

# TRƯỜNG ĐẠI HỌC BÁCH KHOA HÀ NỘI

## ĐỒ ÁN TỐT NGHIỆP

Xây dựng extension giúp ngăn chặn các nội dung độc hại trên website

Nguyễn Đức Hiếu

hieu.nd184099@sis.hust.edu.vn

Ngành Công nghệ thông tin Việt-Nhật

Giảng viên hướng dẫn: TS. Đỗ Quốc Huy

Chữ ký GVHD

Khoa: Công nghệ thông tin Việt-Nhật

Trường: Công nghệ thông tin và Truyền thông

HÀ NỘI, 08/2023

# LỜI CẢM ƠN

Để hoàn thành đồ án tốt nghiệp này, lời đầu tiên em xin cảm ơn TS. Đỗ Bá Lâm, là giảng viên hướng dẫn thực hiện đồ án. Thầy đã dành thời gian theo dõi và góp ý cho em trong suốt quá trình thực hiện, giúp em đảm bảo tiến độ và hoàn thiện hơn đồ án của mình.

Em cũng xin gửi lời cảm ơn tới các thầy cô đã nhiệt tình giảng dạy, truyền đạt kiến thức trong các môn học, các anh chị, bạn bè và người thân đã đồng hành, chia sẻ và giúp đỡ em trong suốt 5 năm học tập tại Đại học Bách khoa Hà Nội.

Do thời gian thực hiện đồ án có hạn và trình độ, kiến thức của bản thân còn hạn chế, đồ án của em có thể còn nhiều thiếu sót. Em rất mong nhận được những ý kiến nhận xét, bổ sung từ các thầy cô và các bạn, giúp em hoàn thiện đồ án và rút kinh nghiệm cho quá trình học hỏi, làm việc sau này. Em xin chân thành cảm ơn!

# TÓM TẮT NỘI DUNG ĐỒ ÁN

Hiện nay, các hình thức mượn sách phổ biến như mượn từ người quen, thư viện trường học, thư viện các tỉnh, thành phố còn một số hạn chế về khả năng dễ tiếp cận, mức độ phong phú, cập nhật các sách mới và khả năng chủ động của những người đọc có mong muốn mượn, cho mượn sách.

Trong đồ án này, em hướng đến xây dựng hệ thống web hỗ trợ mượn sách cộng đồng. Hệ thống có các chức năng chính giúp người dùng tìm kiếm, gửi yêu cầu mượn sách và quản lý những sách mình có thể cho mượn. Dữ liệu thông tin sách cần có cho hệ thống được thu thập từ các trang web bán sách tại Việt Nam. Hệ thống có thể gợi ý sách dựa trên khoảng cách đến vị trí của người dùng và độ tương đồng với sách mà người dùng quan tâm.

Các công việc chính để xây dựng đồ án này bao gồm: Thu thập dữ liệu sách từ các trang web bán sách tại Việt Nam, phân tích, thiết kế và xây dựng hệ thống web hỗ trợ mượn sách cộng đồng, tìm hiểu và áp dụng các phương pháp mã hóa văn bản thành vector để thực hiện gợi ý sách dựa trên độ tương đồng.

## MỤC LỤC

<b>CHƯƠNG 1. GIỚI THIỆU ĐỀ TÀI.....</b>	<b>1</b>
1.1 Đặt vấn đề.....	1
1.2 Mục tiêu và phạm vi đề tài.....	1
1.3 Định hướng giải pháp.....	2
1.4 Bố cục đồ án .....	2
<b>CHƯƠNG 2. KHẢO SÁT VÀ PHÂN TÍCH YÊU CẦU.....</b>	<b>4</b>
2.1 Tổng quan chức năng .....	4
2.1.1 Biểu đồ use case tổng quát .....	4
2.1.2 Biểu đồ phân rã use case Quản lý tài khoản .....	5
2.1.3 Biểu đồ phân rã use case Quản lý sách cá nhân .....	5
2.1.4 Biểu đồ phân rã use case Quản lý sách đang có .....	6
2.1.5 Biểu đồ phân rã use case Quản lý sách cho mượn .....	6
2.1.6 Biểu đồ phân rã use case Quản lý sách mượn .....	6
2.1.7 Biểu đồ phân rã use case Quản lý sách yêu thích .....	7
2.1.8 Biểu đồ phân rã use case Quản lý sách tự thêm .....	7
2.1.9 Biểu đồ phân rã use case Đánh giá sách .....	8
2.1.10 Biểu đồ phân rã use case Quản lý người dùng .....	8
2.2 Đặc tả chức năng .....	9
2.2.1 Đặc tả use case Đăng ký .....	9
2.2.2 Đặc tả use case Quên mật khẩu .....	10
2.2.3 Đặc tả use case Mượn sách.....	11
2.2.4 Đặc tả use case Xác nhận cho mượn sách .....	11
2.2.5 Đặc tả use case Tìm kiếm sách có ở gần .....	12
2.2.6 Đặc tả use case Nhắn tin trò chuyện .....	13

2.2.7 Đặc tả use case Thêm sách đang có.....	14
2.2.8 Đặc tả use case Thêm sách hệ thống chưa có .....	14
2.3 Yêu cầu phi chức năng .....	16
<b>CHƯƠNG 3. THIẾT KẾ HỆ THỐNG VÀ KẾT QUẢ THỰC NGHIỆM .. 17</b>	
3.1 Thiết kế kiến trúc.....	17
3.1.1 Thiết kế tổng quan.....	17
3.1.2 Thiết kế chi tiết gói .....	17
3.2 Thiết kế chi tiết.....	19
3.2.1 Thiết kế giao diện .....	19
3.2.2 Thiết kế lớp .....	22
3.2.3 Thiết kế cơ sở dữ liệu .....	26
3.3 Xây dựng ứng dụng.....	29
3.3.1 Thư viện và công cụ sử dụng.....	29
3.3.2 Kết quả triển khai.....	30
3.4 Kiểm thử chức năng .....	32
3.4.1 Kịch bản kiểm thử chức năng “Thêm sách đang có” .....	32
3.4.2 Kịch bản kiểm thử chức năng “Mượn sách” .....	33
3.4.3 Kịch bản kiểm thử chức năng “Xác nhận cho mượn” .....	34
3.4.4 Kịch bản kiểm thử chức năng “Tìm kiếm sách có ở gần” .....	35
<b>CHƯƠNG 4. CÔNG NGHỆ SỬ DỤNG..... 36</b>	
4.1 Công nghệ sử dụng trong xây dựng hệ thống web .....	36
4.1.1 Framework Angular sử dụng trong xây dựng frontend .....	36
4.1.2 Nền tảng ASP.NET sử dụng trong xây dựng backend.....	36
4.2 Ngôn ngữ Xpath và bộ chọn CSS .....	37
4.3 Các phương pháp trong hệ gợi ý.....	39

4.4 Cơ sở lý thuyết cho các mô hình mã hóa vector.....	41
4.4.1 Mã hóa theo TF-IDF .....	41
4.4.2 Mô hình Word2Vec .....	42
<b>CHƯƠNG 5. TÍCH HỢP DỮ LIỆU VÀ GỌI Ý SÁCH .....</b>	<b>45</b>
5.1 Tích hợp dữ liệu .....	45
5.1.1 Phương pháp thực hiện .....	45
5.1.2 Các nguồn dữ liệu và các giá trị thu thập .....	45
5.1.3 Đổi sánh trùng lặp và đổi sánh thể loại.....	46
5.2 Gọi ý sách.....	49
5.2.1 Thủ nghiệm các phương pháp mã hóa văn bản thành vector .....	49
5.2.2 Thực hiện gọi ý sách .....	58
<b>CHƯƠNG 6. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN.....</b>	<b>60</b>
6.1 Kết luận .....	60
6.2 Hướng phát triển.....	61
<b>TÀI LIỆU THAM KHẢO.....</b>	<b>62</b>
<b>PHỤ LỤC.....</b>	<b>64</b>
<b>A. ĐẶC TẢ USE CASE.....</b>	<b>64</b>
A.1 Đặc tả use case “Đăng nhập” .....	64
A.2 Đặc tả use case “Cập nhật thông tin tài khoản” .....	64
A.3 Đặc tả use case “Đổi mật khẩu”.....	65
A.4 Đặc tả use case “Cập nhật ảnh đại diện”.....	66
A.5 Đặc tả use case “Đánh dấu yêu thích sách” .....	66
A.6 Đặc tả use case “Bỏ yêu thích sách” .....	67
A.7 Đặc tả use case “Đăng cảm nghĩ” .....	67
A.8 Đặc tả use case “Cập nhật cảm nghĩ” .....	68
A.9 Đặc tả use case “Xóa cảm nghĩ” .....	68

A.10	Đặc tả use case “Xem sách đang cho mượn” .....	69
A.11	Đặc tả use case “Đánh dấu sách đã được trả” .....	69
A.12	Đặc tả use case “Xem sách đang yêu cầu mượn” .....	70
A.13	Đặc tả use case “Hủy yêu cầu mượn” .....	71
A.14	Đặc tả use case “Cập nhật thông tin sách” .....	71
A.15	Đặc tả use case “Khóa tài khoản” .....	72
A.16	Đặc tả use case “Mở khóa tài khoản” .....	73
<b>B.</b>	<b>ÁNH XẠ THỂ LOẠI CHO CÁC NGUỒN DỮ LIỆU .....</b>	<b>74</b>
B.1	Ánh xạ thể loại cho nguồn dữ liệu “fahasa.com” .....	74
B.2	Ánh xạ thể loại cho nguồn dữ liệu “khaitam.com” .....	74
B.3	Ánh xạ thể loại cho nguồn dữ liệu “nhasachphuongnam.com” .....	75
B.4	Ánh xạ thể loại cho nguồn dữ liệu “tiki.vn” .....	75
B.5	Ánh xạ thể loại cho nguồn dữ liệu “vinabook.com” .....	76

## DANH MỤC HÌNH VẼ

Hình 2.1	Biểu đồ use case tổng quát . . . . .	4
Hình 2.2	Biểu đồ phân rã use case Quản lý tài khoản . . . . .	5
Hình 2.3	Biểu đồ phân rã use case Quản lý sách cá nhân . . . . .	5
Hình 2.4	Biểu đồ phân rã use case Quản lý sách đang có . . . . .	6
Hình 2.5	Biểu đồ phân rã use case Quản lý sách cho mượn . . . . .	6
Hình 2.6	Biểu đồ phân rã use case Quản lý sách mượn . . . . .	7
Hình 2.7	Biểu đồ phân rã use case Quản lý sách yêu thích . . . . .	7
Hình 2.8	Biểu đồ phân rã use case Quản lý sách tự thêm . . . . .	8
Hình 2.9	Biểu đồ phân rã use case Đánh giá sách . . . . .	8
Hình 2.10	Biểu đồ phân rã use case Quản lý người dùng . . . . .	9
Hình 3.1	Biểu đồ gói thiết kế tổng quan . . . . .	17
Hình 3.2	Chi tiết gói Components . . . . .	17
Hình 3.3	Chi tiết gói Services . . . . .	18
Hình 3.4	Chi tiết gói Controllers . . . . .	18
Hình 3.5	Chi tiết gói Models . . . . .	18
Hình 3.6	Giao diện thiết kế màn hình thông tin sách . . . . .	19
Hình 3.7	Giao diện thiết kế màn hình quản lý sách cá nhân . . . . .	20
Hình 3.8	Giao diện thiết kế màn hình thêm sách chưa có trên hệ thống .	20
Hình 3.9	Giao diện thiết kế màn hình tìm kiếm sách có ở gần . . . . .	20
Hình 3.10	Giao diện thiết kế màn hình nhắn tin . . . . .	21
Hình 3.11	Giao diện thiết kế màn hình quản lý người dùng . . . . .	21
Hình 3.12	Giao diện thiết kế màn hình thống kê hệ thống . . . . .	21
Hình 3.13	Biểu đồ giao tiếp cho use case Mượn sách . . . . .	24
Hình 3.14	Biểu đồ trình tự cho use case Xác nhận cho mượn sách . . . . .	25
Hình 3.15	Biểu đồ trình tự cho use case Hủy yêu cầu mượn . . . . .	25
Hình 3.16	Biểu đồ trình tự cho use case Đánh dấu sách đã được trả . . . . .	25
Hình 3.17	Sơ đồ thiết kế cơ sở dữ liệu . . . . .	26
Hình 3.18	Sơ đồ triển khai hệ thống . . . . .	30
Hình 3.19	Minh họa màn hình thông tin sách . . . . .	31
Hình 3.20	Minh họa màn hình sách gần bạn . . . . .	31
Hình 3.21	Minh họa màn hình sách của bạn . . . . .	32
Hình 4.1	Cấu trúc cây HTML minh họa cho ngôn ngữ XPath [2] . . . . .	38
Hình 4.2	Mô hình Continuous Bag-of-Word [7] . . . . .	42
Hình 4.3	Mô hình Skip-Gram [7] . . . . .	43

Hình 4.4	Mô hình CBOW với một từ ngữ cảnh [7]	43
Hình 5.1	Các bước thực hiện tích hợp dữ liệu	45
Hình 5.2	Các bước thực hiện bài toán phân loại thể loại sách	49
Hình 5.3	Biểu đồ kích thước tập từ vựng và tổng số tần suất xuất hiện trong toàn bộ tập dữ liệu	51

## DANH MỤC BẢNG BIỂU

Bảng 2.1	Đặc tả use case Đăng ký . . . . .	9
Bảng 2.2	Dữ liệu đầu vào trong use case Đăng ký . . . . .	10
Bảng 2.3	Đặc tả use case Quên mật khẩu . . . . .	10
Bảng 2.4	Đặc tả use case Mượn sách . . . . .	11
Bảng 2.5	Đặc tả use case Xác nhận cho mượn sách . . . . .	12
Bảng 2.6	Đặc tả use case Tìm kiếm sách có ở gần . . . . .	13
Bảng 2.7	Đặc tả use case Nhắn tin trò chuyện . . . . .	13
Bảng 2.8	Đặc tả use case Thêm sách đang có . . . . .	14
Bảng 2.9	Đặc tả use case Thêm sách hệ thống chưa có . . . . .	15
Bảng 2.10	Dữ liệu đầu vào trong use case Thêm sách hệ thống chưa có . . . . .	16
Bảng 3.1	Thiết kế lớp BooksController. . . . .	23
Bảng 3.2	Thiết kế lớp BorrowBooksController. . . . .	23
Bảng 3.3	Thiết kế lớp ChatController. . . . .	24
Bảng 3.4	Thiết kế cơ sở dữ liệu bảng "Books". . . . .	27
Bảng 3.5	Thiết kế cơ sở dữ liệu bảng "BookComments". . . . .	27
Bảng 3.6	Thiết kế cơ sở dữ liệu bảng "BorrowBooks". . . . .	27
Bảng 3.7	Thiết kế cơ sở dữ liệu bảng "FavoriteUserBook". . . . .	28
Bảng 3.8	Thiết kế cơ sở dữ liệu bảng "HasUserBook". . . . .	28
Bảng 3.9	Thiết kế cơ sở dữ liệu bảng "Users". . . . .	29
Bảng 3.10	Thiết kế cơ sở dữ liệu bảng "Messages". . . . .	29
Bảng 3.11	Danh sách thư viện và công cụ sử dụng . . . . .	30
Bảng 3.12	Kịch bản kiểm thử chức năng “Thêm sách đang có” . . . . .	33
Bảng 3.13	Kịch bản kiểm thử chức năng “Mượn sách” . . . . .	34
Bảng 3.14	Kịch bản kiểm thử chức năng “Xác nhận cho mượn” . . . . .	35
Bảng 3.15	Kịch bản kiểm thử chức năng “Tìm kiếm sách có ở gần” . . . . .	35
Bảng 4.1	Một số ví dụ về các bộ chọn CSS cơ bản . . . . .	39
Bảng 4.2	Một số ví dụ sử dụng điều kiện trong bộ chọn CSS . . . . .	39
Bảng 5.1	Số lượng sách thu thập tại mỗi nguồn . . . . .	46
Bảng 5.2	Các trường dữ liệu thu thập và ý nghĩa của chúng . . . . .	46
Bảng 5.3	Ánh xạ thể loại cho nguồn dữ liệu “bookbuy.vn” . . . . .	48
Bảng 5.4	Ánh xạ thể loại cho nguồn dữ liệu “cachep.vn” . . . . .	48
Bảng 5.5	Các thể loại sách và số lượng mỗi loại . . . . .	49

Bảng 5.6	Kết quả thử nghiệm mã hóa vector theo TF-IDF với kích thước tập từ vựng 4000 và 6000 . . . . .	52
Bảng 5.7	Kết quả thử nghiệm mã hóa vector theo TF-IDF với kích thước tập từ vựng 8000 và 10000 . . . . .	53
Bảng 5.8	Kết quả thử nghiệm mã hóa vector theo mô hình Word2Vec được huấn luyện trước trên dữ liệu các bài báo tiếng Việt . . . . .	54
Bảng 5.9	Kết quả thử nghiệm mã hóa vector theo mô hình Word2Vec huấn luyện trên dữ liệu sách thu thập được theo phương pháp CBOW	56
Bảng 5.10	Kết quả thử nghiệm mã hóa vector theo mô hình Word2Vec huấn luyện trên dữ liệu sách thu thập được theo phương pháp Skip-Gram . . . . .	57
Bảng 5.11	Kết quả thử nghiệm các phương pháp mã hóa trên tập dữ liệu valid và tập dữ liệu test . . . . .	58
Bảng A.1	Đặc tả use case Đăng nhập . . . . .	64
Bảng A.2	Đặc tả use case Cập nhật thông tin tài khoản . . . . .	65
Bảng A.3	Đặc tả use case Đổi mật khẩu . . . . .	65
Bảng A.4	Dữ liệu đầu vào trong use case Đổi mật khẩu . . . . .	66
Bảng A.5	Đặc tả use case Cập nhật ảnh đại diện . . . . .	66
Bảng A.6	Đặc tả use case Đánh dấu yêu thích sách . . . . .	67
Bảng A.7	Đặc tả use case Bỏ yêu thích sách . . . . .	67
Bảng A.8	Đặc tả use case Đăng cảm nghĩ . . . . .	68
Bảng A.9	Đặc tả use case Cập nhật cảm nghĩ . . . . .	68
Bảng A.10	Đặc tả use case Xóa cảm nghĩ . . . . .	69
Bảng A.11	Đặc tả use case Xem sách đang cho mượn . . . . .	69
Bảng A.12	Đặc tả use case Đánh dấu sách đã được trả . . . . .	70
Bảng A.13	Đặc tả use case Xem sách đang yêu cầu mượn . . . . .	71
Bảng A.14	Đặc tả use case Hủy yêu cầu mượn . . . . .	71
Bảng A.15	Đặc tả use case Cập nhật thông tin sách . . . . .	72
Bảng A.16	Đặc tả use case Khóa tài khoản . . . . .	72
Bảng A.17	Đặc tả use case Mở khóa tài khoản . . . . .	73
Bảng B.1	Ánh xạ thể loại cho nguồn dữ liệu “fahasa.com” . . . . .	74
Bảng B.2	Ánh xạ thể loại cho nguồn dữ liệu “khaitam.com” . . . . .	75
Bảng B.3	Ánh xạ thể loại cho nguồn dữ liệu “nhasachphuongnam.com”	75
Bảng B.4	Ánh xạ thể loại cho nguồn dữ liệu “tiki.vn” . . . . .	76
Bảng B.5	Ánh xạ thể loại cho nguồn dữ liệu “vinabook.com” . . . . .	76

## **DANH MỤC THUẬT NGỮ VÀ TỪ VIẾT TẮT**

<b>Thuật ngữ</b>	<b>Ý nghĩa</b>
API	Giao diện lập trình ứng dụng (Application Programming Interface)
CBOW	Continuous Bag-of-Word
CSS	Cascading Style Sheets
HTML	HyperText Markup Language
IDE	Môi trường phát triển tích hợp (Integrated Development Environment)
JSON	JavaScript Object Notation
KNN	K-Nearest Neighbors
MVC	Model - View - Controller
SPA	Ứng dụng một trang (Single-Page Application)
TF-IDF	Term Frequency – Inverse Document Frequency
W3C	World Wide Web Consortium
XPath	XML Path Language

# CHƯƠNG 1. GIỚI THIỆU ĐỀ TÀI

## 1.1 Đặt vấn đề

Sách là nơi chứa đựng nhiều giá trị về các lĩnh vực trong cuộc sống. Mỗi cuốn sách đều mang trong mình những kiến thức, kinh nghiệm, cùng những cảm xúc và ý nghĩa mà tác giả mong muốn gửi gắm đến người đọc. Không chỉ là sản phẩm cần thiết, phục vụ cho quá trình học tập và làm việc. Sách còn là món ăn tinh thần đem lại nhiều lợi ích. Việc đọc sách giúp nâng cao vốn hiểu biết về nhiều lĩnh vực, nhiều nền văn hóa khác nhau từ khắp các quốc gia trên thế giới. Đọc sách thường xuyên cũng giúp mở rộng vốn từ, tăng khả năng tập trung và giảm căng thẳng.

Với sự phát triển của các thiết bị công nghệ, ngoài việc sử dụng sách giấy, người đọc có thể tiếp cận với sách qua sách điện tử hoặc sách nói. Tuy nhiên, hai hình thức này chưa thể thay thế được hoàn toàn sách giấy do số lượng sách nói và sách điện tử, cả có bản quyền và không có bản quyền, còn hạn chế, chưa đầy đủ như sách giấy. Việc sử dụng sách giấy cũng đem lại cảm giác thoải mái và hứng thú hơn cho người đọc. Điểm khó khăn trong việc tiếp cận sách giấy là vấn đề chi phí, số lượng sách mà mỗi người có thể mua được là có giới hạn. Người đọc có mong muốn mượn sách để có thể tiếp cận với nhiều sách hơn. Ngoài ra, những người sở hữu sách cũng mong muốn có thể chia sẻ sách của mình tới những người xung quanh.

Việc mượn và cho mượn sách ở nước ta hiện nay chủ yếu được thực hiện qua một số hình thức như mượn sách từ người quen, thư viện trường học, thư viện các tỉnh, thành phố. Những hình thức này còn hạn chế về khả năng dễ tiếp cận, mức độ phong phú, cập nhật các sách mới và khả năng chủ động của những người đọc có mong muốn mượn, cho mượn sách. Việc có thể chia sẻ sách, mượn sách từ những người xung quanh giúp mọi người tiếp cận với sách nhiều hơn, nâng cao tính hiệu quả của nguồn sách có trong cộng đồng và tăng thêm tính kết nối giữa các bạn đọc.

## 1.2 Mục tiêu và phạm vi đề tài

Với các nhu cầu được trình bày trong phần 1.1, trong đồ án này, em hướng đến xây dựng hệ thống web hỗ trợ mượn sách cộng đồng. Hệ thống cung cấp các chức năng cho phép người dùng tìm kiếm và gửi yêu cầu mượn sách, nhắn tin trò chuyện với mọi người, quản lý sách mình đang có, sách yêu thích, sách mượn và cho mượn, chia sẻ cảm nghĩ về các cuốn sách. Để có thông tin về sách phục vụ cho các chức năng trên, hệ thống cần thu thập được dữ liệu sách phổ biến tại thị trường trong nước. Ngoài ra, người dùng có thể đăng thêm thông tin về các cuốn sách mà hệ thống chưa có. Từ thông tin về sách và địa chỉ người dùng, hệ thống có khả năng

gợi ý sách dựa trên độ tương đồng và khoảng cách đến vị trí của người dùng.

### 1.3 Định hướng giải pháp

Để thực hiện các mục tiêu đã nêu trong phần 1.2, em xây dựng hệ thống web theo hướng Single-Page Application (SPA). Dữ liệu sách như tên sách, tác giả, mô tả,... được thu thập từ các trang web bán sách. Chức năng gợi ý được thực hiện theo phương pháp gợi ý dựa trên nội dung bằng cách mã hóa các văn bản mô tả thành các vector đại diện cho mỗi cuốn sách. Hệ thống cho phép người dùng lựa chọn địa điểm trên giao diện bản đồ, từ đó sử dụng tọa độ kinh độ và vĩ độ để tính khoảng cách, tìm kiếm các cuốn sách có ở gần vị trí của người dùng.

Hệ thống web được xây dựng theo hướng SPA với frontend sử dụng framework Angular và backend sử dụng các công nghệ trên nền tảng ASP.NET. Các Web API được xây dựng bằng ASP.NET Core. Hệ thống sử dụng cơ sở dữ liệu quan hệ MySQL, việc kết nối và truy vấn đến cơ sở dữ liệu được thực hiện qua Entity Framework. Thư viện SignalR được sử dụng để tạo kết nối Web Socket cho chức năng nhắn tin.

Việc tích hợp dữ liệu từ các trang web bán sách được thực hiện theo kiến trúc kho dữ liệu (Data warehousing). Dữ liệu được thu thập và lưu trữ vào cơ sở dữ liệu dựa trên lược đồ chung được định nghĩa trước. Các thuộc tính được bóc tách từ các nguồn dữ liệu bằng đường dẫn Xpath, CSS Selector hoặc bóc tách từ dữ liệu JSON tùy theo từng nguồn dữ liệu.

Để mã hóa dữ liệu văn bản thành các vector đại diện cho mỗi cuốn sách, phục vụ cho việc tính độ tương đồng khi gợi ý sách, em thử nghiệm một số phương pháp mã hóa như TF-IDF (Term Frequency – Inverse Document Frequency), sử dụng mô hình Word2Vec được huấn luyện sẵn trên tập dữ liệu các bài báo tiếng Việt [1] và huấn luyện lại mô hình Word2Vec trên dữ liệu mô tả của các cuốn sách thu thập được.

### 1.4 Bố cục đồ án

Phần còn lại của báo cáo đồ án tốt nghiệp này được tổ chức như sau.

Chương 2 trình bày về khảo sát và phân tích yêu cầu, giai đoạn đầu tiên của quy trình phân tích và thiết kế hệ thống. Các phần trong chương lần lượt nêu ra các chức năng chính mà hệ thống cần có, tổng quan về các chức năng và các tác nhân của hệ thống, đặc tả chi tiết cho một số use case quan trọng và đưa ra một số yêu cầu phi chức năng cần thực hiện.

