EDIT - feature request #9085

Improve deduplication of parsed names

06/19/2020 11:16 PM - Andreas Müller

Status:	Closed		**Start date:**	
Priority:	Highest		**Due date:**	
Assignee:	Andreas Müller		**% Done:**	100%
Category:	cdmlib		**Estimated time:**	0:00 hour
Target version:	Release 5.16			
Severity:	normal			

Description

Special methods have been implemented during E+M import for improved deduplication (references+authors) of parsed names.

There are still some open issues before this deduplication can be used in other context like TaxEditor name parsing.

- improve MatchStrategyFactory
- switch order of match parameters (in E+M the order was [full name, parsed name] but should be [parsed name, full name] to allow retrieving matches from the database
- improve handling of components like TimePeriod, LSID, etc.
- write tests on persistence level => #9157
- ...

Related issues:

Related to EDIT - feature request #9078: Handle name parsing and deduplicatio...	**Closed**
Related to EDIT - feature request #9022: Implement lifespan for Person Detail...	**Closed**
Related to EDIT - bug #9081: Handle empty Partials correctly	**Closed**
Related to EDIT - bug #7829: Improve deduplication of parsed names and refere...	**Closed**
Related to EDIT - bug #6427: Improve Matching for Reference.title and abbrevT...	**New**
Related to EDIT - task #6417: [DISCUSS] Need and properties of a Nomenclatura...	**New**
Has duplicate EDIT - bug #1119: [PARSER] Duplicate inreferences created durin...	**Duplicate**
Copied to EDIT - bug #9157: Further improve deduplication of names	**In Progress**

Associated revisions

Revision 72eb424b - 06/17/2020 01:20 AM - Andreas Müller

ref #9078 first server side implementation for parsing and deduplication of names

Revision c078cfb8 - 06/19/2020 11:26 PM - Andreas Müller

ref #9085 improvements to MatchStrategyFactory and switch order of match parameters, some improvements to TimePeriod matching

Revision d285f18d - 06/20/2020 09:31 AM - Andreas Müller

ref #9085 improve X_OR_FIRST_NULL handling

Revision a9bd2165 - 06/23/2020 08:36 AM - Andreas Müller

ref #9078, ref #9085 improve deduplication of teams nom.ref. authors

Revision aab00a6d - 06/23/2020 09:06 AM - Andreas Müller

ref #9078, ref #9085 fix deduplication of nom.ref. author

History

#1 - 06/19/2020 11:16 PM - Andreas Müller

- Related to feature request #9078: Handle name parsing and deduplication on server side added

#2 - 06/19/2020 11:17 PM - Andreas Müller

- Related to feature request #9022: Implement lifespan for Person Details View added

#3 - 06/19/2020 11:18 PM - Andreas Müller

- Related to bug #9081: Handle empty Partials correctly added

#4 - 06/20/2020 08:34 AM - Andreas Müller

- Status changed from New to In Progress

- Priority changed from New to Highest

- % Done changed from 0 to 30

#5 - 06/23/2020 01:47 PM - Andreas Müller

- Target version changed from Release 5.15 to Release 5.18

#6 - 07/17/2020 09:12 AM - Andreas Müller

- Copied to bug #9157: Further improve deduplication of names added

#7 - 07/17/2020 09:13 AM - Andreas Müller

- Description updated

- Status changed from In Progress to Closed

- Target version changed from Release 5.18 to Release 5.16

- % Done changed from 30 to 100

Open issues moved to [#9157](#9157)

#8 - 01/20/2021 06:23 PM - Andreas Müller

- Has duplicate bug #1119: [PARSER] Duplicate inreferences created during parsing are NOT merged when data gets saved added

#9 - 01/20/2021 07:24 PM - Andreas Müller

- Related to bug #7829: Improve deduplication of parsed names and references added

#10 - 05/19/2021 11:06 AM - Andreas Müller

- Related to bug #6427: Improve Matching for Reference.title and abbrevTitle added

#11 - 05/19/2021 11:24 AM - Andreas Müller

- Related to task #6417: [DISCUSS] Need and properties of a NomenclaturalReference class added