

Improvements
through
Automation:
MarcEdit



Approach for this session

I reviewed my email and pulled out a couple of common questions that will let me highlight a couple of concepts:

1. Encodings
2. Tasks
3. Unmediated automation

Following this discussion, I'm looking to leave time so folks can ask questions and we can work through edit strategies together.




Thinking about Diacritics


MARC Character Conversions

- Supports moving between any known Windows Characterset and MARC8.
- Can be run from the Breaker/Maker – or as its own standalone utility

Characterset Conversions


Character Conversions


Source File: 

Destination File: 

Batch Process

Encoding Settings

Original Encoding: 

Final Encoding: 

Why Diacritics can sometimes be a pain

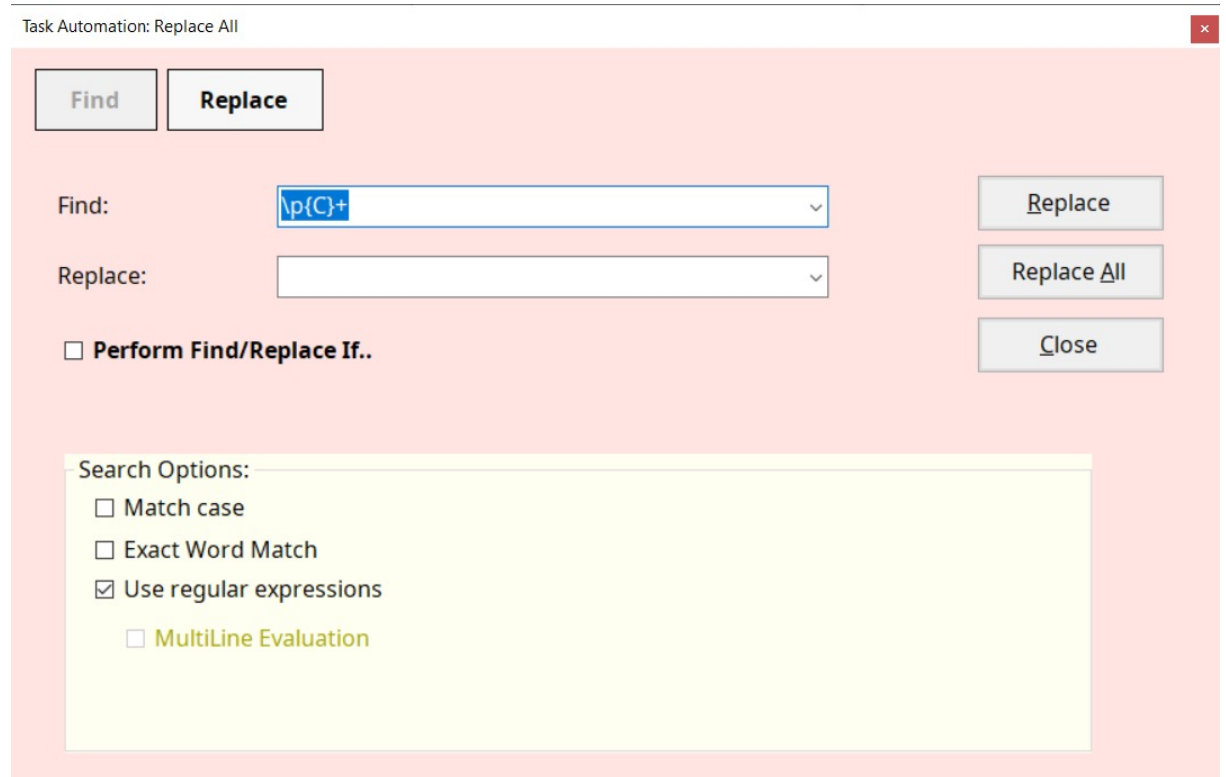
```
=090 \\$pREM  
=100 1\\$adailyBooks.
```

```
=245 10$aWhite fragility--why it's so hard for white people to talk about  
racism<U+0080{acute}><U+008B><U+0080{acute}><U+008B><U+  
0080{acute}><U+008B><U+0080{acute}><U+008B><U+0080{acute}><U+  
008B><U+0080{acute}><U+008B><U+0080{acute}><U+008B> by robin  
diangelo<U+0080{acute}><U+008B><U+0080{acute}><U+008B><U+  
0080{acute}><U+008B><U+0080{acute}><U+008B><U+0080{acute}><U+  
008B><U+0080{acute}><U+008B><U+0080{acute}><U+008B> |
```

```
conversation starters /$cdailyBooks.
```

```
=264 \1$aMilano :$hDaily Books $r2018
```