Trong chương 3, em trình bày về các thiết kế hệ thống và kết quả thực nghiệm. Mở đầu là thiết kế tổng quan về kiến trúc hệ thống. Tiếp theo, em trình bày về các

## CHƯƠNG 1. GIỚI THIỆU ĐỀ TÀI

---

thiết kế giao diện, thiết kế lớp và cơ sở dữ liệu. Sau đó là minh họa cho các kết quả đạt được khi hoàn thành hệ thống và kết quả kiểm thử cho một số chức năng chính của hệ thống.

Chương 4 giới thiệu về một số công nghệ được sử dụng để thực hiện đồ án và một số cơ sở lý thuyết liên quan. Nội dung chương bao gồm các công nghệ để xây dựng hệ thống web như Angular, ASP.NET, giới thiệu về ngôn ngữ XPath và bộ chọn CSS sử dụng trong bóc tách dữ liệu, một số phương pháp sử dụng trong hệ gợi ý và cơ sở lý thuyết cho các mô hình mã hóa văn bản thành vector như TF-IDF, Word2Vec.

Trong chương 5, em trình bày cụ thể hơn về quá trình tích hợp dữ liệu và gợi ý sách. Phần tích hợp dữ liệu trình bày về phương pháp tích hợp Kho dữ liệu (Data warehousing) so với Tích hợp dữ liệu ảo (Virtual data integration), sau đó trình bày về các nguồn dữ liệu, các giá trị thu thập và việc đối sánh trùng lặp, đối sánh thể loại của các cuốn sách thu thập được. Nội dung phần gợi ý mô tả về cách thức lựa chọn các cuốn sách được gợi ý trong các trường hợp khi người dùng xem chi tiết một cuốn sách, danh sách các sách mình đang có, yêu thích, hay khi người dùng tìm kiếm sách ở gần mình. Phần này cũng trình bày về việc thử nghiệm kết quả khi sử dụng các phương pháp mã hóa vector khác nhau thông qua bài toán phân loại thể loại sách, dựa trên dữ liệu hệ thống thu thập được.

Chương 6 là chương cuối của báo cáo đồ án tốt nghiệp này, trình bày về một số kết luận và hướng phát triển thêm cho hệ thống.

## CHƯƠNG 2. KHẢO SÁT VÀ PHÂN TÍCH YÊU CẦU

### 2.1 Tổng quan chức năng

#### 2.1.1 Biểu đồ use case tổng quát

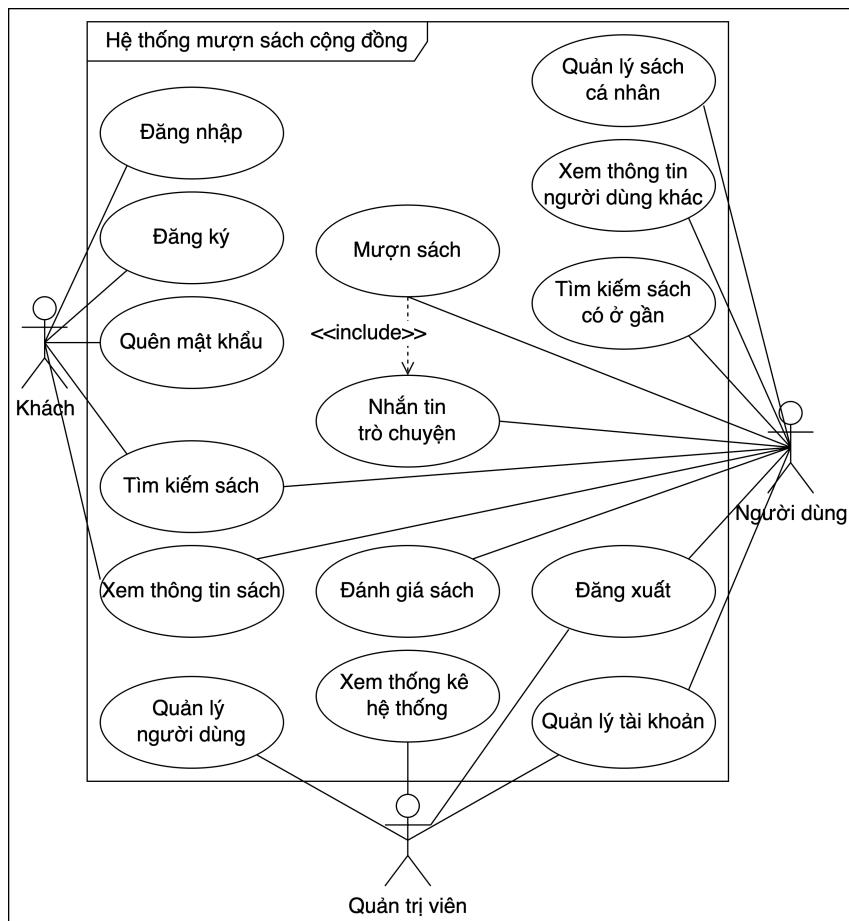
Hình 2.1 thể hiện biểu đồ use case tổng quát của hệ thống.

Hệ thống có 3 tác nhân là Khách, Người dùng, Quản trị viên.

Khách là người truy cập vào hệ thống và chưa đăng nhập, có thể thực hiện tìm kiếm và xem thông tin sách. Sau khi đăng nhập thành công, Khách sẽ trở thành Người dùng.

Người dùng có thể thực hiện quản lý thông tin tài khoản, tìm kiếm sách có ở gần, gửi yêu cầu mượn sách, nhắn tin với các người dùng khác, đánh giá sách, quản lý sách cá nhân, bao gồm sách người dùng đang có, sách yêu thích, sách cho mượn, sách mượn và sách người dùng tự thêm thông tin.

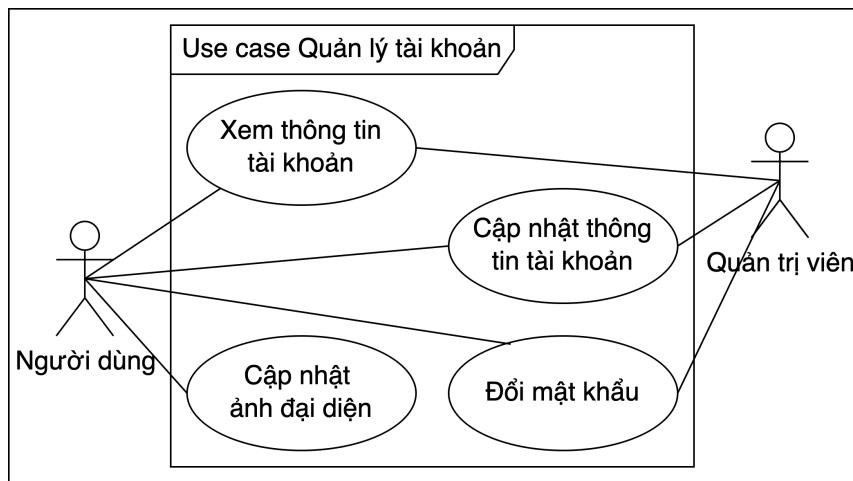
Quản trị viên là người quản trị hệ thống, có thể thực hiện quản lý tài khoản người dùng, xem các thống kê hệ thống, thay đổi tên đăng nhập và mật khẩu.



Hình 2.1: Biểu đồ use case tổng quát

### 2.1.2 Biểu đồ phân rã use case Quản lý tài khoản

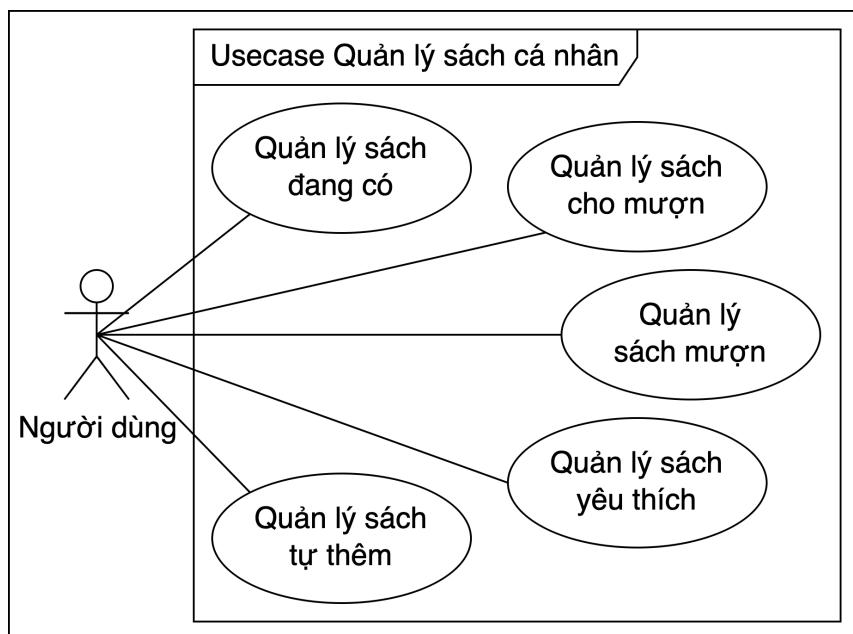
Hình 2.2 là biểu đồ phân rã use case Quản lý tài khoản. Trong use case này, người dùng có thể cập nhật ảnh đại diện, đổi mật khẩu và thay đổi các thông tin chi tiết đã nhập khi đăng ký như tên, email, địa chỉ, số điện thoại. Đối với tài khoản quản trị viên, việc cập nhật thông tin cho phép thay đổi tên đăng nhập.



**Hình 2.2:** Biểu đồ phân rã use case Quản lý tài khoản

### 2.1.3 Biểu đồ phân rã use case Quản lý sách cá nhân

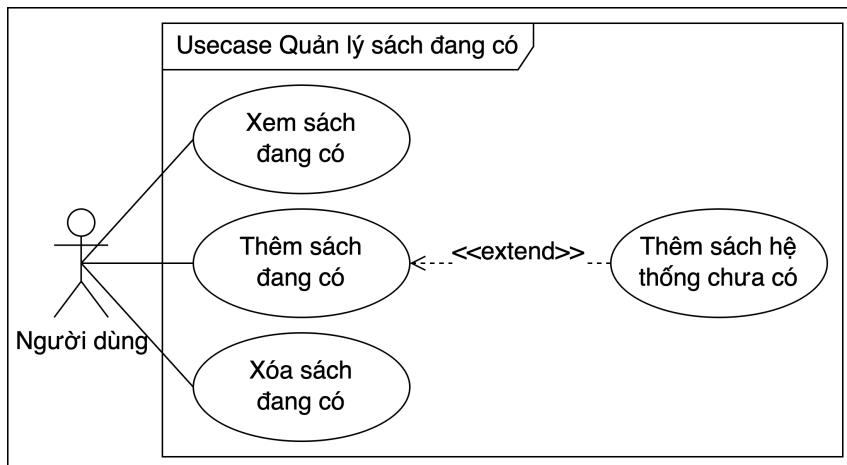
Hình 2.3 là biểu đồ phân rã use case Quản lý sách cá nhân. Use case này được phân rã thành các use case nhỏ hơn theo các danh mục sách đang có, sách yêu thích, sách cho mượn, sách mượn và sách tự thêm. Biểu đồ phân rã use case theo các danh mục này được trình bày trong các phần tiếp theo.



**Hình 2.3:** Biểu đồ phân rã use case Quản lý sách cá nhân

### 2.1.4 Biểu đồ phân rã use case Quản lý sách đang có

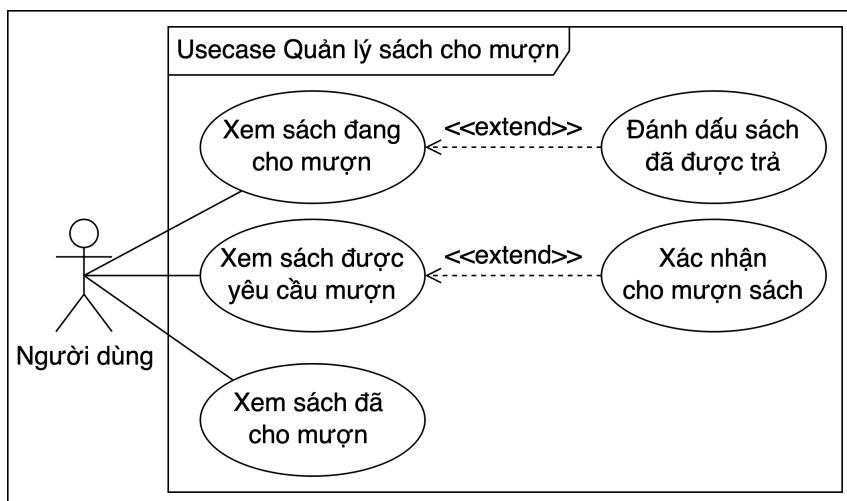
Hình 2.4 là biểu đồ phân rã use case Quản lý sách đang có. Người dùng thực hiện tìm kiếm theo tên sách để chọn thêm những cuốn sách mình có thể cho mượn. Khi không có kết quả phù hợp với tên sách người dùng nhập, hệ thống sẽ gợi ý người dùng thêm thông tin sách mà hệ thống chưa có.



**Hình 2.4:** Biểu đồ phân rã use case Quản lý sách đang có

### 2.1.5 Biểu đồ phân rã use case Quản lý sách cho mượn

Hình 2.5 là biểu đồ phân rã use case Quản lý sách cho mượn. Người dùng có thể xem và xác nhận các yêu cầu mượn hoặc đánh dấu đã được trả sách khi nhận lại sách từ người mượn.

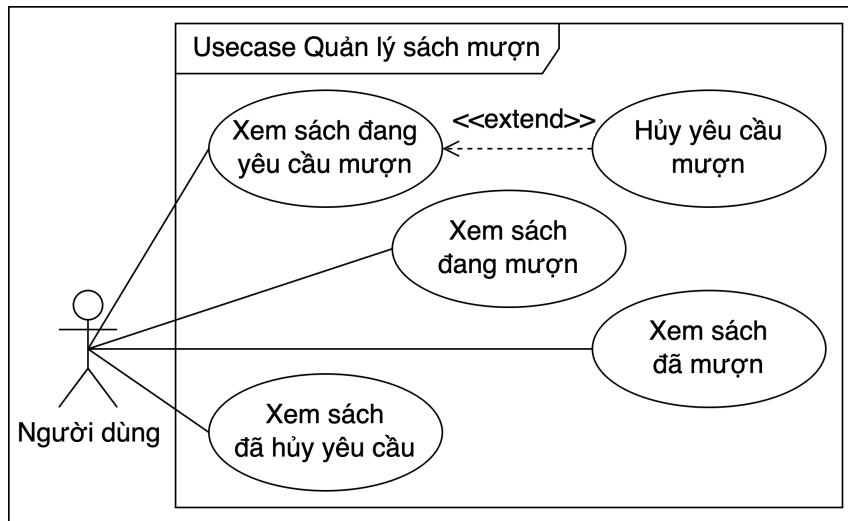


**Hình 2.5:** Biểu đồ phân rã use case Quản lý sách cho mượn

### 2.1.6 Biểu đồ phân rã use case Quản lý sách mượn

Hình 2.6 là biểu đồ phân rã use case Quản lý sách mượn. Người dùng xem sách mượn với các trạng thái: đang yêu cầu, đang mượn, đã mượn, đã hủy yêu cầu. Khi

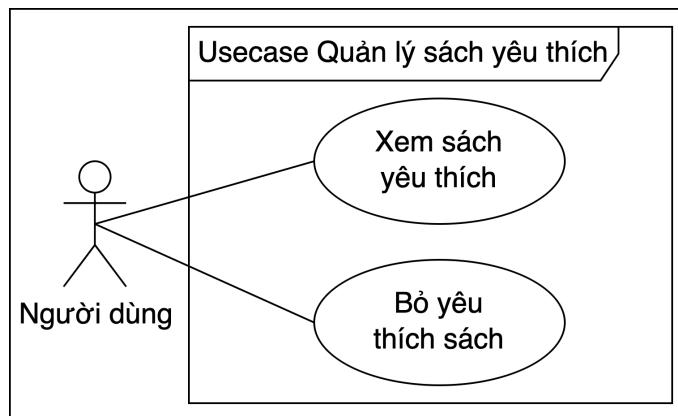
không liên lạc được với người cho mượn hoặc không còn mong muốn mượn, người dùng có thể hủy các yêu cầu mượn đã gửi.



**Hình 2.6:** Biểu đồ phân rã use case Quản lý sách mượn

### 2.1.7 Biểu đồ phân rã use case Quản lý sách yêu thích

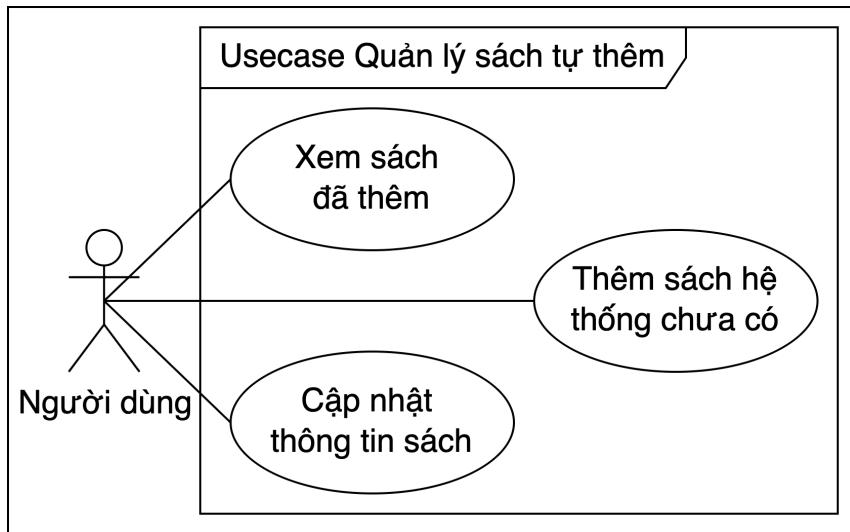
Hình 2.7 là biểu đồ phân rã use case Quản lý sách yêu thích. Người dùng có thể xem lại các cuốn sách mà mình yêu thích, hoặc bỏ các cuốn sách không mong muốn ra khỏi danh sách này.



**Hình 2.7:** Biểu đồ phân rã use case Quản lý sách yêu thích

### 2.1.8 Biểu đồ phân rã use case Quản lý sách tự thêm

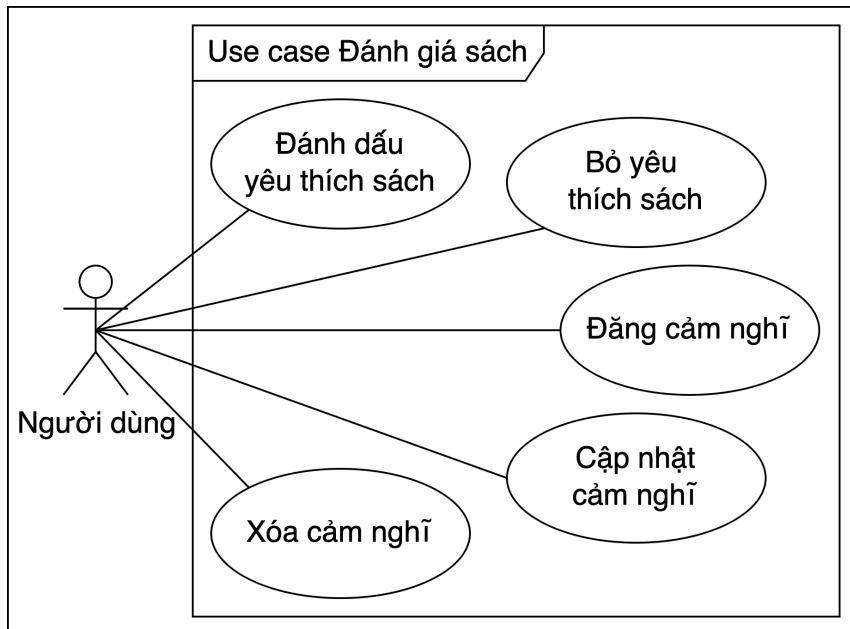
Hình 2.8 là biểu đồ phân rã use case Quản lý sách tự thêm. Người dùng có thể thêm và cập nhật thông tin các cuốn sách mà hệ thống chưa có.



**Hình 2.8:** Biểu đồ phân rã use case Quản lý sách tự thêm

### 2.1.9 Biểu đồ phân rã use case Đánh giá sách

Hình 2.9 là biểu đồ phân rã use case Đánh giá sách. Tại giao diện thông tin chi tiết của một cuốn sách, người dùng có thể đánh dấu yêu thích sách hoặc chia sẻ cảm nghĩ của mình về cuốn sách đó.

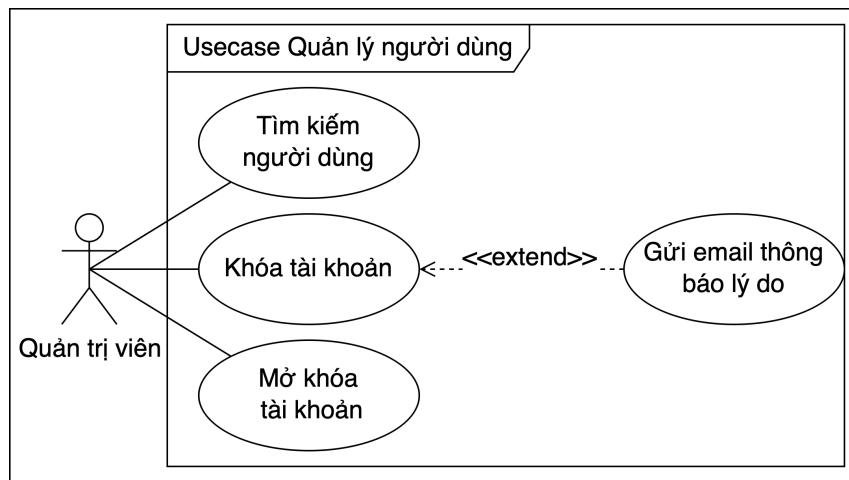


**Hình 2.9:** Biểu đồ phân rã use case Đánh giá sách

### 2.1.10 Biểu đồ phân rã use case Quản lý người dùng

Hình 2.10 là biểu đồ phân rã use case Quản lý người dùng. Trong trường hợp cần thiết, quản trị viên của hệ thống có thể tìm kiếm và khóa hoặc mở khóa một số tài khoản. Khi khóa tài khoản, quản trị viên có thể lựa chọn thêm việc gửi email thông báo lý do khóa đến người dùng.

## CHƯƠNG 2. KHẢO SÁT VÀ PHÂN TÍCH YÊU CẦU



**Hình 2.10:** Biểu đồ phân rã use case Quản lý người dùng

### 2.2 Đặc tả chức năng

#### 2.2.1 Đặc tả use case Đăng ký

Bảng 2.1 và bảng 2.2 đặc tả chi tiết cho use case Đăng ký và mô tả dữ liệu đầu vào sử dụng trong use case này.

Tên use case	Đăng ký		
Tác nhân	Khách		
Tiền điều kiện	Không		
Luồng sự kiện chính	STT	Thực hiện bởi	Hành động
	1	Khách	Ấn chọn đăng ký
	2	Hệ thống	Hiển thị giao diện đăng ký
	3	Khách	Nhập thông tin đăng ký*
	4	Khách	Ấn nút đăng ký
	5	Hệ thống	Thông báo đăng ký thành công
Luồng sự kiện phát sinh	STT	Thực hiện bởi	Hành động
	5a	Hệ thống	Hiển thị các trường dữ liệu không hợp lệ
	5a1	Hệ thống	Thông báo tên đăng nhập hoặc email đã được sử dụng
Hậu điều kiện	Hệ thống lưu thông tin tài khoản mới, hiển thị giao diện đăng nhập		

**Bảng 2.1:** Đặc tả use case Đăng ký

\* Mô tả dữ liệu đầu vào

Trường dữ liệu	Bắt buộc?	Điều kiện hợp lệ	Ví dụ
Tên đăng nhập	Có		anhht

## CHƯƠNG 2. KHẢO SÁT VÀ PHÂN TÍCH YÊU CẦU

Mật khẩu	Có	Tối thiểu 6 ký tự	MatKhau123!@#
Nhập lại mật khẩu	Có	Trùng với mật khẩu	MatKhau123!@#
Họ và tên	Có		Hoàng Thế Anh
Email	Có	Đúng định dạng email	anhht@gmail.com
Địa chỉ	Có		Số 1 Đại Cồ Việt, Bách Khoa, Hai Bà Trưng, Hà Nội
Số điện thoại	Không	10 số, bắt đầu bằng 0	0987654321

**Bảng 2.2:** Dữ liệu đầu vào trong use case Đăng ký

### 2.2.2 ĐẶC TẢ USE CASE QUÊN MẬT KHẨU

Bảng 2.3 đặc tả chi tiết cho use case Quên mật khẩu. Use case này bao gồm hai màn hình nhập email và đặt lại mật khẩu mới. Mã OTP được gửi đến email của người dùng để xác nhận trong quá trình đặt lại mật khẩu mới.

Tên use case	Quên mật khẩu		
Tác nhân	Khách		
Tiền điều kiện	Không		
Luồng sự kiện chính	STT	Thực hiện bởi	Hành động
	1	Khách	Ấn chọn quên mật khẩu
	2	Hệ thống	Hiển thị giao diện quên mật khẩu
	3	Khách	Nhập địa chỉ email
	4	Khách	Ấn nút gửi
	5	Hệ thống	Thông báo gửi email thành công
	6	Hệ thống	Hiển thị giao diện nhập mã OTP và mật khẩu mới
	7	Khách	Nhập mã OTP và mật khẩu mới
	8	Khách	Ấn nút gửi
	9	Hệ thống	Thông báo đặt lại mật khẩu thành công
Luồng sự kiện phát sinh	STT	Thực hiện bởi	Hành động
	9a	Hệ thống	Hiển thị "Vui lòng nhập mật khẩu tối thiểu 6 ký tự" / "Nhập lại mật khẩu không đúng"
	9a1	Hệ thống	Thông báo mã OTP không chính xác
Hậu điều kiện	Hệ thống lưu thông tin mật khẩu mới, hiển thị giao diện đăng nhập		

**Bảng 2.3:** Đặc tả use case Quên mật khẩu

### 2.2.3 ĐẶC TẢ USE CASE MƯỢN SÁCH

Bảng 2.4 đặc tả chi tiết cho use case Mượn sách. Use case này mô tả quá trình người dùng gửi yêu cầu mượn sách.

