Bringing Agrovoc terms into Koha ILS

Eugene Jose T. Espinoza O O O C O O 1

Consultant

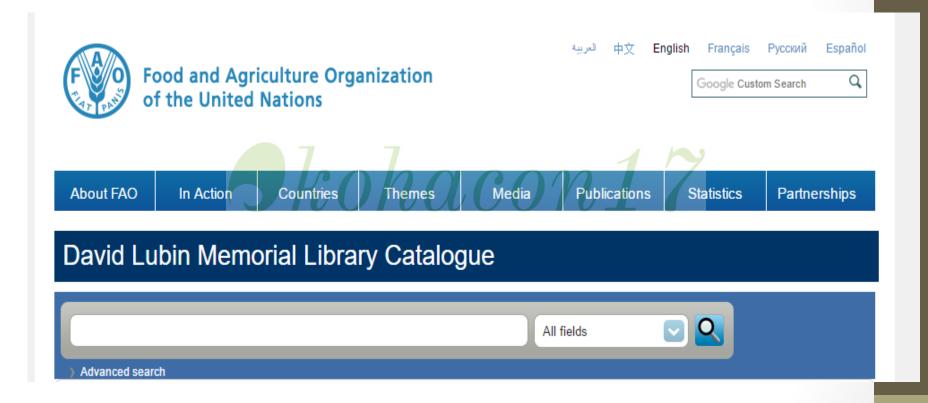
Co-owner and CTO, OnStrike Library Solutions Secretary, Association of Special Libraries of the Philippines Librarian, Chinese eneral Hospital Colleges

For 2017 Koha International Conference (KohaCon2017)

 Library Systems Analyst Consultant for the Food and Agriculture Organization of the United Nations in Rome, Italy (April 2015 – December 2015)

9kohacon17

Backgrounder



About Koha

Server Information Perl modules System information Koha Team Translations Koha Timeline Licenses Server information Koha version: 3.12.12.000 Linux unfao.koha.ptfsadmin.uk0.bigv.io 3.2.0-4-amd64 #1 SMP Debian 3.2.60-1+deb7u3 x86 64 GNU/Linux OS version ('uname -a'): Perl interpreter: /usr/bin/perl Perl version: 5.014002 /home/koha/kohaclone /etc/perl /usr/local/lib/perl/5.14.2 /usr/local/share/perl/5.14.2 /usr/lib/perl5 Perl @INC: /usr/share/perl5 /usr/lib/perl/5.14 /usr/share/perl/5.14 /usr/local/lib/site perl MySQL version: mysql Ver 14.14 Distrib 5.5.38, for debian-linux-gnu (x86 64) using readline 6.2 Apache version: Server version: Apache/2.2.22 (Debian) Zebra 2.0.59 (C) 1994-2014, Index Data Zebra is free software, covered by the GNU General Public License, and you are welcome to change it and/or distribute copies of it under certain conditions. SHA1 Zebra version: ID: c00bfddbf0f3608340d61298acc61dafb167f9b2 Using ICU

Had to do it in 6-8 months, aside from some data cleaning tasks



```
addbiblio.pl 🗵
```

