



**SỔ TAY HƯỚNG DẪN**  
**XỬ LÝ DỮ LIỆU THU THẬP NGOÀI THỰC ĐỊA**  
**TỪ MÁY CHỦ TRÊN PHẦN MỀM OPENREFINE**  
**CỦA GOOGLE**

**Hà Nội, tháng 2 năm 2017**

## LỜI NÓI ĐẦU

Cuốn sổ tay này trình bày chi tiết cách thức xử lý dữ liệu trên phần mềm Open Refine của Google sau khi nhận dữ liệu được chiết xuất từ ODK Aggregate từ tổ quản trị hệ thống. Những nội dung trong các phần “Lưu ý” nhằm nhắc nhở người sử dụng cần phải chú ý, và cần phải nhớ các nội dung này trong quá trình thao tác trên phần mềm.

Các hình ảnh được sử dụng trong sổ tay này là các hình ảnh được sao chép khi sử dụng trên phần mềm Open Refine.

Khi đọc tài liệu này, cách tốt nhất để nắm vững các nội dung là vừa đọc vừa thao tác trên ứng dụng web.

### Giải thích các từ viết tắt, các từ tiếng Anh

| Từ viết tắt/tiếng Anh | Giải thích   |
|-----------------------|--|
| 3D                    | Chế độ không gian ba chiều   |
| 3D Fix                | Đã thu được tín hiệu (GPS) để hiển thị theo 3 chiều                      |
| Accuracy              | Sai số (GPS)   |
| Android               | Tên gọi của hệ điều hành cho các thiết bị di động                        |
| Camera                | Tên của ứng dụng chụp ảnh  |
| CSDL                  | Cơ sở dữ liệu  |
| DVMTR                 | Dịch vụ môi trường rừng  |
| Feets                 | Đơn vị đo chiều dài (1 feet = 0,3048 mét)                                |
| Geo ODK               | Tên của ứng dụng di động thu thập dữ liệu thực địa                       |
| Geotagging            | Gán dữ liệu địa lý   |
| GPS                   | Hệ thống định vị toàn cầu dân dụng của Mỹ                                |
| GPS Test              | Tên của một ứng dụng máy tính bảng dùng để xem dữ liệu tọa độ            |
| In Use                | Thu được tín hiệu (số vệ tinh GPS)                                       |
| In View               | Quan sát được (số vệ tinh GPS)   |
| Inches                | Đơn vị đo chiều dài (1 inch = 2,54 cm).                                  |
| Internet              | Hệ thống thông tin toàn cầu gồm các mạng máy tính được liên kết với nhau |
| MBTiles               | Một định dạng bản đồ dành cho máy tính bảng                              |
| Meters                | Đơn vị đo chiều dài (mét)  |
| No Fix                | Chưa thu được tín hiệu (GPS) để hiển thị theo 3 chiều                    |
| Status                | Tình trạng (tín hiệu GPS)  |
| True North            | Cực bắc của trái đất tính theo trục                                      |
| VNFF                  | Quỹ bảo vệ và phát triển rừng Việt Nam                                   |

## Mục lục

|   |    |
|---|----|
| PHẦN I. TỔNG QUAN VỀ PHẦN MỀM .....                         | 5  |
| 1. Tổng quan về phần mềm OpenRefine.....                    | 5  |
| 2. Ý tưởng.....   | 5  |
| 3. Tải phần mềm.....  | 6  |
| PHẦN II. CÁC TÍNH NĂNG VÀ CÁCH SỬ DỤNG CỦA OPENREFINE ..... | 8  |
| 1. Nhập dữ liệu (Importing Data).....                       | 8  |
| 2. Lọc dữ liệu/Cắt dữ liệu (Filtering/Faceting) .....       | 10 |
| 3. Chỉnh sửa hàng (Editing cells) .....                     | 15 |
| 4. Chỉnh sửa cột (Editing columns).....                     | 22 |
| 5. Hoán vị hàng và cột (Transposing rows and columns) ..... | 25 |
| 6. Xuất dữ liệu (Exporting Data).....                       | 27 |
| 7. Quay lại/Làm lại (Undo/Redo).....                        | 27 |

## PHẦN I. TỔNG QUAN VỀ PHẦN MỀM

### 1. Tổng quan về phần mềm OpenRefine

OpenRefine là công cụ rất mạnh trong xử lý với các dữ liệu chưa được chuẩn hóa trong việc: làm sạch dữ liệu, chuyển đổi dữ liệu từ một định dạng sang các định dạng khác nhau; và được phát triển với dịch vụ web và các dữ liệu bên ngoài. OpenRefine cũng cho phép chuyển đổi dữ liệu từ dạng không có cấu trúc sang dạng dữ liệu có cấu trúc. Là một công cụ mã nguồn mở nên OpenRefine có thể được sử dụng cho nhiều dự án khác nhau và có rất nhiều chức năng như cắt, gộp, sửa các ô bảng tính, điều hòa, mở rộng các dịch vụ web. OpenRefine dễ dàng sử dụng như công cụ excel của Microsoft và rất hữu dụng như cơ sở dữ liệu của phần mềm Access. Nó cho phép thao tác dễ dàng và giúp phân tích dữ liệu thông qua lọc, cắt gọt và chuyển đổi dữ liệu.

Ưu điểm của OpenRefine

- ✓ OpenRefine là một ứng dụng desktop. Nó mở bằng trình duyệt như một giao diện web trên chính máy tính cá nhân. Chính vì vậy, dữ liệu an toàn và không đăng tải lên máy chủ của Google.
- ✓ Nó cho phép cắt gọt và lọc dữ liệu trong các tập con dữ liệu chính vì thế có thể lựa chọn và tổ chức dữ liệu theo ý muốn.
- ✓ Được sử dụng với trình duyệt cơ bản vì vậy có thể thao tác nhiều dữ liệu một cách hiệu quả
- ✓ Có tính năng mạnh cho các dữ liệu mở rộng, người dùng có thể sử dụng để tìm kiếm các dữ liệu mô tả và sử dụng các dữ liệu tương quan khác.

Nhược điểm

- ✓ Giao diện người sử dụng không được thân thiện mặc dù các tính năng và các chức năng rất mạnh.
- ✓ Đáng tiếc là Google đã không phát triển tiếp nên chỉ có một vài chức năng cơ bản không được phong phú.

### 2. Ý tưởng

Ý tưởng của việc chọn OpenRefine như một công cụ là cung cấp tới người sử dụng một công cụ mở miễn phí để có thể xử lý dữ liệu. OpenRefine cung cấp một cách linh hoạt với nhiều chức năng xử lý dữ liệu khác nhau và rất thân thiện với người sử dụng. Người dùng có thể sử dụng công cụ này để thấy một cái nhìn tổng quan về dữ liệu của mình có liên quan đến thống kê về số lượng. Họ cũng có thể xử lý dữ liệu lớn mà không cần lo lắng về các nguy cơ mất dữ liệu vì có thể khôi phục lại dữ liệu tại bất kỳ thời điểm nào. Làm sạch, chuyển đổi, liên kết với các đường dữ liệu cho một bộ dữ liệu có thể dễ dàng ứng dụng tải về trong hệ thống.

**Cái gì** – Một bộ dữ liệu lộn xộn, không có cấu trúc, không phù hợp có thể được khám phá sử dụng bởi sự tinh lọc mở. Nhìn chung, nó rất khó để tìm hiểu dữ liệu thông qua sự dư thừa và sự không phù hợp. Nhưng OpenRefine đưa cho chúng ta các

chức năng thông qua đó có thể lọc dữ liệu, sửa chữa sự không phù hợp và xem dữ liệu. Nó là một công cụ làm sạch dữ liệu.

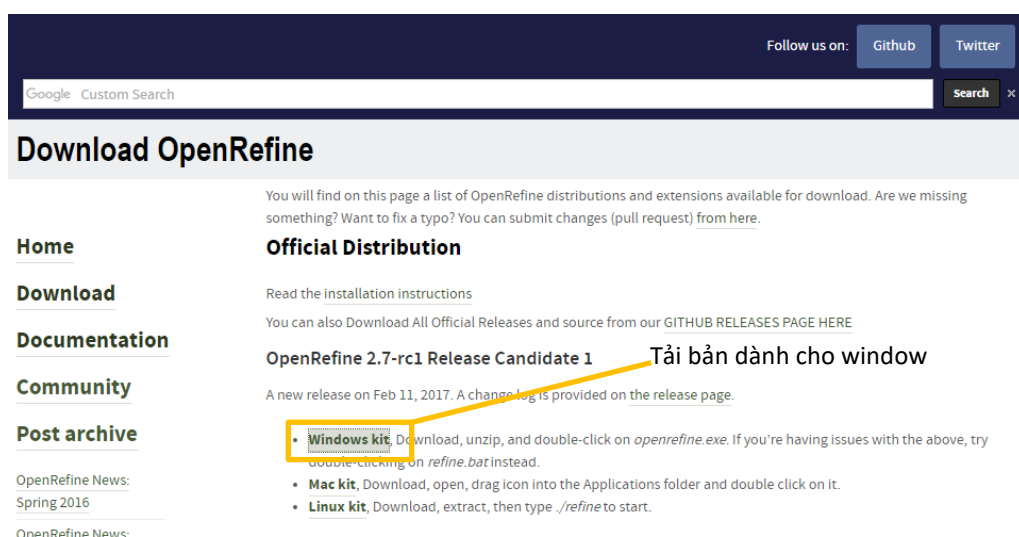
**Tại sao** – Bảng tính Spreadsheet cũng có thể tinh lọc một bộ dữ liệu nhưng Spreadsheets không giống như OpenRefine làm sạch dữ liệu theo cách kiểm soát hệ thống nhiều hơn. Trong khi sử dụng dữ liệu lịch sử, chúng ta có thể gặp các vấn đề như các trường dữ liệu trắng, các bản ghi trùng lặp, các định dạng không phù hợp và sử dụng công cụ OpenRefine có thể giúp giải quyết được vấn đề này.


**Khi nào** – Ngày nay phân tích dữ liệu đóng vai trò quan trọng trong xã hội. Các nhà phân tích dữ liệu cải thiện việc ra quyết định, giảm thiểu chi phí và xác định các cơ hội mới. Sự phân tích dữ liệu là một quá trình kiểm tra, làm sạch, chuyển đổi và mô hình hóa dữ liệu với các mục đích khác nhau cho sự khám phá các thông tin hữu dụng, đề xuất các kiến nghị, và hỗ trợ ra quyết định. Vì vậy để đảm bảo cho độ chính xác của sự phân tích, chúng ta phải làm sạch dữ liệu của chính chúng ta.

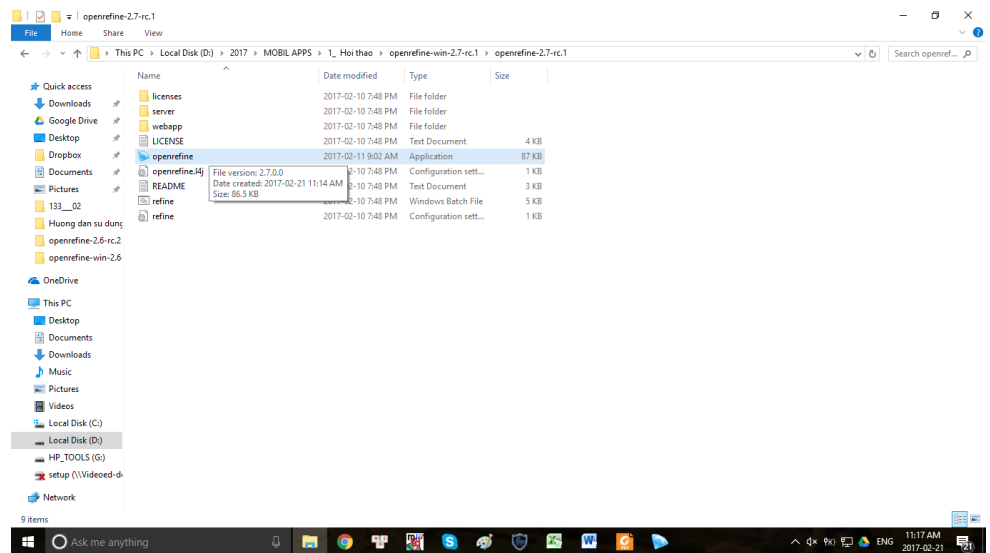
### 3. Tải phần mềm

1. Truy cập đến trang chính thống của tổ chức OpenRefine

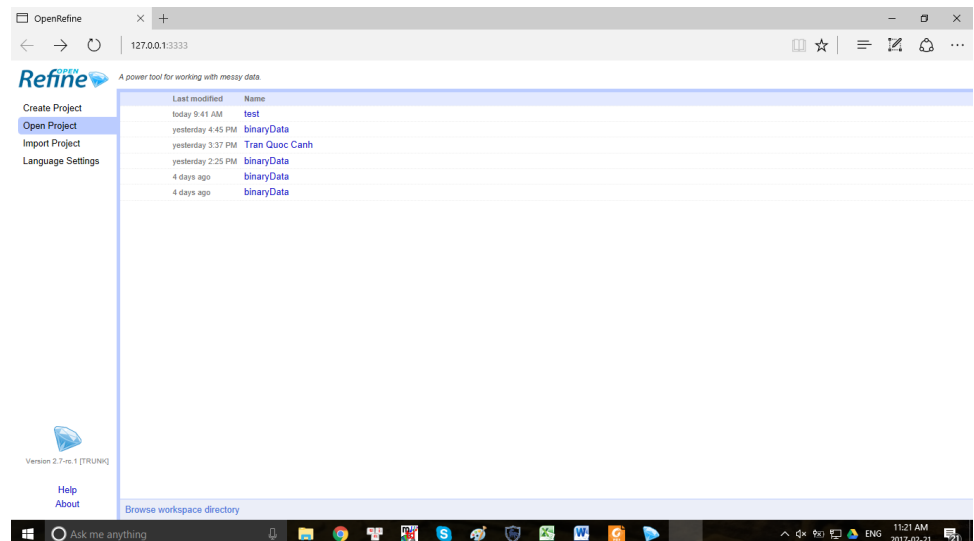
<http://openrefine.org/download.html>



2. Mở thư mục chứa tệp tin tải về. Để khởi động nhấn file chạy có biểu tượng kim cương  để mở trình duyệt. OpenRefine là phần mềm ứng dụng web offline. Chính vì vậy nó sẽ được mở bằng trình duyệt web explorer.



