

ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN
KHOA HỆ THỐNG THÔNG TIN

NGUYỄN ĐỨC TÙNG – 16521396

HÀ THANH ĐOÀN - 16520219

XÂY DỰNG HỆ THỐNG GỢI Ý CHO
WEBSITE REVIEW NHÀ HÀNG/QUÁN ĂN
RECOMMENDER SYSTEM FOR
RESTAURANTS/EATERIES REVIEWING WEBSITE

KỸ SƯ NGÀNH HỆ THỐNG THÔNG TIN

GIẢNG VIÊN HƯỚNG DẪN

THS. THÁI BẢO TRÂN

TP. HỒ CHÍ MINH, 12/2020

DANH SÁCH HỘI ĐỒNG BẢO VỆ KHÓA LUẬN

Hội đồng chấm khóa luận tốt nghiệp, thành lập theo Quyết định số
ngày của Hiệu trưởng Trường Đại học Công nghệ Thông tin.

1. – Chủ tịch.
2. – Thư ký.
3. – Ủy viên.
4. – Ủy viên.

ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC
CÔNG NGHỆ THÔNG TIN

CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM
Độc Lập - Tự Do - Hạnh Phúc

TP. HCM, ngày.....tháng.....năm.....

NHẬN XÉT KHÓA LUẬN TỐT NGHIỆP

(CỦA CÁN BỘ HƯỚNG DẪN)

Tên khóa luận:

XÂY DỰNG HỆ THỐNG GỢI Ý CHO WEBSITE REVIEW NHÀ HÀNG/QUÁN ĂN

Nhóm SV thực hiện:

Nguyễn Đức Tùng

16521396

Hà Thanh Đoàn

16520219

Cán bộ hướng dẫn:

ThS. Thái Bảo Trân

Đánh Giá Khóa Luận:

1. Về cuốn báo cáo:

Số trang

Số chương

Số bảng số liệu

Số hình vẽ

Số tài liệu tham khảo

Sản phẩm

Một số nhận xét về hình thức cuốn báo cáo:

2. Về nội dung nghiên cứu:

3. Về chương trình ứng dụng:

4. Về thái độ làm việc của sinh viên:

Đánh giá chung:

Điểm từng sinh viên:

Nguyễn Đức Tùng:

Hà Thanh Đoàn:

Người nhận xét

ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC
CÔNG NGHỆ THÔNG TIN

CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM
Độc Lập - Tự Do - Hạnh Phúc

TP. HCM, ngày.....tháng.....năm.....

NHẬN XÉT KHÓA LUẬN TỐT NGHIỆP

(CỦA CÁN BỘ PHẢN BIỆN)

Tên khóa luận:

XÂY DỰNG HỆ THỐNG GỢI Ý CHO WEBSITE REVIEW NHÀ HÀNG/QUÁN ĂN

Nhóm SV thực hiện:

Cán bộ phản biện:

Nguyễn Đức Tùng	16521396
Hà Thanh Đoàn	16520219

Đánh giá Khóa luận

1. Về cuốn báo cáo:

Số trang	Số chương
----------	-----------

Số bảng số liệu	Số hình vẽ
-----------------	------------

Số tài liệu tham khảo	Sản phẩm
-----------------------	----------

Một số nhận xét về hình thức cuốn báo cáo:

2. Về nội dung nghiên cứu:

3. Về chương trình ứng dụng:

4. Về thái độ làm việc của sinh viên:

Đánh giá chung:

Điểm từng sinh viên:

Nguyễn Đức Tùng:

Hà Thanh Đoàn:

Người nhận xét

LỜI CẢM ƠN

Chúng em xin gửi lời cảm ơn sâu sắc đến cô Thái Bảo Trân. Cô là người đã hướng dẫn, chỉ lối và gợi ý để chúng em có thể để hoàn thành khóa luận này. Cô luôn nhiệt tình liên lạc, trao đổi và tận tình góp ý giúp chúng em hoàn thành khóa luận với kết quả tốt nhất.

Chúng em xin chân thành biết ơn sự tận tình dạy dỗ và sự giúp đỡ của tất cả quý thầy cô tại trường Đại học Công Nghệ Thông Tin, đặc biệt là các thầy cô trong khoa Hệ Thống Thông Tin, các thầy cô đã tận tình truyền đạt những kiến thức bổ ích, quý giá cho chúng em trong suốt hơn 4 năm học qua, những kiến thức học được từ giảng đường sẽ là hành trang quý báu để chúng em có thể tiếp tục trên con đường học tập, nghiên cứu, làm việc cũng như trong quá trình hoàn thiện bản thân của mình.

Chúng em xin chân thành cảm ơn quý thầy cô!

Xin trân trọng cảm ơn!

ĐỀ CƯƠNG CHI TIẾT

TÊN ĐỀ TÀI:

Xây dựng hệ thống gợi ý cho website review nhà hàng/quán ăn
(Recommender system for restaurants/eateries reviewing website)

Cán bộ hướng dẫn: ThS. Thái Bảo Trân

Thời gian thực hiện: Từ 9/2020 đến 12/2020

Sinh viên thực hiện:

Nguyễn Đức Tùng - 16521396

Hà Thanh Đoàn - 16520219

NỘI DUNG ĐỀ TÀI

1. Lý do chọn đề tài

- Nhu cầu ăn uống là một trong những nhu cầu cần thiết trong cuộc sống. Thế nhưng rất khó để tìm quán ăn phù hợp với sở thích về nhiều yếu tố như quang cảnh, chất lượng, giá cả,... Đặc biệt đối với khách du lịch, việc tìm kiếm một nhà hàng/quán ăn phù hợp luôn là một trong những điều được quan tâm hàng đầu.
- Ăn gì, ở đâu? Là bài toán không còn mới. Tuy nhiên, với một số website review nhà hàng hiện nay, người dùng phải tìm kiếm từng nhà hàng, xem mô tả, đọc các bình

luận,.. Điều này thật sự tốn thời gian. Hơn nữa, việc chỉ dựa vào các đánh giá (chấm điểm, bình luận) của những khách hàng trước đó sẽ không hoàn toàn phù hợp với nhu cầu vì mỗi người có một sở thích, khẩu vị riêng. Do đó, việc website tự động đề xuất những nhà hàng/quán ăn phù hợp với người dùng hoặc gợi ý những nhà hàng tương tự sẽ giúp người dùng dễ dàng tiếp cận những nơi phù hợp với cá nhân.

- Qua khảo sát các website trong nước, chúng tôi nhận thấy hiện nay chưa có website nào xây dựng được chức năng đề xuất, đa số các website đơn giản chỉ hiển thị các nhà hàng gần/mới nhất mà không quan tâm đến nhu cầu của người dùng. Vì vậy, chúng tôi chọn đề tài này với mong muốn thiết kế một trang web, ngoài những chức năng của một website review nhà hàng, hệ thống tập trung giải quyết thêm vấn đề khuyến nghị cho người dùng có thể tìm kiếm quán ăn phù hợp với sở thích của mình nhất.

2. Mục tiêu đề tài

- Nhóm người dùng hướng đến: Dành cho tất cả mọi người, đặc biệt giới trẻ.
- Áp dụng các kỹ thuật trong khai thác dữ liệu để xây dựng hệ thống gợi ý cho website, nhằm mang lại những đánh giá đúng nhất về các nhà hàng dựa trên trải nghiệm người dùng đã chia sẻ.
- Ngoài ra, để tạo ra sự thuận tiện cho người dùng, chúng tôi sẽ tích hợp thêm vào hệ thống chức năng tìm kiếm bằng giọng nói.

3. Phạm vi đề tài

Đề tài tập trung vào các nội dung chính sau:

- Các chức năng cơ bản của website (hiển thị, quản lý, thống kê,...)
- Hệ thống gợi ý nhà hàng phù hợp nhất với người dùng
- Tìm kiếm bằng giọng nói.
- Tư vấn các nhà hàng có khoảng cách gần người dùng nhất.
- Lưu lịch sử thao tác người dùng như: những nhà hàng mà người dùng vào xem, thời gian xem,...

- Hiện tại chúng tôi chỉ xây dựng hệ thống gợi ý dựa trên các dữ liệu đánh giá thu thập được từ khách hàng đã trải nghiệm dịch vụ. Do đó, kết quả hệ thống gợi ý sẽ không khai thác tối đa lợi ích đối với địa điểm không có hoặc có ít thông tin khách hàng đánh giá sau khi họ trải nghiệm.

4. Đối tượng

- Tất cả mọi người.

5. Phương pháp thực hiện

- Khảo sát và phân tích nhu cầu của người dùng.
- Thiết kế lược đồ Cơ Sở Dữ Liệu (CSDL) và xây dựng quy trình nghiệp vụ.
- Tìm hiểu các công nghệ áp dụng để xây dựng các chức năng phù hợp.
- Ứng dụng các thuật toán phù hợp để xây dựng hệ thống gợi ý.
- Xây dựng chức năng tìm kiếm giọng nói bằng một số kỹ thuật xử lý ngôn ngữ tự nhiên.

6. Kết quả mong đợi của đề tài

- Website trực quan, dễ sử dụng, giao diện thân thiện với mọi thiết bị.
- Xây dựng được hệ thống quản lý website (admin)
- Hệ thống gợi ý và tìm kiếm giọng nói với độ chính xác cao.
- Đề tài có tính ứng dụng thực tiễn cao. Đáp ứng nhu cầu của người dùng.

KẾ HOẠCH THỰC HIỆN

Công việc cần thực hiện	Thời gian
<ul style="list-style-type: none"> - Khảo sát hiện trạng - Viết mô tả cho đề tài - Tìm hiểu các công nghệ liên quan áp dụng cho đề tài. 	1 tuần
<ul style="list-style-type: none"> - Thu thập dữ liệu. - Tìm hiểu và thiết kế CSDL cho hệ thống 	1 tuần

- Tìm hiểu các thuật toán, kỹ thuật và hiện thực với dữ liệu có sẵn.	3 tuần
- Thiết kế giao diện cho webiste	2 tuần
- Lập trình backend.	3 tuần
- Kiểm tra và sửa lỗi phát sinh. Cải thiện trang web	2 tuần
- Triển khai hệ thống	1 tuần
- Hoàn chỉnh báo cáo đề tài	2 tuần
Tổng cộng	15 tuần
<p>Xác nhận của CBHD</p> <p>(Ký tên và ghi rõ họ tên)</p> <p>Thái Bảo Trân</p>	<p>TP. HCM, 26 / 07 / 2020</p> <p>Sinh viên</p> <p>(Ký tên và ghi rõ họ tên)</p>

MỤC LỤC

LỜI CẢM ƠN	i
ĐỀ CƯƠNG CHI TIẾT	ii
DANH MỤC CÁC BẢNG.....	iii
DANH MỤC CÁC HÌNH.....	iv
DANH MỤC THUẬT NGỮ VÀ TỪ VIẾT TẮT	v
TÓM TẮT	vi
CHƯƠNG 1: TỔNG QUAN.....	2
1.1. Đặt vấn đề	2
1.2. Mục tiêu đề tài	3
1.3. Khảo sát hiện trạng	4
1.3.1. Khảo sát website review nhà hàng/quán ăn	4
1.3.2. Các nghiên cứu liên quan	6
1.4. Đối tượng, phạm vi nghiên cứu	6
1.4.1. Đối tượng.....	6
1.4.2. Phạm vi nghiên cứu	6
1.5. Nội dung thực hiện	6
1.6. Bố cục báo cáo.....	7
CHƯƠNG 2: CƠ SỞ LÝ THUYẾT	9
2.1. Tổng quan về cơ sở dữ liệu MySQL	9
2.1.1. Giới thiệu về MySQL	9
2.1.2. Phương thức hoạt động của MySQL.....	9
2.1.3. Câu lệnh và kiểu dữ liệu	10
2.2. Công thức Cosine trong tính toán mức độ tương đồng giữa hai vector	11
2.3. Thuật toán lọc cộng tác(Collaborative Filtering):	13

2.3.1.	Tổng quan thuật toán	14
2.3.2.	Các bước thực thi	15
2.3.2.1.	Cách tiếp cận User-based filtering:	15
2.3.2.2.	Cách tiếp cận Item-based filtering:.....	16
2.3.3.	Nhận xét hai cách tiếp cận User-based filtering và Item-based filtering	19
2.4.	Thuật toán Tf-idf.	20
2.5.	Công thức Haversine:	22
2.6.	Các công cụ hỗ trợ hệ thống.....	23
2.6.1.	Python.....	23
2.6.1.1.	Giới thiệu	23
2.6.1.2.	Đặc điểm.....	23
2.6.1.3.	Ưu điểm	24
2.6.1.4.	Nhược điểm	24
2.6.2.	Php.....	25
2.6.2.1.	Giới thiệu	25
2.6.2.2.	Đặc điểm.....	25
2.6.2.3.	Ưu điểm	26
2.6.2.4.	Nhược điểm	26
CHƯƠNG 3: PHÂN TÍCH VÀ THIẾT KẾ HỆ THỐNG		27
3.1.	Phân tích hiện trạng nhu cầu	27
3.1.1.	Yêu cầu chức năng	27
3.1.2.	Yêu cầu phi chức năng:	27
3.2.	Phân tích và thiết kế thành phần dữ liệu.....	28
3.3.	Phân tích thiết kế thành phần xử lý	32
3.3.1.	Sơ đồ Usecase.....	32
3.3.2.	Danh sách Usecase	33
3.3.3.	Sơ đồ tuần tự.....	36

3.4. Xây dựng giao diện:	59
CHƯƠNG 4: CÀI ĐẶT VÀ TRIỂN KHAI HỆ THỐNG	64
4.1. Triển khai server trên Heroku Cloud Platform và dịch vụ Azure Database for MySQL.....	64
4.2. Giải pháp công nghệ phía Client.....	65
4.2.1. Xác thực với Json Web Token (JWT).....	65
4.2.2. Công cụ ghi giọng nói	67
4.3. Giải pháp công nghệ phía Server	67
4.3.1. Ứng dụng thuật toán lọc cộng tác	67
4.3.2. Ứng dụng thuật toán Tf-idf trong điều khiển bằng giọng nói:.....	75
4.3.3. Tìm kiếm nhà hàng có khoảng cách gần nhất.....	78
CHƯƠNG 5: KẾT LUẬN	80
5.1. Kết quả đạt được.....	80
5.2. Ưu điểm của hệ thống:.....	80
5.3. Hạn chế của hệ thống.....	80
5.4. Hướng phát triển.....	81
TÀI LIỆU THAM KHẢO.....	82

DANH MỤC CÁC BẢNG

Bảng 2.1. Ma trận đánh giá	12
Bảng 2.2. Ưu và nhược điểm giữa Top-N và Threshold.....	17
Bảng 2.3. Ví dụ thuật toán tf-idf	21
Bảng 3.1. Mô tả các thuộc tính	31
Bảng 3.2. Danh sách các usecase	35
Bảng 4.1. Ma trận đánh giá	69
Bảng 4.2. Ma trận chuẩn hóa	70
Bảng 4.3 Ma trận tương đồng	71
Bảng 4.4. Tập dữ liệu huấn luyện	75

DANH MỤC CÁC HÌNH

Hình 1.0 Màn hình trang web riviuvn.....	4
Hình 1.1 Giao diện website foody.vn	5
Hình 1.2. Phương thức hoạt động MySQL	9
Hình 2.1 Biểu diễn các vector trên tọa độ Oxy	12
Hình 2.2 Luồng hoạt động của hai cách tiếp cận	15
Hình 2.3 Luồng hoạt động của User-based filtering	16
Hình 2.4 So sánh hai phương pháp Top-N và Threshold	18
Hình 3.1. Sơ đồ usecase	32
Hình 3.2. Sơ đồ tuần tự quản lý admin	36
Hình 3.3. Thêm tài khoản.....	37
Hình 3.4. Sửa tài khoản.....	38
Hình 3.5. Xóa tài khoản	39
Hình 3.6. Quản lý tài khoản	40
Hình 3.7. Thêm tài khoản.....	41
Hình 3.8. Sửa tài khoản.....	42
Hình 3.9. Xóa tài khoản	43
Hình 3.10. Tạo nhà hàng	44
Hình 3.11. Duyệt nhà hàng	45
Hình 3.12. Quản lý nhà hàng	46
Hình 3.13 Sửa nhà hàng	47
Hình 3.14. Xóa nhà hàng.....	48
Hình 3.15. Thống kê nhà hàng	49
Hình 3.16. Đăng ký tài khoản	50
Hình 3.17 Tìm lại mật khẩu	51
Hình 3.18. Đăng nhập	52

Hình 3.19 Đổi mật khẩu	53
Hình 3.20 Quản lý thông tin cá nhân	54
Hình 3.21. Tạo nhà hàng	55
Hình 3.22 Điều khiển giọng nói	56
Hình 3.23 Tìm kiếm nhà hàng.....	57
Hình 3.24. Đánh giá nhà hàng.....	58
Hình 3.25 Giao diện trang chủ	59
Hình 3.26 Giao diện trang chi tiết.....	60
Hình 3.27. Màn hình đánh giá.....	61
Hình 3.28. Màn hình quản lý tài khoản.....	62
Hình 3.29. Màn hình thêm mới địa điểm	63
Hình 4.1. Luồng hoạt động của jwt.....	66
Hình 4.2. Ứng dụng thuật toán lọc cộng tác	68
Hình 4.3. Hiện thị các nhà hàng gần nhất	79

DANH MỤC THUẬT NGỮ VÀ TỪ VIẾT TẮT

STT	Từ viết tắt	Nội dung
1	neighbors	Các item có độ tương đồng cao nhất
2	TF	Term Frequency
3	IDF	Term Frequency
4	CSDL	Cơ sở dữ liệu
5	SQL	Structured Query Language
6	DML	Data Manipulation Language
7	DDL	Data Definition Language
8	DCL	Data Control Language
9	Client	Máy khách
10	Server	Máy chủ

TÓM TẮT

Đề tài khóa luận “Xây dựng hệ thống gợi ý cho website review nhà hàng/quán ăn” mang tính thực tế, hướng đến những đối tượng có nhu cầu tìm các nhà hàng/quán ăn hợp với sở thích hoặc giá tiền nhưng cần các nhận xét hoặc gợi ý chân thực đến từ những người đã từng trải nghiệm các nhà hàng này.

Sau khi tìm hiểu, khảo sát các ứng dụng review nhà hàng/quán ăn, nhóm tác giả đưa ra hướng giải quyết sau:

- Tìm hiểu và xây dựng phương pháp khuyến nghị dựa vào lịch sử nhận xét hoặc vị trí của người dùng, đồng thời tích hợp điều khiển bằng giọng nói.
- Xây dựng ứng dụng với chức năng tìm kiếm và đọc review với giao diện thân thiện, dễ sử dụng. Ngoài ra, ứng dụng còn cung cấp các chức năng khác như: đăng nhập, đăng xuất, tìm kiếm, đánh giá nhà hàng, chia sẻ nhà hàng, tạo nhà hàng.
- Công nghệ sử dụng xây dựng hệ thống: Python, Php (Laravel Framework), Heroku Cloud Platform.

CHƯƠNG 1: TỔNG QUAN

1.1. Đặt vấn đề

Ngay từ khi những dấu chân đầu tiên của con người in dấu lên Trái Đất, ngay từ thời kỳ bình minh của nhân loại, việc ăn uống đã được coi như là một nhu cầu cần thiết để duy trì sự sống của sinh vật nói chung và con người nói riêng. Tuy nhiên, ở thời kỳ cổ đại đó, thức ăn vẫn còn khan hiếm và đơn giản, buộc con người không có quyền lựa chọn thức ăn. Sau này, trải qua hàng triệu năm tiến hóa, thế giới ngày một văn minh hơn. Do đó, những tri thức cơ bản đầu tiên về lĩnh vực ăn uống được hình thành, tạo nên khái niệm đầu tiên về văn hóa ăn uống: Ẩm thực.

Xã hội càng văn minh, càng hiện đại, con người lại càng có góc nhìn và nhu cầu khắt khe hơn về ẩm thực, không còn gói gọn như câu nói ngày xưa của ông bà ta “Ăn no mặc ấm” mà bây giờ đã chuyển thành “Ăn ngon mặc đẹp”.

Theo thống kê của Dcorp R-Keeper Việt Nam¹ vào năm 2019 cho biết cả nước hiện có đến 540.000 cửa hàng ăn uống, trong đó có khoảng 430.000 cửa hàng nhỏ, 7.000 nhà hàng chuyên dịch vụ thức ăn nhanh, 22.000 cửa hàng cà phê, các quầy bar và trên 80.000 nhà hàng được đầu tư, phát triển một cách bài bản. Qua từng năm, con số này tăng lên một cách nhanh chóng và đáng kể.

Theo thống kê của Vnetwork², tính đến tháng 1 năm 2020, có khoảng 68,17 triệu người Việt Nam sử dụng internet, chiếm khoảng 70% dân số cả nước. Điều này chứng tỏ rằng người Việt Nam sử dụng internet trong cuộc sống rất nhiều. Và tất nhiên việc tìm kiếm nhà hàng/quán ăn bằng internet cũng không ngoại lệ.

¹ Tham khảo tại: <https://vietnambiz.vn/nganh-fb-viet-nam-2025-co-the-dat-408-ti-usd-doanh-thu-201912071250533.htm>

² Tham khảo tại: <https://vnetwork.vn/news/thong-ke-internet-viet-nam-2020>

Với sự phát triển của công nghệ thông tin, việc tìm kiếm nhà hàng/quán ăn sẽ dễ dàng hơn cho mọi người vì chỉ cần lên website và tra trên Google. Nhưng rất khó tìm được những trang uy tín và có nhận xét chân thực.

Nhằm hỗ trợ cho việc tìm kiếm nhà hàng/quán ăn, đặc biệt dành cho người Việt, nhóm tác giả cố gắng tạo ra một phương pháp giúp người dùng có thể tìm kiếm nhà hàng/quán ăn phù hợp với mình, đồng thời nhận xét và đánh giá các nhà hàng mà chính người dùng đã trải nghiệm. Từ đó tạo nên một cộng đồng chia sẻ, đánh giá nhà hàng/quán ăn đáng tin cậy.

1.2. Mục tiêu đề tài

Mặc dù đã có rất nhiều trang quảng cáo hoặc nhận xét về chất lượng cũng như giá cả của nhà hàng/quán ăn, nhưng hầu hết lại khá rời rạc và hoàn toàn không thuần túy là trang review nhà hàng/quán ăn. Vì vậy trang web này sẽ giúp người dùng tìm được nhà hàng/quán ăn hợp lý cũng như đưa ra các nhận xét dựa theo trải nghiệm của bản thân.

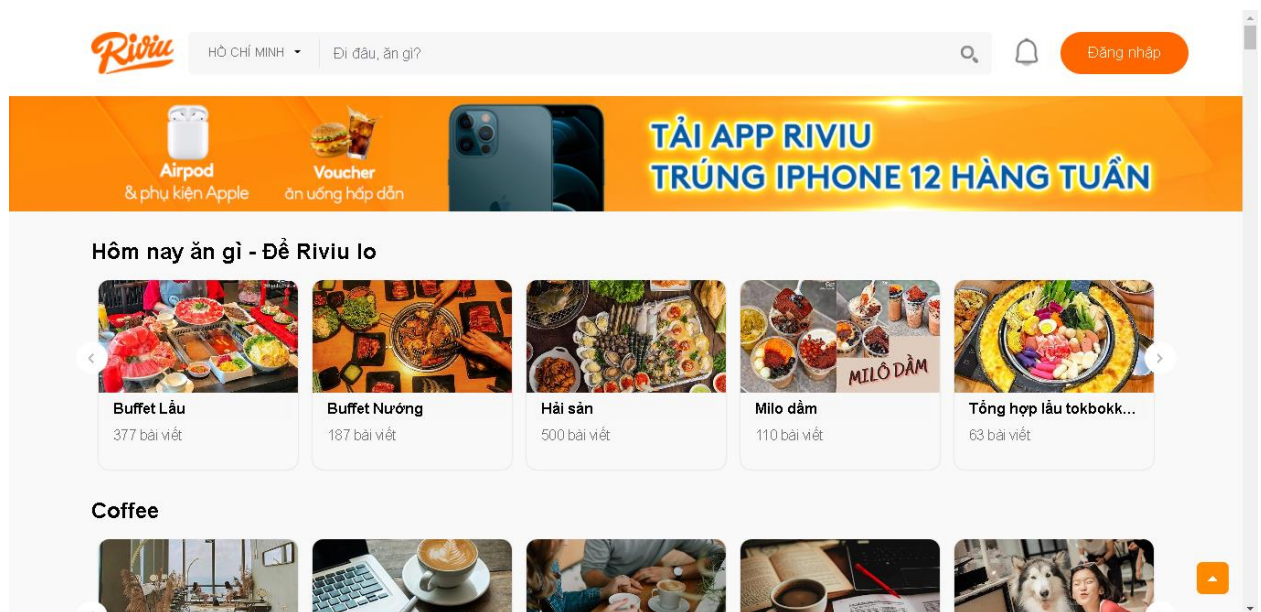
Để tiện lợi cho việc tìm kiếm, nhóm hướng tới một website review nhà hàng/quán ăn để giúp người dùng có thể dễ dàng truy cập trên máy tính, laptop hoặc cả trên điện thoại. Website giúp người dùng dễ dàng sử dụng để tìm kiếm nhà hàng/quán ăn mà mình cần đến, những quán ăn xung quanh mình hoặc đưa ra những thang điểm, nhận xét về nhà hàng mình vừa ăn. Bên cạnh đó, ứng dụng tạo nên một cộng đồng cùng nhau bàn luận cũng như gợi ý về các nhà hàng/quán ăn để tiện lợi hơn cho việc chọn lựa.