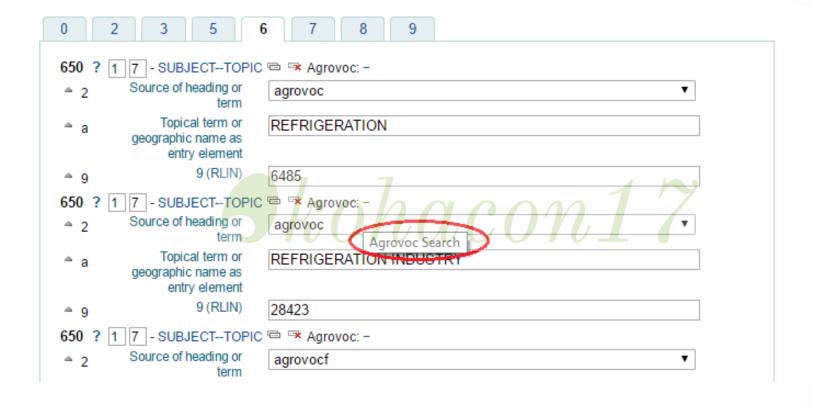
```
push(@loop data, \%tag data );
     } # foreach $field end
# if breeding is empty
else {
   my @subfields data;
   foreach my $subfield ( sort( keys %{ $tagslib->{$tag} } ) ) {
        next if ( length $subfield != 1 );
        next
          if ( ( $tagslib->{$tag}->{$subfield}->{hidden} <= -5 )</pre>
            or (\frac{1}{2} - \frac{1}{2} - \frac{1}{2} - \frac{1}{2} = 4)
          and not ( $subfield eq "9" and
                    exists($tagslib->{$tag}->{'a'}->{authtypecode}) and
                    defined($tagslib->{$tag}->{'a'}->{authtypecode}) and
                    $tagslib->{$tag}->{'a'}->{authtypecode} ne ""
               #check for visibility flag
               # if subfield is $9 in a field whose $a is authority-controlled,
               # always include in the form regardless of the hidden setting - bug 2206
        next
          if ( $tagslib->{$tag}->{$subfield}->{tab} ne $tabloop );
push (
            @subfields data,
            &create input(
                $tag, $subfield, '', $index tag, $tabloop, $record,
                $authorised values sth,$input
        );
    if ( $#subfields data >= 0 ) {
        my %tag data = (
                             => $tag,
            tag
                           => $index tag,
            index
                           => $tagslib->{$tag}->{lib},
            tag lib
            repeatable
                             => $tagslib->{$tag}->{repeatable},
```

```
AgrovocWSService.pm
package AgrovocWSService;
户# Generated by SOAP::Lite (v0.714) for Perl -- soaplite.com
 # Copyright (C) 2000-2006 Paul Kulchenko, Byrne Reese
 # -- generated at [Fri Jan 13 15:44:50 2012]
# -- generated from file:///home/cc/bloody agrovoc/ACSWWebService.WSDL
 my %methods = (
endpoint => 'http://agrovoc.mimos.my/ACSWWebservice/services/ACSWWebService'.
     soapaction => '',
     namespace => 'http://webservice.acsw.fao.org',
     parameters => [
     1, # end parameters
   }, # end getAgrovocDatabaseNames
 getConceptObject => {
     endpoint => 'http://agrovoc.mimos.my/ACSWWebservice/services/ACSWWebService',
     soapaction => '',
     namespace => 'http://webservice.acsw.fao.org',
     parameters => [
       SOAP::Data->new(name => 'clsName', type => 'xsd:string', attr =>
       SOAP::Data->new(name => 'ontology', type => 'xsd;string', attr => {}),
     1, # end parameters
   }, # end getConceptObject
 getTermcodeBvTerm => {
     endpoint => 'http://agrovoc.mimos.my/ACSWWebservice/services/ACSWWebService',
     soapaction => '',
     namespace => 'http://webservice.acsw.fao.org',
     parameters => [
      SOAP::Data->new(name => 'searchTerm', type => 'xsd:string', attr => {}),

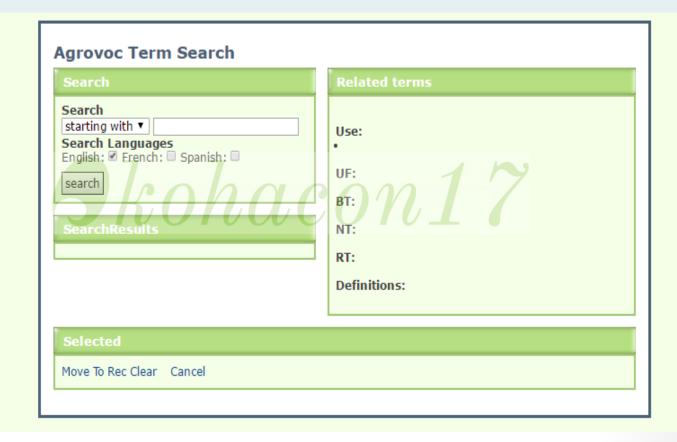
    # end parameters

   }, # end getTermcodeBvTerm
getConceptInfoByURI => {
     endpoint => 'http://agrovoc.mimos.my/ACSWWebservice/services/ACSWWebService'.
     soapaction => '',
     namespace => 'http://webservice.acsw.fao.org'.
     parameters => [
       SOAP::Data->new(name => 'ontologyName', type => 'xsd:string', attr => {}),
```

SOAP::Data->new(name => 'conceptURI', type => 'xsd:string', attr => {}),
SOAP::Data->new(name => 'format', type => 'xsd:string', attr => {}),



① unfao.koha-ptfs.eu:8080/cgi-bin/koha/agrovoc/agrovoc.pl?lang_english=%27EN%27&index=tag_650_157555214276



Agrovoc Term Search

Search Search starting with ▼ rice Search Languages English: ▼ French: □ Spanish: □ search

SearchResults

- rice bean select
- rice moth select
- rice weevil select
- ricefield aquaculture select
- Rice ragged stunt virus select
- rice straw select
- rice cake select
- rice husks select
- rice hulls select
- rice flour select
- rice polishings select
- rice value added product select
- rice (plant) select
- Rice tungro virus select
- Rice tungro bacilliform virus select
- Rice tungro spherical virus select
- Rice yellow mottle sobemovirus select
- Rice yellow mottle virus select
- rice fields select
- rice soil select
- rice select

Related terms

Use:

UF:

BT:

NT:

RT:

Definitions:



Selected

Move To Rec Clear Cancel

Agrovoc Term Search

Search Search starting with ▼ rice Search Languages English: ☑ French: □ Spanish: □ search