Tên use case	Mượn sách		
Tác nhân	Người dùng		
Tiền điều kiện	Không		
Luồng sự kiện chính	STT	Thực hiện bởi	Hành động
	1	Người dùng	Ấn chọn cuốn sách muốn mượn
	2	Hệ thống	Hiển thị thông tin sách
	3	Người dùng	Ấn chọn mượn sách
	4	Hệ thống	Hiển thị danh sách người cho mượn sách
	5	Người dùng	Chọn người cho mượn
	6	Hệ thống	Hiển thị ô nhập lời nhắn mượn sách (có lời nhắn mẫu)
	7	Người dùng	Nhập lời nhắn mượn sách
	8	Người dùng	Ấn chọn Gửi
Luồng sự kiện phát sinh	STT	Thực hiện bởi	Hành động
	4a	Hệ thống	Thông báo hiện tại không có người cho mượn và gợi ý người dùng đánh dấu yêu thích sách
	9a	Hệ thống	Hiển thị “Vui lòng nhập nội dung lời nhắn”
	9a1	Hệ thống	Thông báo bạn đang gửi yêu cầu mượn cuốn sách này rồi
Hậu điều kiện	Hệ thống thêm cuốn sách vào danh sách Sách đang yêu cầu mượn của người mượn và người cho mượn tương ứng		

Bảng 2.4: Đặc tả use case Mượn sách

### 2.2.4 ĐẶC TẢ USE CASE XÁC NHẬN CHO MƯỢN SÁCH

Bảng 2.5 đặc tả chi tiết cho use case Xác nhận cho mượn sách. Để tránh trường hợp người dùng mượn cùng lúc quá nhiều sách, đồng thời khuyến khích người dùng chia sẻ sách để cho mượn nhiều hơn, hệ thống giới hạn số sách người dùng được mượn cùng lúc theo 3 hạng. Thông tin về hạng được hiển thị trong phần thông tin tài khoản và mục quản lý sách mượn của người dùng.

- Hạng 1: được mượn cùng lúc 1 cuốn sách, là những người dùng đã cho mượn sách từ 0 đến 9 lần.
- Hạng 2: được mượn cùng lúc 2 cuốn sách, là những người dùng đã cho mượn

sách từ 10 đến 29 lần.

- Hạng 3: được mượn cùng lúc 3 cuốn sách, là những người dùng đã cho mượn sách từ 30 lần trở lên.

Với mỗi yêu cầu mượn sách, hệ thống tạo một mã xác nhận của yêu cầu và hiển thị cho người mượn. Khi gặp nhau để mượn sách, người mượn cung cấp mã này để người cho mượn xác nhận yêu cầu.

Tên use case	Xác nhận cho mượn sách		
Tác nhân	Người dùng		
Tiền điều kiện	Không		
Luồng sự kiện chính	STT	Thực hiện bởi	Hành động
	1	Người dùng	Ấn chọn mục Sách của bạn
	2	Người dùng	Ấn chọn Sách cho mượn
	3	Hệ thống	Hiển thị danh sách sách được yêu cầu mượn của người dùng
	4	Người dùng	Ấn chọn một cuốn sách
	5	Hệ thống	Hiển thị danh sách người yêu cầu mượn
	6	Người dùng	Ấn chọn đồng ý yêu cầu của một người mượn
	7	Hệ thống	Thông báo nhập mã xác nhận
	8	Người dùng	Nhập mã xác nhận
	9	Người dùng	Ấn chọn gửi
	10	Hệ thống	Thông báo xác nhận cho mượn thành công
Luồng sự kiện phát sinh	STT	Thực hiện bởi	Hành động
	10a	Hệ thống	Thông báo người dùng này hiện tại chỉ có thể mượn 1/2/3 cuốn sách
	10a1	Hệ thống	Thông báo mã xác nhận không chính xác
Hậu điều kiện	Hệ thống lưu cuốn sách vào danh sách đang cho mượn của người dùng và danh sách đang mượn của người mượn		

**Bảng 2.5:** ĐẶC TẢ USE CASE XÁC NHẬN CHO MƯỢN SÁCH

### 2.2.5 ĐẶC TẢ USE CASE TÌM KIẾM SÁCH CÓ Ở GẦN

Bảng 2.6 đặc tả chi tiết cho use case Tìm kiếm sách có ở gần. Trong chức năng này, người dùng có thể tìm kiếm theo tên sách, địa chỉ, khoảng cách và lựa chọn sắp xếp theo khoảng cách hoặc mức độ tương tự với sách của mình (các sách người dùng đang có, đã mượn, yêu thích). Địa chỉ tìm kiếm mặc định là địa chỉ của người dùng khi đăng ký tài khoản. Người dùng có thể chọn địa điểm trên bản đồ, tìm kiếm những địa chỉ khác mà mình mong muốn. Khoảng cách là bán kính khu vực xung quanh địa chỉ tìm kiếm, có giá trị mặc định là 3 km.

## CHƯƠNG 2. KHẢO SÁT VÀ PHÂN TÍCH YÊU CẦU

Tên use case	Tìm kiếm sách có ở gần		
Tác nhân	Người dùng		
Tiền điều kiện	Không		
Luồng sự kiện chính	STT	Thực hiện bởi	Hành động
	1	Người dùng	Ấn chọn Sách gần bạn
	2	Hệ thống	Hiển thị danh sách sách có ở gần địa chỉ của người dùng
	3	Người dùng	Nhập nội dung tìm kiếm
	4	Người dùng	Ấn chọn tìm kiếm
Luồng sự kiện phát sinh	5	Hệ thống	Hiển thị danh sách kết quả
Hậu điều kiện			

Bảng 2.6: ĐẶC TẢ USE CASE TÌM KIẾM SÁCH CÓ Ở GẦN

### 2.2.6 ĐẶC TẢ USE CASE NHẮN TIN TRÒ CHUYỆN

Bảng 2.7 đặc tả chi tiết cho use case Nhắn tin trò chuyện. Người dùng có thể sử dụng chức năng nhắn tin bằng cách chọn mục Tin nhắn trên menu hoặc chọn nút Nhắn tin trong trang thông tin tài khoản của một người dùng khác.

Tên use case	Nhắn tin trò chuyện		
Tác nhân	Người dùng		
Tiền điều kiện	Không		
Luồng sự kiện chính	STT	Thực hiện bởi	Hành động
	1	Người dùng	Ấn chọn mục Tin nhắn
	2	Hệ thống	Hiển thị danh sách các cuộc trò chuyện của người dùng
	3	Người dùng	Ấn chọn một cuộc trò chuyện
	4	Hệ thống	Hiển thị cuộc trò chuyện
	5	Người dùng	Nhập tin nhắn
	6	Người dùng	Ấn nút gửi
Luồng sự kiện phát sinh	7	Hệ thống	Gửi tin nhắn và hiển thị vào cuộc trò chuyện
	1a	Người dùng	Ấn chọn nhắn tin ở trang thông tin tài khoản của người dùng khác
	1b	Hệ thống	Hiển thị cuộc trò chuyện
Hậu điều kiện	Hệ thống lưu tin nhắn vào cuộc trò chuyện giữa hai người dùng		

Bảng 2.7: ĐẶC TẢ USE CASE NHẮN TIN TRÒ CHUYỆN

### **2.2.7 ĐẶC TẢ USE CASE THÊM SÁCH ĐANG CÓ**

Bảng 2.8 đặc tả chi tiết cho use case Thêm sách đang có. Đây là chức năng để người dùng thêm những cuốn sách mình đang có và có mong muốn cho mượn. Từ đó, những người dùng khác tìm kiếm cuốn sách này và có mong muốn mượn có thể gửi yêu cầu mượn đến người sở hữu.

Tên use case	Thêm sách đang có		
Tác nhân	Người dùng		
Tiền điều kiện	Không		
Luồng sự kiện chính	STT	Thực hiện bởi	Hành động
	1	Người dùng	Ấn chọn mục Sách của bạn
	2	Người dùng	Ấn chọn Sách đang có
	3	Hệ thống	Hiển thị danh sách sách đang có của người dùng
	4	Người dùng	Ấn chọn Thêm sách
	5	Hệ thống	Hiển thị giao diện tìm kiếm, nhập tên sách
	6	Người dùng	Nhập tên sách
	7	Hệ thống	Hiển thị danh sách tìm kiếm với tên người dùng nhập
	8	Người dùng	Ấn chọn cuốn sách muốn thêm
	9	Hệ thống	Thông báo thêm sách thành công
Luồng sự kiện phát sinh	STT	Thực hiện bởi	Hành động
	7a	Hệ thống	Hiển thị không có sách tương ứng và nút “Thêm sách chưa có trên hệ thống”
	7b	Người dùng	Ấn chọn “Thêm sách chưa có trên hệ thống”
	7c	Hệ thống	Hiển thị giao diện “Thêm sách chưa có trên hệ thống”
	9a	Hệ thống	Thông báo “Bạn đã thêm cuốn sách này rồi”
Hậu điều kiện	Hệ thống lưu cuốn sách vào danh sách sách đang có của người dùng		

**Bảng 2.8: Đặc tả use case Thêm sách đang có**

### **2.2.8 ĐẶC TẢ USE CASE THÊM SÁCH HỆ THỐNG CHƯA CÓ**

Bảng 2.9 và 2.10 trình bày đặc tả chi tiết và mô tả dữ liệu đầu vào cho use case Thêm sách hệ thống chưa có. Đây là chức năng để người dùng thêm các cuốn sách mà hệ thống chưa có thông tin. Khi người dùng tìm kiếm một cuốn sách để đánh dấu là mình đang có mà không thấy thông tin, hệ thống cũng gợi ý người dùng sử dụng chức năng này.

## CHƯƠNG 2. KHẢO SÁT VÀ PHÂN TÍCH YÊU CẦU

Tên use case	Thêm sách hệ thống chưa có		
Tác nhân	Người dùng		
Tiền điều kiện	Không		
Luồng sự kiện chính	STT	Thực hiện bởi	Hành động
	1	Người dùng	Ấn chọn mục Sách của bạn
	2	Người dùng	Ấn chọn mục Sách tự thêm
	3	Hệ thống	Hiển thị danh sách sách người dùng đã thêm vào hệ thống
	4	Người dùng	Ấn nút Thêm sách
	5	Hệ thống	Hiển thị giao diện nhập thông tin sách
	6	Người dùng	Nhập thông tin sách*
	7	Người dùng	Ấn nút Lưu
	8	Hệ thống	Thông báo thêm sách thành công
	9	Hệ thống	Hỏi người dùng có muốn thêm sách vào danh sách sách đang có
	10	Người dùng	Ấn chọn Đồng ý
	11	Hệ thống	Thông báo thêm sách thành công
Luồng sự kiện phát sinh	STT	Thực hiện bởi	Hành động
	8a	Hệ thống	Hiển thị các trường dữ liệu không hợp lệ
	10a	Người dùng	Ấn chọn đóng
Hậu điều kiện	Hệ thống lưu thông tin sách và thêm vào danh sách sách người dùng đã thêm, đang có		

**Bảng 2.9:** Đặc tả use case Thêm sách hệ thống chưa có

\* Mô tả dữ liệu đầu vào

Trường dữ liệu	Bắt buộc?	Điều kiện hợp lệ	Ví dụ
Tên sách	Có		Tin học đại cương
Mô tả sách	Có		Tài liệu học tập và tham khảo cho sinh viên các trường đại học kỹ thuật, công nghệ

Tác giả	Có		Trần Đình Khang
Thể loại	Có		Giáo dục
Ảnh	Có	Định dạng png, jpg, jpeg. Kích thước < 2 MB	
Nhà xuất bản	Không		NXB Bách Khoa Hà Nội
Thời gian xuất bản	Không		2022
Số trang	Không		246

**Bảng 2.10:** Dữ liệu đầu vào trong use case Thêm sách hệ thống chưa có

### 2.3 Yêu cầu phi chức năng

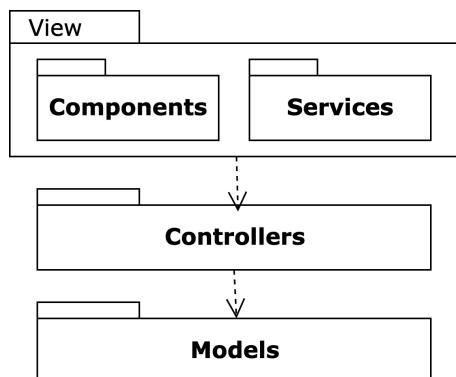
Hệ thống cần đáp ứng một số yêu cầu phi chức năng về bảo mật, giao diện sử dụng. Mật khẩu của người dùng cần được mã hóa trước khi lưu vào cơ sở dữ liệu. Giao diện sử dụng cần có thiết kế thông nhất, trong đó án này sử dụng các khôi giao diện của Ant Design (<https://ng.ant.design>). Ngôn ngữ sử dụng là tiếng Việt, định dạng ngày tháng theo thứ tự ngày/tháng/năm.

## CHƯƠNG 3. THIẾT KẾ HỆ THỐNG VÀ KẾT QUẢ THỰC NGHIỆM

### 3.1 Thiết kế kiến trúc

#### 3.1.1 Thiết kế tổng quan

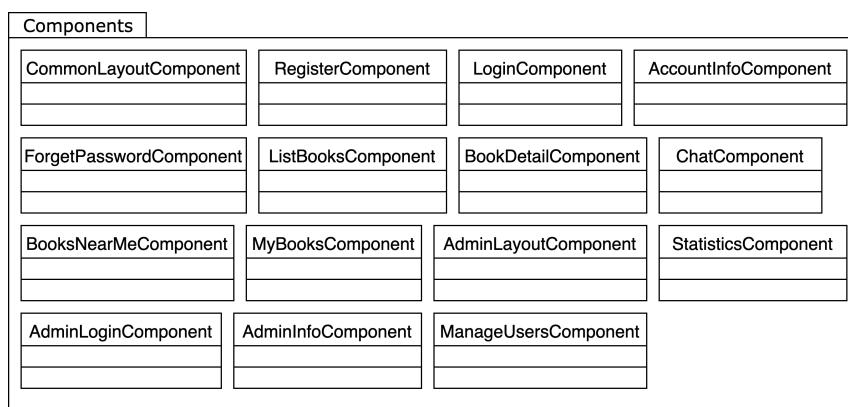
Hình 3.1 là biểu đồ gói thiết kế tổng quan cho hệ thống. Hệ thống được thiết kế theo kiến trúc MVC (Model - View - Controller) với phần Model và Controller thuộc backend và View thuộc frontend. Model là các lớp tương ứng với các thực thể trong cơ sở dữ liệu, Controller là các lớp tạo ra các Web API để giao tiếp với frontend. Phần View bao gồm các component và các service. Các component tương ứng với các khôi giao diện tương tác với người dùng. Các service được sử dụng để gọi đến các API tương ứng ở backend.



**Hình 3.1:** Biểu đồ gói thiết kế tổng quan

#### 3.1.2 Thiết kế chi tiết gói

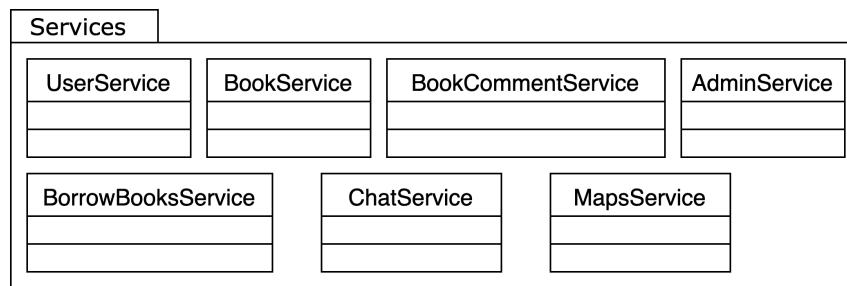
Hình 3.2 là thiết kế chi tiết gói Components với các lớp tương ứng với các khôi giao diện tương tác với người dùng. CommonLayoutComponent và AdminLayoutComponent là hai lớp đặc biệt, được sử dụng làm khung, bao gồm menu và phần thân để chứa các khôi giao diện dành cho khách, người dùng và quản trị viên.



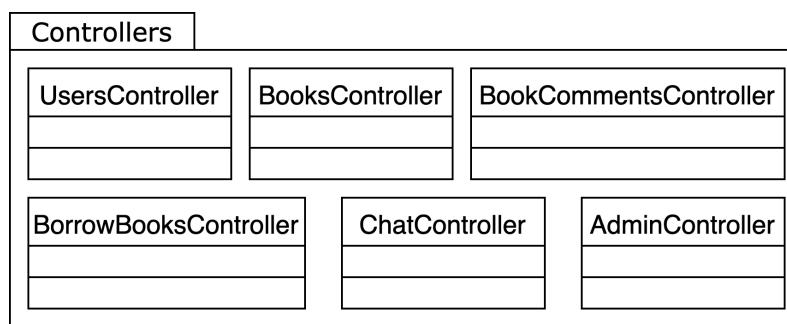
**Hình 3.2:** Chi tiết gói Components

### CHƯƠNG 3. THIẾT KẾ HỆ THỐNG VÀ KẾT QUẢ THỰC NGHIỆM

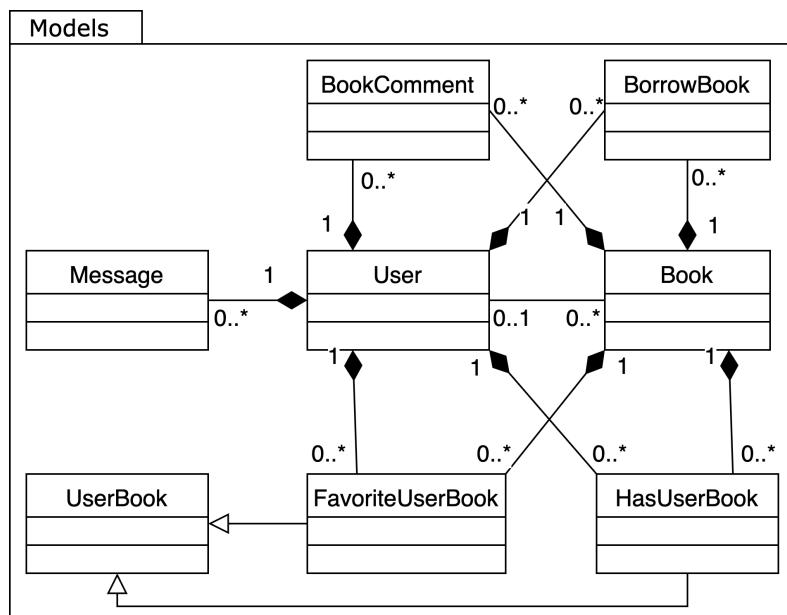
Hình 3.3 và hình 3.4 lần lượt là thiết kế chi tiết gói Services và gói Controllers. Có sự tương ứng giữa các lớp trong hai gói với nhau. Các lớp thuộc gói Controllers cài đặt các logic, tạo ra các Web API, các lớp trong gói Services được sử dụng để gọi đến các API này. Ngoài ra, gói Services có thêm lớp MapsService, dùng để gọi đến các API chuyển đổi giữa tọa độ và địa chỉ. Các API này được cung cấp bởi dịch vụ tích hợp bản đồ tại Việt Nam Goong Maps được mô tả trong bảng 3.11.



**Hình 3.3:** Chi tiết gói Services



**Hình 3.4:** Chi tiết gói Controllers



**Hình 3.5:** Chi tiết gói Models

## CHƯƠNG 3. THIẾT KẾ HỆ THỐNG VÀ KẾT QUẢ THỰC NGHIỆM

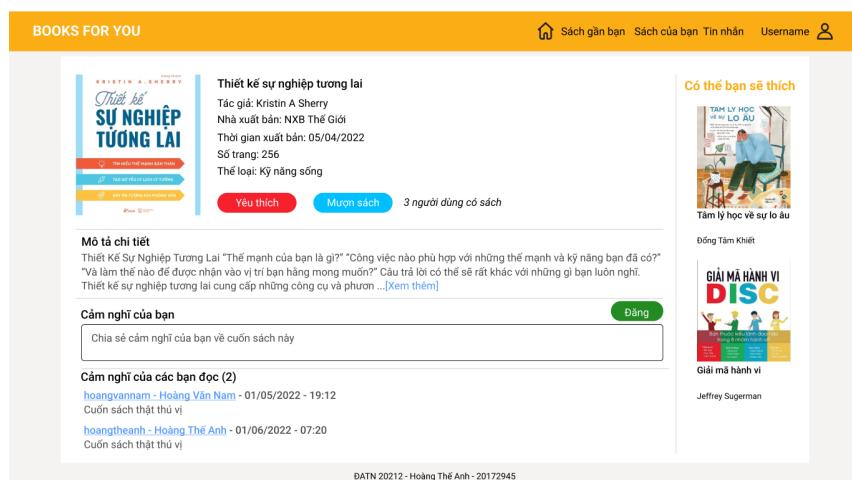
Hình 3.5 là thiết kế chi tiết gói Models. Các lớp trong gói Models tương ứng với các thực thể trong cơ sở dữ liệu.

### 3.2 Thiết kế chi tiết

#### 3.2.1 Thiết kế giao diện

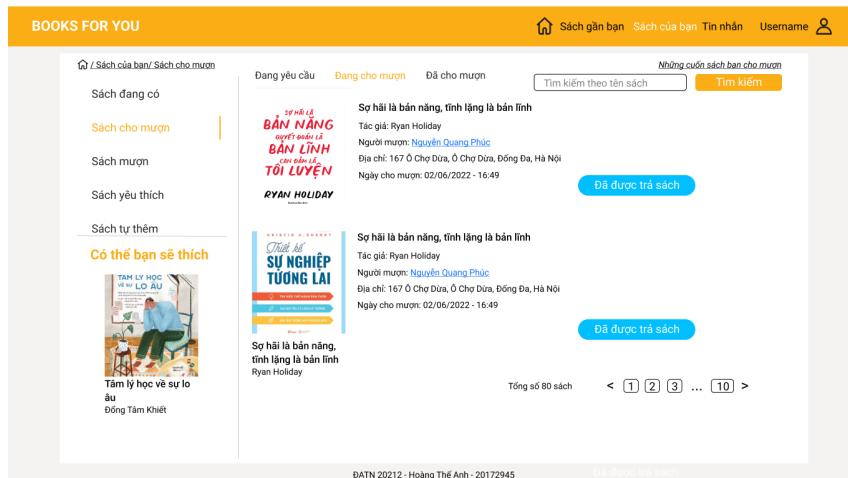
Hệ thống hướng tới thiết kế ban đầu cho màn hình máy tính với kích thước 1366x768 px. Phần giao diện chính có chiều rộng 1200 px, các phần còn lại được thu phóng theo kích thước màn hình hiển thị. Giao diện chính sử dụng nền trắng, màu chủ đạo là màu cam với mã màu #FAAD14. Các thông báo phản hồi được hiển thị phía trên, bên phải của màn hình.

Các hình 3.6, 3.7, 3.8, 3.9, 3.10 là giao diện thiết kế cho một số chức năng chính dành cho người dùng như: Xem thông tin sách, Quản lý sách cá nhân, Thêm sách chưa có trên hệ thống, Tìm kiếm sách có ở gần, Nhắn tin trò chuyện. Các hình 3.11, 3.12 là giao diện thiết kế cho chức năng quản lý người dùng và xem thống kê hệ thống của quản trị viên.

