从元数据到RDF数据
OpenRefine作为RDF数据转换工具

上海图书馆　夏翠娟
2018年6月 @****大学

主要内容

- OpenRefine简介——使用范围和场景
- OpenRefine下载安装
- OpenRefine数据导入
- OpenRefine数据清洗
- OpenRefine的RDF相关模块的使用
- 知识本体映射配置
- RDF数据导出
- 多源数据混搭

OpenRefine简介

- Google Refine2.5
- OpenRefine 2.6 (Oct 13, 2015)
- 数据剖析（data profiling）
- 数据清洗（data cleaning）
- 数据转换（interactive data transformation tools）
- 数据关联与混搭
- 看起来像EXCEL，用起来像database
- 不方便增加新的行

OpenRefine下载

- 下载：http://openrefine.org/download.html
- 版本
  - OpenRefine 2.6-rc2 Release Candidate 2
  - OpenRefine 2.6 beta 1
  - Google Refine 2.5
  - LODRefine
- 插件
  - RDF extension 0.8 for Google Refine 2.5
  - RDF extension 0.9 for OpenRefine 2.6

OpenRefine安装

- **OpenRefine在Windows的安装**

1）ZIP包解压到某个目录；

2）要运行OpenRefine，双击openrefine.exe文件。

- **OpenRefine在Mac的安装**

1）DMG文件打开磁盘镜像，拖动OpenRefine的图标到Applications目录；

2）双击图标以启动OpenRefine。

OpenRefine在Linux的安装

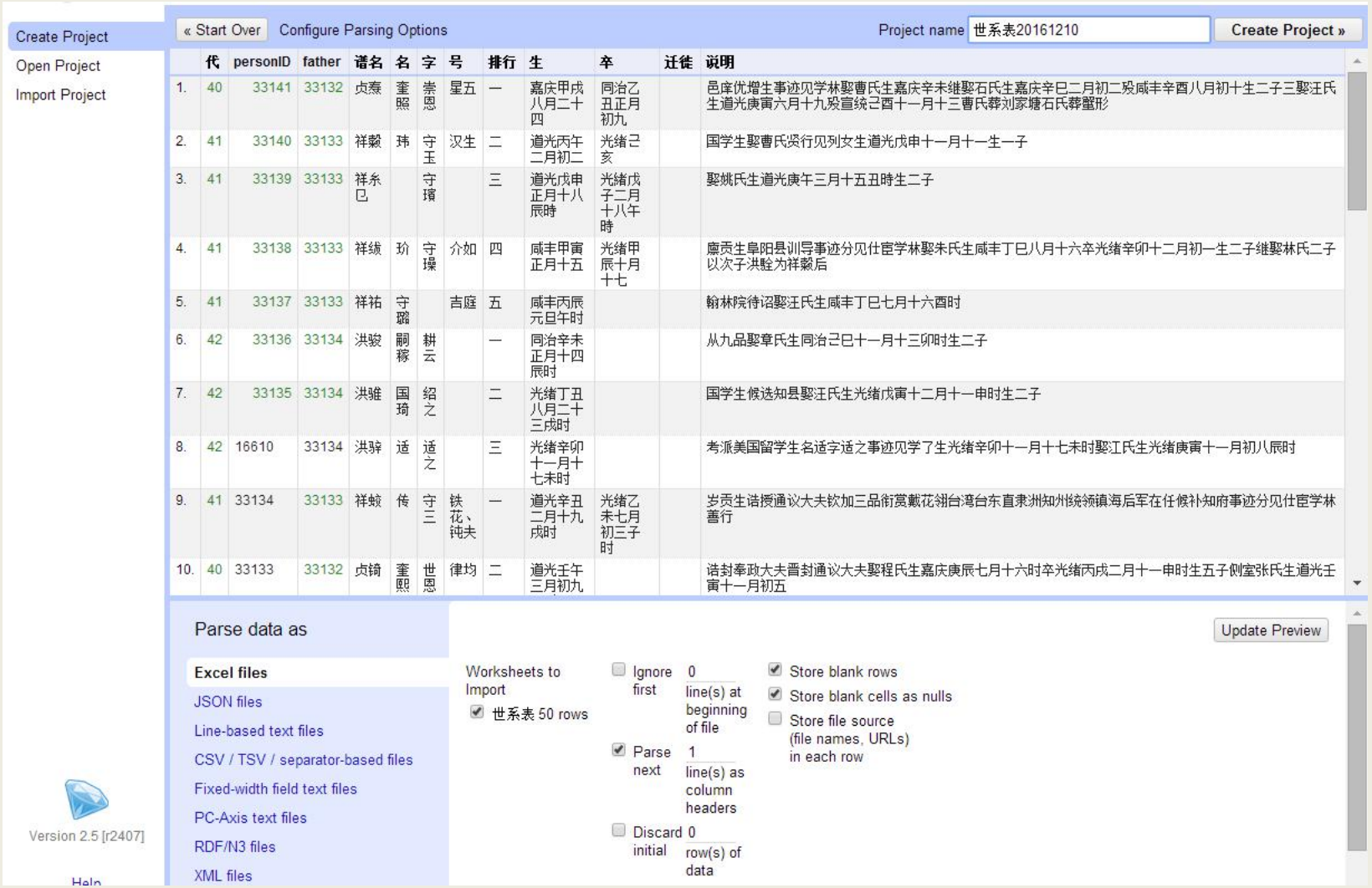