SearchResults

- rice bean select
- rice moth select
- rice weevil select
- ricefield aquaculture select
- Rice ragged stunt virus select
- rice straw select
- rice cake select
- rice husks select
- rice hulls select
- rice flour select
- rice polishings select
- rice value added product select
- rice (plant) select
- · Rice tungro virus select
- Rice tungro bacilliform virus select
- · Rice tungro spherical virus select
- Rice yellow mottle sobemovirus select
- Rice vellow mottle virus select
- rice fields select
- rice soil select
- rice select

Related terms

Use:

UF:

BT:

NT:

RT:

Definitions:

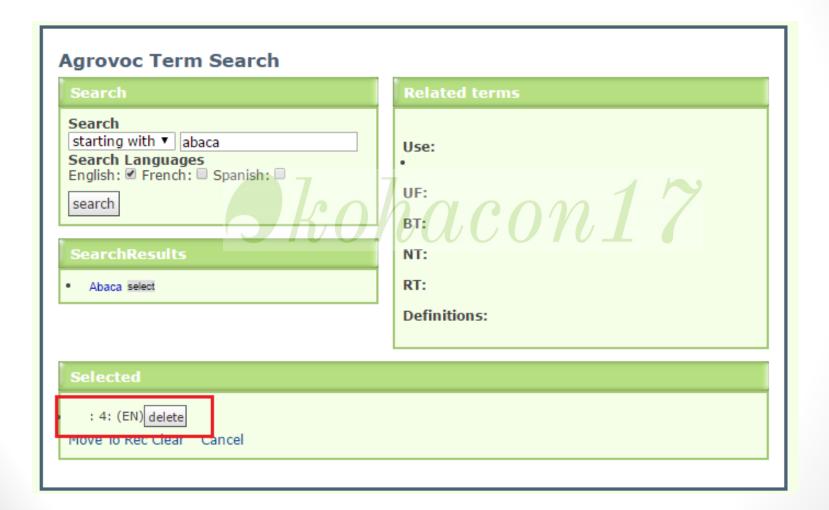


Selected

rice: 6599: (EN) delete

Move To Rec Clear Cancel

unfao.koha-ptfs.eu:8080/cgi-bin/koha/cataloguing/addbiblio.pl?biblionumber=4382&frameworkcode=&op=	
Show MARC tag documentation links	
■ Save ▼ Q Z39.50 search Cancel Change framework: Serials ▼	
0 2 3 5 6 7 8 9	
650 ? 1 7 - SUBJECTTOPIC □ □ Agrovoc: -	
Source of heading or term agrovoc ▼ Topical term or REFRIGERATION	
geographic name as entry element	
9 (RLIN) 6485 650 ? 1 7 - SUBJECTTOPIC ™ ™ Agrovoc: -	
≥ Source of heading or term agrovoc ▼	
Topical term or geographic name as entry element 9 (RLIN)	
650 ? 1 7 - SUBJECTTOPIC → Agrovoc: - Source of heading or agrovoc ▼	
Topical term or geographic name as	
entry element 9 (RLIN) 28423	
650 ? 1 7 - SUBJECTTOPIC □ → Agrovoc: -	
Source of heading or term agrovocf ▼	
Topical term or geographic name as entry element	
9 (RLIN) 428448	
650 ? 1 7 - SUBJECTTOPIC □ □ Agrovoc: - a grovocf ■ agrovocf ■	
Topical term or geographic name as entry element	
9 (RLIN) 6485	



Show MARC tag documentation links Save Q Z39.50 search Cancel Change framework: Serials • 650 ? 1 7 - SUBJECT--TOPIC □ 🗷 Agrovoc: -Source of heading or • a 2 agrovoc Topical term or REFRIGERATION geographic name as entry element 9 (RLIN) 6485 ė 9 650 ? 1 7 - SUBJECT--TOPIC - Agrovoc: -Source of heading or agrovoc Topical term or geographic name as entry element 9 (RLIN) △ 9 650 ? 1 7 - SUBJECT--TOPIC 🗁 🗫 Agrovoc: -Source of heading or • agrovoc Topical term or geographic name as entry element 9 (RLIN) ė 9 650 ? 1 7 - SUBJECT--TOPIC Topics -Source of heading or agrovoc • ė 2 Topical term or rice geographic name as entry element 9 (RLIN) ė 9 650 ? 1 7 - SUBJECT--TOPIC Topic Agrovoc: -Source of heading or • agrovoc 2 term Topical term or REFRIGERATION INDUSTRY geographic name as entry element 9 (RLIN) 28423

- AGROVOC can be accessed in various ways:
- Online Search & browse (LOD)
- Download Core (available in RDF and NT format)
- Download LOD (available in NT and NQ format)
- Web Services (LOD)SPARQL endpoint (LOD)

- AGROVOC is available in the following formats:
- SKOS RDF/XML (English only)
- MySQL (which is used for the AIMS AGROVOC search interface)
- Protégé DB (used internally by VocBench 1.3)
- OWL (the internal model used by VocBench 1.3)
- ISO2709 (used by WebAgris and other CS-ISIS systems)

