

**TRƯỜNG ĐẠI HỌC NAM CẦN THƠ  
KHOA CÔNG NGHỆ THÔNG TIN**



**ĐỖ THỊ BÍCH TRÂM**

**XÂY DỰNG KHO DỮ LIỆU  
LIÊN KẾT MỞ (LOD) THÔNG TIN  
DU LỊCH MEKONG DELTA BẰNG CÁCH  
TRÍCH XUẤT DỮ LIỆU TỪ WEBSITE DU LỊCH**

**KHÓA LUẬN TỐT NGHIỆP ĐẠI HỌC**

**Ngành: Công nghệ Thông tin**

**Mã số ngành: 7480201**

**Tháng 6-2024**

**TRƯỜNG ĐẠI HỌC NAM CẦN THƠ  
KHOA CÔNG NGHỆ THÔNG TIN**



**ĐỖ THỊ BÍCH TRÂM**

**MSSV: 200946**

**LỚP: DH20TIN02**

**XÂY DỰNG KHO DỮ LIỆU  
LIÊN KẾT MỞ (LOD) THÔNG TIN  
DU LỊCH MEKONG DELTA BẰNG CÁCH  
TRÍCH XUẤT DỮ LIỆU TỪ WEBSITE DU LỊCH**

**KHÓA LUẬN TỐT NGHIỆP ĐẠI HỌC  
NGÀNH CÔNG NGHỆ THÔNG TIN**

**Mã số ngành: 7480201**

**CÁN BỘ HƯỚNG DẪN  
ThS. TRẦN VĂN THIỆN**

**Tháng 6-2024**

## CHẤP THUẬN CỦA HỘI ĐỒNG

Khóa luận “Xây dựng kho dữ liệu liên kết mở (LOD) thông tin du lịch Mekong Delta bằng cách trích xuất dữ liệu từ website du lịch”, do sinh viên Đỗ Thị Bích Trâm thực hiện dưới sự hướng dẫn của ThS.Trần Văn Thiện. Khóa luận đã báo cáo và được Hội đồng chấm khóa luận thông qua ngày.....

**Ủy viên**

**Thư ký**

*(ký tên)*

*(ký tên)*

**TS./ThS.....**

**TS./ThS.....**

**Phản biện 1**

**Phản biện 2**

*(ký tên)*

*(ký tên)*

**TS./ThS.....**

**TS./ThS.....**

**Cán bộ hướng dẫn**

**Chủ tịch Hội đồng**

*(ký tên)*

*(ký tên)*

**TS./ThS .....**

**TS./ThS.....**

## LỜI CẢM ƠN

Lời đầu tiên, em xin gửi lời cảm ơn chân thành nhất đến toàn bộ quý thầy cô, ban giám hiệu Trường Đại học Nam Cần Thơ, Quý thầy cô lãnh đạo khoa Công nghệ thông tin, quý thầy cô bộ môn đã dạy dỗ, truyền đạt những kiến thức quý báu cũng như đã tận tình chỉ dạy và tạo điều kiện thuận lợi cho chúng em trong suốt thời gian 4 năm học tập và rèn luyện tại trường. Xin cảm ơn quý thầy cô đã tạo điều kiện thuận lợi để chúng em có thể hoàn thành và bảo cáo khóa luận tốt nghiệp này.

Em xin chân thành cảm ơn và biết ơn sâu sắc nhất đến giảng viên hướng dẫn ThS.Trần Văn Thiện đã tận tình hướng dẫn, đưa ra những lời góp ý và giải đáp những vấn đề trong quá trình thực hiện luận văn tốt nghiệp. Sự quan tâm và hỗ trợ tận tình từ thầy đã giúp tôi vượt qua những thách thức và hoàn thiện dự án của mình một cách tốt nhất.

Cuối cùng em xin gửi lời cảm ơn chân thành đến gia đình và tất cả bạn bè đã giúp đỡ, động viên em trong suốt quá trình học tập và thực hiện đề tài. Sự ủng hộ và động viên nhiệt tình từ họ đã là nguồn động lực lớn lao giúp em hoàn thành mục tiêu của mình.

Một lần nữa em xin chân thành cảm ơn!

*Cần Thơ, ngày..... tháng 6 năm 2024*

**Người thực hiện**

Đỗ Thị Bích Trâm

## **TRANG CAM KẾT**

Tôi xin cam kết khóa luận này được hoàn thành dựa trên các kết quả nghiên cứu của tôi và các kết quả này chưa được dùng cho bất cứ khóa luận cùng cấp nào khác.

Tôi xin cam kết khóa luận này được hoàn thành dựa trên các kết quả nghiên cứu của tôi trong khuôn khổ của đề tài xây dựng kho dữ liệu liên kết mở (LOD) thông tin du lịch Mekong Delta từ việc trích xuất dữ liệu từ website du lịch dưới sự hướng dẫn của ThS.Trần Văn Thiện.