1.3. Khảo sát hiện trạng

1.3.1. Khảo sát website review nhà hàng/quán ăn ¹

➤ riviuvn

Đúng như tên gọi, riviuvn là website ẩm thực, du lịch hàng đầu Việt Nam là diễn đàn để cộng đồng thảo luận, chia sẻ những bài review về những hàng quán ngon tuyệt của dân “sành ăn” và những kinh nghiệm du lịch tuyệt vời đã được cộng đồng trải nghiệm ở khắp mọi miền đất nước.



Hình 1.0 Màn hình trang web riviuvn

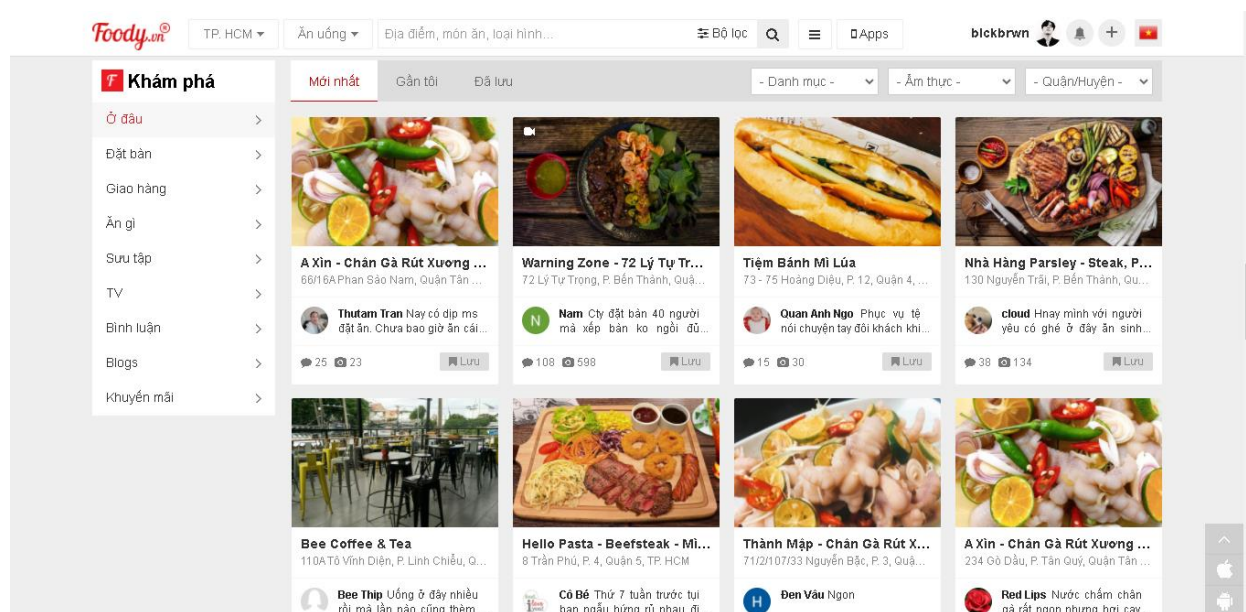
➤ foody.vn

Được xây dựng từ giữa năm 2012 tại TP. HCM, Việt Nam, Foody là cộng đồng tin cậy cho mọi người có thể tìm kiếm, đánh giá, bình luận các địa điểm ăn uống: nhà hàng,

¹ Tham khảo tại: <https://www.cooky.vn/blog/top-10-website-am-thuc-noi-tieng-nhat-viet-nam-2934>

quán ăn, cafe, bar, karaoke, tiệm bánh, khu du lịch... tại Việt Nam, từ website hoặc ứng dụng di động. Tất cả thành viên từ Bắc đến Nam, Foody kết nối những thực khách đến với các địa điểm ăn uống lớn nhỏ cả đất nước.

Đến thời điểm hiện tại, Foody với hàng chục ngàn địa điểm và hàng ngàn bình luận, hình ảnh tại Việt Nam ở hầu hết các tỉnh thành. Foody là cách dễ nhất để bạn có thể tìm kiếm và lựa chọn địa điểm tốt nhất cho mình và bạn bè.



Hình 1.1 Giao diện website foody.vn

Nhận xét: Nhìn chung, chúng ta thấy được các website review nhà hàng/quán ăn hiện nay đều có chức năng chính là hiển thị các nhận xét cũng như số điểm từ người dùng chấm cho các quán ăn. Tuy nhiên, việc tìm kiếm vẫn hoàn toàn phụ thuộc vào người dùng và chưa có bất kỳ một sự gợi ý tự động nào đến từ phía website. Đối với những người dùng không quyết đoán trong việc nên đi đến quán ăn nào thì việc gợi ý khá là quan trọng. Đồng thời, các website trên hoàn toàn chưa có chức năng điều khiển bằng

giọng nói. Việc sử dụng hệ thống gợi ý cũng như điều khiển qua giọng nói sẽ giúp người dùng dễ dàng hơn trong việc tìm kiếm cũng như đưa ra quyết định khi tìm kiếm nhà hàng/quán ăn.

1.3.2. Các nghiên cứu liên quan

- Thuật toán khuyến nghị Lọc cộng tác (Collaborative Filtering).
- Thuật toán Tf-idf.
- Công thức Haversine.

1.4. Đối tượng, phạm vi nghiên cứu

1.4.1. Đối tượng

Mọi đối tượng có sử dụng website và có nhu cầu tìm kiếm các nhà hàng/quán ăn cũng như đưa ra nhận xét, thang điểm cho quán ăn đã trải nghiệm.

1.4.2. Phạm vi nghiên cứu

Về công nghệ:

- Ngôn ngữ lập trình: Python, Php.
- Thư viện: Scikit learn.
- Framework: PhpStorm, PyCharm.
- Hệ quản trị cơ sở dữ liệu: MySQL
- Quản lý source code với Git.

Về nghiên cứu:

- Thuật toán khuyến nghị Lọc cộng tác (Collaborative Filtering).
- Thuật toán Điều khiển bằng giọng nói (Tf-idf).

1.5. Nội dung thực hiện

Bước 1:

- Khảo sát các website review nhà hàng/quán ăn.
- Xác định các chức năng cần có và thứ tự ưu tiên.

- Tìm hiểu các công nghệ, thư viện hỗ trợ việc lập trình ứng dụng.
- Tìm kiếm và nghiên cứu các nguồn dữ liệu.

Bước 2:

- Thiết kế các thành phần hệ thống.
- Thiết kế cơ sở dữ liệu.
- Thiết kế server.
- Thiết kế giao diện.
- Thực hiện lấy dữ liệu.

Bước 3:

- Xây dựng server.
- Tạo API tương tác với server.
- Tạo giao diện
- Tạo các chức năng đã chọn.

Bước 4:

- Xây dựng hệ thống khuyến nghị từ thuật toán tìm hiểu.

Bước 5:

- Hoàn chỉnh và cải thiện hiệu năng hệ thống.
- Nhận xét và đánh giá.

1.6. Bộ cục báo cáo

- *Chương 1:* Tổng quan về đề tài

Giới thiệu tổng quan nhất về nội dung đề tài. Nội dung chương 1 bao gồm: Đặt vấn đề và lý do chọn đề tài, khảo sát hiện trạng, mục tiêu và phạm vi nghiên cứu, công cụ xây dựng hệ thống và bộ cục của báo cáo.

- *Chương 2: Cơ sở lý thuyết*

Giới thiệu lý thuyết, các thư viện, công cụ sử dụng trong đề tài. Trình bày ưu nhược điểm của từng mục và lí do lựa chọn. Trình bày chi tiết lý thuyết áp dụng trong đề tài.

- *Chương 3: Phân tích và thiết kế hệ thống*

Trình bày sơ đồ tổ chức của ứng dụng, phân tích thiết kế hệ thống, thiết kế cơ sở dữ liệu, thiết kế giao diện và luồng chạy của ứng dụng.

- *Chương 4: Cài đặt và triển khai hệ thống*

Trình bày những công cụ đã sử dụng để hiện thực sản phẩm của đề tài kèm ví dụ. Cách áp dụng các thuật toán vào bài toán đề xuất nhà hàng cho người dùng.

- *Chương 5: Kết luận*

Tóm tắt đề tài, thành công và hạn chế, hướng phát triển tương lai về mặt chức năng và khả năng đề xuất cho người dùng.

CHƯƠNG 2: CƠ SỞ LÝ THUYẾT

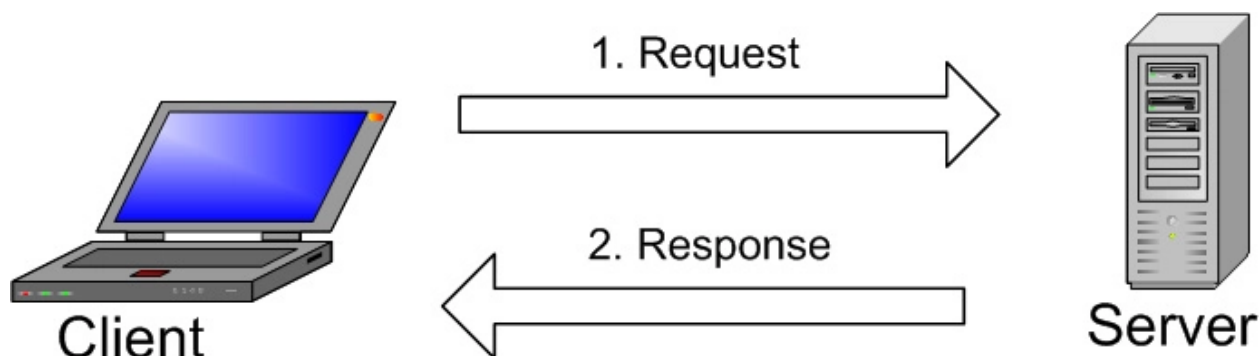
2.1. Tổng quan về cơ sở dữ liệu MySQL

2.1.1. Giới thiệu về MySQL¹

MySQL là chương trình dùng để quản lý hệ thống cơ sở dữ liệu (CSDL), trong đó CSDL là một hệ thống lưu trữ thông tin, được sắp xếp rõ ràng, phân lớp ngăn nắp những thông tin mà mình lưu trữ.

Vì thế, người dùng có thể truy cập dữ liệu một cách thuận lợi, nhanh chóng. MySQL hỗ trợ đa số các ngôn ngữ lập trình. Chính vì thế mà mã nguồn mở phổ biến nhất hiện nay là WordPress đã sử dụng MySQL làm cơ sở dữ liệu mặc định.

2.1.2. Phương thức hoạt động của MySQL²



Hình 1.2. Phương thức hoạt động MySQL

Hình ảnh trên giải thích cấu trúc cơ bản về việc giao tiếp giữa client và server. Một máy client sẽ liên lạc với máy server trong một mạng nhất định. Mỗi client có thể gửi một request từ giao diện người dùng (Graphical user interface – GUI) trên màn hình,

¹ Tham khảo tại: <http://sqladvice.com/mysql-la-gi-tong-quan-ve-he-quan-tri-du-lieu-mysql/>

² Tham khảo tại: <https://openplanning.net/10321/du-lieu-va-cau-truc-trong-mysql>

và server sẽ trả về kết quả như mong muốn. Miễn là cả hai hiểu nhau. Cách vận hành chính trong môi trường MySQL cũng như vậy:

- MySQL tạo ra bảng để lưu trữ dữ liệu, định nghĩa sự liên quan giữa các bảng đó.
- Client sẽ gửi yêu cầu SQL bằng một lệnh đặc biệt trên MySQL.
- Ứng dụng trên server sẽ phản hồi thông tin và trả về kết quả trên máy client.

2.1.3. Câu lệnh và kiểu dữ liệu¹

Có 4 loại câu lệnh trong MySQL:

- Structured Query Language (SQL) là câu lệnh truy vấn dữ liệu.
- Data Manipulation Language (DML) các câu lệnh được sử dụng để quản lý dữ liệu bên trong SCHEME: Insert, update, delete, merge, call, explain plan, lock table.
- Data Definition Language (DDL) là các câu lệnh được sử dụng để định nghĩa cấu trúc database hoặc schema: Create, alter, drop, truncate, comment, rename.
- Data Control Language (DCL) là các lệnh điều khiển truy cập dữ liệu: Grant và revoke.

MySQL có đầy đủ các loại dữ liệu cơ bản cần thiết cho việc lưu trữ:

- Các kiểu số nguyên tiêu chuẩn của SQL như INTEGER (or INT) và SMALLINT đều được hỗ trợ bởi MySQL. Và các mở rộng tiêu chuẩn, MySQL cũng hỗ trợ các kiểu số nguyên khác như TINYINT, MEDIUMINT, và BIGINT.

¹ Tham khảo tại: <https://www.hostinger.vn/huong-dan/mysql-la-gi/>

- Kiểu dữ liệu FLOAT và DOUBLE mô tả gần đúng các giá trị số thực. MySQL sử dụng 4 byte để lưu trữ dữ liệu FLOAT và 8 byte dành cho kiểu dữ liệu DOUBLE.
- Kiểu dữ liệu BIT được sử dụng để lưu trữ trường giá trị bit.
- Các kiểu dữ liệu ngày tháng và thời gian đại diện bao gồm DATE, TIME, DATETIME, TIMESTAMP, and YEAR. Mỗi kiểu có một phạm vi hợp lệ.
- Kiểu dữ liệu CHAR và VARCHAR là giống nhau, nhưng khác nhau ở cách chúng được lưu trữ và truy xuất. Chúng cũng khác nhau về chiều dài tối đa và giữ lại hay không khoảng trắng phía trước (trailing spaces).
- Các kiểu dữ liệu BINARY và VARBINARY tương tự như CHAR và VARCHAR, ngoại trừ việc chúng có chứa các chuỗi nhị phân chứ không phải là chuỗi non-binary.
- BLOB là một đối tượng nhị phân lớn (Binary Large Object) có thể chứa một lượng lớn dữ liệu. Có bốn loại BLOB, TINYBLOB, BLOB, MEDIUMBLOB, và LONGBLOB, nhưng chỉ khác nhau về độ dài tối đa của các giá trị.
- Bốn loại TEXT là TINYTEXT, TEXT, MEDIUMTEXT, và LONGTEXT. Chúng tương ứng với bốn loại BLOB và có độ dài tối đa và các yêu cầu lưu trữ tương tự.

2.2. Công thức Cosine trong tính toán mức độ tương đồng giữa hai vector

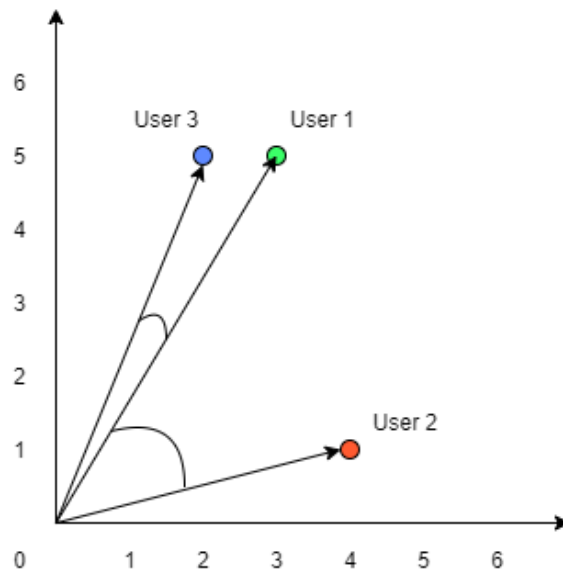
Công thức Cosine thích hợp cho dạng dữ liệu định lượng. Ví dụ dữ liệu điểm đánh giá của các nhà hàng như sau (thang điểm 10):

	Nhà hàng A	Nhà hàng B
User 1	3	5

User 2	4	1
User 3	2	5

Bảng 2.1. Ma trận đánh giá

Công thức Cosine xem các dòng của ma trận đánh giá như một vector không gian. Từ ma trận trên ta biểu diễn trong đồ thị hai chiều như sau.



Hình 2.1 Biểu diễn các vector trên tọa độ Oxy

Từ những góc giữa các vector, dễ dàng nhận ra vector của user 1 và user 3 gần hơn là giữa user 1 và user 2. Từ đó có thể nhận định rằng user 1 và user 3 có sở thích (khẩu vị) giống nhau hơn.

Công thức Cosine được tính như sau: ¹

¹ Badrul Sarwar, et al., "Item-based Collaborative Filtering Recommendation Algorithms", 08-2001

$$sim(i, j) = \cos(\vec{i}, \vec{j}) = \frac{\vec{i} \cdot \vec{j}}{\|\vec{i}\|_2 * \|\vec{j}\|_2}$$

Tuy nhiên khi nói về việc so sánh các điểm đánh giá của người dùng khác nhau, nhiều người lại sử dụng các thang điểm khác nhau khi họ đánh giá (một số người dễ tính sẵn sàng đánh giá 9 hoặc 10 điểm cho nhà hàng mà họ thích, nhưng một số người khó tính lại cho 7 hoặc 8). Do đó công thức Cosine được điều chỉnh lại như sau :¹

$$sim(i, j) = \frac{\sum_{u \in U} (R_{u,i} - \bar{R}_u)(R_{u,j} - \bar{R}_u)}{\sqrt{\sum_{u \in U} (R_{u,i} - \bar{R}_u)^2} \sqrt{\sum_{u \in U} (R_{u,j} - \bar{R}_u)^2}}.$$

Trong đó:

$sim(i, j)$: Độ tương đồng giữa hai đối tượng i và j .

U : tập người dùng(user).

$R_{u,i}$: Điểm đánh giá của user cho đối tượng i .

$R_{u,j}$: Điểm đánh giá của user cho đối tượng j .

\bar{R}_u : Điểm trung bình của user

$sim(i,j) \in [-1,1]$;

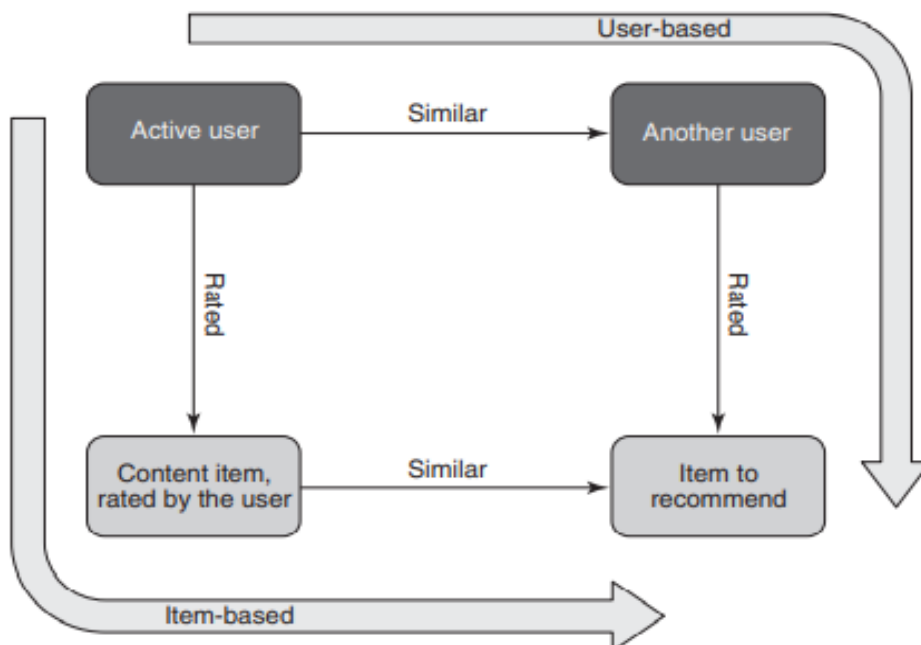
2.3. Thuật toán lọc cộng tác(Collaborative Filtering):

¹ Badrul Sarwar, et al., "Item-based Collaborative Filtering Recommendation Algorithms", 08-2001

2.3.1. Tổng quan thuật toán

Ý tưởng cơ bản của thuật toán lọc cộng tác là xác định mức độ quan tâm của một user tới một item dựa trên các users khác gần giống với user này. Việc gần giống nhau giữa các users có thể được xác định thông qua mức độ quan tâm của các users này tới các items khác mà hệ thống đã biết. Ví dụ, A, B đều thích nhà hàng X, tức đều đánh giá nhà hàng này 10/10 điểm. Ta đã biết A cũng thích nhà hàng Y, vậy nhiều khả năng B cũng thích nhà hàng này.

Thuật toán lọc cộng tác có thể được xử lý trong hai cách. Cách thứ nhất xác định mức độ quan tâm của một user tới một item dựa trên mức độ quan tâm của những user “có cùng khẩu vị” tới item đó (User-based filtering). Cách thứ hai thay vì xác định user tương đồng, hệ thống sẽ xác định các item tương đồng. Từ đó hệ thống gợi ý những item gần giống những item mà user có mức độ quan tâm cao (Item-based filtering).



Hình 2.2 Luồng hoạt động của hai cách tiếp cận¹

2.3.2. Các bước thực thi

2.3.2.1. Cách tiếp cận User-based filtering:

Bước 1: Xây dựng ma trận đánh giá.

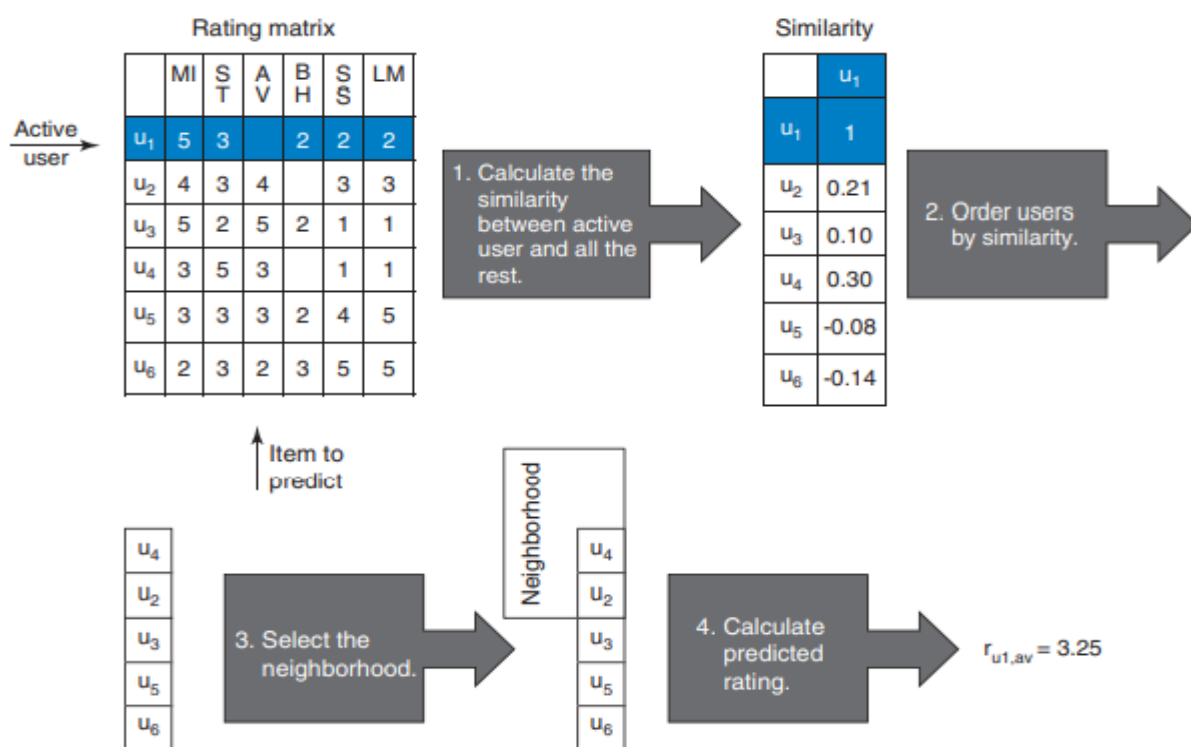
Bước 2: Tính độ tương đồng giữa user hiện tại với tất cả các user còn lại bằng công thức Cosine.