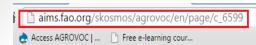
- Among the various ways and formats available, the following were explored:
- AGROVOC Core in RDF and NT format
- AGROVOC LOD in NT and NQ format
- A search on the SPARQL endpoint
- A search on the AGROVOC's web interface
- SKOS RDF/XML (English only)
- MySQL
- Protégé DB

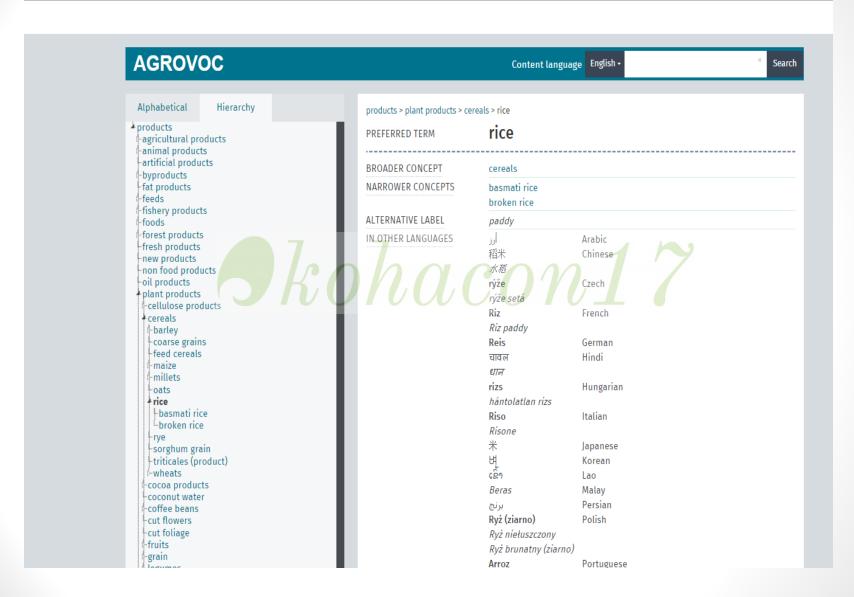
- The AGROVOC Core in RDF and NT formats as well as the AGROVOC LOD in NT and NQ formats was inspected but due to the file not easily opened by regular text editor and Notepad++ (a powerful software used to review the text file of records and use RegEx (regular expression) search and replace to prepare the data for data migration), it was shelved.
- SKOS RDF/XML was not considered because of its exclusivity to English terms. But the pattern was taken into account.

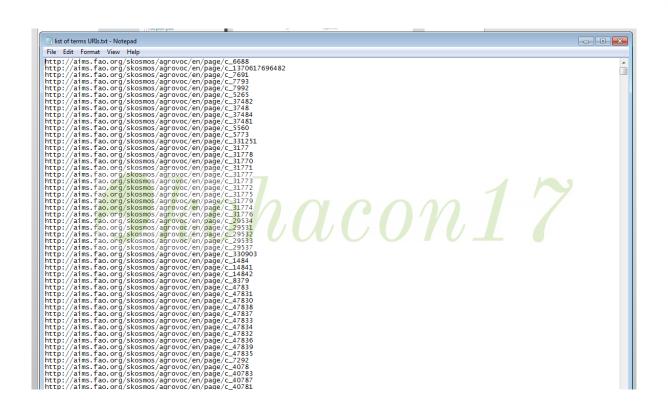
Bringing the Agrovoc terms into MARC format



1	s .	p	0
2	http://aims.fao.org/aos/agrovoc/c_6688>	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://www.w3.org/2004/02/skos/core#Concept
3	http://aims.fao.org/aos/agrovoc/c_1370617696482	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://www.w3.org/2004/02/skos/core#Concept
4	http://aims.fao.org/aos/agrovoc/c_7691	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://www.w3.org/2004/02/skos/core#Concept
5	http://aims.fao.org/aos/agrovoc/c_7793	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://www.w3.org/2004/02/skos/core#Concept
6	http://aims.fao.org/aos/agrovoc/c_7992	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://www.w3.org/2004/02/skos/core#Concept
7	http://aims.fao.org/aos/agrovoc/c_5265>	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://www.w3.org/2004/02/skos/core#Concept
8	http://aims.fao.org/aos/agrovoc/c_37482	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://www.w3.org/2004/02/skos/core#Concept
9	http://aims.fao.org/aos/agrovoc/c_3748>	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://www.w3.org/2004/02/skos/core#Concept
10	http://aims.fao.org/aos/agrovoc/c_37484	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://www.w3.org/2004/02/skos/core#Concept
11	http://aims.fao.org/aos/agrovoc/c_37481	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://www.w3.org/2004/02/skos/core#Concept
12	http://aims.fao.org/aos/agrovoc/c_5560	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://www.w3.org/2004/02/skos/core#Concept
13	http://aims.fao.org/aos/agrovoc/c_5773>	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://www.w3.org/2004/02/skos/core#Concept
14	http://aims.fao.org/aos/agrovoc/c_331251	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://www.w3.org/2004/02/skos/core#Concept
15	http://aims.fao.org/aos/agrovoc/c_3177>	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://www.w3.org/2004/02/skos/core#Concept
16	http://aims.fao.org/aos/agrovoc/c_31778>	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://www.w3.org/2004/02/skos/core#Concept
17	http://aims.fao.org/aos/agrovoc/c_31770>	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://www.w3.org/2004/02/skos/core#Concept
18	http://aims.fao.org/aos/agrovoc/c_31771	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://www.w3.org/2004/02/skos/core#Concept
19	http://aims.fao.org/aos/agrovoc/c 31777>	http://www.w3.org/1999/02/22-rdf-syntax-ns#tyne	http://www.w3.org/2004/02/skos/core#Concent







