|  |  |
| --- | --- |
|  |  |
| **HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG**  **---------------------------------------**  **C:\Users\tchai\Desktop\ptit-logo (1).png** | |
|  | |
|  | |
| **TRƯƠNG CÔNG HẢI** | |
|  | |
|  | |
|  | |
| **ĐỀ CƯƠNG**  **LUẬN VĂN THẠC SĨ KỸ THUẬT** | |
|  | |
|  | |
| HÀ NỘI - 2016 | |

Các bước thực hiện xử lý nội dung bài viết

**Bước 1: Chuẩn bị**

Input: Công cụ lập trình, access\_key

Output: File csv là danh sách các status và thông tin cần thiết: DataFacebook\_v2.csv

|  |  |
| --- | --- |
| Phạm vi | Sử dụng mạng xã hội Facebook |
| Đối tượng | Nội dung bài viết: trạng thái của người dụng mạng xã hội  Note:   * Hiện tại chỉ sử dụng status * Không sử dụng comment user |
| Công cụ sử dụng | * Graph API for Facebook * Java 1.8 * Eclipse juno * Restfb-1.33.0.jar, weka.jar library |
| Các lấy nội dung bài viết | Mỗi người dùng sẽ có một key gọi là access\_key  Key này lấy được bằng cách cho phép facebook truy cập thông tin người dùng (có các options cho người dùng)  Khi có key này, chúng ta có thể sử dụng Graph API để lấy status  Note: key tồn tại tầm 1 đến 2 tiếng.  Tạo 1 project java để lấy status xử lý và xuất ra file CSV |
| Định dạng | Sẽ lấy một danh sách status của người dùng cùng với thông tin cần thiết:  ID, name, gender, status’s ID, status  1 người dùng có nhiều status, tập hợp tất cả status vào 1 file định dạng CSV: **datafacebook\_v2.csv** |
| Tiền xử lý | Trước khi status lưu vào file chúng ta cần xử lý một số vấn đề:   * Loại bỏ các link (http, https…) ra khỏi status * Các từ cách nhau bởi 1 dấu cách (loại bỏ các dấu cách gần nhau, xuống dòng, tab… bằng 1 khoảng trắng) |

**Bước 2: Xử lý**

Input: File datafacebook\_v2.csv

Output: File REFF FacebookData.arff

**Bước 3: Chạy chương trình**

Input: File REFF FacebookData.arff

Output: Thông tin kết quả chạy của weka lưu vào file result.txt

**Bước 4: Đưa ra kết quả**

Input: File result.txt

Output: Đánh giá và kết luận đưa vào luận văn