Using Regular Expressions to find Unprintable data



Microsoft's Regular Expression language

- Concepts:
 - Character escapes
 - Anchors
 - Character classes
 - Grouping
 - Qualifiers
 - Substitutions



Additional Resources:

Language Reference:

<https://docs.microsoft.com/en-us/dotnet/standard/base-types/regular-expression-language-quick-reference>

Printable Copy:

<https://download.microsoft.com/download/D/2/4/D240EBF6-A9BA-4E4F-A63F-AEB6DA0B921C/Regular%20expressions%20quick%20reference.pdf>

Web-Site to Test Expressions:

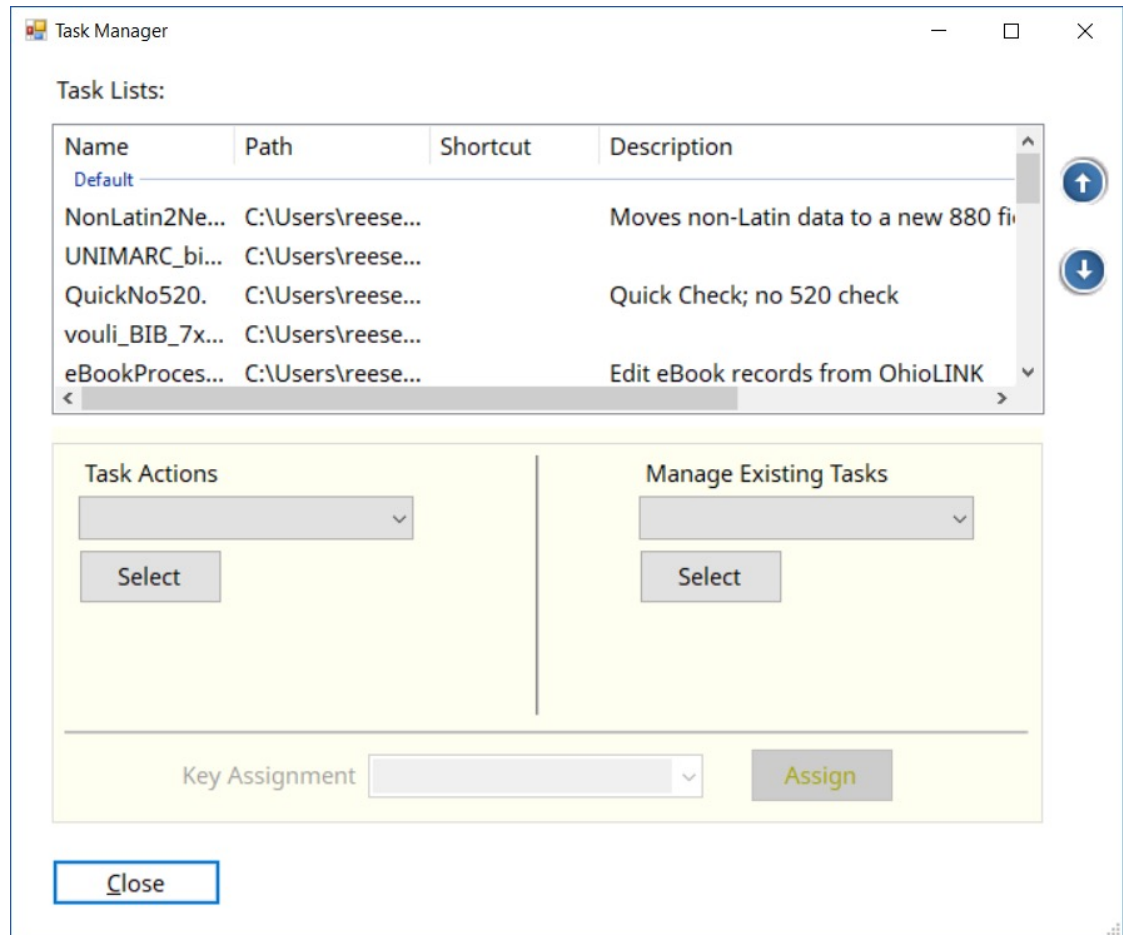
<http://regexstorm.net/tester>

How we use Regular Expressions in MarcEdit

- Your most important parts of the regular expression language are:
 - 1.Character escapes: `\d\w\r\n\$\p`
 - 2.Character Classes `[]` & `[^]`
 - 3.Grouping Elements `()`
 - 4.Anchors: `^$`
 - 5.Quantifiers: `*?+{#}`
 - 6.Substitutions: `$#`

Task Automation

- Managing Tasks
 - Task management works like macro management
 - You can
 - Create new tasks
 - Clone tasks
 - Rename tasks
 - Delete tasks
 - Edit tasks



Task Editor

- Supports copy and paste between tasks
- Adding Comments
- Creating Dynamic Task Lists
- Printing of Tasks
- Control Flow Options For Recursion
 - Counter Loop (does an action X# of times)
 - Results Loop (does action till results are ><= a result)
- User Defined Data added on run

The screenshot shows the 'Edit Task List' window with the following configuration:

- Task List Name:** Book2Electronic
- Description:** Converts book records to Electronic records
- Override Broker Action (not recommended)
- Tasks:**

Action	Count	Field	Value	Condition	Print	Repeat	Other	