3. Sau khi khởi động xong ứng dụng sẽ mở ra giao diện web như sau:



## PHẦN II. CÁC TÍNH NĂNG VÀ CÁCH SỬ DỤNG CỦA OPENREFINE

Có rất nhiều tính năng trong phần mềm OpenRefine. Tuy nhiên, chúng tôi tập trung vào một số ứng dụng chủ yếu và quan trọng của OpenRefine. Chúng được liệt kê như sau:

- a) Nhập dữ liệu (Importing Data)
- b) Lọc dữ liệu/ Cắt dữ liệu (Filtering/Faceting)
- c) Sửa chữa hàng (Editing cells)
- d) Chỉnh sửa cột (Editing columns)
- e) Xuất dữ liệu (Exporting Data)
- f) Quay lại/Làm lại (Undo/Redo)

### 1. Nhập dữ liệu (Importing Data)

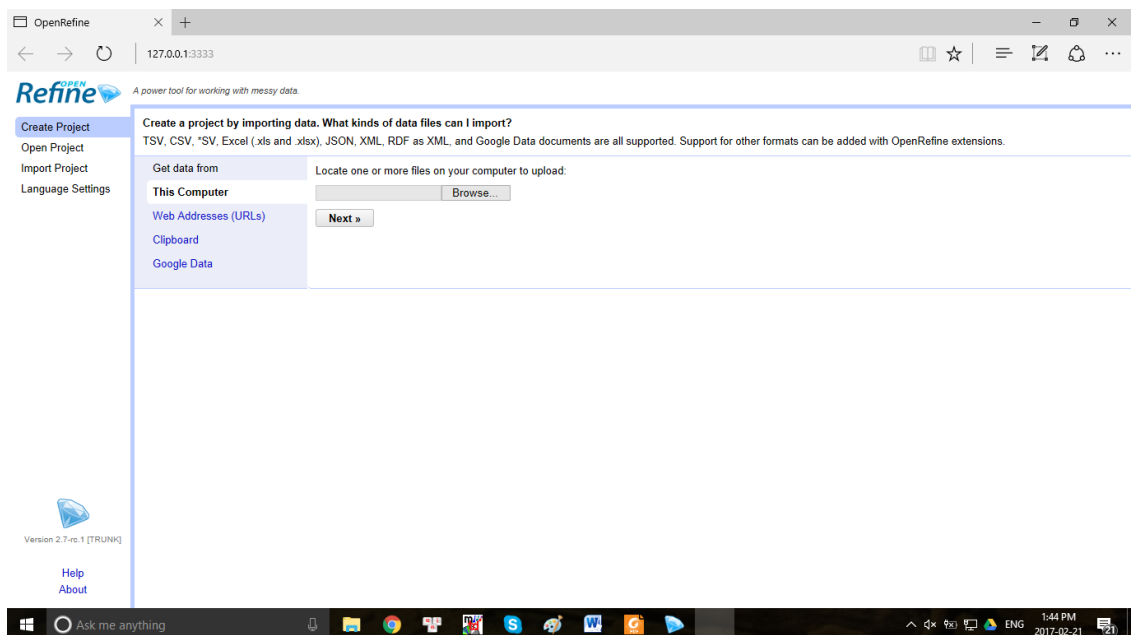
Dữ liệu được nhập từ nhiều nguồn bên ngoài khác nhau. Phần này bao gồm hai phần: tạo đặt tên dự án và phân tích cú pháp dữ liệu.

#### a. Tạo dự án

Mở OpenRefine → ấn vào nhãn tạo dự án “Create project” → Lựa chọn thư mục mà bạn muốn lấy dữ liệu → Nhấn nút để chọn tệp tin “Choose File” → Next → Tạo dự án.

##### Chi tiết:

- ✓ Mở Google Refine
- ✓ Nhấn nút chọn tệp tin để mở tài liệu và chọn tệp tin mà bạn muốn làm việc nếu tệp tin nằm trên máy tính cá nhân
- ✓ Bạn có thể chọn website mà bạn muốn lấy dữ liệu từ trang tin đó hoặc từ clipboard hay thậm chí từ dữ liệu chia sẻ trên google.
- ✓ Chúng ta có thể chỉ ra giao diện của phần mềm như hình dưới.





Có nhiều lựa chọn để tải bộ dữ liệu vào trong OpenRefine. Bên trái màn hình trong ô nội dung có các lựa chọn sau:

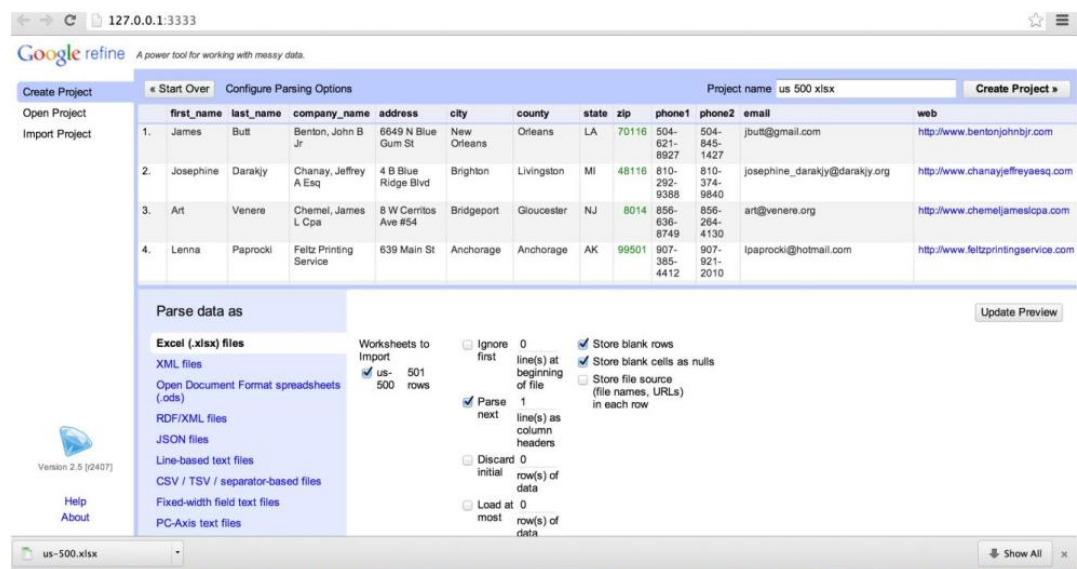
- **This Computer:** Lựa chọn tệp tin được chứa trên máy tính cá nhân
- **Web Addresses (URLs):** Nhập trực tiếp từ một nguồn dữ liệu trực tuyến
- **Clipboard:** Sao chép và dán dữ liệu của bạn vào trong trường dữ liệu dạng ký tự
- **Google Data:** Có thể truy cập đến Google Spreadsheet hoặc Fusion Table (Yêu cầu cần có sự kết nối mạng)

**Lưu ý:** Chúng ta có thể lựa chọn tệp tin từ máy tính dưới các định dạng như: xlsx, tsv, csv, \*sv, xls, json, xml, rdf cũng như các tài liệu google.

- ✓ Nhấn nút Next sau khi lựa chọn tệp tin.
- ✓ Khi tệp tin đã được tải lên, nhập tên tệp tin trong mục “Project name” và kích tạo dự án “Create Project”.

### b. Cấu trúc dữ liệu trong OpenRefine

Dữ liệu sau khi được nhập vào OpenRefine sẽ được thể hiện dưới dạng bảng hàng và cột.



Mỗi cột là một trường thông tin dữ liệu và mỗi hàng là một bản ghi dữ liệu. Nhìn vào màn hình ở dưới bên phải bạn có thể thấy các lựa chọn của cú pháp dữ liệu được chỉ ra như hình bên dưới đây:

|  |   |                              |
|--|---|------------------------------|
| <input type="checkbox"/> Ignore first  | 0 | line(s) at beginning of file |
| <input checked="" type="checkbox"/> Parse next   | 1 | line(s) as column headers    |
| <input type="checkbox"/> Discard initial   | 0 | row(s) of data               |
| <input type="checkbox"/> Load at most  | 0 | row(s) of data               |
| <input checked="" type="checkbox"/> Parse cell text into numbers, dates, ...                               |   |                              |
| <input checked="" type="checkbox"/> Quotation marks are used to enclose cells containing column separators |   |                              |
| <input checked="" type="checkbox"/> Store blank rows   |   |                              |
| <input checked="" type="checkbox"/> Store blank cells as nulls   |   |                              |
| <input type="checkbox"/> Store file source (file names, URLs) in each row                                  |   |                              |

Mặc định dòng đầu tiên của file là dòng tiêu đề của các cột. Tuy nhiên có một số dữ liệu không có tiêu đề hoặc muốn bỏ qua một số các dòng đầu tiên bắt đầu của file thì cần phải đánh dấu vào lựa chọn đầu tiên.

- **Ignore first** : Bỏ qua dòng đầu tiên bắt đầu của file
- **Parse next**: Phân tích cú pháp dòng tiếp theo như tiêu đề của cột
- **Discard initial**: Loại bỏ dữ liệu của dòng
- **Load at most**: Tải dữ liệu nhiều nhất (số dòng mà muốn tải dữ liệu trong file dữ liệu)
- **Quotaion marks are used to enclose cells containing column separators**: Nếu trong tệp dữ liệu có sự phân cách các giá trị giữa các ô bằng tính là dấu ngoặc kép thì tích vào ô này để OpenRefine nhận biết được và phân tách. Ngược lại không có thì không tích.
- **Parse cell text into numbers, dates,...**: Cú pháp này để OpenRefine tự động tìm và thiết lập các cột dữ liệu có định dạng dữ liệu kiểu ký tự, kiểu số hay kiểu ngày tháng
- **Store blank rows**: Trong dữ liệu còn các dòng trống nếu vẫn muốn lưu lại thì tích vào
- **Store blank cells as nulls**: Trong dữ liệu còn các ô trống nếu vẫn muốn lưu lại như một ô có giá trị rỗng thì tích vào
- **Store file source**: Muốn giữ tên file gốc và đường dẫn cho mỗi dòng thì tích vào

## 2. Lọc dữ liệu/Cắt dữ liệu (Filtering/Faceting)

Đây là một phương thức lọc dữ liệu bên trong các tập con để dành cho việc sử dụng. Nó có thể được sử dụng cho dữ liệu dạng ký tự (chữ), dạng số và dạng ngày tháng.

### *Các kiểu cắt gọt*

1. Text (chữ hay gọi là ký tự): Kiểu cắt gọt này giúp lọc các bộ dữ liệu giống nhau trong một nhóm mà nó cho phép dễ sửa chữa và chỉnh sửa dữ liệu.

2. Numeric (dạng số): Kiểu cắt gọt này nhóm các số trong cùng một mức giống nhau. Khi đó chúng ta có thể lựa chọn bất kỳ mức nào cho việc dùng chỉ ra sự liên tiếp của các số.
3. Custom Text Facet: Kiểu cắt này là một kiểu cắt dạng text mà ở đó bạn có thể tách dữ liệu trong cột bằng cách sử dụng biểu thức (`value.split("") [0]`) mà không cần tạo cột mới. Các nhóm sẽ được làm bằng cách tách dữ liệu được sắp xếp bởi sự xác định vị trí của các ký tự.
4. Custom Numerical Facets: Kiểu cắt này cho phép bạn lựa chọn kiểu cắt dạng số. Những giá trị số có thể được nhóm lại theo độ dài của chuỗi ký tự, logs, modulus ...
5. Customized Facets: Có nhiều kiểu cắt có thể lựa chọn được. Chúng bao gồm cắt ký tự, cắt trùng lặp, cắt theo mức giá trị số, cắt độ dài của chữ ký tự, cắt theo mã code ký tự, cắt bởi lỗi, cắt theo loga, cắt bởi các giá trị ô trống...

### Chi tiết

- Faceting-Clustering - Google Chrome

127.0.0.1:3333/project?project=1547120655302

Faceting-Clustering

Facet / Filter Undo / Redo

Extensions: Freebase

500 rows

Show as: rows records Show: 5 10 25 50 rows

« first < previous 1 - 10 next > last »

Using facets and filters

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started?  
[Watch these screencasts](#)

|     | All       | first_name | last_name                | company_name          | address     | city        | county      | state                   | zip          | phone1       | phone2       | email                    | web        |
|-----|-----------|------------|--------------------------|-----------------------|-------------|-------------|-------------|-------------------------|--------------|--------------|--------------|--------------------------|------------|
| 1.  | James     | Butt       | Benton, John B Jr        | 6649 N Blue Gum St    | New Orleans | Orleans     | Facet       | Text facet              |              |              |              | nal.com                  | http://www |
| 2.  | Josephine | Darsky     | Chanay, Jeffrey A Esq.   | 4 B Blue Gum Blvd     | Brighton    | Livingston  | Text filter | Numeric facet           |              |              |              | s_darsky@darasky.org     | http://www |
| 3.  | Art       | Veneri     | Chemel, James L Cpa      | 8 W Centros Ave #54   | Bridgeport  | Gloucester  | Edit cells  | Timeline facet          |              |              |              | ere.org                  | http://www |
| 4.  | Lenna     | Paprocki   | Felz Printing Service    | 639 Main St           | Anchorage   | Anchorage   | Edit column | Scatterplot facet       |              |              |              | @hotmail.com             | http://www |
| 5.  | Donette   | Foler      | Printing Dimensions      | 34 Center St          | Hamilton    | Butler      | Transpose   | Custom text facet...    |              |              |              | foler@cox.net            | http://www |
| 6.  | Simona    | Morasca    | Chapman, Ross E Esq      | 3 McAuley Dr          | Ashland     | Ashland     | Sort...     | Custom numeric facet... |              |              |              | smorasca.com             | http://www |
| 7.  | Mitsue    | Toliner    | Morning Associates       | 7 Eads St             | Chicago     | Cook        | View        | Customized facets       |              |              |              | mitsue_toliner@yahoo.com | http://www |
| 8.  | Leta      | Dillard    | Commercial Press         | 7 W Jackson Blvd      | San Jose    | Santa Clara | Reconcile   | -573-44-14              | 773-924-3500 | 773-924-3500 | 408-813-1105 | leta@hotmail.com         | http://www |
| 9.  | Sage      | Wieser     | Truhler And Truhler Adys | 5 Boston Ave #68      | Sioux Falls | Minnehaha   | CA          | 95111                   | 605-414-2147 | 605-414-2147 | 605-794-4895 | sage_wieser@cox.net      | http://www |
| 10. | Kris      | Manner     | King, Christopher A Esq  | 226 Runamuck Pl #2008 | Baltimore   | Baltimore   | MD          | 21224                   | 410-855-0723 | 410-855-0723 | 410-855-0723 | kris@gmail.com           | http://www |

- ✓ Kích vào Text Facet trong cột “State” và nhóm dữ liệu thành các tập con của một bang “State” như hình dưới. Các nhóm của bang California, Ca, California (ca) được chia ra từng nhóm không phù hợp trong dữ liệu
- ✓ Chúng ta có thể gộp các nhóm thành một nhóm bằng cách bằng cách sửa tên và hợp lại

Google Refine Facets-Clusters

Facet / Filter: state (55 choices)

