
TRƯỜNG ĐẠI HỌC PHENIKA
KHOA CÔNG NGHỆ THÔNG TIN



Báo Cáo Đồ án cơ sở

Đề tài: Contify.ai

Hệ thống viết mô tả sản phẩm tự động sử dụng AI

Thành viên:

- Nguyễn Chí Công – 20010846
- Nguyễn Lý Khánh Linh – 20010965
- Nguyễn Thị Mơ - 20010807

Giảng viên hướng dẫn: TS. Lương Văn Thiện

Hà Nội, Ngày 09 tháng 12 năm 2022

Mục lục

| | |
|--|-----------|
| 1. Yêu cầu (Requirements) | 2 |
| 1.1 Đặt vấn đề (Problem statement) | 2 |
| 1.1.1 Mô tả vấn đề..... | 2 |
| 1.1.2 Giải pháp..... | 2 |
| 1.2 Thuật ngữ (Glossary)..... | 2 |
| 1.3 Thông số kỹ thuật bổ sung..... | 3 |
| 1.4 Mô hình sử dụng hệ thống (Use-case models) | 5 |
| 1.4.1 Generate Product Descriptions | 6 |
| 1.4.2 Create Account..... | 7 |
| 1.4.3 Delete User..... | 8 |
| 1.4.4 Log In..... | 9 |
| 1.4.5 Update profile | 10 |
| 1.4.6 User Reports..... | 11 |
| 1.4.7 View User Reports..... | 12 |
| 1.4.8 Trail..... | 13 |
| 1.4.9 SEO analysis..... | 15 |
| 2. Phân tích trường hợp sử dụng (Use-case analysis) | 17 |
| 2.1 Phân tích kiến trúc hệ thống | 17 |
| 2.1.1 Kiến trúc mức cao của hệ thống..... | 17 |
| 2.1.2 Các đối tượng trừu tượng chính của hệ thống (Key abstractions) | 17 |
| 2.2 Thực thi trường hợp sử dụng (Use-case realizations) | 18 |
| 2.2.1 Các biểu đồ tuần tự (Sequence diagrams)..... | 18 |
| 2.2.2 Góc nhìn của các lớp trong hệ thống (Views of participating classes)..... | 23 |
| 3. Thiết kế (Use-case design)..... | 25 |
| 3.1 Xác định các thành phần thiết kế (Identify design elements) | 25 |
| 3.1.1 Xác định các lớp (Identify classes)..... | 25 |
| 3.1.2 Xác định các hệ thống con và giao diện (Identify subsystems and interfaces)..... | 26 |
| 3.1.3 Xác định các gói (Identify packages)..... | 27 |
| 3.2 Thiết kế trường hợp sử dụng (Use-case design)..... | 29 |
| 3.2.1 Thiết kế các biểu đồ tuần tự (Design sequence diagrams)..... | 29 |
| 3.2.2 Thiết kế biểu đồ lớp (Class diagrams)..... | 34 |
| 3.3 Thiết kế cơ sở dữ liệu (Database design)..... | 37 |

1. Yêu cầu (Requirements)

1.1 Đặt vấn đề (Problem statement)

1.1.1 Mô tả vấn đề

Trong thời điểm xã hội hiện nay, công nghệ và thương mại ngày càng phát triển dẫn đến sự phát triển vô cùng mạnh mẽ của thương mại điện tử. Các trang thương mại điện tử chiếm thị phần lớn. Đặc biệt, sự cạnh tranh khốc liệt về giá khi mua hàng trực tuyến thấp hơn nhiều so với mua hàng kiểu truyền thống. Thương mại điện tử đã và đang là phương thức kinh doanh của các doanh nghiệp.

Phần mô tả của sản phẩm đóng vai trò quan trọng giúp tăng tỷ lệ mua hàng. Mô tả của sản phẩm cần đáp ứng nhiều yêu cầu như là: độ dài, nội dung, cách diễn đạt,... Vì vậy để viết được một miêu tả sản phẩm tốt cần rất nhiều thời gian, bên cạnh đó khi hoàn thiện miêu tả cũng khó đánh giá được mức độ hiệu quả.

1.1.2 Giải pháp

Contify.ai sẽ có nhiệm vụ chính là viết ra những miêu tả cho sản phẩm dựa trên những thông tin cơ bản mà nhà bán hàng đưa ra sao cho đảm bảo được đủ chi tiết cần miêu tả và đảm bảo được sự hấp dẫn giúp có được tỷ lệ mua hàng cao. contify.ai sẽ liên tục đưa ra những phương án tối ưu sao cho phù hợp với xu thế của thời điểm hiện tại.

1.2 Thuật ngữ (Glossary)

Giới thiệu

Tài liệu này được sử dụng để xác định thuật ngữ cụ thể cho miền vấn đề, giải thích các thuật ngữ, có thể không quen thuộc với người đọc các mô tả trường hợp sử dụng hoặc các tài liệu dự án khác. Thông thường, tài liệu này có thể được sử dụng như một từ điển dữ liệu không chính thức, nắm bắt các định nghĩa dữ liệu để các mô tả trong trường hợp sử dụng và các tài liệu dự án khác có thể tập trung vào những gì hệ thống phải làm với thông tin.

Definitions

Thuật ngữ chứa các định nghĩa làm việc cho các khái niệm chính trong trang web Contify.ai.

Account

Một bản ghi về người dùng/quản trị viên có chứa thông tin về tên, địa chỉ email, mật khẩu, số điện thoại và tự giới thiệu tùy chọn. Mỗi tài khoản có ID người dùng duy nhất và mật khẩu, được sử dụng để xác định người dùng/quản trị viên và cấp cho họ quyền truy cập vào các bộ phận bảo mật của hệ thống

Administrator

Người có công việc quản lý hoạt động, quản lý quảng cáo cũng như người dùng. Đảm bảo không có những hoạt động bất thường và quyền lợi của người dùng.

User

Bất kỳ người nào có tài khoản đã đăng ký và mua gói tài khoản trên trang web nhưng không phải là quản trị viên. Người dùng có thể thực hiện nhiều tác vụ với tài khoản và chức năng viết mô tả được cung cấp.

Visitor

Một người quan tâm đến việc xem, tham khảo trên trang web nhưng không có tài khoản hoặc chưa đăng ký gói sản phẩm.

1.3 Thông số kỹ thuật bổ sung

Objectives

Mục đích của tài liệu này là xác định các yêu cầu của hệ thống Contify.ai. Thông số kỹ thuật bổ sung này liệt kê các yêu cầu không được nắm bắt nhanh trong trường hợp sử dụng của mô hình trường hợp sử dụng. Thông số kỹ thuật bổ sung và mô hình trường hợp sử dụng cùng nhau nắm bắt một tập hợp các yêu cầu hoàn chỉnh trên hệ thống.

References

None.

Functionality

Nhiều người dùng phải có thể thực hiện đồng thời công việc của họ.

Usability

Phần mềm phải dễ sử dụng để người dùng mới có thể học cách sử dụng hệ thống trong vòng 1 giờ.

Giao diện người dùng phải thân thiện và trực quan

Reliability

Hệ thống phải có sẵn 24 giờ một ngày, 7 ngày một tuần. Hệ thống cũng phải có thời gian chết ít hơn 5%.

Performance

Hệ thống sẽ hỗ trợ đồng thời người dùng đồng thời đối với cơ sở dữ liệu trung tâm và tối đa 100 người dùng đồng thời đối với các máy chủ cục bộ tại bất kỳ thời điểm nào. Hệ thống sẽ cung cấp quyền truy cập vào cơ sở dữ liệu với độ trễ không quá 5 giây. Hệ thống phải có khả năng hoàn thành 90% tất cả các giao dịch trong vòng 30 giây.

Supportability

Đa nền tảng, ứng dụng web

Security

Hệ thống phải ngăn người dùng đăng nhập nếu họ không có mật khẩu.

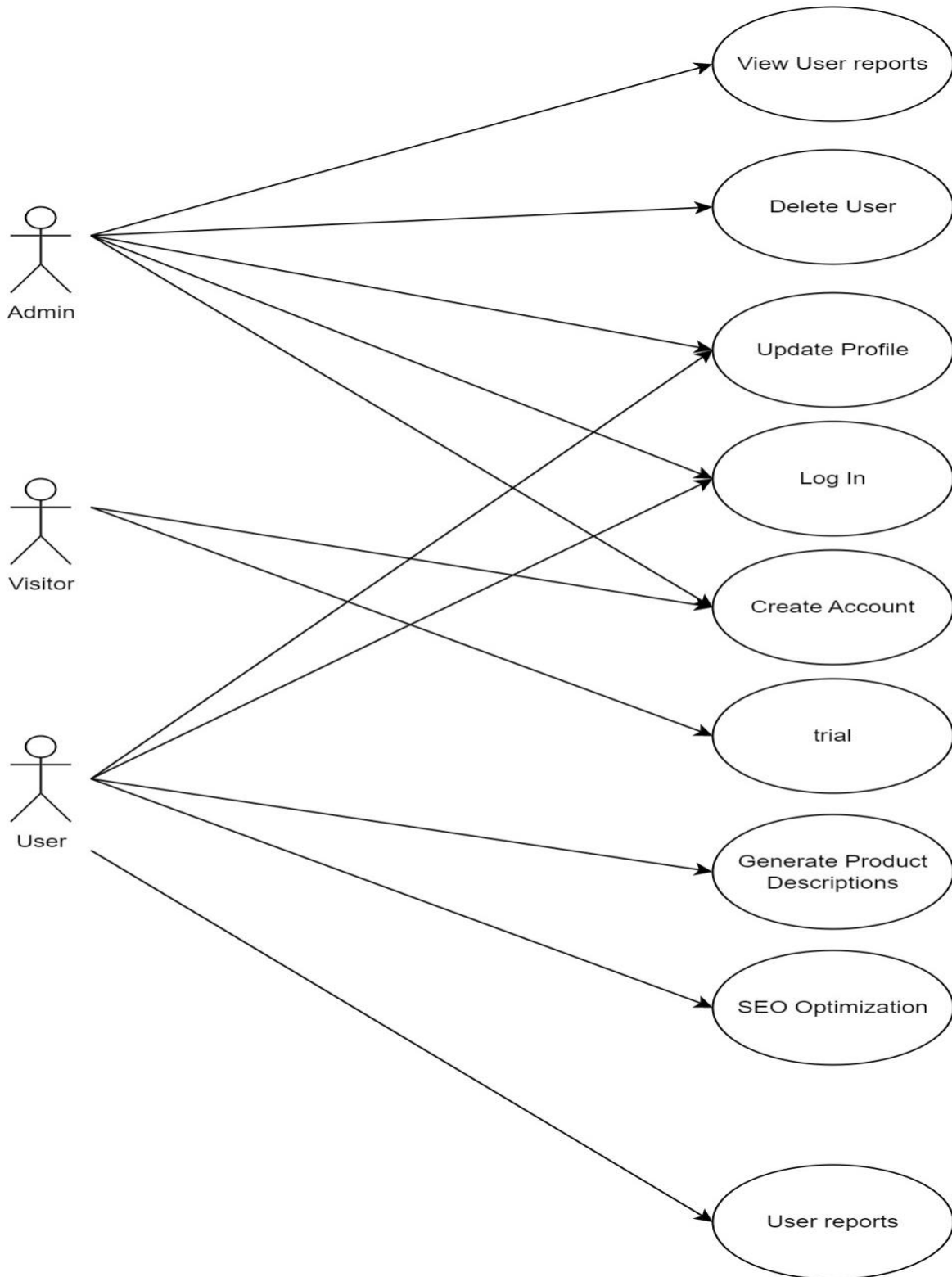
Chỉ quản trị viên mới có thể xóa tài khoản người dùng.

Chỉ có quản trị viên mới có thể phê duyệt giá cũng như những khiếu nại, thay đổi của khách hàng.

Design Constraints

Hệ thống phải cung cấp giao diện dựa trên web đáp ứng có thể sử dụng trên máy tính và thiết bị di động.

1.4 Mô hình sử dụng hệ thống (Use-case models)



Hình 1: Mô hình use-case sử dụng cho hệ thống trang web Contify.ai

1.4.1 Generate Product Descriptions

Brief Description

Người sử dụng nhập những thông tin cần thiết như tên sản phẩm, những tính năng/đặc tính của sản phẩm để thực hiện viết mô tả cho sản phẩm đó dựa trên mô hình đã được đào tạo sẵn.

Flow of Events

Basic Flow

Trường hợp này bắt đầu khi user yêu cầu thực hiện viết một mô tả mới cho một sản phẩm bất kỳ nào đó.

1. Màn hình hệ thống hiển thị các mô tả đã được viết trước đó. User có thể lựa chọn chỉnh sửa/tối ưu những mô tả đã viết hoặc lựa chọn viết mô tả cho một sản phẩm mới.
2. Nếu User chọn chỉnh sửa/tối ưu thì Sub-flow **chỉnh sửa/tối ưu mô tả** sẽ được thực hiện. Nếu User chọn viết mô tả mới thì Sub-flow **viết mô tả mới** cho một sản phẩm sẽ được thực hiện.

Chỉnh sửa/tối ưu mô tả

1. Hệ thống sẽ phân tích SEO, từ khóa của mô tả đã được viết.
2. Hệ thống phân tích từ những đối thủ khác.
3. Thêm những tính năng mới mà user yêu cầu.
4. Viết lại mô tả.

