



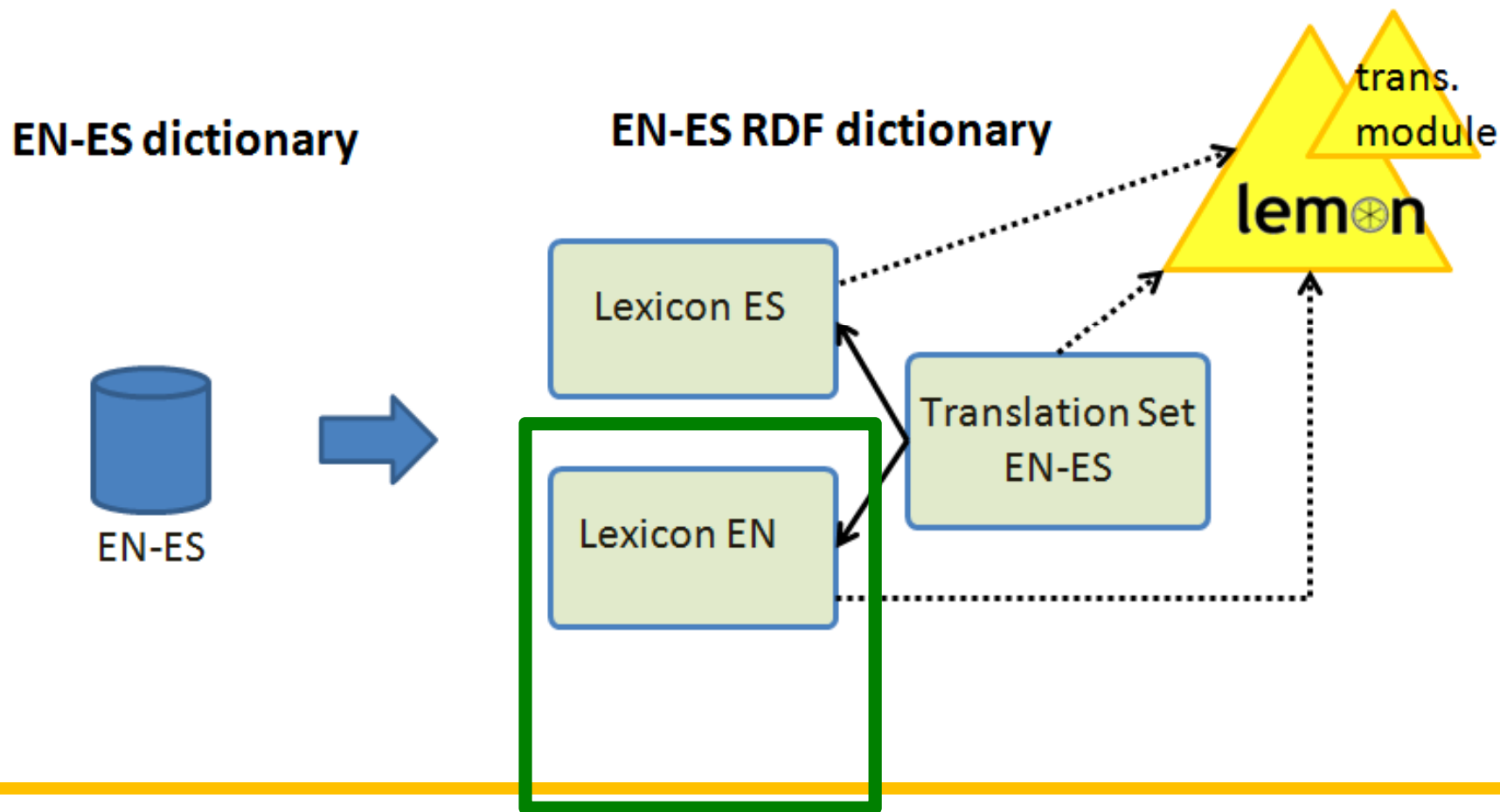
Hands-on session
Multilingual Linked Data generation and
publication

Jorge Gracia, Daniel Vila-Suero

Jgracia, dvila@fi.upm.es

ISWC Tutorial "Multilingual Semantic Web" 2014,
Trentino (Italy)

