

5-24-2018

A Step Forward: Adding Linked Data Vocabularies to Digital Repositories

Sai Deng

University of Central Florida, sai.deng@ucf.edu



Part of the [Cataloging and Metadata Commons](#)

Find similar works at: <https://stars.library.ucf.edu/ucfscholar>

University of Central Florida Libraries <http://library.ucf.edu>

This Conference Presentation is brought to you for free and open access by STARS. It has been accepted for inclusion in Faculty Scholarship and Creative Works by an authorized administrator of STARS. For more information, please contact STARS@ucf.edu.

Original Citation

Deng, S. (2018). A step forward: adding linked data vocabularies to digital repositories. Florida Library Association 2018 Annual Conference. May 24, 2018, Orlando, FL.



A Step Forward: Adding Linked Data Vocabularies to Digital Repositories

Sai Deng, University of Central Florida Libraries

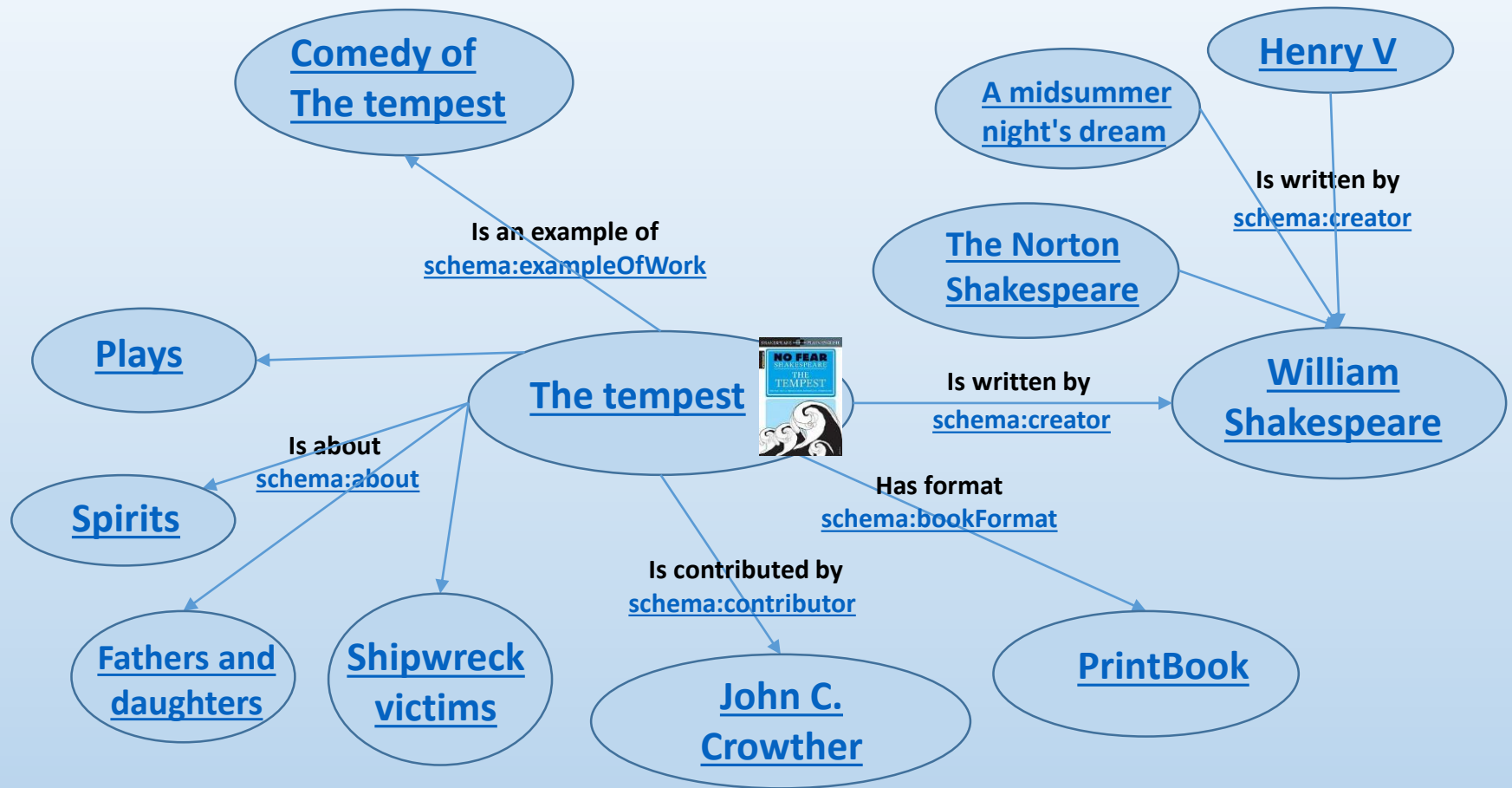
Linked Data Basics

- **Linked data:**
 - “A term used to describe a recommended best practice for exposing, sharing, and connecting pieces of data, information, and knowledge on the [Semantic Web](#) using [URIs](#) and [RDF](#).”
-Wikipedia
- **Resource Description Framework (RDF): Represent information in triples**
 - **Subject:** a resource, identified with a URI
 - **Predicate:** specification of the relationship
 - **Object:** a resource the subject is related to

Linked Data Basics

- **The Four Design Principles of Linked Data**
 - Use Uniform Resource Identifiers (**URIs**) as names for **things**.
 - Use **HTTP URIs** so that people can look up those names.
 - When someone looks up a URI, **provide** useful **information**, using the standards (**RDF, SPARQL**).
 - Include **links to other URIs** so that they can discover more things.
 - Tim Berners-Lee. 2006.
<https://www.w3.org/DesignIssues/LinkedData.html>
- URIs for naming things, RDF for describing things, SPARQL for querying data; link to other URIs to find more things.

Linked Data: An Example



*Based on information from worldcat.org

Linked Open Data Building Blocks

- **The LOD Data Model:** RDF triples
- **Content Rules:** relies on external vocabularies and ontologies
- **Metadata schema**
 - RDF Schema
 - Vocabulary Building Blocks: SKOS, OWL
 - Open Archives Initiative Object Reuse and Exchange (OAI-ORE)
- **Serializations of LOD**
 - RDF Notation-3/Ns, Turtle, N-Triples; RDFa; JSON-LD
- **Exchanging LOD**
 - SPARQL
- Mitchell, Erik T. Building Blocks of Linked Open Data in Libraries.
Library Technology Reports (vol. 49, no. 5)

Five Star-schema of Linked Open Data (LOD)

- **Five star-schema of Linked Open Data** (Tim Berners-Lee)

★ Available **on the web** (whatever format) but with an open licence, to be Open Data

★★ Available as **machine-readable structured data** (e.g. excel instead of image scan of a table)

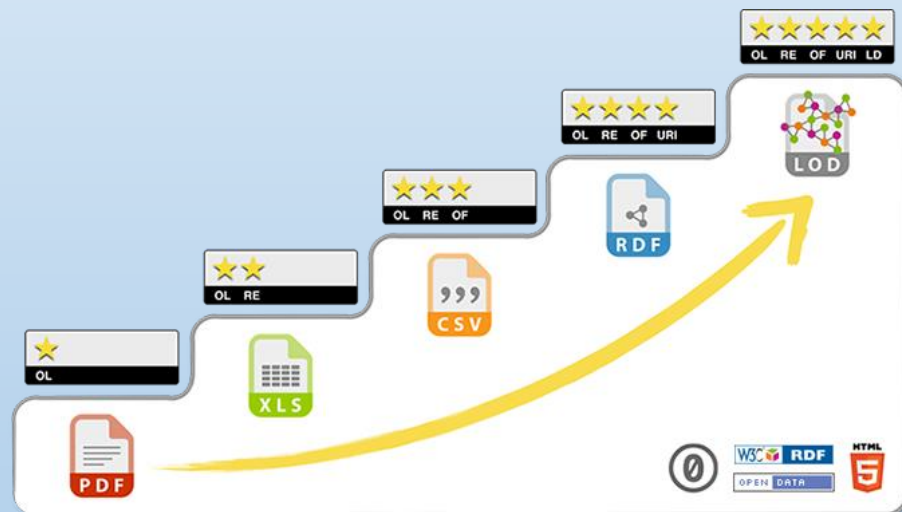
★★★ as (2) plus **non-proprietary** format (e.g. CSV instead of excel)

★★★★ All the above plus, Use open standards from W3C (**RDF** and SPARQL) to **identify things**, so that people can point at your stuff

★★★★★ All the above, plus: **Link** your data to **other people's data** to provide context



Source: <https://www.w3.org/DesignIssues/LinkedData.html>



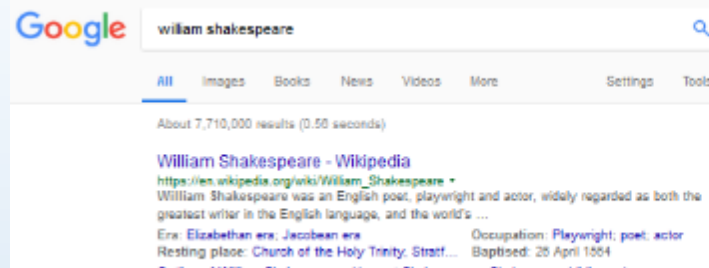
Source: <http://5stardata.info/en/>

Linked Data Examples and Datasets

- **LC Linked Data Service: Authorities and Vocabularies** <https://id.loc.gov/>
- **FAST: Faceted Application of Subject Terminology** <http://id.worldcat.org/fast>
- **Virtual International Authority File (VIAF)** <https://viaf.org/>
- **Worldcat Identities** www.worldcat.org/identities
- **Open Researcher and Contributor ID (ORCID)** <https://orcid.org/>
- **International Standard Name Identifier (ISNI)** <http://www.isni.org/>
- **GeoNames** <http://www.geonames.org/ontology/documentation.html>
- **DBpedia** <http://wiki.dbpedia.org/>
- **Schema.org Vocabulary** <http://schema.org/docs/gs.html#schemaorg>
- **The Linked Open Data Cloud** <https://lod-cloud.net/>

Linked Data Implementations

- Google Knowledge Graph



Quotes

[View 7+ more](#)

All that glitters is not gold

Hell is empty and all the devils are here.

