



CLEANING UP MESSY DATA  
WITH  
**OPEN REFINE**

---

Presented by  
Anjum Najmi & Spencer Keralis

# OVERVIEW

- Introduction
  - Installing Open Refine
  - Features
  - Working with Open Refine
-

# INTRODUCTION

- Public data on important social issues
  - Big pattern thinking
  - Use filters & facets based on common characteristics
  - Edit cells by clustering, columns by extending data
  - Understand expressions [GREL](#) Quick Expressions
-

# INSTALLING OPEN REFINE

- Download zip file, uncompress
  - Run .exe file
  - Command window will run in background
  - Switch to command window use Ctrl-c to exit
-

# Download OpenRefine

## OpenRefine Core

Home

Down

Docu

Com

Post

Using O

manual

OpenRe

Twee



New learning

OpenRefine (ex-Google Refine) is a powerful tool for working with messy data [openrefine.org](http://openrefine.org) #cool

Expand



KOMATSU Issei 9h @isseium

Linked Data を生成を補助するWebツール。昔は

```
C:\Users\an0091\Desktop\google-refine-2.5-r2407\google-refine.exe
23:21:51.921 [ refine ] POST /command/core/get-rows <16ms>
23:21:51.976 [ refine ] POST /command/core/compute-facets <55ms>
)
23:22:01.531 [ refine ] GET /command/core/get-history <9555ms>
23:22:01.570 [ refine ] POST /command/core/get-rows <39ms>
23:22:01.607 [ refine ] POST /command/core/compute-facets <37ms>
)
23:22:22.691 [ refine ] POST /command/core/get-rows <21084ms>
23:22:51.942 [ refine ] POST /command/core/compute-clusters <29251ms>
23:22:51.945 [ compute-clusters_command ] computed clusters [binning,fingerprint]
in 3ms (3ms)
23:23:30.684 [ refine ] POST /command/core/compute-clusters <38739ms>
23:23:30.687 [ compute-clusters_command ] computed clusters [binning,fingerprint]
in 3ms (3ms)
23:24:23.178 [ project_manager ] Saving some modified projects ... <52491ms>
23:24:23.207 [ project_utilities ] Saved project '2060998669216' <29ms>
11:52:42.329 [ refine ] GET /command/core/get-operations <44899122ms>
11:52:42.392 [ project ] Loaded project 1428331511041 from disk
in 0 sec(s) <63ms>
```

ve, try

ssue

The following extension are available for OpenRefine:

- [Crowdsourcing extension](#) - by Zemanta
- [DBpedia extension](#) - by Zemanta
- [History tools, pivot tool and scatterplot tool using D3](#) - by VIB-BITS
- [LMF Extension](#) (to be migrated to Apache Marmotta after summer) - by Salzburg Research
- [Named-Entity Recognition](#) - by Ruben Verborgh (Free Your Metadata)

# FEATURES

- Powerful text search & clustering
  - Find & Replace,
  - Group cells, Group groups
  - Sort, View, Reconcile
  - Full undo/redo support
  - Scripting language, web service & JSON support
-

# Working with Open Refine ...

