



## TÓM TẮT DỰ ÁN CÁ NHÂN

Họ và Tên: La Thị Cẩm Hồng

Ngành học: Hệ thống thông tin

SĐT: 0794929103

Email: [20dh110783@st.huflit.edu.vn](mailto:20dh110783@st.huflit.edu.vn)

Trường: Đại học Ngoại Ngữ - Tin Học  
HCM (HUFLIT)

## LỜI MỞ ĐẦU

Kính gửi Quý Doanh Nghiệp,

Tôi là La Thị Cẩm Hồng, vừa tốt nghiệp chuyên ngành Hệ thống Thông tin tại trường Đại học Ngoại ngữ - Tin học TP.HCM (HUFLIT). Trong suốt quá trình học tập, tôi đã tham gia và thực hiện nhiều dự án cá nhân, từ phát triển hệ thống quản lý bán hàng, ứng dụng di động, đến tối ưu hóa quy trình doanh nghiệp bằng các giải pháp công nghệ hiện đại.

Mặc dù chưa có nhiều trải nghiệm thực tế, tôi tin rằng những dự án mà tôi đã thực hiện sẽ minh chứng cho khả năng áp dụng kiến thức, kỹ năng giải quyết vấn đề và tinh thần làm việc chăm chỉ của tôi. Qua bảng tóm tắt các dự án này, tôi hy vọng Quý Doanh Nghiệp sẽ nhận thấy tiềm năng và sự phù hợp của tôi với nhu cầu của doanh nghiệp.

Tôi xin chân thành cảm ơn Quý Doanh Nghiệp đã dành thời gian xem xét hồ sơ của tôi. Tôi rất mong có cơ hội được hợp tác và đóng góp vào sự phát triển của Quý Doanh Nghiệp.

Trân trọng,

Cẩm Hồng

lacamhong2k2@gmail.com

# **KHÓA LUẬN TỐT NGHIỆP:**

**Đề tài:** ỨNG DỤNG KHAI THÁC DỮ LIỆU VĂN BẢN TRONG VIỆC PHÂN LOẠI NỘI DUNG WEB

**Tên giáo viên hướng dẫn:** TS. Trần Minh Thái

**Link github:**

- Báo cáo: <https://github.com/lacamhong/Project-BA/blob/0e76ce401bde57c1faad5a1ed5886bf1c12e2f15/HUFLIT-%20THESIS-TextMining-final.docx>
- Crawl URL: <https://github.com/lacamhong/Crawler-Spider-Crapy.git>
- Thuật toán: [https://github.com/lacamhong/Thesis\\_code.git](https://github.com/lacamhong/Thesis_code.git)

**Công nghệ sử dụng:**

- Web Crawling

Scrapy: Một framework mạnh mẽ để thực hiện việc crawling các trang web và trích xuất dữ liệu.

Beautiful Soup (BS4): Thư viện Python để phân tích cú pháp HTML và XML, giúp trích xuất dữ liệu từ các trang web.

- Text Processing

NLTK (Natural Language Toolkit): Thư viện Python để xử lý ngôn ngữ tự nhiên, bao gồm các công cụ để tokenization, stemming, và phân tích ngữ pháp.

- Feature Extraction

TF-IDF (Term Frequency-Inverse Document Frequency): Một kỹ thuật phổ biến để trích xuất các đặc trưng từ văn bản.

Word2Vec/GloVe: Các mô hình từ vector để biểu diễn văn bản dưới dạng vector số.

- Machine Learning Algorithms

Scikit-learn: Thư viện Python phổ biến cung cấp các công cụ cho việc học máy, bao gồm các thuật toán như Naive Bayes, Support Vector Machines (SVM), Random Forest, và K-Nearest Neighbors (KNN).

- **Model Evaluation**

Confusion Matrix: Công cụ để đánh giá hiệu suất của mô hình phân loại.

Precision, Recall, F1-Score: Các chỉ số quan trọng để đo lường hiệu suất mô hình phân loại.

**Giới thiệu:** Đề tài "Ứng dụng khai thác dữ liệu văn bản trong việc phân loại nội dung web" tập trung vào việc sử dụng các kỹ thuật khai thác dữ liệu văn bản (text mining) và học máy (machine learning) để tự động phân loại nội dung các trang web. Mục tiêu của nghiên cứu là phát triển một hệ thống tự động có thể xác định và phân loại chính xác nội dung của các trang web dựa trên các đặc trưng văn bản của chúng.