Download the list :P
 wget –i filenameofthelist.txt

Combine the downloaded files:
cat * > merged-filename

Some Regular Expressions Used

Regular Expression (Find)	Regular Expression (Replace)	
<pre>(<!DOCTYPE html>)(.*?)(property-table">?)</pre>	\r\n <html>\r\n\3</html>	
(RDF/XML)(.*?)()(.*?)(?)	\1\2\3\5	
<pre></pre>	<pre></pre>	
()\n(.*?)(">)(.*?)(?)	\1\4\5	
()(PREFERRED	<td< th=""></td<>	
TERM)	class="preflabel- desc" />\2	
	</span 	
(<td <="" class="preflabel-desc" th=""><th><pre><prefterm>\4</prefterm></pre></th></td>	<th><pre><prefterm>\4</prefterm></pre></th>	<pre><prefterm>\4</prefterm></pre>
/>)PREFERRED	fterm>	
TERM(.*?)("prefLabel">?)(.*?)		
(?).*?"preflabel-spacer"?(.*?)(?)		

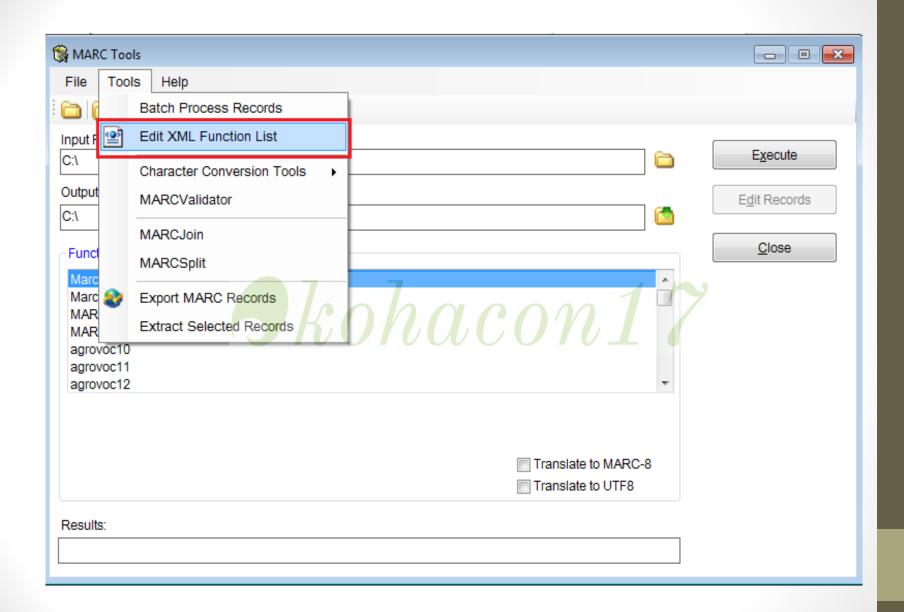
Some Regular Expressions Used (cont.)

