**TÀI LIỆU BÁO CÁO ĐỒ ÁN CHUYÊN NGÀNH**

**Bản ghi thay đổi**

\*A - Added M - Modified D - Deleted

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Effective Date | Changed Items | A\* M, D | Change Description | New Version |
| Sep 10 |  | A | The first version | 1.0 |
| Oct 10 | 3. yêu cầu chức năng.  - M | A,M | The second version | 1.1 |
| Oct 20 |  | M | The second version | 1.1 |
| Oct 30 |  | M | The third version | 1.2 |
| Nov 30 |  | A,M | The fourth version | 1.3 |
|  |  |  |  |  |

[I. Mở đầu 4](#_Toc435564491)

[II. HỆ THỐNG GỢI Ý SỬ DỤNG PHẢN HỒI TIỀM ẨN 5](#_Toc435564492)

[A. Hệ thống gợi ý 5](#_Toc435564493)

[1.1 Mục đích 7](#_Toc435564494)

[1.2 Các kỹ thuật gợi ý 7](#_Toc435564495)

[III. XÂY DỰNG HỆ THỐNG 10](#_Toc435564496)

[1. Các định nghĩa, viết tắt và mô tả 10](#_Toc435564497)

[2. Các yêu cầu ở mức cao 10](#_Toc435564498)

[2.1 Product Perspective 10](#_Toc435564499)

[2.2 Miêu tả tác nhân và các ca sử dụng 11](#_Toc435564500)

[2.2.1 Biểu đồ UC 11](#_Toc435564501)

[2.2.2 Biểu đồ lớp tham gia UC 12](#_Toc435564502)

[2.2.3 Biểu đồ trình tự 14](#_Toc435564503)

[2.2.2 Mô tả tác nhân 17](#_Toc435564504)

[2.2.3 Mô tả UC 17](#_Toc435564505)

[2.2.4 Biểu đồ UC và Actor 18](#_Toc435564506)

[2.3 Môi trường hệ điều hành 19](#_Toc435564507)

[2.4 Giả thiết và các điều kiện lệ thuộc 19](#_Toc435564508)

[3. Yêu cầu chức năng 19](#_Toc435564509)

[3.1 Login 19](#_Toc435564510)

[3.2 Clustering customer 20](#_Toc435564511)

[3.3 Recommend product 21](#_Toc435564512)

[3.4 Retrieve dynamic data 23](#_Toc435564513)

[4.Thiết kế 24](#_Toc435564514)

[4.1 Thiết kế cơ sở dữ liệu 24](#_Toc435564515)

[4.2 Thiết kế thuật toán 29](#_Toc435564516)

[5. Lập trình 29](#_Toc435564517)

[6. Kiểm thử 30](#_Toc435564518)

[7. Vận hành bảo trì 30](#_Toc435564519)

[8. Các yêu cầu phi chức năng 30](#_Toc435564520)

[8.1 Khả năng thích ứng 30](#_Toc435564521)

[8.2 Kiểm tra khả năng 30](#_Toc435564522)

[8.3 Tính đúng đắn 30](#_Toc435564523)

[8.4. Bảo trì 30](#_Toc435564524)

[8.5. Thời gian đáp ứng 31](#_Toc435564525)

[8.6. Khả năng tương thích 31](#_Toc435564526)

[8.7. Mạnh mẽ 31](#_Toc435564527)

[8.8 Sự an toàn 32](#_Toc435564528)

[9.1 Những hạn chế phần cứng 32](#_Toc435564529)

[9.2 Phần mềm hạn chế 32](#_Toc435564530)

[9.3. High-Level Ngôn ngữ 32](#_Toc435564531)

