

Documentatie vnster.nl

Walter A. Kusters

13 december 2023

Inhoudsopgave

1	Inleiding	2
2	De database	2
2.1	Het HTML-formaat	2
2.2	Wiki-pagina's	6
2.3	Goede boeken	7
2.4	Bijzonderheden	7
3	De website	8
3.1	De PHP-scripts	8
3.2	Verfilmingen	9
3.3	Foto's	10
3.4	De foto-versie	12
3.5	De csv-versie	13
3.6	De app	13
3.7	Collectie	13
4	Onderhoud	14
4.1	Juni	14
4.2	Klein onderhoud	15
4.3	Controle	15
4.4	Data en data	15
5	De scripts	16
5.1	Het zoek-script	16
6	De files	18

1 Inleiding

Deze tekst, `docuvnster.pdf`, bevat uitgebreide documentatie over de website www.vnster.nl, en is geschreven door [Walter Kusters](#). De file is te vinden via

www.liacs.leidenuniv.nl/~kusterswa/vnster/docuvnster.pdf

Hier is trouwens ook de L^AT_EX-broncode, `docuvnster.tex`, aan te treffen.

Voorlopig staat een en ander nog in vrij lukrake volgorde. De website bevat zelf ook veel informatie, in het bijzonder in de hulp-pagina's, en in enkele gevallen staat die informatie ook weer in dit document. De nadruk ligt hier echter op onderhoud en technische details.

2 De database

In feite bestaat de database uit vijf files: `vnsterac.html`, `vnsterdg.html`, `vnsterhl.html`, `vnstermr.html` en `vnstersz.html`, vanaf nu de *data-files* genoemd. Ze bevatten respectievelijk de boeken van auteurs met beginletter A-C, D-G, H-L, M-R en S-Z. Ze bevinden zich in de map `public.html`, met rechten 644. Het zijn HTML-files, die gewoon als zodanig bekeken kunnen worden; misschien kan `tidy` gebruikt worden om ze te controleren. Er staan zelfs nog wat overbodige regels in, overblijfselen van oudere versies. In zekere zin gaat het alleen om de auteurs en hun titels.

De files worden op twee manieren benaderd: via het zoek-script `teztt.cgi` (overigens wordt dan de versie `eenvnsteraz.html` gebruikt, zie Hoofdstuk 4.2) en bij het bladeren via de 26 files `a.html` tot en met `z.html`: de *blader-files*. Deze laatste worden gegenereerd door het update-script (zie opnieuw Hoofdstuk 4.2). Daar wordt nog iets aan de vormgeving gedaan, en bijvoorbeeld de tooltips voor de foto's goed gezet; ook worden tooltips voor de IMDb- en wiki-pagina's geplaatst (zie Hoofdstuk 3.2 en Hoofdstuk 2.2); en de sterren worden verfraaid; en er worden landvlaggetjes bij de auteurs geplaatst.

2.1 Het HTML-formaat

De auteurs staan, gescheiden door lege regels, in de data-files op alfabetische volgorde van achternaam. Per beginletter is het een lijst, zoals voor de "A":

```
1 <!-- ===== -->
2 <h2><a name="letter-a"> A </a></h2>
3
4 <dl>
5
6     eerste auteur
7
8     tweede auteur
```

```
9
10 ...
11 </dl>
```

Een vrij willekeurig voorbeeld, de Engelse auteur Paul Adam:

```
1 <!-- u$$ -->
2 <dt><b>Adam, Paul</b> GB [T]
3 <dd>
4   [<A href="http://nl.wikipedia.org/wiki/Paul_Adam" target="_parent">
5     wiki</A>]
6 <ul>
7   <li id="N00084">Unholy trinity 1999 1* (De onzalige drie-eenheid)
8     [<A K="kaft7/adamp1">KAFT</A>]
9 </ul>
```

Merk op dat de afsluitende tags zoals `</dt>`, `</dd>` en `` doorgaans ontbreken. En het gebruik van kleine letters of hoofdletters in tags is ook niet helemaal zoals het hoort.

