

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



**TRẦN VĂN THOẠI**

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

**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**



**TRẦN VĂN THOẠI**

**MSSV: 201065**

**LỚP: DH20TIN02**

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

**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 ĐBSCL bằng cách trích xuất dữ liệu từ Wikipedia”, do sinh viên Trần Văn Thoại 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 28/06/2024.

**Ủy viên**

**Thư ký**

*(ký tên)*

*(ký tên)*

**ThS. Lê Đức Thắng**

**ThS. Trần Văn Thiện**

**Phản biện 1**

**Phản biện 2**

*(ký tên)*

*(ký tên)*

**ThS. Phan Thị Xuân Trang**

**ThS. Võ Văn Phúc**

**Cán bộ hướng dẫn**

**Chủ tịch Hội đồng**

*(ký tên)*

*(ký tên)*

**ThS. Trần Văn Thiện**

**TS. Ngô Hồ Anh Khôi**

## LỜI CẢM TẠ

Trước hết, em xin chân thành cảm ơn thầy hướng dẫn, thầy Trần Văn Thiện đã tận tình hướng dẫn, nhận xét và giải đáp trong quá trình em viết luận văn tốt nghiệp.

Em xin chân thành cảm ơn các thầy cô trường Đại học Nam Cần Thơ đã tận tình hướng dẫn và tạo điều kiện cho em học tập và phát triển tại trường.

Cuối cùng, em xin bày tỏ lòng biết ơn chân thành tới gia đình và bạn bè đã giúp đỡ, động viên em trong quá trình nghiên cứu và thực hiện đồ án này.

*Cần Thơ, ngày..... tháng 6 năm 2024*

**Người thực hiện**

Trần Văn Thoại

## **TRANG CAM KẾT**

Em đảm bảo rằng bài viết này dựa trên nghiên cứu của chính em, chưa được sử dụng trong bất kỳ bài viết nào khác cùng cấp. Em cam kết luận án này sẽ được hoàn thành trong khuôn khổ Dự án xây dựng Kho dữ liệu liên kết mở (LOD) về thông tin du lịch vùng Đồng bằng sông Cửu Long dưới sự hướng dẫn của Thạc sĩ Trần Văn Thiện.

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

**Người thực hiện**

Trần Văn Thoại

# MỤC LỤC

LỜI CẢM ƠN .....	i
TRANG CAM KẾT .....	ii
MỤC LỤC .....	iii
DANH SÁCH BẢNG .....	viii
DANH SÁCH HÌNH.....	ix
DANH MỤC TỪ VIẾT TẮT .....	xi
CHƯƠNG 1 .....	1
GIỚI THIỆU.....	1
1.1 Tổng quan về ngành du lịch DBSCL .....	1
1.1.1 Tình hình phát triển du lịch tại DBSCL .....	1
1.1.2 Đặc điểm kinh tế văn hóa .....	1
1.1.3 Tiềm năng du lịch.....	1
1.1.4 Những thách thức trong quản lý thông tin du lịch .....	1
1.2 Lý do chọn đề tài .....	1
1.2.1 Nhu Cầu Thực Tế Của Ngành Du Lịch.....	2
1.2.2 Lợi ích của Công Nghệ Linked Open Data (LOD) .....	2
1.2.3 Tận Dụng Nguồn Dữ Liệu Mở từ Wikipedia .....	3
1.2.4 Đóng Góp Cho Cộng Đồng và Sự Phát Triển Bền Vững .....	3
1.3 Tổng quan về Linked Open Data (LOD) .....	3
1.4 Lý Do Trích Xuất Dữ Liệu từ Wikipedia .....	4
1.4.1 Đa Dạng và Toàn Diện của Dữ Liệu .....	4
1.4.2 Tính Mở và Khả Năng Truy Cập .....	4
1.4.3 Tính Chính Xác và Độ Tin Cậy của Dữ Liệu.....	4
1.4.4 Khả Năng Tích Hợp và Liên Kết Dữ Liệu .....	5
1.4.5 Tiềm Năng cho Các Ứng Dụng và Nghiên Cứu .....	5
1.4.6 Các Thách Thức và Giải Pháp.....	5