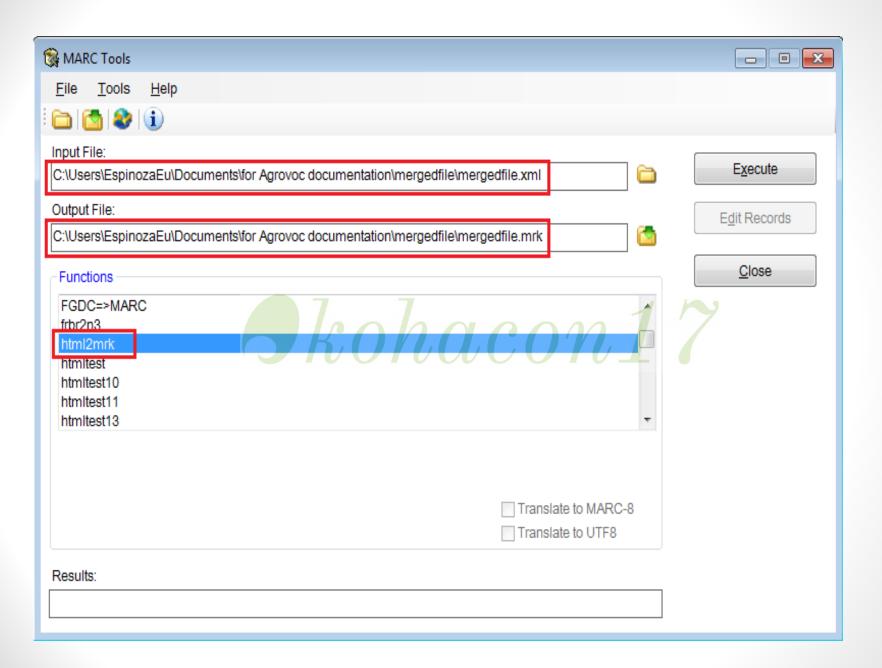
NARROWER <narrow class="versal CONCEPTS property-click" title="Narrower concepts.">NARROW ER CONCEPTS</narrow> class="versal property-click" title="Alternative terms for <span <altlabel class="versal concept.">ALTERNATIVE LABEL property-click" title="Alternative terms for the concept.">ALTERNATIV E LABEL</altlabel> <span class="versal property-click" title="Terms for the concept in other</pre> <otherlang languages." >IN OTHER LANGUAGES class="versal propertytitle="Other click" language" >IN OTHER

