

What I did in using OpenRefine to refine data "HK's Electric's Inspection":

1. Text facet on column "defect\_section\_en"; cluster and edit some inconsistent words with 'fingerprint' and 'ngram-fingerprint' keying function.

"MS", "Main switch", "main switch", "main switches" are refined to "Main Switch"(57)

"FC", "Final circuits" to "Final Circuits" (32)

"RS" to "Rising Mains"(13)

"MP", "Meter position", "Meter positions" to "Meter Position" (39)

"PIC" to "Pre-Inspection Check"(11)

"Other"(3)

Besides, a row "blank" is totally empty without any data, so it was deleted.

2. Text facet, cluster and edit on column "defect\_type".

"EO", and "Electricity ordinance" to "Electricity Ordinance"(64)

"SR" to "Supply Rule" (88)

3. Filling in missing text

Also, it was found that 3 rows are blank in column "defect\_type". But the data in other columns except "defect\_type" exist. So, it is necessary to fill in the missing data.

Meanwhile, the column "HKE\_inspection\_defect\_code" indicates what the missing data should be, because the data codes beginning with "1" are mostly "Electricity Ordinance" in column "defect\_type", and data codes beginning with "2" are mostly "Supply Rule". Thus, the missing text in the 3 rows are refilled.