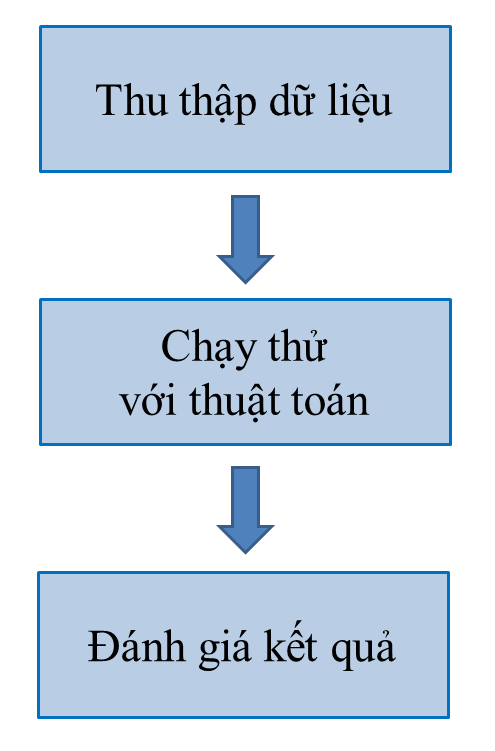
## Thực nghiệm và đánh giá:

Quá trình thực nghiệm và đánh giá được nhóm tác giả thực hiện nhằm mục đích kiểm tra khả năng khuyến nghị của hệ thống trong thực tế với phương pháp đã được trình bày ở chương 3. Toàn bộ quá trình được minh họa khái quát qua hình vẽ sau đây:



Hình 6.13: Mô hình thực nghiệm.

Quá trình gồm ba bước chính là thu thập dữ liệu, chạy thử với thuật toán và cuối cùng là đánh giá kết quả. Cụ thể từng bước được thực hiện chi tiết như sau:

### Thu thập dữ liệu:

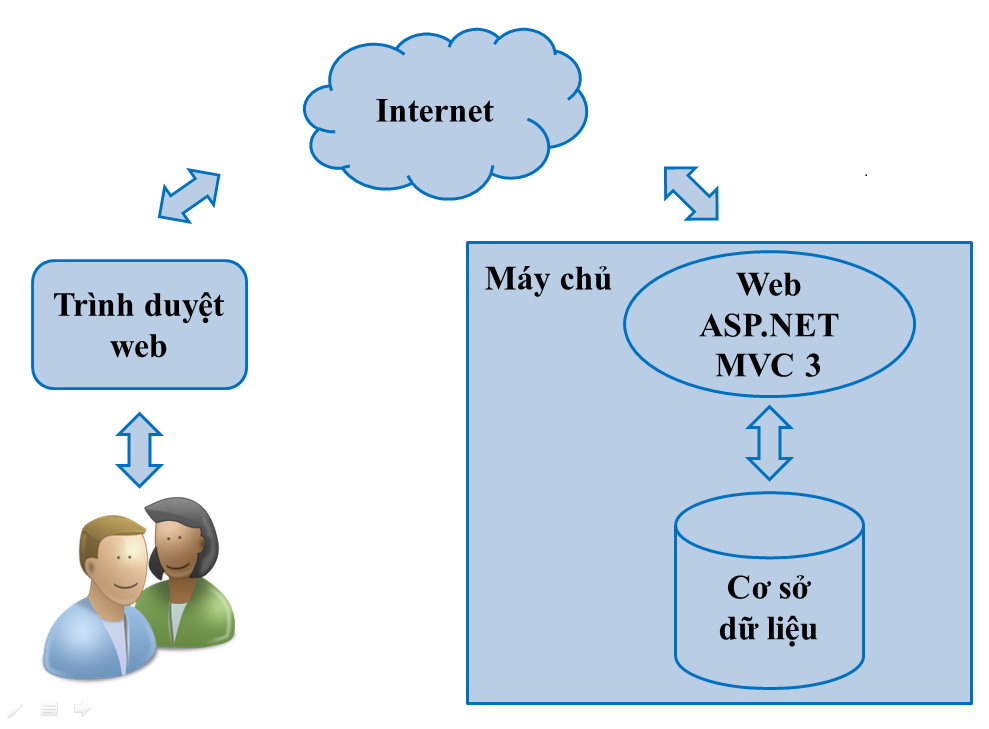
Như đã trình bày ở các chương trước, hệ thống khuyến nghị du lịch được nhóm tác giả xây dựng nhằm mục đích gợi ý những địa điểm du lịch đến người dùng tùy theo điều kiện ngữ cảnh của họ. Để gợi ý được, hệ thống cần có cơ sở dữ liệu các chỉ số đánh giá về các địa điểm du lịch của những người đã đi trước đó. Nhóm tác giả đã tìm kiếm trên Internet nhưng không tìm thấy tập dữ liệu mẫu nào có chứa các điều kiện ngữ cảnh thích hợp có thể sử dụng. Vì vậy, nhóm tác giả đã quyết định tự thu thập dữ liệu từ những người dùng thực tế.