*Cần Thơ, ngày ..... tháng 6 năm 2024*

**Người thực hiện**

**Đỗ Thị Bích Trâm**

# MỤC LỤC

LỜI CẢM ƠN .....	i
TRANG CAM KẾT .....	ii
MỤC LỤC .....	iii
DANH SÁCH BẢNG .....	vii
DANH SÁCH HÌNH.....	viii
DANH MỤC TỪ VIẾT TẮT .....	x
CHƯƠNG 1 .....	1
GIỚI THIỆU.....	1
1.1 Lý do chọn đề tài .....	1
1.2 Lý do trích xuất dữ liệu từ website.....	2
1.3 Mục tiêu nghiên cứu .....	2
1.4 Phạm vi đề tài .....	3
CHƯƠNG 2.....	4
CƠ SỞ LÝ LUẬN VÀ PHƯƠNG PHÁP NGHIÊN CỨU .....	4
2.1 Cơ sở lý luận.....	4
2.2 Phương pháp nghiên cứu .....	4
2.3 Ontology .....	5
2.3.1 Khái niệm Ontology .....	5
2.3.2 Các thành phần trong Ontology .....	5
2.3.3 Tạo ra một Ontology .....	6
2.3.4 Ứng dụng Ontology.....	7
2.3.5 So sánh Ontology và CSDL thông thường.....	7
2.4 Linked Open Data (LOD).....	8
2.4.1 Khái niệm Linked Open Data.....	8
2.4.2 Lợi ích của LOD.....	8
2.4.2.1 Tăng cường khả năng truy cập và khám phá dữ liệu .....	9

2.4.2.2	Tăng cường tính liên kết và tích hợp dữ liệu .....	9
2.4.2.3	Tăng cường khả năng tái sử dụng và mở rộng .....	9
2.4.2.4	Tính mở và duy trì.....	9
2.4.2.5	Khả năng tương tác và liên kết.....	9
2.4.3	Ưu điểm của LOD .....	10
2.4.4	Nhược điểm của LOD .....	10
2.4.5	Xu hướng của LOD .....	11
2.5	Web ngữ nghĩa .....	11
2.5.1	Định nghĩa của web ngữ nghĩa.....	11
2.5.2	Kiến trúc của web ngữ nghĩa.....	12
2.5.3	Hoạt động của web ngữ nghĩa.....	14
2.6	RDF .....	14
2.6.1	Khái niệm RDF .....	14
2.6.2	Các thành phần của RDF.....	15
2.7	RDF Schema (RDFS).....	17
2.7.1	Giới thiệu về RDFS .....	17
2.7.2	Các đặc điểm .....	17
2.8	OWL.....	17
2.8.1	Khái niệm cơ bản về OWL.....	17
2.8.2	Tính năng của OWL.....	18
2.9	SPARQL.....	18
2.9.1	Khái niệm cơ bản về SPARQL .....	18
2.9.2	Các kiểu truy vấn SPARQL .....	19
2.10	Công cụ Protete .....	19
2.10.1	Giới thiệu Protege.....	19
2.10.2	Các đặc điểm của Protege .....	19
2.11	Thư viện Selenium .....	20
2.11.1	Giới thiệu về thư viện Selenium.....	20
2.11.2	Lợi ích của việc sử dụng Selenium .....	21

2.12.1 Giới thiệu về thư viện Scrapy.....	21
2.12.2 Lợi ích của việc sử dụng Scrapy .....	21
CHƯƠNG 3.....	22
THỰC NGHIỆM.....	22
3.1 Phân tích hệ thống .....	22
3.1.1 Sơ đồ quy trình xử lý.....	22
3.1.2 Các trang web trích xuất.....	23
3.1.2.1 vietnamtourism.gov.vn.....	23
3.1.2.2 kenhhomestay.com.....	24
3.1.2.3 sovhttdltuyenquang.vn .....	24
3.1.2.4 luhanhvietnam.com.vn .....	25
3.1.2.5 baibiendep.com.....	25
3.2 Thiết kế Ontology du lịch.....	26
3.2.1 Sử dụng Ontology trong chương trình .....	26
3.2.1.1 Các lớp trong chương trình .....	27
3.2.1.2 Các thuộc tính trong Object Property.....	34
3.2.1.3 Các thuộc tính trong Data Property.....	34
3.2.1.4 Domain và các range trong Ontology .....	35
3.2.1.5 Các tập luật trong ontology .....	36
3.2.2 Trích xuất dữ liệu bằng Python .....	37
3.2.2.1 Trích xuất dữ liệu Ăn Uống .....	39
3.2.2.2 Trích xuất dữ liệu Cơ Sở Lưu Trú.....	40
3.2.2.3 Trích xuất dữ liệu Điểm Đến.....	41
3.2.2.4 Trích xuất dữ liệu Mua Sắm.....	45
3.2.2.5 Trích xuất dữ liệu Phương Tiện Vận Chuyển .....	48
3.2.3 Thêm dữ liệu vào Ontology.....	49
3.2.3.1 Đọc và Chuẩn Bị Dữ Liệu JSON .....	49
3.2.3.2 Thiết Lập Môi Trường RDF.....	49
3.2.3.3 Ánh Xạ Dữ Liệu JSON thành RDF.....	50