Bước 3: Sắp xếp kết quả từ bước 2 giảm dần theo độ tương đồng.

Bước 4: Chọn ra những user lân cận (neighbors). Cách chọn neighbors được trình bày rõ ở phần sau (Cách tiếp cận Item-based filtering).

¹ Kim Falk, "Practical Recommender Systems", Page 184, 2019

Bước 5: Dự đoán điểm đánh giá của user hiện tại với các item. Lấy ra item có điểm dự đoán cao nhất. Công thức dự đoán điểm đánh giá được trình bày rõ ở phần sau (Cách tiếp cận Item-based filtering).



Hình 2.3 Luồng hoạt động của User-based filtering ¹

2.3.2.2. Cách tiếp cận Item-based filtering:

Bước 1: Xây dựng ma trận đánh giá.

Ma trận đánh giá được xây dựng với các cột và các hàng lần lượt là item (Trong đồ án này, nhóm sử dụng item là nhà hàng) và user. Mỗi ô của ma trận là điểm đánh giá của user dành cho nhà hàng.

¹ Kim Falk, "Practical Recommender Systems", Page 187, 2019

Bước 2: Xây dựng ma trận tương đồng giữa các item.

Trước tiên thực hiện chuẩn hóa dữ liệu (normalizing) bằng cách trừ mỗi điểm đánh giá của một nhà hàng cho điểm đánh giá trung bình của nhà hàng đó (được trình bày ở mục 2.2. Công thức Cosine trong tính toán mức độ tương đồng giữa 2 Vector).

Bước 3: Sắp xếp kết quả từ bước 2 giảm dần theo độ tương đồng.

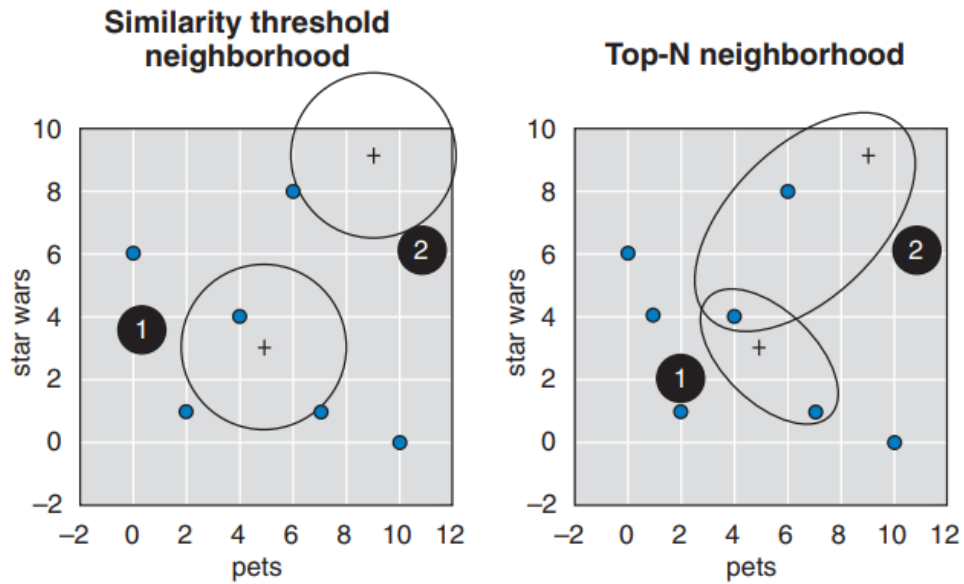
Giả sử một người dùng thích nhà hàng C và đánh giá nhà hàng này với số điểm cao. Khi người dùng này truy cập vào website, sau khi tính ma trận tương đồng, tiến hành sắp xếp các nhà hàng tương đồng với nhà hàng C theo mức độ giảm dần.

Bước 4: Chọn ra những item lân cận (neighbors). Một số cách để chọn ra neighbors:

- Top-N: Đây là cách đơn giản để lấy ra N neighbors mà không cần quan tâm đến độ tương đồng, thậm chí lấy luôn cả những item có độ tương đồng thấp.
- Threshold: Lấy ra những item có độ tương đồng cao hơn một hằng số N cho trước.
- Ưu và nhược điểm giữa hai phương pháp:

	Ưu điểm	Nhược điểm
Top-N	Lấy được nhiều kết quả	Vì không quan tâm đến độ tương tự nên kết quả có thể chứa những item có độ tương tự thấp, dẫn đến chất lượng gợi ý kém hiệu quả.
Threshold	Đảm bảo được kết quả luôn là những item có độ tương đồng cao (tùy theo hằng số N cho trước).	Số lượng neighbors nhỏ.

Bảng 2.2. Ưu và nhược điểm giữa Top-N và Threshold



Hình 2.4 So sánh hai phương pháp Top-N và Threshold ¹

Hình bên trái là phương pháp chọn threshold, xung quanh item hiện tại vẽ vòng tròn bán kính N, những điểm nằm trong vòng tròn chính là neighbors.

Phương pháp Top-N (hình bên phải) không dựa vào khoảng cách, mà nó mở rộng đến khi có đủ N neighbors.

Như vậy có thể nói phương pháp Top-N ưu tiên số lượng hơn chất lượng, Threshold thì ngược lại.

Bước 5: Dự đoán điểm đánh giá. Công thức được tính như sau: ²

$$Pred(u,i) = \bar{r}_u + \frac{\sum_{j \in S_i} (sim(i,j) \times r_{u,j})}{\sum_{j \in S_i} sim(i,j)}$$

¹ Kim Falk, "Practical Recommender Systems", Page 195, 2019

² Kim Falk, "Practical Recommender Systems", Page 198, 2019

Trong đó:

\bar{r}_u : Điểm đánh giá trung bình của user u (user hiện tại).

$r_{u,j}$: Điểm đánh giá của user hiện tại dành cho item j .

S_j : Các item trong vùng lân cận (neighborhood) và đã được đánh giá bởi user hiện tại

$\text{Pred}(u,i)$: Điểm dự đoán của user hiện tại với item i .

$\text{sim}(i,j)$: Độ tương tự giữa hai item i và j .

2.3.3. Nhận xét hai cách tiếp cận User-based filtering và Item-based filtering

Cách tiếp cận User-based filtering có một số hạn chế sau:

- Số lượng users thường lớn hơn số lượng items rất nhiều, dẫn đến kích thước ma trận tương đồng rất lớn .
- Người dùng thường ít sẵn lòng đánh giá, đồng nghĩa với việc ma trận đánh giá có nhiều phần tử mang giá trị rỗng, dẫn đến kết quả gợi ý kém hiệu quả hơn. Đồng thời mỗi khi user thêm hoặc sửa đánh giá, trung bình cộng các rating các rating cũng như vector chuẩn hóa ứng với user này thay đổi nhiều, dẫn đến phải thường xuyên cập nhật ma trận tương đồng, gây tốn nhiều bộ nhớ và thời gian.

Ngược lại, cách tiếp cận Item-based filtering có những lợi ích sau:

- Số lượng item thường nhỏ hơn số lượng user nên kích thước ma trận tương đồng nhỏ hơn nhiều.
- Vì mỗi item thường được đánh giá bởi nhiều user, nên giá trị trung bình của mỗi cột (item) ít bị ảnh hưởng, việc cập nhật ma trận tương đồng cũng có thể ít được cập nhật thường xuyên hơn.

2.4. Thuật toán Tf-idf.¹

Tf-idf (Term Frequency – Inverse Document Frequency)¹ là 1 kỹ thuật sử dụng trong khai phá dữ liệu văn bản. Trọng số này được sử dụng để đánh giá tầm quan trọng của một từ trong một văn bản. Giá trị cao thể hiện độ quan trọng cao và nó phụ thuộc vào số lần từ xuất hiện trong văn bản nhưng bù lại bởi tần suất của từ đó trong tập dữ liệu.

Một vài biến thể của Tf-idf thường được sử dụng trong các hệ thống tìm kiếm như một công cụ chính để đánh giá và sắp xếp văn bản dựa vào truy vấn của người dùng. Tf-idf cũng được sử dụng trong các bài toán như tóm tắt văn bản và phân loại văn bản.

TF: Term Frequency (Tần suất xuất hiện của từ) là số lần từ xuất hiện trong văn bản. Vì các văn bản có thể có độ dài ngắn khác nhau nên một số từ có thể xuất hiện nhiều lần trong một văn bản dài hơn là một văn bản ngắn. Như vậy, term frequency thường được chia cho độ dài văn bản (tổng số từ trong một văn bản).

$$Tf(w, d) = \frac{c(w, d)}{len(d)}$$

Trong đó:

$c(w, d)$: Tần số xuất hiện của từ vựng w trong văn bản d .

$len(d)$: Số từ vựng trong văn bản d .

IDF: Inverse Document Frequency (Nghịch đảo tần suất của văn bản), giúp đánh giá tầm quan trọng của một từ. Khi tính toán TF, tất cả các từ được coi như có độ quan trọng bằng nhau. Nhưng một số từ như “is”, “of” và “that” thường xuất hiện rất nhiều

¹ <http://www.tfidf.com/>

lần nhưng độ quan trọng là không cao. Như thế cần giảm độ quan trọng của những từ này xuống.

$$IDF(w, D) = \log \frac{M}{f(w, D)}$$

Trong đó:

M: Tổng số văn bản trong tập văn bản D.

$f(w, D)$: Tần số của từ w trong tập văn bản D.

Trọng số Tf-idf (w, d, D) sẽ được tính bằng công thức: $Tf(w, d) * IDF(w, D)$.

Ví dụ:

Giả sử có 2 bảng đếm tần suất xuất hiện của các từ trong một kho dữ liệu D (corpus) gồm 2 tài liệu văn bản (document)

Document 1 (d1)		Document 2 (d2)	
Term	Term Count	Term	Term Count
there	1	there	1
is	1	is	1
an	2	a	2
apple	1	table	3

Bảng 2.3. Ví dụ thuật toán tf-idf

tf tính tần suất của từ “there” trong mỗi document:

$$tf("there", d1) = 1/5 = 0.2$$

$$tf("there", d2) = 1/7 = 0.14$$

Cả 2 document đều có chứa từ “there”, do đó ta có idf đánh giá tầm quan trọng của từ “there” ở cả 2 document như sau:

$$idf("there", D) = \log\left(\frac{2}{2}\right) = 0.$$

Ta tính được tf-idf cho từ “there” :

$$Tf-idf("there", d1, D) = 0.2 * 0 = 0$$

$$Tf-idf("there", d2, D) = 0.14 * 0 = 0.$$

Nhận xét: Trọng số Tf-idf của từ “there” trong cả 2 document bằng 0, chứng tỏ “there” không quan trọng và xuất hiện trong tất cả các document trong kho dữ liệu.

Tương tự ta tính được trọng số Tf-idf cho từ “table” như sau:

$$Tf-idf("table", d1, D) = 0/5 * \log\left(\frac{2}{1}\right) = 0.$$

$$Tf-idf("table", d2, D) = 3/7 * \log\left(\frac{2}{1}\right) = 0.13.$$

Nhận xét: Từ “table” có mức độ quan trọng cao hơn trong document 2.

2.5. Công thức Haversine:

Công thức Haversine được sử dụng để tính khoảng cách giữa hai điểm Point 1 và Point 2 trên bề mặt trái đất được xác định bởi kinh độ và vĩ độ. ¹

¹ Dr. P. V. Ingole, Mr. Mangesh K Nichat. “Landmark based shortest path detection by using Dijkstra Algorithm and Haversine Formula”, Vol.3, Issue 3, Page 163, 5-2013.

$$d = 2.r.\sin^{-1}\left(\sqrt{\sin^2\left(\frac{\varphi_2 - \varphi_1}{2}\right) + \cos(\varphi_1)\cos(\varphi_2)\sin^2\left(\frac{\delta_2 - \delta_1}{2}\right)}\right)$$

d : Khoảng cách giữa 2 điểm

φ_1, φ_2 : Lần lượt là vĩ độ của 2 điểm Point 1 và Point 2.

δ_1, δ_2 : Lần lượt là kinh độ của 2 điểm Point 1 và Point 2.

r : Bán kính của trái đất. $r = 6371000$.

2.6. Các công cụ hỗ trợ hệ thống.

2.6.1. Python

2.6.1.1. Giới thiệu¹

- Python là ngôn ngữ lập trình hướng đối tượng thông dụng dùng để viết các tiện ích hệ thống. Nó cũng được sử dụng như ngôn ngữ kết dính đóng vai trò tích hợp C và C++.
- Được tạo ra bởi Guido van Rossum tại Amsterdam năm 1990, Python hoàn toàn tạo kiểu động và dùng cơ chế cấp phát bộ nhớ tự động. Python được phát triển trong một dự án mã nguồn mở, do tổ chức phi lợi nhuận Python Software Foundation quản lý.
- Python là ngôn ngữ có hình thức khá đơn giản và rõ ràng, do đó tạo nên sự dễ dàng tiếp cận cho những lập trình viên mới bắt đầu.

2.6.1.2. Đặc điểm²

- **Interpreted:** nhờ chức năng thông dịch mà trình thông dịch (Interpreter) của Python có thể xử lý lệnh tại thời điểm chạy chương trình (runtime). Nhờ đó mà không cần biên dịch chương trình trước khi thực hiện nó (tương tự như Perl và Php).

¹ Tham khảo tại: <https://www.stdio.vn/python/tong-quan-ve-python-Gyjc>

² Tham khảo tại: <https://www.stdio.vn/python/tong-quan-ve-python-Gyjc>

- **Interactive:** tính năng tương tác của Python giúp tương tác trực tiếp với trình thông dịch của nó ngay tại dấu nhắc lệnh. Cụ thể: Có thể thực hiện lệnh một cách trực tiếp tại dấu nhắc của Python.
- **Object-Oriented:** Python hỗ trợ mạnh cho phong cách lập trình hướng đối tượng và kỹ thuật lập trình gói mã trong đối tượng.
- **Beginner's Language:** Python được xem là ngôn ngữ lập trình dành cho những ai mới làm quen với việc lập trình trên máy tính, nhưng nó hỗ trợ mạnh cho việc phát triển nhiều loại ứng dụng khác nhau, từ các chương trình xử lý văn bản đơn giản đến các ứng dụng web, đến các chương trình game, ...

2.6.1.3. Ưu điểm

- Là một ngôn ngữ có hình thức sáng sủa, cấu trúc rõ ràng, cú pháp ngắn gọn
- Có trên tất cả các nền tảng hệ điều hành từ UNIX, MS – DOS, Mac OS, Windows và Linux và các OS khác thuộc họ Unix.
- Tương thích mạnh mẽ với Unix, hardware, third-party software với số lượng thư viện khổng lồ (400 triệu người sử dụng)
- Python với tốc độ xử lý cực nhanh, python có thể tạo ra những chương trình từ những script siêu nhỏ tới những phần mềm cực lớn như Blender 3D.

2.6.1.4. Nhược điểm

- Python không có các thuộc tính như: protected, private hay public, không có vòng lặp do...while và switch....case.
- Mặc dù tốc độ xử lý của Python nhanh hơn PHP nhưng không bằng JAVA và C++.

2.6.2. Php

2.6.2.1. Giới thiệu¹

- Php: Hypertext Preprocessor, thường được viết tắt thành Php là một ngôn ngữ lập trình kịch bản hay một loại mã lệnh chủ yếu được dùng để phát triển các ứng dụng viết cho máy chủ, mã nguồn mở, dùng cho mục đích tổng quát. Nó rất thích hợp với web và có thể dễ dàng nhúng vào trang HTML.
- Do được tối ưu hóa cho các ứng dụng web, tốc độ nhanh, nhỏ gọn, cú pháp giống C và Java, dễ học và thời gian xây dựng sản phẩm tương đối ngắn hơn so với các ngôn ngữ khác nên Php đã nhanh chóng trở thành một ngôn ngữ lập trình web phổ biến nhất thế giới.

2.6.2.2. Đặc điểm²

- Php rất đơn giản và dễ sử dụng.
- Php có thể kết nối cơ sở dữ liệu hiệu quả và giúp quá trình tải ứng dụng nhanh hơn.
- Php sẽ bảo mật hơn nếu sử dụng các framework.
- Nó quen thuộc hơn đối với các nhà phát triển và hỗ trợ trực tuyến đang được cung cấp cho người mới bắt đầu.
- Php có khả năng tích hợp linh hoạt với các ngôn ngữ lập trình khác.
- Php là nguồn mở và miễn phí.
- Php rất nhẹ và có nhiều Framework có sẵn để lập trình website.
- Php giúp cho quá trình phát triển website trở nên nhanh hơn.

¹ Tham khảo tại: <https://vi.wikipedia.org/wiki/PHP>

² Tham khảo tại: <https://niithanoi.edu.vn/php-la-gi.html>

2.6.2.3. Ưu điểm¹

- Đầu tiên đó là Php được sử dụng miễn phí. Là một lợi thế cực lớn cho ai muốn học lập trình này.
- Cú pháp và cấu trúc của Php tương đối dễ dàng. Nếu muốn học về ngôn ngữ này thì không phải mất quá nhiều thời gian để hiểu được. Đây là một ưu điểm lớn cho mọi người quan tâm và yêu thích hàng đầu về ngôn ngữ lập trình.
- Sẽ rất dễ kiếm được việc làm với mức lương khủng. Đa số công ty thiết kế web hiện nay đều có bộ phận Php. Và đó là điều không thể thiếu. Ngoài ra còn có thêm các công ty về phần mềm. Nhiều công việc lớn khác nhau, cũng như nhiều công ty khác nhau sẽ mang lại thu nhập đáng kể.
- Không chỉ với hiện tại mà trong tương lai, ngôn ngữ lập trình này sẽ còn phát triển mạnh mẽ hơn nữa để khẳng định vị trí của mình luôn là cao nhất trong các bảng khảo sát ngôn ngữ lập trình.

2.6.2.4. Nhược điểm

- Php còn hạn chế về cấu trúc của ngữ pháp. Nó không được thiết kế gọn gàng và không được đẹp mắt như những ngôn ngữ lập trình khác.
- Php chỉ có thể hoạt động và sử dụng được trên các ứng dụng trong web. Đó chính là lý do khiến cho ngôn ngữ này khó có thể cạnh tranh được với những ngôn ngữ lập trình khác nếu như muốn phát triển và nhân rộng hơn nữa trong lập trình.

¹ Tham khảo tại: <http://free-php-editor.com/php-la-gi/>

CHƯƠNG 3: PHÂN TÍCH VÀ THIẾT KẾ HỆ THỐNG

3.1. Phân tích hiện trạng nhu cầu

3.1.1. Yêu cầu chức năng

- Người dùng:
 - Xem thông tin nhà hàng/quán ăn.
 - Tạo nhà hàng/quán ăn.
 - Đánh giá và bình luận nhà hàng/quán ăn.
 - Điều khiển bằng giọng nói.
 - Chia sẻ nhà hàng/quán ăn..
- Hệ thống
 - Đề xuất địa điểm dựa trên lịch sử đánh giá của người dùng.
 - Xử lý điều khiển bằng giọng nói.
 - Gợi ý nhà hàng lên google map.
- Người quản lý:
 - Quản lý nhà hàng/quán ăn.
 - Quản lý người dùng.
 - Tạo biểu mẫu thống kê.

3.1.2. Yêu cầu phi chức năng:

- Giao diện thân thiện.
- Ứng dụng tiện ích.
- Tốc độ xử lý nhanh.

3.2. Phân tích và thiết kế thành phần dữ liệu

Bảng mô tả các thuộc tính.

Tên bảng	Tên thuộc tính	Kiểu dữ liệu	Diễn giải
customers	Id	Int	Id của khách hàng
	Avatar	Varchar (255)	Đường dẫn avatar
	DisplayName	Text	Tên hiển thị
	Status	Tinyint	Trạng thái
	email	Varchar(100)	email
	password	Varchar(255)	Mật khẩu
	Google_Id	Varchar(255)	ID tài khoản google
	CreateAt	Int	Thời gian tạo
	UpdateAt	Int	Thời gian cập nhật
	SavedRes	Text	Lưu id các nhà hàng đã lưu
	CreateDate	DateTime	Thời gian tạo
restaurants	Id	Int	Id nhà hàng
	Address	Varchar(255)	Địa chỉ nhà hàng
	AvgRating	Int	Điểm đánh giá trung bình
	Description	Varchar(255)	Mô tả

	Latitude	Float	Vĩ độ
	Longitude	Float	Kinh độ
	Name	Varchar(255)	Tên nhà hàng
	ResCreatedOn	Varchar(255)	Thời gian tạo
	PriceMax	Int	Giá trị lớn nhất
	PriceMin	Int	Giá trị nhỏ nhất
	Status	Int	Trạng thái
	TotalReviews	Int	Tổng số bình luận
	PositionRating	Float	Điểm đánh giá vị trí
	PriceRating	Float	Điểm đánh giá giá cả
	QualityRating	Float	Điểm đánh giá chất lượng
	ServiceRating	Float	Điểm đánh giá dịch vụ
	SpaceRating	Float	Điểm đánh giá không gian
	IsOpening	Tinyint	Có đang mở cửa không
	CreateDate	DateTime	Thời gian tạo
categories	Id	Int	Id category
	Name	Varchar(255)	Tên

	Status	Int	Trạng thái
	Parent_id	Int	Id category cha
comment_likes	Id	Int	Id
	IdOwner	Int	Id người dùng
	IdComment	Int	Id bình luận
	Status	Int	Trạng thái
	CreatedAt	Int	Thời gian tạo
	UpdatedAt	Int	Thời gian cập nhật
comment_pictures	Id	Int	Id hình
	Url	Varchar(255)	Đường dẫn hình
	CommentId	Int	Id bình luận
comments	Id	Int	Id bình luận
	AvgRating	Float	Điểm đánh giá trung bình
	Description	Varchar(255)	Mô tả
	Owner_id	Int	Id người dùng
	ResId	Int	Id nhà hàng
	TotalLike	Int	Tổng số like
	CreatedAt	Int	Thời gian tạo

	UpdatedAt	Int	Thời gian cập nhật
	PositionRating	Int	Điểm đánh giá vị trí
	PriceRating	Int	Điểm đánh giá giá cả
	QualityRating	Int	Điểm đánh giá chất lượng
	ServiceRating	Int	Điểm đánh giá dịch vụ
	SpaceRating	Int	Điểm đánh giá không gian
search_train	id	Int	ID
	action	Varchar(255)	Hành động
	input	Varchar(255)	Văn bản
similarity	Id	Int	ID
	Created	Date	Ngày tạo
	Source	Int	Id nhà hàng
	Target	Int	Id nhà hàng
	similarity	decimal	Độ tương đồng

Bảng 3.1. Mô tả các thuộc tính

3.3. Phân tích thiết kế thành phần xử lý

3.3.1. Sơ đồ Usecase



Hình 3.1. Sơ đồ usecase

3.3.2. Danh sách Usecase

STT	Usecase	Ý nghĩa
1	Usecase Quản lý tài khoản admin(Admin)	Admin quản lý tài khoản của mình và các admin
2	Usecase Thêm tài khoản admin(Admin)	Admin thêm các tài khoản có quyền admin
3	Usecase Sửa tài khoản admin(Admin)	Admin xem/sửa thông tin các tài khoản admin
4	Usecase Xóa tài khoản admin(Admin)	Admin xóa các tài khoản admin
5	Usecase Quản lý tài khoản người dùng(Admin)	Admin quản lý tài khoản của người dùng
6	Usecase Thêm tài khoản (Admin)	Admin thêm tài khoản người dùng
7	Usecase Sửa tài khoản(Admin)	Admin xem/sửa thông tin tài khoản người dùng
8	Usecase Xóa tài khoản(Admin)	Admin xóa tài khoản người dùng
9	Usecase Tạo nhà hàng(Admin)	Admin tạo nhà hàng/quán ăn mới