42 matching rows (500 total)

|    | first_name | last_name  | company_name                  | address               | city          | county        | state | zip   | phone1       | phone2       | email                             | website  |
|----|------------|------------|-------------------------------|-----------------------|---------------|---------------|-------|-------|--------------|--------------|-----------------------------------|----------|
| 38 | Rozela     | Ostrosky   | Parkway Company               | 17 Morena Blvd        | Camarillo     | Ventura       | CA    | 93012 | 805-832-6163 | 805-809-1531 | rozela.ostrosky@ostrosky.com      | http://w |
| 48 | Kanisha    | Waycott    | Schroer, Gene E Esq           | 5 Tomahawk Dr         | Los Angeles   | Los Angeles   | CA    | 90008 | 323-453-2780 | 323-315-7314 | kanisha_waycott@yahoo.com         | http://w |
| 71 | Kalle      | Blackwood  | Rowley Schlingens Inc         | 701 S Harrison Rd     | San Francisco | San Francisco | CA    | 94104 | 415-315-2761 | 415-404-7609 | kalle.blackwood@gmail.com         | http://w |
| 73 | Bobbie     | Rhym       | Smta, Patricia Garly          | 30 W 80th St #1995    | San Carlos    | San Mateo     | CA    | 94070 | 650-528-5783 | 650-811-9032 | brhym@rhym.com                    | http://w |
| 74 | Micaela    | Rhymes     | H Lee Leonard Attorney At Law | 20932 Hedley St       | Concord       | Contra Costa  | CA    | 94520 | 925-647-3298 | 925-522-7798 | micaela_rhymes@gmail.com          | http://w |
| 84 | Dominique  | Dickerson  | E A I Electronic Assocs Inc   | 89 Marquette Ave      | Hayward       | Alameda       | CA    | 94545 | 510-993-3758 | 510-901-7640 | dominique.dickerson@dickerson.org | http://w |
| 91 | Tamara     | Wardrip    | Jewel My Shop Inc             | 4800 Black Horse Pike | Burlingame    | San Mateo     | CA    | 94010 | 650-803-1936 | 650-216-5075 | twardrip@cox.net                  | http://w |
| 92 | Cory       | Gibes      | Chinese Translation Resources | 63649 W Belmont Ave   | San Gabriel   | Los Angeles   | CA    | 91776 | 626-572-1596 | 626-896-2777 | cory.gibes@gmail.com              | http://w |
| 95 | Elvera     | Benimadho  | Tree Musketiers               | 95385 Charity St #940 | San Jose      | Santa Clara   | CA    | 95110 | 408-783-8505 | 408-445-8447 | elvera.benimadho@cox.net          | http://w |
| 96 | Carma      | Vanheusden | Springfield Div Oh Edison Co  | 68556 Central Hwy     | San Leandro   | Alameda       | CA    | 94577 | 510-503-7169 | 510-452-4835 | carma@cox.net                     | http://w |

- ✓ Có hai nhóm California và California. Chúng ta có thể sửa tên thành CA và nó sẽ gộp với nhóm CA, làm phù hợp dữ liệu. Trước khi gộp với nhóm CA có 42 bản ghi.

Google Refine Facets-Clusters

Facet / Filter: state (55 choices)

42 matching rows (500 total)

California

Apply

- ✓ Sau khi sửa tên nhóm thành CA, tổng số bản ghi là 55 như được chỉ ra bên dưới

Google Refine Facets-Clusters

Facet / Filter: state (53 choices)

55 matching rows (500 total)

|    | first_name | last_name | company_name                  | address               | city          | county        | state | zip   | phone1       | phone2       | email                             | website  |
|----|------------|-----------|-------------------------------|-----------------------|---------------|---------------|-------|-------|--------------|--------------|-----------------------------------|----------|
| 22 | Veronika   | Inouye    | C 4 Network Inc               | 6 Greenleaf Ave       | San Jose      | Santa Clara   | CA    | 95111 | 408-540-1785 | 408-813-4592 | vinouye@aol.com                   | http://w |
| 38 | Rozela     | Ostrosky  | Parkway Company               | 17 Morena Blvd        | Camarillo     | Ventura       | CA    | 93012 | 805-832-6163 | 805-809-1531 | rozela.ostrosky@ostrosky.com      | http://w |
| 48 | Kanisha    | Waycott   | Schroer, Gene E Esq           | 5 Tomahawk Dr         | Los Angeles   | Los Angeles   | CA    | 90008 | 323-453-2780 | 323-315-7314 | kanisha_waycott@yahoo.com         | http://w |
| 58 | Shenka     | Seewald   | East Coast Marketing          | 4 Otis St             | Van Nuys      | Los Angeles   | CA    | 91405 | 818-423-4007 | 818-749-8650 | shenka@gmail.com                  | http://w |
| 71 | Kalle      | Blackwood | Rowley Schlingens Inc         | 701 S Harrison Rd     | San Francisco | San Francisco | CA    | 94104 | 415-315-2761 | 415-404-7609 | kalle.blackwood@gmail.com         | http://w |
| 73 | Bobbie     | Rhym      | Smta, Patricia Garly          | 30 W 80th St #1995    | San Carlos    | San Mateo     | CA    | 94070 | 650-528-5783 | 650-811-9032 | brhym@rhym.com                    | http://w |
| 74 | Micaela    | Rhymes    | H Lee Leonard Attorney At Law | 20932 Hedley St       | Concord       | Contra Costa  | CA    | 94520 | 925-647-3298 | 925-522-7798 | micaela_rhymes@gmail.com          | http://w |
| 84 | Dominique  | Dickerson | E A I Electronic Assocs Inc   | 89 Marquette Ave      | Hayward       | Alameda       | CA    | 94545 | 510-993-3758 | 510-901-7640 | dominique.dickerson@dickerson.org | http://w |
| 87 | Stephane   | Barfield  | Beutelschies & Company        | 47154 Whipple Ave NW  | Gardena       | Los Angeles   | CA    | 90247 | 310-774-7643 | 310-988-1219 | stephane@barfield.com             | http://w |
| 91 | Tamara     | Wardrip   | Jewel My Shop Inc             | 4800 Black Horse Pike | Burlingame    | San Mateo     | CA    | 94010 | 650-803-1936 | 650-216-5075 | twardrip@cox.net                  | http://w |

- ✓ Chúng ta có thể sắp xếp các nhóm theo tên, và tính tìm ra nhóm lớn nhất như bên dưới



Faceting-Clustering - Google

127.0.0.1:3333/project?project=1547120655302

Google refine Faceting-Clustering Permalink

Facet / Filter Undo / Redo 2

Refresh Reset All Remove All

67 matching rows (500 total)

Show as: rows records Show: 5 10 25 50 rows

Extensions: Freebase

« first » previous 1 - 10 next » last »

| state | first_name | last_name | company_name                  | address            | city          | county        | state | zip   | phone1       | phone2       | email                             | web        |
|-------|------------|-----------|-------------------------------|--------------------|---------------|---------------|-------|-------|--------------|--------------|-----------------------------------|------------|
| CA 67 | Leota      | Dillard   | Commercial Press              | 7 W Jackson Blvd   | San Jose      | Santa Clara   | CA    | 95111 | 408-752-3500 | 408-813-1105 | leota@hotmail.com                 | http://www |
| NJ 52 | Kiley      | Caldarera | Feiner Bros                   | 25 E 75th St #89   | Los Angeles   | Los Angeles   | CA    | 90034 | 310-498-5651 | 310-254-3084 | kiley.caldarera@aol.com           | http://www |
| NY 46 | Veronika   | Inouye    | C 4 Network Inc               | 6 Greenleaf Ave    | San Jose      | Santa Clara   | CA    | 95111 | 408-540-1785 | 408-813-4592 | vinouye@aol.com                   | http://www |
| TX 32 | Rozella    | Ostrosky  | Parkway Company               | 17 Morena Blvd     | Camarillo     | Ventura       | CA    | 93012 | 805-832-6193 | 805-609-1531 | rozella.ostrosky@ostrosky.com     | http://www |
| PA 29 | Kaniha     | Waycott   | Schroer, Gene E Esq           | 5 Tomahawk Dr      | Los Angeles   | Los Angeles   | CA    | 90006 | 323-453-2780 | 323-315-7314 | kaniha_waycott@yahoo.com          | http://www |
| FL 27 | Shenika    | Seewald   | East Coast Marketing          | 4 Otis St          | Van Nuys      | Los Angeles   | CA    | 91405 | 818-423-4007 | 818-749-8650 | shenika@gmail.com                 | http://www |
| IL 15 | Kalie      | Blackwood | Rowley Schlingens Inc         | 701 S Harrison Rd  | San Francisco | San Francisco | CA    | 94104 | 415-315-2781 | 415-604-7609 | kalie.blackwood@gmail.com         | http://www |
| MD 15 | Bobbye     | Rhym      | Smts. Patricia Garly          | 30 W 80th St #1995 | San Carlos    | San Mateo     | CA    | 94070 | 650-528-5783 | 650-811-9032 | brhym@rhym.com                    | http://www |
| MI 14 | Micaela    | Rhymes    | H Lee Leonard Attorney At Law | 20932 Hedley St    | Concord       | Contra Costa  | CA    | 94529 | 925-647-3298 | 925-522-7798 | micaela_rhymes@gmail.com          | http://www |
| MA 12 | Dominque   | Dickerson | E A I Electronic Associa Inc  | 69 Marquette Ave   | Hayward       | Alameda       | CA    | 94545 | 510-993-3758 | 510-901-7640 | dominique.dickerson@dickerson.org | http://www |

- ✓ Bạn có thể tạo cắt dữ liệu trong nhiều cột cùng một lúc. Như bên dưới cùng lúc có thể làm cột “State” và cột “City”

Facets-Clusters - Google

127.0.0.1:3333/project?project=1877640074305

Google refine Facets-Clusters Permalink

Facet / Filter Undo / Redo 2

Refresh Reset All Remove All

500 rows

Show as: rows records Show: 5 10 25 50 rows

Extensions: Freebase

« first » previous 1 - 10 next » last »

| state            | first_name | last_name | company_name              | address               | city        | county      | state    | zip   | phone1       | phone2       | email                       | web        |
|------------------|------------|-----------|---------------------------|-----------------------|-------------|-------------|----------|-------|--------------|--------------|-----------------------------|------------|
| AK 6             | James      | Butt      | Benton, John B Jr         | 5649 N Blue Gum St    | New Orleans | Orleans     | LA       | 70116 | 504-621-8927 | 504-845-1427 | jbutt@gmail.com             | http://www |
| AR 1             | Josephine  | Daraky    | Chanay, Jeffrey A Esq     | 4 B Blue Ridge Blvd   | Brighton    | Livingston  | MI       | 48116 | 810-292-9308 | 810-374-9840 | josephine_daraky@daraky.org | http://www |
| AZ 9             | Art        | Venere    | Chemel, James L Cpa       | 8 W Centros Ave #54   | Bridgeport  | Gloucester  | NJ       | 0814  | 856-436-8749 | 856-264-4130 | art@venere.org              | http://www |
| CA 55            | Lenna      | Paprocki  | Feltz Printing Service    | 639 Main St           | Anchorage   | Anchorage   | AK       | 99501 | 907-385-4412 | 907-921-2010 | lpaprocki@hotmail.com       | http://www |
| Cali-CA 4        | Donette    | Foler     | Printing Dimensions       | 34 Center St          | Hampton     | Butler      | OH       | 45011 | 513-576-1893 | 513-549-4581 | donette.foler@cox.net       | http://www |
| California(CA) 5 | Simona     | Morasca   | Chapman, Ross E Esq       | 3 Moulay Dr           | Ashland     | Ashland     | Ohio(oh) | 44805 | 419-503-2484 | 419-800-8759 | simona@morasca.com          | http://www |
| CO 8             | Mitsue     | Tolner    | Morning Associates        | 7 Eads St             | Chicago     | Cook        | IL       | 60632 | 773-673-6914 | 773-924-8585 | mitsue_tolner@yahoo.com     | http://www |
| CT 5             | Leota      | Dillard   | Commercial Press          | 7 W Jackson Blvd      | San Jose    | Santa Clara | Cal-CA   | 95111 | 408-752-3500 | 408-813-1105 | leota@hotmail.com           | http://www |
| DC 1             | Sage       | Wieser    | Truhlar And Truhlar Atlys | 5 Boston Ave #88      | Sioux Falls | Minnehaha   | SD       | 57105 | 605-414-2147 | 605-794-4895 | sage_wieser@cox.net         | http://www |
| FL 27            | Kris       | Manner    | King, Christopher A Esq   | 228 Rutanuck Pl #2008 | Baltimore   | Baltimore   | MD       | 21224 | 410-655-8723 | 410-804-4694 | kris@gmail.com              | http://www |

- ✓ Tương tự nhóm CA, California(ca), Cali-CA, California-CA bản chất giống nhau vì vậy chúng ta có thể gộp chung thành một nhóm với thuộc tính “Cluster” để làm cho dữ liệu phù hợp hơn. Chúng ta có thể gộp như sau:

### Custom text facet:

Cái này được sử dụng khi bạn muốn chỉnh sửa một ô giống như chiết xuất chỉ tên. Bạn cần đặt giá trị value.split(“”) [0]. Áp dụng biểu thức trên ta có trên cột “country\_name” như sau

Google refine us 500.xlsx [Permalink](#)

Facet / FilterUndo / Redo 0

RefreshReset AllRemove All

500 rows

Show as: rows recordsShow: 5 10 25 50 rows

Extensions: Freebase +

« first » previous 1 - 10 next » last »

countrychange

209 choices Sort by: name countCluster

Ad 3

Alamance 1

Alameda 7

Albany 1

Allegheny 1

Allen 1

Anchorage 4

Anne Arundel 2

Apakahoe 1

Ashland 1

Atlantic 4

Allfirst\_namelast\_namecompany\_nameaddresscitycountystatezipphone1phone2emailweb

1JamesButtFacetText facet

2JosephineDarakyText filterNumeric facet

3ArtVenereEdit cellsTimeline facet

4LennaPaprockiEdit columnScatterplot facet

5DonetteFolerTransposeCustom text facet...

6SimonaMorascaSort...Custom numeric facet...

7MitsueTolnerViewCustomized facets

8LeotaDillardReconcileJade St Chicago Cook IL 60632 773-673-6914 773-924-8585

9SageWieserCommercial Press 7 W Jackson Blvd San Jose Santa Clara CA 95111 408-752-3500 408-813-1105

10KrisMannerTruhlar And Truhlar Atlys 5 Boston Ave #88 Sioux Falls Minnehaha SD 57105 605-414-2147 605-794-4895

11King, Christopher A Esq228 Rutanuck Pl #2008 Baltimore Baltimore MD 21224 410-655-8723 410-804-4694

### Custom Facet on column company\_name

Expression Language Google Refine Expression Language (GREL)

`value.split(" ")[0]`

No syntax error.

