

Ongebruikte_thesaurustermen.xsl

Dit stylesheet genereert een SQL-query die kan opzoeken welke thesaurustermen niet voorkomen in een gegeven catalogus-database (collect of document).

Anders gezegd: de query zoekt alle velden in de opgegeven database die zijn gevalideerd in de thesaurus, zoekt daarvan het linkreferentietag op, en zoekt daar de naam van de bijbehorende index bij. Elke index is een SQL-tabel, en als zodanig doorzoekbaar in SQL Server Management Studio ('SSMS'). De query gaat linkref-index voor linkref-index na welke recordnummers in thesau in geen enkele linkref-index in collect voorkomen. Zie ook ThesauTermenNietInGebruik.ppsx

NB: Soms hebben velden die wel gevalideerd zijn in thesau, geen linkreferentie en ook geen index. De beste oplossing is om in Adlib Designer voor die velden een linkreferentie met een tag aan te maken, en een index voor dat tag te maken en op te bouwen (*reindex this index*).

Je kunt ervoor kiezen om dat niet te doen, want het zijn doorgaans velden die weinig worden gebruikt. Dan leveren die velden in de query een syntax-fout op. Je ziet collect_.term in plaats van collect_indexnaam.term. Die regels kun je verwijderen met een editor of in SSMS. Zie stap 2 hieronder.

Gebruik

Ongebruikte_thesaurustermen.xsl is het stylesheet, dat zet je in je Adlib Designer submap [jouw\pad\naar\Designer]\XSL\DatabaseInfo.

1. In Designer -> Application browser -> datamap, geef rechtsklik op collect, en kies Create Documentation. Kies onder View: Ongebruikte_thesaurustermen.
2. Copy-paste de query in MS SQL Server Management Studio (ingesteld op de juiste database, uiteraard). Regels met collect_.term in plaats van collect_indexnaam.term moet je weghalen (zie hierboven). Voer de query uit ('Run query').
3. De query geeft priref en term: SELECT distinct priref, **term**. Dat dient ter geruststelling ('aha, deze termen worden niet gebruikt'), maar je hebt genoeg aan de prirefs. Maak ervan: SELECT distinct priref en draai de query nogmaals.
4. Sla het resultaat van de query op in een csv.
5. Open die csv in een editor (Notepad++) en vervang alle harde returns (\r\n en Extended in Notepad++), om in komma spatie (', '). Dan heb je een lijst prirefs gescheiden door komma spatie.
6. In je editor, plak dit voor die lijst

```
us = "" and bt = "" and nt = "" and et = "" and rt = "" and %0 =
```

Je hebt nu een zoekzin voor termen waarnaar vanuit collect niet wordt verwezen, die niet-voorkeurs- noch bredere noch nauwere noch equivalente noch gerelateerde term zijn.

7. Deze zoekzin plak je in Adlib in de Thesaurus in Geavanceerd zoeken. Dat resultaat kun je al bekijken met Link informatie. Dit zoekresultaat kun je opslaan in een pointerfile.

Deze stappen kun je herhalen voor alle verwijzende databases, dus ook voor document, people, en archive. Dan heb je het leeuwendeel. Al die zoekresultaten kun je met elkaar

AND-en. Dan hou je een lijst over van termen die niet of weinig gebruikt worden. Met Link overzicht kun je die verder opschonen.

Het stylesheet is makkelijk aan te passen voor ongebruikte termen in Personen en Instellingen. Vervang in het stylesheet 'thesau' door 'people' (en geef het resultaat een andere naam).

Rolf Blijleven, november 2018
rolf@rolfblijleven.nl