We bespreken de verschillende regels van onze voorbeeld-auteur. De eerste regel, commentaar in HTML, zegt in welke jaren VN's de auteur vermeld werd. Zo representeert de “u” vlak voor “\$\$” het jaar 2000. Als de auteur in verschillende jaren staat, komen er ingewikkelde notaties in beeld, bijvoorbeeld

```
<!-- v-xE-HJL$$ -->
```

Dit staat voor 2001–03, 2010–13, 2015, 2017. Een streepje staat klaarblijkelijk voor alle tussenliggende letters. Er geldt: a = 1980, b = 1981, ..., z = 2005, A = 2006, ..., N = 2019, O = 2020, P = 2021. Als de auteur nooit in VN stond, staat er

```
<!-- NIET IN VN -->
```

Tevens staat er dan, direct na `<dd>`, een regel met

```
niet in VN vermeld
```

Na de commentaar-regel komt een regel, beginnend met `<dt>`, waarin allereerst de naam van de auteur(s) staat, tussen `` en ``; eerst de achternaam, dan de voornaam. Soms staat er tussen haakjes nog een andere voornaam of iets dergelijks bij. Als de echte naam van een pseudoniem bekend is, staat dat na de tags, met een gelijkteken, tussen haakjes. Daarna komt het oorspronkelijke land van de auteur, of landen als dat er meer zijn. Tot slot komt het genre, als dat bekend is. Het kunnen ook meerdere genres zijn; deze staan dan op alfabetische volgorde. Een enkele keer staat er zelfs — niet geheel consequent, overigens — een genre bij een individueel boek, met name als de auteur series in verschillende genres schrijft. Het

land wordt door het update-script (zie Hoofdstuk 4.2) en het zoek-script (zie Hoofdstuk 5.1) voorzien van een vlaggetje.

Daarna komt er een regel met alleen <dd>, eventueel gevolgd door wat informatie over de auteur:

- (indien van toepassing, zie boven) de regel met “niet in VN vermeld”
- geboorte- en sterfjaar tussen haakjes, indien bekend (alleen als de auteur al overleden is)
- thuis- en wikipagina, voor zover bekend
- eventuele kleine anecdotes, vaak tussen [en]

De regels, behalve de laatste, eindigen met
. In het voorgaande voorbeeld staat alleen een wiki-pagina; een volgend voorbeeld bevat ook de andere zaken.

Daarna komen de boektitels. Deze zijn per serie geordend, zie het volgende voorbeeld. Voor Paul Adam is dat eenvoudig: hij heeft slechts één boek in de lijst. Zijn lijst is van de vorm

```
<ul>
  <li ... enzovoorts
</ul>
```

In de -regel staat alle informatie over het desbetreffende boek. In de -tag staat doorgaans ook een uniek identificatienummer, aangegeven met “id”, zie Hoofdstuk 3.7. Allerst komt dan de originele titel en het jaar van eerste publicatie — in de oorspronkelijke taal. Vervolgens eventuele sterren: een getal gevolgd door een sterretje. Als deze waardering in VN in de loop der jaren veranderde staan de verschillende waarden in tijdsvolgorde, gescheiden door komma’s, waarbij het sterretje bij de hoogste waarde staat. Enkele scripts vertalen het sterretje naar een wat mooier symbool: (2* krijgt dan twee sterretjes, enzovoorts). Daarna, als het een vertaald boek betreft, tussen haakjes de Nederlandse titel. Vervolgens komen eventuele series waarin het boek voorkomt. In volgorde: pocketseries (in het blauw met), kaftjes (zie Hoofdstuk 3.3), prijzen (in een kleur tussen groen en blauw via) en kleine opmerkingen (tussen [en]). Dan komen eventuele verfilmingen, zie Hoofdstuk 3.2. En misschien een citaat, zie Hoofdstuk 2.3.

Een wat uitgebreider voorbeeld:

```
1 <li id="N00209">The mask of Dimitrios = A coffin for Dimitrios 1939
2   5* (Het masker van Dimitrios)
3   <FONT COLOR=BLUE>Pr733</FONT> = <FONT COLOR=BLUE>ZB2213</FONT>
4   [<A K=ambler/masker>KAFT</A>]
5   [<A K=ambler/masker2>KAFT</A>]
6   [<A K=kaft7/amblerdimi>KAFT</A>]
7   [<A K=kaft15/amblerdimi2>KAFT</A>]
```

```

8      [<A K=ambler/ambleromnibus>KAFT</A>]
9      [ook samen met nacht en Topkapi]
10     <BR><A HREF="http://www.imdb.com/title/tt0037055/" TARGET="
11         _parent">IMDb</A>
12     <em>in 1944 verfilmd door Jean Negolescu, met Peter Lorre en
13     Sydney Greenstreet</em>
14     <BLOCKQUOTE>
15     <FONT COLOR=GREEN>
16     Dimitrios kwam binnen.
17     </FONT>
18     </BLOCKQUOTE>