DELETE	33x		0	False	False	False	False	
buildnewfield		=776	1	\\${245\$a}\$z{020\$a}\$x{022\$a}\$w{035\$a}.find				
DELETE	029		0	False	False	False	False	
DELETE	019		0	False	False	False	False	
DELETE	001		0	False	False	False	False	
DELETE	007		0	False	False	False	False	
ADD	588	\\\$aDescription based on print version record.						1
ADD	007	c						100
SUBFIELD_EDIT	300	a	.(^[^:]*)(.*)					\${1}1electr
SUBFIELD_EDIT	090	b	^e					ebook 0 0

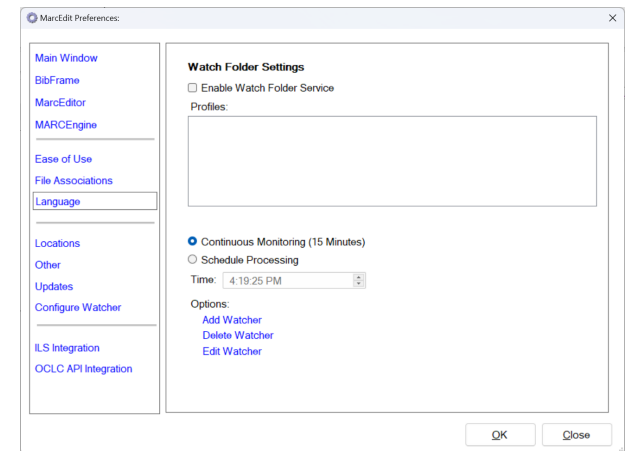
Actions:

Buttons: Save, Debug, Close

MarcEdit's Folder Watcher

Watch Folder is a tool that:

- Automate the download of records from an FTP or other report site
- Join and split records
- Process XML records to MARC
- Automatically apply tasks to all records in a file



Question: Is there a way to find records that specifically have a 650 /4 field (or any field or is missing a field?

- Is there a way to find records that specifically have a 650 /4 field (or any field?
- There are a variety of ways to accomplish this task, and the best method will depend on what the user is looking to do with the data. Available options can be broken down into 2 categories:
 1. Lists
 2. Data Subsets

Question

- Lists:
- From within the MarcEditor, a user can retrieve a records list of all the items matching a specific criteria. In this case, a user looking for a field with specific indicators could utilize the Find All Function with the following regular expression:
- *Expression: (=650.{3}[^7])*
- * Find all 650 fields, where the second indicator isn't a 7.

Replace Text

Find Replace

Find:

Replace:

[Add To Task](#)

Perform Find/Replace If..