1）gzipped包解压到当前用户的home目录；

2）在终端命令行环境，键入./refine以启动OpenRefine。

OpenRefine安装RDF插件

解压到\google-refine-2.5-r2407\webapp\extensions\rdf-extension

OpenRefine导入数据



OpenRefine数据清洗

GREL（Google Refine Expression Language）

- Variables
- GREL Controls
- GREL Functions overview
- GREL Boolean Functions
- GREL String functions, including parsing, splitting, encoding and hashing
- GREL Array functions
- GREL Math functions
- GREL Date functions
- GREL Other functions including JSON and Jsoup

OpenRefine数据清洗

1. 数据类型转换
substring(toString(value),0,indexOf(toString(value),"."))
2. 生成新的列（Edit column->add column based on this column）
substring(value,0,2)
if(indexOf(value,"娶")>-1,substring(value,indexOf(value,"娶")+1,indexOf(value,"氏")+1),"")
3. 分面浏览
4. 批量修改

OpenRefine数据转换

1. 设置Base URI
2. 设置命名空间前缀
3. 设置“主体”的类型
4. 设置“主体”的URI规则
5. 设置“谓词”及元数据字段与本体属性之间的映射
6. 设置“客体”的取值
7. 随时查看转换结果（RDF Preview）
8. 导出RDF数据