```

Een tweede voorbeeldauteur (met enkele regels weggelaten):

```

1      <!-- NIET IN VN -->
2      <dt><b><A NAME="heeresma">Heeresma</A>, Faber (Johannes)</b> NL
3      <dd>
4          niet in VN vermeld<BR>
5          (1939–1969)<BR>
6          [<A HREF="http://nl.wikipedia.org/wiki/Faber_Heeresma" TARGET="
7              _parent">wiki</A>]<BR>
8          <i>Francis Pax</i>
9      <ul>
10         <li id="N05532">Bonbons voor Belinda 1969 <FONT COLOR=BLUE>ZB1196</
11             FONT>
12             [<A K=kaft/zb1196>KAFT</A>]
13             [<A K=kaft/heeresma1>KAFT</A>]
14             [<A K=kaft10/heeresmaomni>KAFT</A>]
15             [ook samen met Spaans, Plotsky en Dublin in omnibus]
16         <li id="N05533">Bijzonder Spaans 1969 <FONT COLOR=BLUE>ZB1260</FONT
17             >
18             etcetera
19     </ul>
20     <i>overig</i>
21     <ul>
22         <li id="N05535">Teneinde in Dublin 1969
23             etcetera
24     </ul>

```

In dit voorbeeld is te zien hoe verschillende series worden afgehandeld. De hoofdpersoon (of hoofdpersonen) staat, tussen `<i>` en `</i>`, aan het begin van de betreffende serie; eventuele overgebleven serieloze titels hebben als hoofdpersoon `overig`. Als het beroep van de hoofdpersoon bekend is, staat dat er meestal bij. En als de boeken zich doorgaans in een zelfde locatie aanspelen, staat deze in ronde haakjes (met “in” erbij) na de hoofdpersoon. Let ook op het inspringen.

Als een schrijver op diverse plekken staat, wordt er verwezen via “ZIE”. Voorbeeld: Conan Doyle, bij C en D. Als in een titel een jaartal staat, wordt er ` ` voor gezet, om matches bij het zoeken op jaartal te voorkomen. Verhalen staan in een ``, en hebben `` in plaats van ``; een `` is echt voor een compleet boek.

2.2 Wiki-pagina’s

Bij een flink aantal auteurs, ruim 2900, is hun wiki-pagina opgenomen. Zie Hoofdstuk 2.1 voor een voorbeeld; het formaat is:

```
[<A HREF="http://nl.wikipedia.org/wiki/Faber_Heeresma" TARGET="
  _parent">wiki</A>]
```

Als er meerdere links bij een auteur staan, gaat de eigen thuis-pagina voorop, en daarna de wiki-pagina; ze staan samen op één regel.

Als er een Nederlandse pagina is, heeft deze de voorkeur. Daarna, bij Engelstalige auteurs, een Engelse bladzijde. Of een Duitse of een Franse. Mochten deze er alle niet zijn, dan zijn er nog een aantal Zweedse, Noorse, Deense, Italiaanse, en verder zeven Spaanse, vier Zuid-Afrikaanse, vijf Finse, twee Friese, twee Tsjechische, twee Poolse, één Portugese, één Welshe en één Baskische.

Het script `eerstewiki.pl` genereert uit de data-files (die in dezelfde map moeten staan) een file `dewikis.txt` met daarin alle wiki-links met de eerste alinea van de desbetreffende pagina. Het script doet daartoe zo’n 3000 internet-verzoeken, en probeert dan de eerste alinea van het resultaat op te slaan. Dit gaat als volgt: haal de hele wiki-pagina op; gooi na tussenkomst van de wiki-API (die de eerste serieuze alinea opspoort) alles tot de eerste `<p>` weg, en idem alles na de eerste daarop volgende `</p>`; verwijder tot slot wat HTML-zaken en vreemde tekens. De file `dewikis.txt` moet met de hand naar de webserver worden overgezet, in de map `vnster/`. (Dit script draait overigens niet op de webserver, dat zou niet zo vriendelijk zijn. In 2023 is de interface van IMDb gewijzigd, en werkt het script helaas niet meer.)

Een voorbeeldregel uit `dewikis.txt` is:

```
http://nl.wikipedia.org/wiki/Faber_Heeresma====Faber Johannes Heeresma (Ams...
```

Bij een 20–30-tal pagina’s loopt er iets fout. Deze zijn daarom met de hand opgeslagen in een speciale file `dewikisextra.txt` (in november 2021: 31 regels). Deze file moet ook naar de webserver overgezet worden, in de map `vnster/`. Als een pagina in beide files voorkomt, wordt die uit deze tweede file gebruikt. De file kan eventueel ook benut worden om met de

hand een enkele nieuwe pagina toe te voegen, zonder helemaal `eerstewiki.pl` te hoeven aanmaken; maar dan moet de betreffende regel er te zijner tijd wel weer worden uitgehaald. Er is een apart script `paarwikiserbij.pl` dat hiervoor gebruikt kan worden.