Preview
History
Starred
Help

| row | value                  | value.split(" ")[0] |
|-----|------------------------|---------------------|
| 1.  | Benton, John B Jr      | Benton,             |
| 2.  | Chanay, Jeffrey A Esq  | Chanay,             |
| 3.  | Chemel, James L Cpa    | Chemel,             |
| 4.  | Feltz Printing Service | Feltz               |
| 5.  | Printing Dimensions    | Printing            |
| 6.  | Chapman, Ross E Esq    | Chapman,            |

OK Cancel

Facet / Filter Undo / Redo
2 matching rows (500 total)
Extensions: Freebase

Refresh Reset All Remove All

**county** change invert reset

2 choices Sort by: name count Cluster

- Lackawanna 1
- Maricopa 1

Facet by choice counts

Show as: rows records Show: 5 10 25 50 rows

|  | id  | first_name | last_name   | company_name                      | address                        | city            | county     | state | zip   | phone1           | phone2           | email              | web                 |
|--|-----|------------|-------------|-----------------------------------|--------------------------------|-----------------|------------|-------|-------|------------------|------------------|--------------------|---------------------|
|  | 126 | Rory       | Papaseerg   | Bailey Cntl Co Div<br>Babcock     | 83 County<br>Road 437<br>#5051 | Clark<br>Summit | Lackawanna | PA    | 16411 | 576-867-<br>7489 | 576-489-<br>8401 | rpapaseerg@cox.net | http://www.baileycd |
|  | 347 | Helene     | Rodenberger | Bailey Transportation Prod<br>Inc | 347 Chestnut<br>St             | Peoria          | Maricopa   | AZ    | 85361 | 623-461-<br>8551 | 623-426-<br>4907 | helene@aol.com     | http://www.baileyt  |

Cắt số để sắp xếp các số. Chúng được đặt các số dưới các hạng mức giá trị của số.

Facet / Filter Undo / Redo
500 rows
Extensions: Freebase

Refresh Reset All Remove All

**county** change invert reset

209 choices Sort by: name count Cluster

- Ada 3
- Alamance 1
- Alameda 7
- Albany 1
- Allegheny 1
- Allen 1
- Anchorage 4
- Anne Arundel 2
- Appomattox 1
- Ashland 1
- Atlantic 4

Show as: rows records Show: 5 10 25 50 rows

|  | id  | first_name | last_name | company_name                 | address                    | city           | county      | state | zip   | phone1           | phone2           | email                  | web       |
|--|-----|------------|-----------|------------------------------|----------------------------|----------------|-------------|-------|-------|------------------|------------------|------------------------|-----------|
|  | 1.  | James      | Bull      | Benton, John B Jr            | 6649 N Blue<br>Gum St      | New<br>Orleans | Orleans     | LA    |       |                  |                  |                        |           |
|  | 2.  | Josephine  | Carakly   | Chanay, Jeffrey A<br>Esq     | 4 B Blue<br>Ridge Blvd     | Brighton       | Livingston  | MI    |       |                  |                  |                        | http://wv |
|  | 3.  | Art        | Venero    | Chemel, James L<br>Cpa       | 8 W Centross<br>Ave #54    | Bridgeport     | Gloucester  | NJ    |       |                  |                  |                        | http://wv |
|  | 4.  | Lenna      | Paprocki  | Feltz Printing<br>Service    | 639 Main St                | Anchorage      | Anchorage   | AK    |       |                  |                  |                        | http://wv |
|  | 5.  | Donette    | Foller    | Printing Dimensions          | 34 Center St               | Hamilton       | Butler      | OH    |       |                  |                  |                        | http://wv |
|  | 6.  | Simona     | Morasca   | Chapman, Ross E<br>Esq       | 3 Mcawley Dr               | Ashland        | Ashland     | OH    |       |                  |                  |                        | http://wv |
|  | 7.  | Mitae      | Tollier   | Working Associates           | 7 Eads St                  | Chicago        | Cook        | IL    |       |                  |                  | mtae_tollier@yahoo.com | http://wv |
|  | 8.  | Leola      | Dillard   | Commercial Press             | 7 W Jackson<br>Blvd        | San Jose       | Santa Clara | CA    | 95111 | 408-752-<br>3500 | 408-813-<br>1105 | leola@hotmail.com      | http://wv |
|  | 9.  | Sage       | Wieser    | Truhlar And Truhlar<br>Allys | 5 Boston Ave<br>#55        | Sioux Falls    | Minnehaha   | SD    | 57105 | 605-414-<br>2147 | 605-794-<br>4895 | sage_wieser@cox.net    | http://wv |
|  | 10. | Kris       | Manner    | King, Christopher A<br>Esq   | 229<br>Runnuck Pl<br>#2608 | Baltimore      | Baltimore   | MD    | 21224 | 410-855-<br>8723 | 410-354-<br>4694 | kris@gmail.com         | http://wv |

|     | first_name | last_name | company_name              | address               | city        | county      | state | zip   | phone1       | phone2       | email                       | we       |
|-----|------------|-----------|---------------------------|-----------------------|-------------|-------------|-------|-------|--------------|--------------|-----------------------------|----------|
| 1.  | James      | Butt      | Benton, John B Jr         | 6549 N Blue Gum St    | New Orleans | Orleans     | LA    | 70116 | 504-421-8927 | 504-845-1427 | jbutt@gmail.com             | http://w |
| 2.  | Josephine  | Daraly    | Chaney, Jeffrey A         | 4 B Blue Ridge Blvd   | Brighton    | Livingston  | RI    | 01518 | 618-362-9389 | 618-374-9840 | josephine_daraly@daraly.org | http://w |
| 3.  | Art        | Venere    | Chenel, James L           | 8 W Cerritos Ave #54  | Bridgeport  | Gloucester  | NJ    | 0814  | 856-636-8749 | 856-264-4130 | art@venere.org              | http://w |
| 4.  | Lenna      | Paproski  | Feltz Printing Service    | 639 Main St           | Anchorage   | Anchorage   | AK    | 99501 | 907-305-4412 | 907-921-2010 | lpaproski@hotmail.com       | http://w |
| 5.  | Donette    | Foler     | Printing Dimensions       | 34 Center St          | Hamilton    | Butler      | OH    | 45011 | 513-578-1893 | 513-545-4581 | donette.foler@cox.net       | http://w |
| 6.  | Simona     | Morasca   | Chapman, Rose E Esq       | 3 Mouley Dr           | Ashland     | Ashland     | OH    | 44805 | 419-503-2464 | 419-800-8759 | simona@morasca.com          | http://w |
| 7.  | Mitau      | Tolner    | Worlong Associates        | 7 Eads St             | Chicago     | Cook        | IL    | 60632 | 773-673-6914 | 773-624-8585 | mitau_tolner@yahoo.com      | http://w |
| 8.  | Leota      | Dillard   | Commercial Press          | 7 W Jackson Blvd      | San Jose    | Santa Clara | CA    | 95111 | 408-752-3550 | 408-813-1195 | leota@hotmail.com           | http://w |
| 9.  | Sage       | Wieser    | Truhlar And Truhlar A/B/C | 5 Boston Ave #88      | Sioux Falls | Minnehaha   | SD    | 57105 | 605-414-2147 | 605-794-4895 | sage_wieser@cox.net         | http://w |
| 10. | Kira       | Manner    | King, Christopher A Esq   | 228 Runamuck Pl #2000 | Baltimore   | Baltimore   | MD    | 21224 | 410-655-8723 | 410-204-4694 | kira@gmail.com              | http://w |

### 3. Chỉnh sửa hàng (Editing cells)

Sửa chữa các ô bảng tính hay còn gọi là các cells bằng cách sử dụng chuyển đổi chung “Commun Transform” trong mục “Edit cells”: Có một vài chức năng được sử dụng để chuyển đổi các giá trị ký tự của cell trong một đợt vận hành. Truy cập tới

#### a) Edit cells → Common Transforms .

Các chức năng trong mục chuyển đổi như sau:

- **Trim leading and trailing whitespace (Xóa khoảng trắng):** Chức năng này được sử dụng nếu trong dữ liệu của chúng ta có khoảng trắng trong một từ của cột.
- **Collapse consecutive whitespace:** Được sử dụng để giảm các khoảng trắng liên tiếp trong cột.
- **Un-escape HTML identities:** Chức năng này giúp loại bỏ các thực thể có gắn với HTML
- **To titlecase:** Được sử dụng để chuyển chữ hoa chữ cái đầu tiên của từ.
- **To uppercase:** Được dùng để chuyển chữ hoa cả từ
- **To lowercase:** Được dùng để chuyển chữ nhỏ cả từ
- **To number:** Chuyển sang định dạng số
- **To date:** Chuyển sang định dạng ngày tháng
- **To text:** Chuyển sang định dạng kiểu ký tự.

#### b) Edit cells → Transforms (Chuyển đổi dữ liệu):

Giả thiết nếu chúng ta đã tải file ký tự cho tất cả các sinh viên của UMD bao gồm các khóa học, năm, tuổi ... Rồi sau đó chúng ta có thể chiết xuất năm hay bất kỳ các dữ liệu nào để tạo một cột mới một cách dễ dàng. Cột mới có thể được tạo từ các cột đã tồn tại.

Chúng ta có thể loại bỏ khoảng trống hay bất kỳ các ký tự không mong muốn từ dữ liệu của các file text đã nhập. File text có một vài dòng trống nhưng bạn có thể lựa chọn không lưu những dòng trống và lọc dữ liệu khi tạo dự án.

Google refine Transform\_data.txt Permalink

Facet / Filter Undo / Redo

Using facets and filters

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started? Watch these screencasts

22 rows

Show as: rows records Show: 5 10 25 50 rows

Extensions: Freebase

| All | Column 1   |
|-----|--|
| 1.  | **1988 James Butt Benton, John B Jr 6649 N Blue Gum St New Orleans Orleans LA 7011 504-621-8927 504-845-1427 jbutt@gmail.com http://www.bentonjohnbjr.com                                |
| 2.  | **1985 Josephine Darakjy Chanay, Jeffrey A Esq 4 B Blue Ridge Blvd Brighton Livingston MI 48116 810-292-9388 810-374-9840 josephine_darakjy@darakjy.org http://www.chanayjeffreyaesq.com |
| 3.  | **1990 Art Venere Chemei, James L Cpa 8 W Cerritos Ave #54 Bridgeport Gloucester NJ 8014 856-636-8749 856-264-4130 art@venere.org http://www.chemejamescpa.com                           |
| 4.  | **1988 Lenna Paprocki Feltz Printing Service 639 Main St Anchorage Anchorage AK 99501 907-385-4412 907-921-2010 lpaprocki@hotmail.com http://www.feltzprintingservice.com                |
| 5.  | **1988 Margaret Lopez  |
| 6.  | **1988 Donette Foler Printing Dimensions 34 Center St Hamilton Butler OH 45011 513-570-1893 513-549-4561 donette.foler@cox.net http://www.printingdimensions.com                         |
| 7.  | **1990 Simona Morasca Chapman, Ross E Esq 3 Mowley Dr Ashland Ashland OH 44805 419-503-2484 419-800-8759 simona@morasca.com http://www.chapmanrossesq.com                                |
| 8.  | **1993 Mitsue Tolner Morlong Associates 7 Eads St Chicago Cook IL 60632 773-573-6914 773-924-8565 mtsue_tolner@yahoo.com http://www.morlongassociates.com                                |
| 9.  | **1983 Lesta Ollard Commercial Press 7 W Jackson Blvd San Jose Santa Clara CA 95111 408-752-3500 408-813-1105 lesta@hotmail.com http://www.commercialpress.com                           |
| 10. | **1984 Sage Wieser Truhlar And Truhlar Atlys 5 Boston Ave #88 Sioux Falls Minnehaha SD 57105 605-414-2147 605-794-4895 sage_wieser@cox.net http://www.truhlarandtruhlaratlys.com         |

✓ Sử dụng biểu thức value.replace ("\*\*", "") để loại bỏ biểu tượng sao

Transform\_data.txt - Google

127.0.0.1:3333/project?project=2152684130791

Google refine Transform\_data.txt Permalink

Facet / Filter Undo / Redo

Using facets and filters

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started? Watch these screencasts

22 rows

Show as: rows records Show: 5 10 25 50 rows

Extensions: Freebase

| All | Column 1    |
|-----|-------------|
| 1.  | Facet       |
| 2.  | Text filter |
| 3.  | Edit cells  |
| 4.  | Edit column |
| 5.  | Transpose   |
| 6.  | Sort...     |
| 7.  | View        |
| 8.  | Reconcile   |
| 9.  |             |
| 10. |             |

### Custom text transform on column Column 1

Expression Language Google Refine Expression Language (GREL)

`value.replace("**", "")` No syntax error.

