



استفاده از ابزارهای فن آوری اطلاعات در ثبت سرطان

دکتر کاظم زنده دل
متخصص اپیدمیولوژی سرطان
مسول برنامه ملی ثبت سرطان جمعیتی
قائم مقام و معاون پژوهشی دانشگاه علوم پزشکی تهران

kzندهh@tums.ac.ir

سایر داده ها
Excel

داده های مرگ و میر
Excel

داده های آسیب شناسی
Excel و سپاس

داده های بخش مدارک پزشکی
Excel و سامانه سپاس

تکرار گیری و جمع
Access, Open Refine

تکرار گیری و جمع
Access, Open Refine

تکرار گیری و جمع
Access, Open Refine

تکرار گیری و جمع
Access, Open Refine

جمع و تکرار گیری و آماده سازی
Access, Open Refine,

IARC Check and Analysis

Follow us on: Github Twitter

OpenRefine



A free, open source, powerful tool for working with messy data

Home

Download

Documentation

Community

Post archive

OpenRefine News:
Spring 2016

#Welcome!

OpenRefine (formerly Google Refine) is a powerful tool for working with messy data: cleaning it; transforming it from one format into another; and extending it with web services and external data.

Please note that since October 2nd, 2012, Google is not actively supporting this project, which has now been rebranded to OpenRefine. Project development, documentation and promotion is now fully supported by volunteers. Find out more about the [history of OpenRefine](#) and how you can [help the community](#).

Using OpenRefine - The Book

Using OpenRefine, by Ruben Verborgh and Max De Wilde, offers a great introduction to OpenRefine. Organized by recipes with hands on examples, the book covers the following topics:

1. Import data in various formats
2. Explore datasets in a matter of seconds

Download OpenRefine

You will find on this page a list of OpenRefine distributions and extensions available for download. Are we missing something? Want to fix a typo? You can submit changes (pull request) [from here](#).

Home

Download

Documentation

Community

Post archive

OpenRefine News:
[Spring 2016](#)

OpenRefine News:
[December 2015](#)

OpenRefine News:
[November 2015](#)

OpenRefine News:
[October 2015](#)

OpenRefine News:
[September 2015](#)

Official Distribution

Read the [installation instructions](#)

OpenRefine 2.6

This is the first beta release of OpenRefine 2.6 on Aug 27, 2013. A change log is provided on [the release page](#).

- **Windows kit**, Download, unzip, and double-click on *google-refine.exe*. If you're having issues with the above, try double-clicking on *refine.bat* instead.
- **Mac kit**, Download, open, drag icon into the Applications folder and double click on it.
- **Linux kit**, Download, extract, then type `./refine` to start.

Google Refine 2.5

Stable version released on Dec, 2011

- **Windows kit**, Download, unzip, and double-click on *google-refine.exe*. If you're having issues with the above, try double-clicking on *refine.bat* instead.
- **Mac kit**, Download, open, drag icon into the Applications folder and double click on it. **NOTE:** If you have issues installing Refine on Mac, please refer to [issue 590 - Google Refine 2.5 for mac support java 6 and 7 only](#)
- **Linux kit**, Download, extract, then type `./refine` to start. **NOTE:** Google Refine 2.5 for linux support java 6 and 7 only

- [OpenRefine News: August 2015](#)
- [OpenRefine News: July 2015](#)
- [OpenRefine News: June 2015](#)
- [Mapping OpenRefine Ecosystem](#)
- [2014 survey results](#)
- [A Governance Model for OpenRefine](#)
- [Using OpenRefine: a manual](#)
- [OpenRefine History](#)
- [Tweets about OpenRefine](#)

Other Distributions

The following distributions have been customized for a specific usage or integration with other technologies.

Distribution	Description	Authors
LODRefine	LODRefine is actually OpenRefine with integrated extensions that make transition from tabular data to Linked Data a bit easier. Integrated extensions are: RDF extension, DBpedia extension, Crowdsourcing extension, Stats extension	Sparkica
OpenDataRise	Tool to cleanse and semantify datasets from CKAN repositories. Based on OpenRefine.	Open Data in Trentino
p3-batchrefine	BatchRefine adds batch processing capabilities to OpenRefine and support multiple back end including spark	SpazioDati
SparkonRefine	RefineOnSpark is a driver program to run OpenRefine jobs on the Spark cluster	SpazioDati
Reconciliation-and-Matching-Framework	A framework to allow the matching of string entities using customised sets of transformations and matchers, plus a tool to produce the necessary configurations and another to expose them as OpenRefine reconciliation services.	RBGKew

List of Extensions:

You can install the following extensions to add functionalities to OpenRefine. See [how to install an extension](#).

Extension Name	Description	Authors
RDF extension 0.8 for Google Refine 2.5	RDF Refine - a Google Refine extension for exporting RDF	DERI
RDF extension 0.9 for OpenRefine 2.6	RDF Refine - a Google Refine extension for exporting RDF	DERI
Opentree	An extension for Google/Open Refine to display phylogenetic trees from the Open Tree of Life	nickynicolson

خصوصیات نرم افزار Open Refine

۱. رایگان است و با زیان جاوا نوشته شده است.
۲. انواع فایل های زیر را پشتیبانی می کند. TSV, CSV, *SV, Excel (.xls and .xlsx), JSON, XML, RDF as XML, and Google Data document
۳. کاربردهای مختلفی دارد و بسته های تکمیلی دارد که می توانید استفاده کنید.
۴. بصورت تک ستون تکرار گیری می کند و با توجه وجود کد ملی در برنامه ما کاربرد دارد.
۵. تغییرات را آشپو می کند و قابل برگشت است.

راهنمای تکرار گیری با Open Refine

۱. نرم افزار را اجرا کنید.
۲. فایل مورد نظر را بخوانید.
۳. یک پروژه بسازید.
۴. بر اساس کد ملی داده ها را تکرار گیری کنید.
 ۱. موس را روی ستون روی فلش کوچک عنوان ستون نگه دارید. منوی دستورات باز می شود.
 ۲. در انتها روی دستور reconciliation کلیک کنید. پنجره مربوطه باز می شود.
 ۳. در ستون سمت چپ روش تکرار گیری عنوان شده که باید انتخاب کرده و سپس دستور start reconciliation از پایین صفحه کلیک کنید.
 ۴. در جدول داد ها تکرار گیری انجام شده است.
 ۵. دوباره روی سرستون بایستد و در ردیف اول منوی دستورات روی facet که اولین دستور است کلیک کنید و text facet را انتخاب کنید. منوس مربوط به لیست موارد تکراری مشخص می شود که می توانید شروع به کار کنید.

راهنمای تکرار گیری با Open Refine

۱. روی هر شماره یا مقداری از جدول کلیک کنید جداول اختصاصی آنها می ماند و بقیه موارد پنهان می شود.
۲. می توانید روی هر سلول کلیک کنید و تغییرات را اعمال کنید.
۳. هر ردیف را که می خواهید نگه دارید انتخاب کنید و پرچم یا ستاره سمت چپ را انتخاب کنید
۴. روی ستون اول بایستید و facet پرچم و یا ستاره را انتخاب کنید تا نتیجه را ببینید.
۵. هر کدام را که بخواهید می توانید انتخاب کنید. مثلا در پنجره پرچم اگر روی موارد fals کلیک کنید، ردیفی که انتخاب کرده اید پنهان می شود و بقیه موارد می ماند.
۶. سپس روی ستون بایستید و edit row را انتخاب کرده و با دستور `remove all matching rows` همه ردیف های باقیمانده را حذف کنید.
۷. نتیجه را به هر فمتی که می خواهید اکسپورت کنید.
۸. اگر بخواهید در سمت چپ می توانید پنجره redo/undo را انتخاب کنید و تغییرات را ببینید و در صورت لزوم بازیابی کنید.