

HƯỚNG DẪN GẮN NHÃN

Mục Lục

| | |
|---|----------|
| Mục Lục..... | 1 |
| 1. Định nghĩa bài toán..... | 2 |
| 2. Định nghĩa các nhãn..... | 2 |
| 2.1. Các nhãn về entity của khách hàng..... | 2 |
| 2.2 Các nhãn về Attribute của bình luận khách hàng..... | 3 |
| 2.3. Các nhãn về cảm xúc của khách hàng..... | 3 |
| 3. Chi tiết và lưu ý gắn nhãn..... | 4 |
| 3.1. Chi tiết về các nhãn..... | 4 |
| 3.1.1. Hotel..... | 4 |
| 3.1.2. Rooms:..... | 5 |
| 3.1.3. Room-amenities:..... | 5 |
| 3.1.4. Facilities:..... | 6 |
| 3.1.5. Service:..... | 6 |
| 3.1.6. Location:..... | 7 |
| 3.1.7. Food & Drinks:..... | 7 |
| 3.2. Lưu ý..... | 8 |
| 4. Giao diện gắn nhãn..... | 8 |

1. Định nghĩa bài toán

Trong mỗi phản hồi của khách hàng về khách sạn đều có thể nói về nhiều vấn đề, với nhiều quan điểm khác nhau nên bài toán sẽ là phân tích đánh giá, quan điểm của khách hàng sau khi sử dụng dịch vụ của khách sạn. Input sẽ là bình luận của khách hàng, Output sẽ là khía cạnh mà khách hàng nhắc đến như chất lượng phòng, thái độ nhân viên,... và cảm xúc của khách hàng về khía cạnh đó.

Bảng 1: Ví dụ về bài toán

| INPUT | OUTPUT |
|--|--|
| Phòng ốc rộng rãi và thoáng. Nhân viên phục vụ rất tận tình, chu đáo. Nói chung là ok. Nếu đưa ra khuyết điểm thì có lẽ là lỗi đi vào buổi tối hơi khó vì gần parking. | ROOMS: positive ROOMS#COMFORT: positive SERVICES#GENERAL: positive HOTEL#GENERAL: positive HOTEL# MISCELLANEOUS: negative |
| Khách sạn giá rẻ, gần biển. Nhân viên lễ tân thiếu lịch sự với khách, thái độ khó chịu. | HOTEL#PRICES: positive LOCATION#GENERAL: positive SERVICES#GENERAL: negative |

2. Định nghĩa các nhãn

Nhãn về khía cạnh của bình luận của khách hàng sẽ là kết hợp giữa entity-attribute

2.1. Các nhãn về entity của khách hàng.

- **HOTEL:** đề cập chung đến khách sạn về giá, hoặc thiết kế, sánh không nhắc đến feature gì phòng cụ thể bên dưới. Các bình luận ko liên quan sẽ quy về **Hotel#Neutral**
- **ROOMS:** đề cập đến phòng khách sạn về các đặc điểm như kích thước, điều kiện phòng nói chung, đồ nội thất, giường ngủ. Hoặc liên quan đến như nhận phòng sai sự thật, sai so với quảng cáo...
- **ROOMS_AMENITIES:** đề cập đến các tiện nghi trong phòng như điều hòa, máy nước nóng-lạnh, tủ lạnh, lò vi sóng, quầy bar, máy sấy tóc, TV, nhà vệ sinh, ban công, máy pha cà phê, nệm, giấy vệ sinh, móc treo đồ,....
- **FACILITIES:** đề cập tới một số tiện nghi của khách sạn được phục vụ kèm theo hoặc cung cấp cho khách hàng như: Netflix , hồ bơi, restaurants, café, parking, information desk, in-room dining, wifi, dịch vụ thuê xe, thang máy...

- **FOOD&DRINKS:** đề cập tới ăn sáng, thức ăn, đồ uống nói chung, hoặc các loại thức ăn, đồ uống cụ thể nào đó chất lượng của nhà hàng bên trong khách sạn.