*To thine own self be true, and it must follow, as the night the day,
thou canst not then be false to any man.*

Plays

[View 35+ more](#)



Macbeth
1623



Hamlet
1616



Romeo and Juliet
1597



Othello
1612



A Midsummer Night's Dream
1619

name Bard of

interesting articles,

People also search for

[View 30+ more](#)



Miguel de Cervantes



Charles Dickens



Christopher Marlowe



Edgar Allan Poe



William Wordsworth



William Shakespeare

English poet

William Shakespeare was an English poet, playwright and actor, widely regarded as both the greatest writer in the English language, and the world's pre-eminent dramatist. He is often called England's national poet, and the "Bard of Avon". [Wikipedia](#)

Born: April 1564, Stratford-upon-Avon, United Kingdom

Died: April 23, 1616, Stratford-upon-Avon, United Kingdom

Spouse: Anna Hathaway (m. 1582-1616)

Did you know: William Shakespeare is the best-selling fiction author of all time (estimated 4 billion copies sold). [wikipedia.org](#)

Quotes

[View 7+ more](#)

All that glitters is not gold

Hell is empty and all the devils are here.

*To thine own self be true, and it must follow, as the night the day,
thou canst not then be false to any man.*

Plays

[View 35+ more](#)



Macbeth
1623



Hamlet
1616



Romeo and Juliet
1597



Othello
1612



A Midsummer Night's Dream
1619

People also search for

[View 30+ more](#)



Miguel de Cervantes



Charles Dickens



Christopher Marlowe



Edgar Allan Poe



William Wordsworth

See results about

Complete Works of Shakespeare (Book by William ...

Author: William Shakespeare

Genre: Drama



Linked Data Implementations

- Libraries, Archives and Museums Community

- Europeana <http://europeana.eu>
- DPLA <http://dp.la>
- LIBRIS <https://libris.kb.se/>
- British National Bibliography <http://bnb.bl.uk/>
- BIBFRAME <http://BIBFRAME.org/>



- The Web and the Library Community

- Schema.org
 - Bibliographic Extension
<http://bib.schema.org/>, e.g., [Audiobook](#), [Thesis](#)
 - Extend Schema.org to represent archives
<https://archival.github.io/schema-org/>



- More Library Experiments/Cases

- OCLC Linked Data <https://www.oclc.org/developer/develop/linked-data.en.html>
- WorldCat Works (schema:exampleOfWork) <http://experiment.worldcat.org/entity/work/data/1080130963>
- Transforming Library Metadata into Linked Library Data: Introduction and Review of Linked Data for the Library Community, 2003–2011 <http://www.ala.org/alcts/resources/org/cat/research/linked-data>
- Check out ALA conference programs on LOD in the past few years, e.g., Linked Library Data IG

Current Digital Repositories and Linked Data

- **Islandora**

- Effort in mapping MODS to RDF;
- Fedora 4, LDP implementation; Islandora CLAW

- **Experiments and Cases:**

- [Islandora CLAW](#)
- [TU Delft repository example: Colonial Architecture](#) (Delftdora module, Fedora 3)
- [A Lightweight Structured Data Implementation Using JSON-LD and Schema.org for Digital Repository](#) (Mak et al.)

Current Digital Repositories and Linked Data

- **CONTENTdm**
- **Not inside the system, but work on exported data;**
- **UIUC.** Challenges of Mapping Digital Collections Metadata to Schema.org (Lampron et. al. 2016)
- **OCLC& CONTENTdm** Metadata Refinery, transforms CONTENTdm metadata to Linked Data, 2016. Not yet scalable. (Washburn et al. 2016)

Linked Open Data Website

Home Browse All

Home > Motley Collection of Theatre and Costume Design (New) > Hamlet > Caesar

External Resources

- Julius Caesar (play)**

The Tragedy of Julius Caesar is a tragedy by William Shakespeare, believed to have been written in 1599. It is one of several plays written by Shakespeare based on true events from Roman history, which also include Coriolanus and Antony and Cleopatra. [more from Wikipedia]

DBPedia
Theatricalia
Theatricalia
- William Shakespeare (Author)**

Gender: male

William Shakespeare (/ˈʃɛkspiəri/; 26 April 1564 (baptised) – 23 April 1616) was an English poet, playwright, and actor, widely regarded as the greatest writer in the English language and the world's pre-eminent dramatist. He is often called England's [more from Wikipedia]

DBPedia
VIAF Record
SNF Record
DNB Record
- Glen Byam Shaw (Director)**

Gender: male

Glencairn Alexander "Glen" Byam Shaw, CBE (13 December 1904 – 29 April 1986) was an English actor and theatre director, known for his dramatic productions in the 1950s and his operatic productions in the 1960s and later. [more from Wikipedia]

DBPedia
VIAF Record
DNB Record
Theatricalia
- Luckham, Cyril (Actor)**

Cyril Alexander Garland Luckham (25 July 1907 – 8 February 1989) was a British film, television and theatre actor. [more from Wikipedia]

DBPedia

Caesar



External Resources

- Julius Caesar (play)**
- Glen Byam Shaw (Director)**

Gender: male

Glencairn Alexander "Glen" Byam Shaw, CBE (13 December 1904 – 29 April 1986) was an English actor and theatre director, known for his dramatic productions in the 1950s and his operatic productions in the 1960s and later. [more from Wikipedia]

DBPedia
VIAF Record
DNB Record
Theatricalia

A UIUC Experiment <http://imagesearch-test1.library.illinois.edu/cdm/ref/collection/motley-new/id/599>
 Project Site <http://publish.illinois.edu/linkedspcollections/>

Current Digital Repositories and Linked Data

- **Omeka**

- Open source web publishing platform. Supports Dublin Core. Can export METS, dc-rdf, JSON etc.
- [Ozmeka \(plug-ins\)/ Omeka linked data](#)

- **DSpace**

- **DSpace Linked Data support** spans all three Layers: “the storage layer with a **triple store**, the business logic with classes to **convert** stored contents into RDF, and the application layer with a module to **publish** RDF serializations.”
 - Donohue. Linked (Open) Data.
(<https://wiki.duraspace.org/display/DSDOC6x/Linked+%28Open%29+Data>)

Current Digital Repositories and Linked Data

- **Digital Commons**

- Allow to add extra fields for links (not RDF)

- **Preparing for Linked Data**

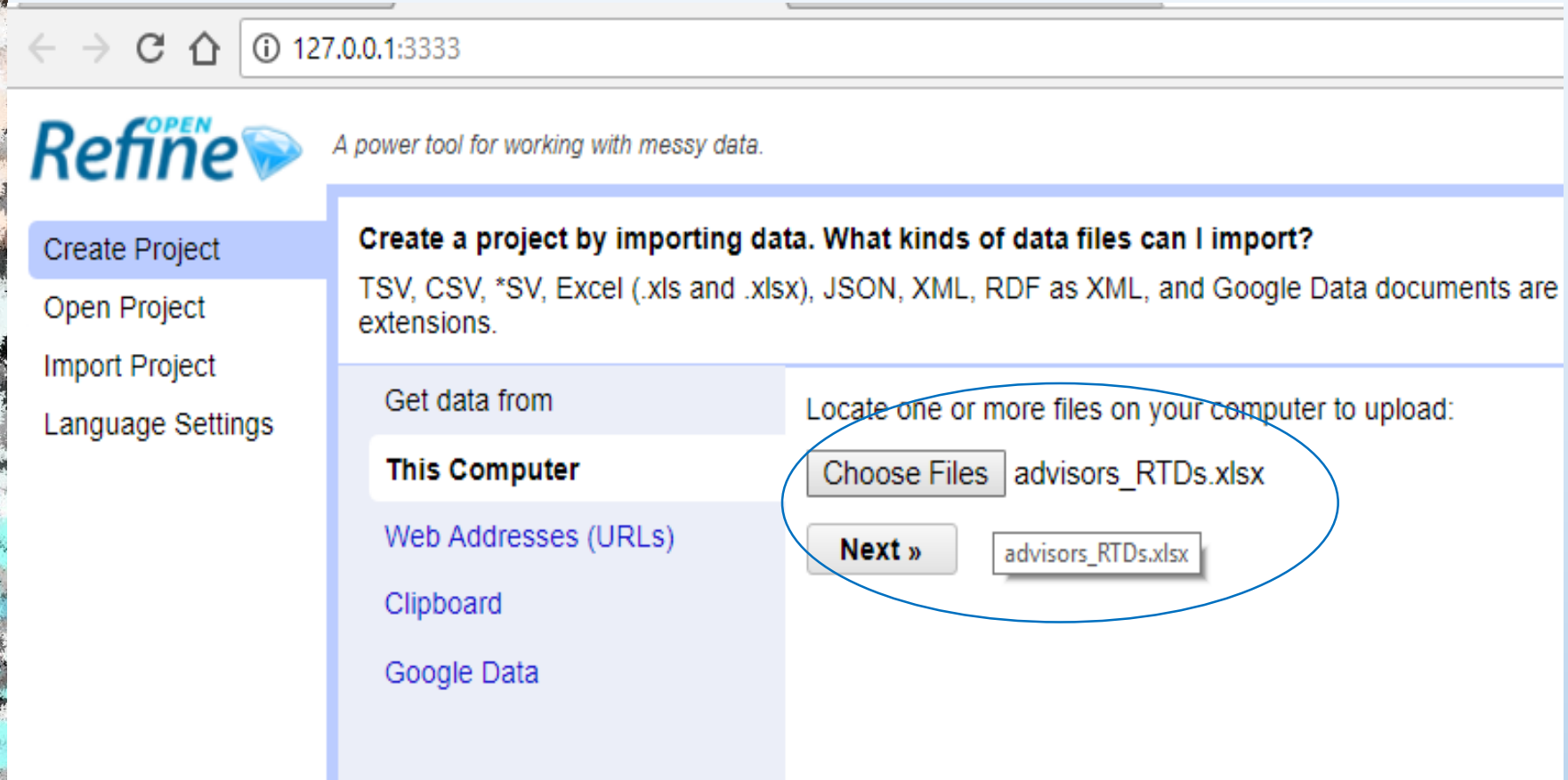
- Include links (e.g., id.loc.gov, viaf.org) in added additional fields for author, advisor, department names etc.;
- Data (in spreadsheet) cleanup and reconciliation using OpenRefine
 - Name review can be tricky and time-consuming: Google, Wikipedia, obituary, university and departmental websites, publications, CVs...