20th October 2014



- Define a resource naming scheme.
- Map TSV data source to RDF re-using a set of vocabularies.
- Find and establish links to a external data source.
- Generate a (small) RDF dataset.
- Upload the RDF dataset into a triple-store
- Query the RDF dataset using SPARQL

- Install Google Refine for your platform (Win, MacOS, Linux)
- Add the RDF extension
- Create project and check RDF extension

- **DO NOT RUN IT YET**
- **MacOS:** *Drag and drop to Applications folder*
- **Windows:**
 - Uncompress it into a folder wherever you want (such as in C:\Google-Refine).
- **Linux:**
 - Execute:

```
tar xzf google-refine.tar.gz  
cd google-refine
```

- Place the rdf-extension folder in:
- **Windows and Linux:**
 - webapp/extensions folder
- **MacOS:**
 - Resources/webapp/extensions folder
- Run Google Refine

- Create Project
- Open Project
- Import Project

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 are all supported. Support for other formats can be added with Google Refine extensions.

Get data from

Locate one or more files on your computer to upload:

This Computer

Elegir archivos Ningún archi...seleccionado

Web Addresses (URLs)

Next »

Clipboard

Google Data

Choose file: data/raw-data-apertium-en-es_lexiconEN



Version 2.5 [r2407]

Help
About



« Start Over
Configure Parsing Options
Project name raw data apertium en es_lexiconEN
Create Project »

	LexicalEntry-id	LexicalEntry-POS	LexicalEntry-writtenForm	sense
1.	abject-adj-en	adj	abject	abject_abyecto-adj-l
2.	abrupt-adj-en	adj	abrupt	abrupt_repentino-adj-l
3.	beabreast-vblex-en	vblex	be abreast	beabreast_estaralcorriente-vblex-l
4.	startafresh-vblex-en	vblex	start afresh	startafresh_empezardenuuevo-vblex-l
5.	abideby-vblex-en	vblex	abide by	abideby_atenera-vblex-l
6.	seaurchin-n-en	n	sea urchin	seaurchin_erizodemar-n-l
7.	abode-n-en	n	abode	abode_morada-n-l
8.	abruptness-n-en	n	abruptness	abruptness_brusquedad-n-l
9.	tradeunion-n-en	n	trade union	tradeunion_sindicato-n-l
10.	tradeunion-n-en	n	trade union	tradeunion_asociaciónobrera-n-l
11.	tradeunionist-n-en	n	trade unionist	tradeunionist_sindicalista-n-l
12.	abate-vblex-en	vblex	abate	abate_calmar-vblex-l
13.	curlyendive-n-en	n	curly endive	curlyendive_escarola-n-l
14.	beabreast-vblex-en	vblex	be abreast	beabreast_estaralcorriente-vblex-l

Parse data as

CSV / TSV / separator-based files

- Line-based text files
- Fixed-width field text files
- PC-Axis text files
- JSON files
- RDF/N3 files
- XML files
- Open Document Format spreadsheets (.ods)
- RDF/XML files

Character encoding

Columns are separated by

commas (CSV)

tabs (TSV)

custom \t

Escape special characters with \

Character encoding: UTF8

[Update Preview](#)

- Ignore 0 line(s) at first beginning of file
- Parse 1 line(s) as next column headers
- Discard 0 row(s) of data initial
- Load at 0 row(s) of data most
- Parse cell text into numbers, dates, ...
- Quotation marks are
- Store blank rows
- Store blank cells as nulls
- Store file

Google refine raw data apertium en es_lexiconEN tsv [Permalink](#)

Open... Export Help

Extensions:

Facet / Filter Undo / Redo 0

25794 rows

Show as: rows records Show: 5 10 25 50 rows

« first < previous 1 - 10 > next >>

Using facets and filters

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started?
[Watch these screencasts](#)