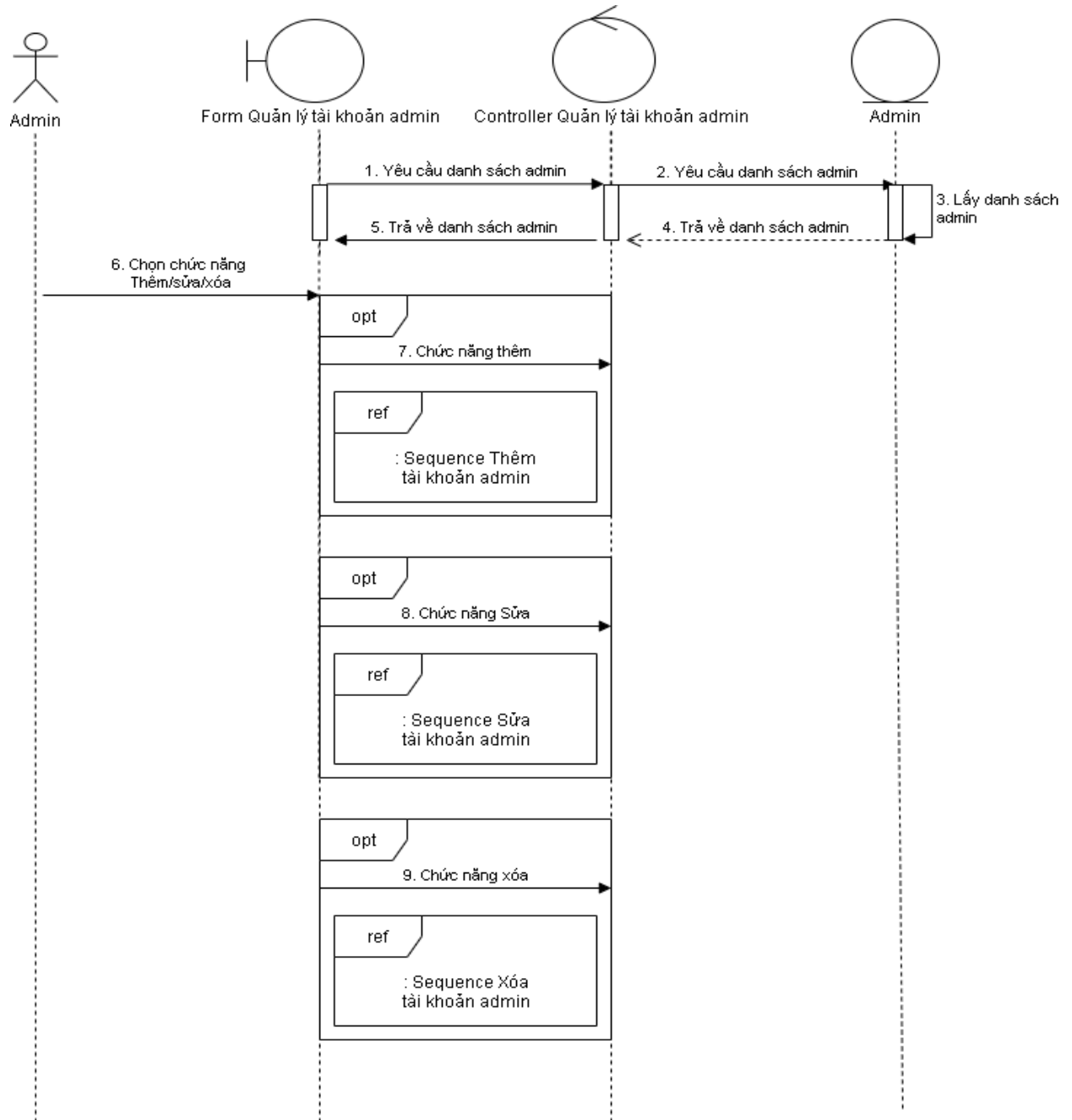
10	Usecase Duyệt nhà hàng(Admin)	Admin duyệt các nhà hàng do người dùng tạo
11	Usecase Quản lý nhà hàng(Admin)	Admin quản lý việc xem/xóa/sửa nhà hàng/quán ăn
12	Usecase Xem/Sửa nhà hàng(Admin)	Admin xem/sửa thông tin nhà hàng/quán ăn
13	Usecase Xóa nhà hàng(Admin)	Admin xóa nhà hàng/quán ăn
14	Usecase Thống kê(Admin)	Admin thống kê thông tin hệ thống
15	Usecase Đăng ký(Người dùng)	Người dùng đăng ký tài khoản
16	Usecase Tìm mật khẩu(Người dùng)	Người dùng tìm mật khẩu của mình
17	Usecase Đổi mật khẩu(Người dùng)	Người dùng đổi mật khẩu của mình
18	Usecase Đăng nhập(Người dùng)	Người dùng đăng nhập vào website
19	Usecase Quản lý thông tin cá nhân(Người dùng)	Người dùng quản lý thông tin cá nhân
20	Usecase Điều khiển bằng giọng nói(Người dùng)	Người dùng điều khiển website bằng giọng nói

21	Usecase Tìm kiếm nhà hàng(Người dùng)	Người dùng tìm kiếm nhà hàng/quán ăn
22	Usecase Xem chi tiết nhà hàng(Người dùng)	Người dùng xem chi tiết nhà hàng/quán ăn
23	Usecase Đánh giá nhà hàng(Người dùng)	Người dùng đánh giá nhà hàng/quán ăn đang xem
24	Usecase Tạo nhà hàng(Người dùng)	Người dùng tạo nhà hàng/quán ăn đề xuất cho website

Bảng 3.2. Danh sách các usecase

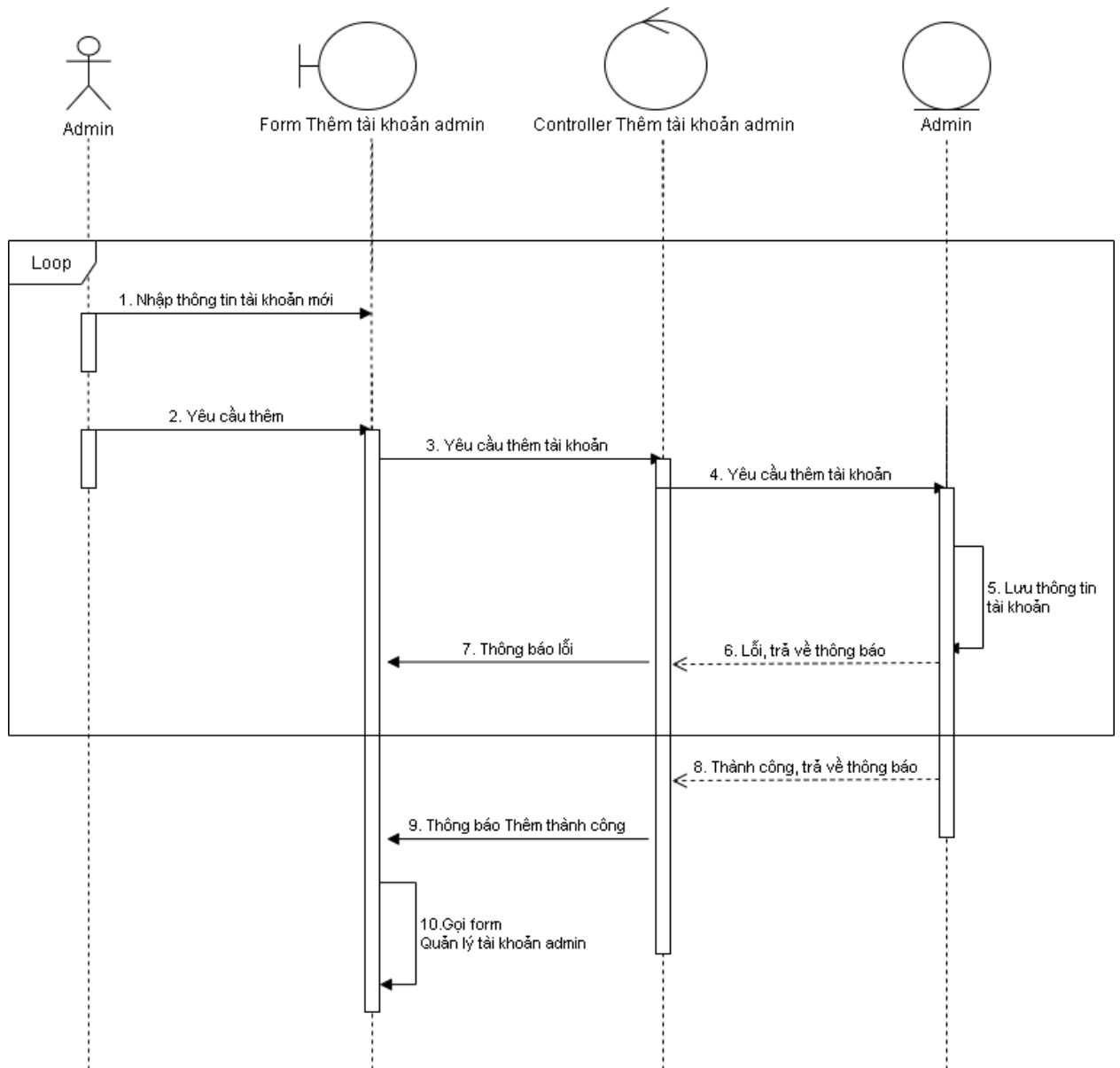
3.3.3. Sơ đồ tuần tự

- Usecase Quản lý tài khoản admin:



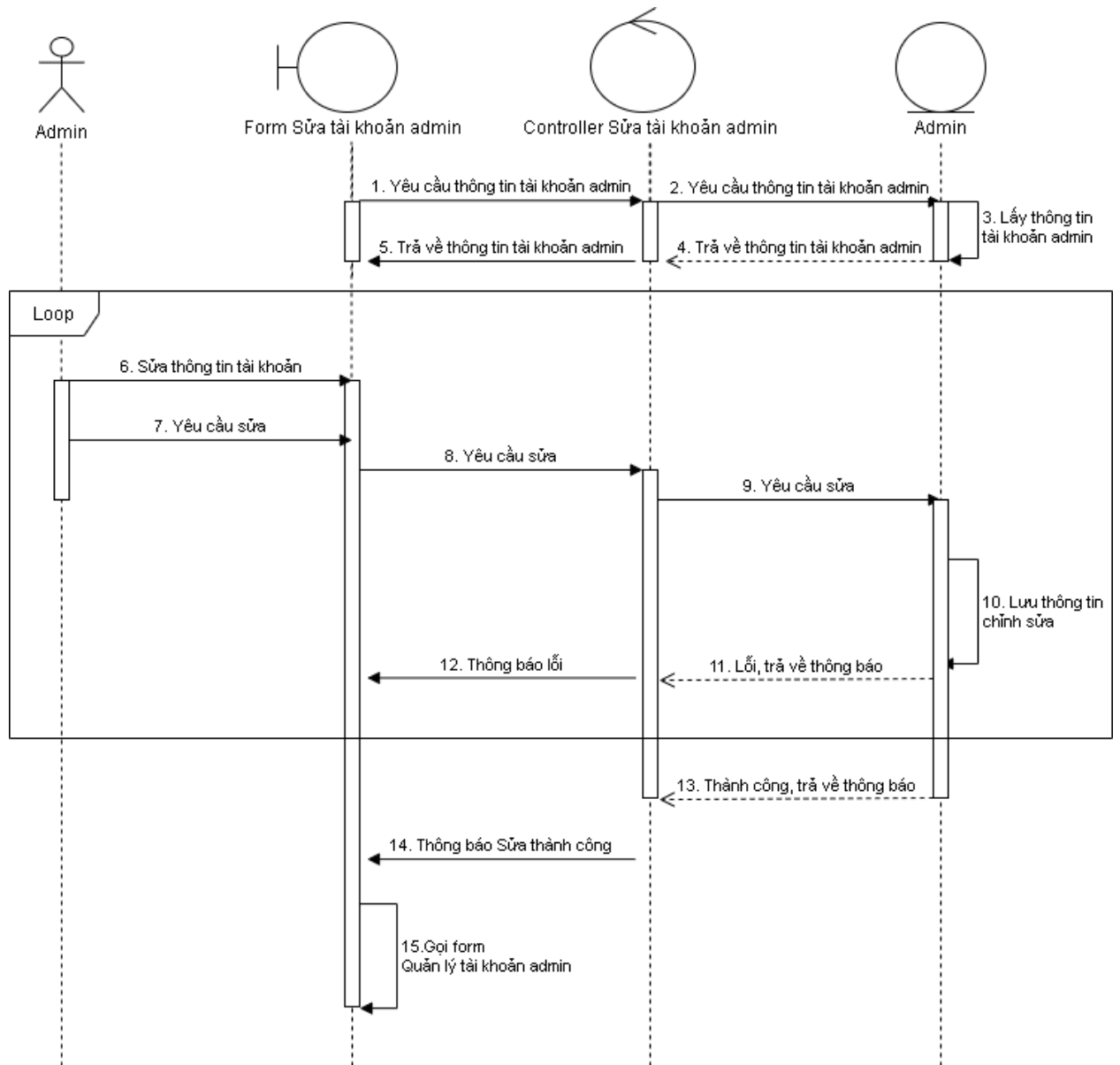
Hình 3.2. Sơ đồ tuần tự quản lý admin

- Usecase Thêm tài khoản admin



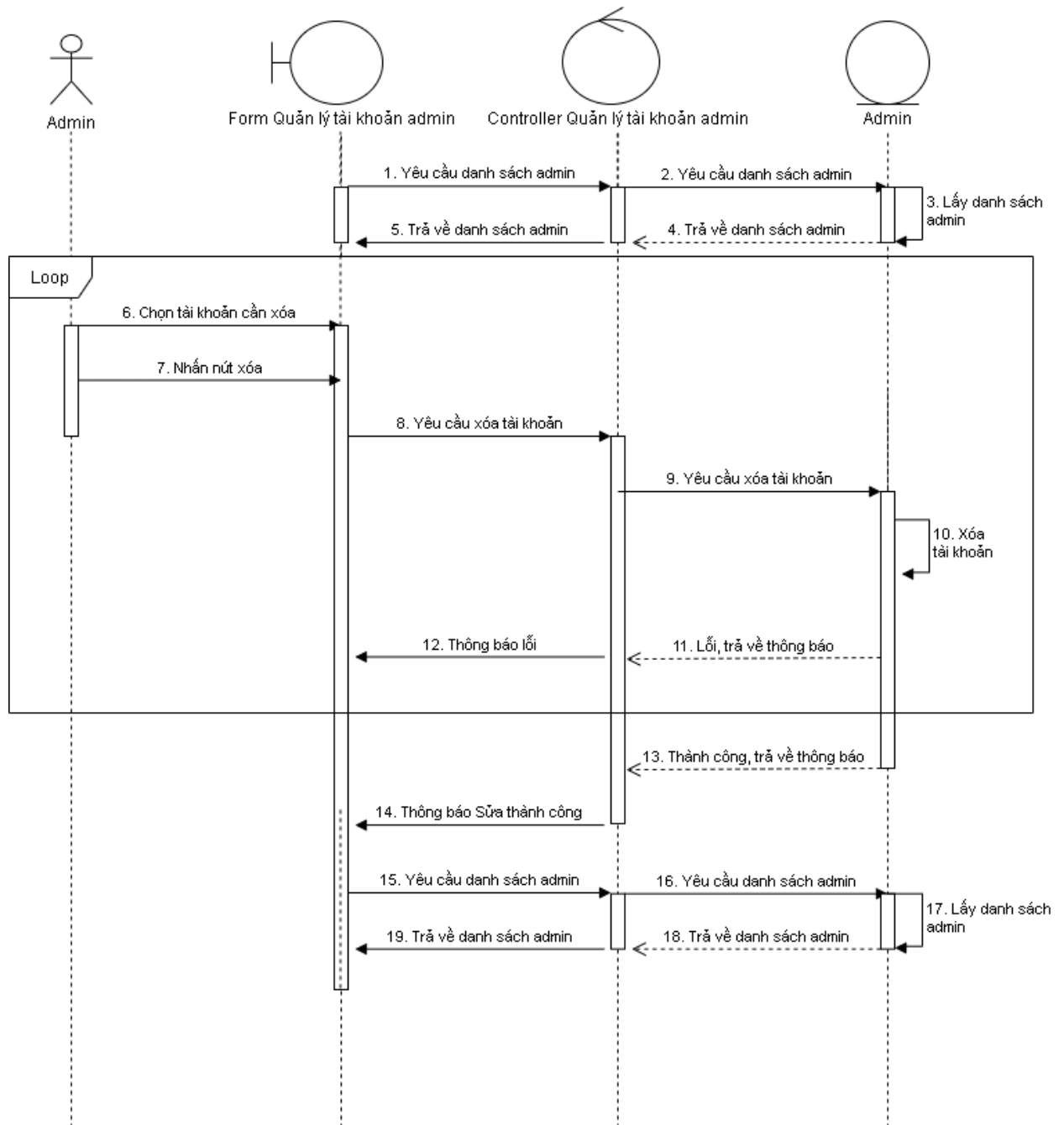
Hình 3.3. Thêm tài khoản

- Usecase Sửa tài khoản admin



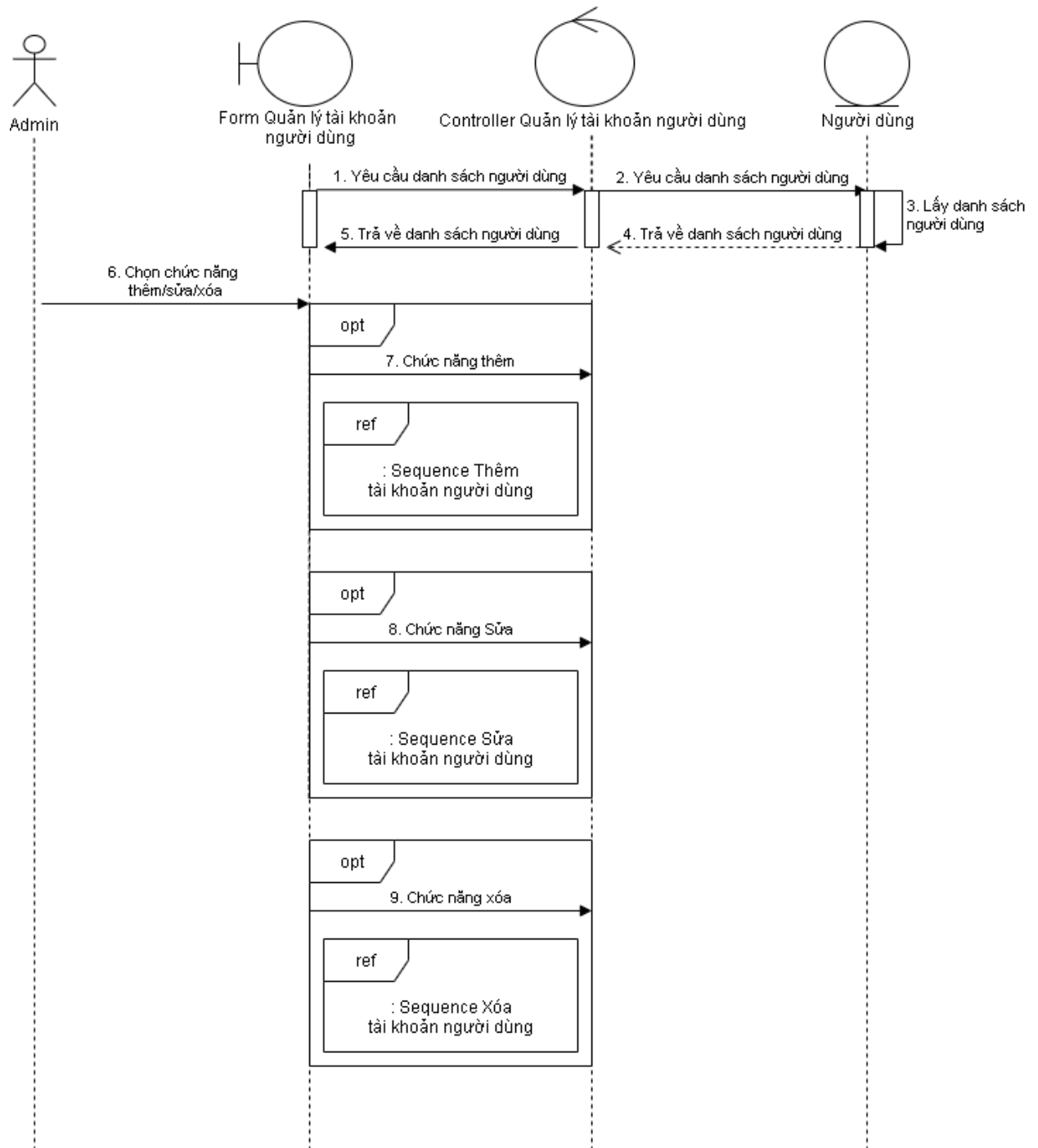
Hình 3.4. Sửa tài khoản

- Usecase Xóa tài khoản admin



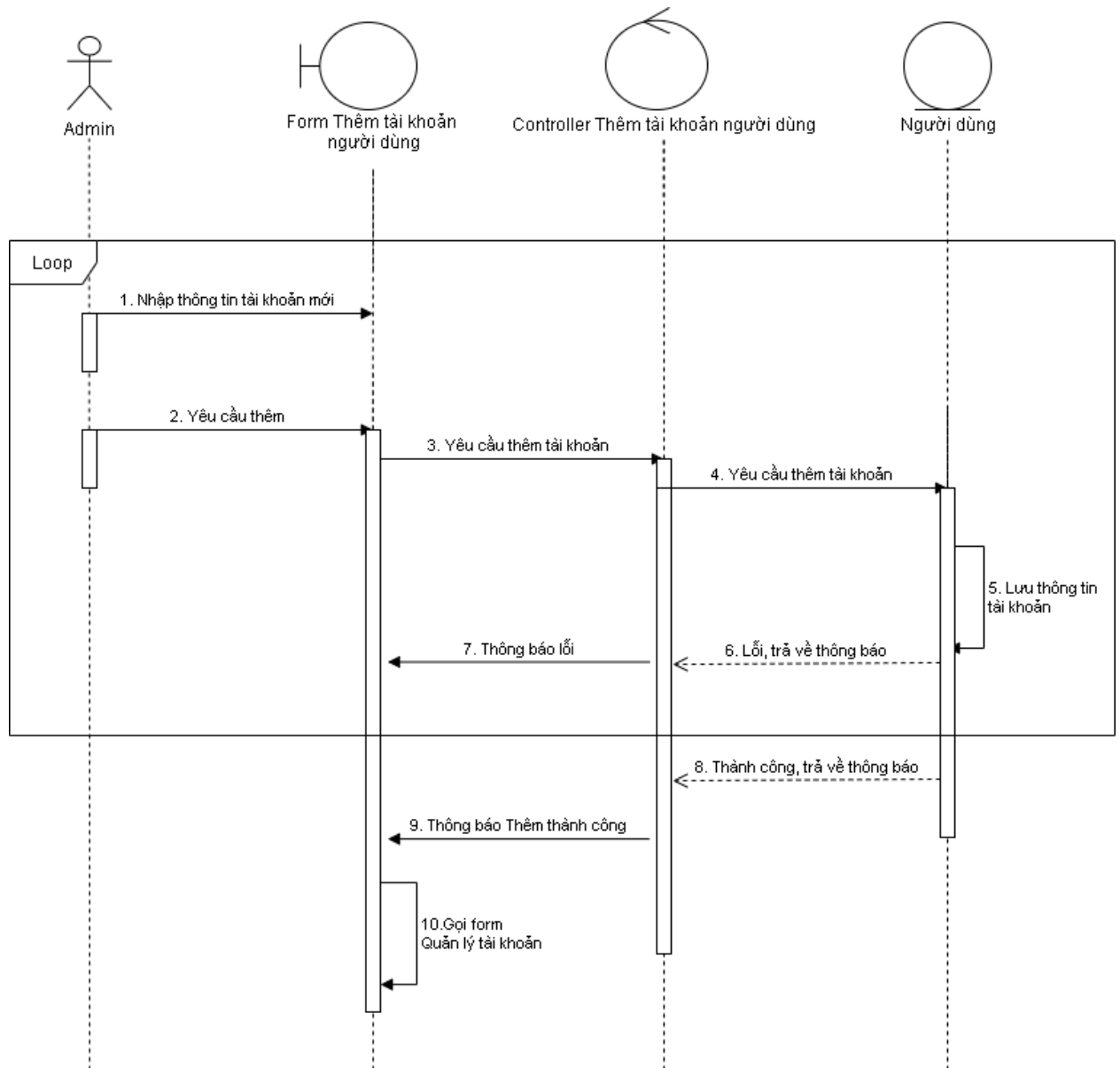
Hình 3.5. Xóa tài khoản

- Usecase Quản lý tài khoản



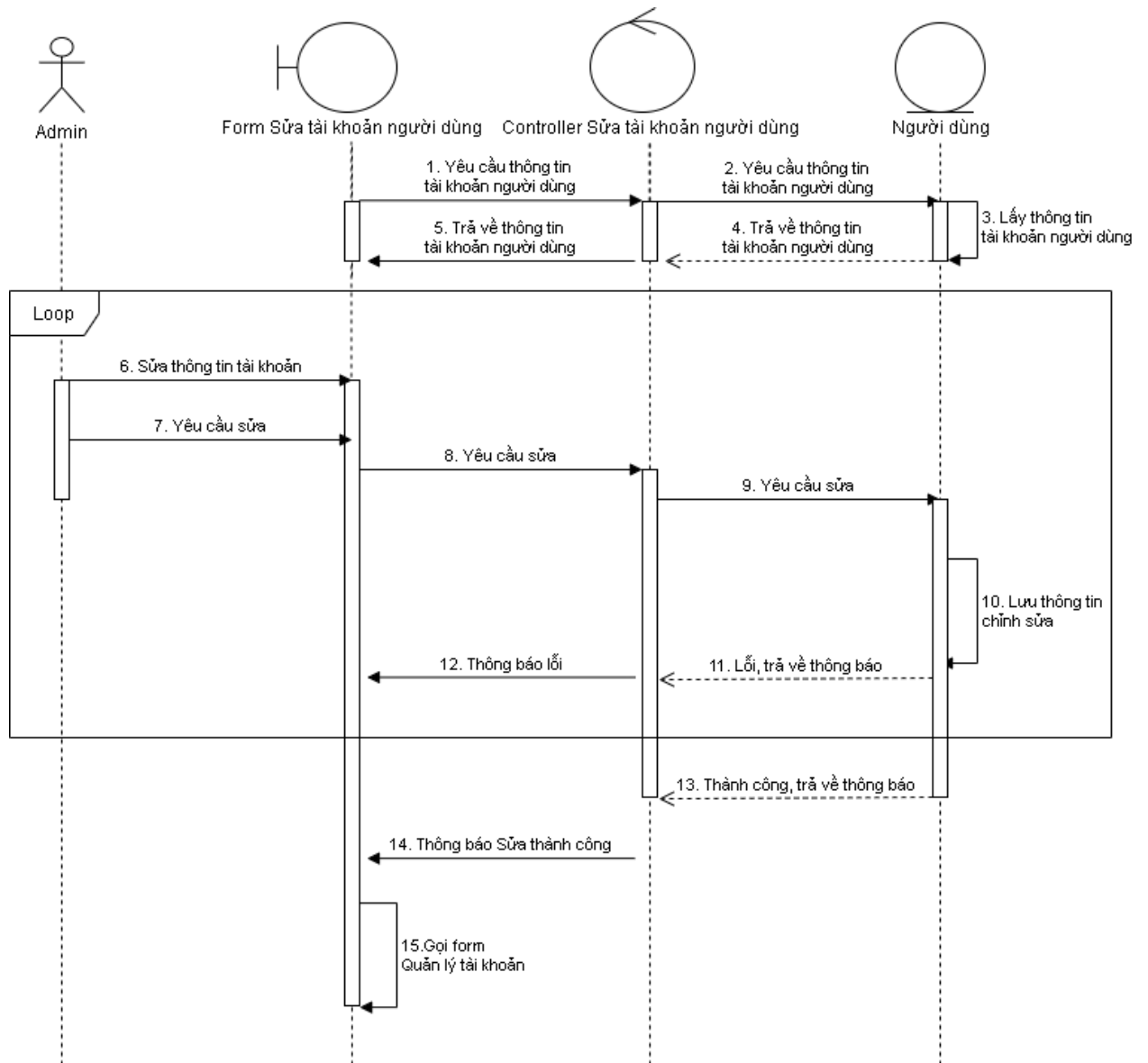
Hình 3.6. Quản lý tài khoản

- Usecase Thêm tài khoản



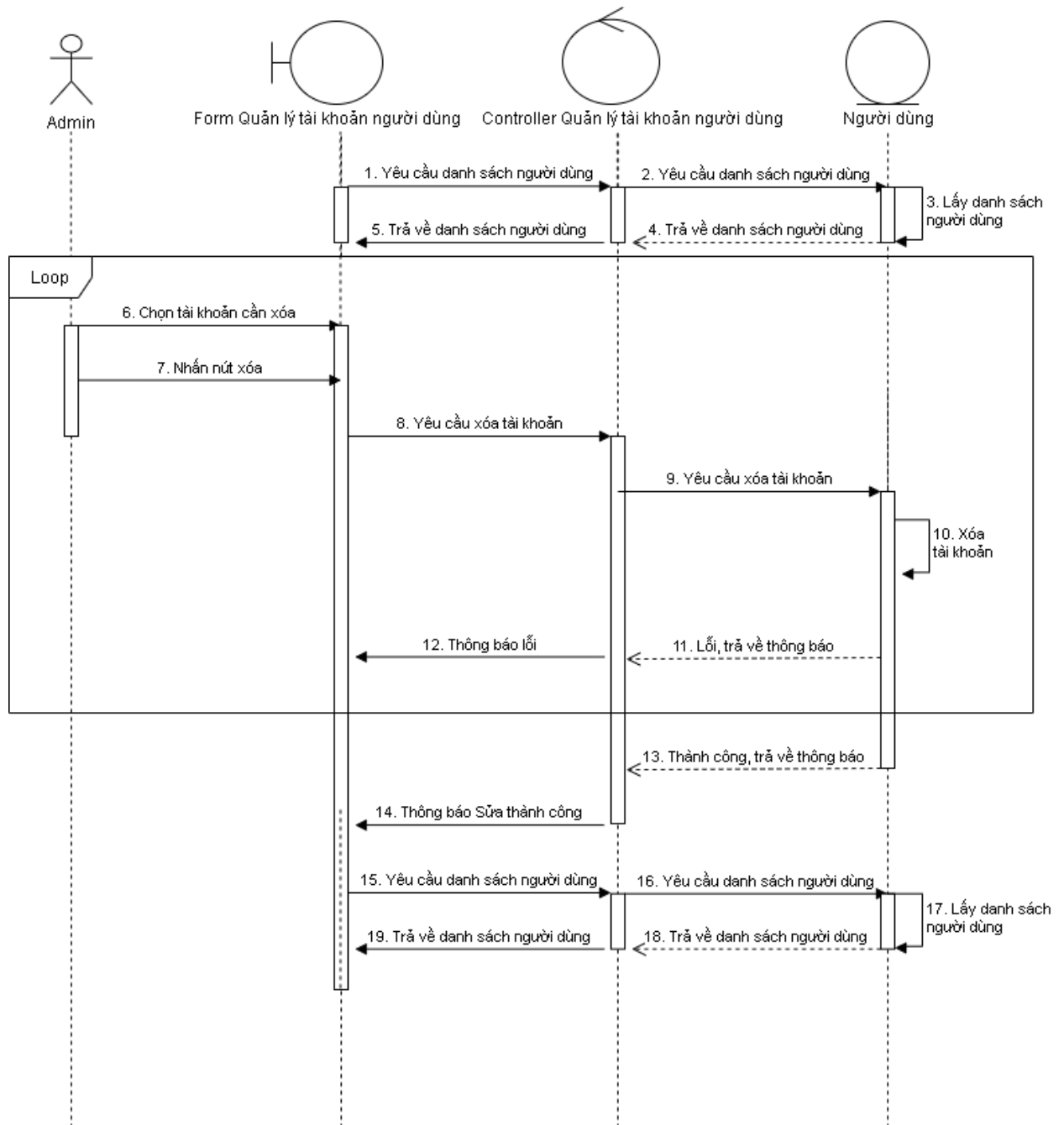
Hình 3.7. Thêm tài khoản

- Usecase Sửa tài khoản



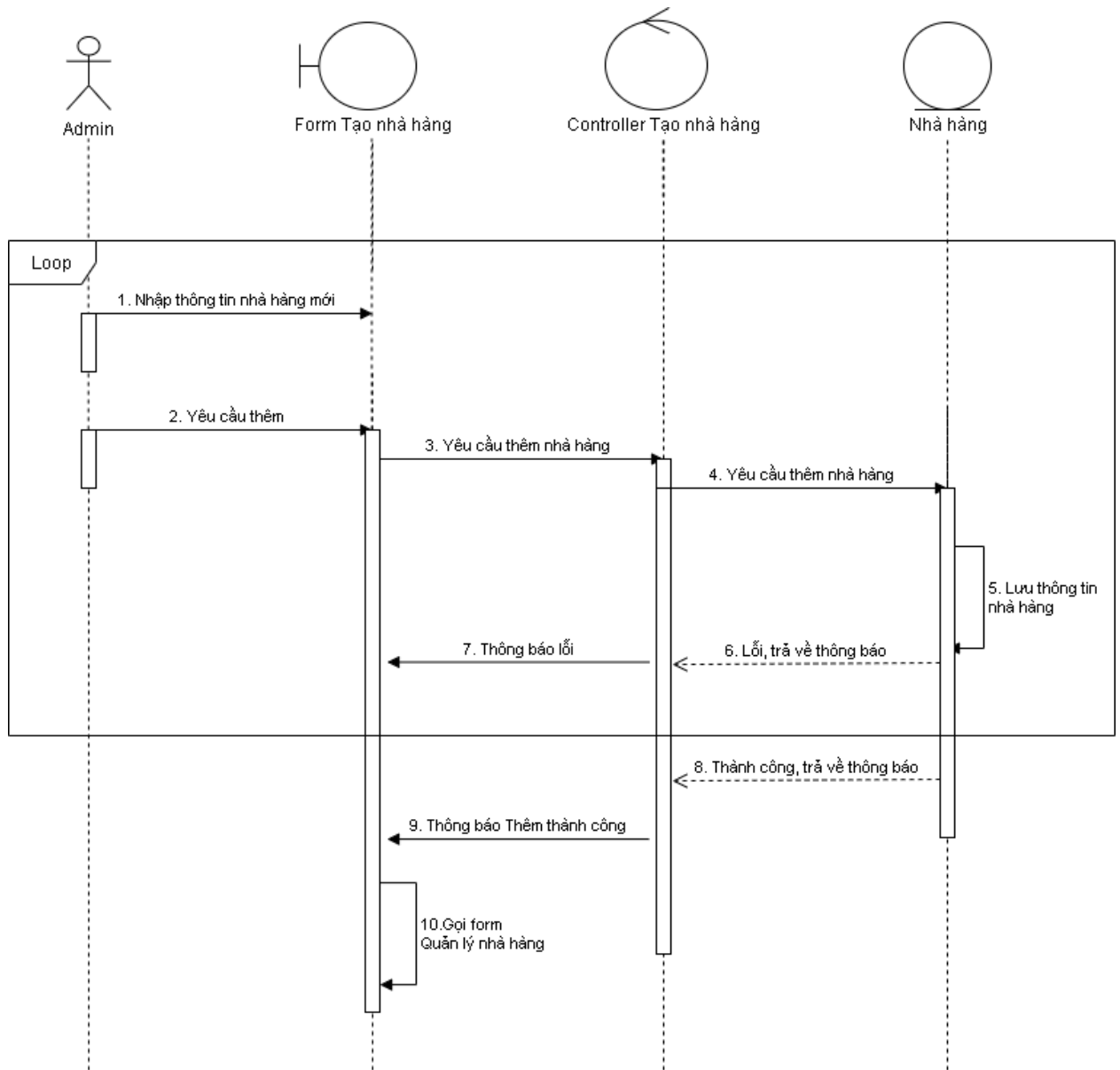
Hình 3.8. Sửa tài khoản

- Usecase Xóa tài khoản



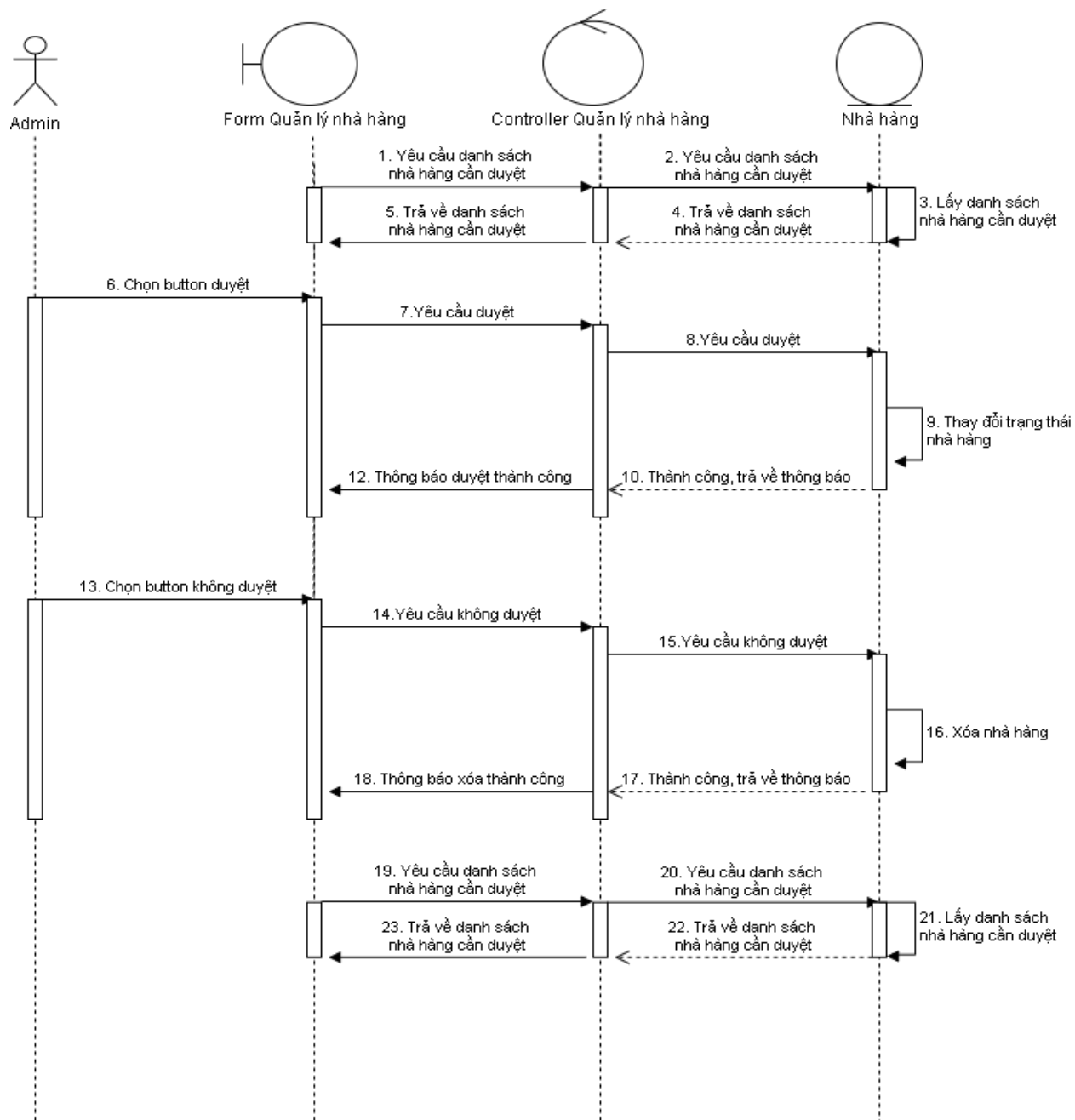
Hình 3.9. Xóa tài khoản

- Usecase Tạo nhà hàng



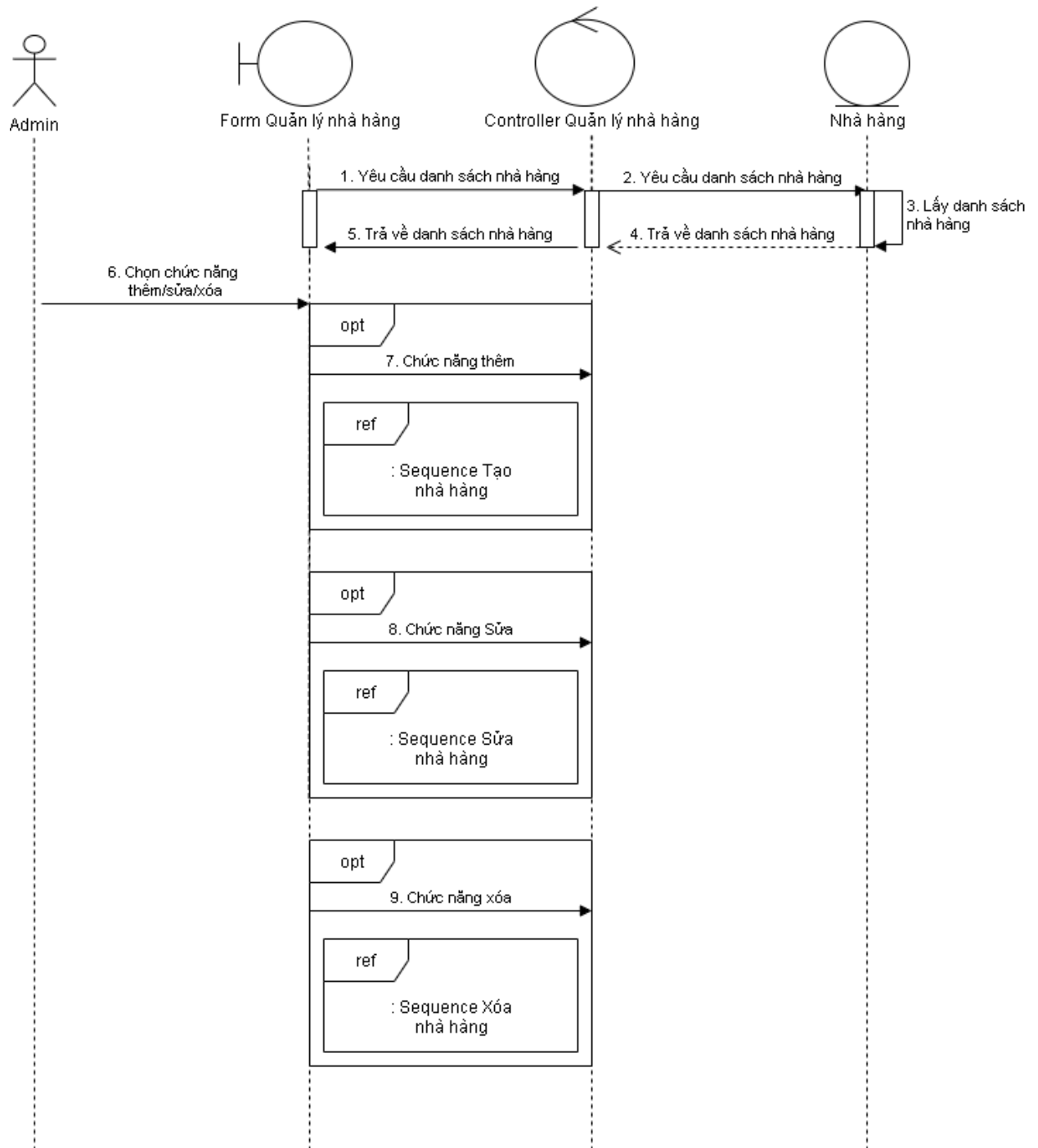
Hình 3.10. Tạo nhà hàng

- Usecase Duyệt nhà hàng



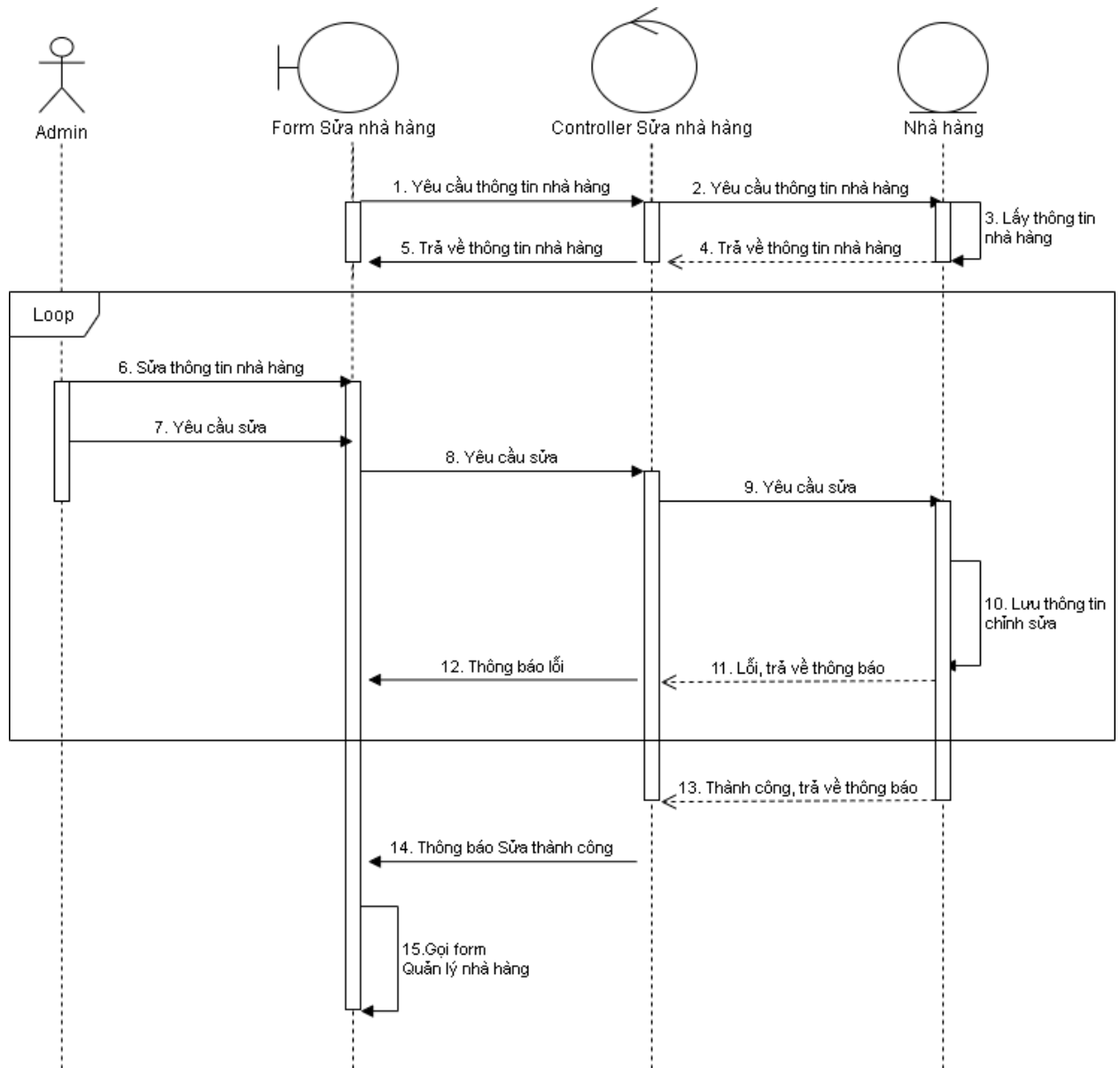
Hình 3.11. Duyệt nhà hàng

- Usecase Quản lý nhà hàng



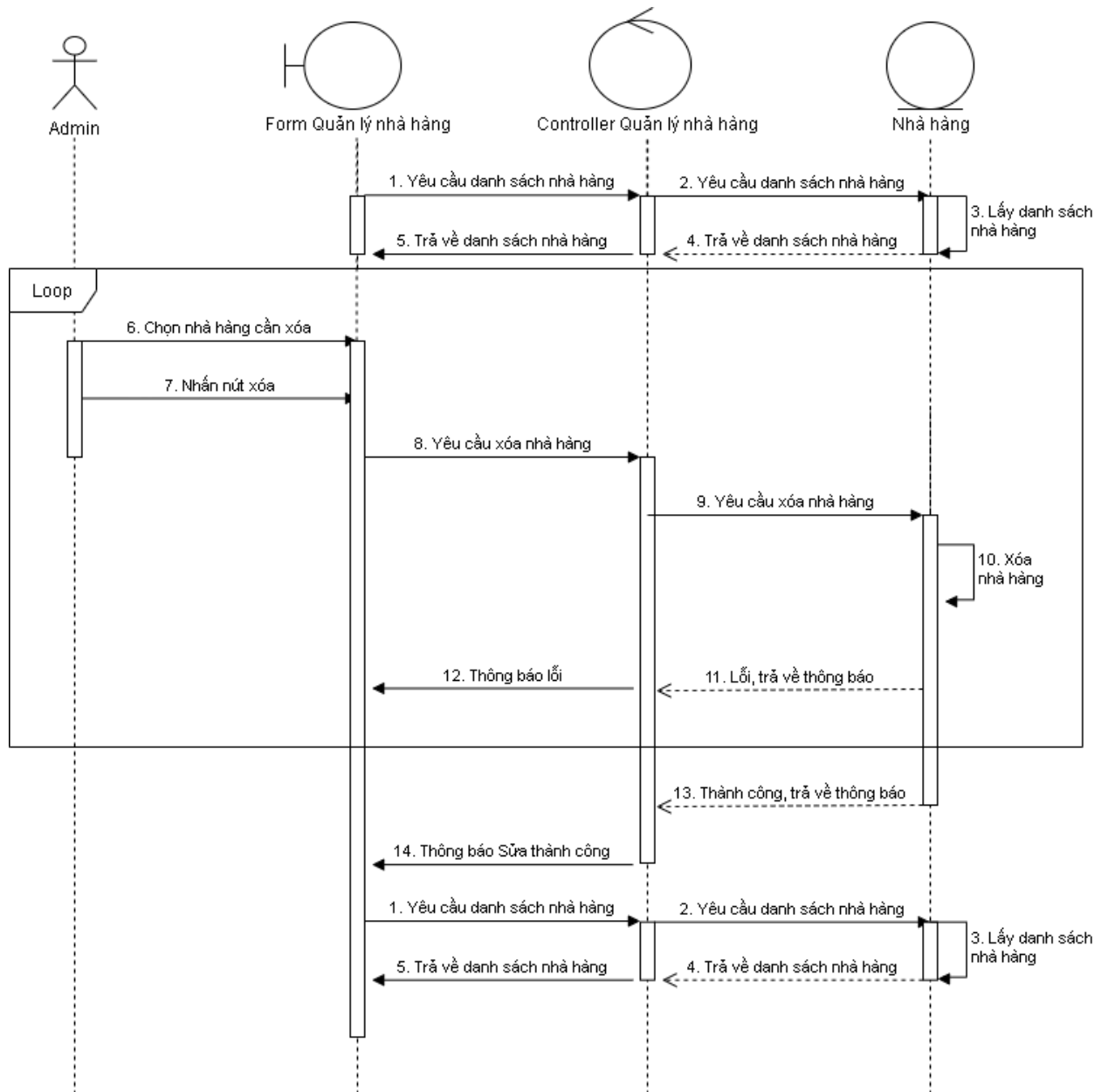
Hình 3.12. Quản lý nhà hàng

- Usecase Xem/sửa nhà hàng



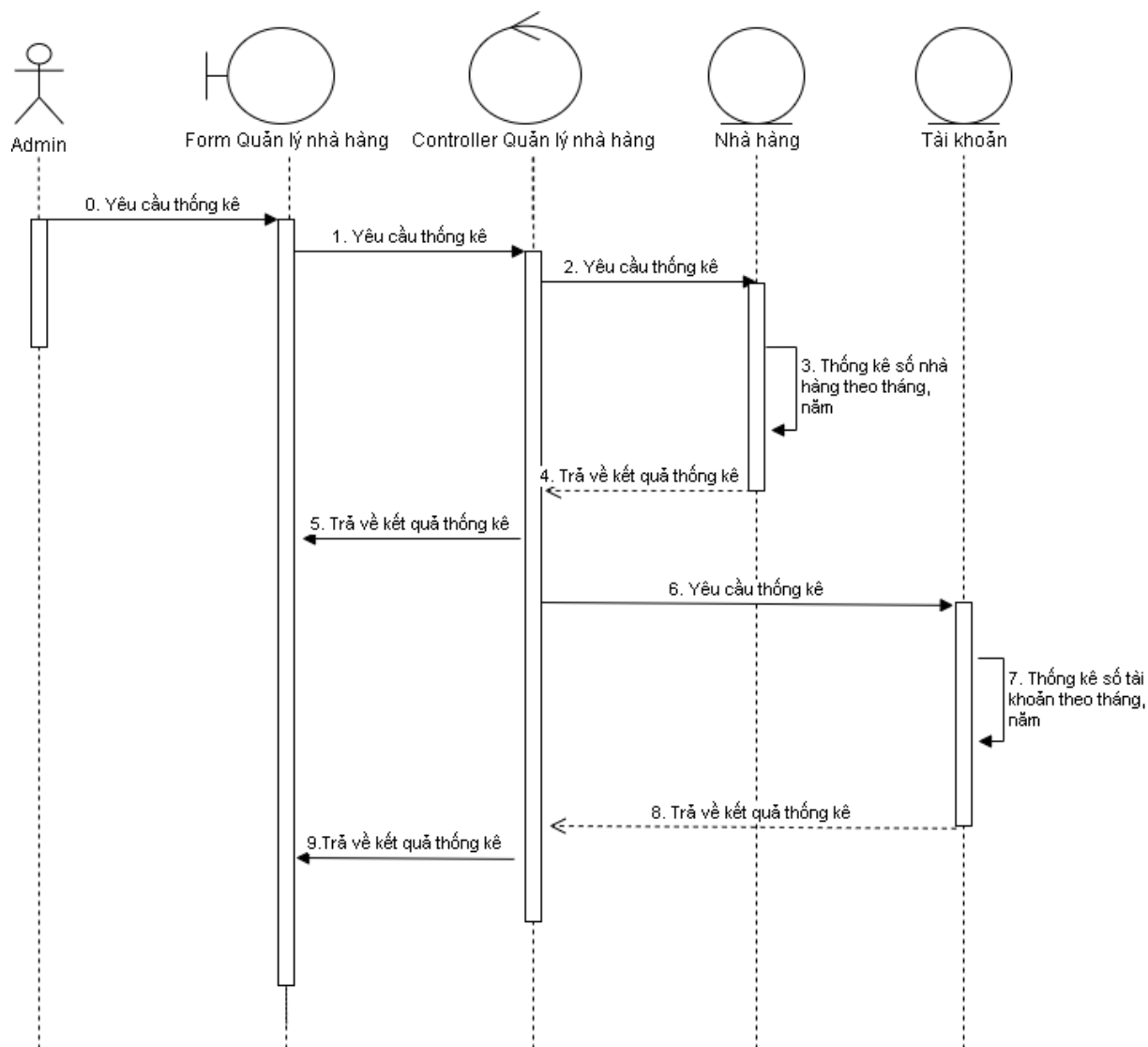
Hình 3.13 Sửa nhà hàng

- Usecase Xóa nhà hàng



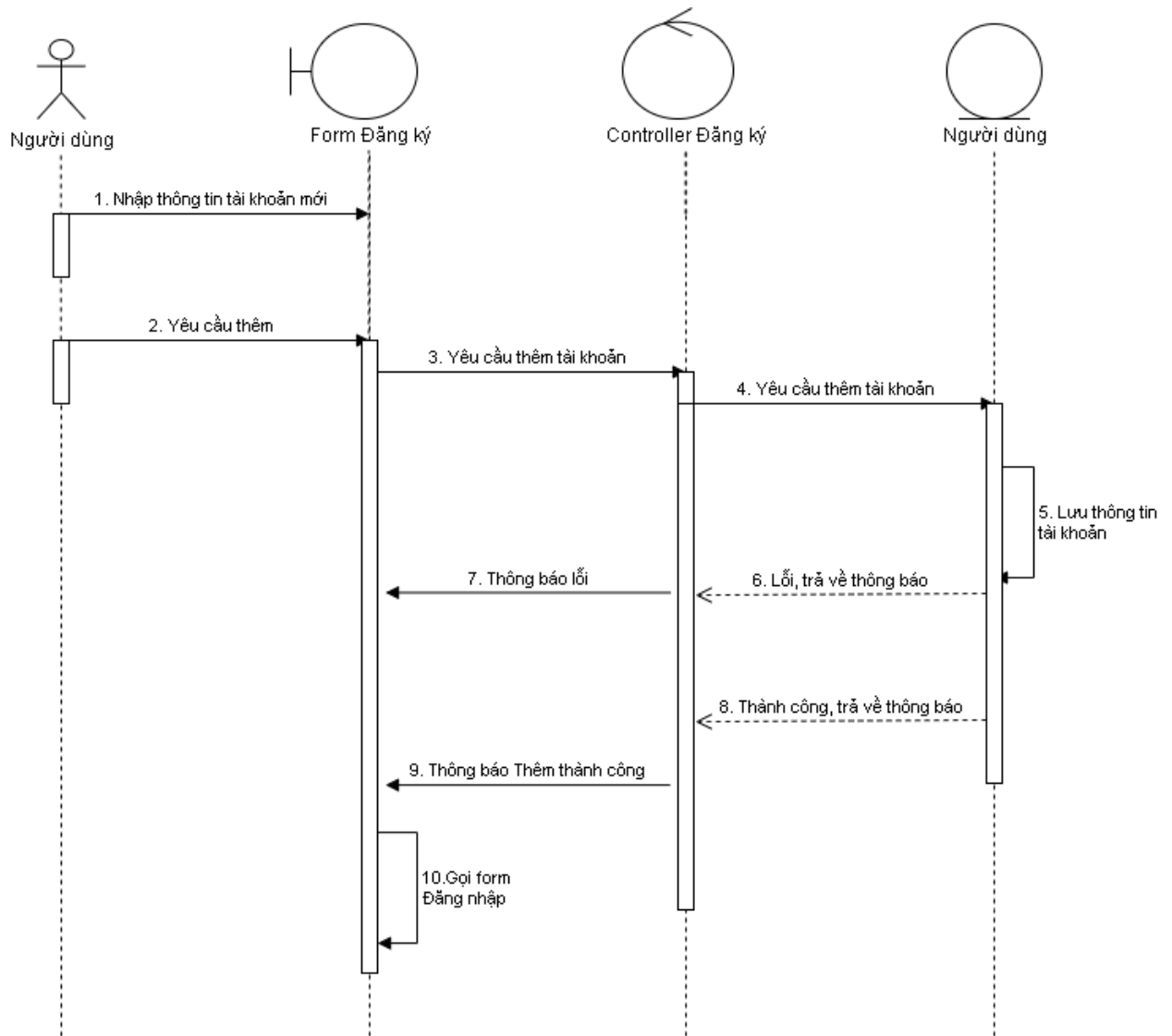
Hình 3.14. Xóa nhà hàng

- Usecase Thống kê nhà hàng



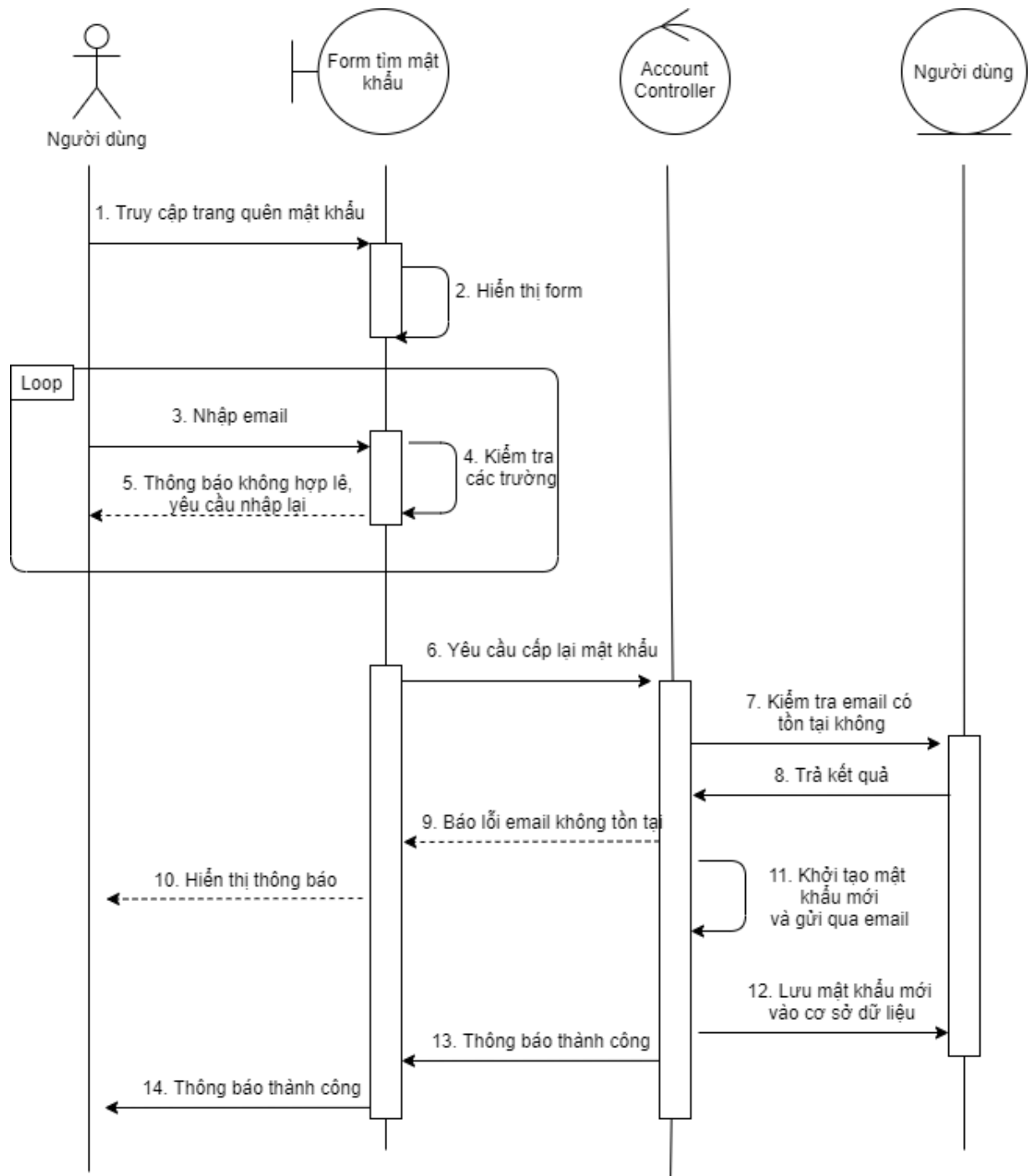
Hình 3.15. Thống kê nhà hàng

- Usecase Đăng ký



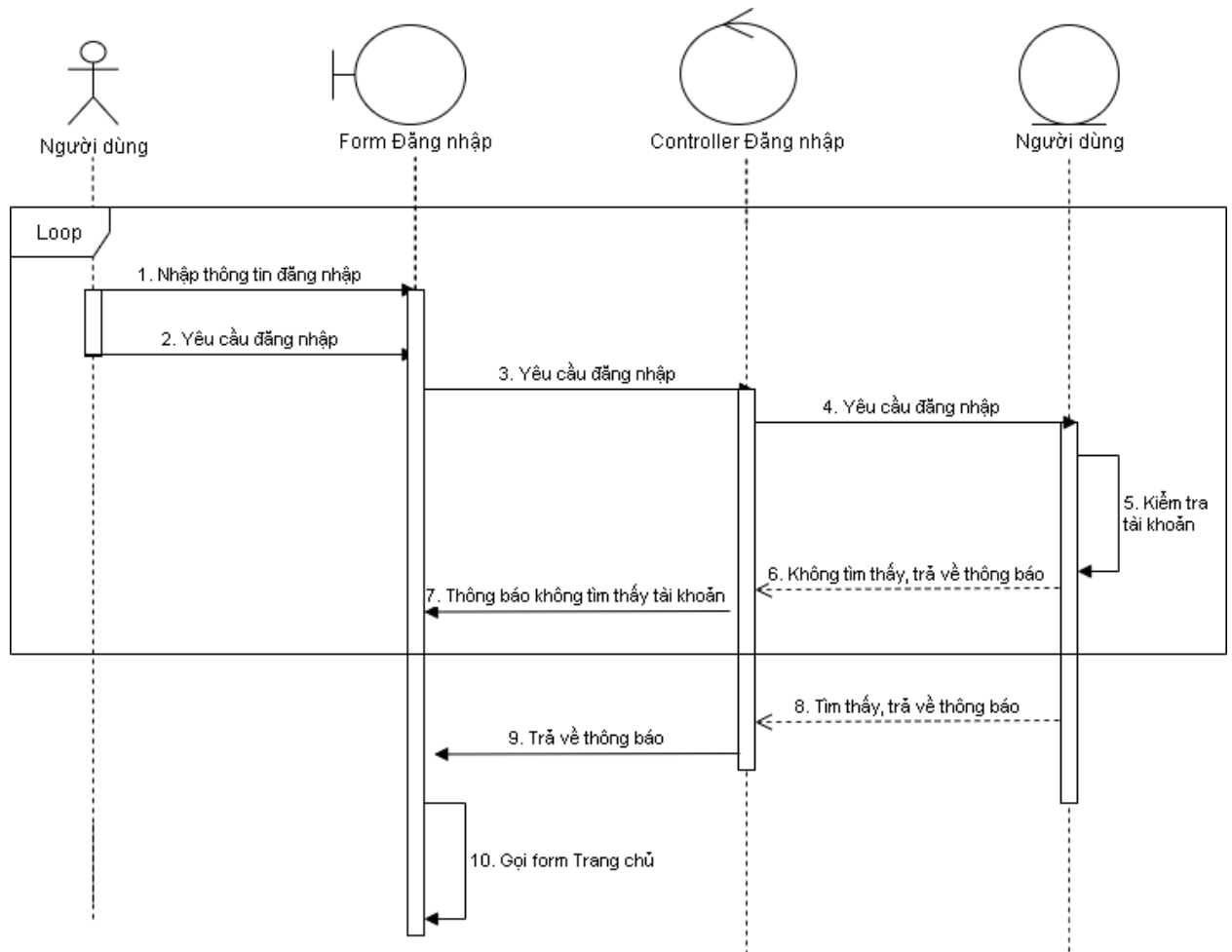
Hình 3.16. Đăng ký tài khoản

- Usecase Tìm mật khẩu



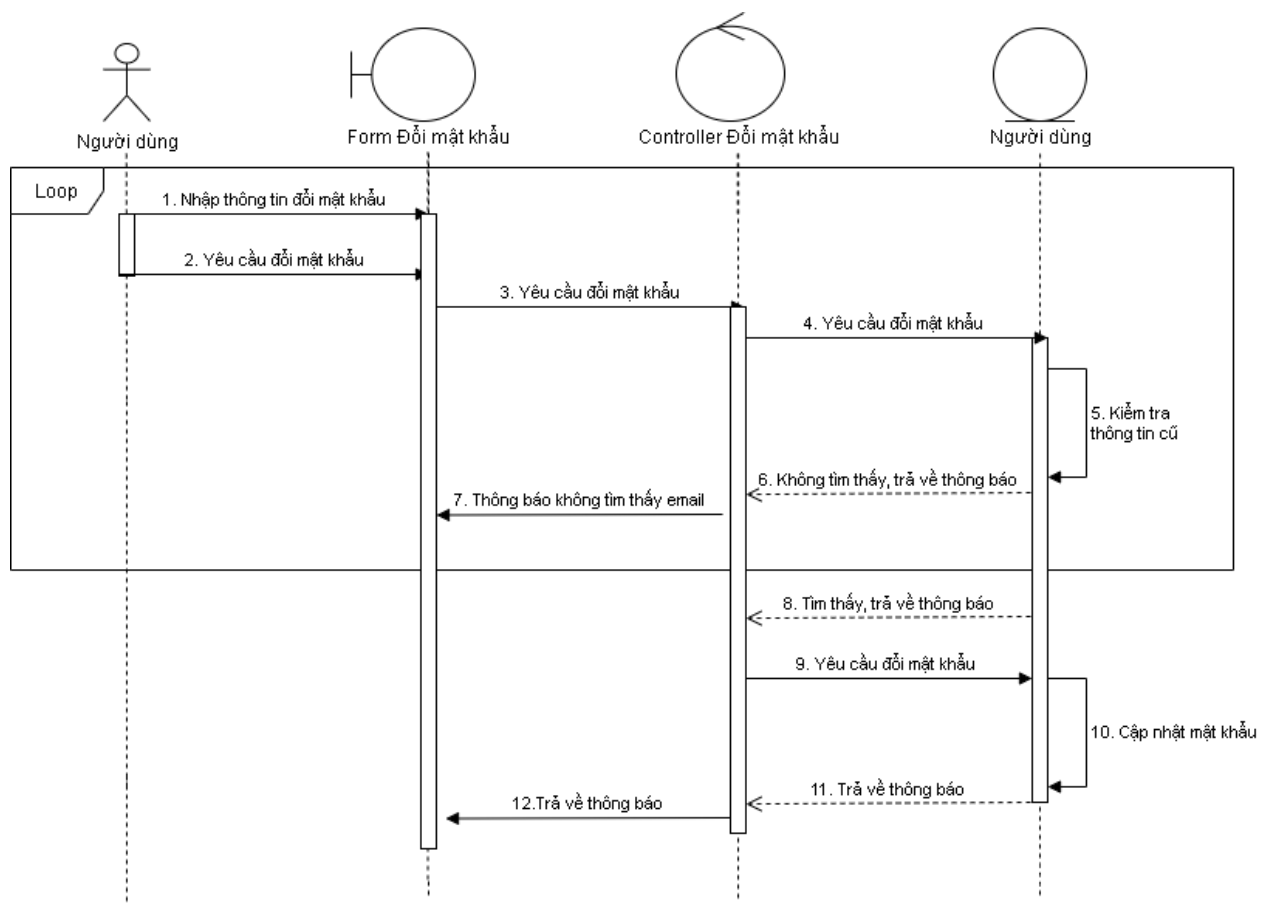
Hình 3.17 Tìm lại mật khẩu

- Usecase Đăng nhập



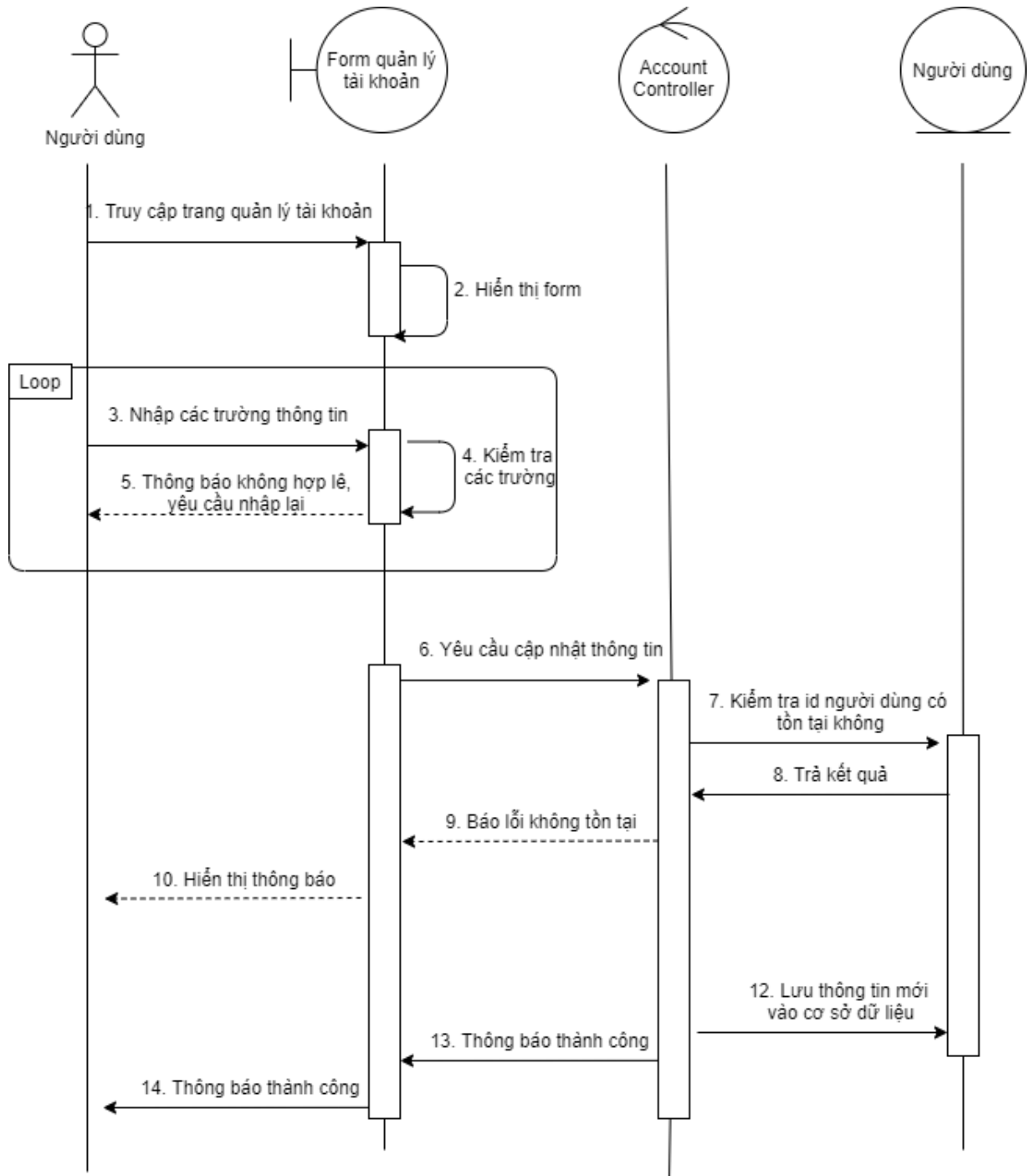
Hình 3.18. Đăng nhập

- Usecase Đổi mật khẩu



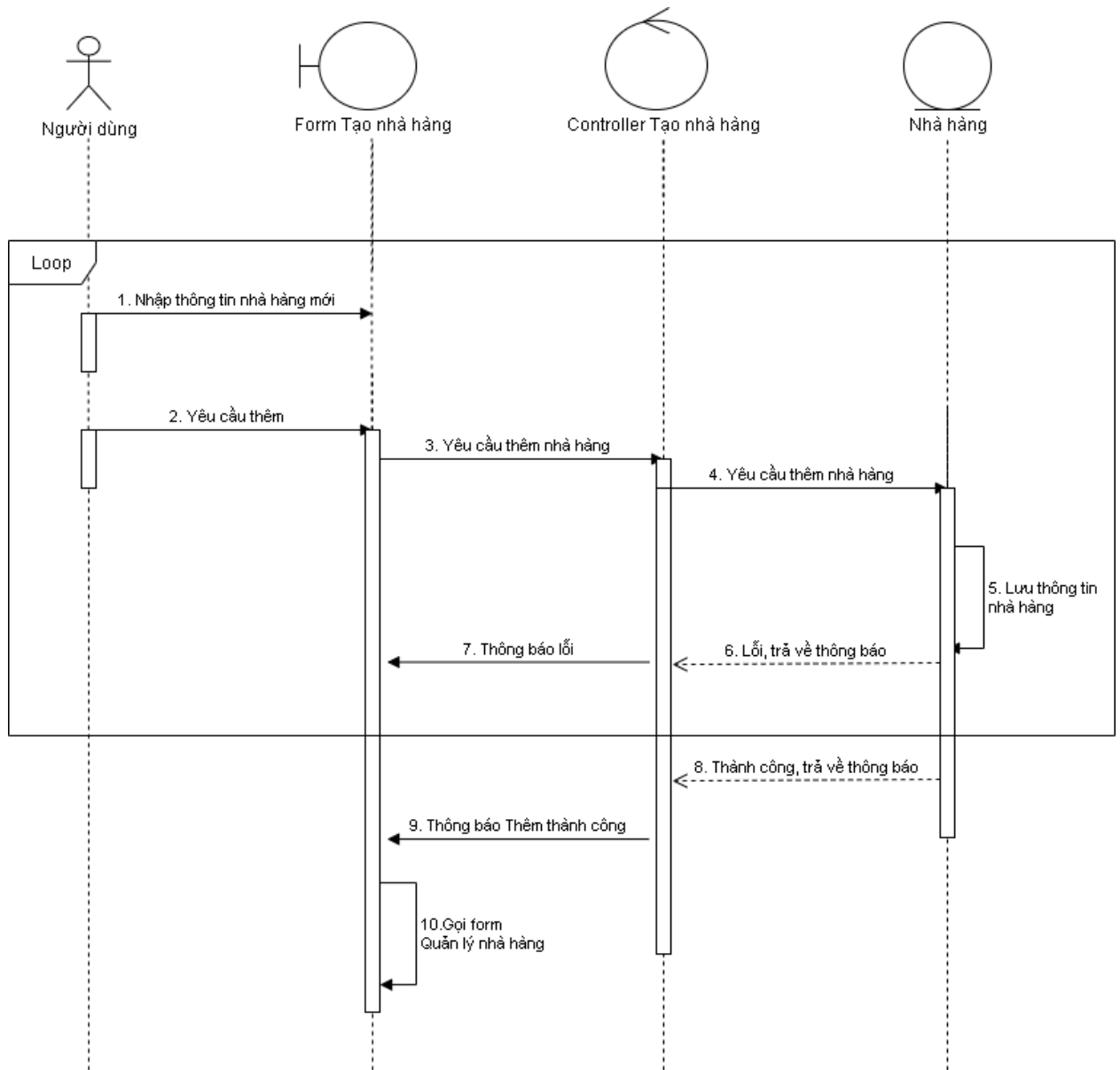
Hình 3.19 Đổi mật khẩu

- Usecase Quản lý thông tin cá nhân



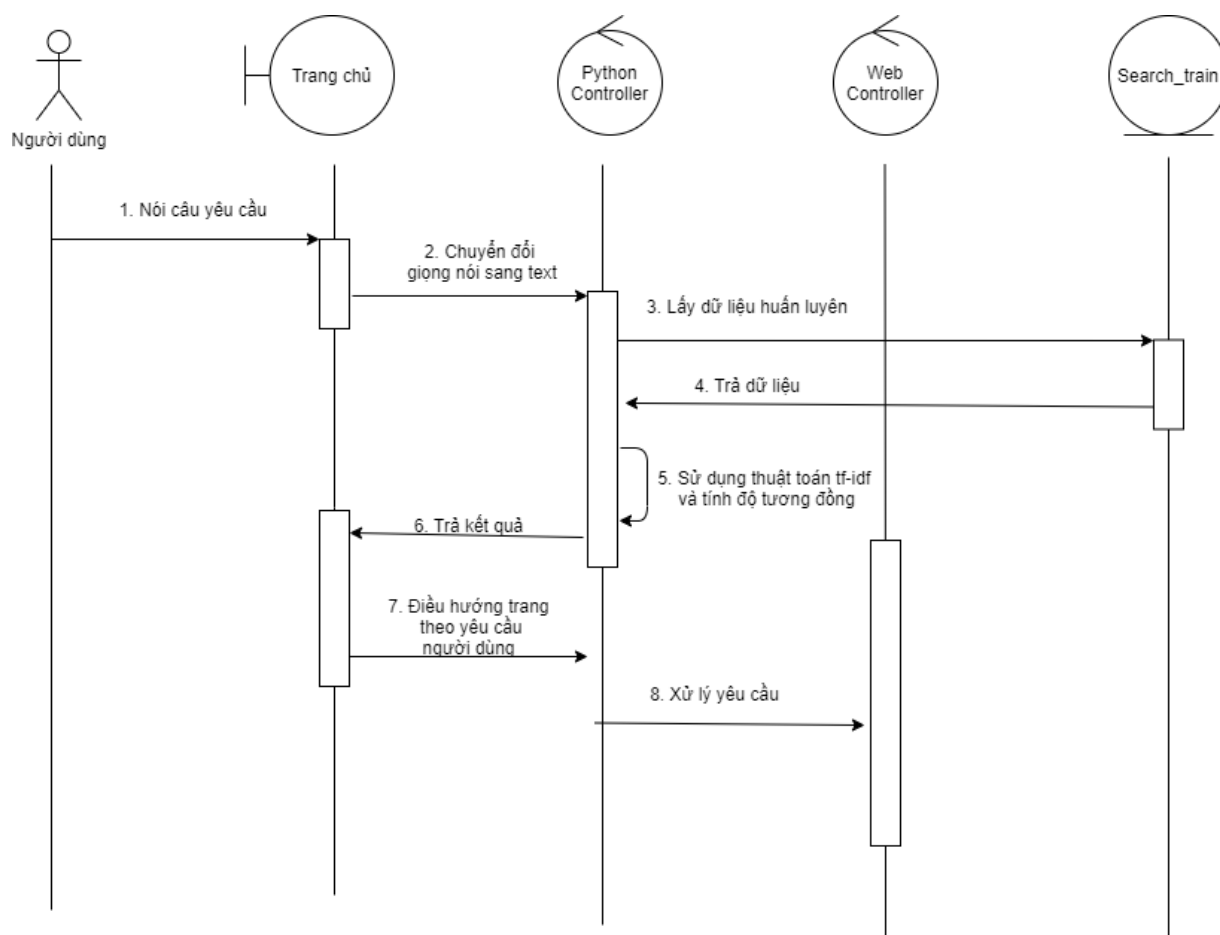
Hình 3.20 Quản lý thông tin cá nhân

- Usecase Tạo nhà hàng (Người dùng)



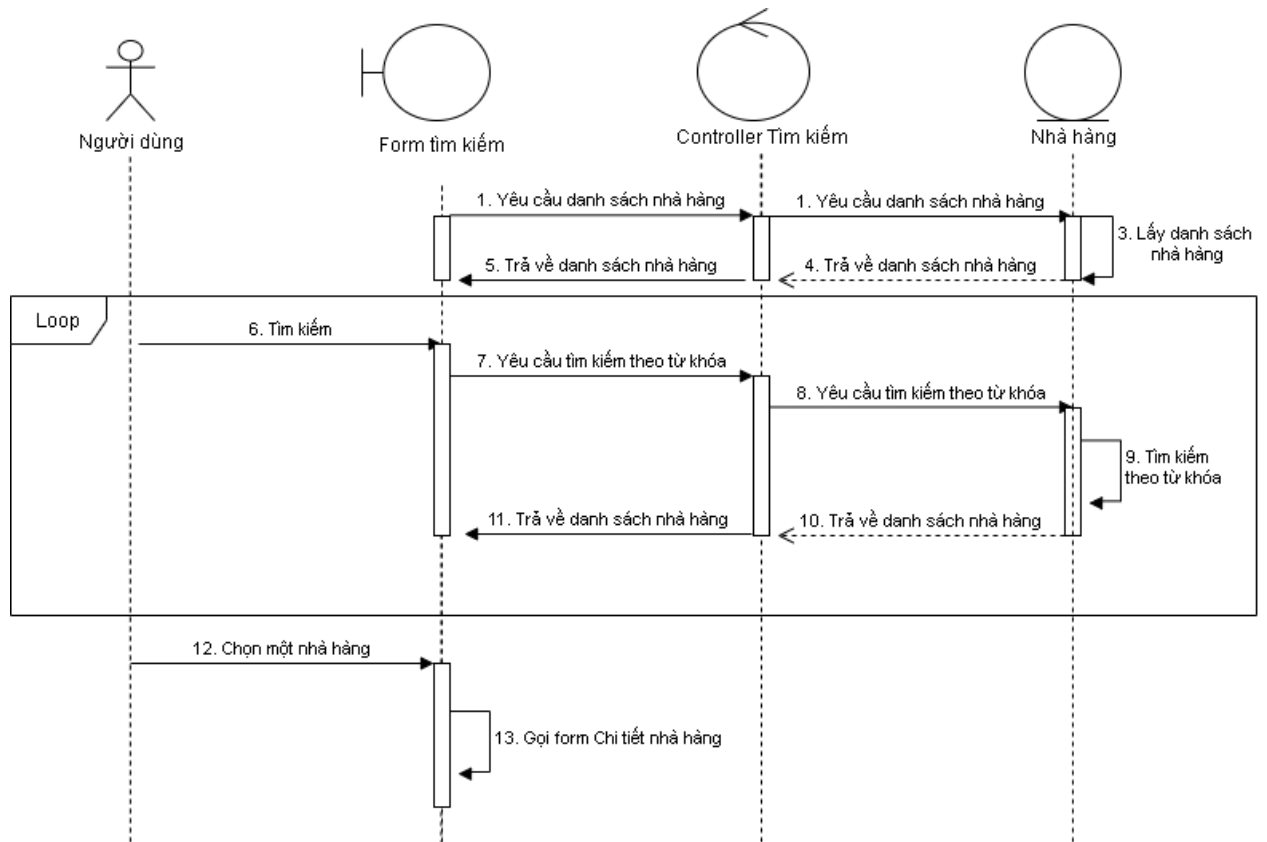
Hình 3.21. Tạo nhà hàng

- Usecase Điều khiển bằng giọng nói



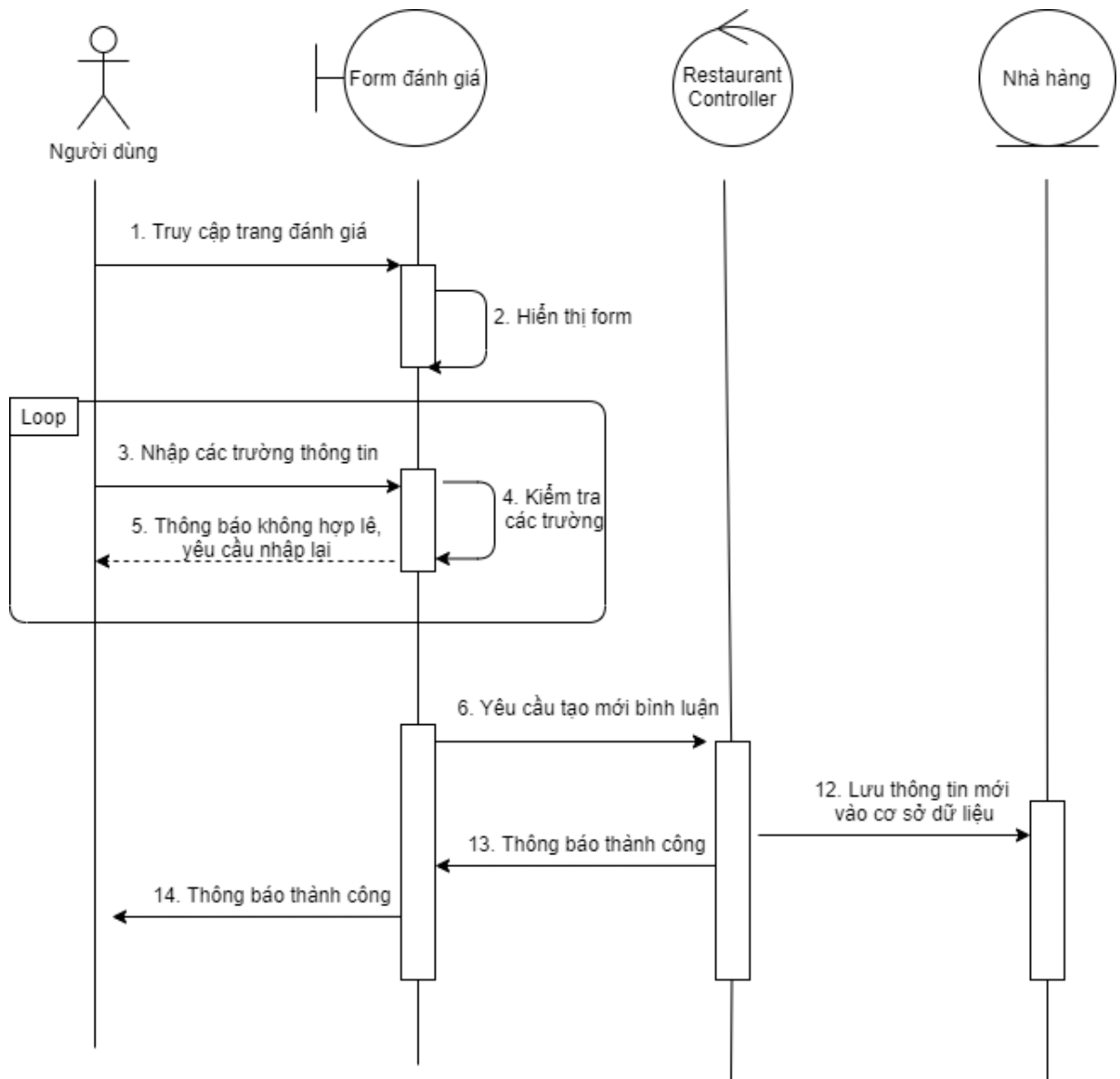
Hình 3.22 Điều khiển giọng nói

- Usecase Tìm kiếm nhà hàng



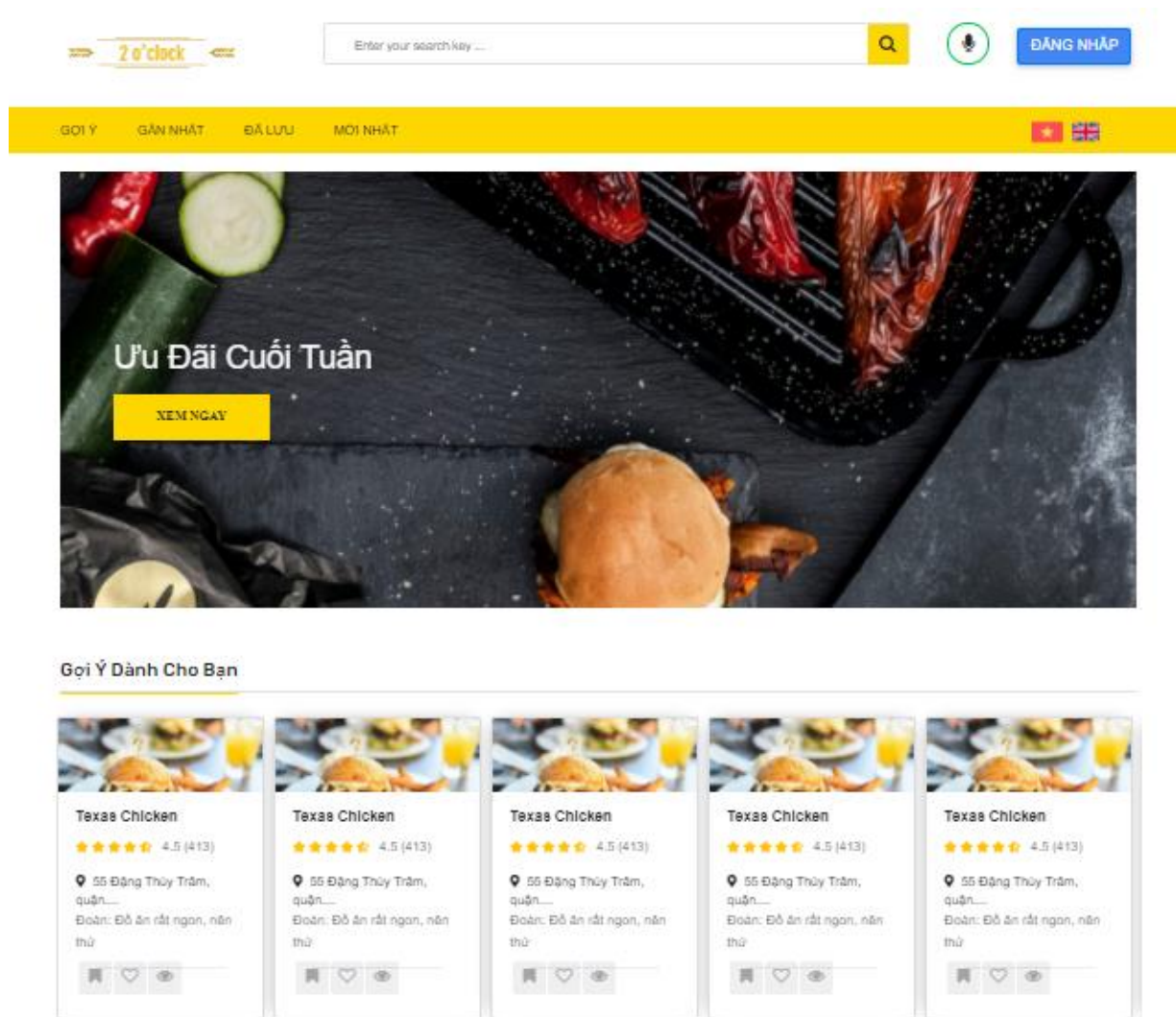
Hình 3.23 Tìm kiếm nhà hàng

- Usecase đánh giá




Hình 3.24. Đánh giá nhà hàng

3.4. Xây dựng giao diện:



Hình 3.25 Giao diện trang chủ



Quán ăn

Cháo Vịt Thanh Đa (Gốc Nhà Lá) - Bình Quới

6

4.5

5.5

9.2

3.2

5.4

Chất lượng

Dịch vụ

Giá cả

Vị trí

Không gian

📍 118 Bình Quới, P. 27, Quận Bình Thạnh, TP. HCM

🕒 07:00 - 23:00

💰 25.000đ - 55.000đ

Bình luận

Add to wishlist

Facebook


Twitter

Google+

Instagram

Reviews

Description



Hương Phạm

08/09/2012


8.4

Hôm qua mình vừa đi ăn với gia đình ở đây. Nghe đồn quán này đã mở mấy chục năm rồi, là quán cháo vịt mở đầu tiên và nổi tiếng nhất khu này. Ăn xong mới thấy đúng là trên cả tuyệt vời. ♥♥♥ Ở đây vịt chỉ bán nguyên con hoặc nửa con, dọn kèm với gói đồ riêng. Ngoài ra, còn có thể gọi cháo, bún, miến ăn kèm với gói vịt. Thịt vịt mềm ơi là mềm, gói cũng được trộn dầu giấm ngon lắm. Đặc biệt nhất là cháo ở đây. Cháo nấu với đậu xanh, ăn vừa thơm vừa béo, rất là ngon luôn. Mình là người cực kì ghét ăn cháo, nhưng hôm qua mình đã ăn hết 2 chén cháo đây. ♥ Ăn xong vẫn còn thèm muốn ăn nữa. :D Miến ở đây thì lại rất bít, ko ngon đặc biệt như cháo. Đến quán cháo thì ăn cháo là lựa chọn hợp lý nhất. :)) Quán rất rộng rãi, thoáng mát và sạch sẽ, tuy đông nhưng ko hề xô bồ. Thích nhất là WC của quán. Sạch sẽ cực kì, ko bẩn bẩn như các quán cháo vịt khác.

3 người đã thích

Like

Report



Nguyễn Việt Hưng

04/11/2012

7.4

90 người dùng đã chia sẻ

25

Tuyệt vời

7

Khá tốt

7

Trung bình

2

Kém

Vị trí

Giá cả

Chất lượng

Dịch vụ

Không gian

6

điểm

Hình 3.26 Giao diện trang chi tiết



Cháo Vịt Thanh Đa (Gốc Nhà Lá) - Bình Quới

118 Bình Quới, P. 27, Quận Bình Thạnh, TP. HCM

Đánh giá

Vị trí	<div></div>	5
Giá cả	<div></div>	5
Chất lượng	<div></div>	5
Dịch vụ	<div></div>	5
Không gian	<div></div>	5

Trung bình: 5

Chọn hình

GỬI

Hình 3.27. Màn hình đánh giá

THÔNG TIN CÁ NHÂN



Choose File No file chosen

Tên

Email