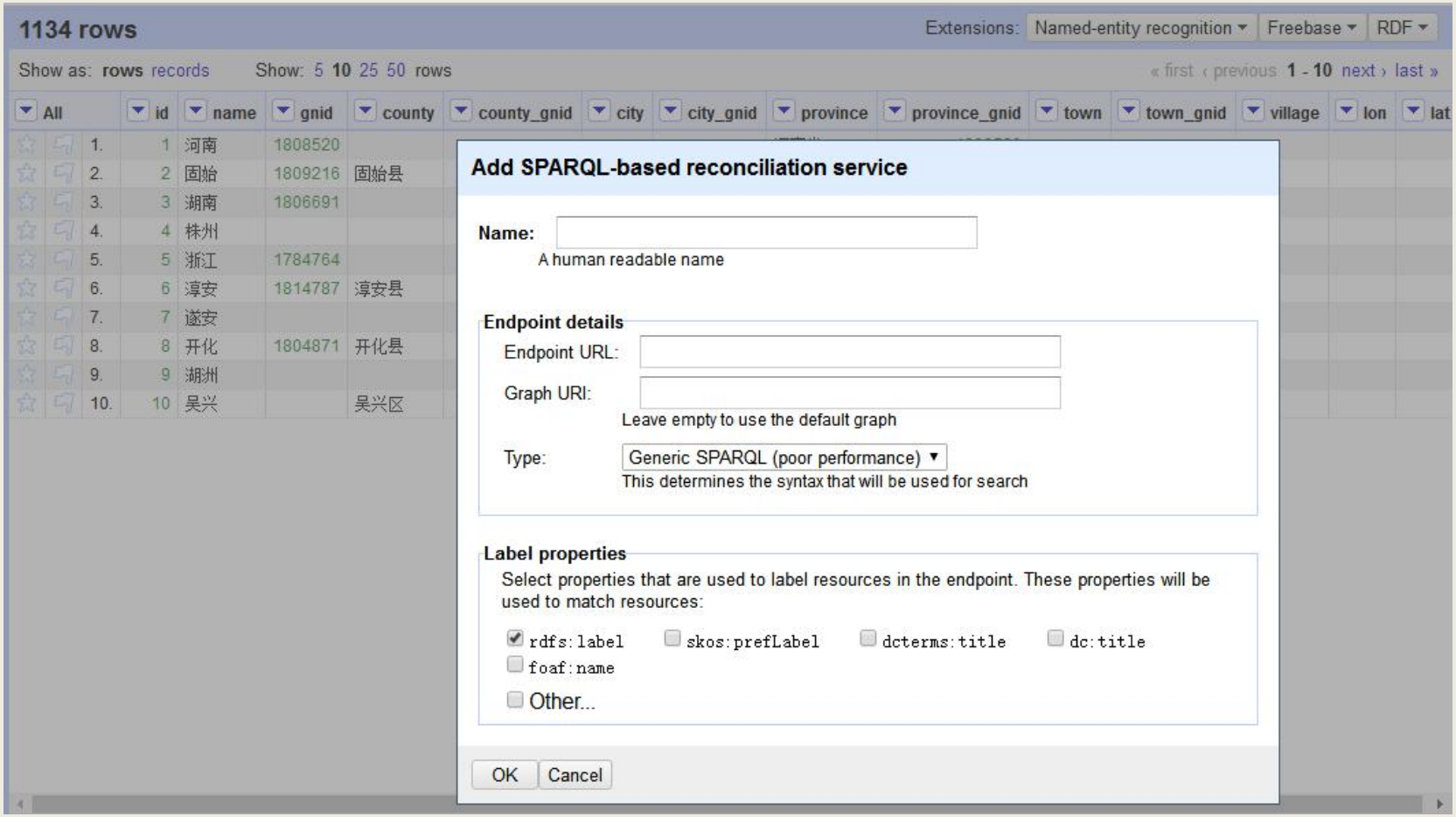
来自本体词表中的“谓词”

属性名（ Name ）	标签（Label）	subPropertyOf	域（Domain）	范围（ Range ）
dfs:label	标签		rdf:Resource	Literal
c:identifier	标识符		shl:Resource	Literal
hl:description	描述		shl:Resource	Literal
hl:temporal	朝代		shl:Resource	shl:Temporal
c:date	日期		shl:Resource	Literal
hl:event	事件		shl:Resource	shl:Event
hl:place	地点		shl:Resource	shl:Place
baf:name	名称		rdf:Resource	Literal
baf:gender	性别		shl:Person	bf:Category
baf:familyName	姓氏		shl:Person	shl:FamilyName
hl:genealogyName	谱名		shl:Person	Literal
hl:name	名称		shl:Agent	shl:Name
hl:courtesyName	字		shl:Person	Literal
hl:pseudonym	号		shl:Person	Literal
hl:orderOfSeniority	行		shl:Person	Literal
hl:posthumousTitle	谥号		shl:Person	Literal
hl:birthday	生于		shl:Person	Literal
hl:deathday	卒于		shl:Person	Literal
hl:birthPlace	出生地		shl:Person	shl:Place
hl:deathPlace	死亡地		shl:Person	shl:Place
hl:ethnicity	民族		shl:Person	bf:Category
hl:nationality	国籍		shl:Person	bf:Category
hl:nativePlace	籍贯		shl:Person	shl:Place
hl:speciality	专长		shl:Person	Literal
hl:briefBiography	小传		shl:Person	Literal
hl:createdWork	著作		shl:Person	Literal