**NỘI DUNG**

# I. Mở đầu

Hệ thống gợi ý (Recommender System - RS) được ứng dụng khá thành công trong thực tiễn giúp người dùng giải quyết vấn đề quá tải thông tin. Hiện nay, hệ thống gợi ý đang được nghiên cứu và ứng dụng ở nhiều lĩnh vực khác nhau đặc biệt là thương mại điện tử. Trên thế giới, đã có nhiều công ty, tổ chức áp dụng thành công hệ thống gợi ý như là một dịch vụ thương mại của mình nhằm gợi ý các dịch vụ, sản phẩm và các thông tin cần thiết đến người dùng như: website mua sắm trực tuyến Amazon (www.amazon.com) cung cấp cho khách hàng những sản phẩm mà họ có thể quan tâm, cổng video clip YouTube (www.youtube.com), giới thiệu phim của MovieLens (www.movielens.org),... Việc gợi ý sản phẩm phù hợp góp phần làm tăng doanh số bán hàng hoặc số lượng truy cập, download của hệ thống. Đồng thời giúp cho khách hàng có thể tìm kiếm được những thông tin thú vị hoặc những sản phẩm mà họ muốn tìm dễ dàng hơn. Hệ thống gợi ý giúp người dùng chọn lựa được thông tin phù hợp nhất cho mình dựa trên những hành vi/phản hồi (feedbacks) mà người dùng đã thực hiện trong quá khứ. Các phản hồi có thể được xác định một cách tường minh (explicit feedback) như thông qua việc đánh giá/xếp hạng (ví dụ, rating từ 1 đến 5; hay like (1) và dislike (0),…) mà người dùng đã bình chọn cho trên sản phẩm – trong trường hợp này gọi là dự đoán xếp hạng (rating prediction) [4] hoặc các phản hồi có thể được xác định một cách không tường minh hay còn gọi là tiềm ẩn (implicit feedback) như số lần click chuột, số lần chọn mua sản phẩm, thời gian mà người dùng đã duyệt/xem sản phẩm,… Rất nhiều hệ thống lớn thu thập thông tin phản hồi từ khách hàng một cách tường minh, như Ebay, Amazon, LastFM, NetFlix,.. ở đó người dùng bình chọn sản phẩm từ đến (rất thích); hay Youtube thu thập thông tin qua like()/ disklike(), và các hệ thống khác [3]. Thông qua việc thu thập phản hồi tường minh, hệ thống dễ dàng xác định rõ mức độ yêu thích của người dùng trên sản phẩm, từ đó dự đoán các sản phẩm tiếp theo mà người dùng có thể thích để gợi ý cho họ. Tuy nhiên, điều này có thể gây bất lợi do không phải người dùng lúc nào cũng sẳn sàng/vui lòng để lại các phản hồi của họ, vì vậy hệ thống phải nên tự xác định người dùng cần gì thông qua phản hồi tiềm ẩn. Trong bài viết này, chúng tôi đề xuất một giải pháp xây dựng hệ thống gợi ý cho bán hàng trực tuyến, sử dụng phản hồi tiềm ẩn từ người dùng (như số lần duyệt/xem sản phẩm, số lần mua sản phẩm). Trước hết chúng tôi đề xuất phương pháp thu thập và khai thác thông tin phản hồi tiềm ẩn từ người dùng, sau đó lựa chọn và đề xuất kết hợp các mô hình sử dụng thông tin phản hồi tiềm ẩn. Kế đến là việc xây dựng hệ thống và tích hợp các giải thuật gợi ý vào hệ thống. Sau khi có hệ thống hoàn chỉnh, chúng tôi thu thập dữ liệu từ người dùng thực nhằm đánh giá hiệu quả của hệ thống gợi ý. Kết quả cho thấy khả năng mà hệ thống gợi ý phù hợp với sở thích của từng người dùng là khá tốt.

Có nhiều phương pháp có thể có thể tiếp cận nhằm nâng cao hiệu quả trang web của bạn để nó được đánh giá cao hơn.Nhưng trong nôi dung tìm hiểu của nhóm.Chúng tôi xin đưa ra một cách tiếp cận dựa trên phản hồi tiềm ẩn.