	All	LexicalEntry-id	LexicalEntry-PO	LexicalEntry-wri	sense
☆	1.	abject-adj-en	adj	abject	abject_abyecto-adj-l
☆	2.	abrupt-adj-en	adj	abrupt	abrupt_repentino-adj-l
☆	3.	beabreast-vblex-en	vblex	be abreast	beabreast_estaralcorriente-vblex-l
☆	4.	startafresh-vblex-en	vblex	start afresh	startafresh_empezardenuuevo-vblex-l
☆	5.	abideby-vblex-en	vblex	abide by	abideby_atenera-vblex-l
☆	6.	seaurchin-n-en	n	sea urchin	seaurchin_erizodemar-n-l
☆	7.	abode-n-en	n	abode	abode_morada-n-l
☆	8.	abruptness-n-en	n	abruptness	abruptness_brusquedad-n-l
☆	9.	tradeunion-n-en	n	trade union	tradeunion_sindicato-n-l
☆	10.	tradeunion-n-en	n	trade union	tradeunion_asociaciónobrera-n-l

1. Select *Edit RDF Skeleton*
2. Set your base URI
3. Add prefix of lemon:
 - `http://lemon-model.net/lemon#`
4. Start creating the mapping (*RDF Schema alignment*)

RDF Schema Alignment

The RDF schema alignment skeleton below specifies how the RDF data that will get generated from your grid-shaped data. The cells in each record of your data will get placed into nodes within the skeleton. Configure the skeleton by specifying which column to substitute into which node.

Base URI: <http://rdf.apertium.org/id/> [edit](#)

RDF Skeleton

RDF Preview

Check results

Available Prefixes: rdfs foaf owl xsd rdf lemon [+ add prefix](#) [manage prefixes](#)

(row index) URI
[add rdf:type](#)

Create instances and add a type

- X [property?→](#) LexicalEntry-id cell
- X [property?→](#) LexicalEntry-POS cell
- X [property?→](#) LexicalEntry-writtenForm cell
- X [property?→](#) sense cell

[add property](#)

Create Properties (literal and objects)

Define object



keleton below specifies how the RDF data that will get generated from your grid-shaped data. The cells in each skeleton. Configure the skeleton by specifying which column to substitute into which node.

RDF Node

Use content from cell...

- (row index)
- LexicalEntry-id
- LexicalEntry-POS
- LexicalEntry-writtenForm
- sense
- Constant Value

Use
column
value or
constant

The cell's content is used ...

- as a URI Create instances and object properties range
- as text
- as language-tagged text
- as integer number
- as non-integer number
- as date (YYYY-MM-DD)
- as dateTime (YYYY-MM-DD HH:MM:SS)
- as boolean
- as custom datatype (specify type URI)
- as a blank node

Literals and language-tagged literals

Use custom expression...

value
preview/edit

Apply transformations
using GREL (GR
expression language)



1. Create a reconciliation service from RDF file
(data/linking-dataset-rdf-wordnet-extract.nt)
2. Duplicate *LexicalEntry-writtenForm* column
3. Reconcile using created service (e.g. *wordnet*)

res_lexiconEN tsv [Permalink](#)

Open... Export Help

25794 rows Extensions: **Freebase** **RDF**

Show as: **rows** records Show: 5 10 25 50 rows

	All	LexicalEntry-id	LexicalEntry-PO	wordnet-pos	LexicalEntry-wri	sense
☆	1.	abject-adj-en	adj	http://wordnet-rdf.princeton.edu	Based on SPARQL endpoint...	
☆	2.	abrupt-adj-en	adj	http://wordnet-rdf.princeton.edu	Based on RDF file...	
☆	3.	beabreast-vblex-en	vblex	http://wordnet-rdf.princeton.edu	Based on a Sindice site search...	
☆	4.	startafresh-vblex-en	vblex	http://wordnet-rdf.princeton.edu	Based on a Apache Stanbol EntityHub...	
☆	5.	abideby-vblex-en	vblex	http://wordnet-rdf.princeton.edu/ontology#vblex		abide by
☆	6.	seurchin-n-en	n	http://wordnet-rdf.princeton.edu/ontology#noun	sea urchin	seurchin_erizodemar-n-l
☆	7.	abode-n-en	n	http://wordnet-rdf.princeton.edu/ontology#noun	abode	abode_morada-n-l
☆	8.	abruptness-n-en	n	http://wordnet-rdf.princeton.edu/ontology#noun	abruptness	abruptness_brusquedad-n-l
☆	9.	tradeunion-n-en	n	http://wordnet-rdf.princeton.edu/ontology#noun	trade union	tradeunion_sindicato-n-l
☆	10.	tradeunion-n-en	n	http://wordnet-rdf.princeton.edu/ontology#noun	trade union	tradeunion_asociaciónobrera-n-l



Add file-based reconciliation service

This will set up a new reconciliation service based on an RDF file that provides entity URIs and entity labels.