Search Options:

- Match case
- Exact Word Match
- Use regular expressions
 - MultiLine Evaluation
- Use External Search/Replace Criteria

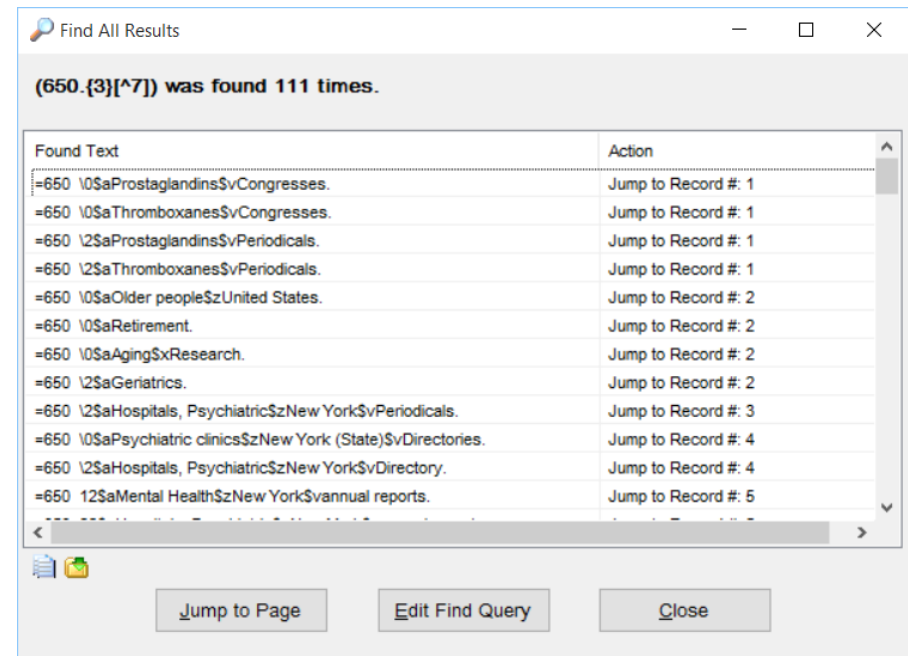
Replace

Replace All

Close

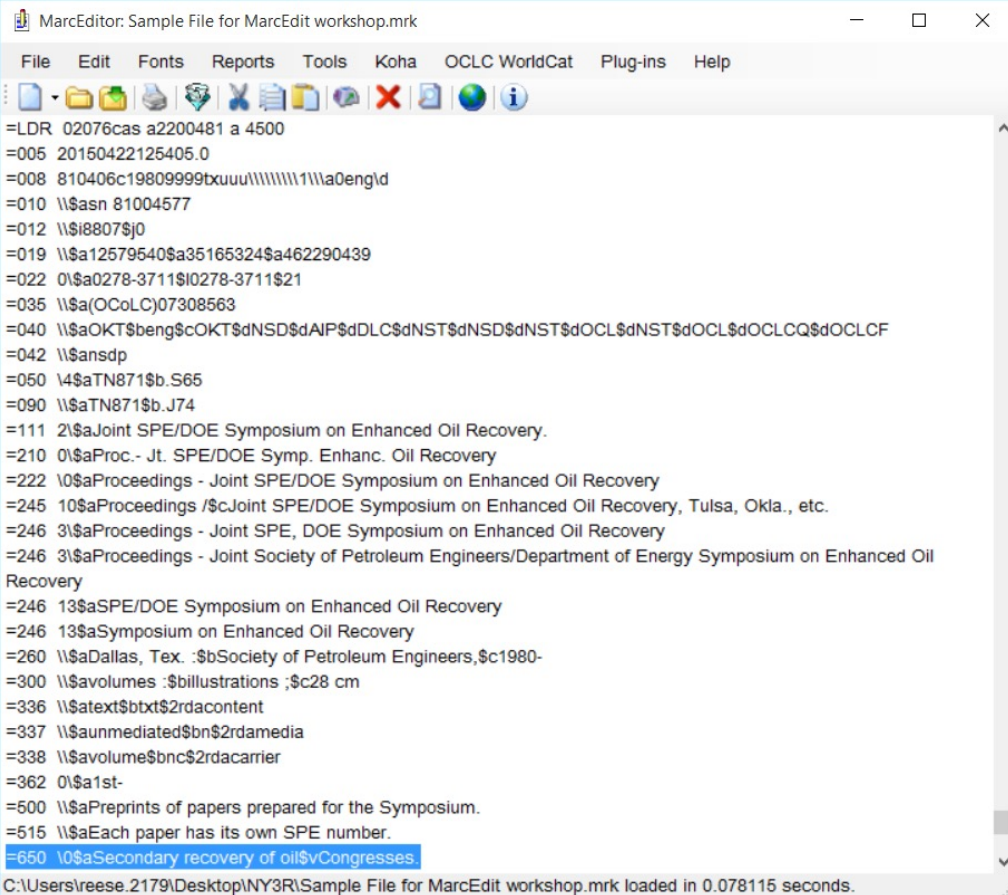
Question

- Lists:
- Expressions using Find All generate a Jump List that allow the user to see their query in context:



Question

- Lists:
- And selecting a record to jump to – takes the user to the record, with their field highlighted for evaluation



MarcEditor: Sample File for MarcEdit workshop.mrk

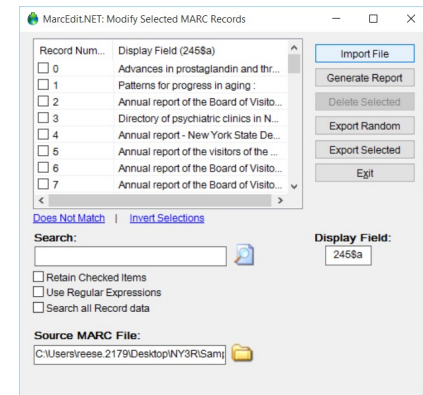
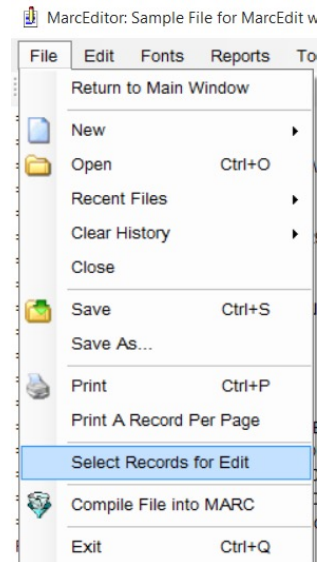
File Edit Fonts Reports Tools Koha OCLC WorldCat Plug-ins Help

```
=LDR 02076cas a2200481 a 4500
=005 20150422125405.0
=008 810406c19809999txuuu\|||||1\|a0engl d
=010 \\\$asn 81004577
=012 \\\$i8807$j0
=019 \\\$a12579540$a35165324$a462290439
=022 0\$\a0278-3711$\i0278-3711$\$21
=035 \\\$(OCoLC)07308563
=040 \\\$aOKT$beng$cOKT$dNSD$dAIP$dDLC$dNST$dNSD$dNST$dOCL$dNST$dOCL$dOCLCQ$dOCLCF
=042 \\\$ansdp
=050 \4$aTN871$b.S65