- **SERVICE:** đề cập tới thái độ phục vụ của nhân viên, tính kịp thời, dễ dàng trong giải quyết vấn đề phát sinh, hoặc trong việc tiếp đón, nhận phòng, trả phòng.

- **LOCATION:** Nói về **vị trí, địa điểm**, gần các địa điểm xung quanh.

2.2 Các nhãn về Attribute của bình luận khách hàng

- **GENERAL:** đề cập tới một loại thực thể nói chung (hotel, room amenities, rooms, facilities, location, service), view của khách sạn, view của phòng.
- **PRICES:** đề cập tới giá phòng, giá thức ăn, đồ uống, giá tiện ích, dịch vụ cung cấp bởi khách sạn, hoặc giá khách sạn nói chung.
- **DESIGN&FEATURES:** đề cập tới thiết kế, trang trí, kích thước của một loại thực thể (hotel, rooms, facilities).
- **CLEANLINESS:** đề cập tới sự vệ sinh, gọn gàng của phòng, của khu vực chung, và của khách sạn nói chung.
- **COMFORT:** nói về các thực thể ở góc độ sự thoải mái, tiện nghi cho khách hàng.
- **QUALITY:** nói về chất lượng đồ ăn, thức uống (ví dụ khẩu vị, sự tươi ngon, nhiệt độ, sự chuẩn bị, v.v), hoặc chất lượng các dịch vụ và tiện ích khách sạn cung cấp.
- **STYLE&OPTIONS:** nói về việc trình bày đồ ăn, thức uống, cách thức phục vụ, các lựa chọn đồ ăn, thực đơn, hoặc tính đa dạng của đồ ăn thức uống phục vụ trong nhà hàng.
- **MISCELLANEOUS:** cho các attributes không nằm trong những cái ở trên

2.3. Các nhãn về cảm xúc của khách hàng.

- **Positive:** Các nhận xét tích cực, khen về các khía cạnh bên trên của khách sạn.
- **Negative:** Các nhận xét tiêu cực, chê về các khía cạnh bên trên của khách sạn.
- **Neutral:** Các nằm giữa tích cực và tiêu cực hoặc không có nhắc đến việc tích cực khen chê về các khía cạnh bên trên của khách sạn.

3. Chi tiết và lưu ý gắn nhãn

3.1. Chi tiết về các nhãn

3.1.1. Hotel

Nhắc đến **khách sạn** nói chung hoặc đề cập đến **giá cả**, có “khách sạn, resort, nhà nghỉ, căn hộ, ...” là thực thể (**ngoại trừ đề cập đến địa điểm, vị trí** của khách sạn).

- Nhãn **Hotel#General**: Không có thực thể rõ ràng, nhận xét chung về khách sạn.
- Nhãn **Hotel#Prices**: Nói về giá của khách sạn.
- Nhãn **Hotel#Designs&Features**:
- Nhãn **Hotel#Cleanliness**:
- Nhãn **Hotel#Comfort**:
- Nhãn **Hotel#Quality**:
- Nhãn **Hotel#Miscellaneous**: Đề cập đến những thứ khác.
 - **Positive**: tốt, đẹp, được, sạch sẽ, ok, thoáng, mát, tuyệt vời,... hoặc nếu đề cập đến **giá cả** như rẻ, tốt, được, hợp lý, rất hợp lý, phải chăng, phù hợp, đáng tiền, khá mềm, vừa phải, hấp dẫn, ok, miễn phí, free...
 - **Negative**: xấu, tồi, tệ, bẩn, chưa được tốt, bí, ẩm thấp, không đẹp, không đầy đủ, cũ, không sang trọng, không sử dụng được... hoặc nếu đề cập đến **giá cả** như đắt, quá đắt, cao, quá cao, hơi cao...
 - **Neutral**: tạm được, bình thường, khá, không có gì đặc sắc,... hoặc có cả 2 ý được đánh giá trái ngược **positive & negative**.