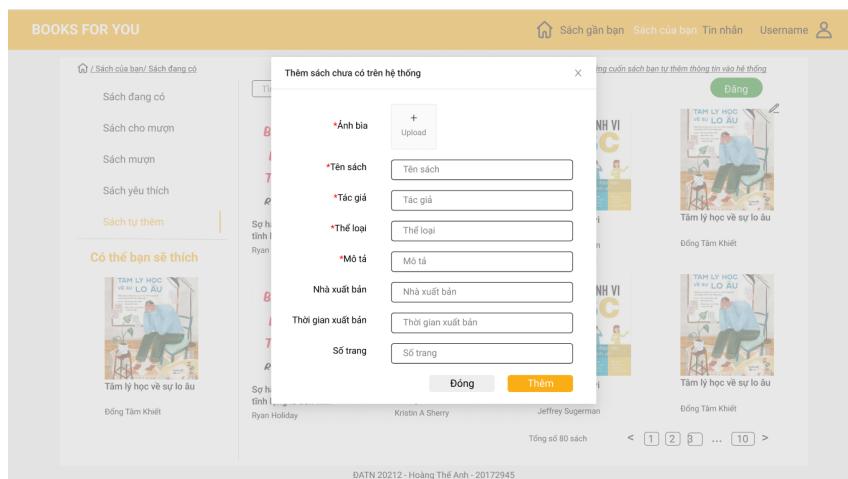


**Hình 3.6:** Giao diện thiết kế màn hình thông tin sách

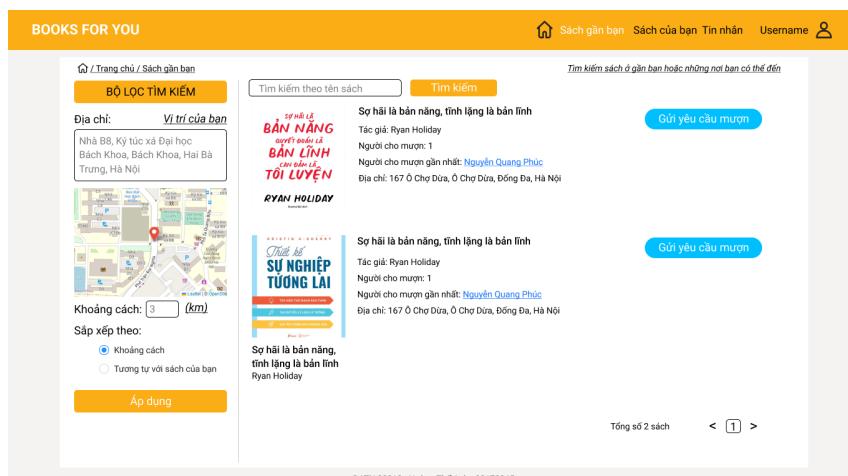
### CHƯƠNG 3. THIẾT KẾ HỆ THỐNG VÀ KẾT QUẢ THỰC NGHIỆM



**Hình 3.7:** Giao diện thiết kế màn hình quản lý sách cá nhân

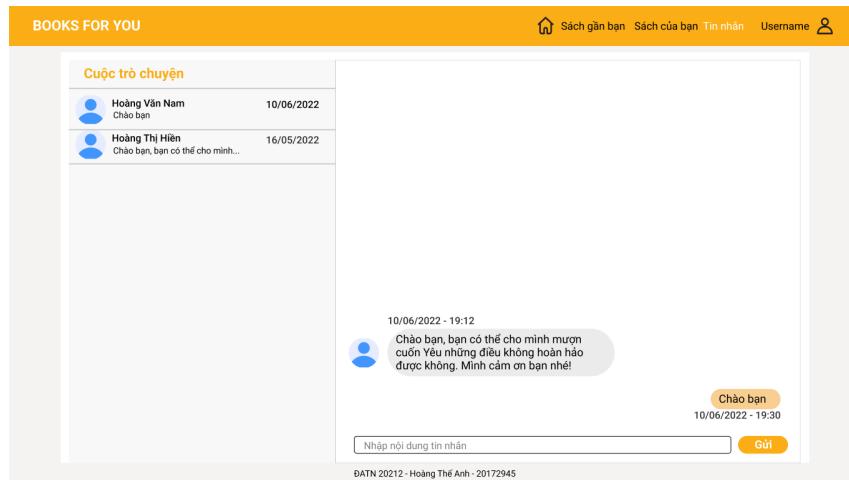


**Hình 3.8:** Giao diện thiết kế màn hình thêm sách chưa có trên hệ thống



**Hình 3.9:** Giao diện thiết kế màn hình tìm kiếm sách có ở gần

### CHƯƠNG 3. THIẾT KẾ HỆ THỐNG VÀ KẾT QUẢ THỰC NGHIỆM



**Hình 3.10:** Giao diện thiết kế màn hình nhắn tin

The screenshot shows a web-based application interface for managing users. The header bar includes 'BOOKS FOR YOU', 'Thống kê hệ thống', 'Quản lý tài khoản', 'Admin', and a user profile icon. Below the header, a breadcrumb navigation shows '/ Quản trị viên / Quản lý tài khoản'. A search bar with placeholder 'Tim kiếm theo tên đăng nhập' and a 'Tim kiếm' button are on the right. The main content area displays a table titled 'Danh sách người dùng (100 users)'. The table has columns for 'STT', 'Họ và tên', 'Tên đăng nhập', 'Email', 'Địa chỉ', 'Trạng thái', and 'Khóa'. There are 7 rows of data, each representing a user named 'Người dùng mới' with different login names ('username1' to 'username7') and email addresses. The 'Trạng thái' column shows various states like 'Hoạt động', 'Khóa', and 'Bị khóa', with some entries in red or green boxes. At the bottom of the table, there are navigation buttons for page numbers (1, 2, 3, ..., 10) and arrows, along with a footer note 'DATN 20212 - Hoàng Thế Anh - 20172945'.

**Hình 3.11:** Giao diện thiết kế màn hình quản lý người dùng



**Hình 3.12:** Giao diện thiết kế màn hình thống kê hệ thống

### **3.2.2 Thiết kế lớp**

Trong phần này, em trình bày thiết kế chi tiết cho một số lớp quan trọng thuộc gói Controller. Các lớp này thể hiện cho một số API cần thiết của hệ thống.

Bảng 3.1 trình bày thiết kế lớp BooksController. Lớp này bao gồm các phương thức về thông tin sách, sách yêu thích, gợi ý sách tương tự và sách ở gần.

Tên lớp	BooksController			
Mô tả	Controller về thông tin sách, sách yêu thích, sách tương tự và sách ở gần			
Phương thức	Tên phương thức	Đầu vào	Đầu ra	Mô tả
	GetBook	Id	BookResponse <sup>1</sup>	Lấy thông tin chi tiết sách
	GetListBooks	Name, Cate- gories, Page, PageSize	List <Book- Response <sup>1</sup> >, Count	Lấy danh sách sách
	GetListMyBooks	Name, Page, Pa- geSize	List <Book- Response <sup>1</sup> >, Count	Lấy danh sách sách đang có
	AddMyBook	BookId	Boolean	Thêm sách đang có
	RemoveMyBook	BookId	Boolean	Xóa sách đang có
	GetList- MyFavorite	Name, Page, Pa- geSize	List <Book- Response <sup>1</sup> >, Count	Lấy danh sách sách yêu thích
	AddMyFavorite	BookId	Boolean	Thêm sách yêu thích
	Remove- MyFavorite	BookId	Boolean	Xóa sách yêu thích
	GetList- MyselfAdded	Name, Page, Pa- geSize	List <Book- Response <sup>1</sup> >, Count	Lấy danh sách sách tự thêm thông tin
	AddMyselfAdded	BookRequest <sup>2</sup>	BookId	Thêm thông tin sách
	Update- MyselfAdded	BookRequest <sup>2</sup> , BookId	Boolean	Cập nhật thông tin sách
	Upload- BookImage	ImageFile	Boolean	Upload file ảnh bìa sách
	GetListBook- Recommend	BookIds	List <Book- Response <sup>1</sup> >	Lấy danh sách sách gợi ý tương tự

### CHƯƠNG 3. THIẾT KẾ HỆ THỐNG VÀ KẾT QUẢ THỰC NGHIỆM

	GetBooksNearMe	Latitude, Longtitude, Distance, SortBy	List<GetBooksNearMeResponse <sup>3</sup> >	Lấy danh sách sách ở gần
--	----------------	--	--	--------------------------

**Bảng 3.1:** Thiết kế lớp BooksController.

<sup>1</sup> BookResponse: Id, Name, Author, Description, Category, ImageUrl, Publisher, PublishTime, Pages, IsUserFavorite, NUserFavorite, Website, Url

<sup>2</sup> BookRequest: Name, Author, Description, Category, ImageUrl, Publisher, PublishTime, Pages

<sup>3</sup> GetBooksNearMeResponse: Id, Name, Author, ImageUrl, NUserLend, NearestUserId, NearestUserName, NearestUserAddress, NearestUserPhoneNumber

Bảng 3.2 trình bày thiết kế lớp BorrowBooksController. Lớp này bao gồm các phương thức về người dùng có sách, yêu cầu mượn sách và cho mượn sách.

Tên lớp	BorrowBooksController			
Mô tả	Controller về người dùng có sách, yêu cầu mượn sách và cho mượn sách			
Phương thức	Tên phương thức	Đầu vào	Đầu ra	Mô tả
	GetUsersHasBook	BookId	List<User-Response <sup>1</sup> >, Count	Lấy danh sách người dùng có sách
	Request-BorrowBook	LendingUserId, BookId, Message	RequestedCode	Gửi yêu cầu mượn sách
	GetListBorrow	Name, Status, Page, PageSize	List<RequestIn-BorrowResponse <sup>2</sup> >, Count	Lấy danh sách sách mượn
	Cancel-BorrowBook	BorrowId	Boolean	Hủy yêu cầu mượn sách
	GetListLend	Name, Status, Page, PageSize	List <RequestIn-BorrowResponse <sup>2</sup> >, Count	Lấy danh sách sách cho mượn
	ConfirmLend	BorrowId, RequestedCode	Boolean	Xác nhận cho mượn sách
	ConfirmReturned	BorrowId	Boolean	Xác nhận được trả sách

**Bảng 3.2:** Thiết kế lớp BorrowBooksController.

<sup>1</sup> UserResponse: Id, Username, FullName, Email, Address, PhoneNumber, Avatar, AvatarPublicId, Latitude, Longitude

<sup>2</sup> RequestInBorrowResponse: BookId, BookName, BookAuthor, BookImage, Users, DateCreated, DateUpdated

### CHƯƠNG 3. THIẾT KẾ HỆ THỐNG VÀ KẾT QUẢ THỰC NGHIỆM

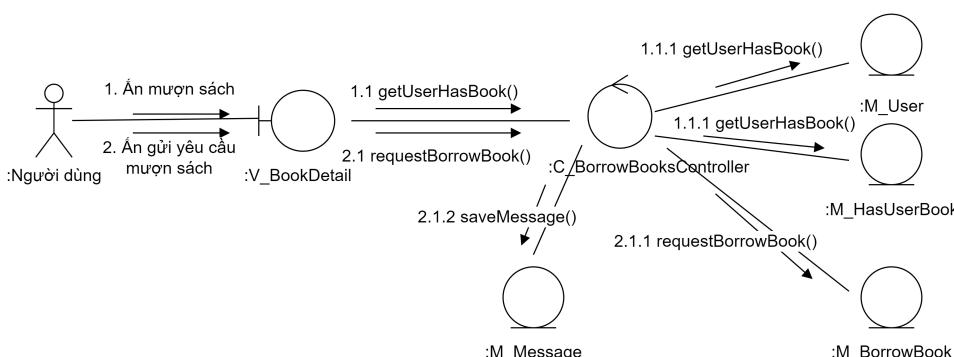
Bảng 3.3 trình bày thiết kế lớp ChatController. Lớp này bao gồm các phương thức cho chức năng nhắn tin trò chuyện.

Tên lớp	ChatController			
Mô tả	Controller cho chức năng nhắn tin trò chuyện			
Phương thức	Tên phương thức	Đầu vào	Đầu ra	Mô tả
	GetConversations		List <Conversation-Response <sup>1</sup> >	Lấy danh sách cuộc trò chuyện
	GetMessages	WithUserId	List <Message <sup>2</sup> >	Lấy tin nhắn của cuộc trò chuyện
	SendMessage	ToUserId, Content	Boolean	Gửi tin nhắn
	SeenMessage	MessageId	Boolean	Đánh dấu đã xem tin nhắn
	CountUnseen- Conversations		Int	Đếm số cuộc trò chuyện chưa xem

**Bảng 3.3:** Thiết kế lớp ChatController.

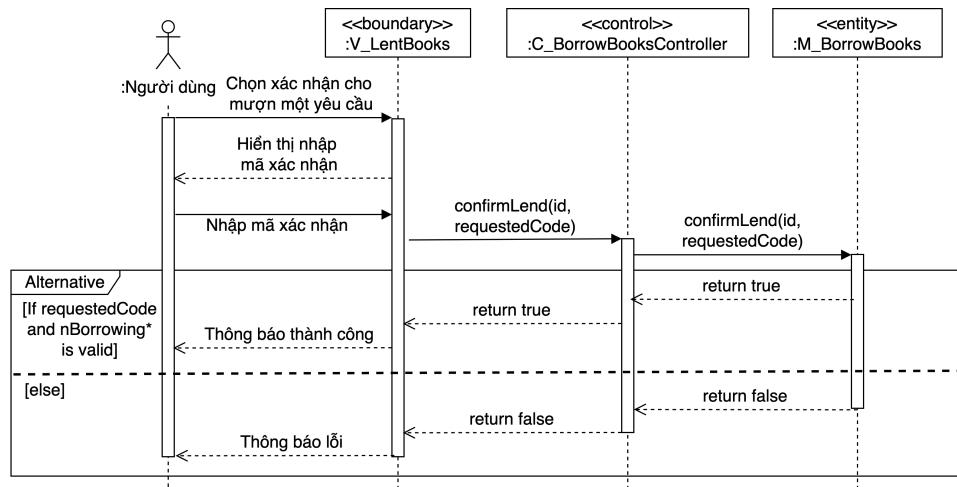
- <sup>1</sup> ConversationResponse: UserId, Username, FullName, Avatar, LastMessageId, LastMessage, LastMessageDate, IsLastOfMe, IsSeen  
<sup>2</sup> Message: Id, FromUserId, ToUserId, Content, DateCreated, IsSeen

Hình 3.13 và hình 3.14 là biểu đồ giao tiếp và biểu đồ trình tự thể hiện luồng sự kiện chính cho các use case “Muốn sách”, “Xác nhận cho muộn sách”, “Hủy yêu cầu muộn” và “Đánh dấu sách đã được trả” minh họa cho thiết kế lớp Borrow-BooksController.

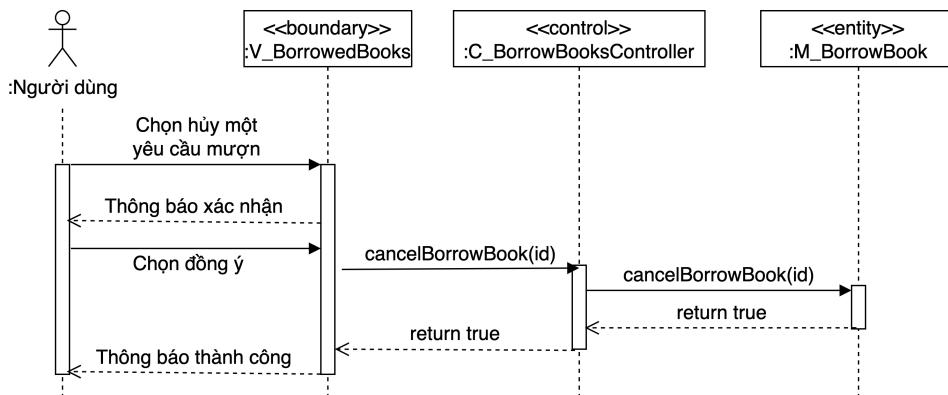


**Hình 3.13:** Biểu đồ giao tiếp cho use case Muốn sách

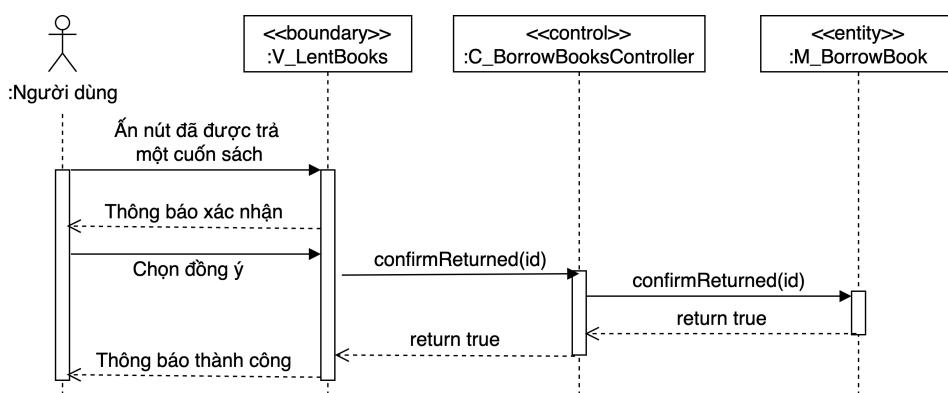
### CHƯƠNG 3. THIẾT KẾ HỆ THỐNG VÀ KẾT QUẢ THỰC NGHIỆM



**Hình 3.14:** Biểu đồ trình tự cho use case Xác nhận cho mượn sách

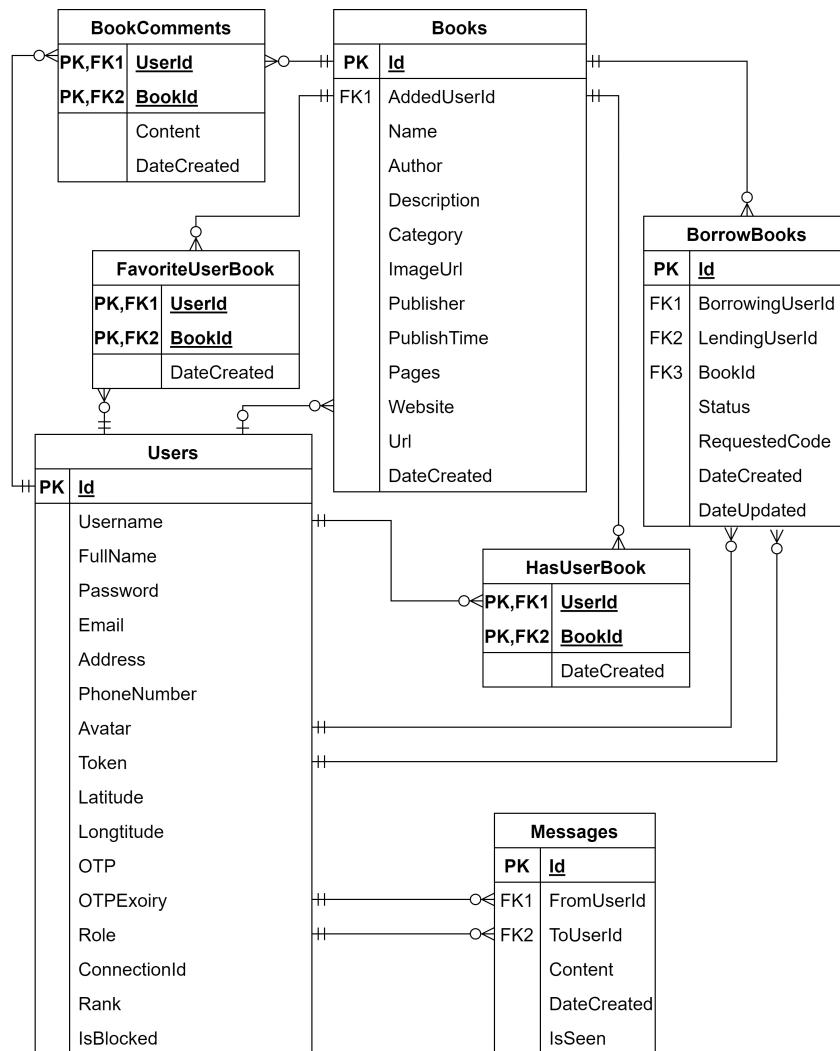


**Hình 3.15:** Biểu đồ trình tự cho use case Hủy yêu cầu mượn



**Hình 3.16:** Biểu đồ trình tự cho use case Đánh dấu sách đã được trả

### 3.2.3 Thiết kế cơ sở dữ liệu



Hình 3.17: Sơ đồ thiết kế cơ sở dữ liệu

Từ sơ đồ thiết kế cơ sở dữ liệu hình 3.17, em thiết kế các bảng cho cơ sở dữ liệu quan hệ SQL như sau.

#### Bảng dữ liệu "Books"

Ý nghĩa: Danh sách các cuốn sách

Thuộc tính	Kiểu dữ liệu	Mô tả	Bắt buộc	Ràng buộc
Id	Int	Mã sách	Có	Khóa chính
AddedUserId	Int	Mã người dùng thêm thông tin sách	Không	Khóa ngoại
Name	String	Tên sách	Có	
Category	String	Thể loại	Có	
Author	String	Tác giả	Không	
Description	String	Mô tả	Không	

### CHƯƠNG 3. THIẾT KẾ HỆ THỐNG VÀ KẾT QUẢ THỰC NGHIỆM

ImageUrl	String	Url ảnh sách	Không	
Publisher	String	Nhà xuất bản	Không	
PublishTime	Datetime	Thời gian xuất bản	Không	
Pages	Int	Số trang	Không	
Website	String	Website thu thập sách	Không	
Url	String	Url thu thập sách	Không	
DateCreated	Datetime	Ngày lưu	Có	

**Bảng 3.4:** Thiết kế cơ sở dữ liệu bảng "Books".

#### Bảng dữ liệu "BookComments"

Ý nghĩa: Danh sách các đánh giá sách

Thuộc tính	Kiểu dữ liệu	Mô tả	Bắt buộc	Ràng buộc
UserId	Int	Mã người dùng đánh giá	Có	Khóa chính, khóa ngoại
BookId	Int	Mã sách	Có	Khóa chính, Khóa ngoại
Content	String	Nội dung	Có	
DateCreated	Datetime	Ngày lưu	Có	

**Bảng 3.5:** Thiết kế cơ sở dữ liệu bảng "BookComments".

#### Bảng dữ liệu "BorrowBooks"

Ý nghĩa: Danh sách các lượt mượn, yêu cầu mượn sách

Thuộc tính	Kiểu dữ liệu	Mô tả	Bắt buộc	Ràng buộc
Id	Int	Mã lượt mượn sách	Có	Khóa chính
Borrowing- UserId	Int	Mã người mượn	Có	Khóa ngoại
LendingUserId	Int	Mã người cho mượn	Có	Khóa ngoại
BookId	Int	Mã sách	Có	Khóa ngoại
Status	Int	Trạng thái lượt mượn sách	Có	
RequestedCode	String	Mã xác nhận yêu cầu	Có	
DateCreated	Datetime	Ngày lưu	Có	
DateUpdated	Datetime	Ngày cập nhật trạng thái	Có	

**Bảng 3.6:** Thiết kế cơ sở dữ liệu bảng "BorrowBooks".

#### Bảng dữ liệu "FavoriteUserBook"

Ý nghĩa: Danh sách người dùng yêu thích sách

### CHƯƠNG 3. THIẾT KẾ HỆ THỐNG VÀ KẾT QUẢ THỰC NGHIỆM

Thuộc tính	Kiểu dữ liệu	Mô tả	Bắt buộc	Ràng buộc
UserId	Int	Mã người dùng	Có	Khóa chính, khóa ngoại
BookId	Int	Mã sách	Có	Khóa chính, khóa ngoại
DateCreated	Datetime	Ngày lưu	Có	

**Bảng 3.7:** Thiết kế cơ sở dữ liệu bảng "FavoriteUserBook".

#### Bảng dữ liệu "HasUserBook"

Ý nghĩa: Danh sách người dùng có sách

Thuộc tính	Kiểu dữ liệu	Mô tả	Bắt buộc	Ràng buộc
UserId	Int	Mã người dùng	Có	Khóa chính, khóa ngoại
BookId	Int	Mã sách	Có	Khóa chính, khóa ngoại
DateCreated	Datetime	Ngày lưu	Có	

**Bảng 3.8:** Thiết kế cơ sở dữ liệu bảng "HasUserBook".

#### Bảng dữ liệu "Users"

Ý nghĩa: Danh sách người dùng

Thuộc tính	Kiểu dữ liệu	Mô tả	Bắt buộc	Ràng buộc
Id	Int	Mã người dùng	Có	Khóa chính
Username	String	Tên đăng nhập	Có	Duy nhất
FullName	String	Tên đầy đủ	Có	
Password	String	Mã băm mật khẩu	Có	
Email	String	Địa chỉ email	Có	Duy nhất
Address	String	Địa chỉ người dùng	Có	
PhoneNumber	String	Số điện thoại	Không	
Avatar	String	Url ảnh đại diện	Không	
Token	String	Mã xác thực phiên đăng nhập	Không	
Latitude	Float	Vĩ độ theo địa chỉ người dùng	Có	
Longitude	Float	Kinh độ theo địa chỉ người dùng	Có	
OTP	String	Mã xác thực email khi quên mật khẩu	Không	

### CHƯƠNG 3. THIẾT KẾ HỆ THỐNG VÀ KẾT QUẢ THỰC NGHIỆM

OTPExpiry	Datetime	Thời điểm OTP hết hạn	Không	
Role	String	Vai trò người dùng hoặc quản trị viên	Có	
ConnectionId	String	Mã kết nối để gửi tin nhắn, thông báo	Không	
Rank	Int	Hạng của người dùng	Có	
IsBlocked	Boolean	Tài khoản có bị khóa hay không	Có	

Bảng 3.9: Thiết kế cơ sở dữ liệu bảng "Users".

#### Bảng dữ liệu "Messages"

Ý nghĩa: Danh sách các tin nhắn của người dùng

Thuộc tính	Kiểu dữ liệu	Mô tả	Bắt buộc	Ràng buộc
Id	Int	Mã tin nhắn	Có	Khóa chính
FromUserId	Int	Mã người gửi	Có	Khóa ngoại
ToUserId	Int	Mã người nhận	Có	Khóa ngoại
Content	Int	Nội dung tin nhắn	Có	
DateCreated	Datetime	Ngày lưu	Có	
IsSeen	Boolean	Tin nhắn đã được xem hay chưa	Có	

Bảng 3.10: Thiết kế cơ sở dữ liệu bảng "Messages".

### 3.3 Xây dựng ứng dụng

#### 3.3.1 Thư viện và công cụ sử dụng

Bảng 3.11 trình bày các thư viện và công cụ em sử dụng trong quá trình xây dựng hệ thống.

