KHOA KỸ THUẬT VÀ CÔNG NGHỆ

**BỘ MÔN CÔNG NGHỆ THÔNG TIN**

****

**THỰC TẬP ĐỒ ÁN CHUYÊN NGÀNH**

**HỌC KỲ I, NĂM HỌC 2024 – 2025**

**XÂY DỰNG WEBSITE TÌM ĐỒ THẤT LẠC**

*Sinh viên thực hiện:*

Họ tên: **Nguyễn Anh Tuấn**

MSSV: **110121123**

Lớp: **DA21TTA**

*Giáo viên hướng dẫn:*

TS.**Nguyễn Nhứt Lam**

***Trà Vinh, ngày tháng 12 năm 2024***

KHOA KỸ THUẬT VÀ CÔNG NGHỆ

**BỘ MÔN CÔNG NGHỆ THÔNG TIN**

****

**THỰC TẬP ĐỒ ÁN CHUYÊN NGÀNH**

**HỌC KỲ I, NĂM HỌC 2024 – 2025**

**XÂY DỰNG WEBSITE TÌM ĐỒ THẤT LẠC**

*Giáo viên hướng dẫn:*TS.**Nguyễn Nhứt Lam**

*Sinh viên thực hiện:*

Họ tên: **Nguyễn Anh Tuấn**

MSSV: **110121123**

Lớp: **DA21TTA**

***Trà Vinh, ngày tháng 12 năm 2024***

**NHẬN XÉT CỦA GIÁO VIÊN HƯỚNG DẪN**

*Trà Vinh, ngày ….. tháng …… năm ……*

**Giáo viên hướng dẫn**

*(Ký tên và ghi rõ họ tên)*

**NHẬN XÉT CỦA THÀNH VIÊN HỘI ĐỒNG**

*Trà Vinh, ngày ….. tháng …… năm ……*

**Thành viên hội đồng**

*(Ký tên và ghi rõ họ tên)*

**LỜI CẢM ƠN**

Lời nói đầu tiên cho em xin gửi lời cảm ơn và biết ơn sâu sắc đến Thầy hướng dẫn TS. Nguyễn Nhứt Lam đã giúp em hoàn thành đề tài “Xây dựng Website tìm đồ thất lạc” và em cũng xin gửi lời cảm ơn chân thành đến các thầy cô đã tận tình hướng dẫn, giảng dạy trong suốt quá trình học tập, nghiên cứu và rèn luyện ở Trường Đại học Trà Vinh.

Mặc dù em đã rất cố gắng để thực hiện đồ án một cách hoàn chỉnh nhất nhưng do thời gian có hạn, trình độ hiểu biết và nhận thức của em chưa cao nên trong đồ án của em không thể tránh khỏi những thiếu sót, em rất mong nhận được sự đóng góp ý kiến của các thầy cô và bạn bè để em có thể hoàn thiện đồ án này tốt hơn

*Em xin chân thành cảm ơn!*

Sinh viên

Nguyễn Anh Tuấn

**MỤC LỤC**

ㅤ

[**MỞ ĐẦU** 1](#_Toc155516481)

[Lý do chọn đề tài 1](#_Toc155516482)

[Mục đích 1](#_Toc155516483)

[Đối tượng 2](#_Toc155516484)

[Phạm vi nghiên cứu 2](#_Toc155516485)

[**CHƯƠNG 1: TỔNG QUAN** 4](#_Toc155516486)

[**1.1. MÔ TẢ VẤN ĐỀ** 4](#_Toc155516487)

[1.1.1. Sự tương đồng trong văn bản 4](#_Toc155516488)

[1.1.2. Thuật toán đánh giá sự tương đồng 4](#_Toc155516489)

[1.1.3. Kết quả của sự tương đồng 4](#_Toc155516490)

[**1.2. MỤC TIÊU NGHIÊN CỨU** 5](#_Toc155516491)

[1.2.1. Tìm hiểu về của hai thuật toán **Jaccard** & **Levenshtein** 5](#_Toc155516492)

[1.2.2. Tìm hiểu về ngôn ngữ **Python** 6](#_Toc155516493)

[1.2.3. Tìm hiểu về những văn bản đầu vào 6](#_Toc155516494)

[1.2.4. Theo sát quá trình so khớp diễn ra như thế nào 6](#_Toc155516495)

[**1.3. HƯỚNG TIẾP CẬN** 6](#_Toc155516496)