Adding Linked Data Vocabularies to UCF Digital Repositories

- **Adding LOD Links to Fields in Digital Repositories**
 - Authors/Creators
 - Contributors, e.g., thesis advisors
 - Corporate names, e.g., university, college and departmental names
 - Subjects
 - Person identifiers
- **OpenRefine**
 - **Download and installation**
<http://openrefine.org/download.html>
 - **Documentation**
<https://github.com/OpenRefine/OpenRefine/wiki/Documentation-For-Users>
 - **Features**
 - Import
 - Facet
 - Edit cells: Clustering...
 - Edit columns: Add Column based on this column; Fetching URLs from Web Services (*Enrich existing data columns)...
 - Expressions
 - Export
 - History...



Adding LC and VIAF Links to Advisor Names



The screenshot shows the OpenRefine web interface in a browser window. The address bar displays '127.0.0.1:3333'. The OpenRefine logo and tagline 'A power tool for working with messy data.' are at the top. On the left, a sidebar lists navigation options: 'Create Project', 'Open Project', 'Import Project', and 'Language Settings'. The main content area is titled 'Create a project by importing data. What kinds of data files can I import?' and lists supported formats: TSV, CSV, *SV, Excel (.xls and .xlsx), JSON, XML, RDF as XML, and Google Data documents. Below this, under 'Get data from', the 'This Computer' option is selected. To the right, the instruction 'Locate one or more files on your computer to upload:' is followed by a 'Choose Files' button and the filename 'advisors_RTDs.xlsx'. A 'Next »' button is also present. A blue circle highlights the 'Choose Files' button, the filename 'advisors_RTDs.xlsx', and the 'Next »' button.

Open OpenRefine — Choose Files — Select advisors-RTDs.xlsx — Click “Next”

Adding LC and VIAF Links to Advisor Names

Refine OPEN A power tool for working with messy data.

Create Project « Start Over Configure Parsing Options Project name: advisors_RTDs.xlsx **Create Project »**

	identifier	advisor
1.	DP0000160	Shah, Mubarak A.
2.	DP0000175	Leckie, Shirley
3.	DP0000186	Armstrong, Jack
4.	DP0000186	Malocha, Donald;
5.	DP0000189	Parsons, Rebecca
6.	DP0000191	Shah, Mubarak;
7.	DP0000191	Wooten, William
8.	DP0000192	Klee, Harold
9.	DP0000193	Dutton, Ronald;
10.	DP0000194	Batarseh, Issa
11.	DP0000195	Leckie, Shirley
12.	DP0000759	Mikhael, Wasfy
13.	DP0000760	Gomez, Fernando
14.	DP0000781	Moharam, Jim
15.	DP0000782	Kasparis, Takis
16.	DP0000784	Christodoulou, Christos G.
17.	DP0000786	Orooji, Ali
18.	DP0000833	Batarseh, Issa
19.	DP0001374	Gelenbe, Erol
20.	DP0001376	Klee, Harold
21.	DP0001377	DeMara, Ronald;
22.	DP0001716	Batarseh, Issa
23.	DP0001717	Shah, Mubarak;
24.	DP0001718	Batarseh, Issa

Parse data as

Excel files Worksheets to Import: ☒ Sheet1 969 rows

JSON files

Line-based text files

CSV / TSV / separator-based files

Fixed-width field text files

PC-Axis text files

MARC files

RDF/N3 files

XML files

Open Document Format

☐ Ignore first 0 line(s) at beginning of file

☒ Parse next 1 line(s) as column headers

☐ Discard initial 0 row(s) of data

☐ Load at most 0 row(s) of data

☒ Store blank rows

☒ Store blank cells as nulls

☐ Store file source (file names, URLs) in each row

Update Preview

Version 2.7 [TRUNK]

Help
About

Click
“Create
Project”

Adding LC and VIAF Links to Advisor Names

Faceting

← → ↺ ↻ ⓘ 127.0.0.1:3333/project?project=2122028945939

Refine OPEN advisors_RTDs.xlsx [Permalink](#)

Facet / Filter Undo / Redo 0

Refresh Reset All Remove All

advisor change
284 choices Sort by: name count Cluster

Abbott, David 2
Abbott, David W. 5
Acierno, Louis J. 2
Adicks, Richard R. 1
Alsaka, Yacoub A. 1
Anderson, Loren A. 3
Armstrong, Jack 1
Armstrong, Lee H. 1
Arnold, Robert 1
Arnold, Robert J. 3
Baldwin, Jr., Vaniah 1
Batarseh, Issa 4

968 rows

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

☒ All ☒ identifier ☒ advisor

Facet
Text filter
Edit cells
Edit column
Transpose
Sort...
View
Reconcile
Extract named entities...

Text facet
Numeric facet
Timeline facet
Scatterplot facet
Custom text facet...
Custom Numeric Facet...
Customized facets

1. 2. 3. 4. 5. 6. 7. 8. 9. 10. 11. 12. 13. 14. 15. 16. 17.

DP0000759 Mikhael, Wasty
DP0000760 Gomez, Fernando
DP0000781 Moharam, Jim
DP0000782 Kasparis, Takis
DP0000784 Christodoulou, Christos G.
DP0000786 Orooji, Ali

After clicking “Facet – Text facet”

Adding LC and VIAF Links to Advisor Names

← → ↺ 🏠 ⓘ 127.0.0.1:3333/project?project=2122028945939

Clustering

Refine OPEN advisors_RTDs.xlsx [Permalink](#)

Facet / Filter Undo / Redo 0

Refresh Reset All Remove All

advisor 284 choices Sort by: name count **Cluster**

Abbott, David 2
Abbott, David W. 5
Acierno, Louis J. 2
Adicks, Richard R. 1
Alsaka, Yacoub A. 1
Anderson, Loren A. 3
Armstrong, Jack 1
Armstrong, Lee H. 1
Arnold, Robert 1
Arnold, Robert J. 3
Baldwin, Jr., Vaniah 1
Batarseh, Issa 4

968 rows

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

| All | identifier | advisor |
|-----|---------------|----------------------------|
| ☆ ↶ | 1. DP0000160 | Facet |
| ☆ ↶ | 2. DP0000175 | Text filter |
| ☆ ↶ | 3. DP0000186 | Edit cells |
| ☆ ↶ | 4. DP0000186 | Edit column |
| ☆ ↶ | 5. DP0000189 | Transpose |
| ☆ ↶ | 6. DP0000191 | Sort... |
| ☆ ↶ | 7. DP0000191 | View |
| ☆ ↶ | 8. DP0000192 | Reconcile |
| ☆ ↶ | 9. DP0000193 | Extract named entities... |
| ☆ ↶ | 10. DP0000194 | Mikhael, Wasty |
| ☆ ↶ | 11. DP0000195 | Gomez, Fernando |
| ☆ ↶ | 12. DP0000759 | Moharam, Jim |
| ☆ ↶ | 13. DP0000760 | Kasparis, Takis |
| ☆ ↶ | 14. DP0000781 | Christodoulou, Christos G. |
| ☆ ↶ | 15. DP0000782 | |
| ☆ ↶ | 16. DP0000784 | |

Facet
Text filter
Edit cells
Edit column
Transpose
Sort...
View
Reconcile
Extract named entities...
Transform...
Common transforms
Fill down
Blank down
Split multi-valued cells...
Join multi-valued cells...
Cluster and edit...