# II. HỆ THỐNG GỢI Ý SỬ DỤNG PHẢN HỒI TIỀM ẨN

## A. Hệ thống gợi ý

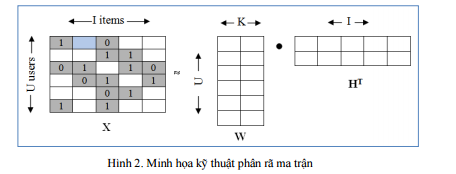
Hệ thống gợi ý (RS) dựa vào sở thích, thói quen, nhu cầu,... trong quá khứ của người sử dụng để dự đoán sở thích trong tương lai của họ. Trong hệ thống gợi ý người ta quan tâm đến 3 đối tượng: người dùng (user), sản phẩm (item - item gọi chung là mục tin nhưng trong bài viết này liên quan đến gợi ý sản phẩm nên từ đây về sau chúng tôi tạm gọi item là sản phẩm) và các đánh giá sở thích của người dùng trên sản phẩm gọi là rating. Thông thường người ta gọi U là tập tất cả người dùng (users) và u là một người dùng cụ thể nào đó (u∈U). I là tập tất cả các sản phẩm (items) sẽ được gợi ý như máy tính, sách, phim ảnh,.. và i là một sản phẩm cụ thể nào đó (i∈I). I là tập các sản phẩm có thể lên đến hàng trăm, hàng nghìn hoặc thậm chí là hàng triệu sản phẩm trong một số ứng dụng, như việc gợi ý về sách, phim ảnh, âm nhạc. Tương tự như vậy, tập người dùng U cũng có thể rất lớn, lên đến hàng triệu trường hợp. R là một tập hợp các giá trị dùng để ước lượng ‘sở thích’ (preference) của người dùng, và rui∈R (R⊂ℜ) là xếp hạng của người dùng u trên sản phẩm i. Giá trị rui có thể được xác định một cách tường minh (explicit feedback) như thông qua việc đánh giá/xếp hạng (ví dụ, rating từ 1 đến 5; hay like (1)/ dislike (0),…) mà u đã bình chọn cho i – trong trường hợp này gọi là dự đoán xếp hạng (rating prediction) hoặc rui có thể được xác định một cách không tường minh hay còn gọi là tiềm ẩn (implicit feedback) như số lần click chuột, số lần chọn mua sản phẩm, thời gian mà u đã duyệt/xem i,… [2][7] Nội dung bài viết này, chỉ quan tâm nhiều đến cách xác định rui không tường minh. Các thông tin này được biểu diễn thông qua một ma trận như trong Hình 1. Trong đó mỗi dòng là một người dùng u, mỗi cột là một sản phẩm i, và giao giữa dòng và cột là các đánh giá rating của người dùng như số lần click chuột hay chọn mua sản phẩm,…. Các ô có giá trị là những item mà các user đã xem đến hoặc chọn mua trong quá khứ. Những ô trống là những item chưa được xem đến (điều đáng lưu ý là mỗi user chỉ click xem hoặc chọn mua cho một vài item trong quá khứ, do vậy có rất nhiều ô trống trong ma trận này – còn gọi là ma trận thưa – sparse matrix). Hình 1. Ma trận biểu diễn xếp hạng của người dùng trên sản phẩm (user-item-rating matrix) Nhiệm vụ chính của RS là dựa vào các ô đã có giá trị trong ma trận này (dữ liệu thu được từ quá khứ), để dự đoán các ô còn trống (của user hiện hành), sau đó sắp xếp kết quả dự đoán (ví dụ, từ cao xuống thấp) và chọn ra Top-N items theo thứ tự, từ đó gợi ý chúng đến người dùng. Một cách hình thức, nếu gọi Dtrain ⊆ U × I × R là tập dữ liệu huấn luyện, Dtest ⊆ U × I × R là tập dữ liệu kiểm thử, và một ánh xạ r: U × I→ R (u, i) ↦ rui Mục tiêu của RS là tìm một hàm ̂: U × I → ℜ sao cho ξ(r, ̂) thỏa mãn một điều kiện nào đó. Ví dụ, nếu ξ là một hàm ước lượng lỗi như Mean Absolute Error (MAE) hay Root Mean Squared Error (RMSE) thì nó cần phải được tối tiểu. ∑ ( ) ∈ = − test r)i,(u, D ui ui test r ˆ D| | 1 MAE r (1) RMSE = ∑( ) ∈ − test Driu test ui iu r r D ),,( 2 ),( ˆ | | 1 (2) Hiện nay, có rất nhiều giải thuật được đề xuất cho hệ thống gợi ý, chúng có thể được gom lại theo 3 nhóm [1][2][7]: - Gợi ý dựa trên cộng tác: người dùng sẽ nhận gợi ý những sản phẩm được ưa thích xuất phát từ những người có cùng thị hiếu và sở thích với mình. Nhóm này dựa vào các phương pháp chủ yếu: o Phương pháp láng giềng (Neighborhood-based, còn gọi là Memory-based), trong đó hoặc là dựa trên dữ liệu quá khứ của người dùng “tương tự - similarity” (user-based approach), hoặc là dựa trên dữ liệu quá khứ của những item “tương tự” (item-based approach). o Dựa trên mô hình (Model-based): Nhóm này liên quan đến việc xây dựng các mô hình dự đoán dựa trên dữ liệu thu thập được trong quá khứ. Như mô hình Bayesian, các mô hình nhân tố tiềm ẩn (latent factor models): trong đó kỹ thuật phân rã ma trận (matrix factorization) là một điển hình. - Gợi ý dựa trên nội dung: người dùng sẽ được gợi ý những sản phẩm tương tự với những sản phẩm đã được người dùng đó ưa thích trước đây. - Gợi ý dựa trên cách tiếp cận kết hợp: kết hợp hai phương pháp tiếp cận dựa trên nội dung và cộng tác. Sau đây chúng tôi tóm lược lại một trong những kỹ thuật trong nhóm lọc cộng tác của hệ thống gợi ý và kỹ thuật sử dụng phản hồi tiềm ẩn, từ đó làm cơ sở cho việc đề xuất mô hình cho hệ thống.

### 1.1 Mục đích

Mục đích của hệ thống gợi ý (RS) là dựa vào sở thích, thói quen, nhu cầu,... trong quá khứ của người sử dụng để dự đoán sở thích trong tương lai của họ. Trong hệ thống gợi ý người ta quan tâm đến 3 đối tượng: người dùng (user), sản phẩm (item) và các đánh giá sở thích của người dùng trên sản phẩm gọi là rating.Tài liệu này chỉ định được dung cho các thành viên trong dự án thực hiện và xác định đúng các chức năng của hệ thống.

### 1.2 Các kỹ thuật gợi ý

B. Kỹ thuật phân rã ma trận (matrix factorization – MF) Kỹ thuật phân rã ma trận (MF) là một trong những phương pháp dựa trên mô hình thành công nhất hiện nay (stateof-the-art) trong RS [1][2]. MF là việc chia một ma trận lớn X thành hai ma trận có kích thước nhỏ hơn W và H, sao cho ta có thể xây dựng lại X từ hai ma trận nhỏ hơn này càng chính xác càng tốt [5]. X ~ WHT như minh họa như trong Hình 2. Trong đó, W ∈ℜ|U|×K là một ma trận mà mỗi dòng u là một véc tơ bao gồm K nhân tố tiềm ẩn (latent factors) mô tả người dùng u; và H ∈ℜ|I|×K là một ma trận mà mỗi dòng i là một véc tơ bao gồm K nhân tố tiềm ẩn mô tả cho item i (lưu ý: K<<

****

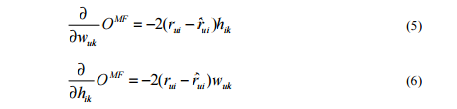
Gọi wuk và hik là các phần tử tương ứng của hai ma trận W và H, khi đó xếp hạng của người dùng u trên mục tin i được dự đoán bởi công thức:

****

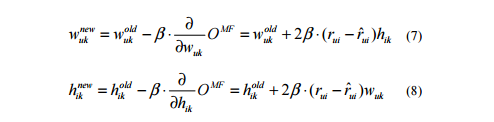
Như vậy, vấn đề chủ chốt của kỹ thuật MF là làm thế nào để tìm được giá trị của hai tham số W và H. Hai tham số này có được bằng cách tối ưu hóa hàm mục tiêu (objective function) như RMSE ở công thức (2). Ta có thể viết lại hàm mục tiêu của MF như sau:

****

Một trong những kỹ thuật có thể dùng để tối ưu hóa hàm mục tiêu là dùng SGD (Stochastic Gradient Descent) [5]. Để tối ưu hóa hàm mục tiêu (4), trước tiên ta khởi tạo các giá trị ngẫu nhiên cho W và H, sau đó từng bước cập nhật giá trị của chúng cho đến khi hàm mục tiêu hội tụ về giá trị nhỏ nhất (convergence). Để làm được điều đó, ta cần phải xác định là nên tăng hay nên giảm các giá trị của W và H qua mỗi lần cập nhật, do vậy cần phải tìm đạo hàm tương ứng của chúng:

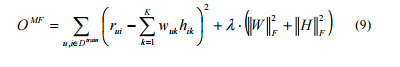
****

Sau khi tìm đạo hàm, các phần tử của W và H sẽ được cập nhật ngược hướng với giá trị của đạo hàm, qua công thức:

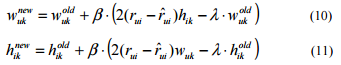
****

Trong đó β là tốc độ học (learning rate **)**Quá trình cập nhật sẽ được thực hiện đến khi nào hàm mục tiêu đạt được giá trị nhỏ nhất.

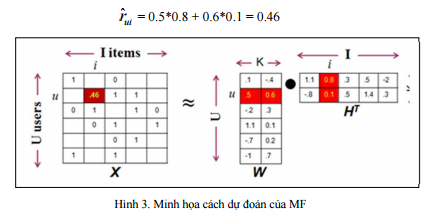
Chính tắc hóa (Regularization): Để ngăn ngừa sự quá khớp hay còn gọi là học vẹt (overfitting – xảy ra khi mô hình dự đoán cho kết quả tốt trên dữ liệu huấn luyện, nhưng cho kết quả kém trên dữ liệu thử nghiệm) người ta thay đổi hàm mục tiêu (4) bằng cách thêm vào một đại lượng gọi là chính tắc hóa (regularization) để điều khiển độ lớn của các giá trị trong W và H. Hàm mục tiêu (4) bây giờ trở thành:

****

Trong đó λ là hệ số chính tắc hóa **** là chuẩn Frobenius. Với hàm mục tiêu mới này, giá trị của wuk và hik được cập nhật qua công thức:

****

Quá trình cập nhật giá trị của W và H được lặp lại cho đến khi đạt độ lỗi chấp nhận hoặc lặp lại đến số lần qui định trước. Quá trình dự đoán: Sau quá trình huấn luyện ta được 2 ma trận W và H đã tối ưu thì quá trình dự đoán được thực hiện theo công thức (3). Ví dụ: Dự đoán xếp hạng của user 2 cho item 2 trong Hình 3:

****

Một biến thể của MF dành cho dữ liệu phản hồi tiềm ẩn là kỹ thuật SVD++ [3] sẽ được trình bày trong phần tiếp theo.

# III. XÂY DỰNG HỆ THỐNG

### 1. Các định nghĩa, viết tắt và mô tả

|  |  |  |
| --- | --- | --- |
| **#** | **Abbreviations/Terms** | **Explanation** |
|  |  |  |
|  |  |  |

Table 1: Abbreviations & Terminologies

Tài liệu

|  |  |  |
| --- | --- | --- |
| # | Tên | Mô tả |
| 1 | PHƯƠNG PHÁP XÂY DỰNG HỆ THỐNG GỢI Ý SẢN PHẨM SỬ DỤNG PHẢN HỒI TIỀM ẨN | Tác giả : Lưu Nguyễn Anh Thư, Nguyễn Thái Nghe  Thuộc: Khoa Công Nghệ Thông Tin và Truyền Thông, Trường Đại Học Cần Thơ |
| 2 |  |  |
| 3 |  |  |

Bảng 2: Tài liệu

### 2. Các yêu cầu ở mức cao

#### 2.1 Product Perspective

Ứng dụng được viết nhằm mục đích xây dựng để tạo ra một bảng kết quả sau đó sẽ đẩy lên web. Do đó, chúng ta có thể dung bất kì hệ điều hành nào hay bất kì ngôn ngữ lập trình nào để xây dựng. Trong tài liệu này chúng tôi xin đề xuất về việc làm trên hệ điều hành window với ngôn ngữ lập trình C# sử dụng nền tảng .Net .Còn môi trường web, trong tài liệu chúng tôi xấy dựng trên mã nguồn mở PHP và sử dụng framework joomla.