Ví dụ: Khách sạn X mới; Khách sạn sạch sẽ. → **Hotel#Positive**
View đẹp → **Hotel#Positive** còn nếu chỉ cụ thể **Phòng view đẹp** → **Rooms#Positive**

- **Khen, chê, đánh giá mà không có chủ ngữ** (mà không nói về vị trí hay con người).

Ví dụ: Rất sạch sẽ; Nói chung là không đạt tiêu chuẩn 3 sao. → **Hotel#Positive**
Chất lượng chưa xứng đáng với giá tiền → **Hotel#Negative**

- **Những câu có dạng** → **Hotel#General**: **Positive**
 - + Tôi đã ở đây 1 lần trước đây và sẽ quay lại nếu có dịp
 - + Nói chung mình sẽ không ở khách sạn X nếu lần sau tới Đà Nẵng.
 - + Chắc chắn tôi sẽ quay lại khách sạn X trong thời gian tới
 - + Nếu bạn đang tìm kiếm một khách sạn cao cấp thì tôi khuyên bạn nên chọn khách sạn này
 - + Sẽ quay lại lần nữa.

Ví dụ: Giờ giấc thoải mái, thích đi về lúc nào cũng được → **Hotel#Positive**

3.1.2. Rooms:

Có “phòng”, “phòng ngủ” là thực thể đề cập đến chi tiết ở **phần 2.1**.

- Nhãn **Rooms#General**
- Nhãn **Rooms#Prices**
- Nhãn **Rooms#Designs&Features**
- Nhãn **Rooms#Cleanliness**
- Nhãn **Rooms#Comfort**
- Nhãn **Rooms#Quality**
- Nhãn **Rooms#Miscellaneous**
 - **Positive**: tốt, đẹp, được, sạch sẽ, ok, thoáng, mát, tuyệt vời,...
 - **Negative**: xấu, tồi, tệ, bẩn, chưa được tốt, bí, ẩm thấp, không đẹp, không đầy đủ, cũ, không sang trọng, không sử dụng được...
 - **Neutral**: tạm được, bình thường, khá, không có gì đặc sắc,... hoặc có cả 2 ý được đánh giá trái ngược **positive** & **negative**.
- Liên quan đến chất lượng giấc ngủ của khách hàng.

Ví dụ:

Ví dụ: Phòng ở đây rất rộng, Phòng đẹp. → **Rooms#Positive**

3.1.3. Room-amenities:

Tiện nghi, nội thất bên trong phòng. Hoặc chỉ **đồ vật cụ thể**: điều hoà, tivi, tủ lạnh, nhà vệ sinh, giường, đệm, gối... đối với villa có thể là phòng bếp, phòng ăn...

- Nhãn **Room-amenities#General**
- Nhãn **Room-amenities#Prices**
- Nhãn **Room-amenities#Designs&Features**
- Nhãn **Room-amenities#Cleanliness**
- Nhãn **Room-amenities#Comfort**
- Nhãn **Room-amenities#Quality**
- Nhãn **Room-amenities#Miscellaneous**
 - **Positive**: tốt, đẹp, được, sạch sẽ, ok, thoáng, mát, tuyệt vời,...
 - **Negative**: xấu, tồi, tệ, bẩn, chưa được tốt, bí, ẩm thấp, không đẹp, không đầy đủ, cũ, không sang trọng, không sử dụng được...

- **Neutral:** tạm được, bình thường, khá, không có gì đặc sắc,... hoặc có cả 2 ý được đánh giá trái ngược **positive** & **negative**.

Ví dụ: Điều máy lạnh cũ rồi, tiếng kêu to, nước tắm yếu lắm, cần khắc phục.
 → **Room-amenities#Comfort** (Do tiếng máy lạnh ồn),
Room-amenities#Quality (Do nước chảy yếu)
 Toilet không có vòi xịt → **Room-amenities#Designs&Features**

3.1.4. Facilities:

Tiện nghi của khách sạn nói chung. Các **dịch vụ kèm theo**: nhà hàng, cafe, gym, spa, đưa đón, giặt là, trông trẻ, bãi biển của resort, các dịch vụ dọn phòng... **Cơ sở vật chất** khách sạn: bể bơi, thang máy, wifi, vườn, chỗ để xe...