Preview History Starred Help

| row | value  | value.replace("**", "")  |
|-----|--|--|
| 1.  | **1988 James Butt Benton, John B Jr 6649 N Blue Gum St New Orleans Orleans LA 7011 504-621-8927 504-845-1427 jbutt@gmail.com http://www.bentonjohnbjr.com                                | 1988 James Butt Benton, John B Jr 6649 N Blue Gum St New Orleans Orleans LA 7011 504-621-8927 504-845-1427 jbutt@gmail.com http://www.bentonjohnbjr.com                                |
| 2.  | **1985 Josephine Darakjy Chanay, Jeffrey A Esq 4 B Blue Ridge Blvd Brighton Livingston MI 48116 810-292-9388 810-374-9840 josephine_darakjy@darakjy.org http://www.chanayjeffreyaesq.com | 1985 Josephine Darakjy Chanay, Jeffrey A Esq 4 B Blue Ridge Blvd Brighton Livingston MI 48116 810-292-9388 810-374-9840 josephine_darakjy@darakjy.org http://www.chanayjeffreyaesq.com |

On error ☒ keep original ☐ set to blank ☐ store error

☐ Re-transform up to  times until no change

OK Cancel



Google Refine Transform\_data.txt Permalink

Text transform on 20 cells in column Column 1:  
grel:value.replace("\*\*\*\*", "") Undo

Facet / Filter Undo / Redo

22 rows

Show as: rows records Show: 5 10 25 50 rows

Extensions: Freebase

Using facets and filters

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started? Watch these screencasts

|     | Column 1   |
|-----|--|
| 1.  | 1988 James Butt Benton, John B Jr 6649 N Blue Gum St New Orleans Orleans LA 7011 504-621-8927 504-845-1427 jbutt@gmail.com http://www.bentonjohnbjr.com                                |
| 2.  | 1985 Josephine Darakjy Chanay, Jeffrey A Esq 4 B Blue Ridge Blvd Brighton Livingston MI 48116 810-292-9388 810-374-9840 josephine_darakjy@darakjy.org http://www.chanayjeffreyaesq.com |
| 3.  | 1990 Art Venere Chemel, James L Cpa 8 W Cerritos Ave #54 Bridgeport Gloucester NJ 8014 856-636-8749 856-264-4130 art@venere.org http://www.chemelamescpa.com                           |
| 4.  | 1988 Lenna Paprocki Felz Printing Service 639 Main St Anchorage Anchorage AK 99501 907-385-4412 907-821-2010 lpaprocki@hotmail.com http://www.felzprintingservice.com                  |
| 5.  | 1988 Margaret Lopez  |
| 6.  | 1988 Donette Foller Printing Dimensions 34 Center St Hamilton Butler OH 45011 513-570-1893 513-549-4561 donette.foller@cox.net http://www.printingdimensions.com                       |
| 7.  | 1990 Simona Morasca Chapman, Ross E Esq 3 McAuley Dr Ashland Ashland OH 44805 419-503-2484 419-800-6759 simona@morasca.com http://www.chapmanrossesq.com                               |
| 8.  | 1993 Mitsue Tolner Morlong Associates 7 Eads St Chicago Cook IL 60632 773-573-8914 773-924-8555 mtsue_tolner@yahoo.com http://www.morlongassociates.com                                |
| 9.  | 1983 Leota Dillard Commercial Press 7 W Jackson Blvd San Jose Santa Clara CA 95111 408-752-3500 408-813-1105 leota@hotmail.com http://www.commercialpress.com                          |
| 10. | 1984 Sage Wieser Truhlar And Truhlar Attys 5 Boston Ave #88 Sioux Falls Minnehaha SD 57105 605-414-2147 605-794-4895 sage_wieser@cox.net http://www.truhlarandtruhlarattys.com         |

- ✓ Để di chuyển năm sang một cột độc lập, lựa chọn “Add column based on this column”

Transform\_data.txt - Google

127.0.0.1:3333/project?project=2152684130791

Google Refine Transform\_data.txt Permalink

Facet / Filter Undo / Redo

22 rows

Show as: rows records Show: 5 10 25 50 rows

Extensions: Freebase

Using facets and filters

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started? Watch these screencasts

|     | Column 1   |
|-----|--|
| 1.  | John B Jr 6649 N Blue Gum St New Orleans Orleans LA 7011 504-621-8927 504-845-1427 jbutt@gmail.com http://www.bentonjohnbjr.com                                |
| 2.  | hanay, Jeffrey A Esq 4 B Blue Ridge Blvd Brighton Livingston MI 48116 810-292-9388 810-374-9840 josephine_darakjy@darakjy.org http://www.chanayjeffreyaesq.com |
| 3.  | James L Cpa 8 W Cerritos Ave #54 Bridgeport Gloucester NJ 8014 856-636-8749 856-264-4130 art@venere.org http://www.chemelamescpa.com                           |
| 4.  | Printing Service 639 Main St Anchorage Anchorage AK 99501 907-385-4412 907-821-2010 lpaprocki@hotmail.com http://www.felzprintingservice.com                   |
| 5.  | Split into several columns...  |
| 6.  | Add column based on this column...   |
| 7.  | 513-570-1893 513-549-4561 donette.foller@cox.net http://www.printingdimensions.com   |
| 8.  | 44805 419-503-2484 419-800-6759 simona@morasca.com http://www.chapmanrossesq.com   |
| 9.  | 573-6914 773-924-8565 mtsue_tolner@yahoo.com http://www.morlongassociates.com  |
| 10. | CA 95111 408-752-3500 408-813-1105 leota@hotmail.com http://www.commercialpress.com  |

Reconcile

Remove this column

Move column to beginning

Move column to end

Move column left

Move column right

- ✓ Sử dụng biểu thức value[1,5] để chỉ ra dãy các ký tự muốn năm muốn tách

Add column based on column Column 1

New column name: Year

On error: ☒ set to blank ☐ store error ☐ copy value from original column

Expression: value[0,5] Language: Google Refine Expression Language (GREL)

No syntax error.

Preview History Starred Help

| row | value  | value[0,5] |
|-----|--|------------|
| 1.  | 1988 James Butt Benton, John B Jr 6649 N Blue Gum St New Orleans Orleans LA 7011 504-621-8927 504-845-1427 jbutt@gmail.com http://www.bentonjohnbjr.com                                | 1988       |
| 2.  | 1985 Josephine Darakjy Chanay, Jeffrey A Esq 4 B Blue Ridge Blvd Brighton Livingston MI 48116 810-292-9388 810-374-9840 josephine_darakjy@darakjy.org http://www.chanayjeffreyaesq.com | 1985       |
| 3.  | 1990 Art Venere Chemel, James L Cpa 8 W Cerritos Ave #54 Bridgeport Gloucester NJ 8014 856-636-8749 856-264-4130 art@venere.org http://www.chemelamescpa.com                           | 1990       |

OK Cancel

22 rows Extensions: Freebase

Show as: rows records Show: 5 10 25 50 rows « first < previous 1 - 10 next > last »

| All | Column 1   | Year |
|-----|--|------|
| 1.  | 1988 James Butt Benton, John B Jr 6649 N Blue Gum St New Orleans Orleans LA 7011 504-621-8927 504-845-1427 jbutt@gmail.com http://www.bentonjohnbjr.com                                | 1988 |
| 2.  | 1985 Josephine Darakjy Chanay, Jeffrey A Esq 4 B Blue Ridge Blvd Brighton Livingston MI 48116 810-292-9388 810-374-9840 josephine_darakjy@darakjy.org http://www.chanayjeffreyaesq.com | 1985 |
| 3.  | 1990 Art Venere Chemel, James L Cpa 8 W Cerritos Ave #54 Bridgeport Gloucester NJ 8014 856-636-8749 856-264-4130 art@venere.org http://www.chemeljameslcpa.com                         | 1990 |
| 4.  | 1988 Lenna Paprocki Feltz Printing Service 639 Main St Anchorage Anchorage AK 99501 907-385-4412 907-921-2010 lpaprocki@hotmail.com http://www.feltzprintingservice.com                | 1988 |
| 5.  | 1988 Margaret lopez  | 1988 |
| 6.  | 1988 Donette Foller Printing Dimensions 34 Center St Hamilton Butler OH 45011 513-570-1893 513-549-4561 donette.foller@cox.net http://www.printingdimensions.com                       | 1988 |
| 7.  | 1990 Simona Morasca Chapman, Ross E Esq 3 McAuley Dr Ashland Ashland OH 44805 419-503-2484 419-800-6759 simona@morasca.com http://www.chapmanrosseeesq.com                             | 1990 |
| 8.  | 1993 Mitsue Tollner Morlong Associates 7 Eads St Chicago Cook IL 60632 773-573-6914 773-924-8565 mitsue_tollner@yahoo.com http://www.morlongassociates.com                             | 1993 |
| 9.  | 1983 Leota Dillard Commercial Press 7 W Jackson Blvd San Jose Santa Clara CA 95111 408-752-3500 408-813-1105 leota@hotmail.com http://www.commercialpress.com                          | 1983 |
| 10. | 1984 Sage Wieser Truhlar And Truhlar Attys 5 Boston Ave #88 Sioux Falls Minnehaha SD 57105 605-414-2147 605-794-4895 sage_wieser@cox.net http://www.truhlarandtruhlarattys.com         | 1984 |

- ✓ Chúng ta có thể loại bỏ năm trong cột gốc bằng “Transform”
- ✓ Sử dụng biểu thức value.substring(5) sẽ loại bỏ năm ở cột gốc.

### Custom text transform on column Column 1

Expression Language Google Refine Expression Language (GREL)

`value.substring(5)` No syntax error.

Preview History Starred Help

| row | value  | value.substring(5)  |
|-----|--|---|
| 1.  | 1988 James Butt Benton, John B Jr 6649 N Blue Gum St New Orleans Orleans LA 7011 504-621-8927 504-845-1427 jbutt@gmail.com http://www.bentonjohnbjr.com                                | James Butt Benton, John B Jr 6649 N Blue Gum St New Orleans Orleans LA 7011 504-621-8927 504-845-1427 jbutt@gmail.com http://www.bentonjohnbjr.com                                |
| 2.  | 1985 Josephine Darakjy Chanay, Jeffrey A Esq 4 B Blue Ridge Blvd Brighton Livingston MI 48116 810-292-9388 810-374-9840 josephine_darakjy@darakjy.org http://www.chanayjeffreyaesq.com | Josephine Darakjy Chanay, Jeffrey A Esq 4 B Blue Ridge Blvd Brighton Livingston MI 48116 810-292-9388 810-374-9840 josephine_darakjy@darakjy.org http://www.chanayjeffreyaesq.com |

On error ☒ keep original ☐ set to blank ☐ store error ☐ Re-transform up to 10 times until no change

OK Cancel

22 rows Extensions: Freebase

Show as: rows records Show: 5 10 25 50 rows « first < previous 1 - 10 next > last »

| All | Column 1  | Year |
|-----|---|------|
| 1.  | James Butt Benton, John B Jr 6649 N Blue Gum St New Orleans Orleans LA 7011 504-621-8927 504-845-1427 jbutt@gmail.com http://www.bentonjohnbjr.com                                | 1988 |
| 2.  | Josephine Darakjy Chanay, Jeffrey A Esq 4 B Blue Ridge Blvd Brighton Livingston MI 48116 810-292-9388 810-374-9840 josephine_darakjy@darakjy.org http://www.chanayjeffreyaesq.com | 1985 |
| 3.  | Art Venere Chemel, James L Cpa 8 W Cerritos Ave #54 Bridgeport Gloucester NJ 8014 856-636-8749 856-264-4130 art@venere.org http://www.chemeljameslcpa.com                         | 1990 |
| 4.  | Lenna Paprocki Feltz Printing Service 639 Main St Anchorage Anchorage AK 99501 907-385-4412 907-921-2010 lpaprocki@hotmail.com http://www.feltzprintingservice.com                | 1988 |
| 5.  | Margaret lopez  | 1988 |
| 6.  | Donette Foller Printing Dimensions 34 Center St Hamilton Butler OH 45011 513-570-1893 513-549-4561 donette.foller@cox.net http://www.printingdimensions.com                       | 1988 |
| 7.  | Simona Morasca Chapman, Ross E Esq 3 McAuley Dr Ashland Ashland OH 44805 419-503-2484 419-800-6759 simona@morasca.com http://www.chapmanrosseeesq.com                             | 1990 |
| 8.  | Mitsue Tollner Morlong Associates 7 Eads St Chicago Cook IL 60632 773-573-6914 773-924-8565 mitsue_tollner@yahoo.com http://www.morlongassociates.com                             | 1993 |
| 9.  | Leota Dillard Commercial Press 7 W Jackson Blvd San Jose Santa Clara CA 95111 408-752-3500 408-813-1105 leota@hotmail.com http://www.commercialpress.com                          | 1983 |
| 10. | Sage Wieser Truhlar And Truhlar Attys 5 Boston Ave #88 Sioux Falls Minnehaha SD 57105 605-414-2147 605-794-4895 sage_wieser@cox.net http://www.truhlarandtruhlarattys.com         | 1984 |

**Một số biểu thức:** OpenRefine hỗ trợ các biểu thức “Expressions”. Và các biểu thức này được sử dụng để chuyển đổi các dữ liệu sẵn có và tạo ra một dữ liệu mới dựa trên các dữ liệu tồn tại. Cái này giống như công thức “Formula” chúng được sử dụng giống như trong Excel. Tuy nhiên có sự khác nhau. Formula trong Excel có thể chỉ được sử dụng để chứa các công thức khác nhau cho mỗi một cell mà nó chỉ định. Ngược lại, Expresión, bằng cách sử dụng “GREL” (Google Refine Expression Language: Ngôn ngữ biểu thức tính lọc của Google) có thể áp dụng cho các hàng và cột bằng cách áp dụng các điều kiện tới các hàng cột đó.

## Edit Cells→Transform

**Chi tiết:** Chúng ta có một bộ dữ liệu sau

| first_name | last_name | company_name          | address              |
|------------|-----------|-----------------------|----------------------|
| James      | Butt      | Benton, John B Jr     | 6649 N Blue Gum St   |
| Josephine  | Darakjy   | Chanay, Jeffrey A Esq | 4 B Blue Ridge Blvd  |
| Art        | Venere    | Chemel, James L Cpa   | 8 W Cerritos Ave #54 |

Khi bạn muốn viện dẫn và đưa biểu thức vào, OpenRefine sẽ đến từng hàng trong dữ liệu (liên kết với các phương thức cắt gọt và lọc), rồi đánh giá biểu thức đó cho từng dòng để đạt được một kết quả của từng dòng đó.

Để chuyển đổi dữ liệu, chúng ta nên chọn một dòng và kích Edit Cells và rồi kích Transform.

| Show as: rows records |            | Show: 5 10 25 50 rows |                             |                    |             | « first < previous 1 - 10 next > last » |       |       |              |              |                      |
|-----------------------|------------|-----------------------|-----------------------------|--------------------|-------------|---|-------|-------|--------------|--------------|----------------------|
| All                   | first_name | last_name             | company_name                | address            | city        | county                                  | state | zip   | phone1       | phone2       | email                |
| 1.                    | James      | Facet                 | n, John B Jr                | 6649 N Blue Gum St | New Orleans | Orleans                                 | LA    | 70116 | 504-621-8927 | 504-845-1427 | jbutt@gmail.com      |
| 2.                    | Josephine  | Text filter           | ay, Jeffrey A               | 4 B Blue           | Brighton    | Livingston                              | MI    | 48116 | 810-292-9388 | 810-374-9840 | josephine_darakjy@   |
| 3.                    | Art        | Edit cells            | Transform...                |                    | idgeport    | Gloucester                              | NJ    | 8014  | 856-636-8749 | 856-264-4130 | art@venere.org       |
| 4.                    | Lenna      | Edit column           | Common transforms           |                    | chorage     | Anchorage                               | AK    | 99501 | 907-385-4412 | 907-921-2010 | lpaprocki@hotmail.c  |
| 5.                    | Donette    | Transpose             | Fill down                   |                    | amilton     | Butler                                  | OH    | 45011 | 513-570-1893 | 513-549-4561 | donette.foller@cox.n |
| 6.                    | Simona     | Sort...               | Blank down                  |                    | hland       | Ashland                                 | OH    | 44805 | 419-503-2484 | 419-800-6759 | simona@morasca.c     |
| 7.                    | Mitsue     | View                  | Split multi-valued cells... |                    | hicago      | Cook                                    | IL    | 60632 | 773-573-6914 | 773-924-8565 | mitsue_tollner@yah   |
| 8.                    | Leota      | Reconcile             | Join multi-valued cells...  |                    | San Jose    | Santa Clara                             | CA    | 95111 | 408-752-3500 | 408-813-1105 | leota@hotmail.com    |
| 9.                    | Sage       |                       | Cluster and edit...         |                    | Sioux Falls | Minnehaha                               | SD    | 57105 | 605-414-2147 | 605-794-4895 | sage_wieser@cox.n    |
| 10.                   | Kris       |                       |                             |                    | Baltimore   | Baltimore City                          | MD    | 21224 | 410-655-8723 | 410-804-4694 | kris@gmail.com       |

Biểu thức 1: value+”(string)”

500 rows

### Custom text transform on column last\_name

Expression Language Google Refine Expression Language (GREL)

"Mr." + value No syntax error.