=090 \\\$aTN871$b.J74
=111 2\$\aJoint SPE/DOE Symposium on Enhanced Oil Recovery.
=210 0\$\aProc.- Jt. SPE/DOE Symp. Enhanc. Oil Recovery
=222 \0$aProceedings - Joint SPE/DOE Symposium on Enhanced Oil Recovery
=245 10\$\aProceedings /$cJoint SPE/DOE Symposium on Enhanced Oil Recovery, Tulsa, Okla., etc.
=246 3\|\aProceedings - Joint SPE, DOE Symposium on Enhanced Oil Recovery
=246 3\|\aProceedings - Joint Society of Petroleum Engineers/Department of Energy Symposium on Enhanced Oil
Recovery
=246 13\$\aSPE/DOE Symposium on Enhanced Oil Recovery
=246 13\$\aSymposium on Enhanced Oil Recovery
=260 \\\$aDallas, Tex. :$bSociety of Petroleum Engineers,$c1980-
=300 \\\$avolumes :$billustrations ;$c28 cm
=336 \\\$atext$btxt$2rdacontent
=337 \\\$aunmediated$bbn$2rdamedia
=338 \\\$avolume$bnc$2rdacarrier
=362 0\|\a1st-
=500 \\\$aPreprints of papers prepared for the Symposium.
=515 \\\$aEach paper has its own SPE number.
=650 \0\$\aSecondary recovery of oil$vCongresses.
```

C:\Users\reese.2179\Desktop\NY3R\Sample File for MarcEdit workshop.mrk loaded in 0.078115 seconds.

