



DISKOMINFO

Dinas Komunikasi, Informatika & Statistik
Pemerintah Kota Banda Aceh



Materi 3. Pembersihan Data



Table of Content

1. Home
2. Table of Content
3. Tools/Software
4. Tahapan Memulai
5. Open Refine : Facet
6. Open Refine : Transform
7. Open Refine : Cluster & Edit
8. Contoh File dan Manual
Open Refine
9. End



Tools/Software

- Notepad++
- Ms Excel (or Libreoffice Calc)
- Open Refine :
<http://openrefine.org/download.html>
- Java Runtime Environmental (JRE)

Deskripsi Open Refine :

- Alat Bantu Pembersihan Data
- Local server java Web Based
- Awalnya bernama Google Refine (s/d 2012)
- Tersedia utk Windows, Linux & Mac
- Format file : TSV, CSV, *SV, Excel (.xls and .xlsx), JSON, XML, RDF as XML, dan Google Data

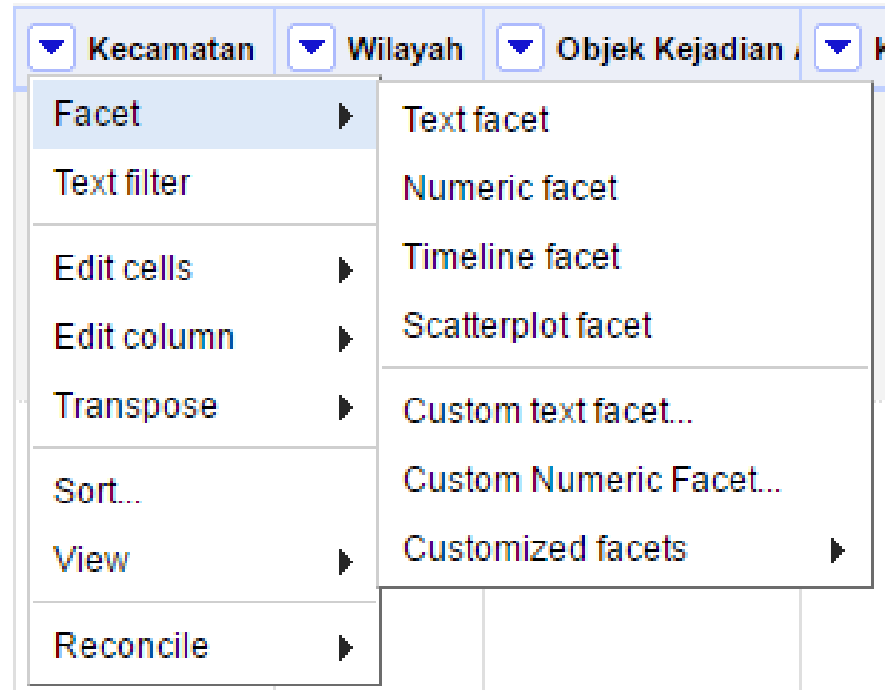


Tahapan Memulai Project Open Refine

1. Extract openrefine-win-2.6-beta.1 yg telah didownload/dicopy
2. Jalankan openrefine.exe
3. Akan terbuka command prompt berisi baris perintah java server
4. Pada browser akan terbuka alamat <http://127.0.0.1:3333/>
5. Open Refine telah siap utk digunakan
6. Pilih **Create Project**
7. Pilih **Get Data from This Computer**
8. Klik Choose File, pilih file pelatihannya kemudian tekan Next
9. Berikan nama Project di Kanan Atas
10. Open Refine telah siap digunakan untuk membersihkan data.

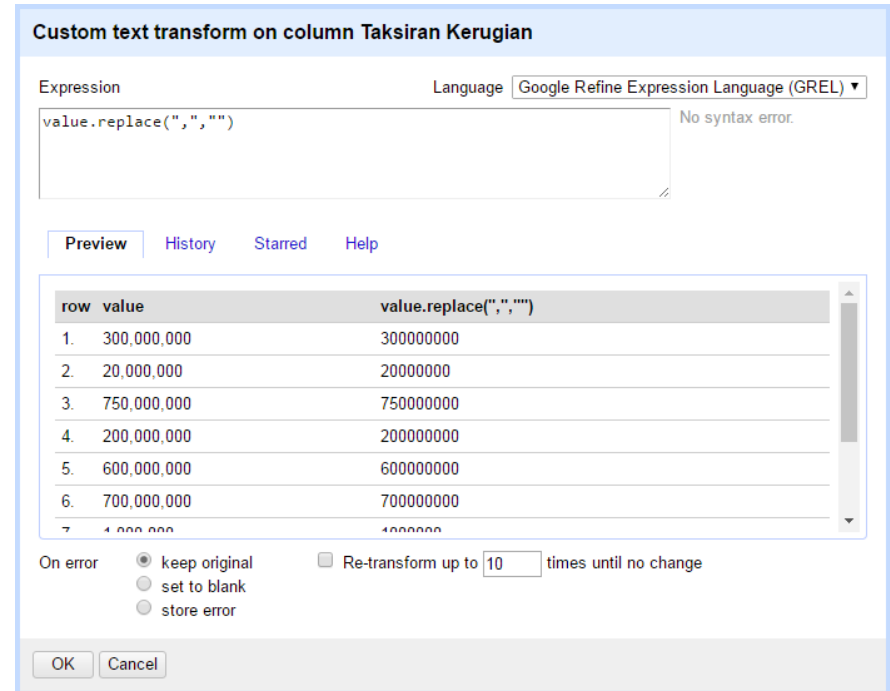
Open Refine : Facet

- Berfungsi utk mengelompokkan data berdasarkan kolom yang dipilih
- Klik panah bawah pada nama kolom yang akan dikelompokkan
- Pilih tipe datanya (text, numeric, etc)
- Maka pada bagian facet/filter akan ditampilkan data apa saja pada kolom tersebut dikelompokkan
- Scroll kebawah dan edit agar data seragam



Open Refine : Transform

- Berfungsi utk mengubah (*transform*) data berdasarkan kolom yang dipilih
- Klik panah bawah pada nama kolom yang akan diubah
- Pilih **Edit Cells**
- Pilih **Transform** (bila ingin mengubah data yang cukup kompleks)
- Akan muncul window baru untuk memasukkan parameter transformasinya.



Open Refine : Cluster and Edit

- Berfungsi utk mengelompokkan (*Cluster*) data berdasarkan kolom yang dipilih
- Klik panah bawah pada nama kolom yang akan dikelompokkan
- Pilih **Edit Cells**
- Pilih **Cluster and Edit**
- Akan muncul window baru dengan kolom-kolom yang diaggap sama/mirip oleh Open Refine.

▼ Alamat Kejadian	▼ RT	▼ RW	▼ Kelurahan
Facet		11	Semper Barat
Text filter			
Edit cells	▶	Transform...	
Edit column	▶	Common transforms	▶
Transpose	▶	Fill down	
Sort...		Blank down	
View	▶	Split multi-valued cells...	
Reconcile	▶	Join multi-valued cells...	
Jl. Indra Loka VI	8	Cluster and edit...	

Cluster & Edit column "Alamat Kejadian"

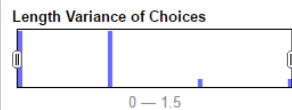
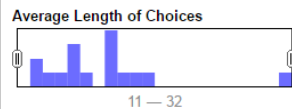
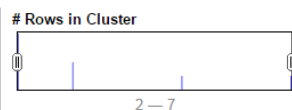
This feature helps you find groups of different cell values that might be alternative representations of the same thing. For example, the two strings "New York" and "new york" are very likely to refer to the same concept and just have capitalization differences, and "GÄfidel" and "Godel" probably refer to the same person. [Find out more](#)

Method key collision

Keying Function fingerprint

16 clusters found

Cluster Size	Row Count	Values in Cluster	Merge?	New Cell Value
2	2	- Jl. P.K.P Raya (1 rows) - Jl. Raya PKP (1 rows)	☐	Jl. P.K.P Raya
2	3	- Jl. T.B Simatupang (2 rows) - Jl. TB. Simatupang (1 rows)	☐	Jl. T.B Simatupang
2	7	- Jl. Raya Bogor (6 rows) - Jl. Raya Bogor (1 rows)	☐	Jl. Raya Bogor
2	7	- Pluit Timur (5 rows) - Pluit Timur (2 rows)	☐	Pluit Timur
2	5	- Jl. Ciputat Raya (4 rows) - Jl. Ciputat Raya (1 rows)	☐	Jl. Ciputat Raya
2	3	- Jl. Kembangan Raya (2 rows) - Jl. Raya Kembangan (1 rows)	☐	Jl. Kembangan Raya
2	5	- Jl. Raya Teluk Gong (4 rows) - Jl. Raya Teluk Gong (1 rows)	☐	Jl. Raya Teluk Gong



Select All Unselect All

Merge Selected & Re-Cluster

Merge Selected & Close

Close



Contoh File dan Manual Open Refine

Contoh File Pembersihan :

- pembersihan_data_tugas1.csv

Referensi Manual Open Refine Lebih Lanjut :

- http://enipedia.tudelft.nl/wiki/OpenRefine_Tutorial



DISKOMINFO

Dinas Komunikasi, Informatika & Statistik
Pemerintah Kota Banda Aceh



Sekian dan Terima Kasih