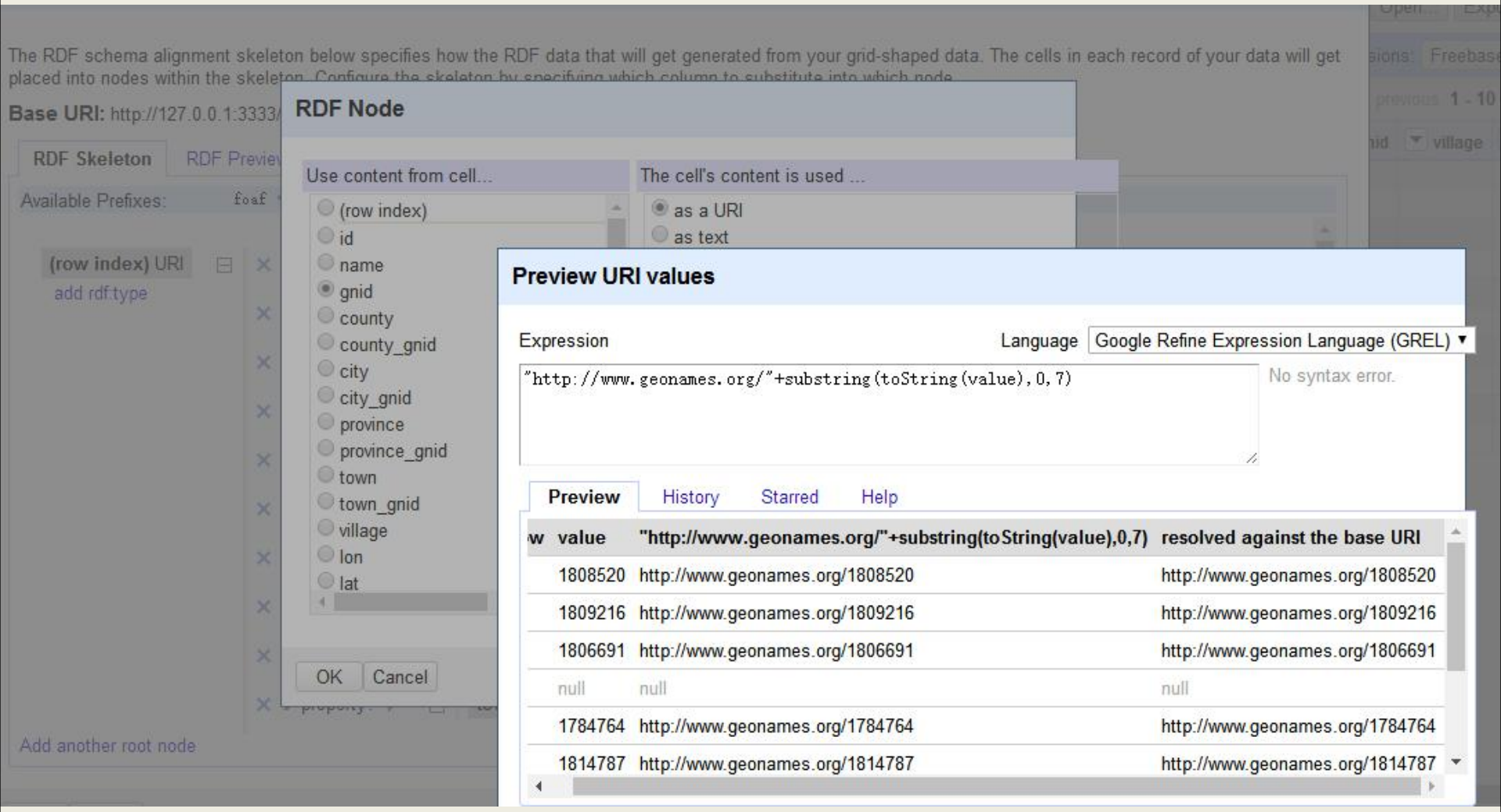
属性名（ Name ）	标签（Label）	subPropertyOf	域（Domain）	范围（ Range ）
hl:briefBiography	小传		shl:Person	Literal
hl:createdWork	著作		shl:Person	Literal
hl:officialExperience	履历		shl:Person	Literal
hl:officialEvent	任职事件	shl:event	shl:Person	shl:OfficialEvent
hl:officialPosition	职衔		shl:OfficialEvent	Literal
prov:startedAtTime	开始时间		prov:Activity	Literal
prov:endedAtTime	结束时间		prov:Activity	Literal
prov:atLocation	发生地点		prov:Activity	prov:Location
hl:migrationEvent	迁徙	shl:event	shl:Person	shl:MigrationEvent
hl:originalLocality	原居地		shl:MigrationEvent	shl:Place
hl:locality	迁居地		shl:MigrationEvent	shl:Place
hl:nameType	名称类别		shl:Name	Literal
el:childOf	父母		shl:Person	shl:Person
el:parentOf	子女		shl:Person	shl:Person
el:spouseOf	配偶		shl:Person	shl:Person
el:friendOf	朋友		shl:Person	shl:Person
hl:source	来源		shl:Resource	shl:Resource
c:source	来源		shl:Resource	Literal
c:references	参考		shl:Resource	Literal
hl:relationSubject	关系主体		shl:Relationship	shl:Person
hl:relationObject	关系人		shl:Relationship	shl:Person
hl:relationType	关系类型		shl:Relationship	Literal

