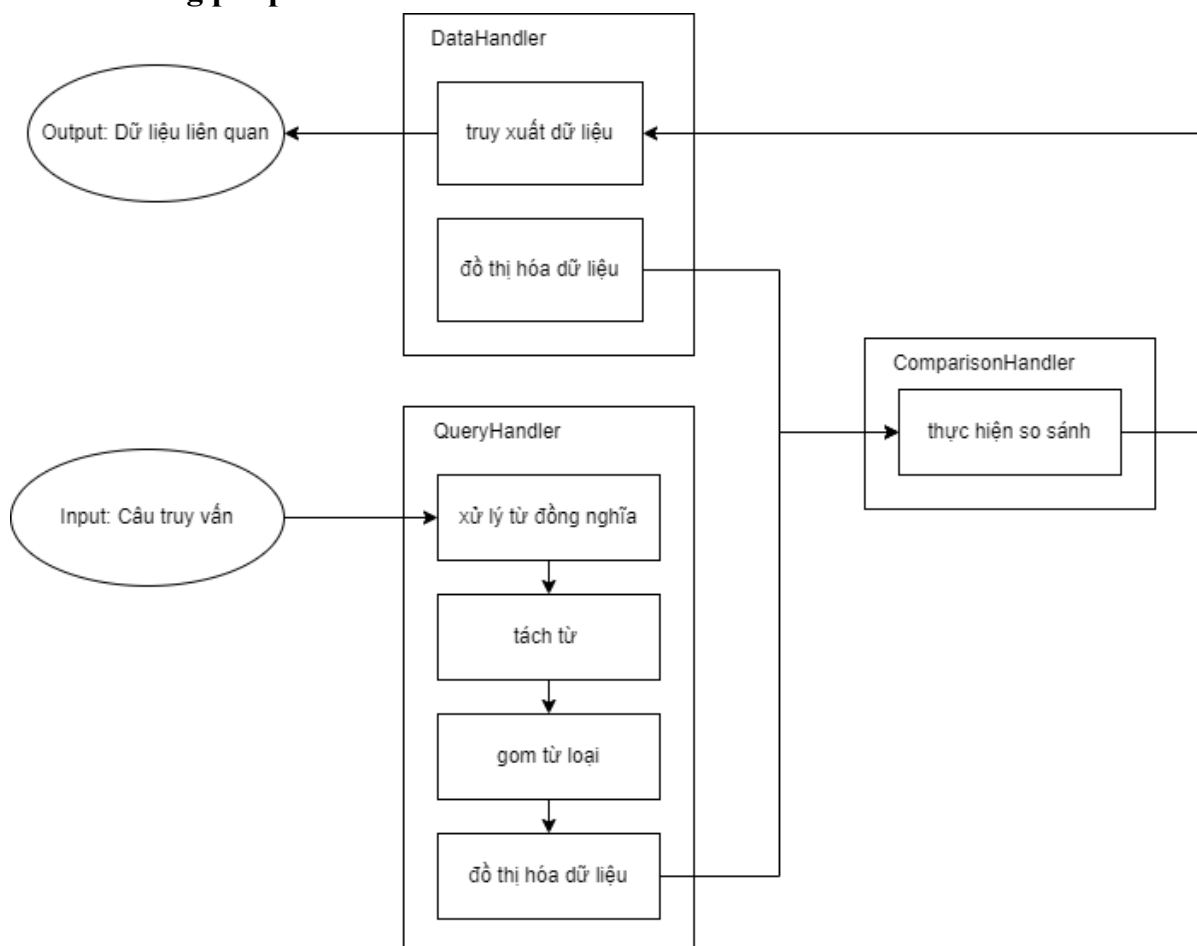
Đề tài thực hiện xử lý câu truy vấn trên ngôn ngữ tự nhiên từ đó trích xuất ra các từ khóa liên quan tới kiến thức luật về bảo hiểm y tế, dựa vào các từ khóa đó sẽ tiến hành tạo các đồ thị khái niệm để đánh giá độ tương đồng về khái niệm mà người dùng muốn tìm kiếm với dữ liệu đã thu thập.