Name:
A human readable name

File details

Load file from URL:

Upload file: Ningún archivo seleccionado

File format:

Label properties

Select properties that are used to label resources in the RDF data. These properties will be used to match resources:

rdfs:label skos:prefLabel dcterms:title dc:title

foaf:name

Other...

es_lexiconEN tsv [Permalink](#)

Open... Export ▾ Help

25794 rows

Extensions: Freebase ▾ RDF ▾

Show as: **rows** records Show: 5 10 25 50 rows

« first < previous **1 - 10** next > last »

▼ All	▼ LexicalEntry-id	▼ LexicalEntry-PO	▼ wordnet-pos	▼ LexicalEntry-wri	▼ linksToWN	▼ sense
☆	1.	abject-adj-en	adj	http://wordnet-rdf.princeton.edu/ontology#adj	abject	Facet ▶ abyecto-adj-l
☆	2.	abrupt-adj-en	adj	http://wordnet-rdf.princeton.edu/ontology#adj	abrupt	Text filter ▶ _repentino-adj-l
☆	3.	beabreast-vblex-en	vblex	http://wordnet-rdf.princeton.edu/ontology#vblex	be abreast	Edit cells ▶ ast_estaralcorriente-vblex-l
☆	4.	startafresh-vblex-en	vblex	http://wordnet-rdf.princeton.edu/ontology#vblex	start afresh	Edit column ▶ esh_empezardenuuevo-vblex-l
☆	5.	abideby-vblex-en	vblex	http://wordnet-rdf.princeton.edu/ontology#vblex	abide by	Transpose ▶ y_atenera-vblex-l
☆	6.	seaurchin-n-en	n	http://wordnet-rdf.princeton.edu/ontology#noun	sea urchin	Sort... ▶ hin_erizodemar-n-l
☆	7.	abode-n-en	n	http://wordnet-rdf.princeton.edu/ontology#noun	abode	View ▶ _morada-n-l
☆	8.	abruptness-n-en	n	http://wordnet-rdf.princeton.edu/ontology#noun	abruptness	Reconcile ▶ hess_brusquedad-n-l
☆	9.	tradeunion-n-en	n	http://wordnet-rdf.princeton.edu/ontology#noun	trade union	Start reconciling...
☆	10.	tradeunion-n-en	n	http://wordnet-rdf.princeton.edu/ontology#noun	trade union	Facets ▶

Google refine raw data spectrum on es-lexiconEN tex-...

Facet / Filter Undo / Redo

Using facets and filters

Use facets and filters to select a subset of your data to act on. Choose from the filter methods from the menu of each data column.

Not sure how to get started? Watch these screencasts

Reconcile column "linksToWN"

» Access Service API

Reconcile each cell to an entity of one of these types:

Sorry, we can't suggest any type for your data. Please specify a type yourself below.

Also use relevant details from other columns:

Column	Include?	As Property
LexicalEntry-id	<input type="checkbox"/>	<input type="text"/>
LexicalEntry-POS	<input type="checkbox"/>	<input type="text"/>
wordnet-pos	<input type="checkbox"/>	<input type="text"/>
LexicalEntry-writtenForm	<input type="checkbox"/>	<input type="text"/>
sense	<input type="checkbox"/>	<input type="text"/>

Reconcile against type:

Reconcile against type:

Reconcile against no particular type

Auto-match candidates with high confidence

Add Standard Service... Add Namespaced Service... Start Reconciling Cancel

nsions: Freebase RDF

« previous 1 - 10 next » last »

se

abyecto-adj-l

repentino-adj-l

ast_estaralcorriente-vblex-l

ash_empezardenuuevo-vblex-l

_atenera-vblex-l

in_erizodemar-n-l

morada-n-l

ess_brusquedad-n-l

ion_sindicato-n-l

ion_asociaciónobrera-n-l

Reconcile against type:
<http://wordnet-rdf.princeton.edu/ontology#Synset>

Google refine raw data apertium en es_lexiconEN tsv [Permalink](#) Open... Export Help