Viết mô tả mới

1. Hệ thống yêu cầu User nhập vào tiêu đề và tên của sản phẩm cần viết mô tả
2. Hệ thống sẽ phân tích SEO, từ khóa của mô tả đã được viết.
3. Hệ thống phân tích từ những đối thủ khác.
4. Thêm những tính năng mới mà user yêu cầu.
5. Viết mô tả dựa trên những thông tin tên.

Alternative Flows

Operation Cancelled

Một trong những Sub-flows, nếu Admin hoặc user lựa chọn “Hủy bỏ” trong quá trình thực hiện các Sub-flows thì các Sub-flows **Chỉnh sửa/tối ưu mô tả** và **Viết mô tả mới** không được thực hiện.

Special Requirements

None.

Pre-Conditions

User, Admin phải đăng nhập vào hệ thống.

Post-Conditions

Trong trường hợp các flow được thực hiện thành công, User/Admin nhận được mô tả sản phẩm. Nếu không thì hệ thống sẽ không thay đổi.

Extension Points

None.

1.4.2 Create Account

Brief Description

Đây là user-case thể hiện một **Visitor** trở thành một **User** của Contify.ai.

Flow of Events

User-case này được bắt đầu khi một Visitor yêu cầu tạo tài khoản và mua một gói sản phẩm được cung cấp trên Website.

1. Màn hình hệ thống sẽ hiển thị một biểu mẫu yêu cầu **Visitor** điền đầy đủ những thông tin dưới đây:
 - Unique user ID
 - Password
 - Name
 - Phone number
 - Email address
 - Brief self-introduction (optional)
2. Khi **Visitor** điền đầy đủ thông tin được yêu cầu, Hệ thống sẽ xác minh user ID là duy nhất và tất cả các trường hợp cần thiết đối với các trường thông tin còn lại. Sau đó, hệ thống thêm một tài khoản mới có thông tin được chỉ định vào cơ sở dữ liệu người dùng.

3. Hệ thống hiển thị những lựa chọn về các gói sản phẩm và dịch vụ, tính năng viết mô tả cung cấp. **Visitor** lựa chọn gói sản phẩm và thanh toán. Sau đó, hệ thống sẽ đưa thông tin tài khoản vào cơ sở dữ liệu **User**.

Alternative Flows

Thiếu thông tin

Nếu bất kỳ trường nào ở trên (ngoại trừ việc tự giới thiệu) không được điền vào, hệ thống sẽ hiển thị một thông báo lỗi. **Visitor** có thể tiếp tục thay đổi mẫu đăng ký hoặc hủy đăng ký, tại thời điểm mà user-case kết thúc.

User ID đã tồn tại

Nếu **User ID** được chỉ định đã tồn tại, hệ thống sẽ hiển thị thông báo lỗi. Khách truy cập có thể tiếp tục thay đổi mẫu đăng ký hoặc hủy đăng ký, tại thời điểm mà user-case kết thúc.

Special Requirements

None.

Pre-Conditions

None.

Post-Conditions

Nếu use-case được thực hiện thành công, một User mới được thêm vào hệ thống. Nếu không thì hệ thống sẽ không thay đổi.

Extension Points

None.

1.4.3 Delete User

Brief Description

Đây là một User-case cho phép các Admin được phép xóa bỏ User ra khỏi hệ thống.

Flow of Events

Basic Flow

User-case này sẽ bắt đầu khi Admin muốn xóa bỏ một User ra khỏi hệ thống.

1. Hệ thống yêu cầu Admin chỉ định User ID.
2. Admin nhập vào User ID đó. Hệ thống truy xuất và hiển thị thông tin của User đó lên màn hình.
3. Hệ thống yêu cầu Admin xác nhận xóa bỏ User.
4. Admin xác nhận xóa bỏ.
5. Hệ thống xóa bỏ User đó ra khỏi hệ thống.

Alternative Flows

User Not Found

Nếu User ID được chỉ định không tồn tại, hệ thống sẽ hiển thị thông báo lỗi. Sau đó, quản trị viên có thể nhập một User ID khác hoặc hủy thao tác, tại thời điểm mà user-case kết thúc.

Delete Cancelled

Nếu Admin quyết định không xóa người dùng, việc loại bỏ sẽ bị hủy và trường hợp sử dụng sẽ được khởi động lại ngay từ đầu.

Pre-Conditions

Admin phải đăng nhập vào trong hệ thống trước khi thực hiện Use-case này. Chức năng chỉ dành cho Admin.

Post-Conditions

Nếu Use-case được thực hiện thành công, User và những mô tả của User đó sẽ được xóa khỏi hệ thống. Nếu không thì hệ thống sẽ không có gì thay đổi.

Extension Points

None.

1.4.4 Log In

Brief Description

Đây là Use-case miêu tả làm cách nào để User hoặc Admin có thể vào được Website với thông tin được lưu trong cơ sở dữ liệu.

Flow of Events

Basic Flow

Use-Case này sẽ bắt đầu khi một ai đó muốn vào Website với thông tin được lưu trong cơ sở dữ liệu.

1. Hệ thống yêu cầu nhập vào User ID và mật khẩu.
2. Hệ thống xác minh ID và một khẩu được nhập vào đã có trong hệ thống chưa. Nếu có, vào Website.

Alternative Flows

Invalid Information

Nếu trong *Basic Flow*, người này nhập sai User ID và/hoặc mật khẩu, hệ thống sẽ hiển thị lên màn hình dòng tin nhắn cảnh báo. Người này có thể chọn thực hiện lại hoặc hủy bỏ, tại thời điểm mà Use-case kết thúc.

Missing Required Information

Nếu bất kỳ trường nào trong biểu mẫu bị bỏ trống, hệ thống sẽ hiển thị thông báo lỗi. Diễn viên có thể tiếp tục sửa đổi biểu mẫu hoặc hủy bỏ thao tác, tại thời điểm mà Use-case kết thúc.

Special Requirements

None.

Pre-Conditions

Hệ thống ở trạng thái đăng nhập và có màn hình đăng nhập được hiển thị.

Post-Conditions

Nếu trường hợp sử dụng thành công, người này được đăng nhập vào hệ thống. nếu không thì hệ thống không có gì thay đổi.

Extension Points

None.

1.4.5 Update profile

Brief Description

Use-case này miêu tả cách mà User/Admin cập nhật lại thông tin của tài khoản đã có trong hệ thống

Flow of Events

Basic Flow

Use-case này bắt đầu khi User/Admin yêu cầu cập nhật lại thông tin tài khoản của User/Admin.

1. Màn hình hệ thống hiển thị những trường thông tin dưới đây và hỏi User/Admin muốn thay đổi gì:
 - Password
 - Name
 - Phone number
 - Email address
 - Brief self-introduction (optional)
2. Khi User cung cấp những thông tin được yêu cầu, hệ thống xác minh lại tất cả những yêu cầu từ các trường thông tin được chỉ định là được cập nhật và cập nhật lại những thông tin đó trong cơ sở dữ liệu.

Special Requirements

None.

Pre-Conditions

User/Admin cần phải đăng nhập vào hệ thống trước khi thay đổi, cập nhật thông tin.

Phải có thông tin trong cơ sở dữ liệu

Post-Conditions

Nếu use-case thực hiện thành công, thông tin của tài khoản được cập nhật. Nếu không thì hệ thống không có gì thay đổi.

Extension Points

None.

1.4.6 User Reports

Brief Description

Use-case này miêu tả cách mà User có thể gửi khiếu nại/báo cáo về những hoạt động không đúng, không hiệu quả của hệ thống về cho Admin.

Flow of Events

Basic Flow

Use case này bắt đầu khi User muốn thực hiện gửi những khiếu nại/báo cáo của mình về cho Admin.

1. Màn hình hiển thị danh sách những vấn đề thường được khiếu nại/báo cáo đính kèm những phương án giải quyết và phần điền khiếu nại/báo cáo mới không có trong danh sách.
2. Nếu User lựa chọn những vấn đề đã có phương án giải quyết, hệ thống sẽ hiển thị phương án giải quyết lên màn hình. Nếu User gửi báo cáo/khiếu nại vấn đề mới, hệ thống sẽ gửi lên cho Admin và hiển thị thông báo vấn đề đã được gửi lên cho Admin.

Alternative Flows

Operation Cancelled

Một trong những Sub-flows, nếu Admin hoặc user lựa chọn “Hủy bỏ” trong quá trình thực hiện các Sub-flows thì các Sub-flows không được thực hiện.

Special Requirements

None.

Pre-Conditions

User cần phải đăng nhập vào hệ thống trước khi gửi báo cáo/khiếu nại về cho Admin của hệ thống.

Post-Conditions

Nếu use-case được thực hiện thành công, vấn đề báo cáo/khiếu nại sẽ được gửi đến cho Admin của hệ thống. Nếu không thì hệ thống không có gì thay đổi.

Extension Points

None.

1.4.7 View User Reports

Brief Description

Use-case này miêu tả cách mà Admin có thể nhìn thấy những khiếu nại/báo cáo của User có thể gửi khiếu nại/báo cáo về những hoạt động không đúng, không hiệu quả của hệ thống về cho Admin.

Flow of Events

Basic Flow

Use case này bắt đầu khi Admin muốn nhìn thấy những khiếu nại của User gửi những khiếu nại/báo cáo của mình về cho Admin.

1. Hệ thống sẽ hiện thị những khiếu nại/báo cáo mà User gửi lên đã nằm trong cơ sở dữ liệu.
2. Admin có thể sửa, xóa khiếu nại/báo cáo cho nhiều mục đích khác nhau. Admin lựa chọn gửi thông báo cho User để thông báo về vấn đề được khiếu nại/báo cáo nếu cần.

Alternative Flows

Operation Cancelled

Một trong những Sub-flows, nếu Admin hoặc user lựa chọn “Hủy bỏ” trong quá trình thực hiện các Sub-flows thì các Sub-flows không được thực hiện.

Special Requirements

None.

Pre-Conditions

Admin cần đăng nhập vào hệ thống.

Post-Conditions

Nếu use-case được thực hiện thành công, Admin có thể nhìn thấy được những khiếu nại mà User gửi lên. Nếu không thì hệ thống không có gì thay đổi.

Extension Points

None.

1.4.8 Trail

Brief Description

Use-case này miêu tả cách mà Visitor có thể sử dụng thử viết mô tả một sản phẩm bằng hệ thống của Contify.ai.

Flow of Events

Basic Flow

Use case này bắt đầu khi Visitor yêu cầu dùng thử chức năng viết mô tả sản phẩm của hệ thống.

1. Màn hình hệ thống sẽ hiển thị một biểu mẫu yêu cầu **Visitor** điền đầy đủ những thông tin dưới đây:
 - Unique user ID
 - Password
 - Name
 - Phone number
 - Email address
 - Brief self-introduction (optional)
2. Khi **Visitor** điền đầy đủ thông tin được yêu cầu, Hệ thống sẽ xác minh user ID là duy nhất và tất cả các trường hợp cần thiết đối với các trường thông tin còn lại. Sau đó, hệ thống thêm một tài khoản mới có thông tin được chỉ định vào cơ sở dữ liệu người dùng.
3. Visitor chọn vào phần Trail được hiển thị trên màn hình của hệ thống. Hệ thống đưa đến phần **Viết mô tả mới**. Hệ thống đưa ra cảnh báo và giới hạn các Visitor chỉ có thể sử dụng chức năng này 3 lần.

Viết mô tả mới

1. Hệ thống yêu cầu User nhập vào tiêu đề và tên của sản phẩm cần viết mô tả
2. Hệ thống sẽ phân tích SEO, từ khóa của mô tả đã được viết.
3. Hệ thống phân tích từ những đối thủ khác.
4. Thêm những tính năng mới mà user yêu cầu.
5. Viết mô tả dựa trên những thông tin trên.

Alternative Flows

Operation Cancelled

Một trong những Sub-flows, nếu Admin hoặc user lựa chọn “Hủy bỏ” trong quá trình thực hiện các Sub-flows thì các Sub-flows không được thực hiện.

Thiếu thông tin

Nếu bất kỳ trường nào ở trên (ngoại trừ việc tự giới thiệu) không được điền vào, hệ thống sẽ hiển thị một thông báo lỗi. **Visitor** có thể tiếp tục thay đổi mẫu đăng ký hoặc hủy đăng ký, tại thời điểm mà user-case kết thúc.

User ID đã tồn tại

Nếu **User ID** được chỉ định đã tồn tại, hệ thống sẽ hiển thị thông báo lỗi. Khách truy cập có thể tiếp tục thay đổi mẫu đăng ký hoặc hủy đăng ký, tại thời điểm mà user-case kết thúc.

Special Requirements

None.

Pre-Conditions

Vistor cần tạo tài khoản và đăng nhập vào hệ thống

Post-Conditions

Nếu use-case được thực hiện thành công, Visitor nhận được mô tả sản phẩm. Nếu không thì hệ thống không có gì thay đổi.

Extension Points

None.

1.4.9 SEO analysis

Brief Description

Use-case này miêu tả cách mà User có thể phân tích SEO bằng hệ thống của Contify.ai.

Flow of Events

Basic Flow

Use case này bắt đầu khi User muốn phân tích SEO cho mô tả có sẵn bằng hệ thống Website.

1. Hệ thống hiển thị những mô tả đã được viết. Đưa ra lựa chọn tối ưu và phân tích SEO cho từng mô tả.
2. Khi User lựa chọn phân tích, hệ thống sẽ phân tích mô tả dựa trên các thông số của SEO Onpage, SEO Offpage và các kỹ thuật SEO.
3. Hệ thống hiển thị độ chuẩn, độ hấp dẫn của mô tả trên thang điểm 10 và đưa ra những khuyến nghị để cải thiện mô tả.

SEO Onpage

Sẽ bao gồm tất cả nội dung được thực hiện và vận hành trên trang Web như: Nội dung, cấu trúc, thẻ meta, hình ảnh, từ khóa, HTML, tiêu đề, mô tả, UX/UI, tên miền,... và một số vấn đề liên quan khác.

SEO Offpage

Hệ thống sẽ phân tích yếu tố liên quan đến các trang website bên ngoài có link trở về website cần phân tích.

Kỹ thuật SEO

Hệ thống phân tích các yếu tố khía cạnh trên Onpage liên quan đến cấu trúc, hoạt động trang web. Thu thập thông tin.

Alternative Flows

Operation Cancelled

Một trong những Sub-flows, nếu Admin hoặc user lựa chọn “Hủy bỏ” trong quá trình thực hiện các Sub-flows thì các Sub-flows không được thực hiện.

Thiếu thông tin

Nếu bất kỳ trường nào ở trên (ngoại trừ việc tự giới thiệu) không được điền vào, hệ thống sẽ hiển thị một thông báo lỗi. **Visitor** có thể tiếp tục thay đổi mẫu đăng ký hoặc hủy đăng ký, tại thời điểm mà user-case kết thúc.

User ID đã tồn tại

Nếu **User ID** được chỉ định đã tồn tại, hệ thống sẽ hiển thị thông báo lỗi. Khách truy cập có thể tiếp tục thay đổi mẫu đăng ký hoặc hủy đăng ký, tại thời điểm mà user-case kết thúc.

Special Requirements

None.

Pre-Conditions

User cần đăng nhập vào hệ thống.

Post-Conditions

Nếu use-case được thực hiện thành công, User nhận được những đánh giá về SEO của mô tả được hệ thống viết ra. Nếu không thì hệ thống không có gì thay đổi.

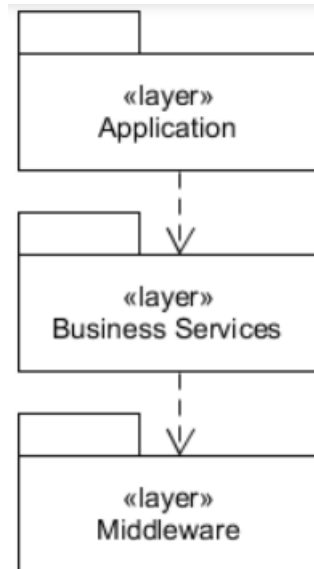
Extension Points

None.

2. Phân tích trường hợp sử dụng (Use-case analysis)

2.1 Phân tích kiến trúc hệ thống

2.1.1 Kiến trúc mức cao của hệ thống



Hình 2: Phương pháp phân lớp

Hình trên mô tả tổ chức cấp cao của hệ thống phần mềm. Hệ thống bao gồm ba lớp:

- Lớp *Application* chứa các yếu tố thiết kế dành riêng cho từng trường hợp sử dụng của hệ thống.
- Lớp *Business Services* gói gọn một số trừu tượng và dịch vụ chính chung cho tất cả các trường hợp sử dụng. Nó có thể truy cập từ lớp *Application*.
- Lớp *Middleware* cung cấp các dịch vụ để cho phép truyền thông và quản lý dữ liệu trên các hệ thống phân tán.

2.1.2 Các đối tượng trừu tượng chính của hệ thống (Key abstractions)



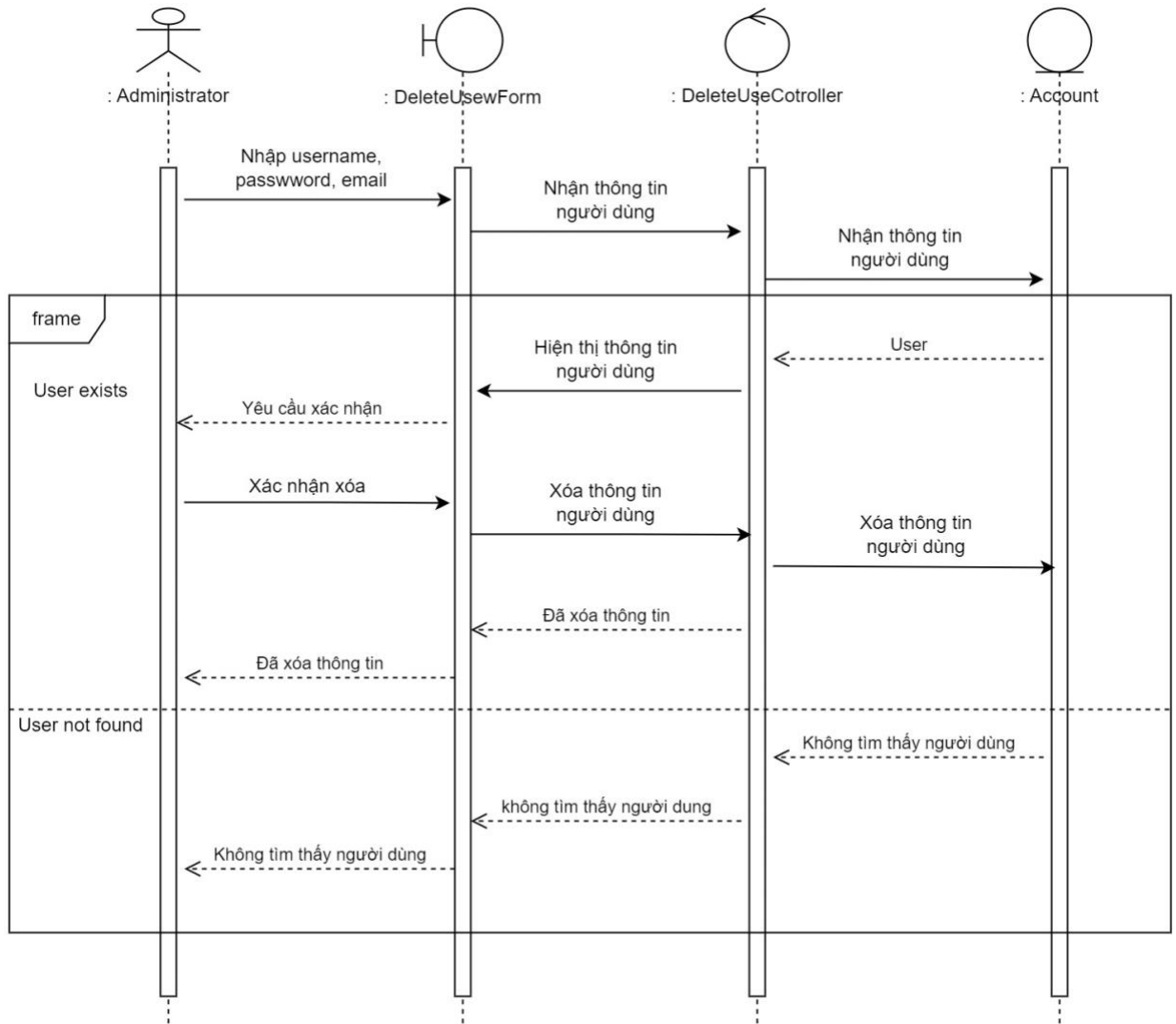
Hình 3: Tóm tắt chính được sử dụng trong ứng dụng

Account: Một bản ghi về User/Admin. Mỗi tài khoản có ID người dùng duy nhất và mật khẩu, được sử dụng để xác định User/Admin và cấp cho họ quyền truy cập vào các bộ phận bảo mật của hệ thống.

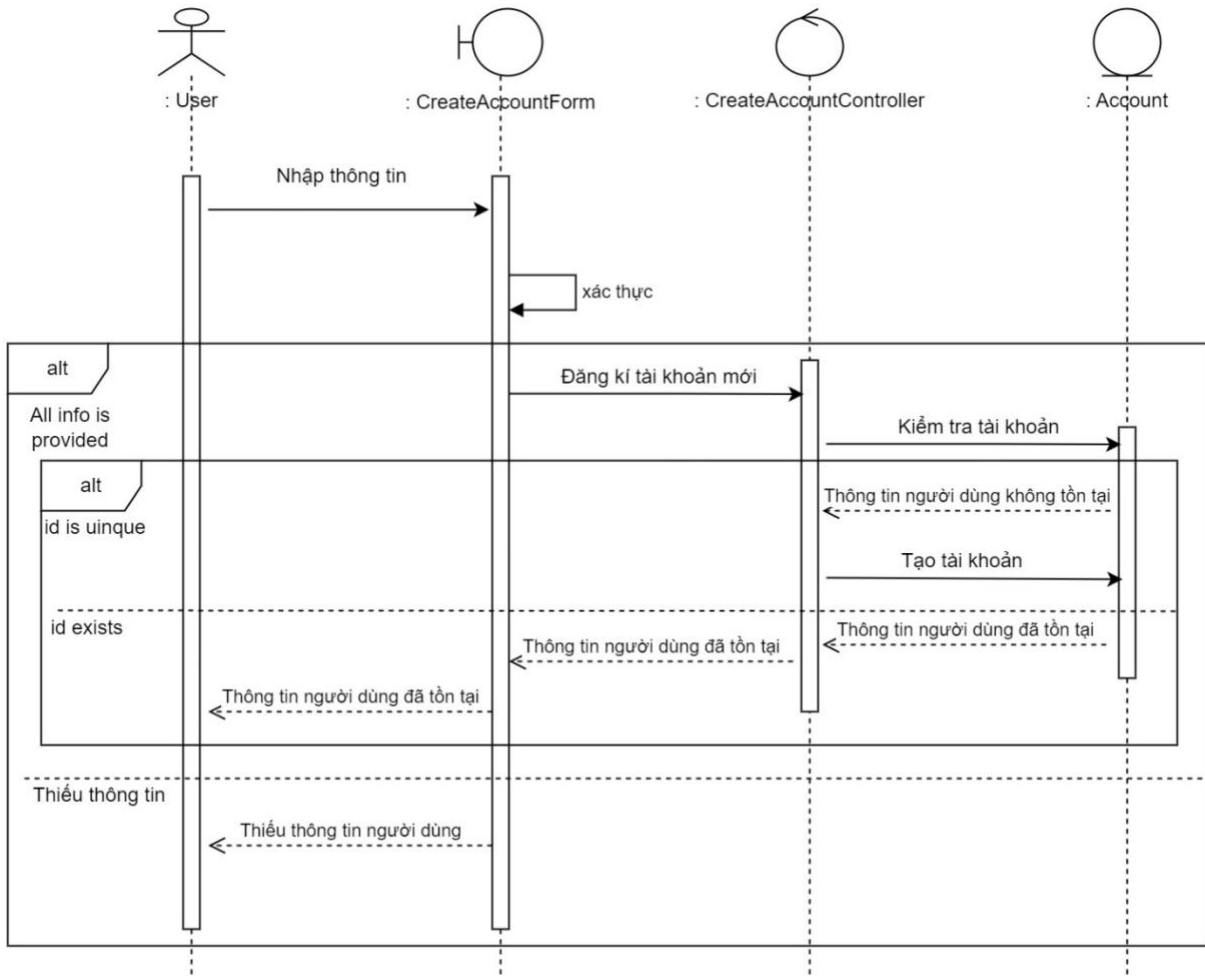
Product Description: Mô tả của sản phẩm được viết từ những thông tin được cung cấp như tiêu đề, từ khóa, chức năng,... Mô tả được tối ưu dựa trên phân tích SEO.

2.2 Thực thi trường hợp sử dụng (Use-case realizations)

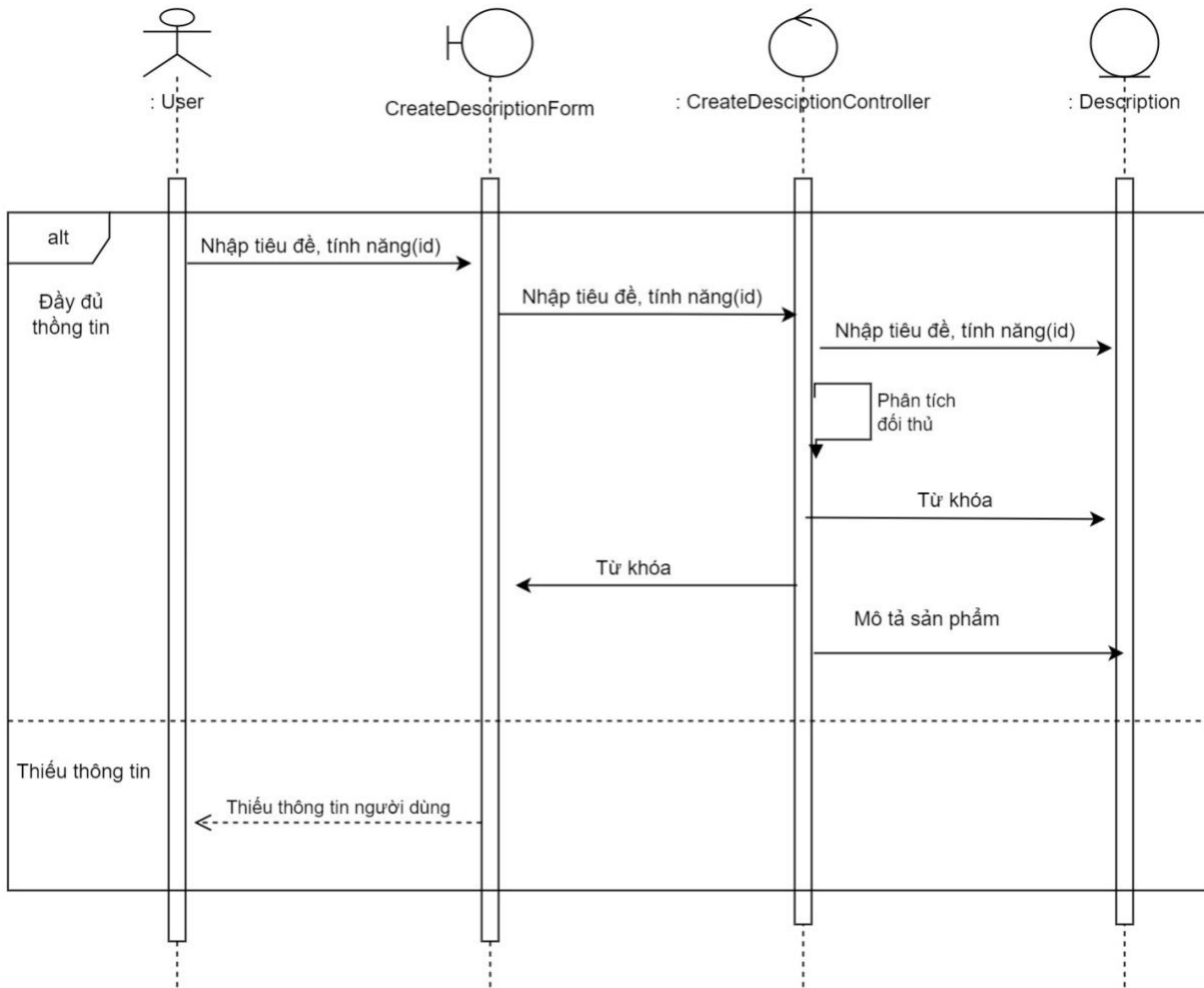
2.2.1 Các biểu đồ tuần tự (Sequence diagrams)



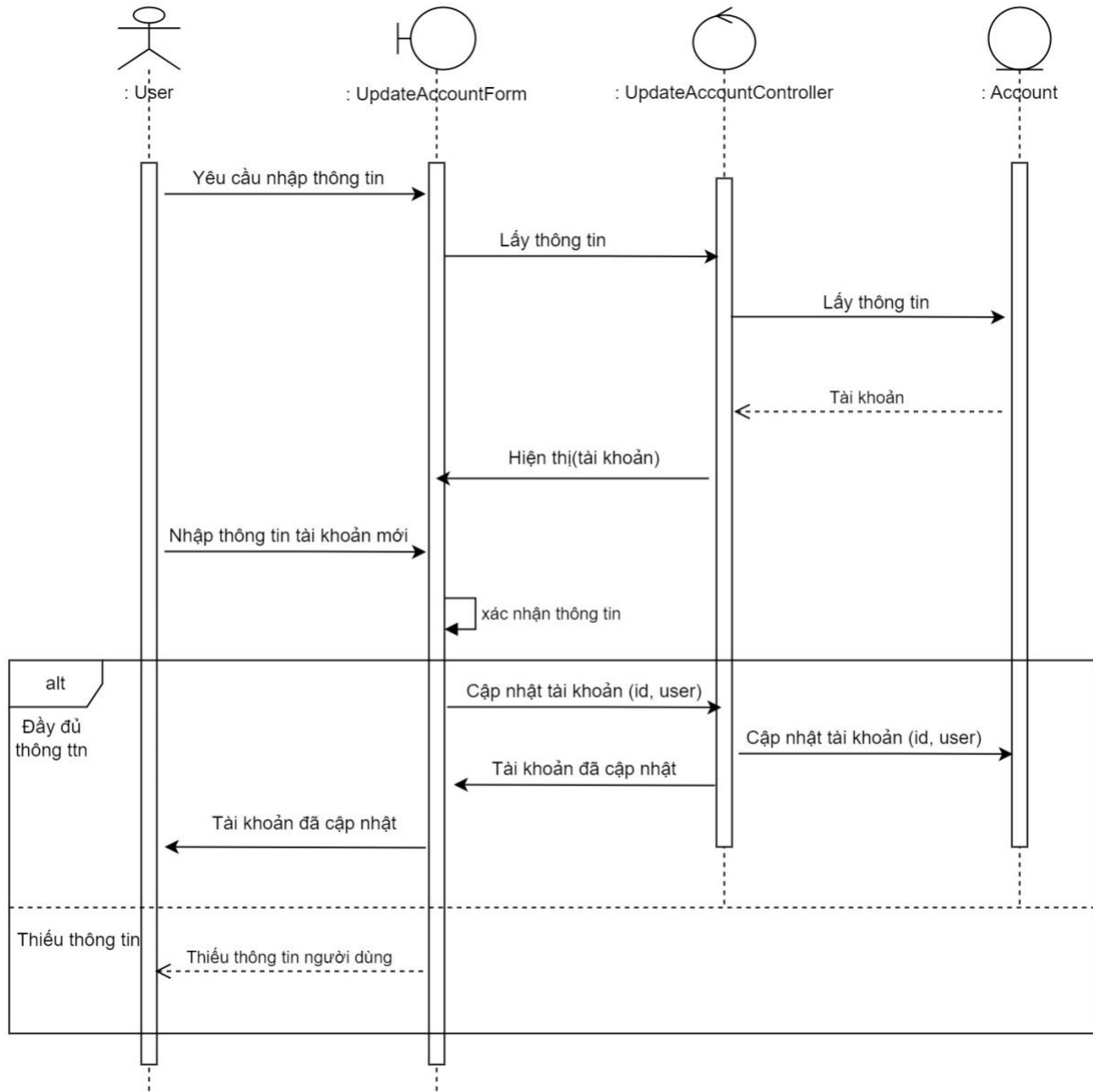
Hình 4: Design sequence diagram for the Delete user use case



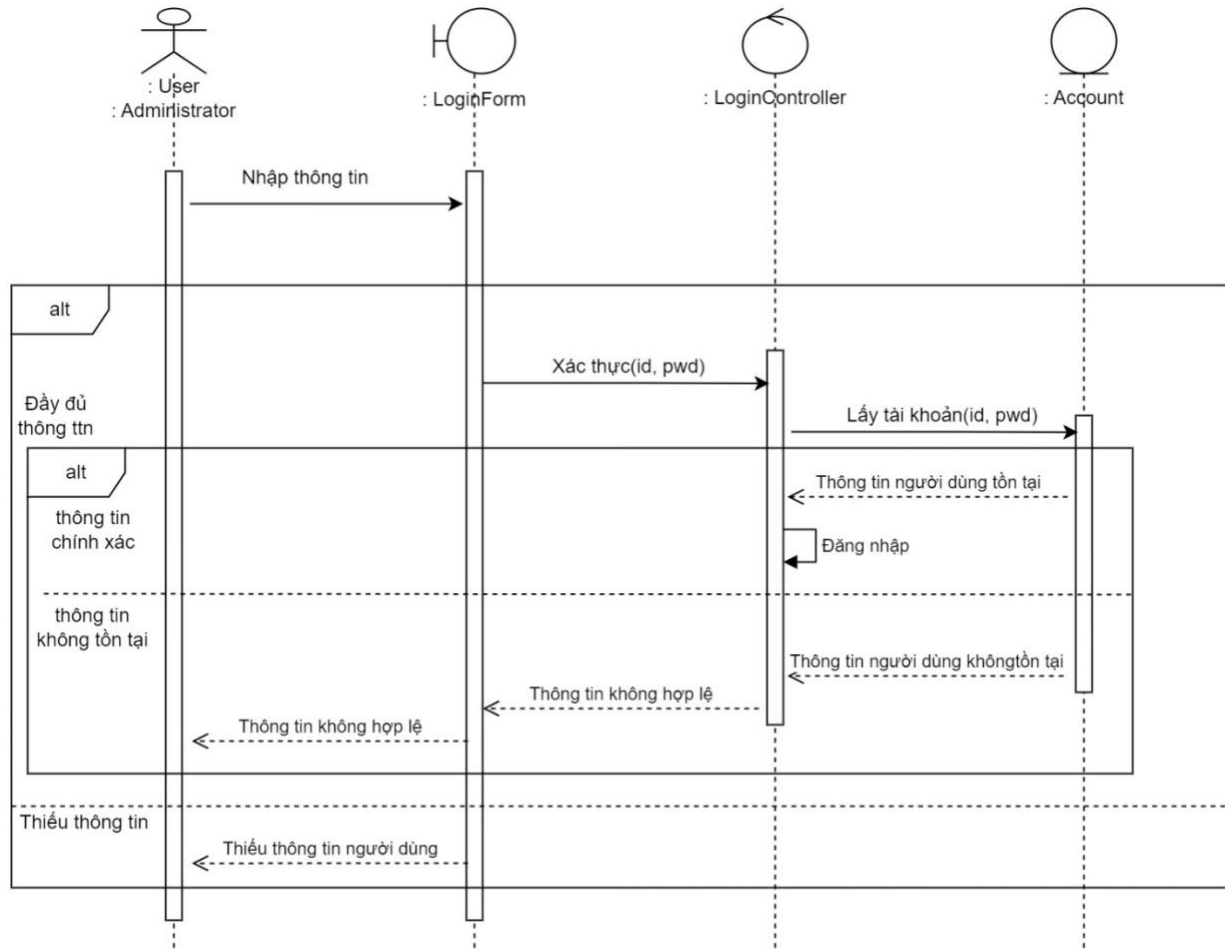
Hình 5: Design sequence diagram for the Create user use case



Hình 6: Design sequence diagram for the Create Description

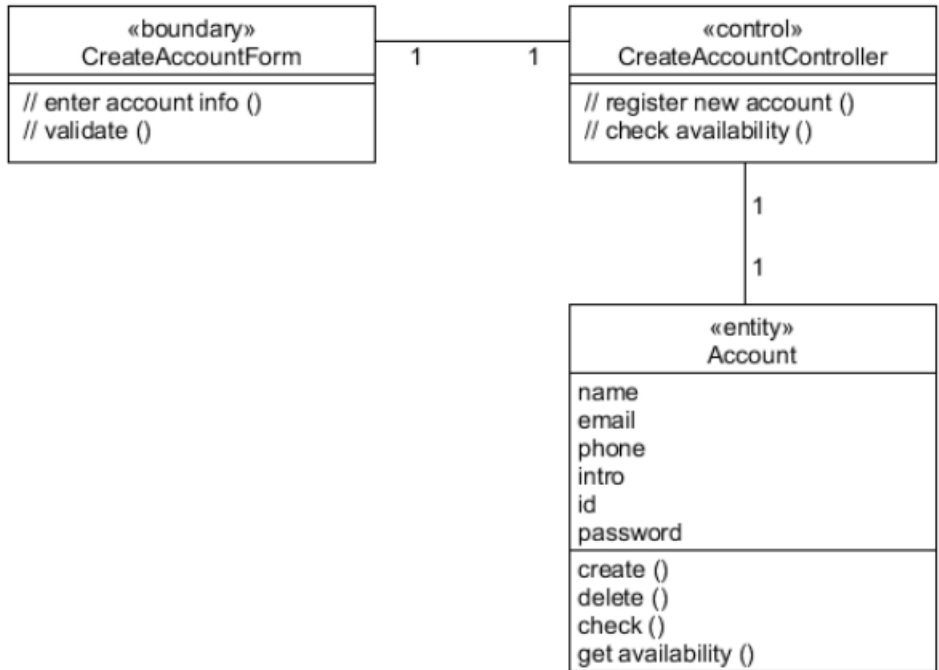


Hình 7: Design sequence diagram for the Create Description

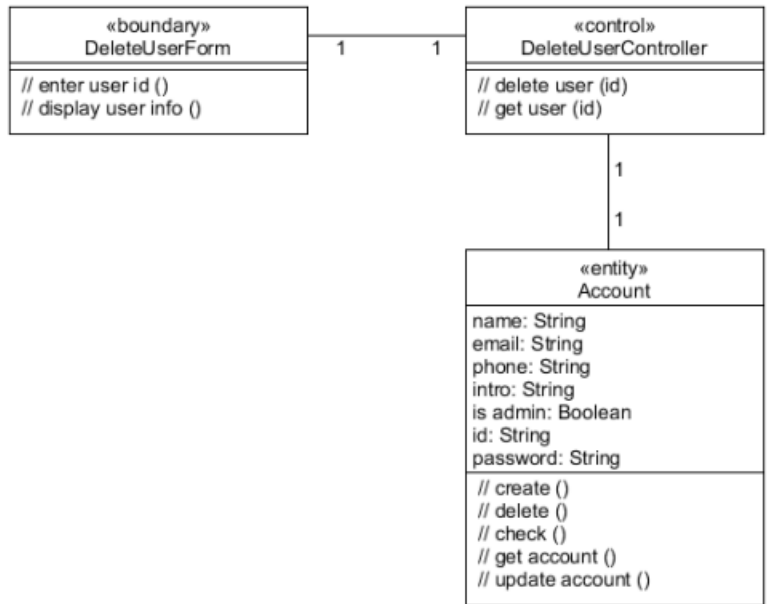


Hình 8: Design sequence diagram for the Login use case

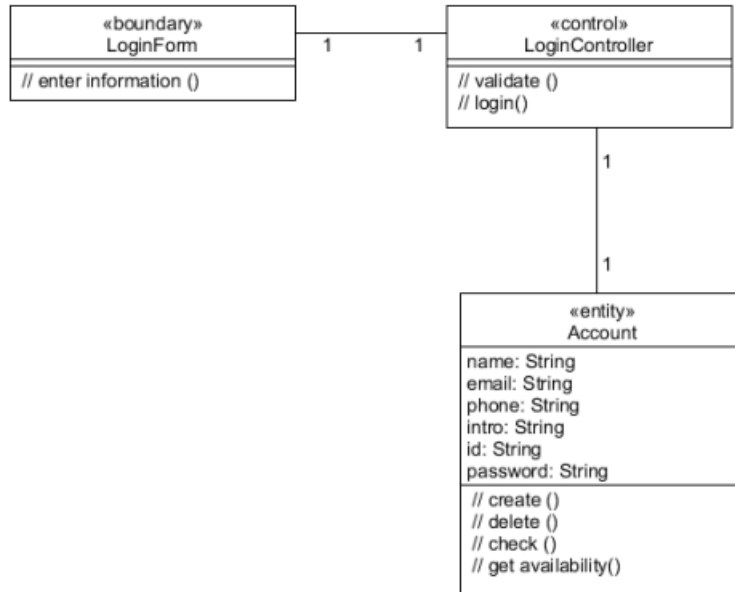
2.2.2 Góc nhìn của các lớp trong hệ thống (Views of participating classes)



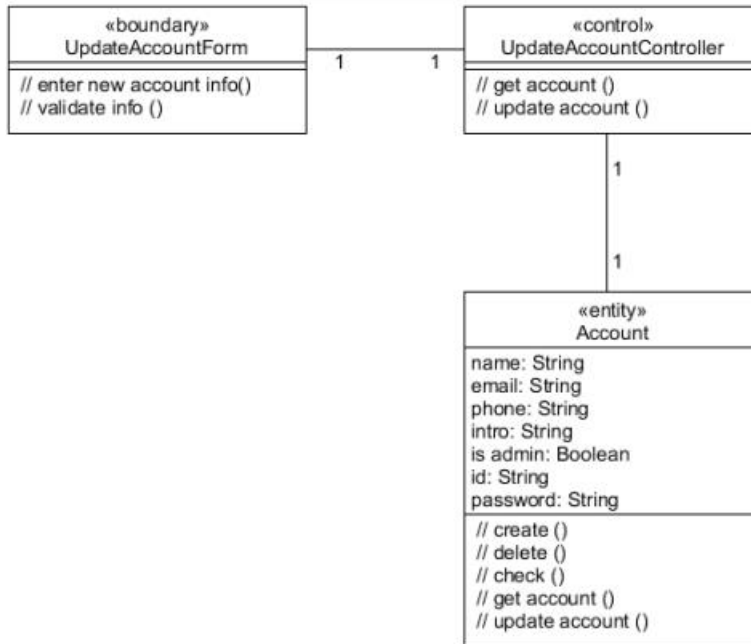
Hình 9: VOPC for the Create Account use case



Hình 10: VOPC for the Delete User use case



Hình 11: VOPC for the Log In use case



Hình 12: VOPC for the Update Account use case

3. Thiết kế (Use-case design)

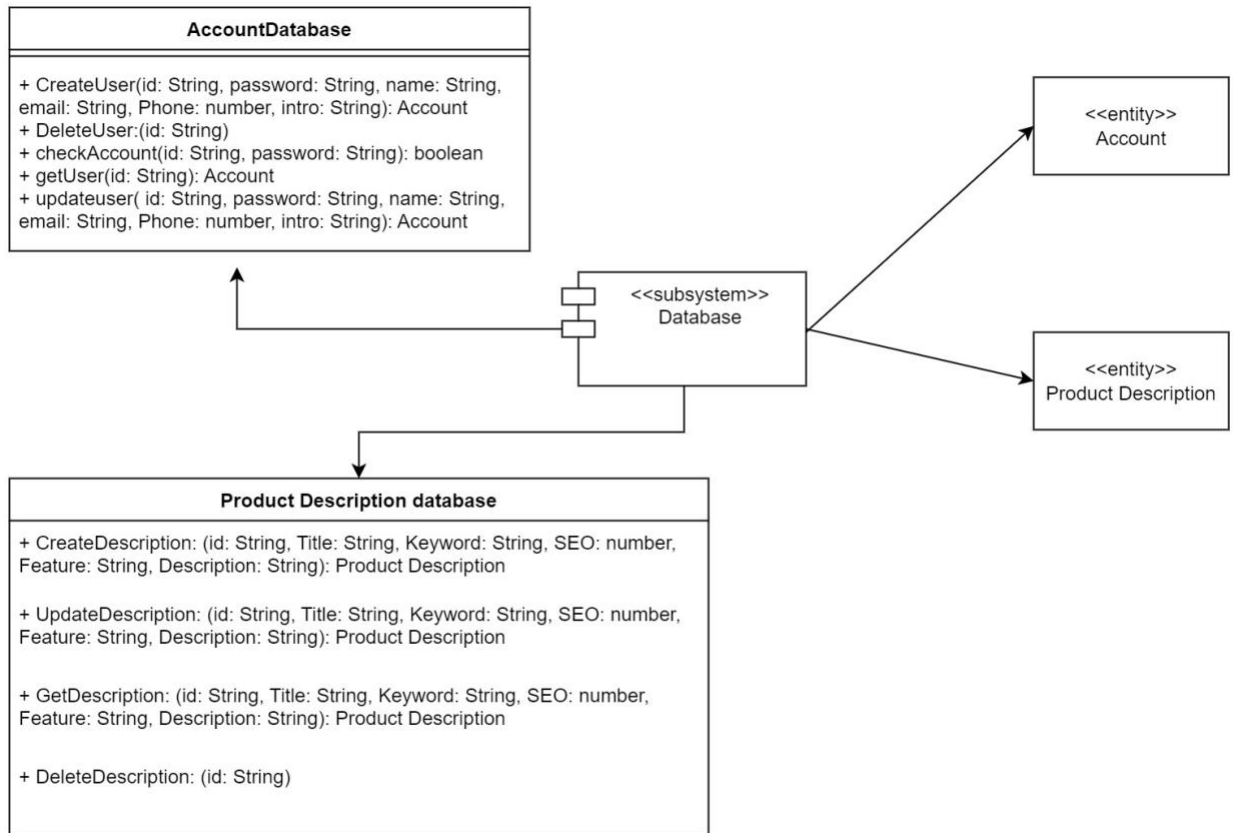
3.1 Xác định các thành phần thiết kế (Identify design elements)

3.1.1 Xác định các lớp (Identify classes)

| Analysis class | Design element |
|-------------------------|--------------------------------|
| Account | Account, Database subsystem |
| ProductDescription | Model, Database subsystem |
| CreateAccountController | Map directly to design classes |
| SearchController | |
| UpdateAccountController | |
| CreateAccountForm | |
| LoginForm | |
| ProcessReportsForm | |
| SearchForm | |
| UpdateAccountForm | |

3.1.2 Xác định các hệ thống con và giao diện (Identify subsystems and interfaces)

Hệ thống con cơ sở dữ liệu cung cấp hỗ trợ cho cơ sở dữ liệu quan hệ được viết bằng ngôn ngữ SQL. Hệ thống con được thiết kế như sau:

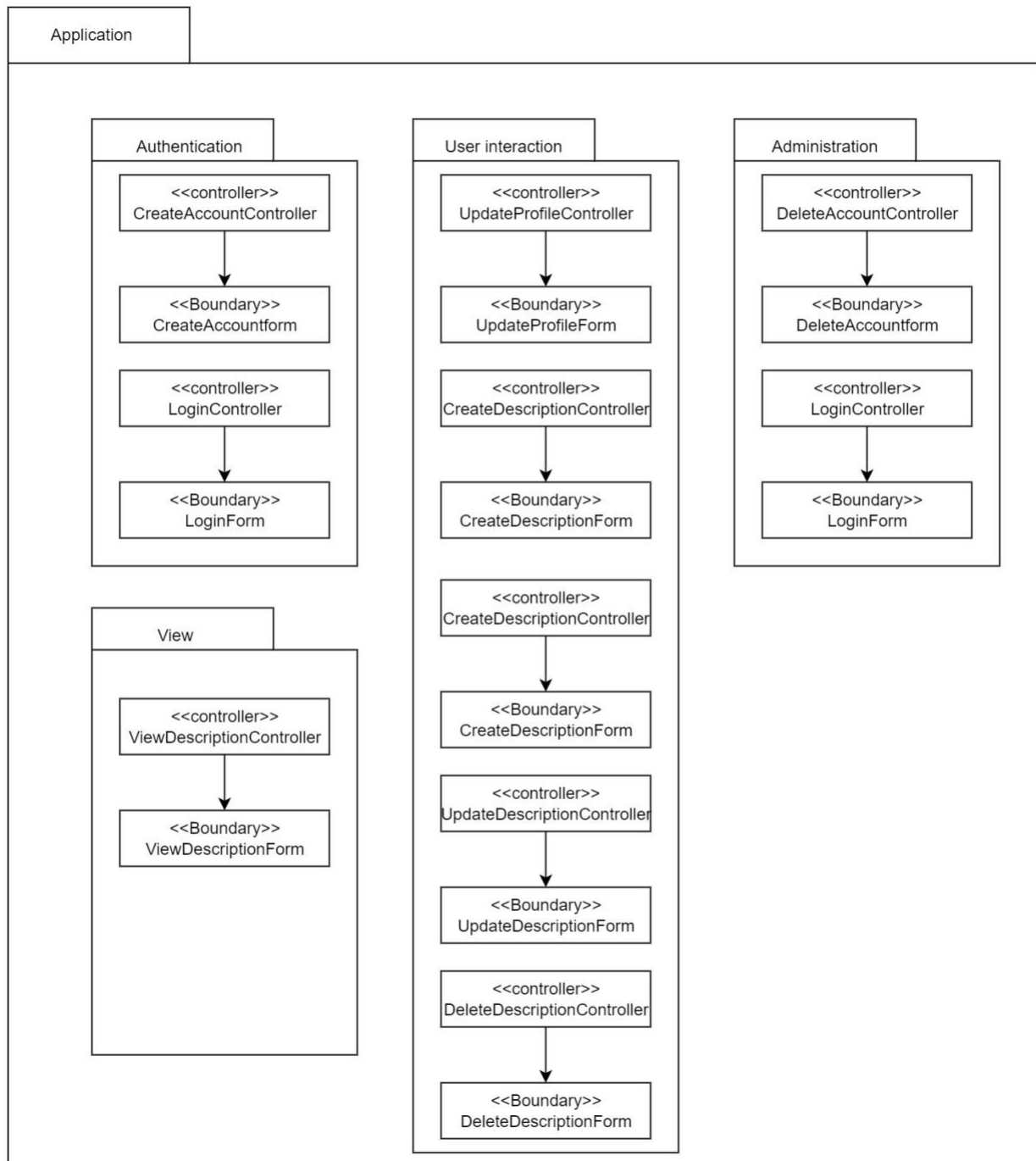


Hình 13: The Database subsystem and its interfaces

3.1.3 Xác định các gói (Identify packages)

Mỗi lớp trong phân tích tương ứng với một gói cấp cao trong hệ thống.

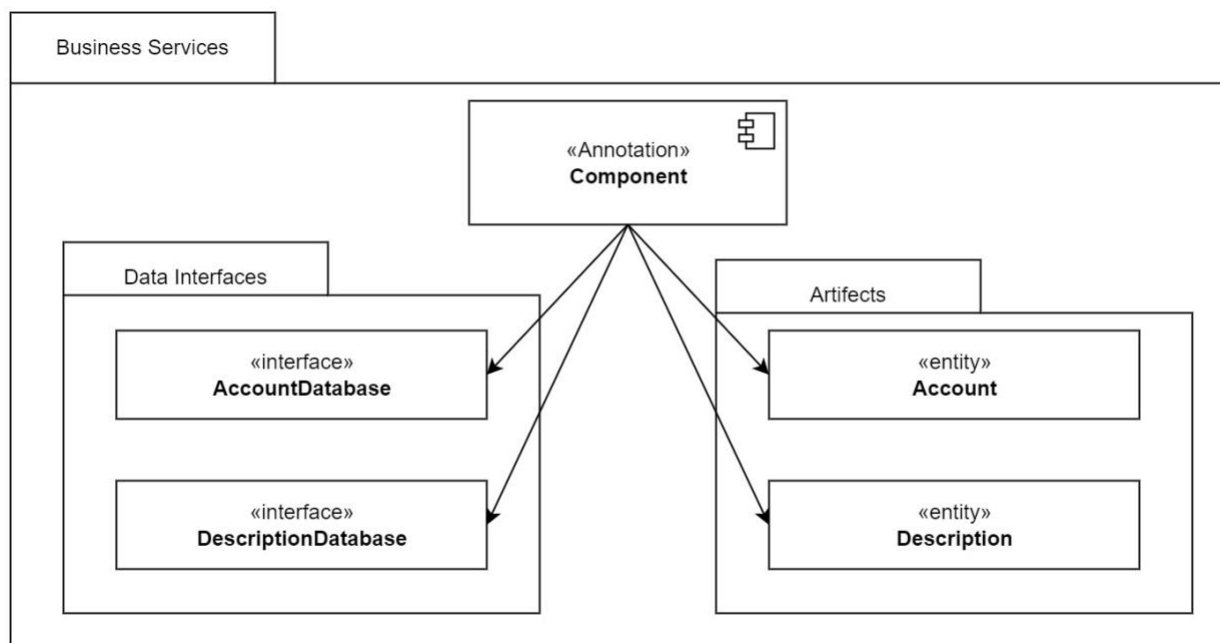
The Application package



Gói Application chứa các lớp ranh giới và điều khiển, có trong ứng dụng máy khách. Nó được chia thành bốn gói phụ, mỗi gói chịu trách nhiệm cho một phần khác nhau của ứng dụng:

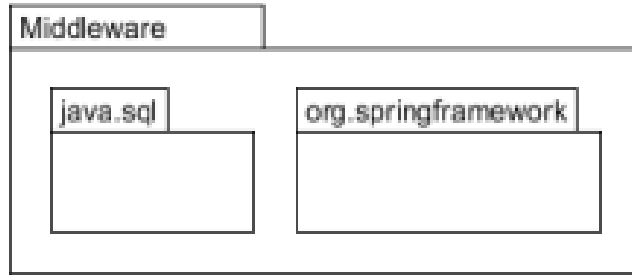
- Authentication: chứa Sub-class của tạo tài khoản và đăng nhập
- User interaction: chứa Sub-class liên quan đến việc viết mô tả sản phẩm như Create, update, delete.
- View: chứa Sub-class thực hiện nhiệm vụ hiển thị mô tả.
- Administration: chứa Sub-class thực hiện nhiệm vụ quản lý User.

The Business Services package



Gói *Business Services* chứa hệ thống cơ sở dữ liệu và các giao diện của nó, cũng như các lớp thực thể. Những yếu tố này là phổ biến cho tất cả các trường hợp sử dụng.

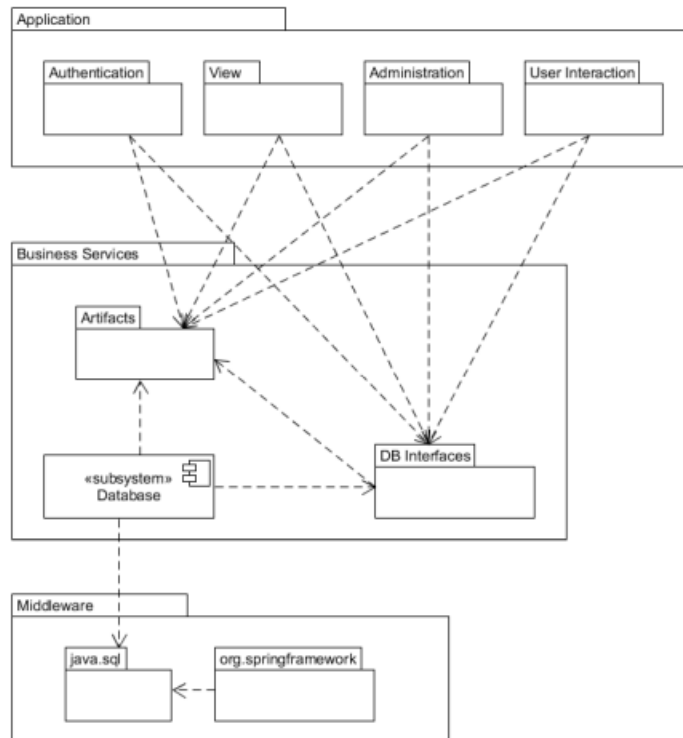
The Middleware package



Gói *Middleware Services* bao gồm các gói Java's SQL package, cung cấp cơ sở dữ liệu.

Packages and their dependencies

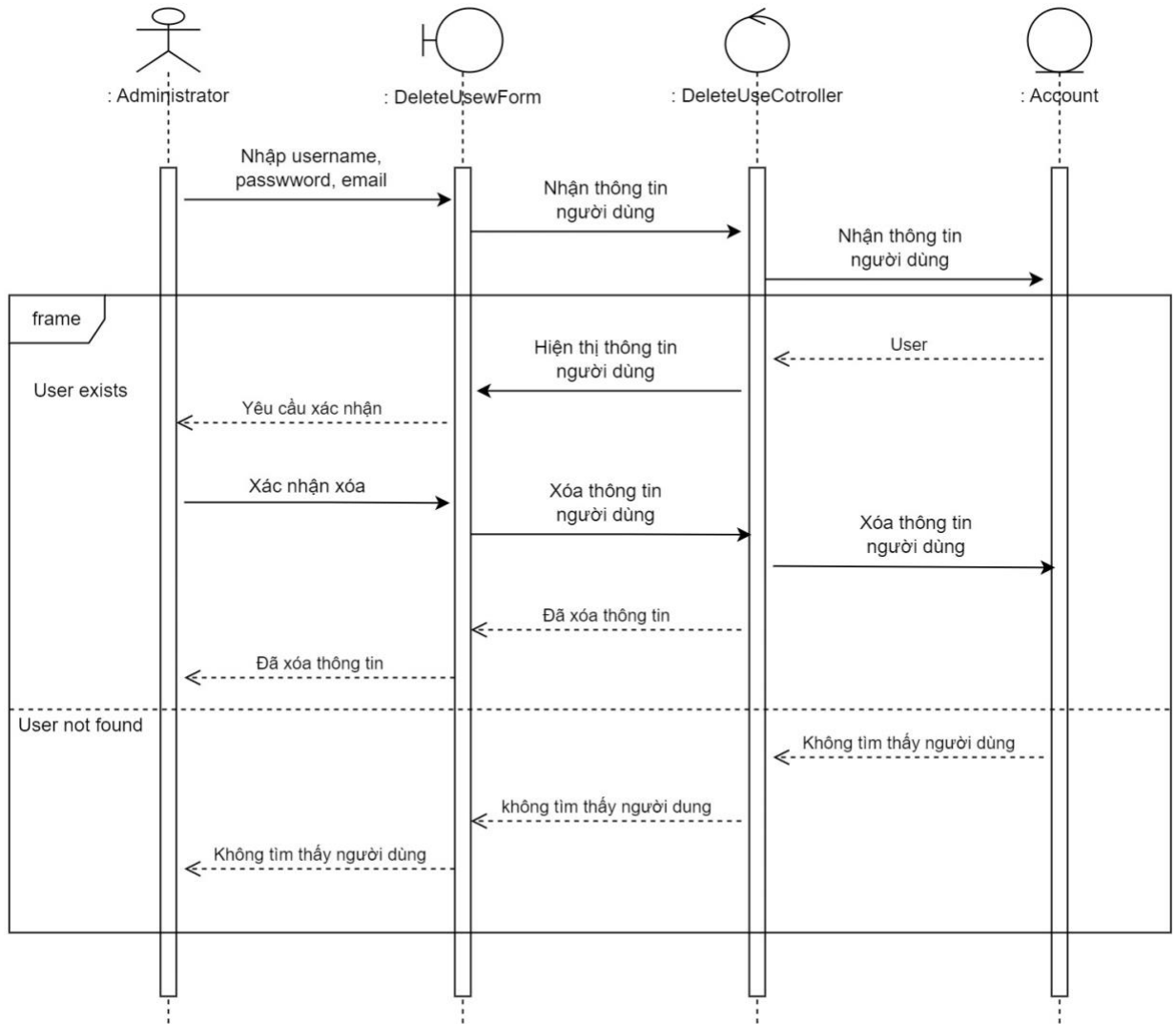
Như đã nêu, gói *Application* vào gói *Business Services*, từ đó phụ thuộc vào gói *Middleware*.



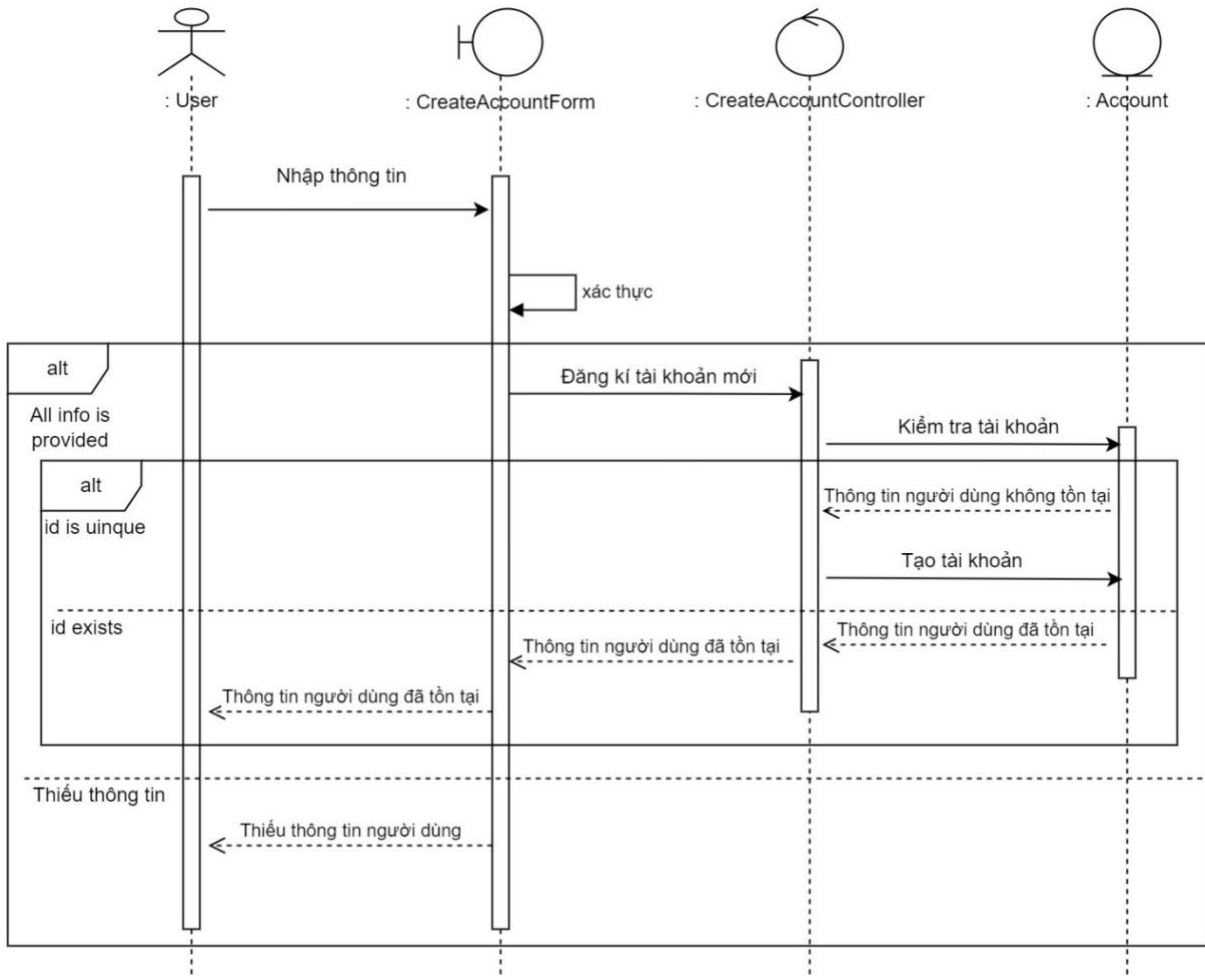
3.2 Thiết kế trường hợp sử dụng (Use-case design)

3.2.1 Thiết kế các biểu đồ tuần tự (Design sequence diagrams)

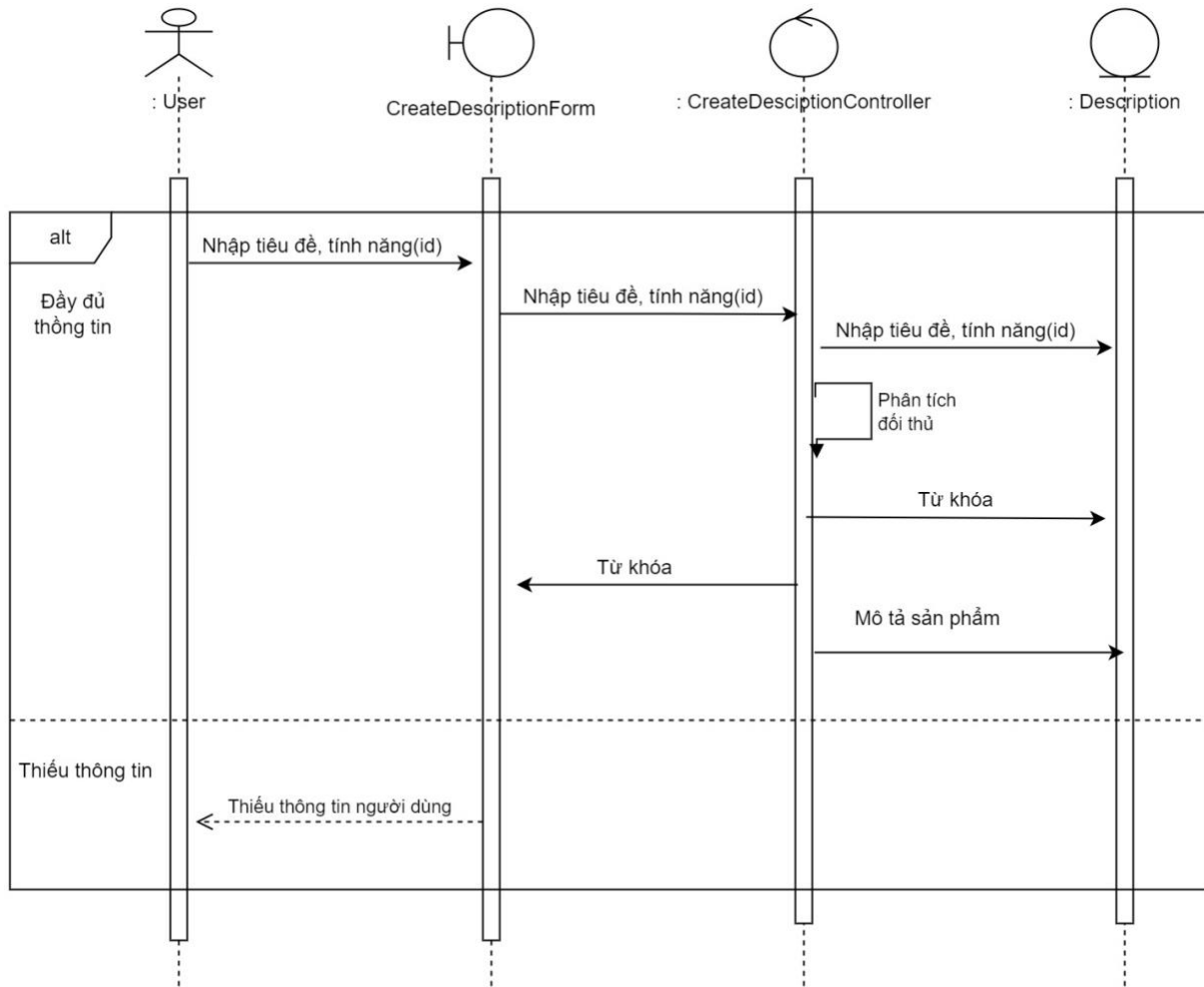
3.2.2 Các biểu đồ tuần tự (Sequence diagrams)



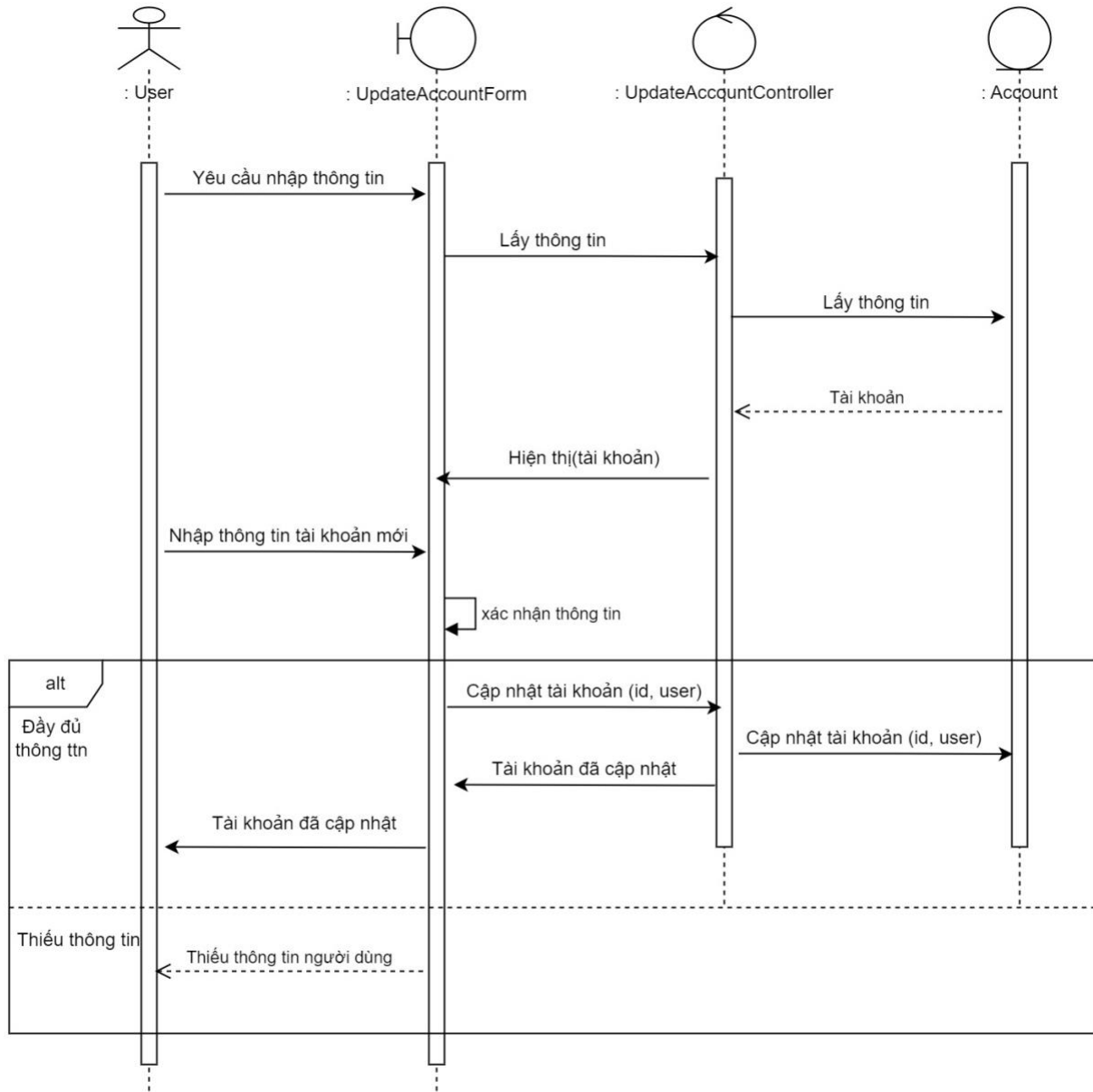
Hình 14: Design sequence diagram for the Delete user use case



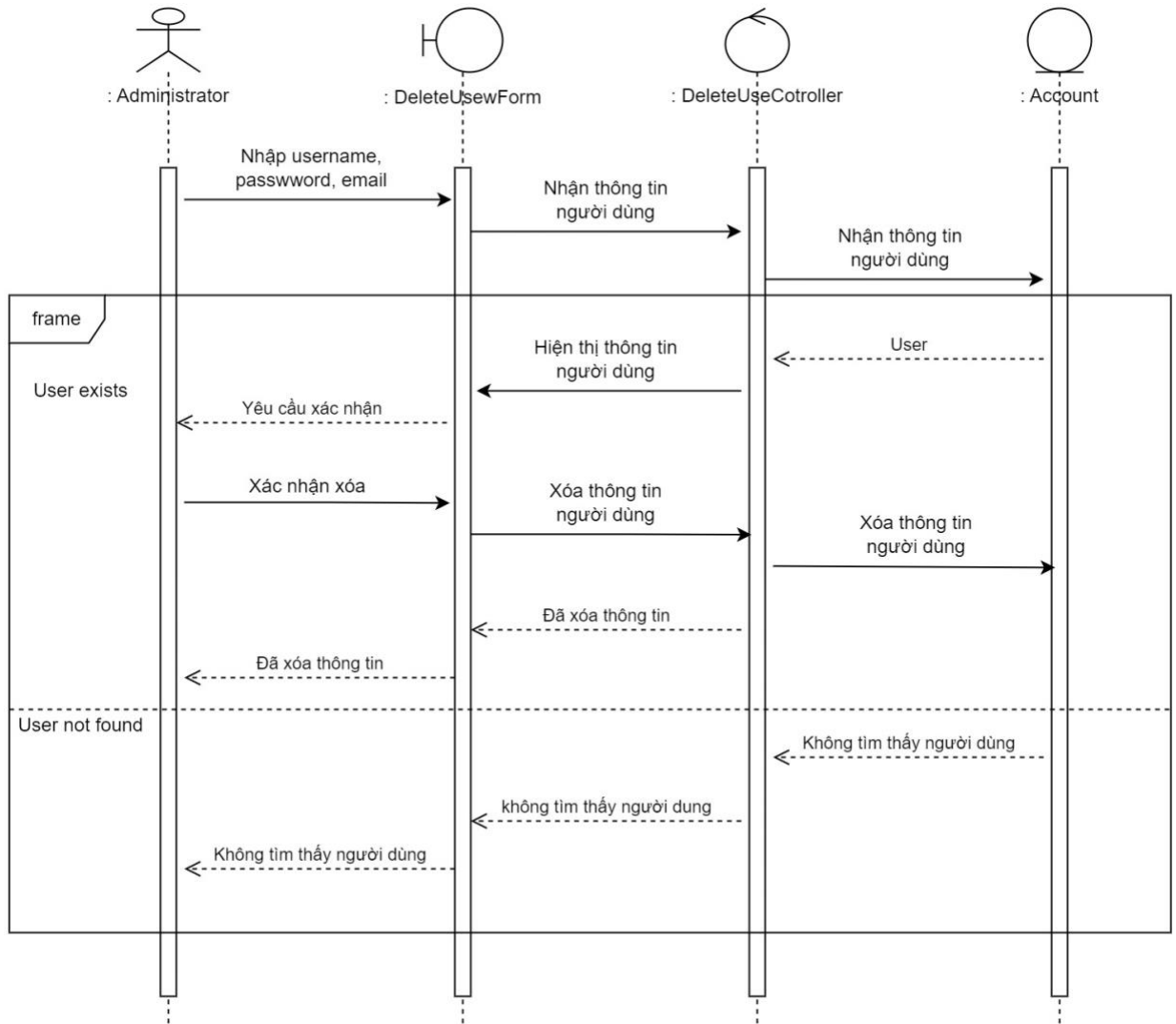
Hình 15: Design sequence diagram for the Create user use case