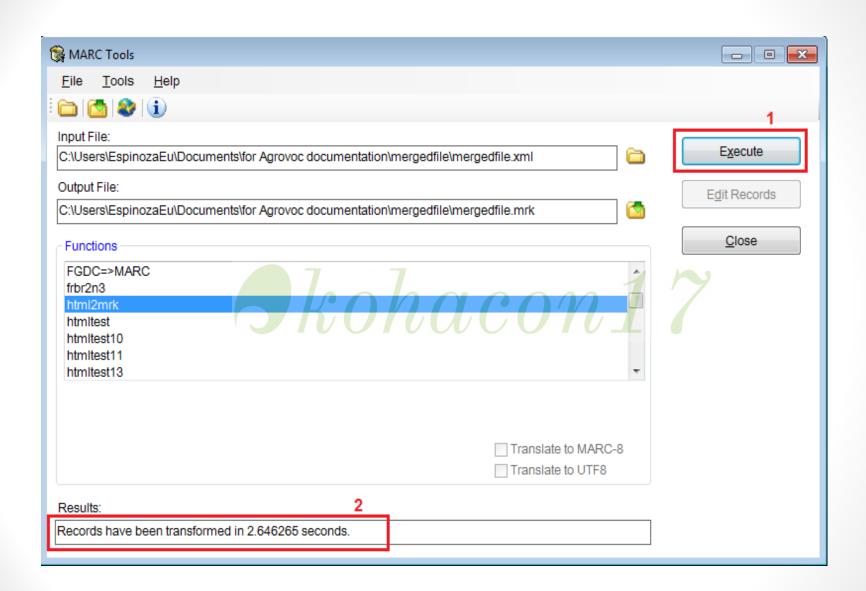
LANGUAGES</otherla

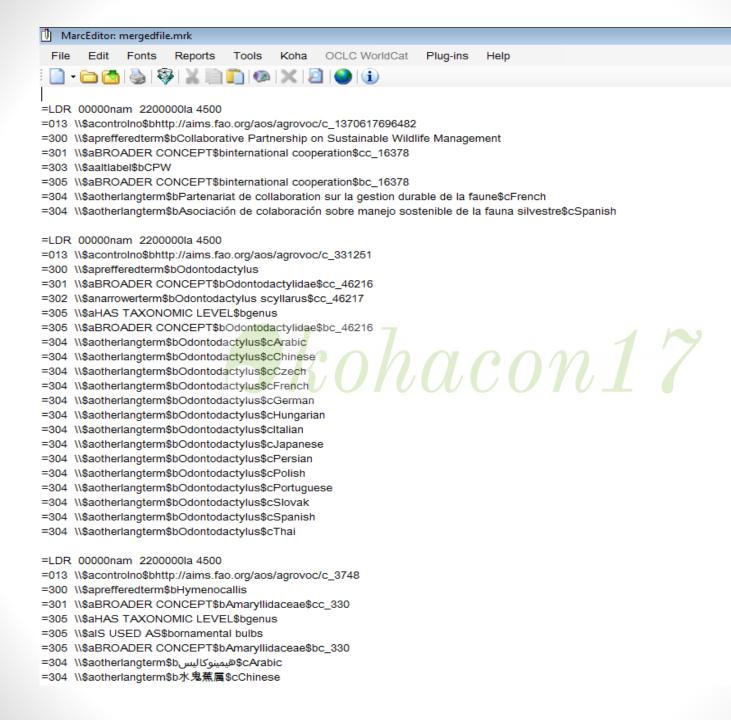
ng>

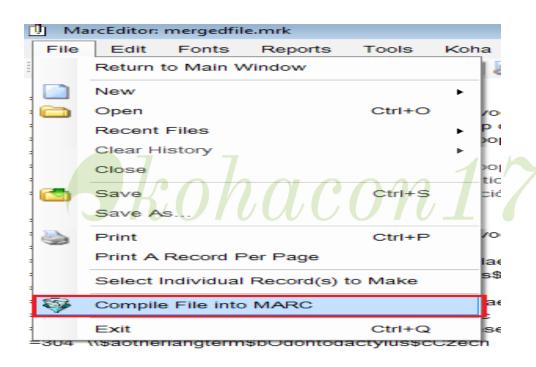












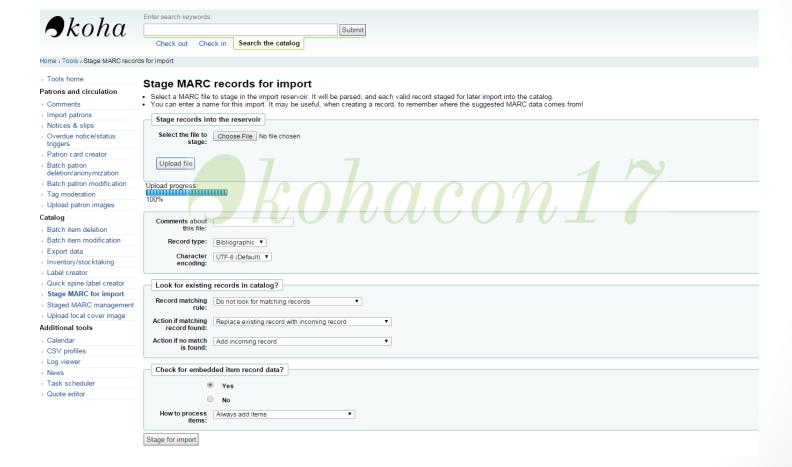
Notes on the MARC fields used

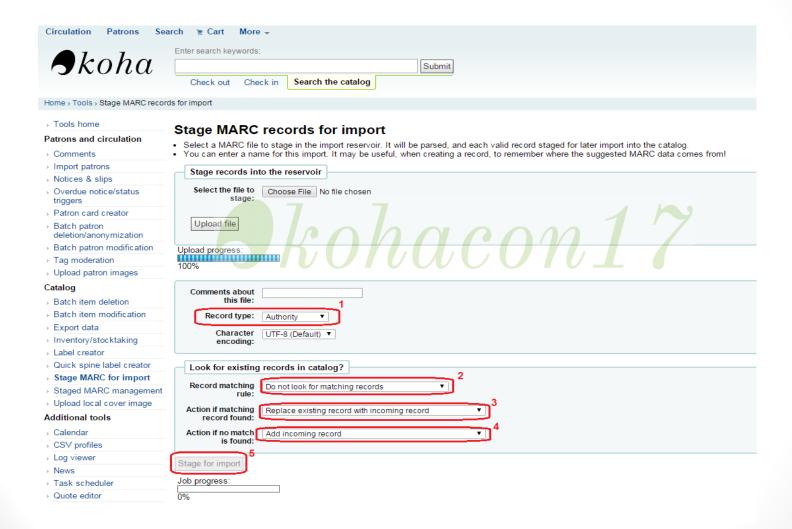
AGROVOC	MARC fields/subfields pattern	Notes
Preferred Term	150\$a	
Broader Term	550\$abroaderterm\$wg	\$w=g e.g. =550 \\\$wg\$aanimal meals
Narrower term	550\$anarrowerterm\$wh	\$w=h e.g. =550 \\\$wh\$aCostae
Notes	667\$a KONA	Definition and all other labels here (e.g. IS USED AS, HAS TAXONOMIC LEVEL). Example: =667 \\\$aIS USED AS: timber trees
Other Language Term	450\$\$wr\$iotherlangterm\$iOther Language Terms (the Language)\$atheterm	Example: =450 \\\$wr\$iOther Language Terms (Arabic)\$a شوح
Control No.	001	
Alternative Label	450 \0\$athealternativelabelterm	=450 \0\$apaddy

