

EDIT - bug #7200

México Distrito Federal spelling and ignore accents in area search

01/18/2018 10:04 AM - Katja Luther

Status:	Worksforme	Start date:	
Priority:	Priority14	Due date:	
Assignee:	Katja Luther	% Done:	100%
Category:	taxeditor	Estimated time:	0:00 hour
Target version:		Found in Version:	
Severity:	normal		
Description			
NK:			
es gibt eine problematische Inkonsistenz bei den level 4 areas in Mexico:			
Sucht man im Area Wizard nach "Mexico" erhält man alles mögliche, auch Mexico Distrito Federal, nicht aber den Bundesstaat Mexico "Mexico State", denn dieser ist als "México State" aufgeführt (also mit Akzent auf dem e, so auch bei TDWG aufgeführt) und wird auch nur gefunden, wenn man den Akzent mitangibt. Ich hatte zuerst angenommen Mexico state wäre vergessen worden. "México Distrito Federal" hingegen müsste korrekterweise mit Akzent geschrieben werden (so auch bei TDWG), ist es aber nicht.			
Ich fände es sinnvoll, wenn auch México Distrito Federal mit Akzent geschrieben würde, aber die Suche sollte unbedingt die Akzente ignorieren.			
Related issues:			
Copied to EDIT - bug #9830: México Distrito Federal needs accent			Closed

History

#1 - 01/18/2018 01:09 PM - Andreas Müller

Be aware that there are a couple of other cases with accents and similar. So we need a general solution for the search.

Some solutions could be to use lucene search (therefore I add AK). But maybe also hibernate allows according search. I even wonder that the version with accent is not found because as far as I know SQL does not distinguish accents. Is there an Java equal after the SQL search? (I have the same problem but other direction for deduplication, where I need to distinguish but SQL "order by" does not).

Also we will need an update script.

#2 - 01/18/2018 01:10 PM - Andreas Müller

- Description updated

#3 - 01/18/2018 02:13 PM - Andreas Kohlbecker

These types of problems are usually rather a question of the collation being used in the database.

See for example this thread <https://stackoverflow.com/questions/28863402/mysql-diacritic-insensitive-search-arabic#28891336>

this simple tests:

```
select 'México' = 'Mexico' COLLATE utf8_unicode_ci
```

gives back a 1 which means MySQL considers these strings being equal.

Maybe the collation in the database in question is not set correctly? It should be utf8_unicode_ci

#4 - 10/26/2021 05:45 PM - Andreas Müller

Andreas Kohlbecker wrote:

These types of problems are usually rather a question of the collation being used in the database.

See for example this thread <https://stackoverflow.com/questions/28863402/mysql-diacritic-insensitive-search-arabic#28891336>

this simple tests:

```
select 'México' = 'Mexico' COLLATE utf8_unicode_ci
```

gives back a 1 which means MySQL considers these strings being equal.

Maybe the collation in the database in question is not set correctly? It should be utf8_unicode_ci

This IMO is not a DB/SQL issue (select 'México' = 'Mexico' returns also 1 on cichorieae so they both are considered equal) but an issue of the later equal testing in java.

#5 - 10/26/2021 05:55 PM - Andreas Müller

- Target version changed from Unassigned CDM tickets to Release 5.45

How to implement java side accent insensitive compare or contains is e.g. explained here:

<https://stackoverflow.com/questions/28833797/compare-strings-ignoring-accented-characters> and <https://stackoverflow.com/questions/8745660/contains-with-collator/8745778#8745778>

So maybe we can implement this for all searches in 5.29?

#6 - 10/26/2021 09:59 PM - Andreas Müller

- Status changed from New to Resolved

- Assignee changed from Katja Luther to Norbert Kilian

- Target version changed from Release 5.45 to Release 5.28

This seems to be fixed in the meanwhile. At least searching for "Mexico" also returns "México State". Norbert, can you please verify?

#7 - 10/27/2021 09:11 AM - Andreas Müller

- Copied to bug #9830: México Distrito Federal needs accent added

#8 - 10/27/2021 09:11 AM - Andreas Müller

NK:

Ja, searching for "Mexico" also returns "México State"

Aber es heißt immer noch „Mexico Distrito Federal“ statt „México Distrito Federal“ (mit Akzent)

#9 - 10/27/2021 09:13 AM - Andreas Müller

- Status changed from Resolved to Worksforme

- Assignee changed from Norbert Kilian to Katja Luther

- Target version deleted (Release 5.28)

- % Done changed from 0 to 100

I split the ticket. Correct spelling for „México Distrito Federal“ is now handled in [#9830](#) as this needs a model change update script.