

Aanwinsten - configuratie

Java Properties-bestanden

Voor de configuratie wordt gebruik gemaakt van Java Properties-bestanden. Het format wordt beschreven bij de [load-method van de Java Properties class](#). In de bestanden die standaard worden meegeleverd wordt steeds het 'key = value'-formaat gebruikt. Regels die beginnen met een hekje ('#') worden als commentaar beschouwd en genegeerd. Lege regels worden ook genegeerd.

Volgorde van laden

Aanwinsten haalt de configuratie uit maximaal vier bronnen die in deze volgorde worden geladen indien beschikbaar:

- Bestand `/etc/aanwinsten.conf` (Linux)
- Bestand `$HOME/.aanwinsten/aanwinsten.conf` (dus bestand `aanwinsten.conf` in een subdirectory genaamd `.aanwinsten` van de home-directory van de gebruiker)
- Bestand gerefereerd in `-Daanwinsten.conf=<pad/naar/bestand>` in java-commando (moet bestaan indien genoemd)
- Bestand `aanwinsten.conf` in de huidige directory

De eerste optie is niet mogelijk in Windows.

Overschrijven en aanvullen

Elk van de beschikbare bronnen wordt geladen. Indien een waarde is gezet in een configuratie-bestand, kan deze in een volgend configuratie-bestand worden aangepast. Indien een key niet wordt gemeld in een configuratie-bestand, maar in een vorig configuratie-bestand al een waarde heeft gekregen, dan blijft die waarde staan. Het is dus mogelijk om een paar instellingen aan te passen in een bestand dat later wordt geladen. Om een eerder gezette waarde te verwijderen, null te maken als het ware, kan je deze een lege waarde geven. Bijvoorbeeld, om te zorgen dat er niet gelogd wordt naar een logfile, zou in bestand `aanwinsten.conf` in de huidige directory de volgende regel kunnen staan:

```
aanwinsten.logfile =
```

De eerder gezette standaardwaarde `/var/log/aanwinsten.log` (in bestand `/etc/aanwinsten.conf`) is nu uitgezet.

Gedaante van de instellingen

De keys beginnen allemaal met 'aanwinsten.'. Dat is niet een eis van Java of zo, maar voorkomt conflicten met systeem-waarden.

Instellingen

Het complete overzicht van waarden, inclusief uitleg, vind je in bestand `src/main/resources/aanwinsten.conf`. Het heeft weinig zin om die lijst hier te herhalen.

Een paar zaken behoeven wel toelichting.

Wachtwoord

Het wachtwoord voor OCLC staat niet direct in de configuratie. De configuratie geeft met `aanwinsten.ftp.password-file` aan waar het bestand staat waarin het wachtwoord staat. In dat bestand staat 1 regel met het wachtwoord, afgesloten met een `<Return>`. De functioneel beheerders kunnen het wachtwoord verstrekken.

Sets

Per collectie kunnen bestanden worden opgehaald in meerdere sets. Vanaf versie 0.22 is dat configureerbaar. Als voorbeeld hier de configuratie zoals die in het default configuratie-bestand staat, zonder commentaar:

```
aanwinsten.ftp.directory-template = /xfer/metacoll/out/ongoing/${directory-of-set}
aanwinsten.ftp.file-template = metacoll.GRG.${set}.D${year}${month}*.mrc*
aanwinsten.sets = new updates merges
aanwinsten.sets.books = new updates merges
aanwinsten.sets.student-theses = new
aanwinsten.set.new.directory-of-set = new
aanwinsten.set.updates.directory-of-set = updates
aanwinsten.set.merges.directory-of-set = updates
```

In setting `aanwinsten.sets` staat de lijst van sets die Aanwinsten moet onderkennen. Geef in `aanwinsten.sets.books` op welke sets voor boeken moeten verwerkt en in `aanwinsten.sets.student-theses` welke voor scripties. Per set geef je met `aanwinsten.set.<SET>.directory-of-set` aan of je bestanden voor deze set in de 'new'- of 'updates'-directory moet zoeken. In het voorbeeld hierboven is dat voor set 'new' dus de directory met 'new' in de naam en voor 'updates' en 'merges' de directory met 'updates' in de naam. De waarde van deze setting wordt ingevuld in het template voor de directory `aanwinsten.ftp.directory-template` als `${directory-of-set}`. De set wordt dan weer gebruikt in de het template voor de bestandsnamen `aanwinsten.ftp.file-template`.

Collecties

Aanwinsten kan lijsten maken voor verschillende collecties. Elke collectie heeft in de configuratie een eigen ID. Elke collectie kan weer geconfigureerd worden met instellingen waarin het ID terugkomt.

In `aanwinsten.collection-ids` som je de collecties op waarvan je wilt dat die worden verwerkt. De ID's die je daar gebruikt, hoeven geen relatie te hebben met namen van bestanden op de FTP-server, maar ze worden wel gebruikt in de namen van lijst-bestanden en in de titels van lijsten.

Per collectie geef je dan de herken-strings op voor de boeken en de scripties, zodat Aanwinsten de juiste bestanden vindt. Het is niet verplicht om de scripties-herken-string op te geven. Geef je die niet op, dan wordt die niet opgehaald. Per collectie kan de volledige namen opgeven in Nederlands en Engels. Deze kunnen worden gebruikt in de document-template(s).

Wil je tijdelijk een collectie niet verwerken, dan zou je de regel met `aanwinsten.collections` kunnen dupliceren, de eerste regel van een hekje kunnen voorzien zodat die genegeerd wordt en de betreffende collectie-ID in de tweede (de actieve regel dus) kunnen wegnemen. Het uitcommentariëren van de eerste regel is niet eens nodig, omdat die tweede de instelling overschrijft. Later kan je de instelling weer herstellen.

Rubriekenlijst

Instelling `aanwinsten.sections-file` geeft aan waar het bestand staat waarin de secties (rubrieken) zijn gedefinieerd. Hoe dat bestand is opgebouwd en hoe de toewijzing aan rubrieken werkt, wordt in een apart document uitgelegd.

Templates

Het maken van de HTML- en de PDF-bestanden gebeurt in twee stappen. Eerst worden zogenaamde `reStructuredText`-bestanden gemaakt en deze `reStructuredText`-bestanden worden omgezet naar HTML en PDF. `reStructuredText` is een Wiki-achtige taal. Instelling `aanwinsten.templates-directory` geeft aan waar de template-bestanden te vinden zijn. Hoe de templates werken wordt in een apart document uitgelegd. Hoe een lijst eruit ziet wordt in templates (mallen) bepaald.

Cleanups

Na het genereren van de `reStructuredText` voor een record op basis van templates, kan er wat rommel achterblijven. Die wordt met een reeks vervangingen opgeruimd. Dat zijn bijvoorbeeld dubbele spaties die vervangen moeten worden door enkele spaties, spaties voor punten weghalen. In bestand gerefereerd door instelling `aanwinsten.cleanups-file` staan de reeksen. Te vervangen strings gescheiden door spaties van vervangende strings.