Click “Cluster” on the left, or click “Edit Cells – Cluster and edit” under “advisor” column

Adding LC and VIAF Links to Advisor Names

Clustering

Cluster & Edit column "advisor" (Method: key collision, Keying Function: fingerprint)

Refine advisors_RTDs.xlsx Permalink

Facet / Filter Undo / Redo 968 rows Extensions: Named-entity recognition RDF

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

Cluster & Edit column "advisor"

This feature helps you find groups of different cell values that might be alternative representations of the same thing. For example, the two strings "New York" and "new york" are very likely to refer to the same concept and just have capitalization differences, and "Gödel" and "Godel" probably refer to the same person. [Find out more ...](#)

Method: key collision Keying Function: fingerprint 20 clusters found

| Cluster | Count | Members | Preview |
|---------|-------|---|-----------------------|
| 2 | 2 | <ul style="list-style-type: none">Hagedoorn, A. Henry (1 rows)Hagedoorn, Henry A. (1 rows) | Hagedoorn, A. Henry |
| 2 | 22 | <ul style="list-style-type: none">Bauer, Christian (21 rows)Bauer, Christian; (1 rows) | Bauer, Christian |
| 2 | 7 | <ul style="list-style-type: none">Harden, Richard C. (5 rows)Harden, Richard C.; (2 rows) | Harden, Richard C. |
| 2 | 12 | <ul style="list-style-type: none">Hartman, J Paul (11 rows)Hartman, J. Paul (1 rows) | Hartman, J. Paul |
| 2 | 28 | <ul style="list-style-type: none">Wanielista, Martin P. (25 rows)Wanielista, Martin P.; (3 rows) | Wanielista, Martin P. |

Rows in Cluster

Average Length of Choices

Length Variance of Choices

Select All Unselect All Export Clusters Merge Selected & Re-Cluster Merge Selected & Close Close

30. DP0003046 Stebbins, Consuelo
31. DP0003459 Rapson, Richard
32. DP0003461 Washington, David W.
33. DP0003462 Yousef, Y. A.

e.g., Hagedoorn, A. Henry,
Hagedoorn, Henry A. →
Hagedoorn, A. Henry

Adding LC and VIAF Links to Advisor Names

Method nearest neighbor ▼

Distance Function

levenshtein ▼

Radius 1.0

Block Chars 6

Clustering

| Cluster Size | Row Count | Values in Cluster | Merge? | New Cell Value |
|--------------|-----------|---|-------------------------------------|----------------------|
| 2 | 8 | <ul style="list-style-type: none">Frank, Fredric D. (7 rows)Frank, Frederic D. (1 rows) | <input checked="" type="checkbox"/> | Frank, Frederic D. |
| 2 | 6 | <ul style="list-style-type: none">Schrader, George F. (5 rows)Schrader, George A. (1 rows) | <input checked="" type="checkbox"/> | Schrader, George F. |
| 2 | 5 | <ul style="list-style-type: none">Snelson, Franklin F. (4 rows)Snelson, Franklin (1 rows) | <input checked="" type="checkbox"/> | Snelson, Franklin F. |

Cluster & Edit column
“advisor”:
Try other algorithms

Adding LC and VIAF Links to Advisor Names

Reconciliation

← → ↺ ↻ 127.0.0.1:3333/project?project=2122028945939

Refine OPEN advisors_RTDs.xlsx Permalink

Facet / Filter Undo / Redo 5

Refresh Reset All Remove All

968 rows Extension

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

▼ All ▼ identifier ▼ advisor

235 choices Sort by: name count Cluster

Abbott, David W. 7
Acierno, Louis J. 2
Adicks, Richard R. 1
Alsaka, Yacoub A. 1
Anderson, Loren A. 3
Armstrong, Jack 1
Armstrong, Lee H. 1
Arnold, Robert J. 4
Baldwin, Jr., Vaniah 1
Batarseh, Issa 4
Bauer, Christian S. 32
Beck, James K. 5
Belkerdid, Madjid A. 8
Bennett, Gladys H. 1
Biegel, John E. 6
Bishop, P.J. 2
Bishop, Patricia J. 4
Blau, Burt 1
Blau, Burton I. 14

49. DP0003503
568. DP0013348
778. DP0015547
847. DP0016479
953. DP0019479
954. DP0019480
955. DP0019481
802. DP0015862
890. DP0017154
146. DP0011933
711. DP0013883
712. DP0013884
881. DP0017018
949. DP0019440
3. DP0000186
664. DP0013700
51. DP0003509
198. DP0012013

Facet
Text filter
Edit cells
Edit column
Transpose
Sort
View
Reconcile
Extract named entities...

Start reconciling...
Reconcile text in this column with topics on Freebase
QA facets
Actions
Copy reconciliation data...
Discover related RDF datasets...

Sort by
advisor;

Click
“Reconcile—
Start
reconciliation”

Adding LC and VIAF Links to Advisor Names

← → ↺ ↻ 127.0.0.1:3333/project?project=2122028945939

Refine advisors_RTDs.xlsx Permalink

Facet / Filter Undo / Redo 5 968 rows Extensions: Named-entity re

Refresh Reset All Remove All Show as: rows records Show: 5 10 25 50 rows Sort

« first « previous

Choose a Reconciliation Service: Need to set up the service first

Reconcile column "advisor"

Services

- NamedEntity
- Wikidata Reconciliation for OpenRefine (en)
- VIAF
- LC (by way of VIAF)
- ORCID
- Open Library
- LoC Reconciliation Service**
- Virtual International Authority File
- Sindice
- LCSH
- VIAF - LC

Pick a Service or Extension on Left

How to set up LoC Reconciliation Service
Click "Add Standard Service";

Enter the service's URL:

<https://lc-reconcile.herokuapp.com/>

Add Service **Cancel**

Add Standard Service... Start Reconciling Cancel

Adding LC and VIAF Links to Advisor Names

Reconciliation

Leave “Names” as default;
Click “Start Reconciling”

The screenshot shows the Refine interface with the 'advisor' column selected for reconciliation. The 'Reconcile column "advisor"' dialog is open, showing 'Names' as the selected entity type. The main table displays 968 rows of data, with the 'advisor' column highlighted. The 'advisor: judgment' facet shows 313 matches, which is circled in blue. The 'advisor: best candidate's score' facet shows a range from 63.00 to 101.00.

Reconcile column "advisor"

Reconcile each cell to an entity of one of these type

- ☒ Names
- ☐ Subjects
- ☐ LoC

Refine advisors_RTDs.xlsx Permalink

Facet / Filter Undo / Redo 6

Refresh Reset All Remove All

advisor change

235 choices Sort by: name count Cluster

Abbott, David W. 7

Acierno, Louis J. 2

Adicks, Richard R. 1

Alsaka, Yacoub A. 1

Anderson, Loren A. 3

advisor: judgment change

2 choices Sort by: name count

matched 313

none 655

Facet by choice counts

advisor: best candidate's score change reset

63.00 — 101.00

☒ Numeric 482 ☐ Non-numeric 0 ☐ Blank 0 ☒ Error 486

968 rows

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

| | | identifier | advisor |
|---|------|------------|--|
| ☆ | 49. | DP0003503 | Abbott, David W.
Choose new match |
| ☆ | 568. | DP0013348 | Abbott, David W.
Choose new match |
| ☆ | 778. | DP0015547 | Abbott, David W.
Choose new match |
| ☆ | 847. | DP0016479 | Abbott, David W.
Choose new match |
| ☆ | 953. | DP0019479 | Abbott, David W.
Choose new match |
| ☆ | 954. | DP0019480 | Abbott, David W.
Choose new match |
| ☆ | 955. | DP0019481 | Abbott, David W.
Choose new match |
| ☆ | 802. | DP0015862 | Acierno, Louis J.
<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Acierno, Louis J., 1920- (86)
Create new topic |
| ☆ | 890. | DP0017154 | Acierno, Louis J.
<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Acierno, Louis J., 1920- (86)
Create new topic |
| ☆ | 146. | DP0011933 | Adicks, Richard R.
<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new topic |
| ☆ | 711. | DP0013883 | Alsaka, Yacoub A.
<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new topic |
| ☆ | 712. | DP0013884 | Anderson, Loren A. |

Matched 313
(about
32.33%).

Adding LC and VIAF Links to Advisor Names

rows records Show: 5 10 25 50 rows Sort ▼

| | identifier | advisor | |
|------|------------|--------------------------------------|------------|
| 49. | DP0003503 | Abbott, David W.
Choose new match | JWS Sort ▼ |
| 568. | DP0013348 | Abbott, David W.
Choose new match | |

| ☆ | 🗨 | 49. | DP0003503 | Abbott, David | edit | |
|---|---|------|-----------|---|------|--|
| | | | | <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Abbott, David W. (100) | | |
| | | | | <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Abbott, David W., 1936- (85) | | |
| | | | | <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Abbott, David W., 1939- (85) | | |
| | | | | <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new topic | | |
| ☆ | 🗨 | 568. | DP0013348 | Abbott, David W. | | |
| | | | | Choose new match | | |
| ☆ | 🗨 | 778. | DP0015547 | Abbott, David W. | | |
| | | | | Choose new match | | |
| ☆ | 🗨 | 847. | DP0016479 | Abbott, David W. | | |
| | | | | Choose new match | | |
| ☆ | 🗨 | 953. | DP0019479 | Abbott, David W. | | |
| | | | | Choose new match | | |
| ☆ | 🗨 | 954. | DP0019480 | Abbott, David W. | | |
| | | | | Choose new match | | |
| ☆ | 🗨 | 955. | DP0019481 | Abbott, David W. | | |
| | | | | Choose new match | | |
| ☆ | 🗨 | 802. | DP0015862 | Acierno, Louis J. | | |
| | | | | <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Acierno, Louis J., 1920- (86) | | |
| | | | | <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new topic | | |

Review

- For names linked, click the **name** to check;
- If it's not the right person, click "Choose new match";
- It opens several options, re-check and apply.

Adding LC and VIAF Links to Advisor Names

Review

| | | | | | |
|---|---|------|-----------|---|------|
| ☆ | 🗨 | 802. | DP0015862 | Acierno, Louis J. | edit |
| | | | | <input checked="" type="checkbox"/> Acierno, Louis J., 1920- (86) | |
| | | | | <input checked="" type="checkbox"/> Create new topic | |
| ☆ | 🗨 | 890. | DP0017154 | Acierno, Louis J. | |

Match this topic to this and all identical cells

Click the link under “Acierno, Louis J.”, it opens:

<http://id.loc.gov/authorities/names/n85099714.html>

Acierno, Louis J., 1920-

Sources

found: His Comprehensive cardiac rehabilitation and prevention, c1985: CIP t.p. (Louis J. Acierno, M.D., F.A.C.C., F.A.C.P.; associate prof. of pub.

health, [Univ. of Central Fla.](#)) data sheet (b. 1920)

found: His A history of cardiology, 1993: CIP t.p. (L.J. Acierno; [Univ. of Florida](#), USA)

- For names unlinked, click the other options to check;
- If it's the right person, click ✓ (apply to this cell only) or ✓✓ (apply to all identical cells) to apply.