Hình 16: Design sequence diagram for the Create Description

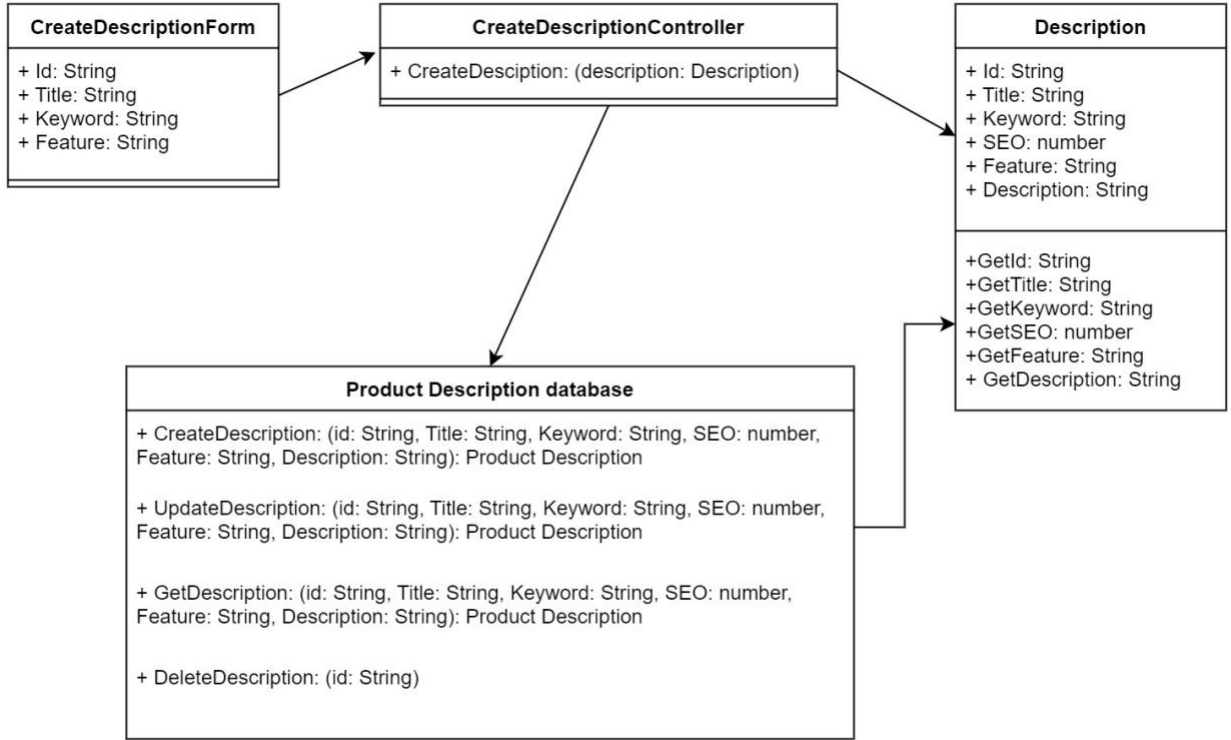


Hình 17: Design sequence diagram for the Create Description

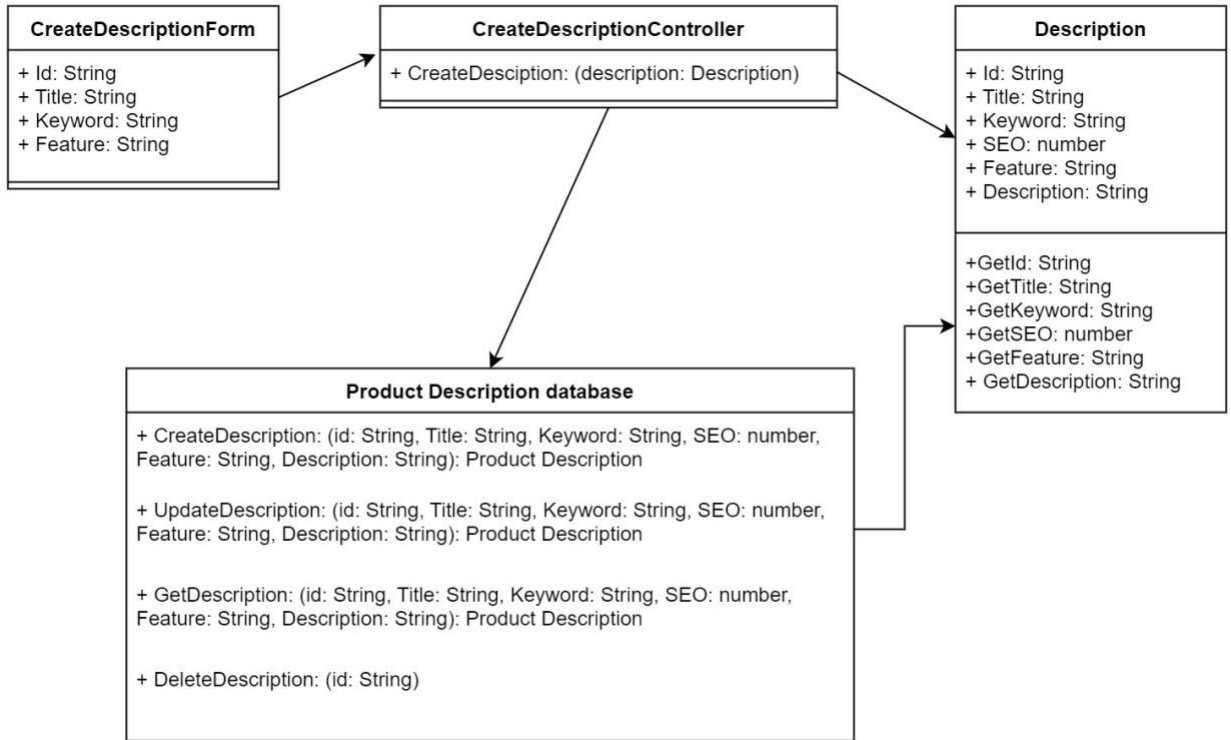


Hình 18: Design sequence diagram for the delete user

3.2.3 Thiết kế biểu đồ lớp (Class diagrams)

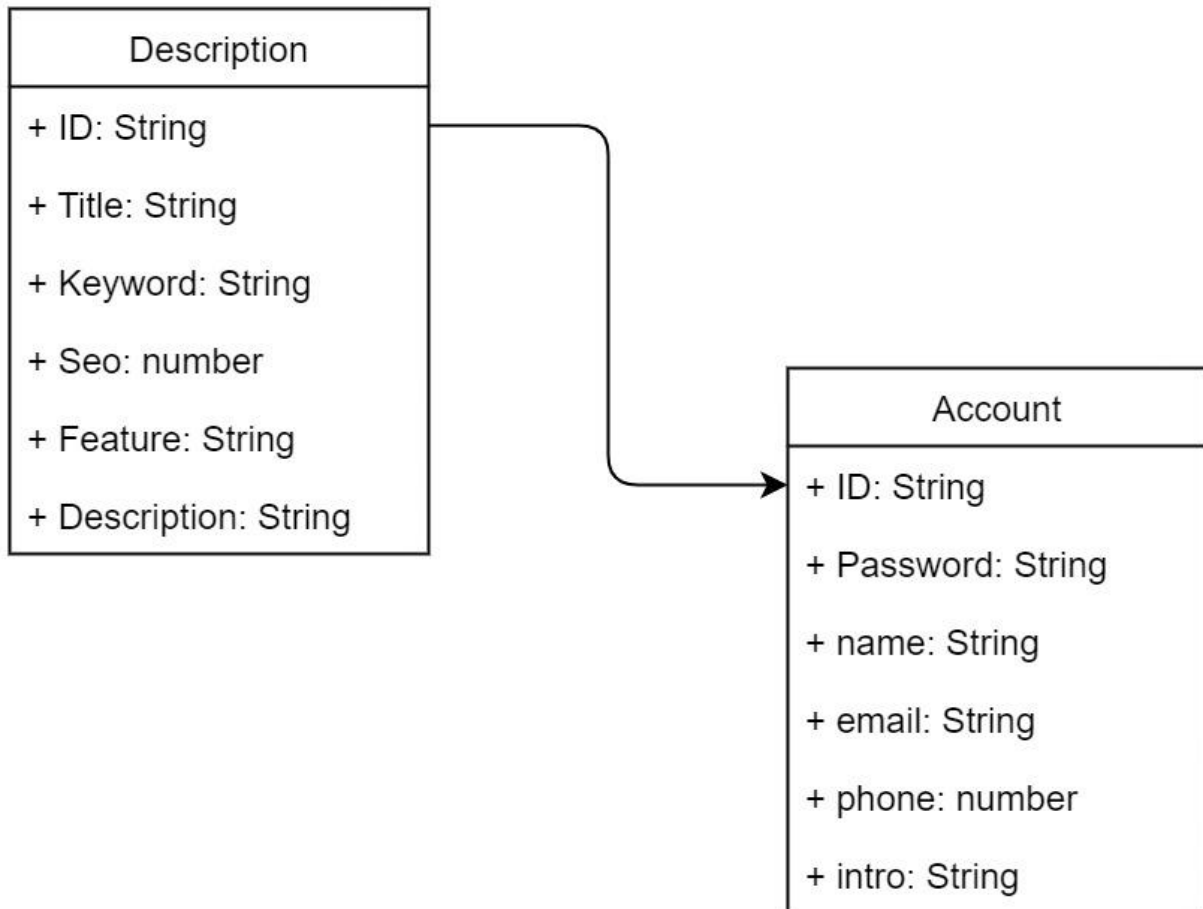


Hình 19: Design VOPC for the Create Description use case



Hình 20: Design VOPC for the Delete Description use case

3.3 Thiết kế cơ sở dữ liệu (Database design)



4. Mô hình GPT

4.1 GPT-1: Improving Language Understanding by Generative Pre-Training

Trước công trình này, hầu hết các mô hình NLP state-of-the-art được huấn luyện chuyên biệt về một nhiệm vụ cụ thể như sentiment classification, textual entailment, v.v. sử dụng phương pháp supervised learning. Tuy nhiên, các mô hình supervised có hai hạn chế chính:

1. Cần một lượng lớn dữ liệu được chú thích để học một nhiệm vụ cụ thể mà thường không dễ dàng có được.
2. Không thể khái quát các nhiệm vụ khác với những gì đã được huấn luyện.

Paper này đề xuất học một mô hình ngôn ngữ chung sử dụng dữ liệu không được gán nhãn và sau đó tinh chỉnh mô hình bằng cách cung cấp các ví dụ về các nhiệm vụ cụ thể.

Đôi với unsupervised learning, GPT-1 sử dụng hàm loss cho mô hình ngôn ngữ chuẩn:

$$L_1(T) = \sum_i \log P(t_i | t_{i-k}, \dots, t_{i-1}; \theta) \quad (i)$$

trong đó T là tập hợp các tokens trong dữ liệu unsupervised $\{t_1, \dots, t_n\}$, k là kích thước của cửa sổ context, θ là các tham số của mạng nơ-ron.

Hàm loss cho finetuning được thiết kế để dự đoán nhãn y dựa vào các đặc trưng x_1, \dots, x_n .

$$L_2(C) = \sum_{x,y} \log P(y | x_1, \dots, x_n) \quad (ii)$$

trong đó C là tập dữ liệu được gán nhãn.

Thay vì chỉ tối đa hóa hàm mục tiêu được đề cập trong phương trình (ii), các tác giả kết hợp tối ưu hai hàm mục tiêu theo công thức:

$$L_3(C) = L_2(C) + \lambda L_1(C) \quad (iii)$$

trong bài báo, λ có giá trị 0,5.

GPT-1 sử dụng *BooksCorpus* dataset chứa 7000 cuốn sách để huấn luyện mô hình ngôn ngữ. Corpus này chứa các đoạn văn bản liền kề lớn, giúp mô hình tìm hiểu các phụ thuộc trên phạm vi rộng.

GPT-1 đạt kết quả tốt hơn các mô hình state-of-the-art trên 9 trong số 12 nhiệm vụ mà các mô hình được so sánh. Quan trọng hơn, bài báo đã chứng minh rằng mô hình ngôn ngữ đóng vai trò trừu tượng hóa các khái niệm giúp cho GPT-1 có khả năng dự đoán zero-shot.

4.2 GPT-2: Language Models are Unsupervised Multitask Learners

Cải tiến của GPT-2 chủ yếu là sử dụng dataset lớn hơn và dùng mô hình nhiều tham số hơn. Chúng ta đã biết, hàm mục tiêu của mô hình ngôn ngữ được xây dựng dưới dạng $P(\text{output}|\text{input})$. Tuy nhiên, GPT-2 hướng đến việc học nhiều tác vụ bằng cách sử dụng cùng một mô hình unsupervised. Để đạt được điều đó, hàm mục tiêu được sửa đổi thành $P(\text{output}|\text{input}, \text{task})$.

Sửa đổi này được gọi là task conditioning, trong đó mô hình tạo ra output khác nhau với cùng một input cho các nhiệm vụ khác nhau.

Để tạo ra một tập dữ liệu lớn và chất lượng, các tác giả rà soát nền tảng Reddit và lấy dữ liệu từ các liên kết ngoài của các paper có uy tín. Tập dữ liệu kết quả được gọi là WebText, có 40GB dữ liệu văn bản từ hơn 8 triệu tài liệu. Tập dữ liệu này được sử dụng để huấn luyện GPT-2, nó lớn hơn nhiều so với tập dữ liệu của Book Corpus được sử dụng để huấn luyện GPT-1.

Các tác giả đã huấn luyện bốn mô hình ngôn ngữ với số các tham số lần lượt là 117 triệu (giống GPT-1), 345 triệu, 762 triệu và 1,5 tỷ tham số (GPT-2). Thông qua thực nghiệm, các mô hình sau có perplexity thấp hơn các mô hình trước, tức là có khả năng mô hình ngôn ngữ tốt hơn.

GPT-2 sau đó được đánh giá trên 8 bộ dataset và đạt kết quả tốt trên 7 bộ trong số chúng. Đối với tác vụ tóm tắt văn bản, GPT-2 hoạt động thậm chí chỉ ngang bằng hoặc kém hơn các cách tiếp cận cổ điển.

GPT-3: Language Models are Few-Shot Learners

GPT-3 có số tham số gấp 10 lần so với mô hình ngôn ngữ Turing NLG mạnh mẽ của Microsoft và gấp 100 lần so với GPT-2. Do có số lượng lớn các tham số và dataset phong phú, GPT-3 thực hiện tốt các tác vụ NLP với zero-shot và few-shot setting. Do dung lượng lớn, nó có khả năng viết các bài viết mà chúng ta khó có thể phân biệt với các bài viết của con người. Nó cũng có thể thực hiện các tác vụ nhanh chóng mà nó chưa bao giờ được huấn luyện một cách rõ ràng.

Trong few-shot setting, mô hình được cung cấp mô tả nhiệm vụ và nhiều ví dụ phù hợp với cửa sổ ngữ cảnh của mô hình. Trong one-shot setting, mô hình được cung cấp chính xác một ví dụ và trong zero-shot setting thì không có ví dụ nào được cung cấp. Với việc tăng dung lượng, khả năng của mô hình trong few, one và zero-shot cũng được cải thiện.

GPT-3 được huấn luyện trên sự kết hợp của năm kho tài liệu khác nhau là Common Crawl, WebText2, Books1, Books2 và Wikipedia.

GPT-3 chứng minh hiệu quả của mình trên một loạt các nhiệm vụ NLP. Ngoài ra, mô hình cũng được đánh giá về các tác vụ tổng hợp như cộng số học, giải mã từ, sinh tin tức, học và sử dụng từ mới v.v. Đối với những nhiệm vụ này, hiệu suất tăng lên cùng với sự gia tăng số lượng các tham số và mô hình hoạt động tốt hơn trong few-shot setting hơn là one hay zero-shot.

Bên cạnh sức mạnh của mình, GPT-3 tiềm ẩn nguy cơ lạm dụng khả năng tạo văn bản giống như con người để lừa đảo, gửi thư rác, truyền bá thông tin sai lệch hoặc thực hiện các hoạt động gian lận khác. Ngoài ra, văn bản do GPT-3 tạo ra có những thành kiến của ngôn ngữ mà nó được huấn luyện. Các bài báo được tạo bởi GPT-3 có thể có thành kiến về giới tính, dân tộc, chủng tộc hoặc tôn giáo. Do đó, điều cực kỳ quan trọng là phải sử dụng các mô hình như vậy một cách cẩn thận và xem xét văn bản được tạo ra trước khi sử dụng.

Hi vọng thông qua bài viết này các bạn đã hiểu thêm về hành trình của các mô hình GPT của OpenAI và có cho mình một số ý tưởng cho nghiên cứu của riêng bạn. Hãy chia sẻ bài viết với những người quan tâm và hãy thường xuyên truy cập website để có những thông tin mới và chuyên sâu về lĩnh vực.