1.5 Mục tiêu nghiên cứu.....	7
1.5.1 Xây Dựng Hệ Thống Kho Dữ Liệu Liên Kết Mở .....	7
1.5.2 Nâng Cao Trải Nghiệm Người Dùng .....	7
1.5.3 Tăng Cường Liên Kết Ngữ Nghĩa.....	8
1.5.4 Đánh Giá và Ứng Dụng Thực Tiễn .....	8
1.5.5 Đóng Góp Cho Nghiên Cứu và Phát Triển Công Nghệ.....	8
1.6 Đối tượng và phạm vi nghiên cứu.....	8
1.6.1 Đối Tượng Nghiên Cứu.....	8
1.6.2 Phạm Vi Nghiên Cứu .....	10
1.7 Phương pháp nghiên cứu và ưu nhược điểm của đề tài .....	10
1.7.1 Phương pháp tiếp cận .....	10
1.7.2 Ưu điểm.....	10
1.7.3 Nhược điểm .....	11
CHƯƠNG 2.....	12
CƠ SỞ LÝ THUYẾT .....	12
2.1 Ontology .....	12
2.1.1 Khái niệm Ontology .....	12
2.1.2 Các thành phần trong Ontology .....	13
2.1.3 Tạo ra một Ontology .....	14
2.1.4 Ứng dụng Ontology.....	15
2.1.5 So sánh Ontology và CSDL .....	16
2.2 Linked Open Data (LOD).....	18
2.2.1 Khái niệm .....	18
2.2.2 Lợi ích của LOD.....	19
2.2.3 Các Công Nghệ và Tiêu Chuẩn trong LOD .....	20
2.2.4 Các ứng dụng của LOD.....	20
2.2.5 Thách Thức và Tương Lai của LOD .....	21
2.2.5.1 Thách Thức.....	21
2.2.5.2 Tương Lai.....	21

2.3 Web ngữ nghĩa (Semantic Web) .....	21
2.3.1 Định nghĩa của Web Ngữ Nghĩa .....	21
2.3.2 Kiến trúc của Web Ngữ Nghĩa .....	22
2.3.2.1 Mô hình ngữ nghĩa (Semantic Model) .....	22
2.3.2.2 Các thành phần cơ bản (Core Components).....	23
2.3.2.3 Giao thức và Tiêu chuẩn (Protocols and Standards) .....	23
2.3.2.4 Dịch vụ và Ứng dụng (Services and Applications).....	23
2.4 Wikipedia .....	24
2.4.1 Lịch Sử và Sự Phát Triển của Wikipedia .....	24
2.4.2 Cách Hoạt Động của Wikipedia .....	24
2.4.3 Dữ Liệu và Cấu Trúc của Wikipedia.....	24
2.4.4 Sự Phổ Biến và Ảnh Hưởng của Wikipedia.....	25
2.4.4.1 Sự phổ biến.....	25
2.4.4.2 Ảnh hưởng đến giáo dục và nghiên cứu.....	25
2.5. RDF .....	25
2.5.1. Khái niệm RDF .....	25
2.5.2. Các thành phần của RDF .....	26
2.6. RDFS (RDF Schema) .....	29
2.6.1. Giới thiệu về RDFS .....	29
2.6.2. Các đặc điểm .....	29
2.7. OWL .....	29
2.7.1. Khái niệm cơ bản về OWL.....	29
2.7.2. Tính năng của OWL .....	30
2.8. SPARQL.....	31
2.8.1. Khái niệm cơ bản về SPARQL .....	31
2.8.2 Các kiểu truy vấn SPARQL .....	31
2.9 Công cụ Protégé .....	32
2.9.1 Giới thiệu Protégé.....	32
2.9.2 Các đặc điểm của Protégé .....	32