Adding LC and VIAF Links to Advisor Names

| | | | |
|------|-----------|---|----------------------|
| 43. | DP0003494 | Clausen, Chris A. | edit |
| | | <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Clausen, Chris A., 1940- (86) | |
| | | <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new topic | |
| 392. | DP0012918 | Claus | |
| | | <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Clausen, Chris A., 1940- (86) | |
| | | <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new topic | |

Match this topic to this and all identical cells

Clausen, Chris A., 1940- <http://id.loc.gov/authorities/names/n78032138.html>

Sources

found: His Principles of industrial ... 1978: t.p. (Chris A. Clausen, Dept. of Chem., Fla. Tech. U., Orlando) pub. info. (b. 12/7/40)

Data type: text

Doering, Robert D.

Apply Apply to All Identical Cells Cancel

Enter Ctrl-Enter Esc

Doering, R. D.
☒ ☒ Create new topic

Doering, R. D.
☒ ☒ Create new topic

Doering, R. D.
☒ ☒ Create new topic

Doering, R. D.
☒ ☒ Create new topic

199. DP0012014

Doering, Robert D., 1925- <http://id.loc.gov/authorities/names/no2018055990.html>

Has Affiliation

Organization: [Florida Technological University](#)

Organization: [University of Central Florida. Department of Industrial Engineering and Management Systems](#)

Associated Locale [Orlando \(Fla.\)](#)

Review

- For names unlinked, click the other options to check (e.g., the LC links below);
- If it's the right person, click ✓ or ✓✓ to apply;
- If no right name found, click "Create new topic";
- Input the right name and apply.

Adding LC and VIAF Links to Advisor Names

Narsingh Deo, 1936-

<http://id.loc.gov/authorities/names/n2006056273.html>

Sources

found: Sysło, Maciej M. **Discrete optimization algorithms**, 2007: CIP t.p.
(Narsingh Deo)

found: LC database, July 20, 2006: (Deo, Narsingh, 1936-)

Wikipedia: https://en.wikipedia.org/wiki/Narsingh_Deo

...**Narsingh Deo** is a professor and [Charles N. Millican](#) Endowed Chair of the Department of Computer Science, [University of Central Florida](#)...

Books

Discrete Optimization Algorithms: With Pascal Programs (with M.M. Sysło and J. S. Kowalik), Prentice-Hall, Englewood Cliffs, N.J., 1983, 542 pages.

| All | identifier | advisor |
|------|------------|--|
| 29. | DP0002215 | Deo, Narsingh
<input checked="" type="checkbox"/> Deo, Narsingh, 1936- (83)
<input checked="" type="checkbox"/> Create new topic |
| 140. | DP0011927 | Dietz
Match this topic to this and all identical cells
Choose new match |

Review

- If LCNAF doesn't provide enough information, further check the web (e.g., Wikipedia, Amazon books, obituaries, university website, newsletter...)
- When confirmed that it's the right person, click ✓ or ✓✓ to apply.

Adding LC and VIAF Links to Advisor Names

Add Column for URLs

- Edit column—Add column based on this column
- In the opened new window, fill in:
New column name: advisor_lc
Expression: cell.recon.match.id

968 rows

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

| | | identifier | advisor |
|---|---|------------|-----------|
| ★ | 🗨 | 49. | DP0003503 |
| ★ | 🗨 | 568. | DP0013348 |
| ★ | 🗨 | 778. | DP0015547 |
| ★ | 🗨 | 847. | DP0016479 |
| ★ | 🗨 | 953. | DP0019479 |

Context menu for 'advisor' column:

- Facet
- Text filter
- Edit cells
- Edit column** (selected)
- Transpose
- Sort

Options for 'Edit column':

- Split into several columns...
- Add column based on this column...** (circled)
- Add column by fetching URLs...

Add column based on column advisor

New column name: **advisor_lc** (circled)

Options: ☒ set to blank ☐ store error ☐ copy value from original column

Expression: **cell.recon.match.id** (circled) Language: General Refine Expression Language (GREL) ▼ No syntax error.

Preview:

| 802. | Acierno, Louis J. | http://id.loc.gov/authorities/names/n85099714 |
|------|--------------------|---|
| 890. | Acierno, Louis J. | http://id.loc.gov/authorities/names/n85099714 |
| 146. | Adicks, Richard R. | Error: Cannot retrieve field from null |
| 711. | Alsaka, Yacoub A. | Error: Cannot retrieve field from null |
| 712. | Anderson, Loren A. | http://id.loc.gov/authorities/names/n80002784 |
| 881. | Anderson, Loren A. | http://id.loc.gov/authorities/names/n80002784 |
| 949. | Anderson, Loren A. | http://id.loc.gov/authorities/names/n80002784 |
| 3. | Armstrong, Jack | Error: Cannot retrieve field from null |

OK Cancel

Adding LC and VIAF Links to Advisor Names

968 rows Extensions

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

| ▼ All | | ▼ identifier | ▼ advisor | ▼ advisor_lc |
|-------|---|--------------|---|---|
| ☆ | 🗨 | 802. | DP0015862
Acierno, Louis J., 1920-
Choose new match | http://id.loc.gov/authorities/names/n85099714 |
| ☆ | 🗨 | 890. | DP0017154
Acierno, Louis J., 1920-
Choose new match | http://id.loc.gov/authorities/names/n85099714 |
| ☆ | 🗨 | 146. | DP0011933
Adicks, Richard R.
<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new topic | |
| ☆ | 🗨 | 711. | DP0013883
Alsaka, Yacoub A.
<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new topic | |
| ☆ | 🗨 | 712. | DP0013884
Anderson, Loren A.
Choose new match | http://id.loc.gov/authorities/names/n80002784 |
| ☆ | 🗨 | 881. | DP0017018
Anderson, Loren A.
Choose new match | http://id.loc.gov/authorities/names/n80002784 |
| ☆ | 🗨 | 949. | DP0019440
Anderson, Loren A.
Choose new match | http://id.loc.gov/authorities/names/n80002784 |
| ☆ | 🗨 | 3. | DP0000186
Armstrong, Jack
<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Armstrong, Jack (100)
<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Armstrong, Jack H. (93)
<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Armstrong, Jack L. (93)
<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new topic | |
| ☆ | 🗨 | 664. | DP0013700
Armstrong, Lee H., 1941-
Choose new match | http://id.loc.gov/authorities/names/n79131864 |

Add Column for URLs: LC Links added

Adding LC and VIAF Links to Advisor Names

Add VIAF Links

Show as: **rows** records Show: 5 10 25 50 rows Sort ▼ « first < previous 951 - 968 next > last »

| ▼ All | ▼ identifier | ▼ advisor | ▼ advisor_c | ▼ advisor_lc |
|-------|--------------|-----------|--|---|
| ☆ | 419. | DP0013092 | Yousef, Yousef A.
<small>Choose new match</small> | http://id.loc.gov/authorities/names/n8202792
<input checked="" type="checkbox"/> Yousef, Hoda A., 1979- Yūsuf, Hudā Aḥmad, 1979-..... 1979 يوسف، هدى أ. (1)
<input checked="" type="checkbox"/> Yousef, Tarek A. (1)
<input checked="" type="checkbox"/> Alqudah, Mohammad Ali Yousef (1)
<input checked="" type="checkbox"/> Yousef, Hassan A. (1)
<input checked="" type="checkbox"/> Yousef, Ahmed Mohamed Fahmy (1)
<input checked="" type="checkbox"/> Mousa, Wael A. Yousef (1)
<input checked="" type="checkbox"/> Yousef, Eltohamy A. A. 1982- (1)
<input checked="" type="checkbox"/> Yousef, Mohamed A. M. (1)
<input checked="" type="checkbox"/> Yousef, Yousef A. (1)
<input checked="" type="checkbox"/> Abu Hammad, Ahmad Hammad Yousef
<input checked="" type="checkbox"/> Hammad, Ahmad Hammad Yousef
<input checked="" type="checkbox"/> At
<input checked="" type="checkbox"/> Cr |

Match this topic to this and all identical cells

Add a duplicate column for advisor

New column name advisor_c

Expression value

[advisor_c] Reconciliation

Select **Virtual International Authority File service**

(*Give more options to choose from if not matched directly)

Or, select **VIAF-LC service**

How to set up VIAF-LC Reconciliation Service

Click “Add Standard Service”;

Enter the service’s URL:

<http://refine.codefork.com/reconcile/viaf/LC>

How to set up Virtual International Authority File service

Enter the service’s URL:

http://iphylo.org/~rpage/phyloinformatics/services/reconciliation_viaf.php

Adding LC and VIAF Links to Advisor Names

Add VIAF Links

New column name:

advisor_viaf

Expression:

"https://viaf.org/viaf/" +
cell.recon.match.id

RTDs.xlsx Permalink

968 rows Extension

Remove All Show as: rows records Show: 5 10 25 50 rows Sort

change All identifier advisor advisor_c advisor_lc

419. DP0013092 Yousef, Yousef A. Yousef, Yousef A. http://id.loc.gov/au

Add column based on column advisor_c

New column name: advisor_viaf

☒ set to blank ☐ store error ☐ copy value from original column

Expression: "https://viaf.org/viaf/" + cell.recon.match.id Language: General Refine Expression Language (GREL) No syntax error.