Dữ liệu được thu thập cần chứa những thông tin sau đây, trong đó có các thông tin ngữ cảnh đã được mô tả ở Chương 1:

* Email: email người dùng được dùng để phân biệt những người dùng khác nhau trong hệ thống.
* Địa điểm du lịch: thông tin địa điểm du lịch người dùng đi.
* Thời tiết: thời tiết khi đi du lịch đến địa điểm được chọn.
* Kinh phí du lịch: kinh phí cho chuyến đi.
* Bạn đồng hành: bạn đồng hành trong chuyến đi.
* Thời gian du lịch: thời gian thực hiện chuyến đi.

Do thời gian hạn hẹp nên nhóm tác giả quyết định chỉ tập trung vào 20 địa điểm nổi bật nhất trong thành phố Hồ Chí Minh được đa số người dùng biết đến. Để đạt được lượng dữ liệu đánh giá cần thiết, nhóm thực hiện cùng lúc hai hình thức thu thập dữ liệu là thu thập qua web và thu thập qua ứng dụng Android. Trước khi tiến hành triển khai ứng dụng thực tế và thu thập, nhóm tác giả phải thực hiện việc cài đặt một máy chủ tại nhà riêng, mở port cho modem, thiết lập các thông số cấu hình cho [Internet Information Services](http://en.wikipedia.org/wiki/Internet_Information_Services) (IIS) phiên bản 7.5, cơ sở dữ liệu SQL Server phiên bản 2008 kèm các công cụ khác như OLAP Cube, Data Warehouse …

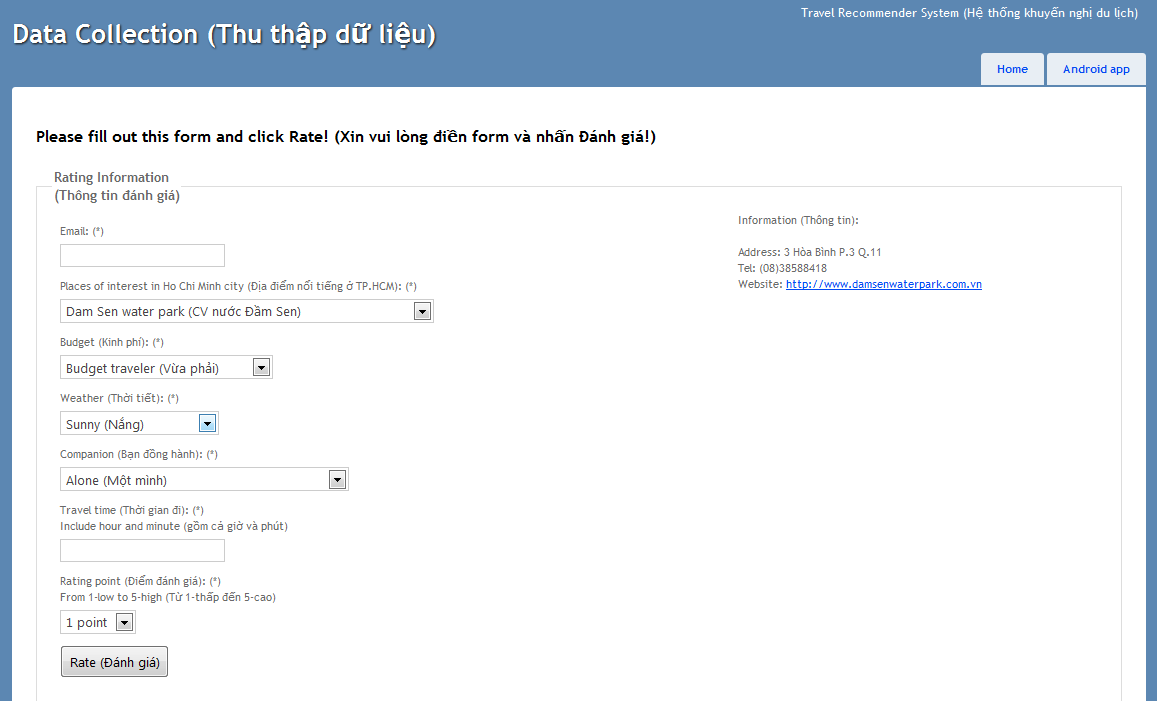
Thu thập qua web: nhóm tác giả xây dựng nhanh một trang web dùng công nghệ [ASP.NET MVC 3](http://www.asp.net/mvc/mvc3) của Microsoft. Mô hình thu thập dữ liệu trên web và đưa dữ liệu vào cơ sở dữ liệu trên máy chủ như sau:



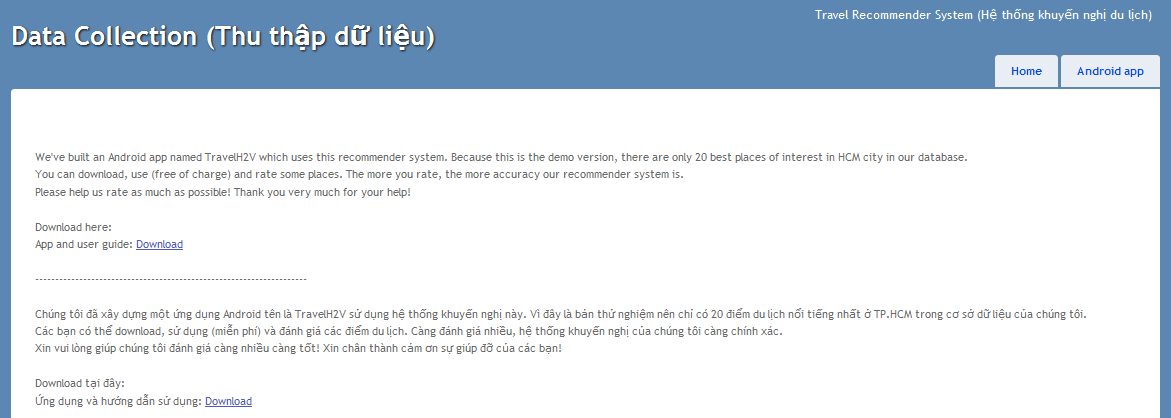
Hình 6.14: Thu thập dữ liệu qua web.

Người dùng truy cập vào trang web thu thập dữ liệu của nhóm, cung cấp các thông tin được yêu cầu. Thông tin được gửi từ trình duyệt web thông qua Internet đến máy chủ và được cập nhật xuống cơ sở dữ liệu. Trang web cũng bố trí một tab cho phép người dùng tải về ứng dụng Android nhóm tác giả xây dựng.

Sau đây là một số hình ảnh của trang web thu thập dữ liệu:

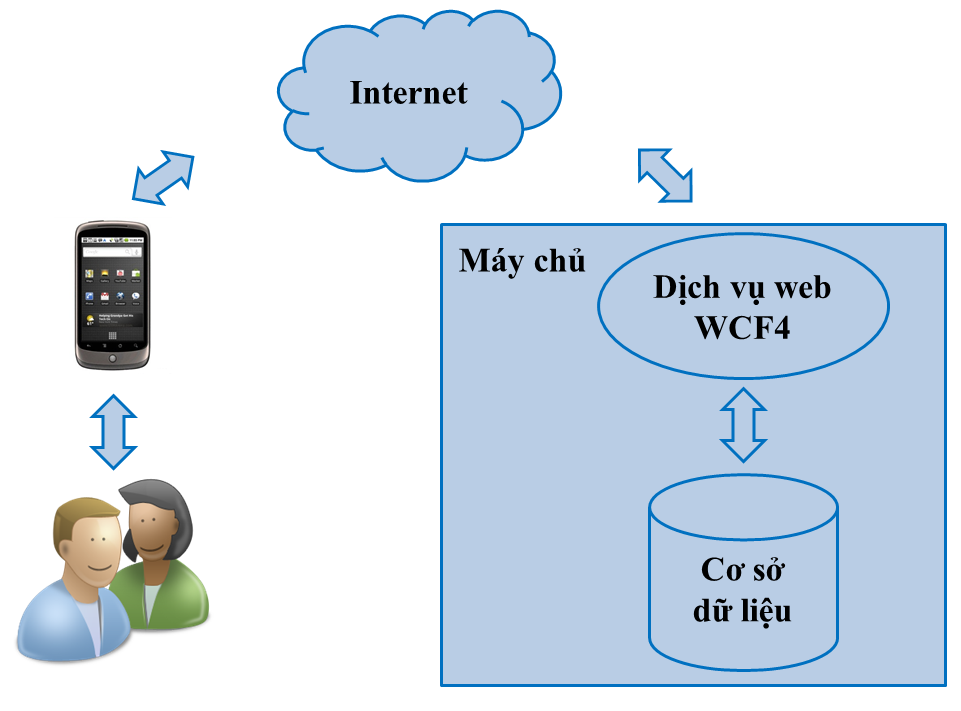


Hình 6.15: Trang thu thập dữ liệu.



Hình 6.16: Trang tải về ứng dụng Android.

Thu thập qua ứng dụng Android: phần trước chúng tôi đã giới thiệu ứng dụng Android. Trong đó, tab Context cho phép người dùng chọn điều kiện ngữ cảnh, email người dùng sẽ là tài khoản Google được thiết lập trên điện thoại. Khi người dùng vào xem thông tin chi tiết địa điểm và đánh giá, những thông đó sẽ được gửi lên Internet đến máy chủ, thông qua dịch vụ web WCF4, được cập nhật xuống cơ sở dữ liệu.



Hình 6.17: Thu thập dữ liệu qua ứng dụng Android.

### Chạy thử với thuật toán:

Ở phần này, nhóm tác giả sẽ tiến hành chạy thử thuật toán theo hai hướng: một là thử nghiệm với tập dữ liệu không chứa các điều kiện ngữ cảnh (dùng tập dữ liệu Movielens có sẵn trên Internet), hai là thử nghiệm với tập dữ liệu chứa các điều kiện ngữ cảnh (dùng tập dữ liệu du lịch được thu thập thực tế ở trên). Qua đó, nhóm tác giả muốn đánh giá độ chính xác của thuật toán trong mỗi trường hợp cũng như đánh giá xem điều kiện ngữ cảnh ảnh hưởng thế nào đến việc khuyến nghị. Kết quả khuyến nghị sẽ ra sao khi có và khi không có điều kiện ngữ cảnh kèm theo?

1. **Thử nghiệm với tập dữ liệu Movielens:**

Movielens là một tập dữ liệu được nhóm nghiên cứu GroupLen (thuộc ngành Khoa Học Máy Tính trường Đại Học Minnesota tại Mỹ) thu thập được từ trang web <http://movielens.umn.edu>. Tập dữ liệu này là những đánh giá của rất nhiều người dùng đối với các bộ phim khác nhau. Có 3 tập với kích cỡ khác nhau có thể được tải về từ trang web <http://www.grouplens.org/>:

* [MovieLens 100k](http://www.grouplens.org/system/files/ml-100k.zip) – gồm 100,000 đánh giá từ 1000 người dùng đối với 1700 phim.
* [MovieLens 1M](http://www.grouplens.org/system/files/ml-1m.zip) – gồm 1 triệu đánh giá từ 6000 người dùng đối với 4000 phim.
* [MovieLens 10M](http://www.grouplens.org/sites/www.grouplens.org/external_files/data/ml-10m.zip) – gồm 10 triệu đánh giá và 100,000 tag applications applied to 10,000 phim bởi 72,000 người dùng.

Nhóm tác giả chọn tập dữ liệu MovieLens 10M để thử nghiệm. Do đây là tập dữ liệu chỉ có hai chiều (người dùng và phim), không có các chiều ngữ cảnh nên nhóm tác giả chỉ sử dụng mô hình hồi quy trong phương pháp khuyến nghị hai chiều được đề cập ở chương 3, không dùng đến kỹ thuật thu giảm số chiều.