Question

- Data Subsets:
- MarcEdit includes a tool to Extract Selected Records...a tool that can be run from within the MarcEditor or outside the MarcEditor. Depending on where the tool is run, it will do different things.
- *Within the MarcEditor:*



Question

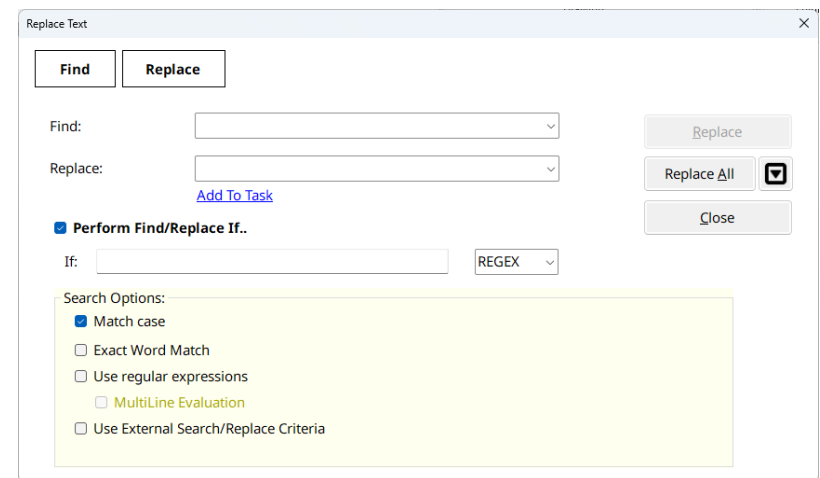
- Data Subsets:
- When run from within the MarcEditor, users can select a subset of data, pull that subset into the MarcEditor, edit just that data, and then save the data back into the original source file by clicking the “Save” button.
- When run from outside the MarcEditor, users can select a subset of data and export that subset into a new file. During export, users can request that the data being exported be removed from the source data file.

Question

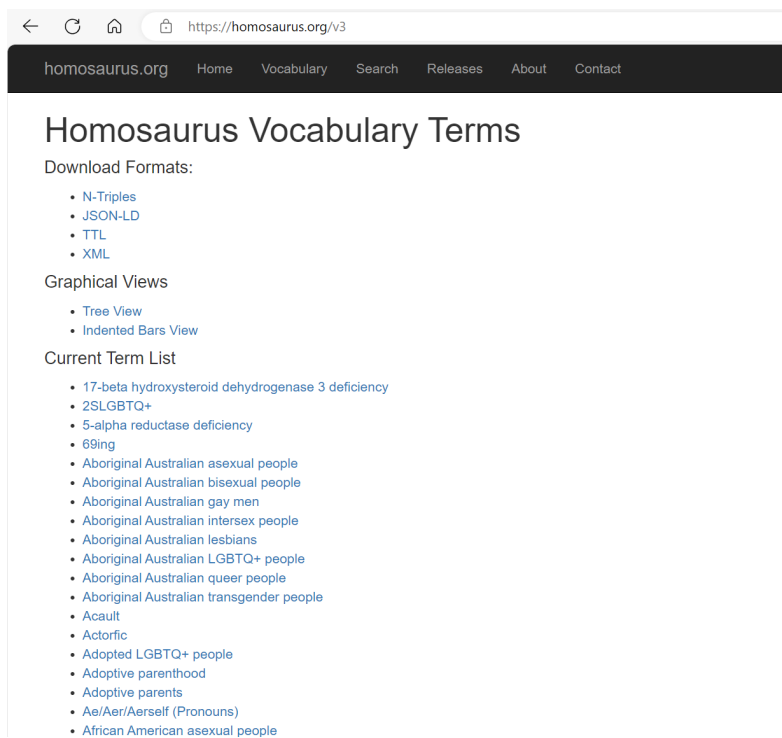
- Data Subsets:
- Extract/Delete Selected Records Search Options:
 1. General search – searches just the item from the display field
 2. Search all record data – searches all record data using either a regular expression or in-string match
 3. Invert selections – invert selected data.
 4. Find records that do not match a specific field
 5. Special Search options:
 1. F#:000\$a [search data]
 2. R#:1-12 [select a range]

Question: Is it possible to make conditional edits (changes based on data in other fields)?