- Nhãn **Facilities#General**
- Nhãn **Facilities#Prices**
- Nhãn **Facilities#Designs&Features**
- Nhãn **Facilities#Cleanliness**
- Nhãn **Facilities#Comfort**
- Nhãn **Facilities#Quality**
- Nhãn **Facilities#Miscellaneous**
 - **Positive:** tốt, đẹp, được, sạch sẽ, ok, thoáng, mát, tuyệt vời,...
 - **Negative:** xấu, tồi, tệ, bẩn, chưa được tốt, bí, ẩm thấp, không đẹp, không đầy đủ, cũ, không sang trọng, không sử dụng được...
 - **Neutral:** tạm được, bình thường, khá, không có gì đặc sắc,... hoặc có cả 2 ý được đánh giá trái ngược **positive** & **negative**.

Ví dụ: Dịch vụ đưa đón thuận tiện/không thuận tiện.

Ví dụ: Wifi ok → **Facilities#Quality**

3.1.5. Service:

- **Kỹ năng, thái độ** phục vụ của nhân viên
- **Giải quyết vấn đề** nhanh chóng, chuyên nghiệp, check in , check out...
- Những từ gắn với yếu tố con người (*thân thiện, vui vẻ, lịch sự, đáng yêu,...*)
 - Nhãn **Service#General**
 - **Positive:** tốt, được, thân thiện, vui vẻ, lịch sự, chuyên nghiệp, ok
 - **Negative:** kém, tồi, không chuyên nghiệp, thiếu chuyên nghiệp, không vui vẻ
 - **Neutral:** tạm được, bình thường + có cả 2 ý được đánh giá trái ngược **positive** & **negative**

Ví dụ:

Ví dụ:

3.1.6. Location:

- Nói về **vị trí, địa điểm**, gần các địa điểm xung quanh.
 - Nhãn **Location#General**
 - **Positive**: trung tâm, gần, tốt, đẹp, ok, thuận lợi.
 - **Negative**: xa, ngõ hẻm,...
 - **Neutral**: tạm được, bình thường hoặc có cả 2 ý được đánh giá trái ngược **positive** & **negative**

Ví dụ:

Ví dụ:

3.1.7. Food & Drinks:

- Tất cả đồ **ăn uống** trong khách sạn, cả ở phòng, ở nhà hàng,... cách thức trình bày, phục vụ, chất lượng, lựa chọn.
 - Nhãn **Food&Drinks#Prices**
 - Nhãn **Food&Drinks#Quality**
 - Nhãn **Food&Drinks#Style&Options**
 - Nhãn **Food&Drinks#Miscellaneous**
 - **Positive**: ngon, tuyệt, hợp khẩu vị, tốt, được, tương đối, đáng tiền, hấp dẫn, ok
 - **Negative**: dở, không ngon, không hợp khẩu vị, mặn/nhạt,
 - **Neutral**: chấp nhận được, tạm ổn, tạm được, bình thường + có cả 2 ý được đánh giá trái ngược **positive** & **negative**

Ví dụ:

Ví dụ:

3.2. Lưu ý

- Các bình luận mang tính **đóng góp ý kiến** về một khía cạnh nào đó sẽ phân vào nhãn **neutral**.

Ví dụ:

- Nếu các ý kiến **không liên quan** tới **khách sạn** hoặc mang tính **spam** chúng tôi cũng sẽ phân vào nhãn **neutral**.

Ví dụ: Tôi có gửi mail nhờ cty đổi lại phòng giường đôi thành giường đơn và đã được cty xác nhận bằng mail là ok nhưng khi nhận phòng thì khách sạn báo là không được quá thất vọng vì nhóm tôi đi tôi toàn là gia đình chưa có bé nhỏ nên mới có nhu cầu nhỏ nhu vậy nhưng công ty không đáp ứng được... (Không đủ nghĩa) → Không cần xét đến

- Các trường hợp “**hơi**” + **1 từ mang tính tích cực** chúng tôi sẽ phân vào nhãn **neutral**.
- Những chi tiết không chỉ rõ thực thể gắn vào nhãn hotel.