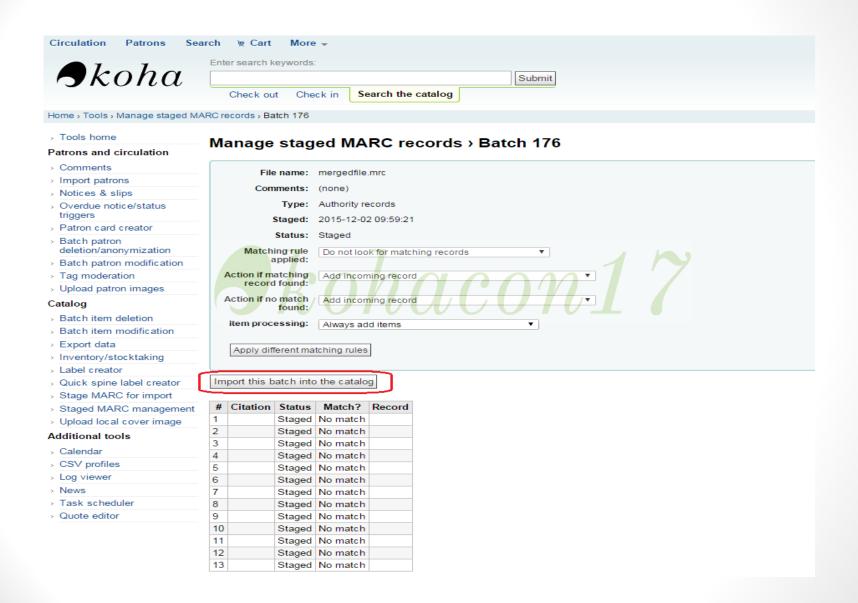


Additional tools

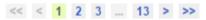
Export data
 Inventory/stocktaking
 Label creator
 Quick spine label creator
 Stage MARC for import
 Staged MARC management
 Upload local cover image







Autnority search results



Results 1 to 20 of 257

Summary		Used in		
19th Session of the International Rice Commission eng	Details	0 records(s)	Edit Delet	
aerobic rice used for/see from: riz aérobic used for/see from: arroz aeróbico see also: Oryza (Broader heading)	Details	0 records(s)	Edit Delet	
Africa Rice Center, Cotonou (Benin) eng	Details	0 records(s)	Edit Delet	
All-Russian Rice Research Institute, Krasnodar (Russian Federation) eng	Details	1 records(s)	Edit	
Asia Rice Foundation, Bangkok (Thailand) eng	Details	0 records(s)	Edit Delet	
Asian Regional Rice-Fish Farming Research and Development Workshop eng	Details	0 records(s)	Edit Delet	
Asian Regional Workshop on Sustainable Development of Irrigation and Drainage for Rice Paddy Fields eng	Details	0 records(s)	Edit Delet	
ASSOCIATION OF RICE RESEARCH WORKERS	Details	1 records(s)	Edit	
Associazione per la Ricerca e lo Studio sui Sistemi Alimentari e Forestali nelle Regioni Emergenti, Viterbo (Italy) ita	Details	0 records(s)	Edit Delet	
Atti - Istituto Superiore per la Protezione e la Ricerca Ambientale (Italy) ita	Details	1 records(s)	Edit	
basmati rice used for/see from: rýže basmati used for/see from: Basmatireis used for/see from: Riso Basmati used for/see from: 바스마티쌀 used for/see from: Basmati pirinci see also: rice (Broader heading)	Details	0 records(s)	Edit Delet	
broken rice used for/see from: lámaná rýže used for/see from: Bruchreis used for/see from: Rotture di riso used for/see from: 쇄미 used for/see from: kırık pirinç see also: rice (Broader heading)	Details	0 records(s)	Edit Delet	
Bulletin - International Rice Research Institute (Tanzania) eng	Details	1	Edit	

		recorda(a)	Delete
Regional Workshop on the Future of Large Rice-based Irrigation Systems in Southeast As	ia Details		Edit
eng		records(s)	
ice	Details		Edit
used for/see from: ارز		records(s)	
used for/see from: 稻米			
used for/see from: rýže			
used for/see from: Riz			
used for/see from: Reis			
used for/see from: चावल			
used for/see from: rizs			
used for/see from: Riso			
used for/see from: 米			
used for/see from: 벼			
used for/see from: เล็ก			
used for/see from: برنج			
used for/see from: Ryż (ziarno)			
used for/see from: Arroz			
used for/see from: рис (зерно)			
used for/see from: ryža			
used for/see from: Arroz			
used for/see from: ข้าว used for/see from: pirinç used for/see from: paddy see a/so: cereals (Broader heading)		1 7	
used for/see from: pirinç			
used for/see from: paddy			
see also: cereals (Broader heading)	4		
see also: basmati rice (Narrower heading)			
see also: broken rice (Narrower heading)			
RICE	Details		Edit
		records(s)	
Rice Blast Workshop eng	Details	_	Edit
		records(s)	Delete
rice cake	Details	0	Edit
used for/see from: 大米饼		records(s)	Delete
used for/see from: rýžové výlisky			
used for/see from: Reiskuchen			
used for/see from: धान की खली			
used for/see from: rizspopgácsa			
used for/see from: Panello di riso			
used for/see from: 餅			
used for/see from: 떡			
used for/see from: ເຂົາໜົນເຄັກເຮັດດ້ວ ນ ເຂົ້າ			
used for/see from: کیك برنج			
used for/see from: рисовые лепешки			
used for/see from: ryžové výlisky			
used for/see from: ข้าวเหนียวปั่น			

Maraming salamat!

Email ad: eugenegf@yahoo.com