

EDIT - task #9364

Cleanup authors with multiple & and protected names and teams

12/22/2020 03:17 PM - Andreas Müller

Status:	Closed	Start date:	
Priority:	Highest	Due date:	
Assignee:	Andreas Müller	% Done:	100%
Category:	data	Estimated time:	0:00 hour
Target version:			
Severity:	normal		
Description			
(high priority only to first issue as it is related to #2200)			
In context of #2200 I already checked names with >1 "&" in nameCache. These are cleanedup. But there are open issues with TeamOrPerson.nomenclaturalTitle:			
Campanula: 43 Diptera: 77 FdAC: 92 Palmae: 266 !			
Guianas: 3(?) => unparsed collectors with 2 collectors having the same family name FM: 59(?) => remaining issues are all MAN with some second author team with unclear semantics needs to be fixed during next import			
<pre>SELECT ab.DTYPE, ab.id, ab.titleCache, ab.nomenclaturaltitle, ab.protectedtitlecache, ab.protected nomenclaturaltitlecache FROM AgentBase ab WHERE ab.nomenclaturaltitle LIKE '%&%&%%' OR ab.titleCache LIKE '%&%&%%';</pre>			
In some databases there are many names with protected caches which may lead to problems when searching for names.			
<pre>SELECT tn.id, tn.titleCache, tn.nameCache, tn.protectedTitleCache, tn.protectedNameCache, tn.prote ctedFullTitleCache, tn.protectedAuthorshipCache , tn.* FROM TaxonName tn WHERE tn.protectednamecache = 1 OR tn.protectedtitlecache = 1 OR tn.protectedFullTitleCache = 1 OR tn.protectedAuthorshipCache ;</pre>			
Especially interesting might be those having an "&" in the protected nameCache which indicates that there is authorship in the nameCache which should not happen.			
Also many Teams have a protected nomenclatural title which is in most cases not necessary.			
<pre>SELECT ab.id, ab.DTYPE, ab.titleCache, ab.nomenclaturaltitle, ab.protectedtitlecache, ab.protected nomenclaturaltitlecache, ab.* FROM AgentBase ab WHERE -- ab.protectedtitlecache = 1 OR ab.protectednomenclaturaltitlecache = 1 ORDER BY ab.nomenclaturaltitle</pre>			
However, in most cases this is not critical.			
Related issues:			
Related to EDIT - task #9648: Remove duplicated extensions		New	
Related to EDIT - task #9650: Update all titleCaches after upgrade of cache s...		In Progress	
Copied to EDIT - task #9658: Cleanup authors with protected names and authors		New	

History

#1 - 12/22/2020 03:56 PM - Andreas Müller

Protected Names (still without fulltitlecache and authorshipcache):

Bromeliaceae: 584
Cichorieae: 476
Cuba: 567
Campanula: 28
Caryo_amaranth: 18
Nepenthes: 8
caryo_genrea: 1
caryo_spp: 268
Greece: 3
Palmae: 66
casearia: 7
COL: >500000
Corvidae: 87
Cyprus: 97 (mostly hybrids)
Diptera: 3
E+M: 801 (mostly hybrids)
E+M cauc: 4
Fl dAfrique Central: 5190
myristicacea: 1
Phycobank: 245 => WHK
Salvador: 3 (ok, highest taxa, but ask Walter)

Guianas: 1487 !!
Flora Malesiana: 535
FM-clean: 2
FM propective: 84
Gabon: 90

RL: All, but many in Animalia, GermanSL, Moose, Standardliste + Plantae(?)

Globis: 7483
Edapho: 4
PiB: All, but most in Ants
Vibrant: 24165

#2 - 12/22/2020 07:29 PM - Andreas Müller

Protected Teams(nomenclaturalTitle):

AlgaTerra: >14000
Cichorieae: 8487
Cuba: 14
Campanula: 1171
Caryo_amaranth: 113
Nepenthes: 65
caryo_genrea: 128
caryo_spp: 126
Palmae: 12592
casearia: 1
COL: >1900000
Corvidae: 487
Cyprus: 1
Diptera: 365
E+M: 16
E+M cauc: 2
Fl dAfrique Central: 25925
Phycobank: 270
Salvador: 19
Mexiko: 101

Guianas: 14465
Flora Malesiana: >121000
FM-clean: 16
FM propective: 55
Gabon: 12940

RL: All, but many in Armeria, Moose, Standardliste + Plantae(?)

Globis: 7483
PiB: few, mostly in chenopodium and spiders
Vibrant: 41202

#3 - 12/22/2020 08:20 PM - Andreas Müller

- Subject changed from *Cleanup protected names and teams* to *Cleanup authors with multiple & and protected names and teams*
- Description updated
- Target version changed from *Unassigned CDM tickets* to *Release 5.19*

#4 - 12/22/2020 08:37 PM - Andreas Müller

- Description updated

#5 - 12/22/2020 08:37 PM - Andreas Müller

- Status changed from *New* to *In Progress*
- Priority changed from *New* to *Highest*
- % Done changed from *0* to *10*

#6 - 12/22/2020 08:38 PM - Andreas Müller

- Description updated

#7 - 01/27/2021 04:10 PM - Andreas Müller

- Target version changed from *Release 5.19* to *Release 5.21*

#8 - 03/10/2021 02:36 PM - Andreas Müller

- Target version changed from *Release 5.21* to *Release 5.22*

#9 - 04/16/2021 10:21 PM - Andreas Müller

- Target version changed from *Release 5.22* to *Release 5.25*

#10 - 05/14/2021 12:36 AM - Andreas Müller

- Tags set to *formatting*

#11 - 06/03/2021 01:37 PM - Andreas Müller

- Related to task #9648: *Remove duplicated extensions added*

#12 - 06/08/2021 11:11 AM - Andreas Müller

- Description updated

#13 - 06/08/2021 01:15 PM - Andreas Müller

- Description updated
- % Done changed from *10* to *20*

#14 - 06/08/2021 01:20 PM - Andreas Müller

- Description updated

#15 - 06/08/2021 01:46 PM - Andreas Müller

- Description updated

#16 - 06/08/2021 05:42 PM - Andreas Müller

- Copied to task #9658: *Cleanup authors with protected names and authors added*

#17 - 06/08/2021 05:44 PM - Andreas Müller

- Related to task #9650: *Update all titleCaches after upgrade of cache strategies and formatters added*

#18 - 06/08/2021 05:46 PM - Andreas Müller

- Status changed from *In Progress* to *Closed*
- Target version deleted (*Release 5.25*)
- % Done changed from *20* to *100*

protected (title)caches moved to new ticket [#9658](#).

All "multiple &" solved.

Potentially dangerous protected caches will be handled in [#9650](#)