2.10 Dbpedia .....	33
2.10.1 Đặc điểm của Dbpedia .....	34
2.10.2 Cấu trúc của Dbpedia .....	34
2.11 Thư viện SPARQLWrapper .....	35
<b>CHƯƠNG 3.....</b>	<b>36</b>
<b>PHÂN TÍCH VÀ THIẾT KẾ ONTOLOGY DU LỊCH DBSCL.....</b>	<b>36</b>
3.1 Phân tích chức năng Ontology .....	36
3.1.1 Mô hình xử lý .....	36
3.1.2 Các chức năng chính .....	36
3.1.3 Yêu cầu phi chức năng .....	37
3.2 Thiết kế Ontology du lịch DBSCL.....	38
3.2.1 Xây dựng cấu trúc của Ontology.....	38
3.2.1.1 Các lớp trong Ontology .....	39
3.2.1.2 Các Object Property trong Ontology.....	44
3.2.1.4 Domain và các range trong Ontology .....	46
3.2.1.5 Các tập luật trong ontology .....	46
3.2.2 Xây dựng hệ thống trích xuất dữ liệu.....	47
3.2.2.1 Trích xuất dữ liệu ăn uống .....	49
3.2.2.2 Trích xuất dữ liệu cơ sở lưu trú.....	50
3.2.2.3 Trích xuất dữ liệu Điểm Đến.....	51
3.2.2.4 Trích xuất dữ liệu Mua Sắm.....	54
3.2.2.5 Trích xuất dữ liệu Phương Tiện Vận Chuyển .....	56
3.2.3 Thêm dữ liệu vào Ontology.....	57
3.2.3.1 Đọc và Chuẩn Bị Dữ Liệu JSON .....	57
3.2.3.2 Thiết Lập Môi Trường RDF.....	57
3.2.3.3 Ánh Xạ Dữ Liệu JSON thành RDF:.....	57
3.2.3.4 Lưu Trữ RDF.....	58
3.3 Kịch bản thử nghiệm và kết quả.....	60
3.3.1 Tìm kiếm cụ thể theo từ khóa.....	60

3.3.2 Tra cứu điểm đến.....	63
3.3.3 Tra cứu ăn uống.....	64
3.3.4 Tra cứu cơ sở lưu trú .....	65
3.3.5 Tra cứu mua sắm .....	66
3.4 Đánh giá thực nghiệm dự án .....	68
3.4.1 So sánh tìm kiếm với Google .....	68
3.4.2 So sánh tìm kiếm với Tripadvisor .....	70
CHƯƠNG 4.....	73
KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN .....	73
4.1 Kết quả đạt được.....	73
4.1.1 Xây dựng Hệ thống Tự động Tạo Ontology .....	73
4.1.2 Phương pháp Trích xuất Dữ liệu từ Wikipedia .....	73
4.1.3 Đóng góp cho Kho Dữ liệu Liên kết Mở .....	73
4.1 Kết luận .....	74
4.1.1 Nhận xét .....	74
4.1.2 Ưu điểm.....	74
4.1.3 Nhược điểm .....	75
4.2 Hướng phát triển.....	76
TÀI LIỆU THAM KHẢO .....	79

## DANH SÁCH BẢNG

<i>Bảng 2. 1: Ví dụ một số namespace phổ biến</i> .....	27
<i>Bảng 2. 2: Một số từ vựng RDF phổ biến</i> .....	29