#### 2.2 Miêu tả tác nhân và các ca sử dụng

##### 2.2.1 Biểu đồ UC

**Biểu đồ UC cho sys**



**Biểu đồ UC cho customer**

****

##### 2.2.2 Biểu đồ lớp tham gia UC

**Biểu đồ cho clustering user**

****

**Biểu đồ recommend product**

****

**Biểu đồ cho retrieve dynamic data**

****

##### 2.2.3 Biểu đồ trình tự

**Biểu đồ trình tự cho clustering user**

****

**Biểu đồ trình tự cho recommend product**

****

**Biểu đồ trình tự cho retrieve dynamic data**

****

**Biểu đồ trình tự cho recommend product**

****

##### 2.2.2 Mô tả tác nhân

|  |  |  |
| --- | --- | --- |
| **#** | **Actor Name** | **Definition & Interests** |
| 1 | Admin | Tương tác với hệ thống sau |
| 2 | Customer | Khách hàng tương tác với trang web |

##### 2.2.3 Mô tả UC

|  |  |  |  |
| --- | --- | --- | --- |
| **#** | **Code** | **Name** | **Brief Description** |
| 1 | UC01 | Login | Cho phép actor đăng nhập vào hệ thống |
| 2 | UC02 | Clustering user | Cho phép actor thực hiện việc phân cụm người dùng |
| 3 | UC03 | Retrieve dynamic data | Cho phép lấy dữ liệu |
| 4 | UC04 | Recommend product | Cho phép gợi ý sản phẩm |
| 5 | UC05 | Recommend ads | Cho phép gợi ý quảng cáo |
| 6 | UC06 | Analyse interactive user | Cho phép phân tích tương tác người dùng |
| 7 | UC07 | Analyse transaction | Cho phép phân tích được lịc sử giao dịch người dùng |
| 8 | UC08 | Analyse comment | Cho phép phân tích bình luận |
| 9 | UC09 | Like product | Cho phép actor like sản phẩm |
| 10 | UC10 | Purchase product | Cho phép actor mua hàng |
| 11 | UC11 | Shipping product | Cho phép actor đặt hàng |
| 12 | UC12 | Rating product | Cho phép actor bình chọn sản phẩm |
| 13 | UC13 | View product | Cho phép actor xem sản phẩm |
| 14 | UC14 | Comment product | Cho phép actor bình luận sản phẩm |
| 15 | UC15 | Sign in | Cho phép actor đăng nhập |

##### 2.2.4 Biểu đồ UC và Actor

|  |  |  |
| --- | --- | --- |
| **Actor**  **Use Case** | **Customer** | **Admin** |
| UC01 : Login | **x** |  |
| UC02 : Clustering user |  | **x** |
| UC03 : Retrieve dynamic data |  | **x** |
| UC04 : Recommend product |  | **x** |
| UC05: Recommend ads |  | **x** |
| UC06 : Analyse interactive user |  | **x** |
| UC07 : Analyse transaction |  | **x** |
| UC08 : Analyse comment |  | **x** |
| UC09 : Like product | **x** |  |
| UC10 : Purchase product | **x** |  |
| UC11 : Shipping product | **x** |  |
| UC12 : Rating product | **x** |  |
| UC13 : View product | **x** |  |
| UC14 : Comment product | **x** |  |
| UC15 : Sign in | **x** |  |

#### 2.3 Môi trường hệ điều hành

Trong tài liệu này chúng tôi đê xuất thông tin cần thiết sau:

OS: yêu cầu HDH window XP trở lên. Đề xuất window 7

Kĩ thuật: RAM 1Gb trở lên, HDD: 10Gb trở lên, chip : Pentium trở lên, Xung: 1Ghz trở lên.

Công cụ: Visual studio, Microsoft SQL server, xmind, MS project, MS visio, Rational Rose. Đề xuất: visual studio 2012 , MS SQL server 2012,

#### 2.4 Giả thiết và các điều kiện lệ thuộc

Các công cụ làm việc tốt, máy móc không bị hỏng hóc.

### 3. Yêu cầu chức năng

#### 3.1 Login

Mô tả UC

|  |  |  |  |
| --- | --- | --- | --- |
| **Name** | Login | **Code** | UC01 |
| **Description** | Cho phép actor đăng nhập vào hệ thống | | |
| **Actor** | Admin/customer | **Trigger** | Actor bấm nút login |
| **Pre-condition** |  | | |
| **Post condition** | Chuyển tới trang default với role tương ứng | | |

Hoạt động

|  |  |  |  |
| --- | --- | --- | --- |
| **Actor** | | **System** | |
| **Main Flow: Login thành công** | | | |
| 1 | Actor nhập tên đăng nhập/mật khẩu và click vào button Login trên trang đăng nhập . |  |  |
|  |  | 2 | Kiểm tra tên đăng nhập/mật khẩu là chính xác sau đó chuyển tới trang default role tương ứng () |

Thông điệp

|  |  |
| --- | --- |
| MS01 | “Tên Đăng Nhập và/hoặc Mật Khẩu của bạn không chính xác. Vui lòng kiểm tra và thử lại.”  Message thông báo khi actor nhập sai tên đăng nhập/mật khẩu |

#### 3.2 Clustering customer

Mô tả UC

|  |  |  |  |
| --- | --- | --- | --- |
| **Name** | Clusteringcustomers | **Code** | UC13 |
| **Description** | Cho phép actor thực hiện công việc thao tác vào hệ thống phân loại khách hàng | | |
| **Actor** | Admin | **Trigger** |  |
| **Pre-condition** | Actor đã đăng nhập vào hệ thống | | |
| **Post condition** | Quá trình hệ thống làm việc đúng trả lại thành công | | |

Hoạt động

|  |  |  |  |
| --- | --- | --- | --- |
| **Actor** | | **System** | |
| **Main Flow: Clustering customers successfully** | | | |
| 1 | Actor sẽ tiến hành đăng nhập vào hệ thống |  |  |
|  |  | 2 | Hiển thị thông báo log thành công |
| 3 | Tiến hành chạy service |  |  |
|  |  | 4 | Hệ thống đang xử lí |
|  |  | 5 | Tiến hành chạy ghi log file và kết thúc là hiển thị thành công |