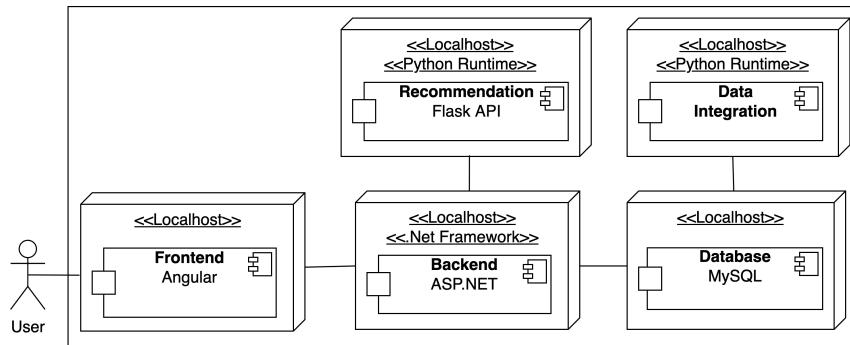
Mục đích	Công cụ	Phiên bản	Địa chỉ URL
IDE lập trình	Visual Studio Code	1.68	code.visualstudio.com
IDE lập trình	Visual Studio	Community 2019	visualstudio.microsoft.com
Xây dựng frontend	Angular	13.3.0	angular.io
Xây dựng giao diện	Ant Design	13.1.1	ng.ant.design
Sử dụng bản đồ	Leaflet	1.8.0	leafletjs.com
Vẽ biểu đồ	Chart.js	3.6.0	chartjs.org
Chuyển đổi giữa tọa độ và địa chỉ	Goong Maps	1.0.9	goong.io
Lưu trữ ảnh	CloudinaryDotNet	1.17.0	cloudinary.com

Xây dựng backend	.NET	5.0	dotnet.microsoft.com/en-us/download/dotnet/5.0
Xây dựng chức năng nhắn tin	SignalR	1.1.0	dotnet.microsoft.com/en-us/apps/aspnet/signalr
Cơ sở dữ liệu	MySQL	8.0.26	mysql.com
Thu thập dữ liệu	Scrapy	2.6.1	scrapy.org
Tiền xử lý văn bản	Pyvi	0.1.1	github.com/trungtv/pyvi
Đối sánh dữ liệu trùng lắp	Python Record Linkage Toolkit	0.15	recordlinkage.readthedocs.io
Huấn luyện mô hình Word2Vec	Gensim	4.2.0	radimrehurek.com/gensim
Xây dựng API gợi ý	Flask	2.0.2	flask.palletsprojects.com

**Bảng 3.11:** Danh sách thư viện và công cụ sử dụng

### 3.3.2 Kết quả triển khai

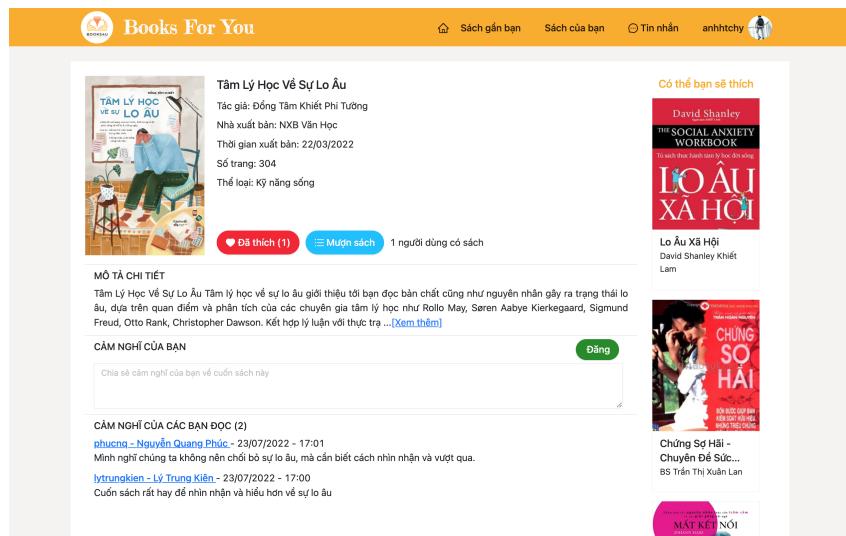
Sau khi hoàn thành, hệ thống bao gồm 4 thành phần là thu thập dữ liệu, frontend, backend và gợi ý. Các thành phần được triển khai trên máy tính cá nhân cùng cơ sở dữ liệu MySQL theo sơ đồ hình 3.18.



**Hình 3.18:** Sơ đồ triển khai hệ thống

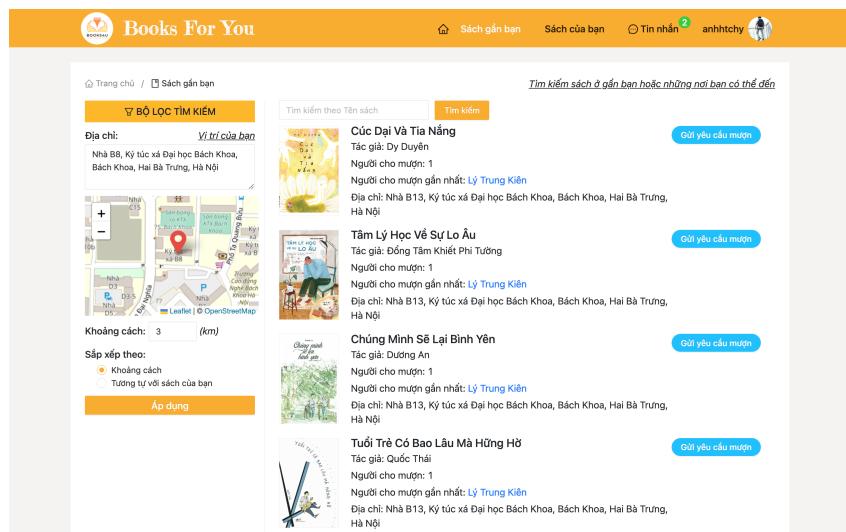
Hình 3.19 minh họa màn hình trang thông tin sách sau khi hoàn thành. Trang này hiển thị ảnh bìa, các thông tin về sách, phần đăng cảm nghĩ và xem cảm nghĩ của các bạn đọc khác. Các phần văn bản mô tả, cảm nghĩ nếu quá dài sẽ được hiển thị một phần và có nút "Xem thêm". Nút mượn sách hiển thị danh sách người dùng có sách. Phần có thể bạn sẽ thích hiển thị một số cuốn sách tương tự với cuốn sách đang được xem.

### CHƯƠNG 3. THIẾT KẾ HỆ THỐNG VÀ KẾT QUẢ THỰC NGHIỆM



Hình 3.19: Minh họa màn hình thông tin sách

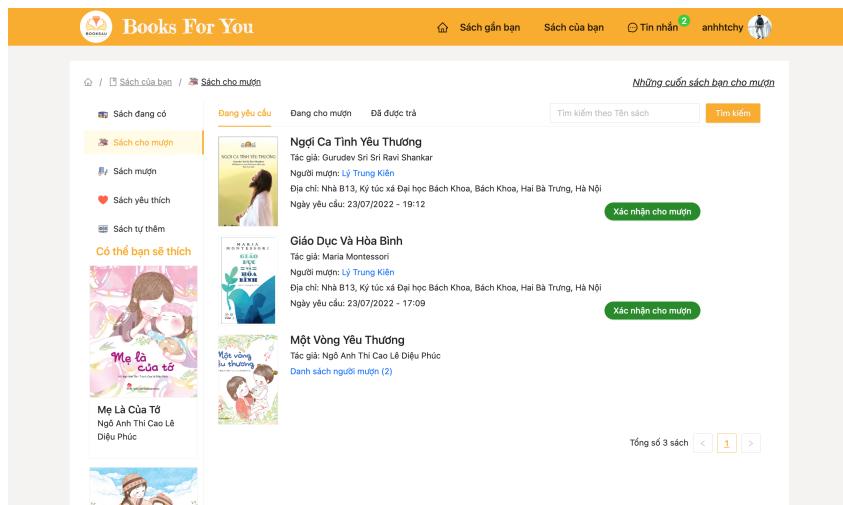
Hình 3.20 minh họa màn hình trang sách gần bạn sau khi hoàn thành. Người dùng có thể tìm kiếm sách ở gần mình và những nơi mình có thể đến theo địa chỉ và khoảng cách xung quanh địa chỉ mong muốn. Kết quả tìm kiếm được sắp xếp theo khoảng cách hoặc mức độ tương tự với sách mà người dùng quan tâm. Địa chỉ tìm kiếm có thể lựa chọn trực tiếp trên bản đồ.



Hình 3.20: Minh họa màn hình sách gần bạn

Hình 3.21 minh họa màn hình trang sách của bạn, mục sách cho mượn sau khi hoàn thành. Phần có thể bạn sẽ thích hiển thị các cuốn sách tương tự với các cuốn sách đang hiển thị trong mục hiện tại. Khi xem sách đang được yêu cầu mượn, với mỗi cuốn sách, nếu chỉ có một người dùng yêu cầu, màn hình hiển thị trực tiếp nút xác nhận cho mượn, nếu có nhiều hơn, màn hình sẽ hiển thị danh sách người mượn.

### CHƯƠNG 3. THIẾT KẾ HỆ THỐNG VÀ KẾT QUẢ THỰC NGHIỆM



**Hình 3.21:** Minh họa màn hình sách của bạn

#### 3.4 Kiểm thử chức năng

##### 3.4.1 Kịch bản kiểm thử chức năng “Thêm sách đang có”

Mô tả	Kịch bản	Kết quả mong muốn	Kết quả
Tìm kiếm sách hệ thống có thông tin	1. Truy cập giao diện sách đang có 2. Ấn thêm sách mới 3. Nhập tên sách hệ thống có thông tin và ấn tìm kiếm	Hiển thị các sách phù hợp	Đạt
Tìm kiếm sách hệ thống chưa có thông tin	1. Truy cập giao diện sách đang có 2. Ấn thêm sách mới 3. Nhập tên sách hệ thống chưa có thông tin và ấn tìm kiếm	Hiển thị không có dữ liệu và gợi ý người dùng sử dụng chức năng Thêm sách hệ thống chưa có	Đạt
Thêm sách đang có hợp lệ	1. Truy cập giao diện sách đang có 2. Ấn thêm sách mới 3. Tìm kiếm và ấn thêm một cuốn sách chưa từng thêm trước đó	Thêm sách vào danh sách sách đang có của người dùng	Đạt
Thêm sách đã từng thêm trước đó	1. Truy cập giao diện sách đang có 2. Ấn thêm sách mới	Thông báo bạn đã thêm cuốn sách này rồi	Đạt

### CHƯƠNG 3. THIẾT KẾ HỆ THỐNG VÀ KẾT QUẢ THỰC NGHIỆM

	3. Tìm kiếm và ấn thêm một cuốn sách đã thêm trước đó		
--	---	--	--

**Bảng 3.12:** Kịch bản kiểm thử chức năng “Thêm sách đang có”

#### 3.4.2 Kịch bản kiểm thử chức năng “Mượn sách”

Mô tả	Kịch bản	Kết quả mong muốn	Kết quả
Hiển thị danh sách người cho mượn	1. Truy cập giao diện thông tin chi tiết một cuốn sách có nhiều người cho mượn 2. Ấn nút mượn sách	Hiển thị các người dùng có sách, sắp xếp theo khoảng cách tăng dần đến địa chỉ người dùng	Đạt
Gửi yêu cầu mượn hợp lệ	1. Truy cập giao diện thông tin chi tiết một cuốn sách có người cho mượn 2. Ấn nút mượn sách 3. Ấn nút gửi yêu cầu mượn 4. Nhập nội dung tin nhắn hoặc giữ nội dung mặc định và ấn gửi 5. Kiểm tra lại mã xác nhận trong danh sách sách đang yêu cầu mượn	Thông báo thành công kèm theo mã xác nhận trùng với mã xác nhận khi kiểm tra lại trong danh sách sách đang yêu cầu mượn	Đạt
Gửi yêu cầu mượn mà không nhập lời nhắn	1. Truy cập giao diện thông tin chi tiết một cuốn sách có người cho mượn 2. Ấn nút mượn sách 3. Ấn nút gửi yêu cầu mượn 4. Xóa nội dung tin nhắn mặc định và ấn gửi	Thông báo “Vui lòng nhập lời nhắn”	Đạt
Gửi yêu cầu mượn một cuốn sách đang gửi yêu cầu rồi	1. Truy cập giao diện thông tin chi tiết một cuốn sách đang gửi yêu cầu mượn 2. Ấn nút mượn sách	Thông báo “Bạn đang gửi yêu mượn cuốn sách này rồi”	Đạt

### CHƯƠNG 3. THIẾT KẾ HỆ THỐNG VÀ KẾT QUẢ THỰC NGHIỆM

	3. Ân nút gửi yêu cầu mượn 4. Nhập nội dung tin nhắn hoặc giữ nội dung mặc định và ân gửi	
--	--	--

**Bảng 3.13:** Kịch bản kiểm thử chức năng “Mượn sách”

#### 3.4.3 Kịch bản kiểm thử chức năng “Xác nhận cho mượn”

Mô tả	Kịch bản	Kết quả mong muốn	Kết quả
Xác nhận cho mượn sách hợp lệ	1. Truy cập giao diện sách đang được yêu cầu mượn trong mục “Sách cho mượn” 2. Ân nút Xác nhận cho mượn với một yêu cầu của một người dùng chưa mượn tới giới hạn số lượng sách được mượn đồng thời theo hạng 1/2/3 3. Nhập đúng mã xác nhận tương ứng với yêu cầu đã chọn	Thông báo xác nhận cho mượn sách thành công và chuyển trạng thái yêu cầu thành “Đang cho mượn”	Đạt
Nhập sai mã xác nhận của yêu cầu mượn sách	1. Truy cập giao diện sách đang được yêu cầu mượn trong mục “Sách cho mượn” 2. Ân nút Xác nhận cho mượn với một yêu cầu 3. Nhập sai mã xác nhận của yêu cầu đã chọn	Thông báo “Mã xác nhận không chính xác”	Đạt
Xác nhận yêu cầu của người dùng đã mượn tới giới hạn số lượng sách được mượn đồng thời	1. Truy cập giao diện sách đang được yêu cầu mượn trong mục “Sách cho mượn”	Thông báo “Hiện tại người mượn chỉ được mượn đồng thời 1/2/3 cuốn sách”	Đạt

### CHƯƠNG 3. THIẾT KẾ HỆ THỐNG VÀ KẾT QUẢ THỰC NGHIỆM

	<p>2. Ấn nút Xác nhận cho mượn với một yêu cầu của một người dùng đã mượn tối giới hạn số lượng sách được mượn đồng thời theo hạng 1/2/3</p> <p>3. Nhập đúng mã xác nhận tương ứng với yêu cầu đã chọn</p>	
--	--	--

**Bảng 3.14:** Kịch bản kiểm thử chức năng “Xác nhận cho mượn”

#### 3.4.4 Kịch bản kiểm thử chức năng “Tìm kiếm sách có ở gần”

Mô tả	Kịch bản	Kết quả mong muốn	Kết quả
Tìm kiếm theo địa chỉ và khoảng cách	<p>1. Truy cập giao diện “Sách gần bạn”</p> <p>2. Nhập địa chỉ (hoặc chọn vị trí trên bản đồ) và khoảng cách tìm kiếm</p> <p>3. Chọn sắp xếp theo khoảng cách</p>	Hiển thị các cuốn sách trong phạm vi tìm kiếm, sắp xếp theo khoảng cách	Đạt
Sắp xếp kết quả theo độ tương đồng với sách của người dùng	<p>1. Truy cập giao diện “Sách gần bạn”</p> <p>2. Nhập địa chỉ và khoảng cách tìm kiếm</p> <p>3. Chọn tiêu chí sắp xếp “Tương tự với sách của bạn”</p>	Hiển thị các cuốn sách trong phạm vi tìm kiếm, sắp xếp theo độ tương đồng với các cuốn sách mà người dùng đang có, yêu thích	Đạt

**Bảng 3.15:** Kịch bản kiểm thử chức năng “Tìm kiếm sách có ở gần”

## CHƯƠNG 4. CÔNG NGHỆ SỬ DỤNG

### 4.1 Công nghệ sử dụng trong xây dựng hệ thống web

#### 4.1.1 Framework Angular sử dụng trong xây dựng frontend

Angular là một framework do Google phát triển để xây dựng các Single Page Application (SPA). Trang web dạng SPA có thể cập nhật một phần dữ liệu hiển thị mà không cần tải lại toàn bộ trang. Phiên bản đầu tiên của Angular sử dụng JavaScript, các phiên bản sau đó sử dụng TypeScript - được coi là sự mở rộng của ngôn ngữ lập trình JavaScript, đặc biệt trong việc quản lý kiểu dữ liệu chặt chẽ hơn.

Khi lập trình frontend với Angular, các thành phần trong giao diện được chia thành các khôi (component). Một component bao gồm hai thành phần chính là phần template HTML và phần code TypeScript xử lý dữ liệu, nằm trong class của component đó. Các component có thể được chia nhỏ và tái sử dụng. Trong template của một component có thể gọi đến các component khác. Giữa class và template có thể cập nhật dữ liệu theo 2 chiều. Khi input trong giao diện được người dùng thay đổi, các biến tương ứng trong class được cập nhật, và ngược lại, khi các biến trong class thay đổi, dữ liệu hiển thị trên giao diện được cập nhật theo.

Một thành phần quan trọng khác của Angular là các service, hỗ trợ việc gọi đến các Web API tương ứng ở backend cũng như bên ngoài hệ thống. Các component sử dụng các service này để truyền đi các yêu cầu theo thao tác của người dùng và nhận lại các dữ liệu cần thiết để hiển thị lên giao diện. Các service có thể được gọi đồng thời và có thể được xử lý bất đồng bộ, không ảnh hưởng đến hoạt động của các service khác. Khi dữ liệu từ một service được trả về, một phần giao diện tương ứng được cập nhật mà không cần tải lại toàn bộ trang.

Trong đồ án này, em sử dụng phiên bản Angular 13 với các component cho các khôi giao diện cơ bản được xây dựng bởi Ant Design (<https://ng.ant.design>). Các component này hỗ trợ việc thống nhất giao diện trong hệ thống, xử lý một số sự kiện theo thao tác người dùng và hỗ trợ trao đổi dữ liệu giữa các template và class trong các component chính của hệ thống.

#### 4.1.2 Nền tảng ASP.NET sử dụng trong xây dựng backend

ASP.NET là tập hợp các công cụ mã nguồn mở được phát triển bởi Microsoft, hỗ trợ xây dựng các ứng dụng đa nền tảng từ các hệ thống Web đầy đủ, các hệ thống Web API đến các ứng dụng máy tính, ứng dụng di động,... Ứng dụng trên môi trường .Net có thể được viết bằng các ngôn ngữ lập trình C#, F#, hoặc Visual Basic.

Trong đồ án này, em sử dụng framework ASP.NET Core 5 để xây dựng backend cho hệ thống. Phần code backend được phân chia thành 2 gói chính là gói Controllers và gói Models. Trong đó, gói Controllers chứa các lớp với các phương thức tương ứng với các API cần sử dụng ở frontend. Các API này có thể linh hoạt kết hợp với các nền tảng khác nhau cho frontend của web hoặc ứng dụng di động, ứng dụng máy tính để trở thành hệ thống hoàn chỉnh. Gói Models chứa các lớp tương ứng với các thực thể trong cơ sở dữ liệu.

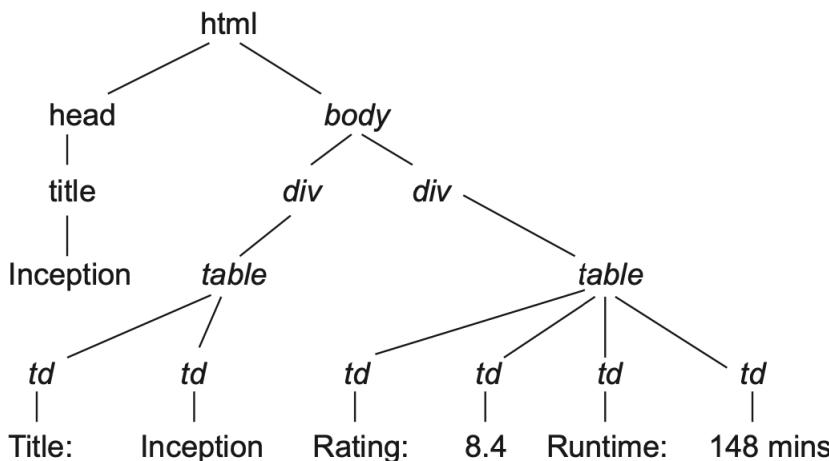
Việc xây dựng backend của hệ thống với ASP.NET Core còn được hỗ trợ thêm bởi hai công cụ là Entity Framework Core và SignalR. Trong đó, Entity Framework Core hỗ trợ việc kết nối đến cơ sở dữ liệu, chuyển đổi các truy vấn được viết bằng ngôn ngữ C# thành các truy vấn đến cơ sở dữ liệu tương ứng, dựa trên các lớp thuộc gói Models. Ngoài ra, framework này còn hỗ trợ việc ánh xạ giữa các lớp trong gói Models và các bảng trong cơ sở dữ liệu theo hai chiều, từ các lớp trong gói Models sinh ra các bảng trong cơ sở dữ liệu (code first), và ngược lại, từ các bảng trong cơ sở dữ liệu sinh ra các lớp trong gói Models (database first). Trong đồ án này, em xây dựng các lớp trong gói Models trước và thực hiện theo hướng code first để sinh ra các bảng trong cơ sở dữ liệu. Tiếp theo, thư viện SignalR là thư viện hỗ trợ xây dựng các chức năng thời gian thực như nhắn tin, gửi thông báo qua giao thức WebSocket. Khác với các chức năng thông thường khi server phản hồi lại các yêu cầu, các chức năng này cho phép server chủ động gửi dữ liệu đến các client đã kết nối khi cần thiết.

### 4.2 Ngôn ngữ Xpath và bộ chọn CSS

Ngôn ngữ Xpath và bộ chọn CSS (CSS Selector) là hai công cụ thường được sử dụng trong quá trình trích xuất dữ liệu từ các tài liệu HTML. XPath là viết tắt của XML Path Language, được sử dụng để truy vấn đến các phần tử trong các tài liệu XML và HTML với cú pháp tương tự như các đường dẫn thư mục trong máy tính, sử dụng dấu gạch chéo “/” để phân tách giữa các phần tử. Bộ chọn CSS cũng là công cụ được sử dụng để truy vấn đến các phần tử trong tài liệu HTML, có cú pháp đơn giản hơn XPath và thường được biết đến với mục đích chỉnh sửa giao diện cho các trang web.

Ngôn ngữ XPath được chia thành hai loại cú pháp là cú pháp tuyệt đối và cú pháp tương đối. Cú pháp tuyệt đối sử dụng một dấu gạch chéo “/” với ý nghĩa là phần tử gốc hoặc phần tử con ngay sau phần tử hiện tại. Cú pháp tương đối sử dụng hai dấu gạch chéo “//” với ý nghĩa truy vấn đến các phần tử ở vị trí bất kỳ sau phần tử hiện tại. Ví dụ, để truy vấn đến nội dung của tiêu đề “Inception” thuộc phần tử `table` thứ nhất trong cấu trúc cây HTML như hình 4.1 [2], đường dẫn XPath với cú

pháp tuyệt đối là /html/body/div[1]/table/td[2]/text(), đường dẫn XPath với cú pháp tương đối là //table[1]/td[2]/text(). Chỉ số trong ngoặc vuông của XPath tương tự như chỉ số sử dụng trong mảng, với giá trị bắt đầu bằng 1. Hàm text() được sử dụng để lấy nội dung văn bản của phần tử. Khi sử dụng để trích xuất dữ liệu từ các trang web, cú pháp tương đối được ưu tiên sử dụng hơn để giảm sự phụ thuộc vào cấu trúc các phần tử trong trang web.



**Hình 4.1:** Cấu trúc cây HTML minh họa cho ngôn ngữ XPath [2]

Cú pháp truy vấn kết hợp với điều kiện trong XPath có dạng:

```
//tagname[@attribute="value"]
```

Trong đó tagname là tên phần tử, attribute là tên thuộc tính và value là giá trị của thuộc tính muốn truy vấn. Ví dụ, đường dẫn //label[@id="hust"] truy vấn tất cả các phần tử label có id bằng "hust". Điều kiện khi truy vấn có thể được kết hợp với hàm contains. Ví dụ, //a[contains(@href, "hust")] truy vấn tất cả các phần tử a có thuộc thuộc tính href chứa chuỗi ký tự "hust". Các chức năng chi tiết hơn của ngôn ngữ XPath được mô tả trong tài liệu của tổ chức W3C (World Wide Web Consortium) [3].

Bộ chọn CSS là một chuỗi mô tả để xác định một hoặc nhiều phần tử trong trang web, dựa trên tên phần tử, các thuộc tính của phần tử hoặc mối liên hệ giữa các phần tử trong trang web. Cú pháp của bộ chọn CSS sử dụng dấu thăng "#" cho id, dấu chấm "." cho class, khoảng trắng thể hiện các phần tử con bất kỳ của một phần tử, dấu lớn hơn ">" thể hiện các phần tử con trực tiếp của một phần tử. Bảng 4.1 trình bày một số ví dụ về các bộ chọn CSS cơ bản.

Bộ chọn CSS	Ý nghĩa
#hust	Phần tử có id="hust"
.soict	Phần tử có class="soict"

#hust .soict	Phần tử có id="hust" và class="soict"
#hust .soict	Phần tử con bất kỳ có class="soict" của phần tử có id="hust"
#hust > .soict	Phần tử con trực tiếp có class="soict" của phần tử có id="hust"
#hust + .soict	Phần tử có class="soict" đứng liền sau phần tử có id="hust"

**Bảng 4.1:** Một số ví dụ về các bộ chọn CSS cơ bản

Cú pháp truy vấn kết hợp với điều kiện trong bộ chọn CSS cũng sử dụng kí hiệu ngoặc vuông tương tự như ngôn ngữ XPath. Bảng 4.2 trình bày một số ví dụ về các điều kiện trong bộ chọn CSS như thuộc tính của phần tử có giá trị bằng, bắt đầu bằng, kết thúc bằng hoặc có chứa một chuỗi ký tự mong muốn. Các cú pháp khác của bộ chọn CSS có thể được tham khảo tại tài liệu của tổ chức W3C (World Wide Web Consortium) [4].

Bộ chọn CSS	Ý nghĩa
a [title="hust"]	Phần tử a có title="hust"
a [href^="https"]	Phần tử a có thuộc tính href bắt đầu bằng "https"
a [href\$="soict"]	Phần tử a có thuộc tính href kết thúc bằng "soict"
a [href*="soict"]	Phần tử a có thuộc tính href có chứa chuỗi "soict"

**Bảng 4.2:** Một số ví dụ sử dụng điều kiện trong bộ chọn CSS

### 4.3 Các phương pháp trong hệ gợi ý

Hệ thống gợi ý thường được sử dụng để gợi ý nội dung cho người dùng trên các diễn đàn, mạng xã hội, các trang thương mại điện tử,... dựa trên các tri thức về sản phẩm, các tương tác của người dùng và các tri thức bổ sung của chuyên gia. Các phương pháp thường được sử dụng trong hệ gợi ý là gợi ý dựa trên nội dung, gợi ý dựa trên lọc cộng tác, gợi ý dựa trên tri thức và kết hợp sử dụng các phương pháp trên [5].

Phương pháp gợi ý dựa trên nội dung gợi ý bằng cách tính độ tương đồng giữa các sản phẩm và đưa ra các sản phẩm có độ tương đồng lớn với các sản phẩm người dùng đang xem hoặc các sản phẩm người dùng đã có tương tác tích cực. Để thực

hiện tính độ tương đồng giữa các sản phẩm, hệ thống cần xây dựng các vector đại diện dựa trên thông tin của mỗi sản phẩm. Các vector này có thể được tính toán và lưu trữ trước để hệ thống thực hiện truy vấn và tính độ tương đồng khi cần thiết. Ngoài xây dựng vector đại diện cho mỗi sản phẩm, hệ thống cũng có thể xây dựng vector đại diện cho mỗi người dùng bằng cách đưa ra một số câu hỏi về sản phẩm hoặc dựa trên các thông tin như độ tuổi, địa chỉ của người dùng. Từ đó, hệ thống gợi ý sản phẩm dựa trên độ tương đồng giữa vector đại diện của các sản phẩm và vector đại diện của người dùng. Phương pháp gợi ý dựa trên nội dung có ưu điểm là sử dụng được khi hệ thống chưa có nhiều thông tin tương tác từ người dùng.