Preview History Starred Help

| row | value |
|------|---|
| 419. | Yousef, Yousef A. https://viaf.org/viaf/272732530 |
| 429. | Yousef, Yousef A. https://viaf.org/viaf/272732530 |

475 968 rows Extensions: Named-entity recognition RDF

500 Show as: rows records Show: 5 10 25 50 rows Sort « first < previous 951 - 968 next > last »

| identifier | advisor | advisor_c | <u>advisor_viaf</u> | advisor_lc |
|------------|---|---------------------------------------|---------------------------------|---|
| 419. | DP0013092 Yousef, Yousef A.
Choose new match | Yousef, Yousef A.
Choose new match | https://viaf.org/viaf/272732530 | http://id.loc.gov/authorities/names/n82027924 |
| 429. | DP0013117 Yousef, Yousef A.
Choose new match | Yousef, Yousef A.
Choose new match | https://viaf.org/viaf/272732530 | http://id.loc.gov/authorities/names/n82027924 |
| 475. | DP0013238 Yousef, Yousef A.
Choose new match | Yousef, Yousef A.
Choose new match | https://viaf.org/viaf/272732530 | http://id.loc.gov/authorities/names/n82027924 |

OK

VIAF links
added

Adding LC and VIAF Links to Advisor Names

- Display in the IR

STARS Citation

Wanielista, Joseph R., "Removal of Color From Surface Water in Central Florida
Retrospective Theses and Dissertations. 80.
<http://stars.library.ucf.edu/rtd/80>

Contributor (Linked data)

[Yousef, Yousef A.](#) [VIAF]

[Yousef, Yousef A.](#) [LC]

[University of Central Florida. College of Engineering](#) [VIAF]

[University of Central Florida. College of Engineering](#) [LC]

Collection (Linked data)

[Retrospective Theses and Dissertations](#)

LOD links
added for:
Advisor name
College name

Home | About | FAQ | My Account
Privacy Copyright

<http://stars.library.ucf.edu/rtd/80/>

*Adding LOD links for the
advisor names is a work in
progress.

STARS Citation

Dory, Ivan Ray, "Solid Waste Resource Recovery Plan for Dekalb County Georgia" (1976).
Retrospective Theses and Dissertations. 211.
<http://stars.library.ucf.edu/rtd/211>

Contributor (Linked data)

[Yousef, Yousef A.](#)

[University of Central Florida. College of Engineering](#)

Collection (Linked data)

[Retrospective Theses and Dissertations](#)

<http://stars.library.ucf.edu/rtd/211/>

*Add both links or
only one type of
link for advisors
(e.g., VIAF, or LC)

Adding LC and VIAF Links to Author Names

- **Using Expression to Format Multiple Links**

- Expression **Language: General Refine Expression Language (GREL)**
- First, run reconciliations and get VIAF and LC links;
- Then add columns, e.g.,
 - Edit column – Add column based on this column
 - **New Column name** author1_viafurl
 - **Expression** "http://viaf.org/viaf/" + cell.recon.match.id
 - Edit column – Add column based on this column
 - **New column name** author1_lcurl
 - **Expression** cell.recon.match.id

The screenshot shows the Refine interface with a table of 102 rows. The table has columns for 'author1' and 'author1_copy'. Two new columns have been added: 'author1_lcurl' and 'author1_viafurl', both circled in blue. The 'author1_lcurl' column contains the URL 'http://id.loc.gov/authorities/names/n2008001755' and the 'author1_viafurl' column contains the URL 'http://viaf.org/viaf/70855412'. The interface also shows a facet for 'author1_copy: judgment' with 1 choice and a sort by 'name count'.

Refine UCFForum_batch2.xlsx Permalink

Facet / Filter Undo / Redo 42

Refresh Reset All Remove All

102 rows

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

Extensions: RDF

| | | ▼ All | ▼ title | ▼ author1 | ▼ author1_copy | ▼ author1_lcurl | ▼ author1_viafurl | ▼ contribu |
|---|-----|-------|------------------------|--------------------------------------|--------------------------------------|---|-------------------------------|------------|
| ★ | 41. | | Where Are Your Papers? | McCloud, Rebekah
Choose new match | McCloud, Rebekah
Choose new match | http://id.loc.gov/authorities/names/n2008001755 | http://viaf.org/viaf/70855412 | |

« first < previous 41 - 50 next > last »

author1_copy: judgment change

1 choices Sort by: name count

none 69

Facet by choice counts

Adding LC and VIAF Links to Author Names

- Using Expression to Format Multiple Links

- Add column based on this column

- New column name author1_viafurl_lcurl

- Expression

cells["author1_viafurl"].value + " " + cells["author1"].value +
"[VIAF] " + cells["author1_lcurl"].value + " " +
cells["author1_copy"].value + "[LC]"

| rows | | | | « first < previous 41 - 50 next > last » | |
|------|---|---|--|--|--|
| | ▼ author1_lcurl | ▼ author1_viafurl | ▼ author1_viafurl_lcurl | ▼ | |
| | http://id.loc.gov/authorities/names/n2008001755 | http://viaf.org/viaf/70855412 | http://viaf.org/viaf/70855412 McCloud, Rebekah
[VIAF]
http://id.loc.gov/authorities/names/n2008001755
McCloud, Rebekah [LC] | ▲ | |

Adding LC Subject Links to IR Records

- Create a project for one “UCF Forum” dataset in OpenRefine;
 - Its keywords field has a mixture of controlled and uncontrolled terms taken from the articles and supplied by the Metadata Librarian;
 - The goal is to add LOD links for the controlled terms.
- Splitting Keywords into Several Columns

The screenshot shows the OpenRefine interface with a dataset of 102 rows. The 'keywords' column is selected, and the 'Edit column' menu is open, showing the 'Split into several columns...' option. The 'Split column keywords into several columns' dialog box is open, showing the 'How to Split Column' section with the 'by separator' option selected. The separator is set to a semicolon (;). The 'After Splitting' section has 'Guess cell type' and 'Remove this column' checked. The 'OK' button is highlighted.

102 rows Extensions: RDF

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

| contributors | description | document_type | publication_date | season | keywords | subjects | disciplines | abstract | comments |
|--------------|---|----------------|----------------------|--------|--|-------------------------------------|-------------|--|----------|
| | The University of Central Florida (UCF) Forum is a weekly series of opinion columns presented by the UCF News & Information. The columns are the opinions of the writers, who serve on the UCF Forum panel of faculty members, staffers and students for a year. A new column is posted each Wednesday at UCF Today and then broadcast between 7:50 | opinion_column | 2016-11-02T00:00:00Z | | Facet
Text filter
Edit cells
Edit column
Transpose
Sort...
View
Reconcile | Communication;
Political Science | | <p>Tuesday, Nov. 8, is Game Day! The election is just days away and the news is all abuzz this morning about "rigged" elections, sexual-assault allegations, FRI | |

Split column keywords into several columns

How to Split Column

☒ by separator
Separator: ; regular expression
Split into columns at most (leave blank for no limit)

☐ by field lengths

After Splitting

☒ Guess cell type
☒ Remove this column

List of integers separated by commas, e.g., 5, 7, 15

OK Cancel

Under “Keywords”, click “Edit column – Split into several columns”; “by separator “;”

Adding LC Subject Links to IR Records

- Create a project for one “UCF Forum” dataset in OpenRefine;
- Keywords splitted into Several Columns

OpenRefine interface showing a dataset with 102 rows. The table has columns: document_type, publication_date, season, keywords 1, keywords 2, keywords 3, keywords 4, keywords 5, keywords 6, keywords 7, and keywords 8. A blue oval highlights the keywords columns.

| document_type | publication_date | season | keywords 1 | keywords 2 | keywords 3 | keywords 4 | keywords 5 | keywords 6 | keywords 7 | keywords 8 |
|----------------|----------------------|--------|------------|------------|------------|------------|----------------|------------|------------|------------|
| opinion_column | 2016-11-02T00:00:00Z | | Election | Game day | Football | Voting | Voting process | Democracy | America | |

OpenRefine interface showing a context menu for the 'keyword1' column. The menu options are: Facet, Text filter, Edit cells, Edit column, Transpose, Sort..., View, and Reconcile. The 'Reconcile' option is selected, and a sub-menu is shown with options: Start reconciling..., Facet, Reconcile text in this column with topics on Freebase, QA facets, Actions, Copy reconciliation data..., and Discover related RDF datasets...

| keyword1 | keyword2 | keyword3 | keyword4 | keyword5 |
|----------|----------|----------|----------------|----------|
| Game day | Football | Voting | Voting process | |

Top graph:

Keywords splitted (by default, e.g., keywords 1, 2, 3...)

Rename the fields, delete the space in field names
keywords 1, 2, 3... → keyword1, 2, 3...

Left graph:

Reconcile column “keyword1”

Reconcile each cell against **LoC Reconciliation Service**

Adding LC Subject Links to IR Records

- Review, select and confirm matched values (when needed)

| keyword1 | keyword2 | keyword3 | keyword4 | keyword5 |
|---|--------------------|------------------|------------|-----------|
| Hispanics edit
<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Hispanics of achievement (55)
<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Hispanics in Philanthropy (53)
<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Hispanics in the United States (46)
<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new topic | Hispanic community | Hispanic culture | Immigrants | Diversity |

Match this topic to this cell

Show: 5 10 25 50 rows

| keyword1 | keyword2 | keyword3 | keyword4 |
|----------|----------|----------|----------|
| Game day | Football | Voting | |

Facet
 Text filter
 Edit cells
 Edit column
 Transpose
 Sort...
 View
 Reconcile

Split into several columns...
 Add column based on this column...
 Add column by fetching URLs...
 Rename this column
 Remove this column
 Move column to beginning
 Move column to end
 Move column left
 Move column right

Edit column – Add column based on this column

Adding LC Subject Links to IR Records

- Adding column for links

New column name

keyword1_url

Expression

cell.recon.match.id + " | "
+ cell.recon.match.name

102 rows

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

| type | publication_date | season | keyword1 | keyword2 | keyword3 | keyword4 | keyword5 |
|------|----------------------|--------|---|----------|----------|----------|----------|
| | 2016-11-02T00:00:00Z | | Election
Election (Wills) (73)
Election (Obama) | Game day | Football | Voting | Voting |

Add column based on column keyword1

New column name

☒ set to blank ☐ store error ☐ copy value from original column

Expression Language No syntax error.