Facet / Filter Undo / Redo 31 **156 matching rows (25794 total)** Extensions: Freebase RDF

Refresh Reset All Remove All Show as: rows records Show: 5 10 25 50 rows « first < previous 1 - 50 next > last »

All	LexicalEntry-id	LexicalEntry-PO	wordnet-pos	LexicalEntry-wri	links2	linksToWN	sense
☆	604. group-vblex-en	vblex	http://wordnet-rdf.princeton.edu/ontology#vblex	group	group Choose new match	group Choose new match	group_agrupar-vblex-
☆	849. whole-adj-en	adj	http://wordnet-rdf.princeton.edu/ontology#adj	whole	whole <input checked="" type="checkbox"/> whole (1) <input checked="" type="checkbox"/> Create new topic Search for match	whole Choose new match	whole_integro-adj-l
☆	850. whole-adj-en	adj	http://wordnet-rdf.princeton.edu/ontology#adj	whole	whole <input checked="" type="checkbox"/> whole (1) <input checked="" type="checkbox"/> Create new topic Search for match	whole Choose new match	whole_entero-adj-l
☆	1261. individual-n-en	n	http://wordnet-rdf.princeton.edu/ontology#noun	individual	individual Choose new match	individual new Choose new match	individual_particular-r
☆	1262. individual-n-en	n	http://wordnet-rdf.princeton.edu/ontology#noun	individual	individual <input checked="" type="checkbox"/> individual (1) <input checked="" type="checkbox"/> Create new topic Search for match	individual Choose new match	individual_individuo-n
☆	1514. state-vblex-en	vblex	http://wordnet-rdf.princeton.edu/ontology#vblex	state	state <input checked="" type="checkbox"/> state (1) <input checked="" type="checkbox"/> Create new topic Search for match	state Choose new match	state_constar-vblex-l
☆	1515. state-vblex-en	vblex	http://wordnet-rdf.princeton.edu/ontology#vblex	state	state <input checked="" type="checkbox"/> state (1) <input checked="" type="checkbox"/> Create new topic Search for match	state Choose new match	state_declararD-vblex
☆	1516. state-vblex-en	vblex	http://wordnet-rdf.princeton.edu/ontology#vblex	state	state <input checked="" type="checkbox"/> state (1) <input checked="" type="checkbox"/> Create new topic Search for match	state Choose new match	state_plantear-vblex-l
☆	1886. group-n-en	n	http://wordnet-rdf.princeton.edu/ontology#noun	group	group Choose new match	group Choose new match	group_cuadrilla-n-l
☆	1887. group-n-en	n	http://wordnet-rdf.princeton.edu/ontology#noun	group	group <input checked="" type="checkbox"/> group (1) <input checked="" type="checkbox"/> Create new topic Search for match	group Choose new match	group_conjunto-n-l
☆	1888. group-n-en	n	http://wordnet-rdf.princeton.edu/ontology#noun	group	group <input checked="" type="checkbox"/> group (1)	group Choose new match	group_grupoD-n-l

links2: best candidate's score change reset

1.00 — 1.01

Numeric 498 Non-numeric 0 Blank 0 Error 25296

Use Reconciliation facets and filter:
E.g. Best candidate's score

Validate!

1. The URI to external source is available using

GREL:

cell.recon.match.id

2. Use RDF Skeleton and match using your preferred linking property:

e.g. **owl:sameAs**, **skos:closeMatch**,
lemon:reference

1. Export to file:

RDF as Turtle menu option

1. Launch Fuseki triplestore
2. Load data
3. Query

- Execute:

fuseki-server --update --mem /handson

Creates a fuseki server storing data in memory with rights to update data

- Server accessible under:

http://localhost:3030/

SPARQL Query

Output:

If XML output, add XSLT style sheet (blank for none):

Force the accept header to `text/plain` regardless.

SPARQL Update

File upload

File: Ningún archivo seleccionado

Graph:

SPARQL Query

Output:

If XML output, add XSLT style sheet (blank for none):

Force the accept header to text/plain regardless.

SPARQL Update

File upload

File: Ningún archivo seleccionado

Graph: