

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



**TRƯỜNG MINH CẢNH**  
**176892**

**NGHIÊN CỨU XÂY DỰNG HỆ THỐNG**  
**TRẢ LỜI TỰ ĐỘNG CHUYÊN NGÀNH**  
**CÔNG NGHỆ THÔNG TIN**

**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**

**TRƯỜNG MINH CẢNH**  
**MSSV: 176892**

**NGHIÊN CỨU XÂY DỰNG HỆ THỐNG**  
**TRẢ LỜI TỰ ĐỘNG CHUYÊN NGÀNH**  
**CÔNG NGHỆ THÔNG TIN**

**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

**NHẬN XÉT CỦA CƠ QUAN THỰC TẬP**

.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....

Cần Thơ, Ngày.....tháng.....năm 2021

**Cán bộ hướng dẫn**  
(Ký tên)

**PHẠM THỊ THỦY**

**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**

## 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**

## MỤC LỤC

NHẬN XÉT CỦA CƠ QUAN THỰC TẬP .....	i
NHẬN XÉT CỦA GIÁO VIÊN HƯỚNG DẪN .....	ii
NHẬN XÉT CỦA GIÁO VIÊN PHẢN BIỆN .....	iii
LỜI CẢM ƠN .....	iv
LỜI CAM ĐOAN .....	v
MỤC LỤC .....	vi
TÓM TẮT .....	xiv
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
1.1.6.1 Sản phẩm phần mềm .....	3
1.1.6.2 Dịch vụ .....	7
1.2 Giới thiệu chung .....	8
1.2.1 Đặt vấn đề nghiên cứu .....	8
1.2.2 Mục tiêu nghiên cứu .....	8
1.2.2.1 Mục tiêu chung.....	8
1.2.2.2 Mục tiêu cụ thể.....	9
1.2.3 Phạm vi nghiên cứu .....	9
1.2.3.1 Không gian .....	9
1.2.3.2 Thời gian .....	9
1.2.3.3 Đối tượng nghiên cứu.....	9
1.2.4 Phạm vi đề tài .....	9
1.2.5 Hướng giải quyết vấn đề.....	9
1.2.6 Bố cục .....	10



CHƯƠNG 2 CƠ SỞ LÝ THUYẾT VÀ PHƯƠNG PHÁP NGHIÊN CỨU .....	11
2.1 Cơ sở lý thuyết .....	11
2.2.1 Trí tuệ nhân tạo .....	11
2.1.1.1 Khái niệm .....	11
2.1.1.2 Phân loại .....	12
2.1.1.3 Ứng dụng.....	13
2.1.2 Chatbots .....	16
2.1.2.1 Khái niệm .....	16
2.1.2.2 Lịch sử hình thành.....	16
2.1.2.3 Cấu tạo chatbots .....	21
2.1.2.4 Phân loại .....	22
2.1.2.5 Cách thức hoạt động.....	23
2.1.2.6 Một số nền tảng hỗ trợ cho việc phát triển chatbots .....	23
2.1.2.7 Một số ứng dụng chatbot.....	24
2.1.2.8 Xu hướng phát triển.....	24
2.1.3 Xử lý ngôn ngữ tự nhiên .....	25
2.1.3.1 Khái niệm .....	25
2.1.3.2 Các bước xử lý .....	25
2.1.3.3 Các bài toán ứng dụng.....	25
2.1.3.4 Các bài toán xử lý.....	27
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 .....	28
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 .....	29
2.3.3 Biến thể .....	29
2.4 Giới thiệu về ngôn ngữ Python .....	30

2.4.1	Khái niệm.....	30
2.4.2	Lịch sử .....	30
2.4.3	Tính năng .....	31
2.4.4	Được ứng dụng .....	31
<b>CHƯƠNG 3 GIẢI THUẬT CRF VÀ CÁC CHIẾN LƯỢC ÁP DỤNG.....</b>		<b>33</b>
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 .....	33
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 chuyển 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
4.1.3.1 Giao diện người dùng (User Interface) .....	49
4.1.3.2 Phân tích câu hỏi (Question Analyzer) .....	49
4.1.3.3 Tìm kiếm dữ liệu (Data Retrieval) .....	49
4.1.3.4 Rút trích câu trả lời (Answer Extraction).....	49
4.1.3.5 Chiến lược xếp hạng (Ranking) .....	50
4.1.3.6 Xác minh câu trả lời (Answer Verification).....	50
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
4.2.1.1 Phân tích câu hỏi .....	50
4.2.1.2 Tách từ.....	51
4.2.1.3 Biểu diễn văn bản.....	52
4.2.2 Phân loại câu hỏi.....	52
4.2.2.1 Rút trích đặc trưng văn bản .....	52
4.2.2.2 Phân loại câu hỏi tiếng Việt .....	54
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 Công Nghệ Thông Tin .....	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 SÁCH BẢNG**

Bảng 1. 1 Ngành nghề kinh doanh .....	2
Bảng 1. 2 Bảng sản phẩm phần mềm .....	3
Bảng 1. 3 Bảng chi tiết dịch vụ .....	7

## DANH SÁCH HÌNH

Hình 2. 1 Trí thông minh nhân tạo .....	11
Hình 2. 2 Mô hình xe không người lái .....	12
Hình 2. 3 Công nghệ nhận thức .....	13
Hình 2. 4 Các hình thức giao hàng của Uber .....	14
Hình 2. 5 Máy bay thiết bị không người lái .....	14
Hình 2. 6 Giải pháp trong giáo dục .....	15
Hình 2. 7 Cách thức hoạt động của Chatbots .....	16
Hình 2. 8 Phép thử Turing .....	17
Hình 2. 9 Tương tác giữa người dùng với ELIZA .....	18
Hình 2. 10 Mô hình cách thức hoạt động của Chatbot .....	23
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 .....	60
Hình 5. 2 Giao diện chính dành cho quản trị viên .....	60
Hình 5. 3 Giao diện 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 cần thêm vào hệ thống .....	62
Hình 5. 6 Tệp TraLoiTuDONGCNTT.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 sao khi nhân viên phản hồi .....	66

## **DANH MỤC TỪ VIẾT TẮT**

**Tên viết tắt**

**Tên đầy đủ**

TNHH

Trách nhiệm hữu hạn

CSDL

Cơ sở dữ liệu

NLP

Natural Language Processing

HOG

Histogram of Oriented Gradients

OCR

Optical Character Recognition

CRFs

Conditional Random Field

AI

Artificial intelligence