Thông điệp

|  |  |
| --- | --- |
|  |  |

#### 3.3 Recommend product

Miêu tả UC

|  |  |  |  |
| --- | --- | --- | --- |
| **Name** | Recommend product | **Code** | UC14 |
| **Description** | Cho phép gợi ý sản phẩm cho khách hàng | | |
| **Actor** | Admin | **Trigger** | Chạy service |
| **Pre-condition** | Actor đã đăng nhập vào hệ thống | | |
| **Post condition** | Hệ thống chạy thành công. | | |

Hoạt động

|  |  |  |  |
| --- | --- | --- | --- |
| **Actor** | | **System** | |
| **Main Flow: Recommend product** | | | |
| 1 | Chuyển tiếp sang recommend product |  |  |
|  |  | 2 | Chuyển hướng lựa chọn theo user |
|  |  | 3 | Tiến hành xử lí, ghi log nội dung output và thông báo cho người dùng biết đầu ra khi hoàn tất |

Business Rules

|  |  |  |
| --- | --- | --- |
| **Rule No.** | **Rule** | **Description** |
|  |  |  |

System Message

N/A

#### 3.4 Retrieve dynamic data

* Use Case Description

|  |  |  |  |
| --- | --- | --- | --- |
| **Name** | Retrieve dynamic data | **Code** | UC15 |
| **Description** | Cho phép lấy dữ liệu từ các db của các trang web khác nhau | | |
| **Actor** | Admin | **Trigger** | Tiến hành thay đổi 1 số câu lệnh theo mặc định cho phép |
| **Pre-condition** | Actor đã đăng nhập vào hệ thống | | |
| **Post condition** | Trả về kết quả thành công | | |

Activities

|  |  |  |  |
| --- | --- | --- | --- |
| **Actor** | | **System** | |
| **Main Flow: retrieve dynamic data successfully** | | | |
| 1 | Chuyển tiếp sang recommend product |  |  |
|  |  | 2 | Chuyển hướng lựa chọn theo user |
|  |  | 3 | Tiến hành xử lí, ghi log nội dung output và thông báo cho người dùng biết đầu ra khi hoàn tất |

Business Rules

|  |  |  |
| --- | --- | --- |
| **Rule No.** | **Rule** | **Description** |
|  |  |  |

System Message

N/A

3.5 Gợi ý quảng cáo: Việc gợi ý quảng cáo sẽ phụ thuộc vào từng trang yêu cầu và do có thể coi quảng cáo là item nên có thể hoàn toàn gần giống với product.

### 4.Thiết kế

#### 4.1 Thiết kế cơ sở dữ liệu

Lập thành các bảng để lưu trữ bảng trong CSDL. Phân theo module ta được như sau:

Module lấy dữ liệu động

* RS\_Item

|  |  |  |  |
| --- | --- | --- | --- |
| Index | Thuộc tính | Kiểu dữ liệu | Mô tả |
| 1 | ItemID | Varchar(50) not null | Primary key |
| 2 | ItemName | Varchar(50) not null |  |
| 3 | MetaItemID | Varchar(50) not null |  |
| 4 | DataID | Varchar(50) not null | Primary key |
| 5 | CreatedDate | datetime |  |
| 6 | ItemCategoryCode | Varchar(50) not null |  |
| 7 | ItemFamillyCode | Varchar(50) not null |  |

* RS\_Users

|  |  |  |  |
| --- | --- | --- | --- |
| Index | Thuộc tính | Kiểu dữ liệu | Mô tả |
| 1 | UserID | Varchar(50) not null | Primary key |
| 2 | UserName | Varchar(50) not null |  |
| 3 | U\_SubCategoryID | Varchar(50) not null |  |
| 4 | U\_CategoryID | Varchar(50) not null |  |
| 5 | CreateDate | datetime |  |
| 6 | CA\_EUROS | float |  |
| 7 | DataID | Varchar(50) not null |  |

* RS\_Transaction

|  |  |  |  |
| --- | --- | --- | --- |
| Index | Thuộc tính | Kiểu dữ liệu | Mô tả |
| 1 | UserID | Varchar(50) not null | Primary key |
| 2 | ItemID | Varchar(50) not null | Primary key |
| 3 | Quantity | int |  |
| 4 | Date | datetime | Primary key |

* DI.DATA\_SOURCE

|  |  |  |  |
| --- | --- | --- | --- |
| Index | Thuộc tính | Kiểu dữ liệu | Mô tả |
| 1 | DataID | Varchar(50) not null | Primary key |
| 2 | CreationDate | datetime |  |
| 3 | ConnectionString | Varchar(MAX) not null |  |
| 4 | Type | Varchar(50) not null |  |

* DI.QUERY

|  |  |  |  |
| --- | --- | --- | --- |
| Index | Thuộc tính | Kiểu dữ liệu | Mô tả |
| 1 | QueryID | Varchar(50) not null | Primary key |
| 2 | DescTable | Varchar(50) not null |  |
| 3 | OrderNo | int |  |
| 4 | String | Varchar(MAX) not null |  |
| 5 | Description | Varchar(MAX) not null |  |
| 6 | DataID | Varchar(50) not null |  |