- Yes, using the Replace function, you have a conditional edit option
- Conditional edits can do HAS or Reg Ex
 - Within these, they can be grouped, and include NOT, AND, and OR keywords



Integrating local vocabularies



← ↻ 🏠 <https://homosaurus.org/v3>

homosaurus.org Home Vocabulary Search Releases About Contact

Homosaurus Vocabulary Terms

Download Formats:

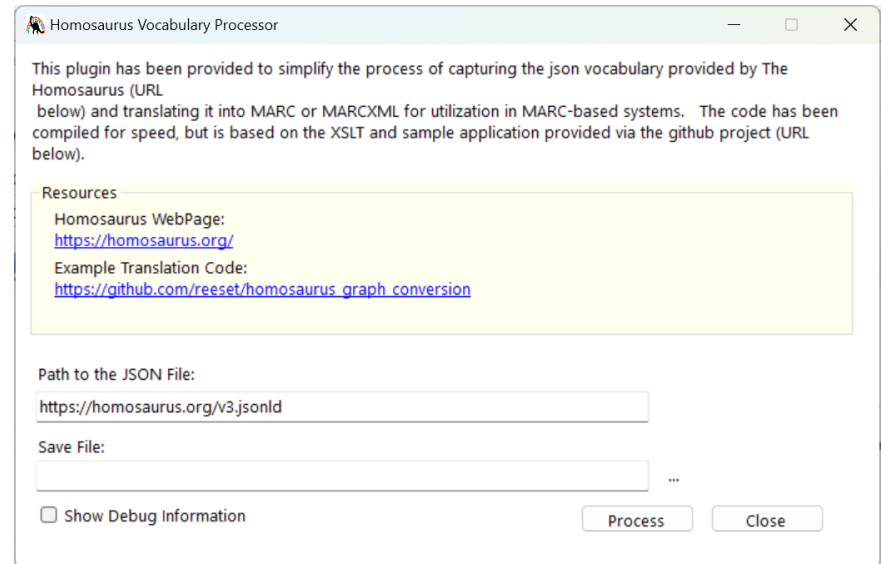
- N-Triples
- JSON-LD
- TTL
- XML

Graphical Views

- Tree View
- Indented Bars View

Current Term List

- 17-beta hydroxysteroid dehydrogenase 3 deficiency
- 2SLGBTQ+
- 5-alpha reductase deficiency
- 69ing
- Aboriginal Australian asexual people
- Aboriginal Australian bisexual people
- Aboriginal Australian gay men
- Aboriginal Australian intersex people
- Aboriginal Australian lesbians
- Aboriginal Australian LGBTQ+ people
- Aboriginal Australian queer people
- Aboriginal Australian transgender people
- Acault
- Actorfic
- Adopted LGBTQ+ people
- Adoptive parenthood
- Adoptive parents
- Ae/Aer/Aerself (Pronouns)
- African American asexual people



Homosaurus Vocabulary Processor

This plugin has been provided to simplify the process of capturing the json vocabulary provided by The Homosaurus (URL below) and translating it into MARC or MARCXML for utilization in MARC-based systems. The code has been compiled for speed, but is based on the XSLT and sample application provided via the github project (URL below).

Resources

Homosaurus WebPage:
<https://homosaurus.org/>

Example Translation Code:
https://github.com/reeset/homosaurus_graph_conversion

Path to the JSON File:

Save File:
 ...

Show Debug Information

Process Close

Resources

- MarcEdit Listserv: <https://listserv.gmu.edu/cgi-bin/wa?A0=marcedit-l>
- Email: reeseet@gmail.com
- Help documentation:
https://marcedit.reeseet.net/archives/category/knowledge_base

Questions