**Preview** History Starred Help

| row | value    | "Mr." + value |
|-----|----------|---------------|
| 1.  | Butt     | Mr.Butt       |
| 2.  | Darakjy  | Mr.Darakjy    |
| 3.  | Venere   | Mr.Venere     |
| 4.  | Paprocki | Mr.Paprocki   |
| 5.  | Foller   | Mr.Foller     |
| 6.  | Morasca  | Mr.Morasca    |

On error ☒ keep original ☐ set to blank ☐ store error ☐ Re-transform up to 10 times until no change

OK Cancel

Chức năng 2: value.trim().length()

### Custom text transform on column last\_name

Expression Language Google Refine Expression Language (GREL)

value.trim().length() No syntax error.

**Preview** History Starred Help

| row | value    | value.trim().length() |
|-----|----------|-----------------------|
| 1.  | Butt     | 4                     |
| 2.  | Darakjy  | 7                     |
| 3.  | Venere   | 6                     |
| 4.  | Paprocki | 8                     |
| 5.  | Foller   | 6                     |
| 6.  | Morasca  | 7                     |

On error ☒ keep original ☐ set to blank ☐ store error ☐ Re-transform up to 10 times until no change

OK Cancel

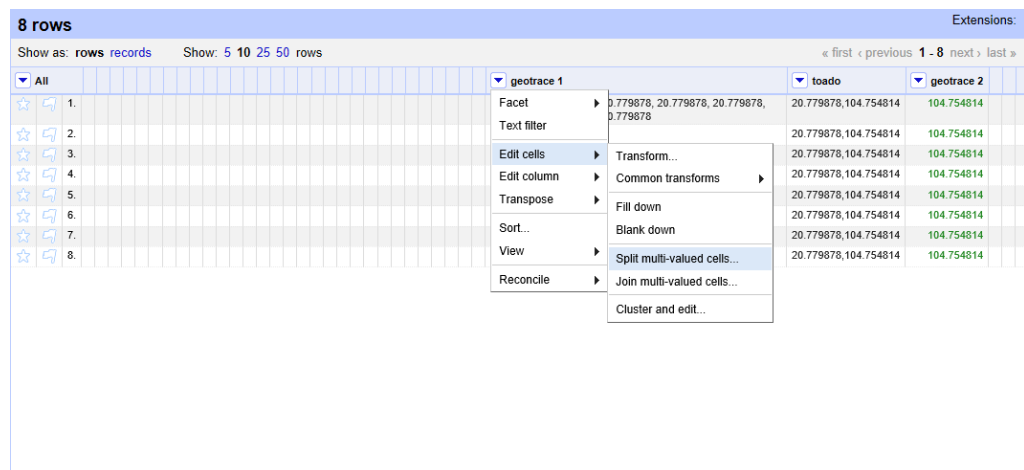
Một vài chức năng:

- Syntax: “Mr.” + value : Được dùng để móc nối hai chuỗi ký tự
- Syntax: value + 3.9 : Được dùng để cộng giá trị nếu như giá trị là dạng số, sẽ cộng chuỗi nếu như là giá trị ký tự.
- Syntax: value.trim().length() : Được dùng để loại bỏ khoảng trắng
- Syntax: value.substring(0,3) : Nhận giá trị ký tự từ vị trí đầu tiên và độ dài ký tự là 3

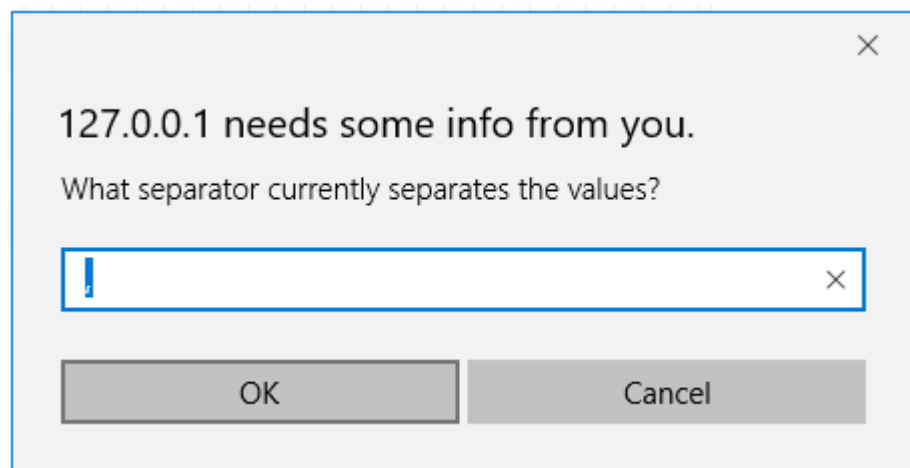
c) **Edit cells → Split multi-valued cells:**

Chức năng này giúp tách nhiều giá trị trong cùng một cell.

Ví dụ: Chúng ta có nhiều giá trị trong cùng một cell, các giá trị được phân cách với nhau bởi dấu phẩy. Vậy thì ta sẽ tách các giá trị đó ra thành nhiều dòng khác nhau bằng cách truy cập đến mục **Edit cells** rồi thì đến **Split multi-valued cells**.



Các giá trị phân cách với nhau bởi biểu tượng gì thì ta nhập vào ô sau



Sau khi nhập biểu tượng vào ta có kết quả trong cột “geotrace 1” như sau

| 8 rows                |    |  |  |  |  |  |  |  |  |                       |  |  |  |  |  |  |  |  |  |  |  |  |  |           |
|-----------------------|----|--|--|--|--|--|--|--|--|-----------------------|--|--|--|--|--|--|--|--|--|--|--|--|--|-----------|
| Show as: rows records |    |  |  |  |  |  |  |  |  | Show: 5 10 25 50 rows |  |  |  |  |  |  |  |  |  |  |  |  |  |           |
| ▼ All                 |    |  |  |  |  |  |  |  |  | ▼ geotrace 1          |  |  |  |  |  |  |  |  |  |  |  |  |  |           |
| ☆ ↻                   | 1. |  |  |  |  |  |  |  |  |                       |  |  |  |  |  |  |  |  |  |  |  |  |  | 20.779878 |
| ☆ ↻                   | 2. |  |  |  |  |  |  |  |  |                       |  |  |  |  |  |  |  |  |  |  |  |  |  | 20.779878 |
| ☆ ↻                   | 3. |  |  |  |  |  |  |  |  |                       |  |  |  |  |  |  |  |  |  |  |  |  |  | 20.779878 |
| ☆ ↻                   | 4. |  |  |  |  |  |  |  |  |                       |  |  |  |  |  |  |  |  |  |  |  |  |  | 20.779878 |
| ☆ ↻                   | 5. |  |  |  |  |  |  |  |  |                       |  |  |  |  |  |  |  |  |  |  |  |  |  | 20.779878 |
| ☆ ↻                   | 6. |  |  |  |  |  |  |  |  |                       |  |  |  |  |  |  |  |  |  |  |  |  |  | 20.779878 |
| ☆ ↻                   | 7. |  |  |  |  |  |  |  |  |                       |  |  |  |  |  |  |  |  |  |  |  |  |  | 20.779878 |
| ☆ ↻                   | 8. |  |  |  |  |  |  |  |  |                       |  |  |  |  |  |  |  |  |  |  |  |  |  | 20.779878 |

#### d) Edit cells → Join multi-valued cells:

Chức năng này giúp gộp giá trị của các cell vào thành một dòng. Truy cập đến **Edit cells → Join multi-valued cells**. Tương tự như **Split multi-valued cells** bạn cần nhập biểu tượng để phân tách các giá trị (có thể là: dấu phẩy, chấm phẩy, hai chấm, hoặc khoảng trắng...). Sau khi nhập xong kết quả cột “geotrace 1” như hình ảnh bên dưới

| 8 rows                |    |  |  |  |  |  |  |  |  |                       |  |  |  |  |  |  |  |  |  |  |  |  |  |  | Extensions:   |
|-----------------------|----|--|--|--|--|--|--|--|--|-----------------------|--|--|--|--|--|--|--|--|--|--|--|--|--|--|---|
| Show as: rows records |    |  |  |  |  |  |  |  |  | Show: 5 10 25 50 rows |  |  |  |  |  |  |  |  |  |  |  |  |  |  | « first < previous 1 - 8 next > last »  |
| ▼ All                 |    |  |  |  |  |  |  |  |  | ▼ geotrace 1          |  |  |  |  |  |  |  |  |  |  |  |  |  |  | ▼ toado   |
| ☆ ↻                   | 1. |  |  |  |  |  |  |  |  |                       |  |  |  |  |  |  |  |  |  |  |  |  |  |  | 20.779878,20.779878,20.779878,20.779878,20.779878,20.779878,20.779878,20.779878 |
| ☆ ↻                   | 2. |  |  |  |  |  |  |  |  |                       |  |  |  |  |  |  |  |  |  |  |  |  |  |  | 20.779878,104.754814  |
| ☆ ↻                   | 3. |  |  |  |  |  |  |  |  |                       |  |  |  |  |  |  |  |  |  |  |  |  |  |  | 20.779878,104.754814  |
| ☆ ↻                   | 4. |  |  |  |  |  |  |  |  |                       |  |  |  |  |  |  |  |  |  |  |  |  |  |  | 20.779878,104.754814  |
| ☆ ↻                   | 5. |  |  |  |  |  |  |  |  |                       |  |  |  |  |  |  |  |  |  |  |  |  |  |  | 20.779878,104.754814  |
| ☆ ↻                   | 6. |  |  |  |  |  |  |  |  |                       |  |  |  |  |  |  |  |  |  |  |  |  |  |  | 20.779878,104.754814  |
| ☆ ↻                   | 7. |  |  |  |  |  |  |  |  |                       |  |  |  |  |  |  |  |  |  |  |  |  |  |  | 20.779878,104.754814  |
| ☆ ↻                   | 8. |  |  |  |  |  |  |  |  |                       |  |  |  |  |  |  |  |  |  |  |  |  |  |  | 20.779878,104.754814  |

## 4. Chỉnh sửa cột (Editing columns)

### a) Chức năng thu hẹp/mở rộng cột (Collapsing/Expanding columns)

Truy cập đến mục **View** chúng ta có 4 lựa chọn:

- Collapse this column : thu hẹp cột này
- Collapse all other columns : thu hẹp tất cả các cột trừ cột này
- Collapse columns to left : thu hẹp các cột bên trái cột này
- Collapse columns to right : thu hẹp các cột bên phải cột này

Sau khi thu hẹp tất cả các cột trừ cột mà chúng ta đang trở tới. Màn hình sẽ như hình bên dưới





e) **Tách dữ liệu từ một cột ra nhiều cột khác nhau**

Chức năng này giúp tách các giá trị trong cùng một cột ra thành nhiều cột khác nhau và riêng biệt. Ví dụ hình ảnh sau là các dữ liệu có trong cùng một cột “geotrace”.

| 2 records             |   |    |  |  |                          |            |      |      |  |
|-----------------------|---|----|--|--|--------------------------|------------|------|------|--|
| Show as: rows records |   |    |  |  | Show: 5 10 25 50 records |            |      |      |  |
| ▼ All                 |   |    |  |  | ▼ geotrace               |            |      |      |  |
| ★                     | 🗨 | 1. |  |  | 20.779878                | 104.754814 | 1060 | 8.0  |  |
| ★                     | 🗨 |    |  |  | 20.779878                | 104.754814 | 1060 | 16.0 |  |
| ★                     | 🗨 |    |  |  | 20.779878                | 104.754814 | 1060 | 16.0 |  |
| ★                     | 🗨 |    |  |  | 20.779878                | 104.754814 | 1060 | 16.0 |  |
| ★                     | 🗨 |    |  |  | 20.779878                | 104.754814 | 1060 | 16.0 |  |
| ★                     | 🗨 |    |  |  | 20.779878                | 104.754814 | 1060 | 16.0 |  |
| ★                     | 🗨 |    |  |  | 20.779878                | 104.754814 | 1059 | 12.0 |  |
| ★                     | 🗨 |    |  |  | 20.779878                | 104.754814 | 1056 | 16.0 |  |
| ★                     | 🗨 | 2. |  |  |                          |            |      |      |  |

Sau khi truy cập đến mục Edit columns → Split into several columns

| 2 records             |   |    |  |  |                          |            |      |      |  |
|-----------------------|---|----|--|--|--------------------------|------------|------|------|--|
| Show as: rows records |   |    |  |  | Show: 5 10 25 50 records |            |      |      |  |
| ▼ All                 |   |    |  |  | ▼ geotrace               |            |      |      |  |
| ★                     | 🗨 | 1. |  |  | 20.779878                | 104.754814 | 1060 | 8.0  |  |
| ★                     | 🗨 |    |  |  | 20.779878                | 104.754814 | 1060 | 16.0 |  |
| ★                     | 🗨 |    |  |  | 20.779878                | 104.754814 | 1060 | 16.0 |  |
| ★                     | 🗨 |    |  |  | 20.779878                | 104.754814 | 1060 | 16.0 |  |
| ★                     | 🗨 |    |  |  | 20.779878                | 104.754814 | 1060 | 16.0 |  |
| ★                     | 🗨 |    |  |  | 20.779878                | 104.754814 | 1060 | 16.0 |  |
| ★                     | 🗨 |    |  |  | 20.779878                | 104.754814 | 1059 | 12.0 |  |
| ★                     | 🗨 |    |  |  | 20.779878                | 104.754814 | 1056 | 16.0 |  |
| ★                     | 🗨 | 2. |  |  |                          |            |      |      |  |

Facet

Text filter

Edit cells

Edit column

Transpose

Sort...

View

Reconcile

Split into several columns...

Add column based on this column...

Add column by fetching URLs...