**Đối tượng nghiên cứu:**

- Trang web: Nghiên cứu tập trung vào việc phân loại nội dung của các trang web.
- Dữ liệu văn bản: Thu thập và phân tích các dữ liệu văn bản từ các trang web để xác định và phân loại chúng.
- Các phương pháp khai thác dữ liệu: Sử dụng các kỹ thuật và công cụ khai thác dữ liệu văn bản để tự động phân loại nội dung trang web.

**Phương pháp nghiên cứu:**

- Phân loại nội dung trang web: Xác định và phân loại các trang web thành các nhóm hoặc danh mục dựa trên nội dung của chúng.
- Dataset: Sử dụng các thuật toán, kỹ thuật tự động trích xuất nội dung cần thiết và xử lý chúng một cách hiệu quả và có thể dùng được cho các bước phân loại.
- Ứng dụng của machine learning: Sử dụng các phương pháp và mô hình machine learning để xây dựng các hệ thống phân loại dựa trên dữ liệu văn bản từ các trang web.

- Đánh giá hiệu suất: Đánh giá hiệu suất của các phương pháp phân loại trang web dựa trên các tiêu chí như độ chính xác, độ phủ, và độ nhớ.

### **Kết quả nghiên cứu dự kiến:**

- Phát triển và triển khai một hệ thống tự động có khả năng phân loại nội dung của các trang web thành các nhóm hoặc danh mục dựa trên dữ liệu văn bản.
- Triển khai được các phương pháp trích xuất dữ liệu như là (url, nội dung, tiêu đề,... )
- Nghiên cứu và áp dụng các phương pháp và mô hình khai thác dữ liệu văn bản, bao gồm machine learning và xử lý ngôn ngữ tự nhiên, để xây dựng các hệ thống phân loại hiệu quả.
- Tiến hành đánh giá hiệu suất của hệ thống phân loại dựa trên các tiêu chí như độ chính xác, độ phủ và độ nhớ, để đảm bảo rằng hệ thống hoạt động hiệu quả trong thực tế.

**Từ khóa:** Khai thác dữ liệu văn bản, phân loại nội dung web, học máy, xử lý ngôn ngữ tự nhiên, BERT, TF-IDF, WebCrawling, DFD.

# MÔN HỌC: CÁC HỆ QUẢN TRỊ CƠ SỞ DỮ LIỆU

**Đề tài:** Xây dựng hệ thống bán sách trên hệ quản trị CSDL Oracle

**Vai trò:** Nhóm trưởng

**Link github:** <https://github.com/lacamhong/Project-DA/blob/f78ec4d03d3726c2284af143c93d8c46a2269549/B%C3%81O-C%C3%81O-CSDL-CamHong.docx>

**Giới thiệu:** Xây dựng hệ thống bán sách trên hệ quản trị CSDL Oracle là tạo ra một hệ thống quản lý bán hàng trực tuyến cho các cửa hàng sách bằng cách sử dụng hệ quản trị CSDL Oracle

**Công nghệ sử dụng:**

- **Ngôn ngữ lập trình:** Sử dụng ngôn ngữ lập trình C#, dùng framework ASP.NET Core 6.0 để viết ứng dụng.
- **Hệ quản trị cơ sở dữ liệu:** hệ quản trị CSDL Oracle

**Đối tượng nghiên cứu:**

- Các cửa hàng sách muốn chuyển đổi từ hệ thống quản lý bán hàng truyền thống sang hệ thống quản lý trực tuyến.
- Khách hàng muốn mua sách trực tuyến một cách thuận tiện và nhanh chóng.

**Phương pháp nghiên cứu:**

- Khảo sát yêu cầu: Tiến hành khảo sát yêu cầu từ các cửa hàng sách và người dùng để hiểu rõ nhu cầu và mong muốn của họ.
- Thiết kế hệ thống: Thiết kế mô hình cơ sở dữ liệu và kiến trúc hệ thống dựa trên yêu cầu đã khảo sát. Đặc biệt là chú tâm vào dữ liệu, phân tích và triển khai hiệu quả, dùng các đối tượng cơ sở dữ liệu (view, trigger, cursor,...) vào đề tài. Phân quyền và backup dữ liệu.
- Phát triển và triển khai: Sử dụng C# và ASP.NET Core 6.0 để phát triển ứng dụng, kết hợp với hệ quản trị CSDL Oracle để quản lý dữ liệu.