Preview History Starred Help

| | | |
|-----|----------------------|---|
| 16. | Ethnicity | http://id.loc.gov/authorities/subjects/sh85045187 Ethnicity |
| 17. | African Americans | http://id.loc.gov/authorities/subjects/sh85001932 African Americans |
| 18. | Black inventors | Error: Cannot retrieve field from null |
| 19. | Segregated community | Error: Cannot retrieve field from null |
| 20. | Kindness | http://id.loc.gov/authorities/subjects/sh85072376 Kindness |
| 21. | Documentary film | http://id.loc.gov/authorities/subjects/sh85088115 Documentary films |
| 22. | Parenting | http://id.loc.gov/authorities/subjects/sh85098009 Parenting |

102 rows

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

Extensions:

| type | publication_date | season | keyword1 | keyword1_url | keyword2 | keyword3 | keyword4 | keyword5 |
|------|----------------------|--------|-------------------------------|---|----------|------------------|------------|----------|
| | 2017-03-08T00:00:00Z | | Fake news
Choose new match | http://id.loc.gov/authorities/subjects/sh2017004659 Fake news | Election | Mainstream media | Bipartisan | News c |

Links added for keyword1;
Repeating the same process for keyword2, 3, 4...

Adding LC Subject Links to IR Records

- Putting Links together

Add column based on column keyword15_url

New column name

☒ set to blank ☐ store error ☐ copy value from original column

Expression

Language

No syntax error.

Preview History Starred Help

| | | |
|----|------|--|
| | | http://id.loc.gov/authorities/subjects/sh99005098 Propaganda |
| 4. | null | http://id.loc.gov/authorities/childrensSubjects/sj96006189 Sports
http://id.loc.gov/authorities/names/n88523062 Equal opportunity in employment
http://id.loc.gov/authorities/subjects/sh85045096 Ethics
http://id.loc.gov/authorities/subjects/sh85038376 Discrimination |
| 5. | null | http://id.loc.gov/authorities/names/no2008007024 NFL Network
http://id.loc.gov/authorities/subjects/sh85038376 Discrimination |
| 6. | | |
| 7. | | |

OK

Putting all subjects and their links together in a column

New column name

Expression

(*can add more values)

Subject links display together

* Use "|" or other dividers to separate subjects

Adding LOD Links to IR Records

Violence is Not a Virus, but it is Transmitted

Roberto Hugh Potter, University of Central Florida

Description

The University of Central Florida (UCF) Forum is a weekly series of opinion columns presented by the UCF News & Information. The columns are the opinions of the writers, who serve on the UCF Forum panel of faculty members, staffers and students for a year. A new column is posted each Wednesday at UCF Today and then broadcast between 7:50 and 8 a.m. Sunday on WUCF-FM (89.9).

Publication Date

9-5-2012

Keywords

Violence; Mass shooting; Firearms; Gun violence; Assailant; Firearm; Criminal behavior; Criminal violence; Campus violence

Introduction

Recent mass shootings have again brought out the "violence is a disease" crowd. The thinking is that if we reduce the availability of firearms, we will decrease violence.

Subjects

Violence; Mass shootings; Firearms; Criminal behavior

Creator (Linked Data)

Potter, Roberto Hugh(VIAF); Potter, Roberto Hugh(LC)

Collection

UCF Forum

Publisher

University of Central Florida

Source

<https://today.ucf.edu/violence-is-not-a-virus-but-it-is-transmitted/> (UCF Today: UCF Forum)

Type

article


STARS Citation

Potter, Roberto (2012). Violence is not a virus, but it is transmitted. UCF Today, 2012-09-05. Retrieved from <https://today.ucf.edu/violence-is-not-a-virus-but-it-is-transmitted/>

UCF Forum note

Information presented on this website is considered public information (unless otherwise noted) and may be distributed or copied. Use of appropriate byline/photo/image credit is

 Download

 INCLUDED IN

[Communication Commons](#),
[Criminology Commons](#)

SHARE



An Example in the IR
<http://stars.library.ucf.edu/ucf-forum/186/>

Subjects

Violence; Mass shootings; Firearms; Criminal behavior

Creator (Linked Data)

Potter, Roberto Hugh(VIAF); Potter, Roberto Hugh(LC)

Collection

UCF Forum

Adding FAST by Fetching URLs

review.xlsx Permalink

45 rows

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

▼ All ▼ Topical Subject ▼ loc_url

Add column by fetching URLs based on column Topical Subject

New column name Throttle delay milliseconds

On error ☒ set to blank ☐ store error ☒ Cache responses

Formulate the URLs to fetch:

Expression Language

General Refine Expression Language (GREL) ▼

No syntax error.

Preview History Starred Help

| row | value | |
|-----|-------------------|--|
| 1. | African Americans | http://experimental.worldcat.org/fast/search?query=oclc.heading+exact+%22African%20Americans%22&httpAccept=application/rdf%2bxml |
| 2. | Collectibles | http://experimental.worldcat.org/fast/search?query=oclc.heading+exact+%22Collectibles%22&httpAccept=application/rdf%2bxml |
| 3. | Blacks | http://experimental.worldcat.org/fast/search?query=oclc.heading+exact+%22Blacks %22&httpAccept=application/rdf%2bxml |
| 4. | United | http://experimental.worldcat.org/fast/search? |

OK Cancel

Add column by fetching URLs based on Topical Subject column

(Click “Edit column—add column by fetching URLs” under “Topical_Subject”)

New column name
fast_url


Expression
"http://experimental.worldcat.org/fast/search?query=oclc.heading+exact+%22"+value.replace(/(\s)/, '%20')+"%22&httpAccept=application/rdf%2bxml"

*Query FAST for RDFXML return

Adding FAST by Fetching URLs

Text Transformation

← → ↺ ⬆ 127.0.0.1:3333/project?project=1655121992645 ☆ 📄 🌐 ⋮

Refine  CarolMundyfieldreview.xlsx Permalink

Open... Export ▾ Help

Facet / Filter Undo / Redo

45 rows

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

Extensions: Named-entity recognition ▾ RDF ▾

« first | previous 1 - 10 next | last »

| ▼ All | ▼ Topical Subject | ▼ fast_url | ▼ loc_url |
|--|-------------------|--|---|
| 1. African Americans
Choose new match | | <pre><rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:coord="http://whatever.java/ORG.oclc.util.CoordToDecimal" xmlns:naco="http://whatever.java/ORG.oclc.util.NacoNormalize" xmlns:norm="http://whatever.java/ORG.oclc.util.NormalFormC" xmlns:schema="http://schema.org/" xmlns:madsrdf="http://www.loc.gov/mads/rdf/v1#" xmlns:skosx="http://www.w3.org/2008/05/skos-xl#" xmlns:dct="http://purl.org/dc/terms/" xmlns:foaf="http://xmlns.com/foaf/0.1/" xmlns:void="http://rdfs.org/ns/void#" xmlns:owl="http://www.w3.org/2002/07/owl#" xmlns:mx="http://www.loc.gov/MARC21/slim" xmlns:skos="http://www.w3.org/2004/02/skos/core#" xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#" xml:base="http://id.worldcat.org/fast" <rdf:Description rdf:about="799558"> <dct:identifier>799558</dct:identifier> <dct:replaces rdf:resource="1939238"/> <skos:inScheme rdf:resource="ontology/1.0/#fast"/> <rdf:type rdf:resource="http://schema.org/Intangible"/> <skos:inScheme rdf:resource="ontology/1.0/#facet-Topical"/> <skos:prefLabel>African Americans</skos:prefLabel> <schema:name>African Americans</schema:name> <skos:altLabel>Afro-Americans</skos:altLabel> <schema:name>Afro- Americans</schema:name> <skos:altLabel>Black Americans</skos:altLabel> <schema:name>Black Americans</schema:name> <skos:altLabel>Colored people (United States)</skos:altLabel> <schema:name>Colored people (United States) </schema:name> <skos:altLabel>Negroes</skos:altLabel> <schema:name>Negroes</schema:name> <skos:related> <rdf:Description rdf:about="799938"> <rdfs:label>Africans</rdfs:label> </rdf:Description> </skos:related> <skos:related> <rdf:Description rdf:about="1204155"> <rdfs:label>United States</rdfs:label> </rdf:Description> </skos:related> <skos:related> <rdf:Description rdf:about="916106"> <rdfs:label>Ethnology</rdfs:label> </rdf:Description> </skos:related> <skos:related> <rdf:Description rdf:about="833880"> <rdfs:label>Blacks</rdfs:label> </rdf:Description> </skos:related> <schema:sameAs> <rdf:Description rdf:about="http://id.loc.gov/authorities/subjects/sh85001932"> <rdfs:label>African</pre> | http://id.loc.gov/authorities/subjects/sh85001932 |

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](#)

Need to extract FAST ID from the returned content:
`<dct:identifier>799558</dct:identifier>`
...
`<dc:format>application/rdf+xml</dc:format>`

Adding FAST by Fetching URLs


Text Transformation

Refine OPEN CarolMundyfieldreview.xlsx [Permalink](#)

Facet / Filter Undo / Redo 5

45 rows Extension

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

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 | Topical Subject | fast_url |
|-----|--|---|
| 1. | <div>Facet</div> <div>Text filter</div> <div>Edit cells</div> <div>Edit column</div> <div>Transpose</div> <div>Sort...</div> <div>View</div> <div>Reconcile</div> <div>Extract named entities...</div> | <div><rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:coord="http://whatever/java/ORG.oclc.util.CoordToDecimal" va/ORG.oclc.util.NacoNormalize" va/ORG.oclc.util.NormalFormC" org/" xmlns:madsrdf="http://www.loc.gov/mads/rdf/v1#" rg/2008/05/skos-xl#" rms/" xmlns:foaf="http://xmlns.com/foaf/0.1/" oid#" xmlns:owl="http://www.w3.org/2002/07/owl#" MARC21/slim" y/2004/02/skos/core#" /2000/01/rdf-schema#" g/fast/"> <rdf:Description rdf:about="799558"> ntifier> <dc:replaces rdf:resource="1939238"/> "ontology/1.0/#fast"/> <rdf:type g/Intangible"/> <skos:inScheme et-Topical"/> <skos:prefLabel>African Americans</skos:prefLabel> <schema:name>African Americans</schema:name> <skos:altLabel>Afro-Americans</skos:altLabel> <schema:name>Afro-Americans</schema:name> <skos:altLabel>Black Americans</skos:altLabel> <schema:name>Black Americans</schema:name> <skos:altLabel>Colored people (United States)</skos:altLabel> <schema:name>Colored people (United States)</schema:name> <skos:altLabel>Negroes</skos:altLabel></div> |

Edit cells – Transform...