* DI.PARAMETER

|  |  |  |  |
| --- | --- | --- | --- |
| Index | Thuộc tính | Kiểu dữ liệu | Mô tả |
| 1 | ParameterID | Varchar(50) not null | Primary key |
| 2 | QueryID | Varchar(50) not null | Primary key |
| 3 | Value | Varchar(MAX) not null |  |
| 4 | Type | Varchar(10) not null |  |

Module phân cụm user

* RS.Rating\_Matrix

|  |  |  |  |
| --- | --- | --- | --- |
| Index | Thuộc tính | Kiểu dữ liệu | Mô tả |
| 1 | ClusterID | Varchar(50) not null |  |
| 2 | UserID | Varchar(50) not null | Primary key |
| 3 | MetaItemID | Varchar(50) not null | Primary key |
| 4 | Value | float |  |

* RS.Clustering\_User

|  |  |  |  |
| --- | --- | --- | --- |
| Index | Thuộc tính | Kiểu dữ liệu | Mô tả |
| 1 | ClusterID | Varchar(50) not null | Primary key |
| 2 | U\_SubCategoryID | Varchar(50) not null |  |

* RS.Cluster\_Setting

|  |  |  |  |
| --- | --- | --- | --- |
| Index | Thuộc tính | Kiểu dữ liệu | Mô tả |
| 1 | Keys | Varchar(50) not null | Primary key |
| 2 | DataType | Varchar(50) not null |  |
| 3 | Description | Varchar(50) not null |  |

* RS.Cluster\_Setting\_Value

|  |  |  |  |
| --- | --- | --- | --- |
| Index | Thuộc tính | Kiểu dữ liệu | Mô tả |
| 1 | ValueID | int | Primary key |
| 2 | Keys | Varchar(50) not null |  |
| 3 | Value | int |  |
| 4 | isDefault | bit |  |

Module gợi ý

* RS.Rating\_Matrix

|  |  |  |  |
| --- | --- | --- | --- |
| Index | Thuộc tính | Kiểu dữ liệu | Mô tả |
| 1 | ClusterID | Varchar(50) not null |  |
| 2 | UserID | Varchar(50) not null | Primary key |
| 3 | MetaItemID | Varchar(50) not null | Primary key |
| 4 | Value | float |  |

* RS.Recommend\_Setting\_TBL

|  |  |  |  |
| --- | --- | --- | --- |
| Index | Thuộc tính | Kiểu dữ liệu | Mô tả |
| 1 | [Key] | Varchar(50) not null | Primary key |
| 2 | DataType | Varchar(50) not null |  |
| 3 | Description | Varchar(MAX) not null |  |

* RECOMMEND\_SETTING\_VALUE\_TBL

|  |  |  |  |
| --- | --- | --- | --- |
| Index | Thuộc tính | Kiểu dữ liệu | Mô tả |
| 1 | ValueID | int | Primary key |
| 2 | Keys | Varchar(50) not null |  |
| 3 | Value | int |  |
| 4 | isDefault | bit |  |

* RECOMMENDATION\_SCHEDULE\_TBL

|  |  |  |  |
| --- | --- | --- | --- |
| Index | Thuộc tính | Kiểu dữ liệu | Mô tả |
| 1 | ScheduleID | int | Primary key |
| 2 | StartTime | datetime |  |
| 3 | StopTime | datetime |  |
| 4 | [Log] | varchar(MAX) |  |
| 5 | LoginID | varchar(10) |  |

* RECOMMENDATION\_TBL

|  |  |  |  |
| --- | --- | --- | --- |
| Index | Thuộc tính | Kiểu dữ liệu | Mô tả |
| 1 | RecomID | int | Primary key |
| 2 | ScheduleID | int |  |
| 3 | MetaItemID | varchar(50) |  |
| 4 | UserID | varchar(50) |  |
| 5 | Quantity | float |  |
| 6 | Score | float |  |
| 7 | RecommendType | varchar(50) |  |
| 8 | ItemFamillyCode | varchar(50) |  |

* WEIGHT\_MATRIX

|  |  |  |  |
| --- | --- | --- | --- |
| Index | Thuộc tính | Kiểu dữ liệu | Mô tả |
| 1 | ClusterID | varchar(50) |  |
| 2 | MetaItemSource | varchar(50) |  |
| 3 | MetaItemDestination | varchar(50) |  |
| 4 | WEIGHT | float |  |

* CONFIDENT\_MATRIX

|  |  |  |  |
| --- | --- | --- | --- |
| Index | Thuộc tính | Kiểu dữ liệu | Mô tả |
| 1 | ClusterID | varchar(50) |  |
| 2 | MetaItemSource | varchar(50) |  |
| 3 | MetaItemDestination | varchar(50) |  |
| 4 | Confident | float |  |

* DISTANCE\_MATRIX

|  |  |  |  |
| --- | --- | --- | --- |
| Index | Thuộc tính | Kiểu dữ liệu | Mô tả |
| 1 | ClusterID | varchar(50) |  |
| 2 | MetaItemSource | varchar(50) |  |
| 3 | MetaItemDestination | varchar(50) |  |
| 4 | Distance | float |  |