Để thực hiện xử lý ngôn ngữ tự nhiên em chọn sử dụng thư viện `underthesea` sử dụng phương pháp `word-embedding` để tách từ và gắn các `pos tag` vào các từ được tách nhằm hỗ trợ cho quá trình xác định nghĩa của từ cùng với đó em sẽ dùng thư viện `networkx` để xây dựng đồ thị dựa trên các từ khóa đã xử lý. Cuối cùng là dùng thư viện `pyqt5` để xây dựng GUI cho ứng dụng.

Kết quả thu nhận được là ứng dụng có thể xử lý câu truy vấn để đưa ra bảng xếp hạng so sánh độ tương đồng của câu trên với nội dung dữ liệu pháp luật đã thu thập

2. NỘI DUNG

2.1. Phương pháp.



Hình 2-1: Sơ đồ mô tả phương thức hoạt động của ứng dụng

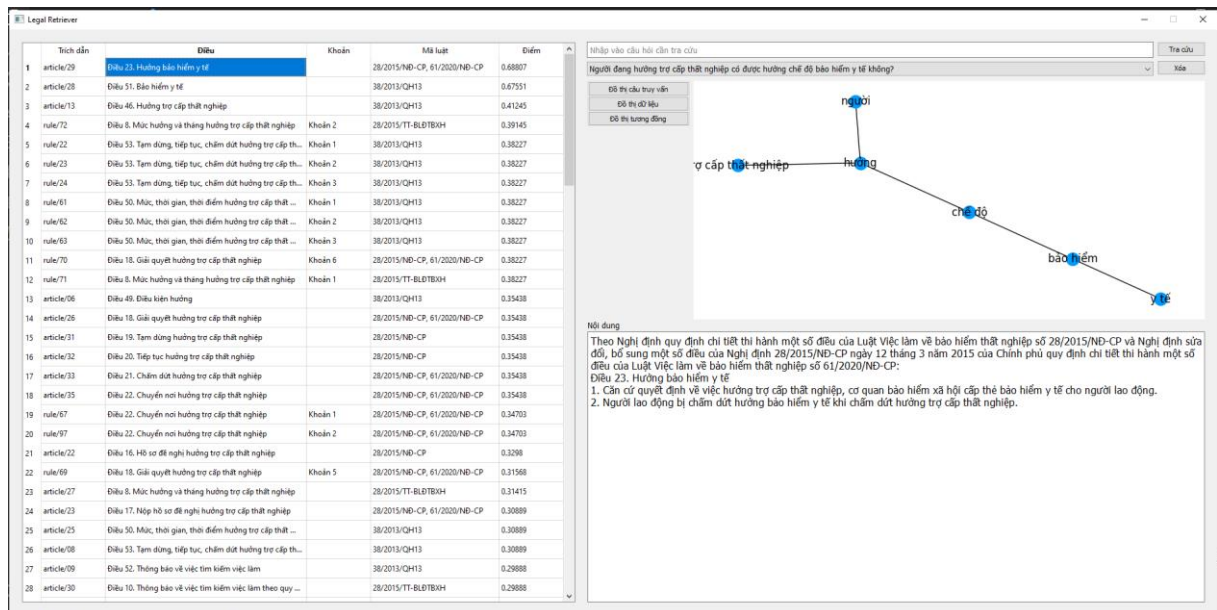
2.1.1. Xử lý ngôn ngữ tự nhiên

Đầu tiên từ câu truy vấn gốc em sẽ dùng underthesea để tách các từ trong câu ra và gán pos tag cho từng từ bằng phương pháp word-embedding. Do từ được tách ra vẫn có độ chưa chính xác cao thì em phải tiếp tục thao tác xác định lại pos tag cho một số từ và loại bỏ các từ dừng. Sau đó, từ dữ liệu đã thu thập em sẽ xác định các từ khái niệm chính của luật bảo hiểm thất nghiệp và thực hiện gom các từ đã được trên lại theo các khái niệm này.

2.1.2. Xây dựng đồ thị khái niệm

Từ các từ khóa đã được xử lý em sẽ tiến hành xây dựng các đồ thị khái niệm bằng networkx và thực hiện tính toán độ tương đồng thông qua hệ số độ tương đồng về khái niệm và quan hệ giữa đồ thị này với các đồ thị khái niệm được xây dựng từ dữ liệu đã thu thập.

