EDIT - task #1404

 task # 4338 (In Progress): [MASTER] Handle abbreviated title and protected cache correctly in TaxEditor

References cache strategy should always render bibliographic references

01/15/2010 05:44 PM - Niels Hoffmann

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

Description

This is a longer story.

The first implementation of the ReferenceFilteredSelection and the underlying getUuidAndTitleCache() was to return the "titleCache". This approach however turned out to be problematic, because titleCache also contains the author which is not intuitive when searching for a reference.

Therefore we added an option to let getUuidAndTitleCache() to either return "title" or "titleCache", because "title" would not start with the author as requested. The problem with "title" is, that it is the title of the reference only (e.g. the name of the book) and does not contain crucial information such as volume, year s.o.

It seems like we would need another cache in ReferenceBase that contains the information in "titleCache" but without the author. Is there any other approach?

History

#1 - 01/16/2010 12:12 AM - Andreas Müller

Yes this is a problem. I have no concrete idea how to solve it but here is just what came into my mind when I thought about it for a moment:

- Having a second cache stored in the db is not so nice therefore I am looking for an other solution.

- When filling the filteredSelection can't we cut of the author part and add it to the end?

The later could either be done by

- searching for the first substring that starts like the title, presuming that this will be the place where the title really starts, taking away the preceding, adding it to the end and do some extra work to handle the inbetween characters like ',', '-' etc

- Define a new cacheStrategy and call this strategy for all those records in the selection that match via title. If this is not performant for large result set a more sophisticated algorithm could be found that works asynchronously or handles only those record that are of highest interest.

I don't know if this is somehow possible to implement for the filteredSelections as they are quite generic as far as I understand them. So it's just an idea ...

#2 - 01/18/2010 01:30 PM - Niels Hoffmann

Just an explanation: FilteredSelectionDialogs is an eclipse technique. We feed it with all strings that need filtering and do not have to bother with the internal logic and implementation. Th filtering itself has a very high performance and we hit the database only once.

Since we always get the full set Proposed solutions are not really suitable. On the fly alteration or calculation of thousands of strings will not be performant and also rather redundant (this is where you would introduce a cache).

Could you please explain the problems with having another cache field?

The current titleCache (strategy) for references seems to be highly optimized for the use as nomenclatural reference and in generating the fullTitleCache of a taxon name. I think we are still missing a lot here.

#3 - 01/18/2010 10:13 PM - Andreas Müller

Another cache field makes the model more complicated and we try to keep the model as simple as possible. It also blows up the amount of data that has to be stored without giving any extra information. That does not mean that a new cache field should be avoided in any case. But we should check carefully if it is really needed.

Some questions that come into my mind:

- Is this the only new cache we need or are there other caches we need for other use-cases.

- Shouldn't we use the title cache for your purposes by using another cacheStrategy for the title cache as the titleCache is the cache you are usually using for a quick search and therefore it is stored in the database. As you say the existing title cache is used more to cite a reference or for creating the fullTitleCache of a taxon name. This can more easily be done on the fly so maybe we do not need the cache starting with the author stored in the db. Or do you want to allow both searches?

- Are you sure that easy alteration of thousands of strings is really not performant? (We will gain performance here but will loose it when storing new or updated data - this may be an issue for imports).

- We should at least check with Ben and maybe with the cdm-edit list to know what others think about this issue.

I agree that caches are one of the most difficult issues in the CDM (we will realize this once we will start with real multi-linguality) and there will always be a trade-off between the different needs

#4 - 01/19/2010 10:22 AM - Niels Hoffmann

Replying to a.mueller:

> Shouldn't we use the title cache for your purposes by using another cacheStrategy for the title cache as the titleCache is the cache you are usually using for a quick search and therefore it is stored in the database. As you say the existing title cache is used more to cite a reference or for creating the fullTitleCache of a taxon name. This can more easily be done on the fly so maybe we do not need the cache starting with the author stored in the db. Or do you want to allow both searches?

Are you suggesting here to change the titleCache field to be more generic and generate the fullTitleCache on the fly?

This would be my desired way to solve the problem. Should we start discussing it on cdm-edit?

#5 - 01/19/2010 02:38 PM - Andreas Müller

Replying to n.hoffmann:

> Replying to a.mueller:
>
> > Shouldn't we use the title cache for your purposes by using another cacheStrategy for the title cache as the titleCache is the cache you are usually using for a quick search and therefore it is stored in the database. As you say the existing title cache is used more to cite a reference or for creating the fullTitleCache of a taxon name. This can more easily be done on the fly so maybe we do not need the cache starting with the author stored in the db. Or do you want to allow both searches?
>
> Are you suggesting here to change the titleCache field to be more generic and generate the fullTitleCache on the fly?
>
> This would be my desired way to solve the problem. Should we start discussing it on cdm-edit?