## DANH SÁCH HÌNH

Hình 2. 1: Ví dụ minh họa biểu diễn tri thức trong Ontology .....	13
Hình 2. 2: Kỹ thuật xử lý các câu hỏi .....	16
Hình 2. 3: Một phần Lod-cloud .....	19
Hình 2. 4: Web ngữ nghĩa giúp máy có thể khai thác tốt các thông tin trên website.....	22
Hình 2. 5: Minh họa các thành phần của ngôn ngữ Web ngữ nghĩa. ....	23
Hình 2. 6: Biểu đồ RDF đơn giản mô tả mối quan hệ. ....	26
Hình 2. 7: Ví dụ cấu trúc mệnh đề RDF .....	27
Hình 2. 8: Ví dụ về đồ thị rdf .....	28
Hình 2. 9: Ví dụ một owl:DatatypeProperty.....	30
Hình 2. 10: Ví dụ một câu truy vấn Sparql .....	31
Hình 2. 11: Giao diện chương trình Protégé .....	33
Hình 2. 12: Ví dụ một trang Dbpedia.....	34
Hình 3. 1: Mô hình xử lý.....	36
Hình 3. 2: Số lượng các lớp, thuộc tính trong ứng dụng .....	39
Hình 3. 3: Các lớp cha trong chương trình .....	40
Hình 3. 4: Các lớp con trong lưu_trú .....	40
Hình 3. 5: Các lớp con trong lớp mua_sắm.....	41
Hình 3. 6: Các lớp con trong lớp vận_chuyển.....	42
Hình 3. 7: Các lớp con trong lớp ăn_uống .....	42
Hình 3. 8: Các lớp con trong lớp điểm_đến .....	43
Hình 3. 9: Thuộc tính trong Object Property.....	44
Hình 3. 10: Các thuộc tính trong Data Property.....	45
Hình 3. 11: Một phần Ontograf trong ứng dụng .....	46
Hình 3. 12: Giao diện khi thêm mới một thực thể.....	46
Hình 3. 13: Tập luật trong chương trình rotégé .....	47
Hình 3. 14: Kết quả một phần sau khi chạy tập luật .....	47

<i>Hình 3. 15: Sơ đồ thuật toán trích xuất dữ liệu .....</i>	<i>49</i>
<i>Hình 3. 16: Ví dụ cấu trúc json .....</i>	<i>57</i>
<i>Hình 3. 17: Ví dụ ánh xạ các trường dữ liệu .....</i>	<i>58</i>
<i>Hình 3. 18: Ontology trước khi thêm dữ liệu .....</i>	<i>59</i>
<i>Hình 3. 19: Ontology sau khi thêm dữ liệu .....</i>	<i>60</i>
<i>Hình 3. 20: Tìm kiếm địa điểm theo từ khóa .....</i>	<i>61</i>
<i>Hình 3. 21: Kết quả tra cứu địa điểm du lịch theo từ khóa .....</i>	<i>62</i>
<i>Hình 3. 22: Kết quả hiển thị chi tiết địa điểm .....</i>	<i>63</i>
<i>Hình 3. 23: Kết quả hiển thị tra cứu theo điểm đến .....</i>	<i>64</i>
<i>Hình 3. 24: Kết quả hiển thị tra cứu theo ăn uống .....</i>	<i>65</i>
<i>Hình 3. 25: Kết quả hiển thị tra cứu theo cơ sở lưu trú .....</i>	<i>66</i>
<i>Hình 3. 26: Kết quả hiển thị tra cứu theo mua sắm .....</i>	<i>68</i>
<i>Hình 3. 27: Kết quả tìm kiếm từ khóa “Nhà Hàng Lion City” từ Google .....</i>	<i>69</i>
<i>Hình 3. 28: Kết quả tìm kiếm từ khóa “Nhà Hàng Lion City” từ ứng dụng ....</i>	<i>69</i>
<i>Hình 3. 29: Kết quả hiển thị khuyến nghị theo từ khóa “Nhà Hàng Lion City” từ ứng dụng .....</i>	<i>70</i>
<i>Hình 3. 30: Kết quả tìm kiếm từ khóa “Trung Tâm Thương Mại Cái Khế” từ Tripadvisor .....</i>	<i>71</i>
<i>Hình 3. 31: Kết quả tìm kiếm từ khóa “Trung Tâm Thương Mại Cái Khế” từ ứng dụng .....</i>	<i>71</i>
<i>Hình 3. 32: Kết quả hiển thị khuyến nghị theo từ khóa “Trung Tâm Thương Mại Cái Khế” từ ứng dụng .....</i>	<i>72</i>

## DANH MỤC TỪ VIẾT TẮT

<b>Từ viết tắt</b>	<b>Giải thích</b>
UC	Usecase
NSD	Người sử dụng
CSDL	Cơ sở dữ liệu
ĐBSCL	Đồng bằng sông Cửu Long
BGD&ĐT	Bộ Giáo dục và Đào tạo
RDF	Resource Description Framework
OWL	Web Ontology Language
Sparql	PARQL Protocol and RDF Query Language
LOD	Linked Open Data