Phương pháp gợi ý dựa trên lọc cộng tác không chỉ dựa trên độ tương đồng giữa các sản phẩm hay giữa sản phẩm và người dùng mà còn dựa trên sự tương đồng giữa các người dùng với nhau. Dữ liệu cần thiết cho phương pháp lọc cộng tác là ma trận đánh giá của người dùng và sản phẩm. Đánh giá của người dùng có thể là đánh giá trực tiếp như số điểm, số sao dành cho sản phẩm, hoặc kết hợp từ các tương tác như bỏ qua sản phẩm, xem hoặc mua,... Phương pháp lọc cộng tác có thể được chia thành hai cách tiếp cận là lọc cộng tác dựa trên mô hình và lọc cộng tác dựa trên láng giềng. Với cách tiếp cận dựa trên mô hình, ma trận đánh giá được sử dụng làm đầu vào để huấn luyện mô hình, từ đó dự đoán mức độ quan tâm của người dùng với các sản phẩm để thực hiện gợi ý. Phương pháp lọc cộng tác dựa trên láng giềng tính toán độ tương đồng giữa các người dùng, từ đó chọn ra tập người dùng láng giềng có độ tương đồng lớn nhất và đưa ra dự đoán đánh giá của người dùng cho các sản phẩm. Công thức tính độ tương đồng giữa hai người dùng  $a$  và  $b$  thường được sử dụng là công thức Pearson (4.1) [5]. Dự đoán đánh giá của người dùng  $a$  cho sản phẩm  $p$  được tính theo công thức (4.2) [5].

$$sim(a, b) = \frac{\sum_{p \in P} (r_{a,p} - \bar{r}_a)(r_{b,p} - \bar{r}_b)}{\sqrt{\sum_{p \in P} (r_{a,p} - \bar{r}_a)^2} \sqrt{\sum_{p \in P} (r_{b,p} - \bar{r}_b)^2}} \quad (4.1)$$

$$pred(a, p) = \bar{r}_a + \frac{\sum_{b \in N} sim(a, b) * (r_{b,p} - \bar{r}_b)}{\sum_{b \in N} sim(a, b)} \quad (4.2)$$

Trong đó,  $P$  là tập sản phẩm cùng được đánh giá bởi cả hai người dùng,  $r_{a,p}$  và  $r_{b,p}$  là đánh giá của người dùng  $a$  và người dùng  $b$  cho sản phẩm  $p$ ,  $\bar{r}_a$  và  $\bar{r}_b$  là đánh giá trung bình của hai người dùng trên các sản phẩm,  $N$  là tập các người dùng láng giềng của người dùng  $a$ .

Phương pháp gợi ý dựa trên tri thức sử dụng các ràng buộc, các thông tin về người dùng và ngữ cảnh, giúp cải thiện kết quả gợi ý và xử lý trong trường hợp hệ thống không có tương tác thường xuyên từ mỗi người dùng, như các hệ thống

bán đồ điện tử, sản phẩm có giá trị cao, sử dụng trong thời gian dài. Ngữ cảnh được sử dụng có thể là các khoảng thời gian trong năm, các sự kiện đặc biệt,... Các ràng buộc được xây dựng dựa trên các phân tích, kiến thức của chuyên gia, giúp hệ thống đưa ra các gợi ý như vào khoảng thời gian nào, nếu người dùng xem sản phẩm thuộc nhóm nào thì nên gợi ý các sản phẩm gì. Hệ thống có thể kết hợp sử dụng một số yêu cầu tìm kiếm của người dùng để đưa ra các kết quả gợi ý phù hợp.

Các phương pháp kể trên đều có những ưu điểm và yếu điểm riêng. Các hệ thống thường kết hợp các phương pháp để cải thiện kết quả gợi ý, đáp ứng được cho nhiều đối tượng người dùng kể cả những người dùng mới và ít tương tác. Các phương pháp cũng có thể được kết hợp theo trọng số để tính toán đánh giá của người dùng cho các sản phẩm hoặc kết hợp trình tự với đầu ra của phương pháp này là đầu vào cho phương pháp tiếp theo [5].

#### 4.4 Cơ sở lý thuyết cho các mô hình mã hóa vector

Các mô hình mã hóa văn bản thành vector được sử dụng để mã hóa mô tả của các cuốn sách thành các vector đại diện, phục vụ cho việc tính độ tương đồng để gợi ý các cuốn sách tương tự khi người dùng sử dụng hệ thống. Trong phần 4.4 này, em trình bày về cơ sở lý thuyết cho hai phương pháp là mã hóa theo TF-IDF và theo mô hình Word2Vec. Với mô hình Word2Vec, khi sử dụng có thể dùng các mô hình đã được huấn luyện trước và được công bố trong cộng đồng, hoặc tiến hành huấn luyện lại mô hình trên dữ liệu trong bài toán của mình.

##### 4.4.1 Mã hóa theo TF-IDF

Phương pháp TF-IDF mã hóa mỗi văn bản thành một vector có kích thước bằng kích thước tập từ vựng của bộ dữ liệu. Tập từ vựng có thể bao gồm toàn bộ các từ trong bộ dữ liệu hoặc sử dụng  $k$  từ có tần suất xuất hiện nhiều nhất, ví dụ  $k = 5000, 10000, \dots$ .

Xét tập từ vựng của bộ dữ liệu  $D$  có kích thước  $k$ , bao gồm các từ  $t_1, t_2, \dots, t_k$ . Mỗi văn bản  $d$  thuộc bộ dữ liệu  $D$  được mã hóa thành vector  $w = (w_1, w_2, \dots, w_k)$ . Các giá trị  $w_i$  được tính theo công thức (4.3) [6].

$$w_i = TF(t_i, d) * IDF(t_i) \quad (4.3)$$

Trong đó,  $TF(t_i, d)$  là số lần xuất hiện của từ  $t_i$  trong văn bản  $d$  (TF - Term Frequency).  $IDF(\text{Inverse Document Frequency})$  được tính theo công thức (4.4).

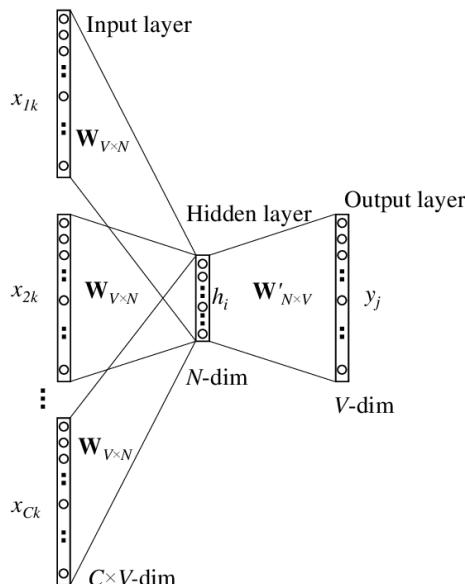
$$IDF(t_i) = \log\left(\frac{|D|}{DF(t_i)}\right) \quad (4.4)$$

Với  $|D|$  là tổng số văn bản trong bộ dữ liệu,  $DF(t_i)$  là số văn bản có xuất hiện từ  $t_i$ . Từ  $t_i$  càng xuất hiện trong nhiều văn bản thì giá trị  $IDF(t_i)$  càng nhỏ.

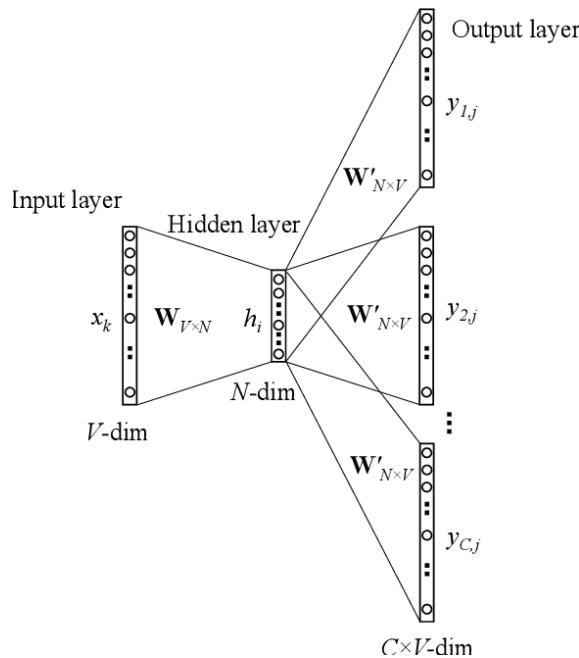
Theo cách tính toán trên, vector mã hóa cho ra bởi phương pháp TF-IDF thể hiện được mức độ quan trọng của các từ xuất hiện trong mỗi văn bản. Các từ xuất hiện nhiều trong một văn bản được xem là quan trọng hơn trong văn bản đó và có trọng số lớn hơn. Ngược lại, một từ xuất hiện trong nhiều văn bản thì được coi là ít quan trọng hơn. Phương pháp này có nhược điểm là cho ra vector mã hóa thưa, có kích thước lớn và không thể hiện được thứ tự xuất hiện của các từ trong văn bản do chỉ xét đến tần suất xuất hiện của các từ khi tính toán.

#### 4.4.2 Mô hình Word2Vec

Việc học vector mã hóa theo mô hình Word2Vec có thể khắc phục được các nhược điểm của phương pháp TF-IDF. Có 2 phương pháp để huấn luyện mô hình Word2Vec là CBOW (Continuous Bag-of-Word) và Skip-Gram. Phương pháp CBOW thực hiện dự đoán từ đích dựa trên các từ ngữ cảnh xung quanh. Ngược lại, phương pháp Skip-Gram thực hiện dự đoán các từ xung quanh dựa trên một từ cho trước. Cả 2 phương pháp đều được huấn luyện theo mạng nơ-ron gồm 1 tầng ẩn được minh họa trong hình 4.2 và hình 4.3 [7].

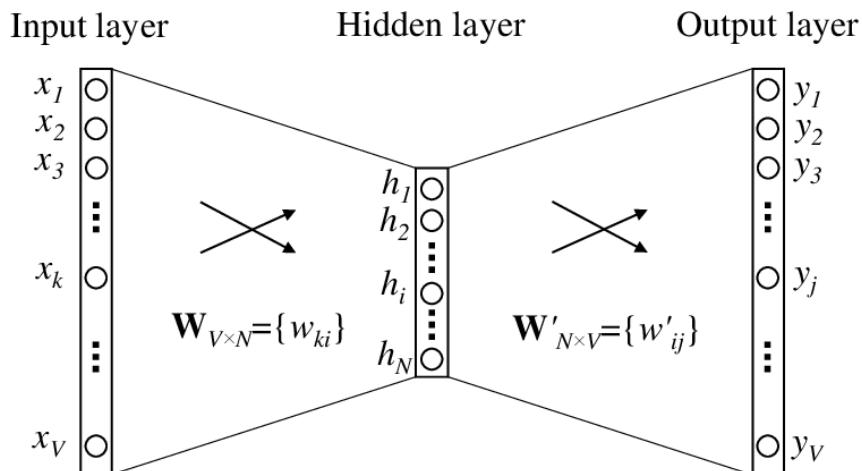


**Hình 4.2:** Mô hình Continuous Bag-of-Word [7]



**Hình 4.3:** Mô hình Skip-Gram [7]

Để thực hiện việc huấn luyện, đầu tiên, mỗi từ trong tập từ vựng được mã hóa thành một vector one-hot có kích thước bằng kích thước  $V$  của tập từ vựng. Vector này có một phần tử có giá trị bằng 1 là phần tử tại vị trí tương ứng với vị trí của từ được mã hóa trong tập từ vựng, các phần tử còn lại có giá trị bằng 0. Các vector one-hot này được sử dụng làm đầu vào và đầu ra mong muốn cho mạng nơ-ron.



**Hình 4.4:** Mô hình CBOW với một từ ngữ cảnh [7]

Hình 4.4 minh họa mô hình CBOW với chỉ một từ ngữ cảnh [7]. Tầng ẩn của mạng nơ-ron có kích thước  $N$ , chính là kích thước của vector mã hóa các từ thu được sau quá trình huấn luyện. Mạng nơ-ron này bao gồm 2 ma trận trọng số gồm ma trận  $W$  có kích thước  $V \times N$  giữa tầng đầu vào và tầng ẩn và ma trận  $W'$  có kích thước  $N \times V$  giữa tầng ẩn và tầng đầu ra. Cả tầng đầu vào và tầng đầu ra của mạng

đều có kích thước  $1 \times V$ . Tầng ẩn không sử dụng một hàm kích hoạt phi tuyến nào, đầu ra của tầng ẩn chính là kết quả của phép nhân ma trận giữa tầng đầu vào và ma trận  $W$ . Trong trường hợp chỉ có một từ ngữ cảnh, kết quả đó chính là một hàng của ma trận  $W$ , tương ứng với phần tử có giá trị 1 duy nhất trong vector đầu vào. Kết quả của tầng đầu ra được tính qua một hàm softmax theo công thức 4.5 [7] để thu được vector thể hiện từ được dự đoán, có tổng các phần tử bằng 1.

$$y_j = \frac{e^{u_j}}{\sum_{j'=1}^V e^{u_{j'}}} \quad (4.5)$$

Trong đó,  $u$  là kết quả của phép nhân giữa tầng ẩn và ma trận  $W'$ . Kết quả của tầng đầu ra và vector one-hot của từ đích được sử dụng để tính toán hàm mất mát, từ đó cập nhật 2 ma trận trọng số  $W$  và  $W'$ . Chi tiết về quá trình cập nhật tham số được trình bày trong nghiên cứu giải thích về quá trình học tham số mô hình Word2Vec của Xin Rong [7]. Sau quá trình huấn luyện, ma trận  $W$  được sử dụng làm kết quả mã hóa, với mỗi hàng tương ứng với một vector mã hóa cho một từ. Với mô hình CBOW tổng quát hơn, có đầu vào là nhiều từ ngữ cảnh, đầu ra của tầng ẩn là trung bình kết quả của các phép nhân giữa vector one-hot của mỗi từ đầu vào và ma trận  $W$ . Với mô hình Skip-Gram, hướng đến dự đoán nhiều từ ngữ cảnh, hàm mất mát được tính theo kết quả của tầng đầu ra và tất cả các vector one-hot của các từ ngữ cảnh.

## CHƯƠNG 5. TÍCH HỢP DỮ LIỆU VÀ GỢI Ý SÁCH

### 5.1 Tích hợp dữ liệu

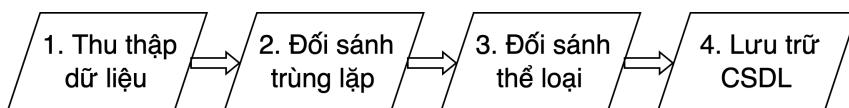
#### 5.1.1 Phương pháp thực hiện

Với mục đích cung cấp thông tin sách để người dùng tìm kiếm, lựa chọn sách mình đang có và gợi ý sách cho người dùng, hệ thống cần tích hợp dữ liệu về thông tin sách từ nhiều nguồn khác nhau. Các hệ thống tích hợp dữ liệu có thể áp dụng nhiều kỹ thuật khác nhau, nhưng một cách tổng quan, có hai kiến trúc tích hợp dữ liệu là Kho dữ liệu (Data warehousing) và Tích hợp dữ liệu ảo (Virtual data integration) [2].

Phương pháp Kho dữ liệu thực hiện thu thập dữ liệu từ các nguồn và lưu trữ vào một cơ sở dữ liệu chung. Sau đó, việc truy vấn dữ liệu được thực hiện trên cơ sở dữ liệu chung này. Nhờ đó, quá trình truy vấn có thể cho kết quả nhanh và ổn định. Cách thực hiện này có hai nhược điểm chính là yêu cầu bộ nhớ lưu trữ và dữ liệu được tích hợp không phải là dữ liệu mới nhất.

Với phương pháp Tích hợp dữ liệu ảo, thay vì lưu trữ các bản sao của dữ liệu vào cơ sở dữ liệu chung, hệ thống chỉ lưu thông tin về các nguồn dữ liệu. Khi nhận được truy vấn, hệ thống tiến hành phân tích truy vấn thành các truy vấn riêng cho từng nguồn, sau đó trả lại kết quả truy vấn được tổng hợp. Do không phải thu thập và lưu trữ tất cả dữ liệu từ các nguồn, phương pháp này không yêu cầu về bộ nhớ lưu trữ và dữ liệu được tích hợp cũng là dữ liệu mới nhất. Tuy nhiên, quá trình truy vấn cần nhiều thời gian và kém ổn định hơn so với phương pháp Kho dữ liệu, do phải tiến hành phân tích, xử lý, tổng hợp truy vấn từ các nguồn.

Với các ưu, nhược điểm như trên, trong đồ án này, em lựa chọn tích hợp theo phương pháp Kho dữ liệu để có được ưu điểm về thời gian và tính ổn định khi truy vấn. Các bước thực hiện được minh họa trong hình 5.1.



Hình 5.1: Các bước thực hiện tích hợp dữ liệu

#### 5.1.2 Các nguồn dữ liệu và các giá trị thu thập

Em đã tiến hành thu thập dữ liệu cho hệ thống từ 7 trang web bán sách phổ biến tại Việt Nam. Tổng số lượng sách thu thập được là 194681 cuốn, với số lượng tại mỗi nguồn cụ thể trong bảng 5.1. Các trường dữ liệu được trích xuất từ các trang

web bằng đường dẫn Xpath hoặc truy vấn CSS Selector phù hợp. Riêng với nguồn dữ liệu từ tiki.vn, các trường dữ liệu được trích xuất từ dữ liệu dạng JSON trả về qua các API truy vấn sản phẩm.

Nguồn dữ liệu	Số lượng thu thập
vinabook.com	60945
fahasa.com	44576
cachep.vn	27862
tiki.vn	19439
nhasachphuongnam.com	17944
khaitam.com	14654
bookbuy.vn	9261

**Bảng 5.1:** Số lượng sách thu thập tại mỗi nguồn

Bảng 5.2 trình bày các trường dữ liệu thu thập và ý nghĩa của chúng. Các trường dữ liệu về tên, tác giả, mô tả, thể loại được sử dụng trong việc đối sánh trùng lặp và gợi ý sách.

Trường dữ liệu	Ý nghĩa
name	tên sách
author	tác giả
description	nội dung mô tả sách
category	thể loại
image_url	đường dẫn đến hình ảnh của cuốn sách
publisher	tên nhà xuất bản sách
publish_time	thời gian xuất bản sách
pages	số trang
website	nguồn dữ liệu
url	url thu thập

**Bảng 5.2:** Các trường dữ liệu thu thập và ý nghĩa của chúng

### 5.1.3 Đối sánh trùng lặp và đối sánh thể loại

Do dữ liệu được thu thập từ nhiều nguồn, có những cuốn sách được thu thập nhiều lần từ các nguồn khác nhau. Để loại bỏ các bản ghi trùng lặp cho mỗi sách, em thực hiện so sánh tên sách theo cách tính độ tương đồng dựa trên khoảng cách chỉnh sửa Levenshtein.

Khoảng cách chỉnh sửa Levenshtein được định nghĩa bằng số chỉnh sửa tối thiểu cần thực hiện để chuyển đổi hai xâu thành giống nhau. Có 3 loại chỉnh sửa có thể thực hiện bao gồm: thêm 1 ký tự, xóa 1 ký tự, chuyển 1 ký tự thành ký tự khác. Khoảng cách chỉnh sửa giữa hai chuỗi có thể được tính theo phương pháp quy

hoạch động. Với 2 chuỗi  $x = x_1x_2\dots x_n$  và  $y = y_1y_2\dots y_m$  có độ dài lần lượt là  $n$  và  $m$ , khoảng cách chỉnh sửa được tính theo phương trình (5.1) [2].

$$d(i, j) = \min \begin{cases} d(i-1, j-1), \text{ trường hợp } x_i = y_j \\ d(i-1, j-1) + 1, \text{ trường hợp thay } x_i \text{ thành } y_j \\ d(i-1, j) + 1, \text{ trường hợp thêm } x_i \text{ vào chuỗi } y \\ d(i, j-1) + 1, \text{ trường hợp thêm } y_j \text{ vào chuỗi } x \end{cases} \quad (5.1)$$

Trong đó,  $d(i, j)$  là khoảng cách chỉnh sửa giữa 2 chuỗi con  $x = x_1x_2\dots x_i$  và  $y = y_1y_2\dots y_j$ .

Độ tương đồng dựa trên khoảng cách chỉnh sửa được tính theo phương trình (5.2).

$$1 - \frac{d(n, m)}{\max(n, m)}. \quad (5.2)$$

Độ tương đồng có khoảng giá trị trong khoảng từ 0 đến 1, hai xâu càng giống nhau thì độ tương đồng càng lớn, hai xâu giống nhau có độ tương đồng bằng 1.

Trong đồ án này, em thực hiện so sánh độ tương đồng dựa trên tên sách, sau khi loại bỏ các ký tự đặc biệt và đưa tên sách về chữ thường, với ngưỡng so sánh là 0.95, được lựa chọn qua quan sát thực nghiệm. Để tránh loại bỏ nhầm những cuốn sách có cùng tên nhưng số tập khác nhau, em thực hiện trích xuất thêm trường dữ liệu về số tập trong tên sách. Các cuốn sách có độ tương đồng tên sách lớn hơn ngưỡng trên và có số tập giống nhau được coi là trùng lặp. Trường số tập được trích xuất và sắp xếp theo thứ tự các từ khóa: “phần”, “tập”, “quyển”, “số”, “vol”. Việc sắp xếp giúp xử lý trường hợp các cuốn sách có nhiều hơn một từ khóa được trích xuất. Ví dụ, cuốn sách “phần 4 - tập 3” được viết ở các nguồn dữ liệu khác nhau thành “phần 4 - tập 3”, “tập 3 - phần 4”.

Sau khi loại bỏ trùng lặp, số lượng sách còn lại là 127021 cuốn.

Để thực hiện bước đối sánh thể loại, em tổng hợp thể loại của các nguồn dữ liệu và lựa chọn ra 18 thể loại chung, sau đó, thực hiện ánh xạ từng thể loại của mỗi nguồn về các thể loại chung. Các cuốn sách không có thể loại thuộc ánh xạ được xếp vào thể loại “Khác”. Bảng 5.3 và bảng 5.4 trình bày ánh xạ thể loại cho hai nguồn dữ liệu “bookbuy.vn” và “cachep.vn”. Ánh xạ thể loại cho các nguồn dữ liệu còn lại được trình bày trong phần phụ lục B.

Thể loại nguồn	Thể loại chung
Văn học	Văn học
Kỹ năng sống	Kỹ năng sống
Tuổi teen	Tâm lý - Giới tính

## CHƯƠNG 5. TÍCH HỢP DỮ LIỆU VÀ GỢI Ý SÁCH

Trẻ em	Thiếu nhi
Kinh tế	Kinh tế - Chính Trị
Sức khỏe	Sức khỏe

**Bảng 5.3:** Ánh xạ thể loại cho nguồn dữ liệu “bookbuy.vn”

Thể loại nguồn	Thể loại chung
Thiếu Nhi	Thiếu nhi
Văn Học	Văn học
Tâm Lý - Kỹ Năng Sống	Kỹ năng sống
Sách Học Ngoại Ngữ	Tin Học - Ngoại Ngữ
Kinh Tế	Kinh tế - Chính Trị
Giáo Khoa - Tham Khảo	Sách giáo Khoa - Giáo Trình
Lịch Sử - Địa Lý - Tôn Giáo	Lịch sử - Địa lý - Tôn giáo
Nuôi Dạy Con	Nuôi dạy con
Khoa Học Kỹ Thuật	Khoa Học - Kỹ Thuật
Tiểu Sử - Hồi Ký	Tiểu sử - Hồi ký
Chính Trị - Pháp Lý - Triết Học	Kinh tế - Chính Trị
Từ Điển	Từ Điển
Văn Hóa - Nghệ Thuật - Du Lịch	Văn Hóa - Nghệ thuật
Báo - Tạp Chí	Báo - Tạp Chí
Thể Dục Thể Thao - Giải Trí	Sức khỏe
Nữ Công Gia Chánh	Gia Đình
Âm Nhạc - Mỹ Thuật - Thời Trang	Văn Hóa - Nghệ thuật

**Bảng 5.4:** Ánh xạ thể loại cho nguồn dữ liệu “cachep.vn”

Các thể loại chung và số lượng cụ thể của mỗi thể loại sau khi đối sánh được tổng hợp trong bảng. 5.5.

Thể loại	Số lượng
Thiếu nhi	31944
Văn học	22659
Sách giáo Khoa - Giáo Trình	17454
Kỹ năng sống	11098
Kinh tế - Chính Trị	9087
Tin Học - Ngoại Ngữ	7678
Khác	5013
Lịch sử - Địa lý - Tôn giáo	3870

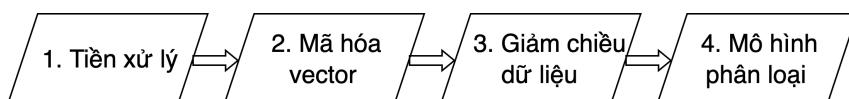
Truyện Tranh	3777
Gia Đình	2808
Khoa Học - Kỹ Thuật	2587
Văn Hóa - Nghệ thuật	2322
Nuôi dạy con	1814
Báo - Tạp Chí	1700
Sức khỏe	1341
Tâm lý - Giới tính	679
Từ Điển	605
Tiểu sử - Hồi ký	585

**Bảng 5.5:** Các thể loại sách và số lượng mỗi loại

## 5.2 Gợi ý sách

### 5.2.1 Thủ nghiệm các phương pháp mã hóa văn bản thành vector

Để thực hiện chức năng gợi ý sách, em cần tiến hành mã hóa các văn bản mô tả thành các vector đại diện cho mỗi cuốn sách. Dựa trên dữ liệu thu thập được, với các thể loại sách được thống nhất theo bảng 5.5, em tiến hành thử nghiệm một số phương pháp mã hóa văn bản thành vector bằng cách áp dụng các phương pháp này vào bài toán phân loại thể loại sách. Các bước thực hiện bài toán phân loại thể loại sách được mô tả trong hình 5.2.

