

TRƯỜNG ĐẠI HỌC NAM CẦN THƠ
KHOA KỸ THUẬT – CÔNG NGHỆ



NGUYỄN HOÀNG ĐẠT
176899

**NGHIÊN CỨU XÂY DỰNG HỆ THỐNG TRẢ
LỜI TỰ ĐỘNG CHUYÊN NGÀNH PHẬT HỌC**

ĐỒ ÁN THỰC TẬP
Ngành Công nghệ Thông tin
Mã số Ngành: 7480201

Cần Thơ, tháng 05 năm 2021

TRƯỜNG ĐẠI HỌC NAM CẦN THƠ
KHOA KỸ THUẬT – CÔNG NGHỆ

NGUYỄN HOÀNG ĐẠT
MSSV: 176899

**NGHIÊN CỨU XÂY DỰNG HỆ THỐNG TRẢ
LỜI TỰ ĐỘNG CHUYÊN NGÀNH PHẬT HỌC**

ĐỒ ÁN THỰC TẬP
Ngành Công Nghệ Thông Tin
Mã số Ngành: 7480201

GIẢNG VIÊN HƯỚNG DẪN
TS. NGÔ HỒ ANH KHÔI

Cần Thơ, tháng 05 năm 2021

CHẤP THUẬN CỦA HỘI ĐỒNG

Đồ án thực tập “Nghiên cứu xây dựng hệ thống trả lời tự động chuyên ngành phật học”, do sinh viên Nguyễn Hoàng Đạt dưới sự hướng dẫn của TS. Ngô Hồ Anh Khôi. Đồ án đã được báo cáo và Hội đồng chấm khóa luận thông qua ngày... tháng... năm 2021.

Ủy viên

Thư ký

GHI CHÚC DANH, HỌ, TÊN

GHI CHÚC DANH, HỌ, TÊN

Phản biện 1

Phản biện 2

GHI CHÚC DANH, HỌ, TÊN

GHI CHÚC DANH, HỌ, TÊN

Cán bộ hướng dẫn

Chủ tịch hội đồng

GHI CHÚC DANH, HỌ, TÊN

GHI CHÚC DANH, HỌ, TÊN

LỜI CẢM ƠN

Tôi xin chân thành gửi lời cảm ơn đến TS. Ngô Hồ Anh Khôi - Giảng viên Trường Đại học Nam Cần Thơ đã nhiệt tình hướng dẫn trong suốt thời gian làm đề tài thực tập tốt nghiệp. Thầy là người đã tạo ra rất nhiều điều kiện thuận lợi và cho em những lời khuyên quý báu nhằm giúp chúng em hoàn thành tốt đề tài thực tập tốt nghiệp này.

Xin gửi lời cảm ơn chân thành đến thầy cô giáo trong Trường Đại học Nam Cần Thơ nói chung và các thầy cô trong Bộ môn Công nghệ thông tin nói riêng đã nhiệt tình giảng dạy, truyền đạt những kiến thức uyên thâm và tạo cho em nhiều điều kiện để giúp đỡ trong suốt quá trình theo học tại trường những năm học qua, giúp em có được những cơ sở lý thuyết vững vàng để có thể thực hiện đề tài ngày hôm nay.

Lời cuối, em xin chân thành tha thiết gửi lời cảm ơn đến bạn bè và gia đình, đã luôn quan tâm, tạo điều kiện khi em cần, giúp đỡ, ủng hộ và động viên em trong quá trình học tập và hoàn thành đồ án thực tập tốt nghiệp.

Xin chân thành cảm ơn!

Cần Thơ, ngày tháng năm 2021

Sinh viên thực hiện

LỜI CAM ĐOAN

Tôi xin cam kết rằng đây là một sản phẩm hoàn toàn do quá trình học tập và vận dụng kiến thức đã được học tại Trường Đại học Nam Cần Thơ để hoàn thành sản phẩm của riêng của mình trong khuôn khổ của đề tài. Về hình thức, những kết luận được trình bày trong đồ án thực tập này là hoàn toàn trung thực.