3.2.3.4 Lưu Trữ RDF.....	50
3.3 Kết quả thực nghiệm .....	52
3.3.1 Tìm kiếm cụ thể theo từ khóa.....	52
3.3.2 Tra cứu điểm đến.....	56
3.3.3 Tra cứu ăn uống.....	58
3.3.4 Tra cứu cơ sở lưu trú .....	61
3.3.5 Tra cứu ăn uống.....	64
3.3.6 Tra cứu vận chuyển .....	66
CHƯƠNG 4.....	69
KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN .....	69
4.1 Cài đặt và đánh giá .....	69
4.1.1 Cài đặt dự án.....	69
4.1.2 Đánh giá thực nghiệm dự án .....	69
4.2 Kết luận .....	72
4.2.1 Nhận xét .....	72
4.2.2 Ưu điểm.....	72
4.2.3 Nhược điểm .....	72
4.3 Hướng phát triển.....	73
TÀI LIỆU THAM KHẢO .....	74

## DANH SÁCH BẢNG

Bảng 2.1: Ví dụ một số namespace phổ biến.....	15
Bảng 2.2: Một số từ vựng RDF phổ biến.....	17

## DANH SÁCH HÌNH

Hình 2. 1: Ví dụ minh họa biểu diễn tri thức trong Ontology.....	5
Hình 2.2: Web ngữ nghĩa giúp máy khai thác tốt các thông tin trên website ...	12
Hình 2.3: Minh họa các thành phần của ngôn ngữ Web ngữ nghĩa .....	12
Hình 2.4: Mạng lưới RDF .....	13
Hình 2.5: Biểu đồ RDF đơn giản mô tả mối quan hệ .....	15
Hình 2.6: Ví dụ cấu trúc mệnh đề RDF .....	16
Hình 2.7: Giao diện chương trình Protégé .....	20
Hình 3.1: Mô hình xử lý.....	22
Hình 3.2: Số lượng các lớp, thuộc tính trong ứng dụng.....	26
Hình 3.3: Các lớp chính trong chương trình .....	27
Hình 3.4: Các lớp con trong lưu_trú .....	28
Hình 3.5: Các lớp con trong lớp mua_sắm .....	29
Hình 3.6: Các lớp con trong lớp vận_chuyên .....	30
Hình 3.7: Các lớp con trong lớp ăn_uống.....	31
Hình 3.8: Các lớp con trong lớp điếm_đến.....	32
Hình 3.9: Thuộc tính trong Object Property .....	34
Hình 3.10: Các thuộc tính trong Data Property .....	34
Hình 3.11: Một phần Ontograp trong ứng dụng .....	35
Hình 3.12: Giao diện khi thêm mới một thực thể .....	35
Hình 3.13: Truy vấn SPARQL trong chương trình Protégé .....	36
Hình 3.14: Tập luật trong chương trình rotégé .....	36
Hình 3.15: Ontology trước khi thêm dữ liệu.....	51
Hình 3.16: Ontology sau khi thêm dữ liệu.....	51
Hình 3.17: Tìm kiếm địa điểm theo từ khóa .....	52
Hình 3.18: Kết quả tra cứu địa điểm du lịch theo từ khóa .....	53
Hình 3.19: Kết quả hiển thị tra cứu địa điểm theo từ khóa.....	54

Hình 3.20: Kết quả hiển thị theo gợi ý.....	55
Hình 3.21: Kết quả hiển thị tra cứu theo điểm đến .....	57
Hình 3.22: Kết quả hiển thị khi tra cứu.....	58
Hình 3.23: Kết quả hiển thị tra cứu theo ăn uống .....	59
Hình 3.24: Kết quả hiển thị khi tra cứu.....	60
Hình 3.25: Kết quả hiển thị tra cứu theo cơ sở lưu trú .....	62
Hình 3.26: Kết quả hiển thị khi tra cứu.....	63
Hình 3.27: Kết quả hiển thị tra cứu theo mua sắm .....	65
Hình 3.28: Kết quả hiển thị khi tra cứu.....	66
Hình 3.29: Kết quả hiển thị tra cứu theo vận chuyển .....	67
Hình 3.30: Kết quả hiển thị khi tra cứu.....	68
Hình 3.31: Kết quả tìm kiếm từ khóa “Khách sạn Mừng Thanh” từ google..	70
Hình 3.32: Kết quả tìm kiếm “Khách sạn Mừng Thanh” từ Ontology.....	71
Hình 3.33: Kết quả hiển thị gợi ý theo từ khóa “Khách sạn Mừng Thanh” từ Ontology.....	71

## DANH MỤC TỪ VIẾT TẮT

<b>Từ viết tắt</b>	<b>Giải thích</b>
NSD	Người sử dụng
CSDL	Cơ sở dữ liệu
ĐBSCL	Đồng bằng sông Cửu Long