Tiêu chí đánh giá được lựa chọn là Mean Absolute Error (MAE). Với MAE, ta có thể dễ dàng hiểu và nhận thấy được độ sai lệch trong kết quả dự đoán của thuật toán. Ví dụ MAE = 1 nghĩa là thuật toán có khả năng dự đoán các chỉ số với độ chính xác (hay còn gọi là sai số) là .

Để thử nghiệm, nhóm tác giả dùng kỹ thuật Cross Validation với n = 10 đã được trình bày ở chương 3. Đầu tiên, tập dữ liệu MovieLens được chia ngẫu nhiên ra thành 10 phần bằng nhau (mỗi phần có 1 triệu đánh giá). Sau đó, nhóm tác giả tiến hành chạy thử thuật toán 10 lần:

* Lần 1: tập dữ liệu đánh giá là phần 1, tập dữ liệu huấn luyện là các phần còn lại 2, 3, 4, 5, 6, 7, 8, 9, 10.
* Lần 2: tập dữ liệu đánh giá là phần 2, tập dữ liệu huấn luyện là các phần còn lại 1, 3, 4, 5, 6, 7, 8, 9, 10.
* Lần 3: tập dữ liệu đánh giá là phần 3, tập dữ liệu huấn luyện là các phần còn lại 1, 2, 4, 5, 6, 7, 8, 9, 10.
* …
* Lần 10: tập dữ liệu đánh giá là phần 10, tập dữ liệu huấn luyện là các phần còn lại 2, 3, 4, 5, 6, 7, 8, 9, 10.

Ở mỗi lần chạy, chỉ số MAE được tính toán và ghi nhận. Sau 10 lần, nhóm tác giả sẽ lấy trung bình cộng và được chỉ số MAE trung bình. Cụ thể, bảng kết quả như sau:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Lần | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| MAE |  |  |  |  |  |  |  |  |  |  |

MAE trung bình = … = 0.85.

1. **Thử nghiệm với tập dữ liệu du lịch thu thập thực tế:**

Phương pháp thử nghiệm tương tự với tập dữ liệu MovieLens là dùng kỹ thuật Cross Validation với n = 10 và chỉ số MAE. Nhưng ở phần này nhóm tác giả sử dụng tập dữ liệu du lịch thu thập thực tế (có chứa các điều kiện ngữ cảnh là thời gian, thời tiết, bạn đồng hành, kinh phí). Không gian khuyến nghị ở đây không còn là hai chiều nữa mà trở thành nhiều nhiều. Do đó, nhóm tác giả dùng pương pháp thu giảm số chiều kết hợp mô hình hồi qui đã được trình bày ở chương 3 để tiến hành thử nghiệm.

Pha thứ nhất của phương pháp là xác định các phân khúc dữ liệu mạnh. Để thực hiện điều này, nhóm tác giả tiến hành áp dụng thuật toán hồi qui trên phân khúc đó và trên toàn bộ dữ liệu (cùng một bộ dữ liệu kiểm thử) để so sánh đánh giá. Sau khi có được phân khúc dữ liệu mạnh, pha thứ hai là dự đoán các chỉ số đánh giá và trả về kết quả cho người dùng. Kết quả của thử nghiệm này được thể hiện qua hai bảng sau:

Bảng tổ hợp tất cả các phân khúc dữ liệu có thể có:

|  |  |  |
| --- | --- | --- |
| Phân khúc dữ liệu | Vùng dữ liệu | MAE |
| … | Phân khúc | … |
| Toàn bộ dữ liệu | … |
| … | Phân khúc | … |
| Toàn bộ dữ liệu | … |
| … | Phân khúc | … |
| Toàn bộ dữ liệu | … |

Bảng các phân khúc dữ liệu mạnh:

|  |  |  |
| --- | --- | --- |
| Phân khúc dữ liệu | Vùng dữ liệu | MAE |
| … | Phân khúc | … |
| Toàn bộ dữ liệu | … |
| … | Phân khúc | … |
| Toàn bộ dữ liệu | … |
| … | Phân khúc | … |
| Toàn bộ dữ liệu | … |

Qua đó, ta thấy …

### Đánh giá kết quả:

Qua hai đợt thử nghiệm trên, nhóm tác giá có đánh giá và nhận xét chung như sau: …