[1.3.1. Sau khi tìm hiểu thuật toán và ngôn ngữ lập trình 6](#_Toc155516497)

[1.3.2. Tối ưu quá trình thực hiện 6](#_Toc155516498)

[1.3.3. Khắc phục những lỗi sai nếu có 6](#_Toc155516499)

[**1.4. ĐỐI TƯỢNG VÀ PHẠM VI NGHIÊN CỨU** 6](#_Toc155516500)

[1.4.1. Người dùng cá nhân 6](#_Toc155516501)

[1.4.2. Tổ chức 7](#_Toc155516502)

[1.4.3. Doanh nghiệp 7](#_Toc155516503)

[1.4.4. Nhà nghiên cứu xử lý ngôn ngữ tự nhiên NLP 7](#_Toc155516504)

[1.4.5. Phạm vi nghiên cứu 7](#_Toc155516505)

[**CHƯƠNG 2: NGHIÊN CỨU LÝ THUYẾT** 9](#_Toc155516506)

[**2.1. LÝ THUYẾT** 9](#_Toc155516507)

[2.1.1. Giới thiệu 9](#_Toc155516508)

[2.1.2. Độ đo **Jaccard** 11](#_Toc155516509)

[2.1.3. Độ đo **Levenshtein** 13](#_Toc155516510)

[**2.2. LÝ LUẬN** 15](#_Toc155516511)

[2.2.1. Sự quan trọng của tương đồng văn bản 15](#_Toc155516512)

[2.2.2. Tính ứng dụng rộng rãi 16](#_Toc155516513)

[**2.3. GIẢ THUYẾT KHOA HỌC** 19](#_Toc155516514)

[2.3.1. Giả thuyết chính 19](#_Toc155516515)

[2.3.2. Giả thuyết phụ 20](#_Toc155516516)

[2.3.3. Dự đoán lợi ích về mọi mặt 21](#_Toc155516517)

[**2.4. PHƯƠNG PHÁP NGHIÊN CỨU** 21](#_Toc155516518)

[2.4.1. Xây dựng thuật toán 21](#_Toc155516519)

[2.4.2. Tối ưu hóa và đánh giá 23](#_Toc155516520)

[2.4.3. So sánh và phân tích 24](#_Toc155516521)

[2.4.4. Ứng dụng thực tế 25](#_Toc155516522)

[**CHƯƠNG 3: HIỆN THỰC HÓA NGHIÊN CỨU** 27](#_Toc155516523)

[**3.1. CHUẨN BỊ NỀN TẢNG NGHIÊN CỨU** 27](#_Toc155516524)

[3.1.1. Tìm hiểu cơ sở lý thuyết 27](#_Toc155516525)

[3.1.2. Xây dựng nền tảng lý luận 29](#_Toc155516526)

[**3.2. THIẾT KẾ THUẬT TOÁN** 29](#_Toc155516527)

[3.2.1. Xây dựng thuật toán 29](#_Toc155516528)

[3.2.2. Tối ưu hóa và điều chỉnh 30](#_Toc155516529)

[**3.3. HIỆN THỰC HÓA THUẬT TOÁN** 31](#_Toc155516530)

[3.3.1. Ngôn ngữ lập trình và môi trường 31](#_Toc155516531)

[3.3.2. Cài đặt thuật toán 32](#_Toc155516532)

[**3.4. ĐÁNH GIÁ VÀ THỬ NGHIỆM** 32](#_Toc155516533)

[3.4.1. Chuẩn bị dữ liệu kiểm thử 32](#_Toc155516534)

[3.4.2. Thực hiện thử nghiệm 32](#_Toc155516535)

[**3.5. SO SÁNH VÀ PHÂN TÍCH KẾT QUẢ** 33](#_Toc155516536)

[3.5.1. So sánh kết quả 33](#_Toc155516537)

[3.5.2. Phân tích hiệu suất 33](#_Toc155516538)

[**3.6. KẾT LUẬN VÀ TÍNH ỨNG DỤNG** 33](#_Toc155516539)

[3.6.1. Rút kinh nghiệm và kết luận 33](#_Toc155516540)

[3.6.2. Tính ứng dụng và khả năng mở rộng 34](#_Toc155516541)

[**CHƯƠNG 4: KẾT QUẢ NGHIÊN CỨU** 35](#_Toc155516542)

[**4.1. HIỆU NĂNG** 35](#_Toc155516543)

[4.1.1. Hiệu suất đạt được 35](#_Toc155516544)

