

Aanwinsten - rubrieken

Uitleg over de opzet van bestand `sections.xml` en hoe records worden toegewezen aan de rubrieken.

Plaats bestand in configuratie

In de configuratie geef je de plek van het bestand op met instelling `aanwinsten.sections-file`.

Boomstructuur

De rubrieken- of sectielijst is een boomstructuur. Er is één root-sectie. Deze heeft subsecties. Subsecties kunnen ook weer subsecties hebben. In principe kan dat oneindig diep zijn, maar de lijst met 'admonitions' (onderstrepingsvarianten voor titels) in Aanwinsten is maar beperkt. Afhankelijk van de instellingen kan je maximaal 4 of 5 niveau's hebben, inclusief de root. Elke rubriek en subrubriek kan een lijstje met holding-records bevatten.

Nieuwe structuur per collectie

Voor elke nieuwe collectie die verwerkt wordt, wordt het bestand opnieuw ingelezen. Dan zijn de lijstjes met holding-records in elke rubriek weer leeg. Via setting `aanwinsten.collection.<collection-id>.main-sections` kan je opgeven welke hoofdrubrieken je actief wilt hebben en welke niet. Is de setting er niet, dan blijft de rubriekenlijst zoals die is ingelezen. Is de setting er wel (lijst van eenvoudige reguliere expressies, gescheiden door spaties), dan wordt voor elke subsectie direct onder de root gekeken naar het ID. Matcht het ID met een van de patronen, dan blijft die subsectie en alle kinderen ervan ongewijzigd. Matcht het ID niet, dan wordt voor de hele tak elke sectie gemarkeerd als `ignore="true"`. Alle secties met `ignore="true"` vooraf al gezet, blijven dat predikaat altijd houden.

Indexatie

Bibliografische records en holding records worden in een map bijgehouden. Het betreft `HashMap<String, RecordBundle>` records in `CollectieAanwinsten`. Scripties worden bij laden direct aan de gereserveerde rubriek voor scripties gekoppeld. Boeken nog niet. Nadat alles is geladen, wordt er geïndexeerd. Van alle holding records wordt bij indexatie gekeken in welke rubriek ze geplaatst kunnen worden. Dat is de diepste rubriek die matcht met het vindplaats-veld. Daar wordt dan het OCN van het bibliografische record geplaatst. Verder wordt de rubrieks-ID gekoppeld aan het bibliografische record in de records-structuur. Daardoor kan achteraf bepaald worden of een boek in meerdere rubrieken voorkomt.

XML-bestand

Voorbeeld begin XML-bestand van huidige configuratie. Een aantal regels is verdeeld over meerdere regels omdat ze anders te lang werden. Sommige titels zijn ingekort.

```
<?xml version="1.0" encoding="UTF-8"?>
<section match=".*" title-nl="Aanwinsten" title-en="Accessions">
  <section id="KOEPELTITELS" match="[Zz]ie afzonderlijke"
    title-nl="" title-en="" ignore="true" />
  <section id="BENG" match="^BENG"
    title-nl="Beeld en Geluid" title-en="Sound and Vision" />
  <section id="BRA" match="^BRA"
```

```

    title-nl="Boeken over Brabant" title-en="Books on the province of Brabant" />
<section id="CIF (A )" match="^CIF (A )"
title-nl="Handleidingen" title-en="Manuals">
  <section id="CIF A[A-Z]" match="^CIF A[A-Z]"
    title-nl="Naslagwerken" title-en="Reference books" />
  <section id="CIF B" match="^CIF B"
    title-nl="Bibliografie&uml;n" title-en="Bibliographies" />
  <section id="CIF C" match="^CIF C"
    title-nl="Filosofie; algemeen" title-en="Philosophy; general">
    <section id="CIF C1" match="^CIF C1"
      title-nl="Metafysica - Natuurfilosofie"
      title-en="Metaphysics and Philosophy of nature" />
    <section id="CIF C2" match="^CIF C2"
      title-nl="Filosofische antropologie - Filosofische psychologie"
      title-en="Philosophical anthropology and Philosophical psychology" />
    <section id="CIF C3" match="^CIF C3"
      title-nl="Cultuurfilosofie - Geschiedenisfilosofie - Filosofie ..."
      title-en="Philosophy of culture, Philosophy of history ..." />
    <section id="CIF C4" match="^CIF C4"
      title-nl="Esthetica - Kunstfilosofie - Filosofie en Literatuur"
      title-en="Esthetics, Philosophy of art and Philosophy and literature" />
    <section id="CIF C51" match="^CIF C51"
      title-nl="Ethiek" title-en="Ethics" />
    <section id="CIF C52" match="^CIF C52"
      title-nl="Sociale en politieke filosofie - Sociale en politieke ethiek"
      title-en="Social and political philosophy and Social and political ethics">
      <section id="CIF C521" match="^CIF C521"
        title-nl="Bio-ethiek / Medische ethiek"
        title-en="Bio-ethics / Medical ethics" />
      <section id="CIF C522" match="^CIF C522"
        title-nl="Bedrijfsethiek" title-en="Business ethics" />

```

De opzet in dit voorbeeld is sterk gekoppeld aan de hierboven beschreven lijst van class members van class Section. Alle booleans zijn default 'false'.

Scripties

De allerlaatste subsectie van de root is gereserveerd voor de scripties. Hier een voorbeeld van het einde van het XML-bestand:

```

<section id="CBM" match="^CBM" title-nl="Boeken in het gesloten magazijn" title-en="Books in closed stocks" only-unique="true">
  <section id="CBM TF ([A-D])" match="^CBM TF ([A-D])" title-nl="Werken uit het (gesloten) magazijn" title-en="Foo" only-unique="true" />
</section>
<section id="THES" match="^niets matcht op dit$" title-nl="Scripties" title-en="Student theses" />
<!-- Speciaal geval, altijd de laatste -->
</section>

```