Adding FAST by Fetching URLs

preview.xlsx Permalink

45 rows

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

All Topical Subject fast_url

Custom text transform on column fast_url

Expression Language General Refine Expression Language (GREL) ▼

`"http://experimental.worldcat.org/fast/"+value.parseHtml().select('dct|identifier')[0].htmlText()` No syntax error.

Preview History Starred Help

| row | value |
|-----|--|
| 1. | <p><rdf.RDF
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:coord="http://whatever/java/ORG.oclc."
xmlns:naco="http://whatever/java/ORG.oclc."
xmlns:norm="http://whatever/java/ORG.oclc."
xmlns:schema="http://schema.org/"
xmlns:madsrdf="http://www.loc.gov/mads/rdf"</p> <p><u>http://experimental.worldcat.org/fast/799558</u></p> |

On error ☒ keep original ☐ Re-transform up to 10 times until no change ☐ set to blank ☐ store error

OK Cancel

Text Transformation
(Extract FAST identifier from the returned content)

Expression
"http://experimental.worldcat.org/fast/"+value.parseHtml().select('dct|identifier')[0].htmlText()

Adding FAST by Fetching URLs

| 45 rows | | | Extensions |
|-----------------------|---|--|--|
| Show as: rows records | | | Show: 5 10 25 50 rows |
| All | Topical Subject | fast_url | |
| 1. | African Americans
Choose new match | http://experimental.worldcat.org/fast/799558 | |
| 2. | Collectibles
Choose new match | http://experimental.worldcat.org/fast/867139 | |
| 3. | Blacks
Choose new match | http://experimental.worldcat.org/fast/833680 | |
| 4. | United States
<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> United States Courthouse (Austin, Tex.) (54)
<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> United States Capitol (Washington, D.C.) (53)
<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> United States Courthouse (Orlando, Fla.) (53)
<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new topic | http://experimental.worldcat.org/fast/799558 | <div> <div>Topical Subject</div> <div> 1. African Americans
Choose new match
 2. Collectibles
Choose new match
 3. Blacks
Choose new match </div> </div> |
| 5. | History
Choose new match | http://experimental.worldcat.org/fast/958235 | |
| 6. | Civil War, 1861-1865
<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new topic | <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:coord="http://whatever/java/ORG.oclc.util.CoordToDecimal" xmlns:naco="http://whatever/java/ORG.oclc.util.NacoNormalize" xmlns:norm="http://whatever/java/ORG.oclc.util.NormalFormC" xmlns:schema="http://schema.org/" xmlns:madsrdf="http://www.loc.gov/mads/rdf/v1#" xmlns:skosd="http://www.w3.org/2008/05/skos-xl#" xmlns:dct="http://purl.org/dc/terms/" xmlns:foaf="http://xmlns.com/foaf/0.1/" xmlns:void="http://rdfs.org/ns/void#" xmlns:owl="http://www.w3.org/2002/07/owl#" xmlns:mx="http://www.loc.gov/MARC21/slim" xmlns:skos="http://www.w3.org/2004/02/skos/core#" xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#" xml:base="http://id.worldcat.org/fast/1/"> <void:Dataset rdf:about="http://id.worldcat.org/fast/void/0.1/"> <dct:title>FAST (Faceted Application of Subject Terminology) Data</dct:title> <dct:license></dct:license> <rdf:Description rdf:about="http://www.oclc.org/research/activities/fast/odcby.html"> <dct:title>Open Data Commons Attribution License (ODC-By) v1.0</dct:title> <rdfs:comment>This is OCLC's "local copy" of the ODC-By license</rdfs:comment> <rdfs:seeAlso rdf:resource="http://opendatacommons.org/licenses/by/#toc-anchor-337-3"/> <owl:sameAs rdf:resource="http://opendatacommons.org/licenses/by/1-0/"> </rdf:Description> </dct:license> </void:Dataset> </rdf:RDF> | |
| 7. | African Americans in art
Choose new match | http://experimental.worldcat.org/fast/799722 | |
| 8. | African American children
<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new topic | http://experimental.worldcat.org/fast/799071 | |
| 9. | Children, Black, in art
Choose new match | http://experimental.worldcat.org/fast/856910 | |
| 10. | Racism in art | http://experimental.worldcat.org/fast/1086652 | |

FAST links added

*If returned text doesn't include a FAST identifier, it'll display the original text; If more than one match, will pick up the first one (index[0])

Other Possibilities

- **Augment Data with Reconciliation Services**

- Reconcile genre values to AAT terms
- Reconcile places to GeoNames
- Reconcile ORCID...

- **Reconcile Against Local Vocabularies**

- **reconciliation-csv**

- <http://okfnlabs.org/reconcile-csv/>
- Possibilities: Reconcile various fields in OpenRefine against local controlled term lists, e.g., medals, ranks for Veterans Oral History

- **Next Steps**

- Keep adding more LOD links to personal names, subjects and college/department names to digital collections;
- Continue to explore OpenRefine and related services to enrich metadata and improve workflow;
- Promote, explore and link authors/researchers to personal identifiers.

Resources

- Berners-Lee, Tim. Linked Data <https://www.w3.org/DesignIssues/LinkedData.html>
- BIBFRAME <http://BIBFRAME.org/>
- DBpedia <http://wiki.dbpedia.org/>
- DPLA <http://dp.la>
- Europeana <http://europeana.eu>
- FAST: Faceted Application of Subject Terminology <http://id.worldcat.org/fast>
- FAST Linked Data API. <https://www.oclc.org/developer/develop/web-services/fast-api/linked-data.en.html>
- Free Your Metadata. Linked Data. <http://freeyourmetadata.org/>
- GeoNames <http://www.geonames.org/ontology/documentation.html>
- International Standard Name Identifier (ISNI) <http://www.isni.org/>
- LC Linked Data Service: Authorities and Vocabularies <https://id.loc.gov/>
- Mak, Lucas et al. [A Lightweight Structured Data Implementation Using JSON-LD and Schema.org for Digital Repository.](#)
- Mitchell, Erik T. Building Blocks of Linked Open Data in Libraries. Library Technology Reports (vol. 49, no. 5)
- Mitchell, Erik T. Three Case Studies in Linked Open Data. Library Technology Reports (vol. 49, no. 5)
- Motley Collection of Theatre and Costume Design (New)—Hamlet--Caesar. <http://imagesearch-test1.library.illinois.edu/cdm/ref/collection/motley-new/id/599>
- OCLC Linked Data <https://www.oclc.org/developer/develop/linked-data.en.html>
- Open Researcher and Contributor ID (ORCID) <https://orcid.org/>

Resources

- OpenRefine. <http://openrefine.org/>
- OpenRefine. Documentation For Users. <https://github.com/OpenRefine/OpenRefine/wiki/Documentation-For-Users>
- OpenRefine Recipes. <https://github.com/OpenRefine/OpenRefine/wiki/Recipes>
- OpenRefine. Reconcilable data sources. <https://github.com/OpenRefine/OpenRefine/wiki/Reconcilable-Data-Sources>
- OpenRefine. Reconciliation. <https://github.com/OpenRefine/OpenRefine/wiki/Reconciliation>
- Orbis Cascade Alliance. OpenRefine for Metadata Cleanup. <https://www.orbiscascade.org/openrefine-for-metadata-cleanup/>
- Orbis Cascade Alliance. OpenRefine for Metadata Enrichment. <https://www.orbiscascade.org/using-openrefine-for-metadata-enrichment/>
- Ozmeka. <https://github.com/ozmeka/ozmeka>
- Reconcile-csv: <http://okfnlabs.org/reconcile-csv/>
Reconcile-csv in GitHub: <https://github.com/okfn/reconcile-csv>
- Schema.org Vocabulary <http://schema.org/docs/gs.html#schemaorg>
- Tay, Aaron. OpenRefine, & reconciliation services - learning the ropes. <http://musingsaboutlibrarianship.blogspot.com/2017/03/openrefine-reconciliation-services.html>
- The Linked Open Data Cloud <https://lod-cloud.net/>
- Transforming Library Metadata into Linked Library Data: Introduction and Review of Linked Data for the Library Community, 2003–2011 <http://www.ala.org/alcts/resources/org/cat/research/linked-data>
- Virtual International Authority File (VIAF) <https://viaf.org/>
- Williamson, Evan Peter. Fetching and Parsing Data from the Web with OpenRefine. <https://programminghistorian.org/lessons/fetch-and-parse-data-with-openrefine>
- Worldcat Identities www.worldcat.org/identities

Acknowledgements

- Lucas Mak, Metadata and Catalog Librarian, Michigan State University Libraries
- Systems, University of Central Florida Libraries
- Image credits
 - 5 Star Open data (<http://5stardata.info/en/>)
 - Linked Data
(<https://www.w3.org/DesignIssues/LinkedData.html>)
 - Linked Open Data Cloud (<https://lod-cloud.net/>)



Thank you!

Contact

Sai Deng

Metadata Librarian & Associate Librarian

University of Central Florida Libraries

sai.deng@ucf.edu