OpenRefine数据混搭

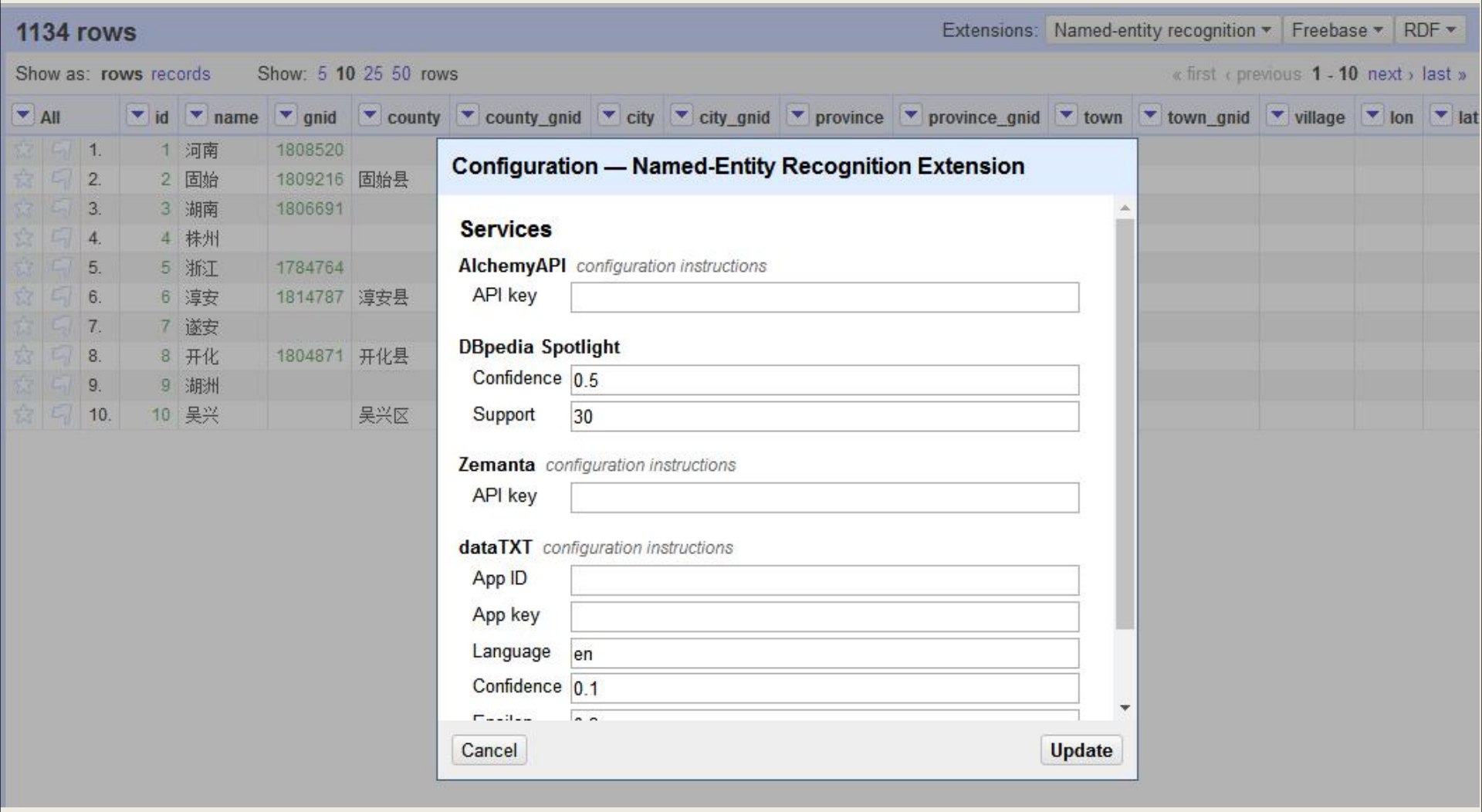
- RDF->adding a reconciliation service



OpenRefine数据关联



OpenRefine实体提取



谢谢！

Q&A

Contact: cjxia@libnet.sh.cn