Ví dụ: View đẹp → **Hotel#Positive**

View phòng đẹp → **Rooms#Positive**

4. Giao diện gắn nhãn

- Chúng tôi sử dụng Label Studio để phục vụ cho việc gắn nhãn. Giao diện gồm có Comment và các nhãn cần gắn (**hình 2**). Tương ứng với từng khía cạnh chúng tôi sẽ chọn vào cảm xúc phù hợp với khía cạnh đó.

Ví dụ:

Input: rất sạch sẽ, nhân viên cực kỳ dễ thương và nhiệt tình, vị trí cực thuận tiện, lần nào về Việt Nam đều ghé đến đây ở và 1 điều đặc biệt nữa là giấc ngủ tại ks rất ngon, nệm rất êm và yên tĩnh, cảm ơn dịch vụ của các bạn.

Output: **Facilities#General:** **Positive**

Comment

Chung cư mới xây, đầy đủ nội thất, có nhiều tiện ích như hồ bơi, khu vui chơi cho trẻ em,... Phòng này phù hợp với nhóm bạn hoặc 5-6 người

HOTEL

| General | Price | Designs_Features | Cleanliness | Comfort | Quality | Miscellaneous |
|--|--|--|--|--|--|--|
| <input type="checkbox"/> Positive ^[1] | <input type="checkbox"/> Positive ^[4] | <input type="checkbox"/> Positive ^[7] | <input type="checkbox"/> Positive ^[0] | <input type="checkbox"/> Positive ^[e] | <input type="checkbox"/> Positive ^[s] | <input type="checkbox"/> Positive ^[a] |
| <input type="checkbox"/> Negative ^[2] | <input type="checkbox"/> Negative ^[5] | <input type="checkbox"/> Negative ^[8] | <input type="checkbox"/> Negative ^[a] | <input type="checkbox"/> Negative ^[f] | <input type="checkbox"/> Negative ^[d] | <input type="checkbox"/> Negative ^[z] |
| <input type="checkbox"/> Neutral ^[3] | <input type="checkbox"/> Neutral ^[6] | <input type="checkbox"/> Neutral ^[9] | <input type="checkbox"/> Neutral ^[w] | <input type="checkbox"/> Neutral ^[a] | <input type="checkbox"/> Neutral ^[f] | <input type="checkbox"/> Neutral ^[v] |

Rooms

| General | Price | Designs_Features | Cleanliness | Comfort | Quality | Miscellaneous |
|--|--|--|--|-----------------------------------|-----------------------------------|-----------------------------------|
| <input type="checkbox"/> Positive ^[c] | <input type="checkbox"/> Positive ^[v] | <input type="checkbox"/> Positive ^[p] | <input type="checkbox"/> Positive ^[l] | <input type="checkbox"/> Positive | <input type="checkbox"/> Positive | <input type="checkbox"/> Positive |
| <input type="checkbox"/> Negative ^[v] | <input type="checkbox"/> Negative ^[l] | <input type="checkbox"/> Negative ^[l] | <input type="checkbox"/> Negative ^[n] | <input type="checkbox"/> Negative | <input type="checkbox"/> Negative | <input type="checkbox"/> Negative |
| <input type="checkbox"/> Neutral ^[b] | <input type="checkbox"/> Neutral ^[o] | <input type="checkbox"/> Neutral ^[k] | <input type="checkbox"/> Neutral ^[m] | <input type="checkbox"/> Neutral | <input type="checkbox"/> Neutral | <input type="checkbox"/> Neutral |

Hình 2: Ví dụ về gắn nhãn

Tài Liệu Tham Khảo:

[1] VLSP 2018 - Aspect Based Sentiment Analysis (VABSA 2018)

[2] Annotation Guideline