Cần Thơ, ngày tháng năm 2021

Sinh viên thực hiện

NHẬN XÉT CỦA GIÁO VIÊN HƯỚNG DẪN

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

Cần Thơ, Ngày.....tháng.....năm 2021

Giáo viên hướng dẫn

(Ký tên)

TS. Ngô Hồ Anh Khôi

NHẬN XÉT CỦA GIÁO VIÊN PHẢN BIỆN

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

Cần Thơ, Ngày....tháng.....năm 2021

Giáo viên phản biện

(Ký tên)

ThS. Huỳnh Bá lộc

MỤC LỤC

LỜI CẢM ƠN	i
LỜI CAM ĐOAN.....	ii
NHẬN XÉT CỦA GIÁO VIÊN HƯỚNG DẪN	iii
NHẬN XÉT CỦA GIÁO VIÊN PHẢN BIỆN	iv
MỤC LỤC	v
DANH MỤC BẢNG	ix
DANH MỤC HÌNH	x
DANH MỤC TỪ VIẾT TẮT	xi
TÓM TẮT	xii
CHƯƠNG 1 GIỚI THIỆU	1
1.1 Tổng quan về cơ quan thực tập	1
1.1.1 Thông tin về công ty.....	1
1.1.2 Trụ sở chính.....	1
1.1.3 Người đại diện theo pháp luật:	1
1.1.4 Thông tin liên hệ.....	1
1.1.5 Ngành nghề kinh doanh.....	2
1.1.6 Sản phẩm và dịch vụ đang kinh doanh.....	3
1.2 Giới thiệu chung	8
1.2.1 Đặt vấn đề nghiên cứu.....	8
1.2.2 Mục tiêu nghiên cứu.....	8
1.2.3 Phạm vi nghiên cứu.....	8
1.2.4 Phạm vi đề tài	9
1.2.5 Hướng giải quyết	9
1.2.6 Bố cục	10
CHƯƠNG 2 CƠ SỞ LÝ LUẬN VÀ PHƯƠNG PHÁP NGHIÊN CỨU	11
2.1 Các cơ sở nghiên cứu	11
2.1.1 Trí tuệ nhân tạo	11

2.1.2	Xử lý ngôn ngữ tự nhiên	24
2.2	Phương pháp nghiên cứu	27
2.2.1	Phương pháp nghiên cứu lý thuyết.....	27
2.2.2	Phương pháp nghiên cứu thực nghiệm.....	27
2.2.3	Phương pháp điều tra.....	28
2.3	Giới thiệu về giải Conditional Random Field	28
2.3.1	Giới thiệu chung	28
2.3.2	Mô tả	28
2.3.3	Biến thể	29
2.4	Giới thiệu về ngôn ngữ Python	29
CHƯƠNG 3 GIẢI THUẬT CRF VÀ CÁC CHIẾN LƯỢC ÁP DỤNG		33
3.1	Trích xuất tính năng văn bản và tài liệu	33
3.1.1	Mã hóa	33
3.1.2	Ngưng từ.....	33
3.1.3	Viết hoa	33
3.1.4	Tiếng lóng và từ viết tắt.....	34
3.1.5	Loại bỏ tiếng ồn.....	34
3.1.6	Sửa chính tả	34
3.1.7	Gốc	34
3.1.8	Bổ sung.....	34
3.1.9	Word2Vec	35
3.1.10	Vectơ toàn cầu cho biểu diễn từ (GloVe)	35
3.1.11	Trình bày từ ngữ theo ngữ cảnh.....	35
3.1.12	FastText.....	37
3.1.13	Tần suất kỳ hạn	37
3.1.14	Tần suất tài liệu nghịch đảo thuật ngữ	37
3.2	Giảm kích thước	38
3.3	Kỹ thuật phân loại văn bản.....	38

