

Vytváříme RDF z tabulek

Budeme využívat program OpenRefine (<https://openrefine.org/>), který je open source. OpenRefine slouží k transformaci a čištění dat. Abychom mohli pracovat s RDF, potřebujeme rozšíření RDF Transform.

Instalace


Chcete-li instalovat OpenRefine, zvolte verzi podle svého operačního systému také podle toho, zda máte nainstalovanou Javu (stačí JRE). Pokud si nejste jistí, zvolte verzi s Javou.

Rozšíření můžete nainstalovat tak, že stáhnete zip z <https://openrefine.org/extensions> (v našem případě najdete Latest Release na <https://github.com/AtesComp/rdf-transform>) a rozbalíte jeho obsah do adresáře openrefine-<verze>/webapp/extensions. Takže v našem případě přibude do webapp/extensions adresář rdf-transform.

Vytvoření projektu

1. Spustíte OpenRefine
2. Vytvoříte projekt tím, že nahrajete data:

← → ↻ ⓘ 127.0.0.1:3333


 **OpenRefine** *A power tool for working with messy data.*

Create project
Open project
Import project
Language settings

Create a project by importing data. What kinds of data files can I import?
TSV, CSV, *SV, Excel (.xls and .xlsx), JSON, XML, RDF as XML, and Google Data documents a

Get data from: Locate one or more files on your computer to upload:
This Computer Choose Files No file chosen
Web Addresses (URLs)
Clipboard
Database
Google Data

Next »


Version 3.7.6 [182a17a]
Preferences
Help
About

Máte k dispozici soubor OpenRefineExample.csv. Jde o formát Comma Separated Values (CSV), který lze snadno otevřít v textovém editoru i v Excelu.

Zkontrolujte nahraná data a pokud je všechno ok, zvolte Create project.

3. Zkonvertujte data narození na typ datum: Zvolte All/Edit All columns/To date a vyberte sloupec Birthdate.
4. Vytvořte sloupec Name, který bude obsahovat jméno a příjmení pro každého člověka. U Given Name zvolte Edit column/Join columns. Jako separator dejte mezeru:

Join columns

Select and order columns to join

- Given Name
- Family Name**
- Title
- Birthdate
- Organisation

Select all

Deselect all

Select options

Separator between the content of each column

Enter one or more characters, or keep blank to join the columns without separator.

Replace nulls with

Enter one or more characters, or keep blank to replace nulls with blank strings.

Skip nulls.

In separator and nulls substitutes, use \n for new lines, \t for tabulation, \\n for \n, \\t for \t.

Write result in selected column.

Write result in new column named

Delete joined columns.

OK

Cancel

5. Otevřete RDF Transform/Edit RDF. Podívejte se na případný výsledek transformace v Preview.
6. Změňte Base IRI na něco, co dává smysl: <http://examplepersons.org/ontology#> například.
7. Upravte způsob, jakým se indexují instance kliknutím na R: [Index].

RDF Node

Prefix: Choose / None ▾

Content...

[Index]

Family Name

Given Name

Name

Title

Birthdate

Organisation

H-Index

Constant value

Content used...

IRI

Text

Language:

Integer

Double

Date (yyyy-MM-dd)

DateTime (yyyy-MM-ddTHH:mm:ss.SSS)

Boolean

Custom (specify type IRI)

Blank

Expression...

row.index

[Edit & Preview](#)

OK

Cancel

Změňte row.index na "Person"+value.

8. Podívejte se na Preview.
9. Upravte datum na typ date. Podobně upravte H-Index na celé číslo.

RDF Node

Prefix:

Content...

- [Index]
- Family Name
- Given Name
- Name
- Title
- Birthdate
- Organisation
- H-Index
- Constant value

Content used...

- IRI
- Text
- Language:
- Integer
- Double
- Date (yyyy-MM-dd)
- DateTime (yyyy-MM-ddTHH:mm:ss.SSS)
- Boolean
- Custom (specify type IRI)
-
- Blank

Expression...

value

[Edit & Preview](#)

OK

Cancel

10. Propojte třídy se známými ontologiemi. Nabízí se FOAF. Přidejte prefix foaf a URI. Můžete použít <http://prefix.cc/> pro vyhledání URI.
11. Klikněte na :Name a změňte jej na foaf:name, podobně Family Name, Given Name a title.
12. Zkontrolujte Preview.
13. Propojte data s ontologií schema.org - ta obsahuje birthDate i affiliation.
14. Bohužel ani FOAF ani schema.org nenabízejí relaci "má H-Index". Vytvořte ji tedy v naší ontologii - změňte "H-Index" na ":hasHIndex".
15. Zkontrolujte Preview a uložte.
16. Bylo by dobré mít organizace propojené s entitami s URI. K tomu použijte Reconciliation na sloupec Organization. Zvolte Start Reconciliation.
17. Vyberte Wikidata. Prohlédněte nalezené možnosti. Vyberte organization a zahajte hledání. Některé organizace se našly jednoznačně, některé je třeba zjednotřit ručně.