[4.2.2. Độ chính xác 35](#_Toc155516545)

[**4.2. CÀI ĐẶT THUẬT TOÁN** 36](#_Toc155516546)

[4.2.1. Thuật toán **Jaccard** 36](#_Toc155516547)

[4.2.2. Thuật toán **Levenshtein** 37](#_Toc155516548)

[**4.3. TRẢI NGHIỆM NGƯỜI DÙNG** 39](#_Toc155516549)

[4.3.1. Giao diện sử dụng 39](#_Toc155516550)

[**CHƯƠNG 5: KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN** 43](#_Toc155516551)

[**5.1. KẾT QUẢ** 43](#_Toc155516552)

[**5.2. HƯỚNG PHÁT TRIỂN** 44](#_Toc155516553)

[**DANH MỤC TÀI LIỆU THAM KHẢO** 45](#_Toc155516554)

[**PHỤ LỤC** 45](#_Toc155516555)

**DANH MỤC HÌNH ẢNH – BẢNG BIỂU**

**Hình 1.** Ví dụ về độ đo Jaccard

**Hình 2.** Ví dụ về độ đo Levenshtein

**Hình 3.** Giải thích từ viết tắt NLP

**Hình 4.** Mô hình tính độ tương đồng văn bản

**Hình 5.** Quá trình xử lý chuyên sâu

**Hình 6.** Các bước tính độ tương đồng văn bản sử dụng độ đo Jaccard

**Hình 7.** Các bước tính độ tương đồng văn bản sử dụng độ đo Levenshtien

**Hình 8.** Thể hiện các mô hình trong lĩnh vực y tế

**Hình 9.** Văn bản pháp luật

**Hình 10.** Chuyển đổi số trong lĩnh vực giáo dục

**Hình 11.** Sự nhận xét góp phần phát triển

**TÓM TẮT NIÊN LUẬN/ĐỒ ÁN CƠ SỞ NGÀNH**

**Những nghiên cứu về hai thuật toán cần sử dụng:**

**Hướng tiếp cận vấn đề:**

**Cách giải quyết vấn đề:**

**Kết quả đạt được sau khi nghiên cứu và triển khai**

# MỞ ĐẦU

## Lý do chọn đề tài

## Mục đích

## Đối tượng

## Phạm vi nghiên cứu

# CHƯƠNG 1: TỔNG QUAN

## 1.1. MÔ TẢ VẤN ĐỀ

### 1.1.1.

### 1.1.2.

### 1.1.3.

## 1.2. MỤC TIÊU NGHIÊN CỨU

### 1.2.1.

### 1.2.2.

### 1.2.3.

### 1.2.4.

## 1.3. HƯỚNG TIẾP CẬN

### 1.3.1.

### 1.3.2.

### 1.3.3.

## 1.4. ĐỐI TƯỢNG VÀ PHẠM VI NGHIÊN CỨU

### 1.4.1.

**1.4.2.**

**1.4.3.**

**1.4.4.**

**1.4.5.**

# CHƯƠNG 2: NGHIÊN CỨU LÝ THUYẾT

## 2.1. LÝ THUYẾT

### 2.1.1. Giới thiệu

### 2.1.2.

### 2.1.3.

## 2.2. LÝ LUẬN

### 2.2.1.

### 2.2.2.

## 2.3. GIẢ THUYẾT KHOA HỌC

### 2.3.1.

### 2.3.2.

### 2.3.3.

## 2.4. PHƯƠNG PHÁP NGHIÊN CỨU

### 2.4.1.

### 2.4.2.

### 2.4.3.

### 2.4.4.

# CHƯƠNG 3: HIỆN THỰC HÓA NGHIÊN CỨU

## 3.1. CHUẨN BỊ NỀN TẢNG NGHIÊN CỨU

### 3.1.1. Tìm hiểu cơ sở lý thuyết

**3.1.1.1.**

**3.1.1.2.**

### 3.1.2.

# CHƯƠNG 4: KẾT QUẢ NGHIÊN CỨU

## 4.1. HIỆU NĂNG

### 4.1.1.

### 4.2.2.

# CHƯƠNG 5: KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

## 5.1. KẾT LUẬN

**5.1.1.**

**5.1.2.**

## 5.2. HƯỚNG PHÁT TRIỂN

**5.2.1**.

**5.2.2.**

# DANH MỤC TÀI LIỆU THAM KHẢO

# PHỤ LỤC