Rename this column

Remove this column

Move column to beginning

Move column to end

Move column left

Move column right

Split column geotrace into several columns

How to Split Column

☒ by separator

Separator

☐ regular expression

Split into  columns at most (leave blank for no limit)

☐ by field lengths

List of integers separated by commas, e.g., 5, 7, 15

After Splitting

☒ Guess cell type

☒ Remove this column

OK

Cancel



Các giá trị trong cột “geotrace” được tách biệt với nhau bởi khoảng trắng và có 4 giá trị. Chính vì vậy biểu tượng sẽ là **dấu cách và tách ra làm 4 cột (Split into 4 columns)**

| 2 records             |    |                          |  |  |  |  |  |  |  |  |                  |  |
|-----------------------|----|--------------------------|--|--|--|--|--|--|--|--|------------------|--|
| Show as: rows records |    | Show: 5 10 25 50 records |  |  |  |  |  |  |  |  | « first < previo |  |
| ▼ All                 |    |                          |  |  |  |  |  |  |  |  |                  |  |
| ☆                     | 1. |                          |  |  |  |  |  |  |  |  |                  |  |
| ☆                     |    |                          |  |  |  |  |  |  |  |  |                  |  |
| ☆                     |    |                          |  |  |  |  |  |  |  |  |                  |  |
| ☆                     |    |                          |  |  |  |  |  |  |  |  |                  |  |
| ☆                     |    |                          |  |  |  |  |  |  |  |  |                  |  |
| ☆                     |    |                          |  |  |  |  |  |  |  |  |                  |  |
| ☆                     |    |                          |  |  |  |  |  |  |  |  |                  |  |
| ☆                     |    |                          |  |  |  |  |  |  |  |  |                  |  |
| ☆                     |    |                          |  |  |  |  |  |  |  |  |                  |  |
| ☆                     |    |                          |  |  |  |  |  |  |  |  |                  |  |
| ☆                     | 2. |                          |  |  |  |  |  |  |  |  |                  |  |

Kết quả cho thấy dữ liệu trong cột “geotrace” sẽ được chia tách thành bốn cột như hình trên

5. Hoán vị hàng và cột (Transposing rows and columns)



Dữ liệu đôi khi không sắp xếp theo hàng và cột như ý muốn. Chính vì vậy có nhiều cách khác nhau để sắp xếp. Sự lựa chọn phụ thuộc vào tình huống.  
Ví dụ: Nhìn vào dữ liệu chúng ta thấy có sự tách biệt là có hai cột “geotrace” và “repeat”

1 records

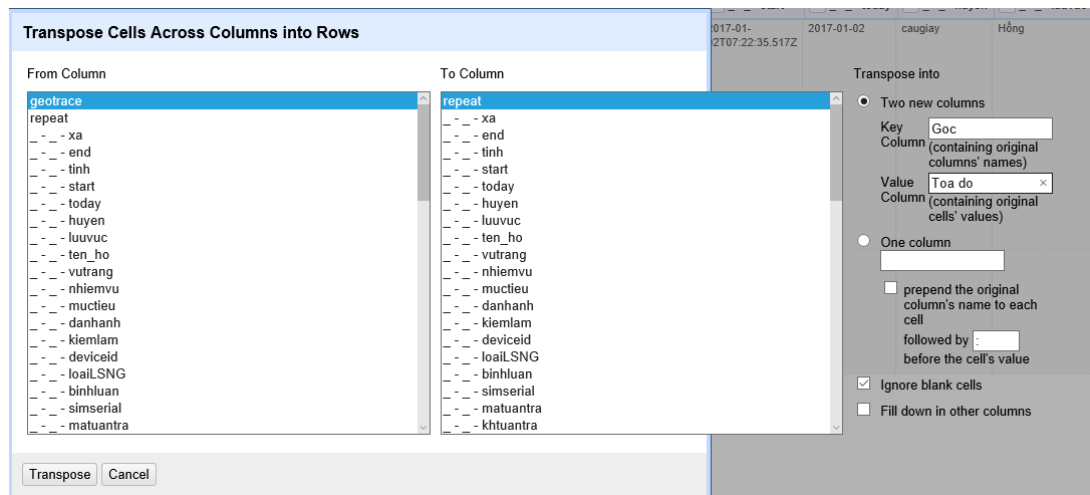
Show as: rows records

Show: 5 10 25 50 records

« first < previous 1 - 1 next > last »

| ▼ All   | ▼ geotrace   | ▼ repeat   | ▼ _ _ _ - xa | ▼ _ _ _ - end            | ▼ _ _ _ - tỉnh | ▼ _ _ _ - start          | ▼ _ _ _ - today      | ▼ _ _ _ - huyen | ▼ _ _ _ - lưu vực | ▼ _ _ _ - ten_ho |    |
|---|--|--|--------------|--------------------------|----------------|--------------------------|----------------------|-----------------|-------------------|------------------|----|
| ☆  | 1. 20.985466 105.829<br>-5 8.0;20.985267<br>105.82919299999999<br>-5 8.0;20.985163<br>105.82937 -3<br>8.0;20.985106<br>105.829467 -2<br>8.0;20.985208<br>105.829599 -1<br>8.0;20.98534<br>105.829787 -1<br>8.0;20.985332<br>105.829983 -2<br>8.0;20.985231<br>105.83015599999999<br>-1 8.0;20.985231<br>105.830155 | 20.985307<br>105.830083 2<br>24.0;20.985415<br>105.83017 1<br>8.0;20.985585<br>105.830079 1<br>12.0;20.985751<br>105.829979 1<br>16.0;20.985844<br>105.829893 1<br>32.0;20.985951<br>105.829939 1<br>32.0;20.986047<br>105.830078 1 16.0;<br><br>20.985809 105.83045<br>2 32.0;20.985888<br>105.83054999999999<br>2 48.0;20.985888<br>105.83054999999999<br>0 48.0;20.985888<br>105.83054999999999<br>0 48.0;20.985877<br>105.830574 1 16.0; | quanhwa      | 2017-01-02T07:40:36.437Z | hanoi          | 2017-01-02T07:22:35.517Z | <a href="#">edit</a> | 2017-01-02      | caugiaiay         | Hồng             | no |
| ☆  |  |  |              |                          |                |                          |                      |                 |                   |                  |    |

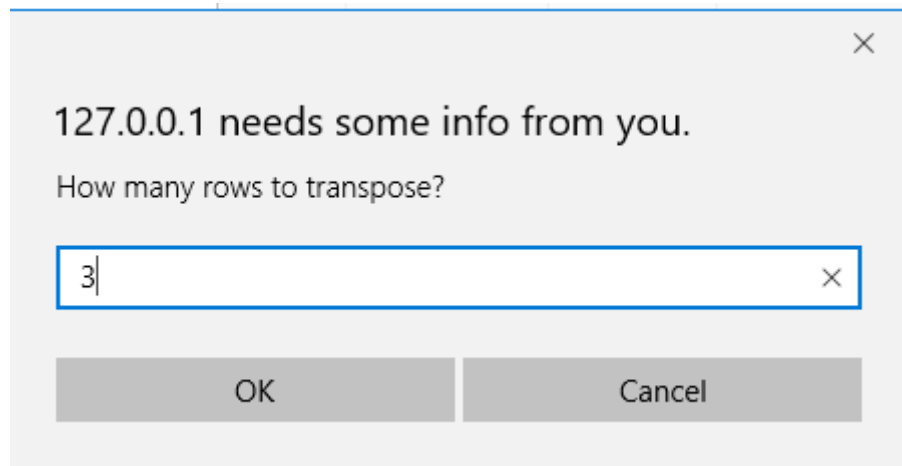
Chúng ta cần gộp hai cột làm một cột. Truy cập đến **Transpose → Transpose cells across columns into rows** . Phần mềm sẽ mở ra hội thoại



Chúng ta sẽ chọn trong ô “From Column” trường thông tin “geotrace “ và trong ô “To Column” là trường thông tin “repeat” rồi chọn “Transpose info” lựa chọn thứ nhất “Two new column” như ảnh trên. Sau đó kích vào nút “Transpose”. Kết quả như ảnh dưới ta được giá trị gộp chung vào một cột

| All | Goc | Toa do              | - - - xa | - - - end | - - - tinh | - - - start | - - - today | - - - huyen | - - - luuvuc | - - - ten_ho | - - - |
|-----|-----|---------------------|----------|-----------|------------|-------------|-------------|-------------|--------------|--------------|-------|
|     |     | 105.82937 -3        |          |           |            |             |             |             |              |              |       |
|     |     | 8.0.20.985106       |          |           |            |             |             |             |              |              |       |
|     |     | 105.829467 -2       |          |           |            |             |             |             |              |              |       |
|     |     | 8.0.20.985208       |          |           |            |             |             |             |              |              |       |
|     |     | 105.829599 -1       |          |           |            |             |             |             |              |              |       |
|     |     | 8.0.20.98534        |          |           |            |             |             |             |              |              |       |
|     |     | 105.829787 -1       |          |           |            |             |             |             |              |              |       |
|     |     | 8.0.20.985332       |          |           |            |             |             |             |              |              |       |
|     |     | 105.829983 -2       |          |           |            |             |             |             |              |              |       |
|     |     | 8.0.20.985231       |          |           |            |             |             |             |              |              |       |
|     |     | 105.8301559999999   |          |           |            |             |             |             |              |              |       |
|     |     | -1 8.0.20.985231    |          |           |            |             |             |             |              |              |       |
|     |     | 105.830155          |          |           |            |             |             |             |              |              |       |
|     |     | 20.985307           |          |           |            |             |             |             |              |              |       |
|     |     | 105.830083 2        |          |           |            |             |             |             |              |              |       |
|     |     | 24.0.20.985415      |          |           |            |             |             |             |              |              |       |
|     |     | 105.83017 1         |          |           |            |             |             |             |              |              |       |
|     |     | 8.0.20.985585       |          |           |            |             |             |             |              |              |       |
|     |     | 105.830079 1        |          |           |            |             |             |             |              |              |       |
|     |     | 12.0.20.985751      |          |           |            |             |             |             |              |              |       |
|     |     | 105.829879 1        |          |           |            |             |             |             |              |              |       |
|     |     | 16.0.20.985844      |          |           |            |             |             |             |              |              |       |
|     |     | 105.829893 1        |          |           |            |             |             |             |              |              |       |
|     |     | 32.0.20.985951      |          |           |            |             |             |             |              |              |       |
|     |     | 105.829939 1        |          |           |            |             |             |             |              |              |       |
|     |     | 32.0.20.986047      |          |           |            |             |             |             |              |              |       |
|     |     | 105.830078 1 16.0;  |          |           |            |             |             |             |              |              |       |
|     |     | 20.985809 105.83045 |          |           |            |             |             |             |              |              |       |
|     |     | 2.32.0.20.985888    |          |           |            |             |             |             |              |              |       |
|     |     | 105.8305499999999   |          |           |            |             |             |             |              |              |       |
|     |     | 2.48.0.20.985888    |          |           |            |             |             |             |              |              |       |
|     |     | 105.8305499999999   |          |           |            |             |             |             |              |              |       |
|     |     | 0.48.0.20.985888    |          |           |            |             |             |             |              |              |       |
|     |     | 105.8305499999999   |          |           |            |             |             |             |              |              |       |
|     |     | 0.48.0.20.985877    |          |           |            |             |             |             |              |              |       |
|     |     | 105.830574 1 16.0   |          |           |            |             |             |             |              |              |       |

Tương tự nếu muốn hoán vị cột thành hàng ta truy cập đến **Transpose → Transpose cells in rows into columns**. Nhập số cột mà muốn tách










Kết quả cột “Tọa độ” sẽ được tách ra thành 3 cột như sau

3 records

Show as: rows records

Show: 5 10 25 50 records

| All   | Goc | Toa do 1 | Toa do 2  | Toa do 3   | _ - _ - xa   | _ - _ - end |  |
|---|-----|----------|---|--|--|-------------|--|
|   | 1.  | geotrace | 20.985466 105.829<br>-5 8.0;20.985267<br>105.82919299999999<br>-5 8.0;20.985163<br>105.82937 -3<br>8.0;20.985106<br>105.829467 -2<br>8.0;20.985208<br>105.829599 -1<br>8.0;20.98534<br>105.829787 -1<br>8.0;20.985332<br>105.829983 -2<br>8.0;20.985231<br>105.83015599999999<br>-1 8.0;20.985231<br>105.830155 | 20.985307<br>105.830083 2<br>24.0;20.985415<br>105.83017 1<br>8.0;20.985585<br>105.830079 1<br>12.0;20.985751<br>105.829979 1<br>16.0;20.985844<br>105.829893 1<br>32.0;20.985951<br>105.829939 1<br>32.0;20.986047<br>105.830078 1<br>16.0; | 20.985809<br>105.83045 2<br>32.0;20.985888<br>105.83054999999999<br>2 48.0;20.985888<br>105.83054999999999<br>0 48.0;20.985888<br>105.83054999999999<br>0 48.0;20.985877<br>105.830574 1 16.0; | quanhua     | 2017-01-02T07:40:36.437Z  |
|   | 2.  | repeat   |   |  |  |             |  |
|   | 3.  | repeat   |   |  |  |             |  |

## 6. Xuất dữ liệu (Exporting Data)

Bạn có thể xuất dữ liệu từ OpenRefine sau khi đã hoàn tất xử lý dưới các định dạng sau:

- Tab separated values (TSV): Các cột dữ liệu cách nhau khoảng cách trắng tab
- Comma separated values (CSV): Các cột dữ liệu cách nhau bởi dấu phẩy
- Excel : Định dạng file Excel
- HTML table : Dưới dạng bảng HTML

Để xuất dữ liệu từ một dự án, chúng ta chỉ cần kích nút “Export” ở góc phía trên bên phải và lựa chọn định dạng mà bạn muốn xuất dữ liệu.

The screenshot shows the OpenRefine web interface. At the top, there's a navigation bar with 'OpenRefine' logo, a test project name, and a 'Permalink' button. Below this is a 'Filter' section with 'Undo / Redo' buttons and a '178' count. The main area displays a table with 2 records. The table has columns: 'Ma diem', 'trung hop', 'Lat', 'Long', 'Time', 'Date', 'Ten tinh', 'Ten xa', 'Ten huyen', and 'Lu'. The first record is 'Gn1' with 'TH1' and coordinates '20.985485', '105.83013'. The second record is 'Gn2' with 'TH2' and coordinates '20.985466', '105.829'. The 'Export' button is visible in the top right corner, and a dropdown menu is open, showing options: 'Export project', 'Tab-separated value', 'Comma-separated value', 'HTML table', 'Excel (.xls)', 'Excel 2007+ (.xlsx)', 'ODF spreadsheet', 'Triple loader', 'MQLWrite', 'Custom tabular exporter...', and 'Templating...'. The 'Export project' option is highlighted.

## 7. Quay lại/Làm lại (Undo/Redo)

**Chức năng Undo/Redo:** Chức năng này cho bạn linh hoạt để chỉnh sửa và làm lại các bước làm trước khi làm lỗi. Để quay lại một vài bước trước khi xảy ra lỗi, lựa chọn các hoạt động trong danh sách mà bạn muốn quay lại và bạn kích lên số bước mà bạn muốn quay lại. Công cụ sẽ tự động đưa bạn đến các bước làm trước đó một cách dễ dàng hơn. Tất cả các bước mà bạn chọn có thể được làm lại hoặc ngược lại. Thậm chí chức năng Redo (làm lại) cho phép bạn đi đến bất kỳ bước nào bạn muốn.