6 rows								Extensions	RDF Transf
Show as: rows records		Show: 5 10 25 50 100 500 1000 rows				« first		« previous	1
All	Family Name	Given Name	Name	Title	Birthdate	Organisation	H-Index		
☆	1.	Mueller	Frank	Frank Mueller	Prof.	1971-12-18T00:00:00Z	International Federation for Human Rights	21	
							<input type="checkbox"/> International Federation for Human Rights (100) <input type="checkbox"/> International Helsinki Federation for Human Rights (90) <input type="checkbox"/> Federation of International Human Rights Museums (83) <input type="checkbox"/> International Federation of ACAT (68) <input type="checkbox"/> Create new item <input type="checkbox"/> or match <input type="checkbox"/> Every International <input type="checkbox"/> new match <input type="checkbox"/> n		
							<input type="checkbox"/> International Federation for Human Rights <input checked="" type="checkbox"/> International Federation for Human Rights (100) <input checked="" type="checkbox"/> International Helsinki Federation for Human Rights (90) <input checked="" type="checkbox"/> Federation of International Human Rights Museums (83) <input checked="" type="checkbox"/> International Federation of ACAT (68) <input checked="" type="checkbox"/> Create new item Search for match		
☆	2.	Wickrey					Český klub skeptiků Sisyfos	16	
☆	3.	Ross					Post Bellum	172	
☆	4.	Chandler	Rosie	Rosie Chandler	Dr.	1990-05-25T00:00:00Z	International Federation for Human Rights	3	
☆	5.	Norman	Lilian	Lilian Norman		1992-09-14T00:00:00Z	Choose new match	4	
☆	6.	Gilbert	Soizic	Soizic Gilbert	Dr.	2001-02-21T00:00:00Z	Choose new match	11	

18. Přidejte sloupec OrganisationID, kde bude ID z Wikipedie. Zvolte znovu Reconcile u Organisation, potom Add entity identifiers column.
19. Přidejte zemi, ve které organizace sídlí. Zvolte Organisation/Edit Column/Add columns from reconciled values. Zvolte country.
20. Vyberte sloupec OrganisationID a zvolte Reconcile/Use values as identifiers.

Use values as identifiers

Reconciliation identifiers will not be validated.

Choose a reconciliation service

Wikidata reconcil.link (en)

Cancel

OK

21. Jděte zpět na RDF Transform/Edit transform. Přidejte jmenný prostor wd (WikiData <http://www.wikidata.org/entity/>)
22. Změňte objekt affiliation na OrganisationID místo Organisation. Zvolte typ IRI a správný jmenný prostor.

RDF Node

Prefix: ▼

Content...

- [Index]
- Family Name
- Given Name
- Name
- Title
- Birthdate
- Organisation
- country
- OrganisationID
- H-Index
- Constant value

Content used...

- IRI
- Text
- Language:
- Integer
- Double
- Date (yyyy-MM-dd)
- DateTime (yyyy-MM-ddTHH:mm:ss.SSS)
- Boolean
- Custom (specify type IRI)
-
- Blank

Expression...

value

[Edit & Preview](#)

OK

Cancel

23. Přidejte třídu foaf:Person k lidem. Přidejte třídu schema:Organization k organizacím.

The screenshot shows the 'RDF Transform' dialog box. The 'Available Namespaces' section lists various namespaces, including 'foaf:Person' and 'schema:Organization', which are highlighted with red arrows. The 'Transform' tab is active, and the 'Preview' tab is also visible. The dialog includes buttons for 'Add Root Node', 'Import Template', 'Export Template', and 'Save'.

24. Exportujte tabulku ve zvoleném formátu RDF pomocí Export/RDF Transform (vpravo nahoře). Můžete si ontologii prohlédnout v Protégé nebo vygenerovat graf ve validátoru

<https://medium.com/wallscope/creating-linked-data-31c7dd479a9e>