**Hình 5.2:** Các bước thực hiện bài toán phân loại thể loại sách

Bước tiền xử lý thực hiện loại bỏ các đường dẫn url, các kí tự đặc biệt, chữ số ra khỏi văn bản mô tả và chuyển văn bản mô tả về chữ thường, sau đó thực hiện tách từ trong văn bản. Đối với tiếng Việt, các từ có thể bao gồm nhiều tiếng, bước tách từ nối các tiếng trong một từ bằng dấu gạch dưới, ví dụ "xin chào" được nối thành "xin\_chào". Từ đó, các văn bản mô tả sách có thể được chia thành các từ theo khoảng trắng, để sử dụng cho bước tiếp theo, mã hóa văn bản thành vector. Trong đồ án này, em thực hiện bước tách từ bằng thư viện pyvi (Python Vietnamese Toolkit) được xây dựng bởi tác giả Trần Việt Trung [8].

Các phương pháp mã hóa em tiến hành thử nghiệm bao gồm: mã hóa theo TF-IDF, mã hóa theo mô hình Word2Vec đã được huấn luyện trước trên tập dữ liệu các bài báo tiếng Việt [1] và mã hóa theo mô hình Word2Vec được huấn luyện trên dữ

liệu là mô tả của các cuốn sách thu thập được. Cơ sở lý thuyết cho các phương pháp này được trình bày trong phần 4.4.

Để có sự tương tự giữa bài toán được thử nghiệm và việc áp dụng các phương pháp mã hóa vector để tính độ tương đồng khi gợi ý, em thực hiện thử nghiệm theo mô hình phân loại KNN (K-Nearest Neighbors) sử dụng khoảng cách giữa các vector là khoảng cách cosine. Với 2 vector  $x, y$  có cùng kích thước  $n$ , khoảng cách cosine được tính theo công thức 5.3.

$$d(x, y) = 1 - \frac{x \cdot y}{\|x\| \|y\|} = 1 - \frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n x_i^2} \sqrt{\sum_{i=1}^n y_i^2}} \quad (5.3)$$

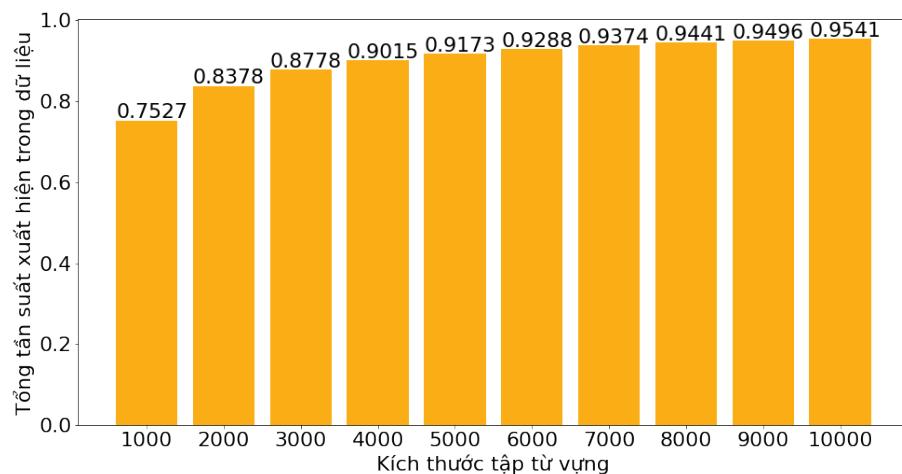
Mô hình KNN thực hiện phân loại bằng cách tìm  $k$  bản ghi trong tập huấn luyện có khoảng cách gần nhất với bản ghi cần phân loại, sau đó, đưa ra kết quả phân loại là nhãn lớp xuất hiện nhiều nhất trong  $k$  bản ghi.

Để giảm bớt khối lượng tính toán cần thực hiện, em áp dụng các phương pháp giảm chiều PCA và Truncated SVD để giảm kích thước các vector mã hóa. Hai phương pháp này có sự tương đồng trong cách tính toán, trong đó, phương pháp Truncated SVD có thể áp dụng được cho các dữ liệu thừa nên được sử dụng để giảm chiều cho các vector mã hóa theo phương pháp TF-IDF. Phương pháp PCA được sử dụng để giảm chiều cho các vector mã hóa theo mô hình Word2Vec. Các bước thực hiện chi tiết của hai phương pháp được phân tích trong nghiên cứu "A Tutorial on Principal Component Analysis" của tác giả Jonathon Shlens [9].

Tổng số dữ liệu thử nghiệm là 127021 sách, được chia thành các tập train, valid, test theo tỉ lệ lần lượt là 60%, 20%, 20% và được giữ theo tỉ lệ sách thuộc mỗi thể loại. Cụ thể tập train gồm 76212 sách, tập valid gồm 25404 sách và tập test gồm 25405 sách.

Cả 3 phương pháp mã hóa được thử nghiệm với mô hình KNN có số hàng xóm là 5, 10, 15, 20 và dữ liệu được giảm chiều về 50, 100, 200, 300.

Hình 5.3 thể hiện tổng số tần suất xuất hiện trong toàn bộ tập dữ liệu với các kích thước tập từ vựng khác nhau. Tổng số từ trong toàn bộ tập dữ liệu là 204174. Tuy nhiên, sự xuất hiện của các từ chủ yếu trong khoảng 10000 từ có tần suất xuất hiện lớn nhất với tỉ lệ 95.41%. Với 4000 từ có tần suất xuất hiện lớn nhất, tỉ lệ này là 90.15%.



**Hình 5.3:** Biểu đồ kích thước tập từ vựng  
và tổng số tần suất xuất hiện trong toàn bộ tập dữ liệu

Em tiến hành thử nghiệm mô hình TF-IDF với các kích thước tập từ vựng là 4000, 6000, 8000 và 10000. Các kết quả thử nghiệm được trình bày trong bảng 5.6 và 5.7. Kết quả tốt nhất trên tập dữ liệu valid khi mã hóa vector theo mô hình TF-IDF là 71.26%, với các tham số kích thước tập từ vựng 8000, kích thước sau khi giảm chiều 200 và số hàng xóm trong mô hình phân loại KNN là 5.

Kích thước tập từ vựng	Kích thước sau khi giảm chiều	Số hàng xóm trong KNN	Độ chính xác trên tập valid
4000	50	5	69.76%
4000	50	10	69.97%
4000	50	15	69.48%
4000	50	20	68.74%
4000	100	5	70.99%
4000	100	10	70.47%
4000	100	15	70.01%
4000	100	20	69.29%
4000	200	5	70.73%
4000	200	10	70.23%
4000	200	15	69.45%
4000	200	20	69%
4000	300	5	70.26%
4000	300	10	70.01%
4000	300	15	69.19%
4000	300	20	68.57%
6000	50	5	70.05%

## CHƯƠNG 5. TÍCH HỢP DỮ LIỆU VÀ GỢI Ý SÁCH

6000	50	10	70.18%	
6000	50	15	69.5%	
6000	50	20	68.8%	
6000	100	5	71.05%	
6000	100	10	70.45%	
6000	100	15	70.02%	
6000	100	20	69.46%	
6000	200	5	70.9%	
6000	200	10	70.58%	
6000	200	15	69.7%	
6000	200	20	69.06%	
6000	300	5	70.47%	
6000	300	10	70.03%	
6000	300	15	69.21%	
6000	300	20	68.66%	

**Bảng 5.6:** Kết quả thử nghiệm mã hóa vector theo TF-IDF  
với kích thước tập từ vựng 4000 và 6000

Kích thước tập từ vựng	Kích thước sau khi giảm chiều	Số hàng xóm trong KNN	Độ chính xác trên tập valid
8000	50	5	70.02%
8000	50	10	70.08%
8000	50	15	69.51%
8000	50	20	69.12%
8000	100	5	71.11%
8000	100	10	70.48%
8000	100	15	70.17%
8000	100	20	69.34%
<b>8000</b>	<b>200</b>	<b>5</b>	<b>71.26%</b>
8000	200	10	70.58%
8000	200	15	69.76%
8000	200	20	69.16%
8000	300	5	70.66%
8000	300	10	70.17%
8000	300	15	69.39%
8000	300	20	68.78%
10000	50	5	70.24%

10000	50	10	69.99%
10000	50	15	69.51%
10000	50	20	69%
10000	100	5	70.79%
10000	100	10	70.5%
10000	100	15	70.24%
10000	100	20	69.53%
10000	200	5	71.11%
10000	200	10	70.56%
10000	200	15	70.07%
10000	200	20	69.44%
10000	300	5	70.73%
10000	300	10	70.2%
10000	300	15	69.39%
10000	300	20	68.8%

**Bảng 5.7:** Kết quả thử nghiệm mã hóa vector theo TF-IDF với kích thước tập từ vựng 8000 và 10000

Bảng 5.8 trình bày kết quả thử nghiệm phương pháp mã hóa vector theo mô hình Word2Vec được huấn luyện trước (pretrained Word2Vec) trên dữ liệu các bài báo tiếng Việt theo nghiên cứu của tác giả Vũ Xuân Sơn [1]. Tác giả đã công bố hai mô hình bao gồm mô hình được huấn luyện với kích thước cửa sổ 2, kích thước vector mã hóa 300 và mô hình được huấn luyện với kích thước cửa sổ 5, kích thước vector mã hóa 400 [10]. Với mô hình đã được huấn luyện, mỗi từ được mã hóa thành một vector, các văn bản được mã hóa thành vector là trung bình của các vector mã hóa các từ xuất hiện trong văn bản, các từ không xuất hiện trong tập từ vựng của mô hình được bỏ qua. Kết quả tốt nhất trên tập dữ liệu valid khi thực hiện mã hóa theo mô hình Word2Vec được huấn luyện trước là 69.57% với mô hình có kích thước cửa sổ 5, kích thước vector mã hóa 400, kích thước sau khi giảm chiều là 300 và số hàng xóm trong mô hình phân loại KNN là 5.

Kích thước vector	Kích thước cửa sổ	Kích thước sau khi giảm chiều	Số hàng xóm trong KNN	Độ chính xác trên tập valid
300	2	50	5	67.65%
300	2	50	10	66.89%
300	2	50	15	66.1%
300	2	50	20	65.6%
300	2	100	5	68.5%

## CHƯƠNG 5. TÍCH HỢP DỮ LIỆU VÀ GỢI Ý SÁCH

300	2	100	10	67.6%
300	2	100	15	66.68%
300	2	100	20	66.12%
300	2	200	5	68.9%
300	2	200	10	67.7%
300	2	200	15	66.66%
300	2	200	20	66.02%
300	2	300	5	69.08%
300	2	300	10	67.57%
300	2	300	15	66.61%
300	2	300	20	66.09%
400	5	50	5	68.42%
400	5	50	10	67.71%
400	5	50	15	67.08%
400	5	100	5	69.08%
400	5	100	10	68.08%
400	5	100	15	67.54%
400	5	100	20	66.79%
400	5	200	5	69.44%
400	5	200	10	68.3%
400	5	200	15	67.64%
400	5	200	20	66.89%
<b>400</b>	<b>5</b>	<b>300</b>	<b>5</b>	<b>69.57%</b>
400	5	300	10	68.28%
400	5	300	15	67.46%
400	5	300	20	66.69%

**Bảng 5.8:** Kết quả thử nghiệm mã hóa vector theo mô hình Word2Vec  
được huấn luyện trước trên dữ liệu các bài báo tiếng Việt

Tiếp theo là kết quả thử nghiệm phương pháp mã hóa vector theo mô hình Word2Vec huấn luyện trên dữ liệu mô tả của các cuốn sách thu thập được. Mô hình được huấn luyện thử nghiệm theo cả hai phương pháp CBOW và Skip-Gram với các kích thước cửa sổ và kích thước vector được thử nghiệm lần lượt là 2 và 300, 5 và 400. 5 và 500. Bảng 5.9 trình bày kết quả khi huấn luyện mô hình theo phương pháp CBOW. Bảng 5.10 trình bày kết quả khi huấn luyện mô hình theo phương pháp Skip-Gram. Kết quả tốt nhất khi thử nghiệm trên tập valid theo mô

hình Word2Vec huấn luyện trên dữ liệu mô tả của các cuốn sách thu thập được là 73.21% với phương pháp Skip-Gram, kích thước vector 500, kích thước cửa sổ 5, kích thước sau khi giảm chiều 300 và số hàng xóm trong mô hình phân loại KNN là 5.

Kích thước vector	Kích thước cửa sổ	Mô hình huấn luyện	Kích thước sau khi giảm chiều	Số hàng xóm trong KNN	Độ chính xác trên tập valid
300	2	CBOW	50	5	69.9%
300	2	CBOW	50	10	69.6%
300	2	CBOW	50	15	68.92%
300	2	CBOW	50	20	68.44%
300	2	CBOW	100	5	70.69%
300	2	CBOW	100	10	70.19%
300	2	CBOW	100	15	69.54%
300	2	CBOW	100	20	68.95%
300	2	CBOW	200	5	70.88%
300	2	CBOW	200	10	70.16%
300	2	CBOW	200	15	69.28%
300	2	CBOW	200	20	68.64%
300	2	CBOW	300	5	70.6%
300	2	CBOW	300	10	69.97%
300	2	CBOW	300	15	69.18%
300	2	CBOW	300	20	68.5%
400	5	CBOW	50	5	70.2%
400	5	CBOW	50	10	69.92%
400	5	CBOW	50	15	69.12%
400	5	CBOW	50	20	68.78%
400	5	CBOW	100	5	70.9%
400	5	CBOW	100	10	70.27%
400	5	CBOW	100	15	69.39%
400	5	CBOW	100	20	68.7%
400	5	CBOW	200	5	71.02%
400	5	CBOW	200	10	70.32%
400	5	CBOW	200	15	69.41%
400	5	CBOW	200	20	68.87%
400	5	CBOW	300	5	71.08%
400	5	CBOW	300	10	70.07%
400	5	CBOW	300	15	69.31%
400	5	CBOW	300	20	68.59%
500	5	CBOW	50	5	70.5%
500	5	CBOW	50	10	69.7%

## CHƯƠNG 5. TÍCH HỢP DỮ LIỆU VÀ GỢI Ý SÁCH

500	5	CBOW	50	15	69.25%
500	5	CBOW	50	20	68.75%
500	5	CBOW	100	5	70.95%
500	5	CBOW	100	10	70.23%
500	5	CBOW	100	15	69.52%
500	5	CBOW	100	20	68.99%
500	5	CBOW	200	5	71.17%
500	5	CBOW	200	10	70.26%
500	5	CBOW	200	15	69.35%
500	5	CBOW	200	20	68.78%
500	5	CBOW	300	5	70.84%
500	5	CBOW	300	10	70.27%
500	5	CBOW	300	15	69.4%
500	5	CBOW	300	20	68.66%

**Bảng 5.9:** Kết quả thử nghiệm mã hóa vector theo mô hình Word2Vec huấn luyện trên dữ liệu sách thu thập được theo phương pháp CBOW

Kích thước vector	Kích thước cửa sổ	Mô hình huấn luyện	Kích thước sau khi giảm chiều	Số hàng xóm trong KNN	Độ chính xác trên tập valid
300	2	Skip-Gram	50	5	72.16%
300	2	Skip-Gram	50	10	72.05%
300	2	Skip-Gram	50	15	71.55%
300	2	Skip-Gram	50	20	70.99%
300	2	Skip-Gram	100	5	72.22%
300	2	Skip-Gram	100	10	72.03%
300	2	Skip-Gram	100	15	71.74%
300	2	Skip-Gram	100	20	71.19%
300	2	Skip-Gram	200	5	72.43%
300	2	Skip-Gram	200	10	71.76%
300	2	Skip-Gram	200	15	71.44%
300	2	Skip-Gram	200	20	70.85%
300	2	Skip-Gram	300	5	72.35%
300	2	Skip-Gram	300	10	71.67%
300	2	Skip-Gram	300	15	71.07%
300	2	Skip-Gram	300	20	70.37%
400	5	Skip-Gram	50	5	72.59%
400	5	Skip-Gram	50	10	72.47%
400	5	Skip-Gram	50	15	71.94%
400	5	Skip-Gram	50	20	71.47%
400	5	Skip-Gram	100	5	73.06%

400	5	Skip-Gram	100	10	72.94%
400	5	Skip-Gram	100	15	72.22%
400	5	Skip-Gram	100	20	71.66%
400	5	Skip-Gram	200	5	73.12%
400	5	Skip-Gram	200	10	72.59%
400	5	Skip-Gram	200	15	71.96%
400	5	Skip-Gram	200	20	71.66%
400	5	Skip-Gram	300	5	73.11%
400	5	Skip-Gram	300	10	72.65%
400	5	Skip-Gram	300	15	71.85%
400	5	Skip-Gram	300	20	71.23%
500	5	Skip-Gram	50	5	72.98%
500	5	Skip-Gram	50	10	72.42%
500	5	Skip-Gram	50	15	71.79%
500	5	Skip-Gram	50	20	71.41%
500	5	Skip-Gram	100	5	73.08%
500	5	Skip-Gram	100	10	72.76%
500	5	Skip-Gram	100	15	72.35%
500	5	Skip-Gram	100	20	71.75%
500	5	Skip-Gram	200	5	73.16%
500	5	Skip-Gram	200	10	72.62%
500	5	Skip-Gram	200	15	71.99%
500	5	Skip-Gram	200	20	71.53%
<b>500</b>	<b>5</b>	<b>Skip-Gram</b>	<b>300</b>	<b>5</b>	<b>73.21%</b>
500	5	Skip-Gram	300	10	72.61%
500	5	Skip-Gram	300	15	71.97%
500	5	Skip-Gram	300	20	71.45%

**Bảng 5.10:** Kết quả thử nghiệm mã hóa vector theo mô hình Word2Vec huấn luyện trên dữ liệu sách thu thập được theo phương pháp Skip-Gram

Bảng 5.11 tổng hợp kết quả tốt nhất của ba phương pháp mã hóa trên tập dữ liệu valid và kết quả thử nghiệm lại các tham số tương ứng với các kết quả này trên tập dữ liệu test. Kết quả thử nghiệm trên tập dữ liệu test tương tự kết quả thử nghiệm trên tập dữ liệu valid. Việc sử dụng mô hình Word2Vec được huấn luyện trước trên dữ liệu các bài báo cho kết quả thử nghiệm thấp nhất. Kết quả cao hơn khi sử dụng phương pháp TF-IDF và cao nhất khi huấn luyện lại mô hình Word2Vec trên dữ liệu thu thập được.

Phương pháp mã hóa	Độ chính xác trên tập valid	Độ chính xác trên tập test
-----------------------	--------------------------------	-------------------------------

TF-IDF	71.26%	71.24%
Pretrained Word2Vec	69.57%	69.70%
Word2Vec	<b>73.21%</b>	<b>73.45%</b>

**Bảng 5.11:** Kết quả thử nghiệm các phương pháp mã hóa trên tập dữ liệu valid và tập dữ liệu test

### 5.2.2 Thực hiện gợi ý sách

Sau khi thử nghiệm các phương pháp mã hóa vector trên, em lựa chọn thực hiện mã hóa vector theo mô hình Word2Vec được huấn luyện lại trên toàn bộ bộ dữ liệu thu thập được. Các vector sau khi huấn luyện được lưu trữ và sử dụng để tính độ tương đồng giữa các cuốn sách khi người dùng xem thông tin sách hoặc tìm kiếm sách có ở gần.

Khi người dùng xem thông tin sách tại trang chi tiết của một cuốn sách hoặc tại các trang quản lý sách cá nhân như sách đang có, sách yêu thích,... hệ thống thực hiện gợi ý các cuốn sách tương tự với các cuốn sách đang hiển thị bằng cách tính toán  $k_1$  cuốn sách có trung bình độ tương đồng cao nhất với các cuốn sách đang hiển thị, sau đó chọn ra  $k_2$  cuốn sách có thể mượn ở gần nhất theo vị trí của người dùng,  $k_2$  là số lượng sách gợi ý cho người dùng trên giao diện. Trong đồ án này, em triển khai hệ thống với  $k_1 = 100$  và  $k_2 = 10$ . Độ tương đồng giữa các cuốn sách được tính theo độ tương đồng cosine bằng công thức 5.4 với  $x$  và  $y$  là 2 vector có cùng kích thước  $n$ .

$$s(x, y) = \frac{x \cdot y}{\|x\| \|y\|} = \frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n x_i^2} \sqrt{\sum_{i=1}^n y_i^2}} \quad (5.4)$$

Với các cuốn sách do người dùng thêm vào hệ thống, chưa được tính toán vector đại diện, hệ thống thực hiện gợi ý theo các cuốn sách có cùng thể loại hoặc cùng tác giả.

Khi người dùng sử dụng chức năng tìm kiếm sách có ở gần, hệ thống tìm kiếm các cuốn sách có trong phạm vi bán kính mà người dùng tìm kiếm (ví dụ 3km, 5km). Người dùng có thể lựa chọn sắp xếp kết quả tìm kiếm theo khoảng cách, hoặc theo độ tương đồng với các cuốn sách cá nhân. Hệ thống tính toán khoảng cách theo đơn vị km dựa trên kinh độ và vĩ độ của 2 điểm bằng phương pháp Haversine với các bước thực hiện lần lượt theo các công thức 5.5, 5.6 và 5.7 [11].

$$a = \sin^2\left(\frac{\Delta lat}{2}\right) + \cos(lat1) * \cos(lat2) * \sin^2\left(\frac{\Delta long}{2}\right) \quad (5.5)$$

$$c = 2 * \text{atan2}(\sqrt{a} * \sqrt{1 - a}) \quad (5.6)$$

$$d = R * c \quad (5.7)$$

Trong đó,  $R \approx 6378.1370\ km$  là bán kính Trái Đất, các kinh độ và vĩ độ cần chuyển đổi từ độ về radian. Hàm  $\text{atan2}$  là mở rộng của hàm  $\arctan$  theo công thức 5.8.

$$\text{atan2}(y, x) = \begin{cases} \arctan\left(\frac{y}{x}\right) & \text{với } x > 0 \\ \arctan\left(\frac{y}{x}\right) + \pi & \text{với } x < 0, y \geq 0 \\ \arctan\left(\frac{y}{x}\right) - \pi & \text{với } x < 0, y < 0 \\ \frac{\pi}{2} & \text{với } x = 0, y > 0 \\ -\frac{\pi}{2} & \text{với } x = 0, y < 0 \\ \text{không xác định} & \text{với } x = 0, y = 0 \end{cases} \quad (5.8)$$

## CHƯƠNG 6. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

### 6.1 Kết luận

Qua các chương trước của báo cáo này, em đã giới thiệu và trình bày nội dung đồ án tốt nghiệp "Ứng dụng tích hợp dữ liệu và hệ gợi ý trong xây dựng hệ thống hỗ trợ mượn sách cộng đồng" của mình, được thực hiện dưới sự hướng dẫn của TS. Đỗ Bá Lâm. Đồ án đã đạt được các mục tiêu đề ra như đã thu thập dữ liệu sách từ các trang web bán sách trong nước, sử dụng làm cơ sở dữ liệu cho các chức năng về sách, gợi ý được các cuốn sách dựa trên độ tương đồng và khoảng cách đến vị trí của người dùng. Hệ thống xây dựng trên môi trường web với các chức năng có thể đáp ứng được một phần nhu cầu mượn và chia sẻ sách trong cộng đồng. Trong các phần tiếp theo của chương, em trình bày cụ thể hơn về các kết quả đạt được và chưa đạt được, từ đó đưa ra các hướng phát triển để hoàn thiện hơn các chức năng của hệ thống.

Về tích hợp dữ liệu, em đã thu thập được 194681 cuốn sách, từ 7 nguồn dữ liệu khác nhau và tích hợp theo phương pháp Kho dữ liệu. Dữ liệu được đối sánh trùng lặp theo độ tương đồng về khoảng cách chỉnh sửa dựa trên tên sách để loại bỏ các cuốn sách được thu thập lặp lại tại nhiều nguồn. Sau khi đối sánh trùng lặp, số lượng sách còn lại là 127021 cuốn. Thể loại sách được thống nhất về 18 thể loại chung để có thể áp dụng vào bài toán phân loại thể loại sách, nhằm thử nghiệm các phương pháp mã hóa văn bản thành vector.

Về hệ thống web, quá trình thực hiện đồ án cũng là lần đầu em tự mình xây dựng một hệ thống hoàn chỉnh từ bước lên ý tưởng đến khi hoàn thành sản phẩm. Em đã học hỏi thêm được các kỹ năng trong quá trình phân tích yêu cầu, cụ thể hóa các chức năng, thiết kế các thành phần chi tiết cho hệ thống như giao diện, các lớp và cơ sở dữ liệu. Quá trình lập trình triển khai hệ thống cũng đòi hỏi em phải hoàn thiện hơn các kiến thức về lập trình web với Angular và ASP.NET. Một số chức năng mới khi thực hiện em phải tìm hiểu nhiều hơn như việc tích hợp bản đồ để hiển thị và lựa chọn vị trí, chức năng nhắn tin và gửi thông báo.

Để thực hiện chức năng tìm kiếm sách ở gần, em đã thực hiện cách tính khoảng cách giữa hai điểm bằng kinh độ và vĩ độ theo phương pháp Haversine. Với chức năng gợi ý sách tương tự, em đã tìm hiểu và thử nghiệm một số phương pháp mã hóa văn bản thành vector như TF-IDF, sử dụng mô hình Word2Vec được huấn luyện trước [1] và huấn luyện lại mô hình Word2Vec trên dữ liệu các cuốn sách thu thập được. Các phương pháp mã hóa văn bản thành vector được so sánh thử nghiệm qua bài toán phân loại thể loại sách với mô hình phân loại KNN theo độ đo khoảng

cách cosine.