- Kiểm thử và đánh giá: Thực hiện các bước kiểm thử hệ thống để đảm bảo tính ổn định và hiệu quả, sau đó thu thập phản hồi từ người dùng để hoàn thiện hệ thống.

### **Kết quả nghiên cứu dự kiến:**

- Xây dựng thành công hệ thống bán sách trực tuyến, cho phép các cửa hàng sách quản lý và theo dõi hoạt động kinh doanh một cách hiệu quả.
- Nắm rõ các kiến thức phân tích và triển khai dữ liệu. Đáp ứng đủ các yêu cầu về ràng buộc dữ liệu, phân quyền và triển khai hệ thống áp dụng các "Database Objects" (Đối tượng cơ sở dữ liệu) như là view, stored procedures, trigger, cursor. Ngoài ra còn phải đáp ứng được yêu cầu backup dữ liệu.
- Nâng cao hiệu suất quản lý dữ liệu và tăng tính bảo mật thông tin cho các cửa hàng sách.

**Từ khóa:** oracle, view, stored procedures, trigger, cursor, hệ thống bán sách, quản lý tồn kho,...

# MÔN HỌC: HỆ HOẠCH ĐỊNH NGUỒN LỰC DOANH NGHIỆP

**Đề tài:** Thiết kế hệ thống thông tin Doanh nghiệp bán quần áo NHIÊN

**Vai trò:** Nhóm trưởng

**Link github:** [https://github.com/lacamhong/Project-BA/blob/0e76ce401bde57c1faad5a1ed5886bf1c12e2f15/NHOM1\\_ket\\_thuc.docx](https://github.com/lacamhong/Project-BA/blob/0e76ce401bde57c1faad5a1ed5886bf1c12e2f15/NHOM1_ket_thuc.docx)

**Giới thiệu:** Hệ thống tích hợp giúp quản lý các quy trình kinh doanh cốt lõi. Đề tài này tập trung vào việc thiết kế một hệ thống thông tin cho doanh nghiệp bán quần áo NHIÊN, sử dụng Odoo, một phần mềm ERP mã nguồn mở phổ biến. Hệ thống này sẽ giúp doanh nghiệp quản lý hiệu quả các hoạt động kinh doanh, bao gồm quản lý kho, bán hàng, mua hàng, và kế toán.

**Công nghệ sử dụng:**

- **Phần mềm ERP:** Odoo
- **Ngôn ngữ lập trình:** Python (sử dụng cho phát triển Odoo)
- **Cơ sở dữ liệu:** PostgreSQL (sử dụng bởi Odoo)
- **Framework:** Odoo Framework

**Đối tượng nghiên cứu:**

- Doanh nghiệp bán quần áo NHIÊN
- Các nhân viên và quản lý trong doanh nghiệp
- Khách hàng mua sắm tại NHIÊN

**Phương pháp nghiên cứu:**

- Khảo sát yêu cầu: Thu thập yêu cầu từ doanh nghiệp, nhân viên và khách hàng để xác định các chức năng cần thiết.
- Thiết kế hệ thống: Sử dụng Odoo để thiết kế các module cần thiết như quản lý kho, bán hàng, mua hàng và kế toán.
- Phát triển và triển khai: Tùy chỉnh và triển khai các module của Odoo theo yêu cầu của doanh nghiệp.



- Kiểm thử và đánh giá: Thực hiện các bài kiểm thử để đảm bảo hệ thống hoạt động hiệu quả và thu thập phản hồi từ người dùng để cải tiến.

**Kết quả nghiên cứu dự kiến:**

- Xây dựng thành công hệ thống thông tin doanh nghiệp sử dụng Odoo, giúp doanh nghiệp NHIÊN quản lý các hoạt động kinh doanh một cách hiệu quả.
- Cải thiện quy trình quản lý kho, bán hàng, mua hàng và kế toán của doanh nghiệp.
- Nâng cao trải nghiệm người dùng cho nhân viên và khách hàng của doanh nghiệp.

**Từ khóa:** ERP, odoo, kế toán, sơ đồ quy trình hoạt động, sơ đồ quan hệ phòng ban, doanh nghiệp bán quần áo,...