EDIT - task #9648

Remove duplicated extensions

06/03/2021 01:37 PM - Andreas Müller

Status:	New	**Start date:**	
Priority:	New	**Due date:**	
Assignee:	Andreas Müller	**% Done:**	0%
Category:	data	**Estimated time:**	0:00 hour
Target version:	Release 5.46		
Severity:	normal		

Description

In cichorieae (maybe also E+M) some reference records had a large number of similar extensions.
Most of them (in cich) have already been removed (but not in AUD).

To find all of them use the following sql (adapt Reference_id in second sql, and adapt type_id for other ex-types or in E+M)

```
SELECT MN.*, e.*

FROM Reference_Extension MN INNER JOIN Extension e ON e.id = MN.extensions_id
WHERE MN.Reference_id IN (

SELECT id FROM (
SELECT r.id, r.titleCache,  e.value, e.type_id, COUNT(*) AS n
FROM Reference r INNER JOIN Reference_Extension MN ON MN.Reference_id = r.id INNER JOIN Extension
e ON e.id = MN.extensions_id
GROUP BY r.id, r.titleCache, e.type_id, e.value
HAVING n > 2
ORDER BY r.titleCache, r.id DESC
) AS tmp)
-- AND MN.Reference_id = 4234
ORDER BY MN.Reference_id
;


SELECT e.id
FROM Reference_Extension MN INNER JOIN Extension e ON e.id = MN.extensions_id
WHERE MN.Reference_id = 10911 AND e.type_id IN (2075, 2076)
```

For removal use something like

```
SELECT *
FROM Extension e
WHERE e.id IN (  44536 , 44537, ...)
;

SELECT *
FROM Reference_Extension MN
WHERE MN.extensions_id IN ( 44536 , 44537, ... )
```

which can be filled by results from above second sql

Related issues:

Related to EDIT - task #9364: Cleanup authors with multiple & and protected n...	**Closed**

History

#1 - 06/03/2021 01:37 PM - Andreas Müller

- Related to task #9364: Cleanup authors with multiple & and protected names and teams added