Giới tính

☒ Nam ☐ Nữ ☐ Khác

Mật khẩu mới

Mật khẩu cũ

LƯU

Hình 3.28. Màn hình quản lý tài khoản


Thêm nhà hàng mới

Tên nhà hàng

Địa chỉ

Địa chỉ ảnh

Tọa độ nhà hàng



Vĩ độ

Kinh độ

Loại hình

Quận

Thành phố

Giờ mở cửa

Phút mở cửa

Giờ đóng cửa

Phút đóng cửa

Giá tiền thấp nhất

Giá tiền cao nhất

LƯU

Hình 3.29. Màn hình thêm mới địa điểm

CHƯƠNG 4: CÀI ĐẶT VÀ TRIỂN KHAI HỆ THỐNG

4.1. Triển khai server trên Heroku Cloud Platform và dịch vụ Azure Database for MySQL

Heroku là nền tảng đám mây cho phép các cá nhân, doanh nghiệp xây dựng và phát triển ứng dụng. Sử dụng Heroku là cách nhanh nhất để chuyển từ ý tưởng sang URL, bỏ qua tất cả những cơ sở hạ tầng phức tạp khó hiểu.

Một số tính năng của Heroku:

- Heroku Runtime: Ứng dụng sẽ chạy trên Smart Container được quản lý xuyên suốt thời gian chạy. Heroku sẽ xử lý những thứ quan trọng như: Giao diện, điều phối, cân bằng tải, bảo mật,...
- Heroku Redis: Hỗ trợ dịch vụ Redis, giúp lưu trữ trên bộ nhớ với tốc độ nhanh phổ biến.
- Add-ons: Giúp quản lý các ứng dụng với dịch vụ được tích hợp sẵn như MongoDB, ClearDB,...
- Github Integration: Tích hợp với github, có thể pull, push, commit...hoạt động với mọi branch

Azure Database for MySQL là một dịch vụ cơ sở dữ liệu quan hệ trên nền tảng đám mây Microsoft. Azure Database for MySQL mang lại một số tính năng:

- Tính khả dụng cao.
- Bảo vệ dữ liệu với việc backup tự động.
- Tự động bảo trì hệ thống đảm bảo hệ thống được an toàn và luôn được cập nhật.
- Quản lý truy cập thông qua các luật (rules) với tường lửa.

4.2. Giải pháp công nghệ phía Client

4.2.1. Xác thực với Json Web Token (JWT)¹

Token-based authentication là phương thức xác thực bằng chuỗi mã hóa. Một hệ thống sử dụng Token-based authentication cho phép người dùng nhập user/password để nhận về 1 chuỗi token. Chuỗi Token này được sử dụng để “xác minh” quyền truy cập vào tài nguyên mà không cần phải cung cấp lại username/password nữa.

JSON Web Token là một chuỗi mã hóa mà nguồn gốc ban đầu là một chuỗi JSON.