#### 4.2 Thiết kế thuật toán

// viet cong thuc tinh chung o day

### 5. Lập trình

Để thuận lợi cho việc phát triển có hiệu quả.Nhóm đã phân chia công việc lập trình như sau:

|  |  |
| --- | --- |
| Mô tả công việc | Thành viên |
| Lấy dữ liệu động | Vũ Tuấn Khắc –Vũ Tiến Dũng |
| Xây dựng công thức | Dương Lê Phúc Nguyện –Vũ Tiến Dũng |
| Phân cụm người dùng | Dương Lê Phúc Nguyện –Vũ Tiến Dũng |
| Gợi ý sản phẩm | Vũ Tuấn Khắc - Dương Lê Phúc Nguyện |

### 6. Kiểm thử

Các công việc kiểm thử sẽ tiến hành song song với công việc lập trình. Do tính chất công việc nên nhóm quyết định đê đội coder lập trình và kiểm thử đơn giản và sau đó sẽ thuê 1 nhân viên kiểm thử chuyên để kiểm tra trong quá trình viết mã.

### 7. Vận hành bảo trì

Quá trình vận hành sẽ được tiến hành ngay sau khi hoàn tất và công việc bảo trì là diễn ra theo định kì tháng/lần.

### 8. Các yêu cầu phi chức năng

#### 8.1 Khả năng thích ứng

Mục này quy định các yêu cầu sau đây liên quan đến mức độ mà hệ thống phải được tiếp cận:

Hỗ trợ thích ứng với các cơ sở dữ liệu hiện có.

#### 8.2 Kiểm tra khả năng

Mục này quy định các yêu cầu sau đây liên quan đến mức độ mà hệ thống chạy đọc lập với trang web hiện tại để không làm ảnh hưởng tới quá trình chạy của trang web

#### 8.3 Tính đúng đắn

Mục này quy định các yêu cầu sau đây gắn liền với sự dễ dàng mà hệ thống có thể được tích hợp với các hệ thống khác (ví dụ, các trình duyệt, ứng dụng, cơ sở dữ liệu và yêu cầu).

• Các CMR sẽ tương thích với các trình duyệt sau:

 IOP-1) Internet Explorer 11

 IOP-2) Google Chrome 34

 IOP-2) Mozilla Firefox 12

#### 8.4. Bảo trì

Mục này quy định các yêu cầu sau đây gắn liền với sự dễ dàng mà hệ thống có thể được duy trì:

• M-1) Cho phép các trao đổi và nâng cấp các phần cứng mà không có thời gian xuống.

• M-2) Cho phép nâng cấp của phần mềm mà không có thời gian xuống.

5.6. Khả năng hoạt động

Tiểu mục này quy định các yêu cầu sau kết hợp với tốc độ mà hệ thống sẽ hoạt động.

#### 8.5. Thời gian đáp ứng

Tiểu mục này quy định các yêu cầu sau đây liên quan đến cách hành quyết rất nhiều hoạt động của một hệ thống nhất định hoặc sử dụng trường hợp đường dẫn hệ thống phải có khả năng thực hiện trong một đơn vị thời gian:

• TBD

#### 8.6. Khả năng tương thích

Tiểu mục này quy định các yêu cầu sau đây gắn liền với sự dễ dàng mà hệ thống có thể được di chuyển từ một môi trường (ví dụ, phần cứng, hệ điều hành) khác.

• Các CRM sẽ cho phép người dùng sử dụng các môi trường sau đây (ví dụ, nền tảng và hệ điều hành) để tương tác với The CRM:

 POR-1) PC với tối thiểu của chip Celeron, 2 GB RAM, 256 kbps modem ADSL

 Hệ điều hành:

 POR-2) Windows 7

 POR-3) Fedora Linux 16

 POR-4) Ubuntu 14

#### 8.7. Mạnh mẽ

Tiểu mục này quy định các yêu cầu sau đây liên quan đến mức độ mà hệ thống vẫn tiếp tục hoạt động đúng theo các hoàn cảnh bất thường.

• ROB-1) Các CRM một cách duyên dáng nên xử lý đầu vào không hợp lệ (tức là phát hiện đầu vào không hợp lệ, yêu cầu đầu vào hợp lệ, và không sụp đổ) từ tất cả externals:

 Các diễn viên của con người.

 Các bộ xử lý Authorization Gateway.

• ROB-2) Các CRM một cách duyên dáng nên xử lý lỗi phần cứng (tức là cung cấp chuyển đổi dự phòng nóng, thông báo cho các nhà điều hành hệ thống, và không sụp đổ).

#### 8.8 Sự an toàn

Tiểu mục này quy định các yêu cầu sau đây liên quan đến mức độ mà hệ thống không trực tiếp hoặc gián tiếp (ví dụ, thông qua không hoạt động) gây ra thiệt hại do tai nạn đối với người hoặc tài sản (ví dụ như, mất tiền hoặc dữ liệu).

• SAF-1) Các CRM sẽ không vô tình bị mất thông tin tài khoản người dùng.

#### 9.1 Những hạn chế phần cứng

Các tài liệu tiểu mục tất cả các yêu cầu hạn chế liên quan đến phần cứng tối thiểu hoặc thực tế.

#### 9.2 Phần mềm hạn chế

Các tài liệu tiểu mục tất cả các yêu cầu hạn chế phần mềm.

#### 9.3. High-Level Ngôn ngữ

Các tài liệu tiểu mục tất cả các yêu cầu hạn chế thiết kế kết hợp với việc sử dụng các ngôn ngữ lập trình cấp cao.