I don't know if it is more generic, I think it is just different. But if we use a cache strategy that creates caches that start with the title, include volume etc, and add the author to the end, and use this for the title cache generation your problem may be solved. We even do not have to discuss it on the cdm-edit list because we do not change the model at all.

But we have to discuss it with the portal people and with yourself because at other places we already got used to the old cachestrategy. At all places where the old cache strategy is used we have to declare explicitly to use it instead of using it as a default as we do now.

Anyway it would be interesting to know if Ben has a similar problem or at least a solution so I think contacting him is a good idea.

I'am thinking about storing the guid of the cache strategy that has been used for creating the caches anyway. This may also solve some other problems. But this should be discussed when I am back in office.

How urgent is your problem? A model change will take time because everyone hates model changes !

#6 - 03/26/2010 06:21 PM - Niels Hoffmann

- Target version set to SPRINT Cichorieae Portal 2

- Priority changed from Priority13 to Highest

#7 - 05/18/2010 07:52 PM - Niels Hoffmann

- Priority changed from Highest to Priority14

#8 - 06/08/2010 09:30 PM - Niels Hoffmann

- Priority changed from Priority14 to Priority08

- Tracker changed from bug to feature request

We solved te problem by altering the title cache on the fly, replacing the author to the end of the string. Although this does not solve the problem completely it does solve the initial problem, where the references title cache starts with an author which was not very intuitive. We will have to see how we can run on this solution.

Changing this to be a feature request and decreasing priority.

#9 - 06/12/2010 02:26 PM - Niels Hoffmann

- Priority changed from Priority08 to Highest

- Subject changed from ReferenceFilteredSelection should show more that the references title to References cache strategy should always render bibliographic references

- Tracker changed from feature request to task

The problem is not solved at all. Any references that are part of other references, an article for example will not be found, because the title of the article does not go into the nomenclatural citation.

We agreed on changing the cache strategy to generate bibliographical references as a nomenclatural reference will be needed in the names full title cache only.

#10 - 09/07/2010 05:31 PM - Niels Hoffmann

- Priority changed from Highest to Lowest

I think the cache strategy was changed. Decreasing priority. Please close if you know that this is solved.

#11 - 01/05/2015 02:19 PM - Andreas Müller

- Assignee changed from Niels Hoffmann to Andreas Kohlbecker

could you please check if this is still an open issue?

#12 - 01/06/2015 09:17 AM - Andreas Kohlbecker

- Assignee changed from Andreas Kohlbecker to Andreas Müller

- Target version changed from cdm_dataportal - Next Major Release to TaxEditor Next Major Release

this ticket is related to the taxeditor, so i am changing the milestone accordingly.

From reading the comments in this ticket it seems to me as if you, Andreas Müller, would be the better candidate to review this ticket since you discussed this issue with Niels in the past.

#13 - 05/26/2016 06:54 PM - Andreas Müller

- Target version changed from TaxEditor Next Major Release to Release 4.2

- Priority changed from Lowest to Priority14

#14 - 07/06/2016 03:23 PM - Andreas Müller

- Target version changed from Release 4.2 to Release 4.3

#15 - 09/23/2016 03:35 PM - Andreas Müller

- Target version changed from Release 4.3 to Release 4.4

#16 - 12/01/2016 02:05 PM - Andreas Müller

- Target version changed from Release 4.4 to Release 4.5

#17 - 01/19/2017 01:36 PM - Andreas Müller

- Target version changed from Release 4.5 to Release 4.6

#18 - 03/11/2017 02:39 AM - Andreas Müller

- Description updated

- Priority changed from Priority14 to Highest

#19 - 12/18/2020 10:48 PM - Andreas Müller

- Target version changed from Release 4.6 to Release 5.19

#20 - 12/18/2020 10:48 PM - Andreas Müller

- Private changed from Yes to No

#21 - 01/27/2021 03:43 PM - Andreas Müller

- Target version changed from Release 5.19 to Release 5.21

#22 - 03/10/2021 02:34 PM - Andreas Müller

- Target version changed from Release 5.21 to Release 5.22

#23 - 04/16/2021 09:51 PM - Andreas Müller

- Target version changed from Release 5.22 to Release 5.25

#24 - 05/14/2021 12:38 AM - Andreas Müller

- Tags set to formatting

#25 - 05/15/2021 01:27 AM - Andreas Müller

- Status changed from New to Rejected

- Target version deleted (Release 5.25)

This is not relevant anymore. Current implementation works for users. Also there are 2 caches now for abbrev and full title.

#26 - 05/15/2021 01:36 AM - Andreas Müller

- % Done changed from 0 to 100