2.2. Thiết kế



Hình 2-2: Mô tả giao diện của ứng dụng

- Edit text để nhập vào câu truy vấn
- Combo box để chọn dựa theo câu truy vấn có sẵn
- Button “Tra cứu” để thực hiện chạy ứng dụng dựa trên câu truy vấn (nếu edit text rỗng thì sẽ chạy theo lựa chọn trong combo box)
- Button “Xóa” để bỏ câu truy vấn đã nhập vào
- Table để hiển thị thông tin các dữ liệu được so sánh và điểm tương đồng giữa đồ thị của dữ liệu này với đồ thị của câu truy vấn

- Canvas để hiển thị đồ thị khái niệm được tạo ra từ các từ khóa, có 3 lựa chọn đồ thị ở đây là theo câu truy vấn, theo dữ liệu ô được chọn trong bảng và theo sự tương đồng giữa 2 đồ thị được xét
- Text Browser để hiển thị chi tiết thông tin điều luật của dữ liệu được chọn trên bảng so sánh

2.3. Kết quả

Thành công tách được câu truy vấn dạng ngôn ngữ tự nhiên và thực hiện xử lý loại bỏ các từ không cần thiết và gom cách từ khái niệm liên quan để tạo ra danh sách từ khóa làm nguồn gốc để xây dựng đồ thị khái niệm trên nền tảng với mục đích nhằm so sánh độ tương đồng về ngữ nghĩa giữa câu truy vấn với dữ liệu pháp luật đã thu thập từ đó có thể đưa ra được bảng so sánh đánh giá hướng tới sự tương thích với nội dung các điều luật mà người dùng muốn tìm kiếm

3. KẾT LUẬN

Trải qua quá trình phân tích từ câu truy vấn dưới dạng ngôn ngữ tự nhiên, thực hiện kiểm tra và xử lý các từ bị xác định sai, loại bỏ các từ dừng, xác định và thực hiện ghép các từ khái niệm quan trọng đã trích xuất được các từ khóa quan trọng từ câu truy vấn nhập vào của người dùng.

Và việc ứng dụng xây dựng đồ thị khái niệm dựa trên các từ khóa đã trích xuất và tính toán so sánh đồ thị tương đồng giữa đồ thị câu truy vấn và đồ thị các dữ liệu đã đưa ra được điểm số về mức tương đồng ngữ nghĩa giữa các thành phần được so sánh.

Cuối cùng dựa trên các phương thức đã thực hiện thì tiến hành xây dựng giao diện để thể hiện rõ nét cho người dùng bảng thống kê so sánh độ tương đồng giữa câu truy vấn nhập vào và dữ liệu luật liên quan

Kết quả đạt được là đã thành công xử lý ngôn ngữ tự nhiên và áp dụng phương pháp đồ thị tri thức vào ứng dụng từ đó tính toán, đánh giá, thể hiện được độ tương đồng về ngữ nghĩa của câu truy vấn so với dữ liệu luật liên quan.

TÀI LIỆU THAM KHẢO

- [1] Aurelio Lopez-Lopez, Alexander Gelbukh, Information Retrieval with Conceptual Graph Matching, 2002
- [2] Svetlana Hensman, Construction of Conceptual Graph representation of texts, 2009
- [3] Catherine Comparot, Nathalie Hernandez, Azuraliza Ollivier Haemmerle, An Easy Way of Expressing Conceptual Graph Queries from Keywords and Query Patterns, 2010
- [4] Wan-Hsuan Lin, Chun-Shien Lu, Automated Graph Generation at Sentence Level for Reading Comprehension Based on Conceptual Graphs, 2020
- [5] John Sowa, Conceptual Graphs For Representing Conceptual Structures, 2009

PHỤ LỤC

PHỤ LỤC PHÂN CÔNG NHIỆM VỤ

STT	Thành viên	Nhiệm vụ
1	Nguyễn Trùng Dương	Thực hiện toàn bộ quy trình