3.3.1	Thúc đẩy.....	39
3.3.2	Đóng bao.....	39
3.3.3	Mạng thần kinh sâu.....	41
3.3.4	Mạng thần kinh tái diễn (RNN).....	41
3.3.5	Mạng thần kinh hợp pháp (CNN).....	42
3.3.6	Mạng thần kinh chuyên đổi lặp lại (RCNN).....	42
3.3.7	Học sâu đa mô hình ngẫu nhiên (RMDL).....	42
3.3.8	Học sâu phân cấp cho văn bản (HDLTex).....	43
3.4	Đánh giá.....	43
3.5	Văn bản và bộ dữ liệu tài liệu.....	44
3.6	Ứng dụng phân loại văn bản.....	45
3.7	Hỗ trợ phân loại văn bản.....	46
CHƯƠNG 4 MÔ HÌNH HỆ THỐNG.....		48
4.1	Tổng quan.....	48
4.1.1	Khái quát.....	48
4.1.2	Sơ lược lịch sử phát triển.....	48
4.1.3	Kiến trúc chung của hệ thống hỏi đáp.....	48
4.1.4	Một số vấn đề quan tâm khi thiết kế hệ thống hỏi đáp.....	50
4.1.5	Các yêu cầu của hệ thống hỏi đáp.....	50
4.2	Tiền xử lý và phân loại câu hỏi.....	50
4.2.1	Tiền xử lý dữ liệu.....	50
4.2.2	Phân loại câu hỏi.....	52
4.3	Xây dựng mô hình hệ thống.....	55
4.3.1	Lựa chọn mô hình và thuật toán.....	55
4.3.2	Mô hình kiến trúc hệ thống.....	56
4.3.3	Cài đặt.....	58
CHƯƠNG 5 KẾT QUẢ NGHIÊN CỨU.....		59
5.1	Hệ thống.....	59

5.1.1 Thao tác với bộ dữ liệu chuyên ngành phát học.....	59
5.1.2 Giao diện hệ thống chương trình	60
5.2 Hướng dẫn cài đặt	62
5.3 Hướng dẫn sử dụng	63
5.3.1 Thêm dữ liệu	63
5.3.2 Thực thi chương trình.....	64
CHƯƠNG 6 KẾT LUẬN.....	67
TÀI LIỆU THAM KHẢO	68

DANH MỤC BẢNG

Bảng 1.1 Bảng ngành nghề kinh doanh.....	2
Bảng 1.2 Bảng sản phẩm phần mềm	3
Bảng 1.3 Bảng dịch vụ	7

DANH MỤC HÌNH

Hình 3.1 Trí tuệ nhân tạo.....	11
Hình 3.2 Cách thức hoạt động của Chatbots	15
Hình 3.3 Tương tác giữa người dùng với ELIZA	16
Hình 4.1 Hệ thống tìm kiếm thông tin IR	48
Hình 4.2 Kiến trúc chung hệ thống hỏi đáp	49
Hình 4.3 Mô hình hoạt động hệ thống	56
Hình 4.4 Giai đoạn phân tích câu hỏi.....	56
Hình 4.5 Tiến trình phân lớp	57
Hình 4.6 Tiến trình tìm kiếm câu hỏi	57
Hình 4.7 Tiến trình tìm kiếm câu trả lời.....	57
Hình 4.8 Tiến trình tìm kiếm câu hỏi đến chuyên gia.....	57
Hình 5.1 Giao diện chính của chương trình dành cho người dùng	60
Hình 5.2 Giao diện chính dành cho quản trị viên.....	60
Hình 5.3 Menu thêm dữ liệu	61
Hình 5.4 Giao diện chọn tệp dữ liệu cần thêm vào hệ thống	61
Hình 5.5 Giao diện chọn thư mục chứa các tệp dữ liệu cần thêm vào hệ thống.....	62
Hình 5.6 Tệp TraLoiTuDongPhatHoc.exe	62
Hình 5.7 Cửa sổ mới cài đặt.....	62
Hình 5.8 Biểu tượng chạy chương trình ở Desktop	63
Hình 5.9 Tệp chạy chương trình trong thư mục đã cài đặt.....	63
Hình 5.10 Menu “Thêm dữ liệu”.....	63
Hình 5.11 Giao diện chọn đường dẫn của tệp dữ liệu cần thêm vào	64
Hình 5.12 Giao diện chính của form người dùng.....	65
Hình 5.13 Giao diện sau khi thực thi.....	65
Hình 5.14 Giao diện của nhân viên khi người dùng gửi yêu cầu.....	66
Hình 5.15 Giao diện người dùng sau khi nhân viên phản hồi.....	66

DANH MỤC TỪ VIẾT TẮT

TNHH	Trách nhiệm hữu hạn
CSDL	Cơ sở dữ liệu
NLP	Natural Language Processing
HOG	Histogram of Oriented Gradients
OCR	Optical Character Recognition