**a) Trước khi Undo/Redo:**

Bạn có thể nhìn thấy ở ô bên trái của màn hình có 5 bước mà chúng được thực hiện bởi người sử dụng. Ngoài những cái trên, cái cuối cùng là màu xám. Bước số 5 là màu xám chỉ ra rằng Undo đã được thực thi một lần. Để chỉ ra điểm này, chúng ta thực hành Undo một lần nữa.

Vì vậy sau đây là một hình trước khi thực thi chức năng Undo và theo sau là dữ liệu được hiển thị.

The screenshot shows the Google Refine interface. At the top, it says "Google refine us 500 Copy.xlsx Perm". Below this is a "Facet / Filter" section with a tab for "Undo / Redo 4". There are buttons for "Extract..." and "Apply...". A "Filter:" input field is present. Below the filter, a list of steps is shown:

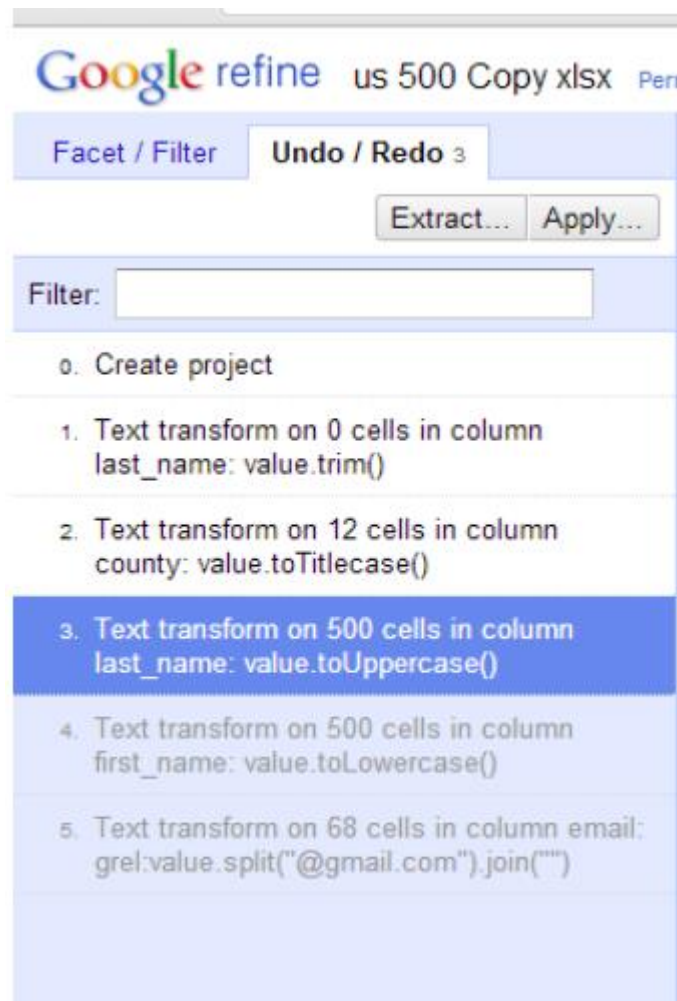
0. Create project
1. Text transform on 0 cells in column last\_name: value.trim()
2. Text transform on 12 cells in column county: value.toTitlecase()
3. Text transform on 500 cells in column last\_name: value.toUppercase()
4. Text transform on 500 cells in column first\_name: value.toLowercase()
5. Text transform on 68 cells in column email: grel.value.split("@gmail.com").join("")

Below the list is a table with 500 rows. The table has columns: first\_name, last\_name, company\_name, address, city, county, state, zip, phone1, phone2, email, and web. The first 10 rows are visible:

|     | first_name | last_name | company_name              | address               | city        | county         | state | zip   | phone1       | phone2       | email                         | web        |
|-----|------------|-----------|---------------------------|-----------------------|-------------|----------------|-------|-------|--------------|--------------|-------------------------------|------------|
| 1.  | james      | BUTT      | Benton, John B Jr         | 6649 N Blue Gum St    | New Orleans | Orleans        | LA    | 70116 | 504-621-8927 | 504-645-1427 | jbutt@gmail.com               | http://www |
| 2.  | Josephine  | DARAKJY   | Chanay, Jeffrey A         | 4 B Blue Ridge Blvd   | Brighton    | Livingston     | MI    | 48116 | 810-292-9388 | 810-374-9840 | josephine_darakjy@darakjy.org | http://www |
| 3.  | art        | VENERE    | Chemel, James L           | 8 W Cerritos Ave #54  | Bridgeport  | Gloucester     | NJ    | 8014  | 856-636-8749 | 856-264-4130 | art@venere.org                | http://www |
| 4.  | lenna      | PAPROCKI  | Feltz Printing Service    | 639 Main St           | Anchorage   | Anchorage      | AK    | 99501 | 907-385-4412 | 907-921-2010 | lpaprocki@hotmail.com         | http://www |
| 5.  | donette    | FOLLER    | Printing Dimensions       | 34 Center St          | Hamilton    | Butler         | OH    | 45011 | 513-570-1893 | 513-549-4561 | donette.foller@cox.net        | http://www |
| 6.  | simona     | MORASCA   | Chapman, Ross E           | 3 Mcauley Dr          | Ashland     | Ashland        | OH    | 44805 | 419-503-2404 | 419-800-6759 | simona@morasca.com            | http://www |
| 7.  | mitsue     | TOLLNER   | Mortong Associates        | 7 Eads St             | Chicago     | Cook           | IL    | 60632 | 773-573-6914 | 773-924-8565 | mitsue_tollner@yahoo.com      | http://www |
| 8.  | leota      | DILLIARD  | Commercial Press          | 7 W Jackson Blvd      | San Jose    | Santa Clara    | CA    | 95111 | 408-752-3500 | 408-813-1105 | leota@hotmail.com             | http://www |
| 9.  | sage       | WIESER    | Truhlar And Truhlar Attys | 5 Boston Ave #88      | Sioux Falls | Minnehaha      | SD    | 57105 | 605-414-2147 | 605-794-4895 | sage_wieser@cox.net           | http://www |
| 10. | kris       | MARRER    | King, Christopher A Esq   | 228 Runamuck Pl #2808 | Baltimore   | Baltimore City | MD    | 21224 | 410-655-8723 | 410-804-4694 | kris@gmail.com                | http://www |

**b) Sau khi Undo/Redo**

Bạn có thể nhìn thấy 5 bước giống nhau với 2 bước màu xám: bước số 4 và số 4. Ở đây chúng ta có thể thấy rằng chức năng Undo được thực thi và bạn cũng có thể nhìn thấy dữ liệu được biến đổi tới bước trước đó; Chữ cái đầu tiên của cột “first\_name” đã được biến đổi thành chữ hoa.



500 rows

Extensions: Freebase

Show as: rows records

Show: 5 10 25 50 rows

« first » previous 1 - 10 next » last »

| ▼ All | ▼ first_name | ▼ last_name | ▼ company_name            | ▼ address            | ▼ city      | ▼ county       | ▼ state | ▼ zip | ▼ phone1     | ▼ phone2     | ▼ email                       | ▼ web      |
|-------|--------------|-------------|---------------------------|----------------------|-------------|----------------|---------|-------|--------------|--------------|-------------------------------|------------|
| ☆ 1.  | James        | BUTT        | Benton, John B Jr         | 6649 N Blue Gum St   | New Orleans | Orleans        | LA      | 70116 | 504-621-8927 | 504-845-1427 | jbutt@gmail.com               | http://www |
| ☆ 2.  | Josephine    | DARAKJY     | Chanay, Jeffrey A Esq     | 4 B Blue Ridge Blvd  | Brighton    | Livingston     | MI      | 48116 | 810-292-9388 | 810-374-9840 | josephine_darakjy@darakjy.org | http://www |
| ☆ 3.  | Art          | VENERE      | Chemel, James L Cpa       | 8 W Cerritos Ave #54 | Bridgeport  | Gloucester     | NJ      | 8014  | 856-636-8749 | 856-264-4130 | art@venere.org                | http://www |
| ☆ 4.  | Lenna        | PAPROCKI    | Feltz Printing Service    | 639 Main St          | Anchorage   | Anchorage      | AK      | 99501 | 907-385-4412 | 907-921-2010 | lpaprocki@hotmail.com         | http://www |
| ☆ 5.  | Donette      | FOLLER      | Printing Dimensions       | 34 Center St         | Hamilton    | Butler         | OH      | 45011 | 513-570-1893 | 513-549-4561 | donette.foller@cox.net        | http://www |
| ☆ 6.  | Simona       | MORASCA     | Chapman, Ross E Esq       | 3 McAuley Dr         | Ashland     | Ashland        | OH      | 44805 | 419-503-2484 | 419-800-6759 | simona@morasca.com            | http://www |
| ☆ 7.  | Mitsue       | TOLLNER     | Morlong Associates        | 7 Eads St            | Chicago     | Cook           | IL      | 60632 | 773-573-6914 | 773-924-8565 | mitsue_tollner@yahoo.com      | http://www |
| ☆ 8.  | Leota        | DILLIARD    | Commercial Press          | 7 W Jackson Blvd     | San Jose    | Santa Clara    | CA      | 95111 | 408-752-3500 | 408-813-1105 | leota@hotmail.com             | http://www |
| ☆ 9.  | Sage         | WIESER      | Truhlar And Truhlar Attys | 5 Boston Ave #88     | Sioux Falls | Minnehaha      | SD      | 57105 | 605-414-2147 | 605-794-4895 | sage_wieser@cox.net           | http://www |
| ☆ 10. | Kris         | MARRER      | King, Christopher A Esq   | 228 Runamuck Pl      | Baltimore   | Baltimore City | MD      | 21224 | 410-655-8723 | 410-804-4694 | kris@gmail.com                | http://www |

Tương tự, chức năng Redo cũng được thực thi như vậy. Chức năng Redo được sử dụng để làm lại một hoạt động mà bạn đã không làm.



**Phụ lục I. MÃ KHÓA CỦA CÁC TRƯỜNG DỮ LIỆU**

| STT | Mã khóa          | Trường dữ liệu   |
|-----|------------------|--|
| 1   | matuantra        | Mã hiệu tuần tra   |
| 2   | khtuantra        | Kế hoạch tuần tra  |
| 3   | phuongthuc       | Phương thức tuần tra   |
| 4   | phuongtien       | Phương tiện tuần tra   |
| 5   | vutrang          | Tuần tra có được trang bị vũ khí   |
| 6   | tinhh            | Tên tỉnh   |
| 7   | huyen            | Tên huyện  |
| 8   | xa               | Tên xã   |
| 9   | luuvuc           | Tên lưu vực  |
| 10  | tentruongnhom    | Tên trưởng nhóm  |
| 11  | ten_ho           | Số hộ/cá nhân đi tuần tra nhiều hơn 1  |
| 12  | songuoithamgia   | Số người tham gia  |
| 13  | repeat_ten       | Tên hộ/cá nhân   |
| 14  | nhomttbvr        | Tên nhóm tuần tra BVR  |
| 15  | nhiemvu          | Nhiệm vụ tuần tra  |
| 16  | muctieu          | Mục tiêu tuần tra  |
| 17  | danhanh          |  |
| 18  | suthamgiakl      | Sự tham gia của cán bộ Kiểm lâm  |
| 19  | kiemlam          | Tên cán bộ Kiểm lâm  |
| 20  | suthamgiaCBLN    | Sự tham gia của cán bộ Lâm nghiệp  |
| 21  | lamnghiep        | Tên cán bộ Lâm nghiệp  |
| 22  | kethoplayLSNG    | Kết hợp lấy lâm sản ngoài gỗ khi tuần tra  |
| 23  | loaiLSNG         | Loại lâm sản ngoài gỗ  |
| 24  | nguoi1a1         | Người lạ trong thôn  |
| 25  | nguoi1a2         | Người lạ ngoài thôn  |
| 26  | yes              | Có   |
| 27  | no               | Không  |
| 28  | mucdich1         | Lấy lâm sản ngoài gỗ (Lấy măng và rau, lấy củi, lấy mật ong, lấy cây thuốc và nấm, lấy song mây,...) |
| 29  | mucdich2         | Lấy gỗ   |
| 30  | mucdich3         | Săn bắn  |
| 31  | mucdich4         | Các hình thức khác   |
| 32  | dungcu1          | Súng   |
| 33  | dungcu2          | Cưa máy  |
| 34  | dungcu3          | Cưa tay  |
| 35  | dungcu4          | Gùi  |
| 36  | dungcu5          | Dụng cụ khác   |
| 37  | hinhhthucx11     | Tuyên truyền vận động  |
| 38  | hinhhthucx12     | Gọi điện cho Kiểm lâm/công an  |
| 39  | hinhhthucx13     | Báo cáo sau khi kết thúc tuần tra  |
| 40  | hinhhthucx14     | Không báo cáo gì cả  |
| 41  | hinhhthucxlchay1 | Báo cáo Kiểm lâm   |
| 42  | hinhhthucxlchay2 | Huy động thôn chữa cháy  |
| 43  | hinhhthucxlchay3 | Hình thức khác   |
| 44  | dauhieuchatcay1  | Ken cây trái phép  |
| 45  | dauhieuchatcay2  | Phát quang cây bụi   |

|    |                     |                                       |
|----|---------------------|---------------------------------------|
| 46 | dauhieuchatcay3     | Lấn trại                              |
| 47 | dauhieuchatcay4     | Nghe thấy tiếng cưa máy               |
| 48 | dauhieuchatcay5     | Bẫy                                   |
| 49 | dauhieuchatcay6     | Dấu vết khác                          |
| 50 | tinhttrang1         | Chặt nhiều nhất                       |
| 51 | tinhttrang2         | Đã đánh dấu                           |
| 52 | hinhtthucxlchatcay1 | Tuyên truyền vận động                 |
| 53 | hinhtthucxlchatcay2 | Gọi điện cho Kiểm lâm/công an         |
| 54 | hinhtthucxlchatcay3 | Báo cáo sau khi kết thúc tuần tra     |
| 55 | hinhtthucxlchatcay4 | Không báo cáo gì cả                   |
| 56 | nuonggray1          | Mới                                   |
| 57 | nuonggray2          | Rẫy cũ đang bỏ hoang                  |
| 58 | nuonggray3          | Rẫy đang làm                          |
| 59 | TH1                 | Bắt gặp con người trực tiếp           |
| 60 | TH2                 | Điểm dễ xảy ra cháy                   |
| 61 | TH3                 | Điểm nóng phá rừng                    |
| 62 | TH4                 | Điểm nóng mất rừng                    |
| 63 | TH5                 | Điểm nóng chuyển đổi mục đích sử dụng |