Cấu trúc của một JWT gồm 3 phần: Header, Payload, Signature. Có dạng như sau:

<base64-encoded header>.<base64-encoded payload>.<HMACSHA256(base64-encoded signature)>

Header: Gồm 2 phần chính:

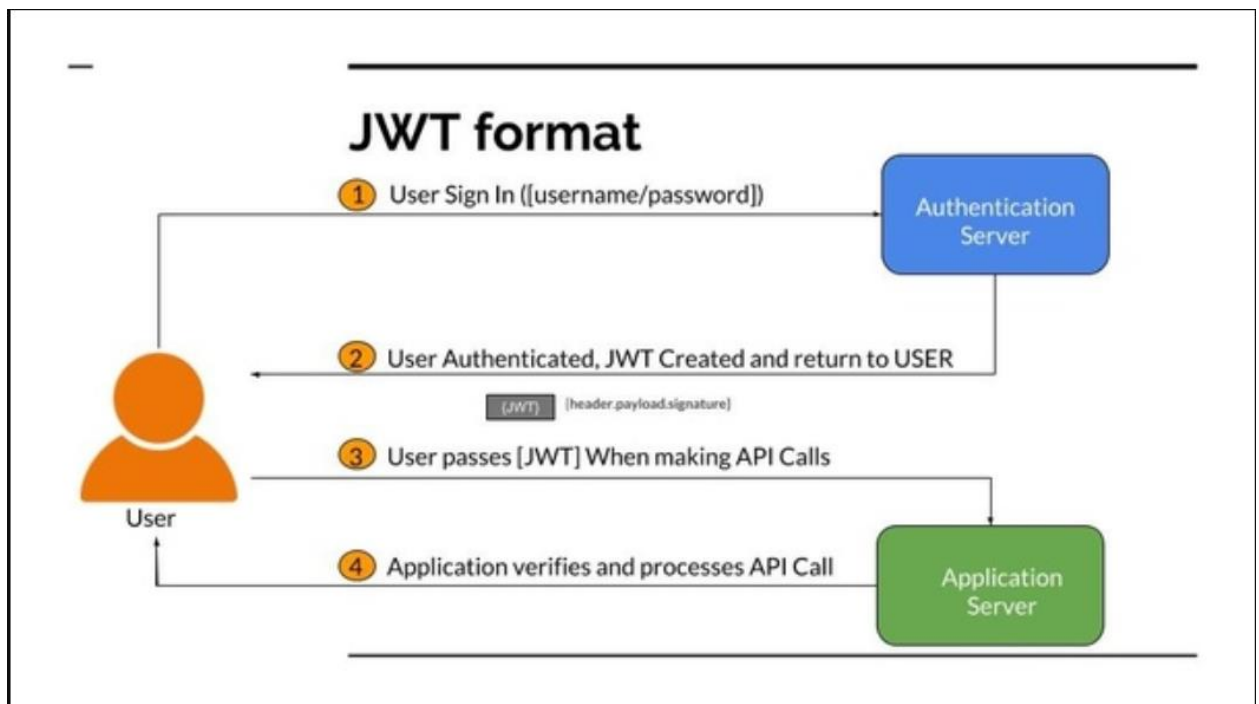
- Typ: Loại Token (mặc định là JWT- cho biết đây là một Token JWT).
- Alg: Thuật toán dùng để mã hóa

Payload: Là nơi chứa các nội dung của thông tin (claim). Thông tin truyền đi có thể là mô tả của 1 thực thể (ví dụ như người dùng) hoặc cũng có thể là các thông tin bổ sung thêm cho phần Header.

Signature: Phần chữ ký được tạo bằng cách kết hợp 2 phần Header + Payload, rồi mã hóa nó lại bằng 1 giải thuật encode bất kỳ ví dụ như HMAC SHA-256.

Luồng hoạt động:

¹Tham khảo tại <https://topdev.vn/blog/json-web-token-la-gi>



Hình 4.1. Luồng hoạt động của jwt

- **Bước 1:** User thực hiện login bằng cách gửi id/password hay sử dụng các tài khoản mạng xã hội lên phía Authentication Server (Server xác thực)
- **Bước 2:** Authentication Server tiếp nhận các dữ liệu mà User gửi lên để phục vụ cho việc xác thực người dùng. Trong trường hợp thành công, Authentication Server sẽ tạo một JWT và trả về cho người dùng thông qua response.
- **Bước 3:** Người dùng nhận được JWT do Authentication Server vừa mới trả về làm “chìa khóa” để thực hiện các “lệnh” tiếp theo đối với Application Server.
- **Bước 4:** Application Server trước khi thực hiện yêu cầu được gọi từ phía User, sẽ xác thực JWT gửi lên. Nếu hợp lệ, tiếp tục thực hiện yêu cầu được gọi.

4.2.2. Công cụ ghi giọng nói

Cộng đồng W3C vào năm 2012 đã giới thiệu API Web Speech. Nhưng Google Chrome dường như là công cụ tìm kiếm duy nhất tận dụng các nguyên tắc này. Do đó, công cụ nhận dạng giọng nói của Google có thể được sử dụng để phát triển các tính năng nhận dạng giọng nói sáng tạo. Sự hiện diện của Google hiện cũng làm cho nó miễn phí.

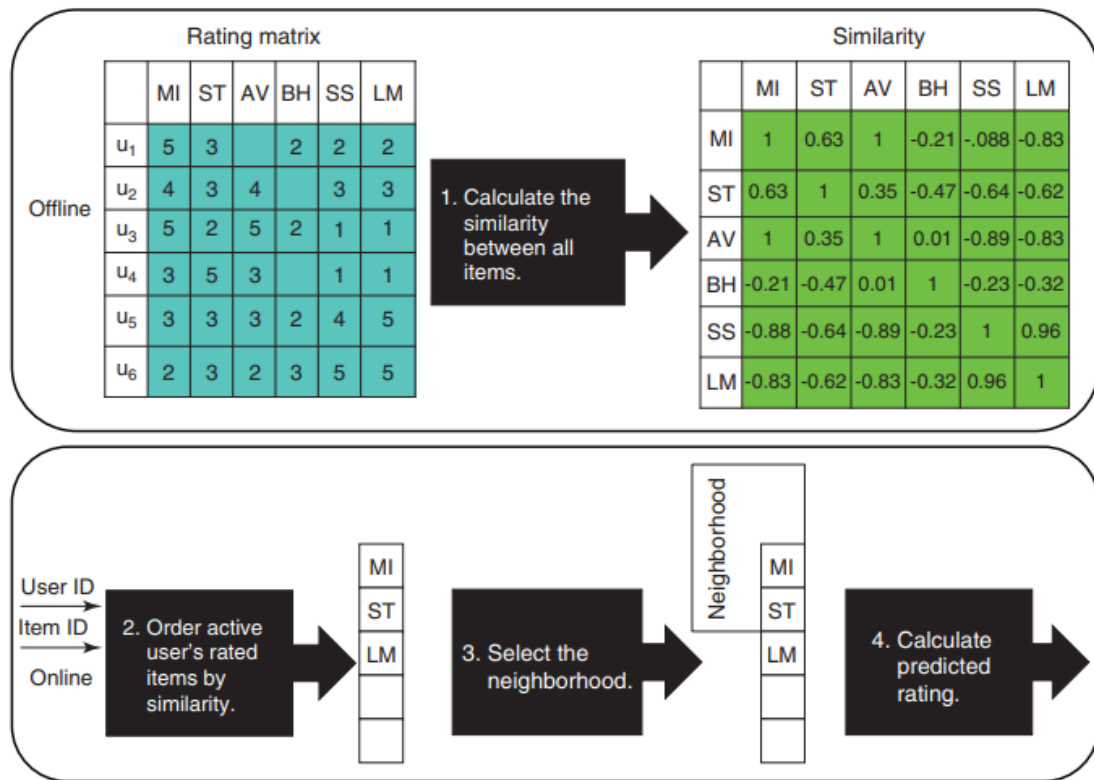
Web Speech API có thể cho phép chúng ta nhập giọng nói và tạo ra một dạng văn bản tương ứng. Văn bản sau đó sẽ thúc đẩy các tính năng khác nhau trên trang web như kiểm soát tab, tìm kiếm dựa trên giọng nói, đặt chỗ, thanh toán, đọc mật khẩu, v.v.

4.3. Giải pháp công nghệ phía Server

4.3.1. Ứng dụng thuật toán lọc cộng tác

Hệ thống gợi ý bằng thuật toán lọc cộng tác được thực hiện qua hai phiên:

- Xây dựng sẵn tập dữ liệu tương đồng giữa các nhà hàng và lưu vào một bảng trong cơ sở dữ liệu (Offline)
- Đưa ra gợi ý khi một user truy cập vào hệ thống (Online).



Hình 4.2. Ứng dụng thuật toán lọc cộng tác¹

Phiên Offline

Bước 1: Xây dựng ma trận đánh giá với mỗi ô là điểm đánh giá của user cho từng nhà hàng. Giả sử ta được ma trận đánh giá như sau:

¹ Kim Falk, "Practical Recommender Systems", Page 188, 2019

	Nhà hàng A	Nhà hàng B	Nhà hàng C	Nhà hàng D	Nhà hàng E
User 1	8	9		2	1
User 2	8	8	9	3	2
User 3	1	2	10	8	9
User 4	2	3	9	8	2
User 5	8				
Trung bình	5.4	5.5	9.3	5.25	3.5

Bảng 4.1. Ma trận đánh giá

- Nhìn vào ma trận đánh giá trên, nhận thấy rằng User 1 và User 2 có “khẩu vị” gần giống nhau. Cả hai đều đánh giá cao cho 2 nhà hàng A, B và cả hai đều không thích nhà hàng D, E với điểm đánh giá rất thấp. Ta cũng thấy User 2 đánh giá cao cho nhà hàng C, như vậy rất có thể User 1 cũng sẽ thích nhà hàng này và ta có thể đề xuất nhà hàng C cho user 1. Đây chính là hướng tiếp cận theo User-based filtering.
- Ma trận trên cũng cho thấy, nhà hàng A và B có mức độ tương đồng khá lớn khi nhận được điểm đánh giá cao từ user 1 và user 2, đồng thời nhận được điểm đánh giá rất thấp từ hai user 3 và 4. Như vậy ta có đề xuất nhà hàng B cho user 5, vì user này đánh giá cao cho hàng A và rất có thể cũng sẽ thích nhà hàng B. Đây là cách tiếp cận item-based filtering.

Bước 2: Xây dựng ma trận tương đồng giữa các item.

Trước tiên cần xây dựng ma trận chuẩn hóa, ta được kết quả:

	Nhà hàng A	Nhà hàng B	Nhà hàng C	Nhà hàng D	Nhà hàng E
User 1	2.6	3.5		-3.25	-2.5
User 2	2.6	2.5	-0.3	-2.25	-1.5
User 3	-4.4	-3.5	0.7	2.75	5.5
User 4	-3.4	-2.5	-0.3	2.75	-1.5
User 5	2.6				
Trung bình	5.4	5.5	9.3	5.25	3.5

Bảng 4.2. Ma trận chuẩn hóa

Trước khi thực hiện tính độ tương đồng giữa hai nhà hàng bằng công thức Cosine, cần quy định trước số lượng overlapping user nhỏ nhất (tức là số lượng user cùng đánh giá cho cả hai nhà hàng).

Lý do: Giả sử nhà hàng A và nhà hàng B chỉ có user 1 đánh giá, lần lượt có số điểm bằng -0.5 và 2.5 (Giá trị đã được chuẩn hóa). Nếu áp dụng công thức cosine ta được:

$$\text{Sim}(A,B) = \frac{(-0.5)*2.5}{\sqrt{(-0.5)^2} * \sqrt{2.5^2}} = 1$$

Theo kết quả trên, nhà hàng A và B hoàn toàn tương đồng nhau, như vậy là hoàn toàn sai vì theo điểm đánh giá 2 nhà hàng này là gần như trái ngược.

Với ma trận chuẩn hóa trên, giả sử quy định min overlapping user = 3. Ta có độ tương đồng giữa nhà hàng A và B:

$$\text{Sim}(A,B) = \frac{2.6*3.5 + 2.6*2.5 + (-4.4)*(-3.5) + (-3.4)*(-2.5) + 5.4*5.5}{\sqrt{2.6^2+2.6^2+4.4^2+3.4^2+5.4^2} * \sqrt{3.5^2+2.5^2+3.5^2+2.5^2+5.5^2}}$$

$$= 0.98$$

Tương tự cho các nhà hàng còn lại, được ma trận tương đồng như sau:

	Nhà hàng A	Nhà hàng B	Nhà hàng C	Nhà hàng D	Nhà hàng E
Nhà hàng A	1	0.98	-0.41	-0.97	-0.7
Nhà hàng B	0.98	1	-0.42	-0.85	-0.27
Nhà hàng C	-0.41	-0.42	1	0.48	0.98
Nhà hàng D	-0.97	-0.85	0.48	1	0.63
Nhà hàng E	-0.7	-0.27	0.98	0.63	1

Bảng 4.3 Ma trận tương đồng

Bước 3: Lưu kết quả vào bảng “similarity” những nhà hàng có mức độ tương đồng lớn hơn 0. Ta được kết quả như sau:

<input type="checkbox"/>	id	created	source	target	similarity
<input type="checkbox"/>	173	2020-12-25	643	640	0.2606513
<input type="checkbox"/>	174	2020-12-25	687	640	0.1513515
<input type="checkbox"/>	175	2020-12-25	1043	640	0.0286865
<input type="checkbox"/>	176	2020-12-25	1393	640	0.1448058
<input type="checkbox"/>	177	2020-12-25	1422	640	0.2239976
<input type="checkbox"/>	178	2020-12-25	1967	640	0.4181058
<input type="checkbox"/>	179	2020-12-25	640	643	0.2606513
<input type="checkbox"/>	180	2020-12-25	1043	643	0.1184530
<input type="checkbox"/>	181	2020-12-25	1256	643	0.2074143
<input type="checkbox"/>	182	2020-12-25	1393	643	0.1579632
<input type="checkbox"/>	183	2020-12-25	1399	643	0.1128499
<input type="checkbox"/>	184	2020-12-25	1422	643	0.1289046
<input type="checkbox"/>	185	2020-12-25	1432	643	0.0128593
<input type="checkbox"/>	186	2020-12-25	1776	643	0.0921987
<input type="checkbox"/>	187	2020-12-25	1821	643	0.1697605

Phiên Online: Giả sử một user có id=43037 truy cập vào hệ thống

Bước 1: Lấy dữ liệu đánh giá của user này.

Bước 2: Lấy ra những nhà hàng tương đồng với những nhà hàng mà user này đã đánh giá dựa vào dữ liệu bảng “similarity”.

Bước 3: Sắp xếp kết quả từ bước 2 theo chiều giảm dần của độ tương đồng.

Bước 4: Chọn ra những item lân cận (neighbors) bằng phương pháp Threshold với giá trị tương đồng nhỏ nhất bằng 0 và dự đoán điểm đánh giá. Ta được kết quả như sau:

```
[
  [
    275791,
    {
      "prediction": "8.8",
      "sim_items": [
```

```

        {
            "id": 273807,
            "sim": "0.1126375"
        }
    ]
}
],
[
    706330,
    {
        "prediction": "8.5",
        "sim_items": [
            {
                "id": 707162,
                "sim": "0.2433587"
            }
        ]
    }
],
[
    705584,
    {
        "prediction": "8.5 ",
        "sim_items": [
            {
                "id": 707162,
                "sim": "0.1375636"
            }
        ]
    }
],
[
    706609,
    {

```

```

        "prediction": "8.5",
        "sim_items": [
            {
                "id": 707162,
                "sim": "0.0928382"
            }
        ]
    },
],
[
    646337,
    {
        "prediction": "8.5",
        "sim_items": [
            {
                "id": 647052,
                "sim": "0.0799900"
            }
        ]
    }
],
[
    34009,
    {
        "prediction": "8.4",
        "sim_items": [
            {
                "id": 36729,
                "sim": "0.4496109"
            }
        ]
    }
]
]

```

Trong đó

- prediction là điểm dự đoán.
- sim_items: Danh sách các id nhà hàng tương đồng.
- sim: Giá trị tương đồng.

4.3.2. Ứng dụng thuật toán Tf-idf trong điều khiển bằng giọng nói:

Tạo tập huấn luyện như sau:

Id	Action	Input
8	LOGOUT	Đăng xuất
9	LOGOUT	Hãy đăng xuất
10	LOGOUT	Thoát tài khoản
11	LOGOUT	Thoát ra
12	NEAREST	Tìm nhà hàng gần nhất
13	NEAREST	Quán nào gần đây
14	NEAREST	Gần đây có quán nào không
15	NEAREST	Xung quanh đây có nhà hàng nào
16	NEAREST	Tìm quán xung quanh đây
17	NEAREST	Tìm nhà hàng gần tôi
18	NEAREST	Quán nào gần nhất
18	NEAREST	Nhà hàng có khoảng cách ngắn nhất
20	NEAREST	Quán nào tới nhanh nhất
21	OPEN	Còn quán nào mở cửa không
22	OPEN	Giờ này còn quán nào mở không
23	OPEN	Nhà hàng nào chưa đóng cửa
24	OPEN	Tìm kiếm nhà hàng còn mở cửa

Bảng 4.4. Tập dữ liệu huấn luyện

Dưới client gửi câu yêu cầu lên server (sau khi đã chuyển giọng nói sang câu văn bản).

Đầu tiên cần vector hóa các document (tức là các giá trị input). Toàn bộ dataset có 33 từ vựng -> Mỗi document là một vector có độ dài 33.

Với 1 vector, gán giá trị Tf-idf tương ứng của từ đó trong document.

Ví dụ người dùng gửi lên câu yêu cầu như sau: “Còn nhà hàng nào mở cửa không”.

Ta được vector cho câu yêu cầu này:

```
[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1.04982212, 1.04982212, 0, 0, 0.7985077 ,0, 1
.60943791, 0, 0, 0, 0, 0, 0, 0, 0, 1.60943791, 1.60943791, 1.60943791, 0, 0,
0, 0, 0]
```

Tương tự ta có được các vector của các document còn lại.

Dùng công thức Cosine để tính độ tương đồng giữa hai vector, ta được kết quả là danh sách các document được sắp xếp giảm dần theo mức độ tương đồng.

```
"data": [
    {
        "id": "21",
        "action": "OPEN",
        "sim": 0.8701591450460358
    },
    {
        "id": "24",
        "action": "OPEN",
        "sim": 0.5915077246154561
    },
    {
        "id": "22",
        "action": "OPEN",
        "sim": 0.4416062334253969
    },
    {
        "id": "23",
        "action": "OPEN",
```

```

    "sim": 0.3091748981663338
  },
  {
    "id": "14",
    "action": "NEAREST",
    "sim": 0.2518563137455107
  },
  {
    "id": "12",
    "action": "NEAREST",
    "sim": 0.1988511257945403
  },
  {
    "id": "15",
    "action": "NEAREST",
    "sim": 0.17650271215801447
  },
  {
    "id": "17",
    "action": "NEAREST",
    "sim": 0.15314490642013093
  },
  {
    "id": "19",
    "action": "NEAREST",
    "sim": 0.10207836588290996
  },
  {
    "id": "13",
    "action": "NEAREST",
    "sim": 0.07018064801446758
  },
  {
    "id": "18",
    "action": "NEAREST",
    "sim": 0.07018064801446758
  },
  {

```

```
"id": "20",  
"action": "NEAREST",  
"sim": 0.037177055311148156  
}]
```

Theo kết quả trên, câu yêu cầu từ người dùng gần giống nhất với document có id bằng 21 với nội dung : “Còn quán nào mở cửa không”. Server tiếp tục điều hướng thực hiện yêu cầu của người dùng.

4.3.3. Tìm kiếm nhà hàng có khoảng cách gần nhất

Sử dụng công thức Haversine tính khoảng cách giữa người dùng với tất cả các nhà hàng. Sau đó lấy ra những nhà hàng gần nhất.

Tuy nhiên, công thức Haversine dùng để tính khoảng cách giữa hai điểm tọa độ trên mặt cầu, do đó bỏ qua các yếu tố địa hình, giao thông,... => Kết quả có độ chính xác không cao.

Giải pháp: Sau khi có được kết quả từ công thức Haversine, tiếp tục sử dụng api tính khoảng cách của Google Map (Api của google map giới hạn số lượng tham số đầu vào, do đó không thể sử dụng để tính khoảng cách cho tất cả nhà hàng trong cơ sở dữ liệu)

https://maps.googleapis.com/maps/api/distancematrix/json?units=imperial&origins=40.6655101,-73.89188969999998&destinations=40.6905615%2C-73.9976592%7C40.6905615%2C-73.9976592%7C40.6905615%2C&key=YOUR_API_KEY

Trong đó:

- origins: Tọa độ người dùng.
- destinations: Tọa độ của các nhà hàng

Kết quả được hiển thị trên website:



Hình 4.3. Hiện thị các nhà hàng gần nhất

CHƯƠNG 5: KẾT LUẬN

5.1. Kết quả đạt được

Thông qua quá trình thực hiện đề tài khóa luận, nhóm đã học hỏi được nhiều kiến thức về các thuật toán gợi ý, đặc biệt là thuật toán lọc cộng tác và ứng dụng được thuật toán Tf-idf cho chức năng điều khiển ứng dụng bằng giọng nói. Ngoài ra nhóm cũng trau dồi được nhiều kiến thức liên quan đến cloud platform và kỹ năng lập trình giao diện.

Về mặt chức năng, trang web đã hoàn thành các chức năng đã đặt ra:

- Chức năng bình luận, chia sẻ, đánh giá nhà hàng.
- Tìm kiếm nhà hàng gần nhất.
- Đề xuất các nhà hàng phù hợp với sở thích người dùng.
- Chức năng điều khiển bằng giọng nói.
- Quản lý admin.
- Giao diện website có hỗ trợ cả tiếng Anh và tiếng Việt.
- Quản lý lịch sử hoạt động người dùng.

5.2. Ưu điểm của hệ thống:

- Giao diện thân thiện, dễ sử dụng.
- Chức năng điều khiển bằng giọng nói khá hữu ích khi người dùng sử dụng trang web ngoài đường.
- Hệ thống đề xuất cũng như tìm kiếm nhà hàng gần nhất giúp cải thiện chất lượng phục vụ.

5.3. Hạn chế của hệ thống

- Hệ thống hiện tại chỉ đề xuất dựa trên điểm đánh giá người dùng, trong khi nhiều người dùng không sẵn lòng đánh giá sau khi họ trải nghiệm.
- Phiên bản hiện tại chỉ có trên nền tảng web.
- Giao diện website chưa thật sự đẹp.

- Chức năng tìm kiếm giọng nói còn một số hạn chế về ngôn ngữ.

5.4. Hướng phát triển

Trong tương lai, nhóm sẽ cố gắng khắc phục những điểm trên bằng cách:

- Hỗ trợ đa ngôn ngữ cho chức năng điều khiển giọng nói.
- Ghi lại lịch sử thao tác người dùng nhằm nâng cao hiệu quả của tính năng gợi ý.
- Có phiên bản dành cho điện thoại.
- Cải thiện giao diện website.
- Tích hợp thêm một số chức năng như liên kết các nhà hàng, đặt món trực tuyến,..

TÀI LIỆU THAM KHẢO

Tiếng Việt

[1] Thuật toán lọc cộng tác

<https://machinelearningcoban.com/>

[2] Ứng dụng thuật toán tf-idf

<https://viblo.asia/p/tf-idf-algorithm-text-retrieval-and-search-engines-1Je5EmGY5nL>

[3] Triển khai dự án Laravel lên Heroku

<https://viblo.asia/p/deploy-mot-project-laravel-len-heroku-maGK7j1A5j2>

Tiếng Anh

[1] Dr. P. V. Ingole, Mr. Mangesh K Nichat. “Landmark based shortest path detection by using Dijkstra Algorithm and Haversine Formula”, Vol.3, Issue 3, Page 163, 5-2013.

https://www.researchgate.net/publication/282314345_Landmark_based_shortest_path_detection_by_using_Dijkstra_Algorithm_and_Haversine_Formula

[2] TF-IDF Algorithm

<http://www.tfidf.com/>

[3] Kim Falk, “Practical Recommender Systems”, 2019

[4] Badrul Sarwar, et al., “Item-based Collaborative Filtering Recommendation Algorithms”, 08-2001

files.grouplens.org/papers/www10_sarwar.pdf

[5] scipy library

<https://docs.scipy.org/doc/scipy/reference/sparse.html>

[6] How to add speech recognition to a website

https://www.paymill.com/en/blog/how-to-add-speech-recognition-to-your-website/?fbclid=IwAR2w6_5Zg01FrTfqrZnvLAF7Dxi7rTVhKrdbAWYS3lHPBznX0A5BN9MxzW0

[7] Deploy a python project to Heroku

https://www.youtube.com/watch?v=6DI_7Zja8Zc