Sau khi thực hiện các công việc của đồ án, em nhận thấy hệ thống còn một số hạn chế như sau. Việc trích xuất các trường dữ liệu về sách khi thu thập từ các nguồn phụ thuộc vào các bộ chọn CSS Selector và các đường dẫn Xpath được xây dựng thủ công. Điều này gây khó khăn khi cấu trúc các trang web không thống nhất, thay đổi theo thời gian. Hệ thống web đã được triển khai các chức năng cần thiết nhưng chưa có giao diện tối ưu riêng dành cho điện thoại. Do chưa có dữ liệu thực tế về tương tác của người dùng, chức năng gợi ý sách của hệ thống mới chỉ được thực hiện theo phương pháp gợi ý dựa trên nội dung bằng cách tính độ tương đồng sử dụng các vector mã hóa mà chưa áp dụng được các kỹ thuật chuyên sâu trong hệ gợi ý.

### 6.2 Hướng phát triển

Cuối cùng, với các kết quả nêu trên, em xin đưa ra một số công việc cần thực hiện và định hướng phát triển để hoàn thiện hơn đồ án của mình.

Để thực hiện việc tích hợp dữ liệu tốt hơn, có khả năng tự động cập nhật cao hơn, hệ thống cần được triển khai các phương pháp trích xuất dữ liệu và đối sánh trùng lặp tự động dựa trên các mô hình học máy. Hệ thống web cần hoàn thiện các chức năng và có giao diện tối ưu dành cho điện thoại. Chức năng nhắn tin cũng cần được xây dựng phong phú hơn với các tiện ích như gửi kèm biểu tượng, hình ảnh khi trò chuyện. Chức năng tìm kiếm sách ở gần theo khoảng cách cần cải tiến để tính được khoảng cách theo quãng đường thay vì phạm vi bán kính xung quanh. Đặc biệt, hệ thống cần được thiết kế để lưu lại các tương tác của người dùng, từ đó tìm hiểu và áp dụng các phương pháp gợi ý chuyên sâu hơn như lọc cộng tác, phân tích ma trận, phân tích ma trận dựa trên học sâu.

## TÀI LIỆU THAM KHẢO

- [1] X.-S. Vu, T. Vu, S. N. Tran, and L. Jiang, “Etnlp: A visual-aided systematic approach to select pre-trained embeddings for a downstream task,” *In: Proceedings of the International Conference Recent Advances in Natural Language Processing (RANLP)*, 2019.
- [2] A. Doan, A. Halevy, and Z. Ives, *Principles of data integration*. Elsevier, 2012.
- [3] J. Clark, S. DeRose, *et al.*, *Xml path language (xpath)*, 1999.
- [4] W3C (World Wide Web Consortium), *Selectors*. [Online]. Available: <https://www.w3.org/TR/CSS2/selector.html> (visited on 08/02/2022).
- [5] D. Jannach, M. Zanker, A. Felfernig, and G. Friedrich, *Recommender systems: an introduction*. Cambridge University Press, 2010.
- [6] L.-P. Jing, H.-K. Huang, and H.-B. Shi, “Improved feature selection approach tfidf in text mining,” in *Proceedings. International Conference on Machine Learning and Cybernetics*, vol. 2, 2002, 944–946 vol.2. DOI: 10.1109/ICMLC.2002.1174522.
- [7] X. Rong, “Word2vec parameter learning explained,” *arXiv preprint arXiv: 1411.2738*, 2014.
- [8] T. V. Trung, *Python vietnamese toolkit*. [Online]. Available: <https://github.com/trungtv/pyvi> (visited on 08/02/2022).
- [9] J. Shlens, “A tutorial on principal component analysis,” *arXiv preprint arXiv: 1404.1100*, 2014.
- [10] X.-S. Vu, *Pre-trained word2vec models for vietnamese*, 2016. [Online]. Available: <https://github.com/sonvx/word2vecVN> (visited on 08/02/2022).
- [11] E. Winarno, W. Hadikurniawati, and R. N. Rosso, “Location based service for presence system using haversine method,” in *2017 international conference on innovative and creative information technology (ICITech)*, IEEE, 2017, pp. 1–4.

# **PHỤ LỤC**

## A. ĐẶC TẢ USE CASE

### A.1 Đặc tả use case “Đăng nhập”

Tên use case	Đăng nhập		
Tác nhân	Khách		
Tiền điều kiện	Không		
Luồng sự kiện chính	STT	Thực hiện bởi	Hành động
	1	Khách	Ấn chọn đăng nhập
	2	Hệ thống	Hiển thị giao diện đăng nhập
	3	Khách	Nhập tên đăng nhập và mật khẩu
	4	Khách	Ấn nút đăng nhập
	5	Hệ thống	Thông báo đăng nhập thành công
Luồng sự kiện phát sinh	STT	Thực hiện bởi	Hành động
	5a	Hệ thống	Thông báo tên đăng nhập hoặc mật khẩu không đúng
Hậu điều kiện	Hệ thống tạo một phiên hoạt động, cho phép người dùng sử dụng các chức năng sau khi đăng nhập		

Bảng A.1: Đặc tả use case Đăng nhập

### A.2 Đặc tả use case “Cập nhật thông tin tài khoản”

Tên use case	Cập nhật thông tin tài khoản		
Tác nhân	Người dùng		
Tiền điều kiện	Không		
Luồng sự kiện chính	STT	Thực hiện bởi	Hành động
	1	Người dùng	Ấn chọn Thông tin tài khoản
	2	Hệ thống	Hiển thị thông tin tài khoản
	3	Người dùng	Ấn nút Cập nhật thông tin
	4	Hệ thống	Hiển thị giao diện cập nhật thông tin
	5	Người dùng	Cập nhật thông tin
	6	Người dùng	Ấn nút Lưu
	7	Hệ thống	Thông báo cập nhật thành công

Luồng sự kiện phát sinh	STT	Thực hiện bởi	Hành động
	7a	Hệ thống	Hiển thị các trường dữ liệu không hợp lệ
Hậu điều kiện	Hệ thống lưu thông tin người dùng cập nhật, hiển thị thông tin mới		

Bảng A.2: Đặc tả use case Cập nhật thông tin tài khoản

### A.3 Đặc tả use case “Đổi mật khẩu”

Tên use case	Đổi mật khẩu		
Tác nhân	Người dùng		
Tiền điều kiện	Không		
Luồng sự kiện chính	STT	Thực hiện bởi	Hành động
	1	Người dùng	Ấn chọn Thông tin tài khoản
Luồng sự kiện phát sinh	2	Hệ thống	Hiển thị thông tin tài khoản
	3	Người dùng	Ấn nút Đổi mật khẩu
	4	Hệ thống	Hiển thị giao diện đổi mật khẩu
	5	Người dùng	Nhập thông tin*
	6	Người dùng	Ấn nút Lưu
	7	Hệ thống	Thông báo cập nhật thành công
	7a	Hệ thống	Hiển thị các trường dữ liệu không hợp lệ
Hậu điều kiện	Hệ thống lưu mật khẩu mới, hiển thị giao diện thông tin tài khoản		

Bảng A.3: Đặc tả use case Đổi mật khẩu

\* Mô tả dữ liệu đầu vào

Trường dữ liệu	Bắt buộc?	Điều kiện hợp lệ	Ví dụ
Mật khẩu cũ	Có		MatKhau123!@#

Mật khẩu mới	Có	Tối thiểu 6 ký tự	MatKhau123!@#
Nhập lại mật khẩu mới	Có	Trùng với mật khẩu mới	MatKhau123!@#

**Bảng A.4:** Dữ liệu đầu vào trong use case Đổi mật khẩu**A.4 Đặc tả use case “Cập nhật ảnh đại diện”**

Tên use case	Cập nhật ảnh đại diện		
Tác nhân	Người dùng		
Tiền điều kiện	Không		
Luồng sự kiện chính	STT	Thực hiện bởi	Hành động
	1	Người dùng	Ấn chọn Thông tin tài khoản
	2	Hệ thống	Hiển thị thông tin tài khoản
	3	Người dùng	Ấn chọn cập nhật ảnh đại diện
	4	Hệ thống	Hiển thị giao diện chọn ảnh
	5	Người dùng	Chọn ảnh đại diện
	6	Người dùng	Ấn nút Lưu
Luồng sự kiện phát sinh	STT	Thực hiện bởi	Hành động
	7a1	Hệ thống	Thông báo ảnh không đúng định dạng png, jpg, jpeg. Kích thước < 2 MB
Hậu điều kiện	Hệ thống lưu ảnh người dùng cập nhật, hiển thị ảnh đại diện mới		

**Bảng A.5:** Đặc tả use case Cập nhật ảnh đại diện**A.5 Đặc tả use case “Đánh dấu yêu thích sách”**

Tên use case	Đánh dấu yêu thích sách		
Tác nhân	Người dùng		
Tiền điều kiện	Không		
Luồng sự kiện chính	STT	Thực hiện bởi	Hành động
	1	Người dùng	Ấn chọn một cuốn sách
	2	Hệ thống	Hiển thị thông tin sách
	3	Người dùng	Ấn nút Yêu thích

	4	Hệ thống	Thay đổi trạng thái biểu tượng yêu thích
Luồng sự kiện phát sinh			
Hậu điều kiện	Hệ thống thêm cuốn sách vào danh sách yêu thích của người dùng		

**Bảng A.6:** Đặc tả use case Đánh dấu yêu thích sách

#### A.6 Đặc tả use case “Bỏ yêu thích sách”

Tên use case	Bỏ yêu thích sách		
Tác nhân	Người dùng		
Tiền điều kiện	Người dùng đã chọn yêu thích cuốn sách muốn bỏ yêu thích		
Luồng sự kiện chính	STT	Thực hiện bởi	Hành động
	1	Người dùng	Ấn chọn Bỏ yêu thích trong trang thông tin sách
	2	Hệ thống	Thay đổi trạng thái biểu tượng yêu thích
Luồng sự kiện phát sinh	STT	Thực hiện bởi	Hành động
	1a	Người dùng	Ấn chọn mục Sách của bạn
	1b	Người dùng	Ấn chọn mục Sách yêu thích
	1c	Hệ thống	Hiển thị danh sách yêu thích của người dùng
	1d	Người dùng	Ấn bỏ yêu thích một cuốn sách
	1e	Hệ thống	Hiển thị thông báo xác nhận
	1f	Người dùng	Ấn Xác nhận
	1g	Hệ thống	Thông báo Bỏ yêu thích thành công
Hậu điều kiện	Hệ thống xóa cuốn sách khỏi danh sách yêu thích của người dùng		

**Bảng A.7:** Đặc tả use case Bỏ yêu thích sách

#### A.7 Đặc tả use case “Đăng cảm nghĩ”

Tên use case	Đăng cảm nghĩ
--------------	---------------

Tác nhân	Người dùng		
Tiền điều kiện	Không		
Luồng sự kiện chính	STT	Thực hiện bởi	Hành động
	1	Người dùng	Ấn chọn một cuốn sách
	2	Hệ thống	Hiển thị thông tin sách
	3	Người dùng	Nhập cảm nghĩ
	4	Người dùng	Ấn nút Đăng
	5	Hệ thống	Thông báo Đăng cảm nghĩ thành công
Luồng sự kiện phát sinh			
Hậu điều kiện	Hệ thống lưu cảm nghĩ của người dùng và hiển thị ở trang thông tin sách		

**Bảng A.8:** Đặc tả use case Đăng cảm nghĩ**A.8 Đặc tả use case “Cập nhật cảm nghĩ”**

Tên use case	Cập nhật cảm nghĩ		
Tác nhân	Người dùng		
Tiền điều kiện	Người dùng đã đăng cảm nghĩ về cuốn sách		
Luồng sự kiện chính	STT	Thực hiện bởi	Hành động
	1	Người dùng	Ấn nút Chỉnh sửa bên cạnh cảm nghĩ của mình
	2	Hệ thống	Hiển thị ô cập nhật cảm nghĩ
	3	Người dùng	Nhập cảm nghĩ
	4	Người dùng	Ấn nút Lưu
	5	Hệ thống	Thông báo Cập nhật cảm nghĩ thành công
Luồng sự kiện phát sinh			
Hậu điều kiện	Hệ thống lưu cảm nghĩ của người dùng sau khi cập nhật và hiển thị ở trang thông tin sách		

**Bảng A.9:** Đặc tả use case Cập nhật cảm nghĩ**A.9 Đặc tả use case “Xóa cảm nghĩ”**

Tên use case	Xóa cảm nghĩ		
Tác nhân	Người dùng		
Tiền điều kiện	Người dùng đã đăng cảm nghĩ về cuốn sách		
Luồng sự kiện chính	STT	Thực hiện bởi	Hành động
	1	Người dùng	Ấn nút Xóa bên cạnh cảm nghĩ của mình
	2	Hệ thống	Hiển thị thông báo xác nhận xóa
	3	Người dùng	Ấn chọn xác nhận
	4	Hệ thống	Thông báo Xóa cảm nghĩ thành công
Luồng sự kiện phát sinh	STT	Thực hiện bởi	Hành động
	3a	Người dùng	Ấn chọn hủy
Hậu điều kiện	Hệ thống xóa cảm nghĩ của người dùng khỏi các cảm nghĩ của cuốn sách		

Bảng A.10: Đặc tả use case Xóa cảm nghĩ

#### A.10 Đặc tả use case “Xem sách đang cho mượn”

Tên use case	Xem sách đang cho mượn		
Tác nhân	Người dùng		
Tiền điều kiện	Không		
Luồng sự kiện chính	STT	Thực hiện bởi	Hành động
	1	Người dùng	Ấn chọn mục Sách của bạn
	2	Người dùng	Ấn chọn Sách cho mượn
	3	Người dùng	Ấn chọn Sách đang cho mượn
	4	Hệ thống	Hiển thị danh sách sách đang cho mượn của người dùng
Luồng sự kiện phát sinh			
Hậu điều kiện			

Bảng A.11: Đặc tả use case Xem sách đang cho mượn

#### A.11 Đặc tả use case “Đánh dấu sách đã được trả”

Tên use case	Đánh dấu sách đã được trả
--------------	---------------------------

Tác nhân	Người dùng		
Tiền điều kiện	Không		
Luồng sự kiện chính	STT	Thực hiện bởi	Hành động
	1	Người dùng	Ấn chọn mục Sách của bạn
	2	Người dùng	Ấn chọn Sách cho mượn
	3	Người dùng	Ấn chọn Sách đang cho mượn
	4	Hệ thống	Hiển thị danh sách sách đang cho mượn của người dùng
	5	Người dùng	Ấn nút đã được trả sách tại cuốn sách tương ứng
	6	Hệ thống	Thông báo xác nhận
	7	Người dùng	Ấn chọn xác nhận
	8	Hệ thống	Thông báo Xác nhận trả sách thành công
Luồng sự kiện phát sinh	STT	Thực hiện bởi	Hành động
	7a	Người dùng	Ấn chọn hủy
Hậu điều kiện	Hệ thống xóa cuốn sách khỏi danh sách đang cho mượn, thêm vào danh sách đã cho mượn của người dùng		

**Bảng A.12:** Đặc tả use case Đánh dấu sách đã được trả**A.12 Đặc tả use case “Xem sách đang yêu cầu mượn”**

Tên use case	Xem sách đang yêu cầu mượn		
Tác nhân	Người dùng		
Tiền điều kiện	Không		
Luồng sự kiện chính	STT	Thực hiện bởi	Hành động
	1	Người dùng	Ấn chọn mục Sách của bạn
	2	Người dùng	Ấn chọn Sách mượn
	3	Người dùng	Ấn chọn Sách đang yêu cầu mượn
	4	Hệ thống	Hiển thị danh sách sách đang yêu cầu mượn của người dùng
Luồng sự kiện phát sinh			

Hậu điều kiện	
---------------	--

**Bảng A.13:** Đặc tả use case Xem sách đang yêu cầu mượn

### A.13 Đặc tả use case “Hủy yêu cầu mượn”

Tên use case	Hủy yêu cầu mượn		
Tác nhân	Người dùng		
Tiền điều kiện	Không		
Luồng sự kiện chính	STT	Thực hiện bởi	Hành động
	1	Người dùng	Ấn chọn mục Sách của bạn
	2	Người dùng	Ấn chọn Sách mượn
	3	Người dùng	Ấn chọn Sách đang yêu cầu mượn
	4	Hệ thống	Hiển thị danh sách sách đang yêu cầu mượn của người dùng
	5	Người dùng	Ấn nút hủy tại một yêu cầu tương ứng
	6	Hệ thống	Thông báo xác nhận
	7	Người dùng	Ấn chọn xác nhận
Luồng sự kiện phát sinh	STT	Thực hiện bởi	Hành động
	7a	Người dùng	Ấn chọn hủy
Hậu điều kiện	Hệ thống cập nhật trạng thái yêu cầu mượn thành đã hủy		

**Bảng A.14:** Đặc tả use case Hủy yêu cầu mượn

### A.14 Đặc tả use case “Cập nhật thông tin sách”

Tên use case	Cập nhật thông tin sách		
Tác nhân	Người dùng		
Tiền điều kiện	Không		
Luồng sự kiện chính	STT	Thực hiện bởi	Hành động
	1	Người dùng	Ấn chọn mục Sách của bạn
	2	Người dùng	Ấn chọn mục Sách tự thêm
	3	Hệ thống	Hiển thị danh sách sách người dùng đã thêm vào hệ thống
	4	Người dùng	Ấn chọn cập nhật một cuốn sách

	5	Hệ thống	Hiển thị giao diện nhập thông tin sách
	6	Người dùng	Nhập thông tin sách
	7	Người dùng	Ấn nút Lưu
	8	Hệ thống	Thông báo cập nhật thành công
Luồng sự kiện phát sinh	STT	Thực hiện bởi	Hành động
	8a	Hệ thống	Hiển thị các trường dữ liệu không hợp lệ
Hậu điều kiện			Hệ thống lưu thông tin sách đã cập nhật

**Bảng A.15:** Đặc tả use case Cập nhật thông tin sách**A.15 Đặc tả use case “Khóa tài khoản”**

Tên use case	Khóa tài khoản		
Tác nhân	Quản trị viên		
Tiền điều kiện	Tài khoản bị khóa đang ở trạng thái Hoạt động		
Luồng sự kiện chính	STT	Thực hiện bởi	Hành động
	1	Quản trị viên	Ấn chọn Quản lý tài khoản
	2	Hệ thống	Hiển thị danh sách tài khoản của người dùng
	3	Quản trị viên	Ấn nút khóa một tài khoản
	4	Hệ thống	Thông báo xác nhận
	5	Quản trị viên	Ấn chọn xác nhận
	6	Hệ thống	Thông báo khóa tài khoản thành công
	7	Hệ thống	Thông báo hỏi có muốn gửi email lý do khóa tài khoản hay không
	8	Quản trị viên	Nhập lý do khóa tài khoản
	9	Quản trị viên	Ấn nút Gửi
	10	Hệ thống	Thông báo Gửi email thành công
Luồng sự kiện phát sinh	STT	Thực hiện bởi	Hành động
	5a	Quản trị viên	Ấn chọn hủy
	8a	Quản trị viên	Ấn chọn đóng
Hậu điều kiện			Thông báo có lỗi xảy ra khi gửi email
			Hệ thống cập nhật trạng thái tài khoản thành bị khóa

**Bảng A.16:** Đặc tả use case Khóa tài khoản

**A.16 Đặc tả use case “Mở khóa tài khoản”**

Tên use case	Mở khóa tài khoản		
Tác nhân	Quản trị viên		
Tiền điều kiện	Tài khoản mở khóa đang ở trạng thái bị khóa		
Luồng sự kiện chính	STT	Thực hiện bởi	Hành động
	1	Quản trị viên	Ấn chọn Quản lý tài khoản
	2	Hệ thống	Hiển thị danh sách tài khoản của người dùng
	3	Quản trị viên	Ấn nút khóa một tài khoản
	4	Hệ thống	Thông báo xác nhận
	5	Quản trị viên	Ấn chọn xác nhận
	6	Hệ thống	Thông báo mở khóa tài khoản thành công
Luồng sự kiện phát sinh	STT	Thực hiện bởi	Hành động
	7a	Quản trị viên	Ấn chọn hủy
Hậu điều kiện	Hệ thống cập nhật trạng thái tài khoản thành hoạt động		

**Bảng A.17:** Đặc tả use case Mở khóa tài khoản

## B. ÁNH XẠ THỂ LOẠI CHO CÁC NGUỒN DỮ LIỆU

### B.1 Ánh xạ thể loại cho nguồn dữ liệu “fahasa.com”

Thể loại nguồn	Thể loại chung
Thiếu nhi	Thiếu nhi
Giáo khoa - Tham khảo	Sách giáo Khoa - Giáo Trình
Văn học	Văn học
Manga - Comic	Truyện Tranh
Tâm lý - Kỹ năng sống	Kỹ năng sống
Sách học ngoại ngữ	Tin Học - Ngoại Ngữ
Khoa học kỹ thuật	Khoa Học - Kỹ Thuật
Lịch Sử - Địa Lý - Tôn Giáo	Lịch sử - Địa lý - Tôn giáo
Kinh Tế	Kinh tế - Chính Trị
Nuôi Dạy Con	Nuôi dạy con
Chính Trị - Pháp Lý - Triết Học	Kinh tế - Chính Trị
Nữ Công Gia Chánh	Gia Đình
Phong Thủy - Kinh Dịch	Lịch sử - Địa lý - Tôn giáo
Tiểu Sử Hồi Ký	Tiểu sử - Hồi ký
Văn Hóa - Nghệ Thuật - Du Lịch	Văn Hóa - Nghệ thuật
Từ điển	Từ Điển
Báo - Tạp Chí	Báo - Tạp Chí
Âm Nhạc - Mỹ Thuật - Thời Trang	Văn Hóa - Nghệ thuật
Thể Dục Thể thao - Giải Trí	Sức khỏe
Giáo trình	Sách giáo Khoa - Giáo Trình

**Bảng B.1:** Ánh xạ thể loại cho nguồn dữ liệu “fahasa.com”

### B.2 Ánh xạ thể loại cho nguồn dữ liệu “khaitam.com”

Thể loại nguồn	Thể loại chung
Văn học	Văn học
Sách gia đình	Gia Đình
Kinh tế - Chính trị	Kinh tế - Chính Trị
Tâm lý - Nghệ thuật sống	Kỹ năng sống
Văn hóa - Nghệ thuật	Văn Hóa - Nghệ thuật
Kỹ năng	Kỹ năng sống
Lịch sử	Lịch sử - Địa lý - Tôn giáo

Y học	Sức khỏe
-------	----------

Bảng B.2: Ánh xạ thể loại cho nguồn dữ liệu “khaitam.com”

### B.3 Ánh xạ thể loại cho nguồn dữ liệu “nhasachphuongnam.com”

Thể loại nguồn	Thể loại chung
Thiếu Nhi	Thiếu nhi
Children’s Books	Thiếu nhi
Truyện Tranh/Manga/Comic	Truyện Tranh
Văn Học	Văn học
Kinh Tế/Kinh Doanh	Kinh tế - Chính Trị
Sách Giáo Khoa	Sách giáo Khoa - Giáo Trình
Kỹ Năng/Sống Đẹp	Kỹ năng sống
Teen	Tâm lý - Giới tính
Sách Học Ngoại Ngữ/Từ Điển	Tin Học - Ngoại Ngữ
Tâm Lý/Giáo Dục	Kỹ năng sống
Tâm Linh/Tôn Giáo	Lịch sử - Địa lý - Tôn giáo
Sách Chuyên Ngành	Khoa Học - Kỹ Thuật
Thường Thức/Đời Sống	Gia Đình
Kiến Trúc/Hội Họa/Điện Ảnh	Văn Hóa - Nghệ thuật
Văn Hoá/Du Lịch	Văn Hóa - Nghệ thuật
Self-help & Psychology	Kỹ năng sống
Tạp Chí	Báo - Tạp Chí

Bảng B.3: Ánh xạ thể loại cho nguồn dữ liệu “nhasachphuongnam.com”

### B.4 Ánh xạ thể loại cho nguồn dữ liệu “tiki.vn”

Thể loại nguồn	Thể loại chung
Khoa Học - Kỹ Thuật	Khoa Học - Kỹ Thuật
Sách tham Khảo	Sách giáo Khoa - Giáo Trình
Chính Trị - Pháp Lý	Kinh tế - Chính Trị
Sách giáo Khoa - Giáo Trình	Sách giáo Khoa - Giáo Trình
Tôn Giáo - Tâm Linh	Lịch sử - Địa lý - Tôn giáo
Thiếu nhi	Thiếu nhi
Lịch sử	Lịch sử - Địa lý - Tôn giáo
Học Ngoại Ngữ	Tin Học - Ngoại Ngữ
Y Học	Sức khỏe
Thường Thức - Gia Đình	Gia Đình

## PHỤ LỤC B. ÁNH XẠ THỂ LOẠI CHO CÁC NGUỒN DỮ LIỆU

Bà mẹ - Em bé	Nuôi dạy con
Kinh tế	Kinh tế - Chính Trị
Kỹ năng sống	Kỹ năng sống
Văn học	Văn học
Văn Hóa - Địa Lý - Du Lịch	Văn Hóa - Nghệ thuật
Truyện Tranh	Truyện Tranh
Điện Ảnh - Nhạc - Họa	Văn Hóa - Nghệ thuật
Từ Điển	Từ Điển
Tạp Chí - Catalogue	Báo - Tạp Chí
Tâm lý - Giới tính	Tâm lý - Giới tính
Thể Dục - Thể Thao	Sức khỏe
Công Nghệ Thông Tin	Tin Học - Ngoại Ngữ

**Bảng B.4:** Ánh xạ thể loại cho nguồn dữ liệu “tiki.vn”

### B.5 Ánh xạ thể loại cho nguồn dữ liệu “vinabook.com”

Thể loại nguồn	Thể loại chung
Sách Chuyên Ngành	Sách giáo Khoa - Giáo Trình
Sách Thiếu Nhi	Thiếu nhi
Sách Văn học Nước Ngoài	Văn học
Sách Văn học Trong Nước	Văn học
Sách Tin Học - Ngoại Ngữ	Tin Học - Ngoại Ngữ
Sách Phát Triển Bản Thân	Kỹ năng sống
Sách Kinh Tế	Kinh tế - Chính Trị
Sách Giáo Khoa - Giáo Trình	Sách giáo Khoa - Giáo Trình
Tạp chí	Báo - Tạp Chí
Sách Giáo Trình Cao Đẳng - Đại Học	Sách giáo Khoa - Giáo Trình
Sách Ngoại Văn	Văn học

**Bảng B.5:** Ánh xạ thể loại cho nguồn dữ liệu “vinabook.com”