**Đặc tả dữ liệu đầu vào:**

1. *Dữ liệu hành trình một ngày của xe buýt:*

Dữ liệu mô tả sự di chuyển trong một ngày của xe buýt TP HCM (tạm gọi là hành trình). Hành trình được biễu diễn dưới dạng tập hợp các điểm GPS theo tọa độ Latitude và Longitude.

Các điểm này được hộp đen xe buýt ghi lại. Dữ liệu này bao gồm ba cột:

* Lat: Vĩ độ
* Long: Kinh độ
* Receiving time: Thời gian nhận được tín hiệu GPS mà hộp đen xe buýt gửi lên. Tính theo hệ Unix Epoch.

Ví dụ:

|  |  |  |  |
| --- | --- | --- | --- |
| **Lat** | **Long** | **Receiving time** | **Mô tả** |
| 10.82935238 | 106.7039795 | 1472749215 | Xe buýt đã đi qua điểm có:  Vĩ độ: 10.82935238  Kinh độ: 106.7039795  Và hệ thống ghi nhận vào thời gian 1472749215 |

1. *Dữ liệu tuyến xe buýt:*

Dữ liệu mô tả các tuyến xe buýt của TPHCM. Các tuyến được biểu diễn bởi tập hợp các trạm, mỗi trạm được miêu tả bằng một điểm GPS theo tọa độ Latitude và Longitude.

Một dữ liệu tuyến bao gồm các cột sau:

* Route\_Id: Số hiệu của tuyến
* Station\_Id: Số hiệu của trạm (trong toàn bộ các trạm TPHCM)
* Station\_Code: Mã trạm
* Station\_Direction: Hướng trạm (0 – Trạm nằm trên chiều đi, 1 – Tram nằm trên chiều về)
* Station\_Order: Thứ tự của trạm trong mỗi chiều của tuyến
* Station\_Name: Tên trạm
* Station\_Address: Địa chỉ trạm
* Lat: Vĩ độ
* Lng: Kinh độ
* Polyline: Tập hợp các điểm GPS biểu diễn lộ trình di chuyển từ trạm liền trước tới trạm hiện tại.
* Distance: Khoảng cách từ trạm hiện tại tới trạm liền trước (tính theo mét)

**Kết quả cần đạt được:**

* Do tài xế **có thể** giữ xe ở những địa điểm không thuộc tuyến. Khi đến giờ làm việc, tài xế lấy xe từ điểm giữ và chạy tới trạm đầu tiên hoặc chạy ra địa điểm giữ xe sau khi hết giờ làm việc. Nên trong tập dữ liệu hành trình có thể tồn tại những điểm GPS không thuộc tuyến (tạm gọi là những điểm nhiễu), khiến cho việc so trùng có thể bị sai. Vì thế cần phát hiện và loại bỏ những điểm GPS nhiễu này.
* Chỉ ra hành trình của xe buýt tương ứng với tuyến có số hiệu nào. Và độ tương tự là bao nhiêu phần trăm (theo cách thức và công thức tự đề xuất).

**Tham khảo:**

* Biểu diễn trực quan các điểm GPS trên bản đồ:http://www.hamstermap.com/quickmap.php