2.3 Goede boeken

Zoals gemeld in Hoofdstuk 3.3 zijn er ook “goede” boeken. Deze worden gekenmerkt doordat bij het boek een citaat staat (de tag `<BLOCKQUOTE>`), en de citaten zijn in het groen via de tag ``. Momenteel (mei 2023) zijn dit er 324. Bij een flink aantal boeken uit Keating’s lijst moet overigens nog een dergelijk citaat gezocht worden.

Wie/wat bepaalt of een boek “goed” is? Een aantal criteria worden gebruikt: allereerst alle titels die in VN ooit vijf sterren kregen (in mei 2023: 226 stuks). Verder alle boeken uit Symons’ top-10, en de in het Nederlands vertaalde boeken uit Keating’s top-101 (met uitzondering van de bovengenoemde titels). Dan nog een klein aantal boeken die Walter erg goed vindt, en een —wederom volgens Walter— stel representatieve titels.

De kaftjes van deze “goede” boeken worden ook nog apart gebruikt, zie Hoofdstuk 3.3.

2.4 Bijzonderheden

Er zijn verschillende eigenaardigheden:

- Bij auteurs uit België staat bij de Vlaamse auteurs een B, bij de Waalse staat voluit België. Zo komt Simenon goed terecht, en levert zoeken naar Belgische boeken alleen Nederlandstalig werk op.
- Bij auteurs uit Wales en Schotland staat er ook nog altijd GB bij.
- Bij jaartallen in titels staat er meestal ` ` voor, om te vermijden dat bij zoeken naar jaartallen het betreffende boek onterecht wordt gevonden.
- Boeken van meerdere auteurs staan soms apart vermeld, maar soms ook bij één van de auteurs. Als voorbeeld: de boeken van Tom Clancy, al sinds jaren door anderen geschreven, staan nog bij Clancy genoemd. De verschillende auteurs worden in de `<dt>`-regel overigens met “en” gescheiden, niet met komma’s.
- Sommige boeken met titels van verschillende auteurs worden apart vermeld, sommige niet.
- De plaats in het alfabet wisselt soms ietwat, zie bijvoorbeeld Mary Higgins Clark (H of C?) en Martin Cruz Smith (M of C?). Doorgaans staat er dan een verwijzing met “ZIE” bij.
- Een tweetal verhalenbundels staat bij Xx, terwijl ze in VN bij een van de auteurs vermeld staan. De vermeldingen bij Xx en Xy zijn niet altijd consequent.

- De titels van verhalen worden met aangeduid, in tegenstelling tot gewone titels: daar wordt <li ... gebruikt.
- Sommige auteurs die onder verschillende namen publiceren, hebben aparte vermeldingen.
- En dan hebben we de “series”. Om verschillende redenen zijn een heel stel titels niet apart, met een eigen <li ..., opgenomen. Ze worden op dezelfde manier afgehandeld als verhalen. Het betreft: Baantjer (secundair werk), Jerry Cotton (2×), Van Gulik (de strips), Inter-Pol, Alistair MacLean (navolgers), Lou Merryl, Spionage-reeks.

3 De website

De website bestaat uit een aantal PHP-script's, bij elkaar gehouden door `index.php`. Ze staan allemaal in de map `vnster/` op de webserver. Deze zelfde map bevat ook de blader-files, de plaatjes van de vlaggetjes en de plaatjes van de voorkanten in de map `vnster/images/`.

3.1 De PHP-scripts

De website bestaat uit de volgende PHP-scripts:

`index.php` De hoofdpagina. Bevat foto's van de voorkanten van een willekeurige en de huidige gids (de 43ste), zie Hoofdstuk 3.3.

`collectie.php` De pagina voor collectie-vorming, zie Hoofdstuk 3.7. Bevat foto's van de voorkanten van gidsen 41 en 42.

`app.php` De pagina voor de app, zie Hoofdstuk 3.6.

`mobile.php` De pagina voor de mobiele versie; ook `mobiel.php`. Deze laatste is namelijk een symbolische link (gemaakt met `ln -s`, zie Hoofdstuk 3.3) naar `mobile.php`.

`legenda.php` Uitleg over de opzet. Bevat foto's van de voorkanten van de gidsen 31 en 34.

`zoeken.php` Geeft toelichting over het zoeken. Bevat foto's van de voorkanten van de gidsen 33, 35, 36 en 37.

`search.php` De zoekpagina. Met `?by=title` kan er naar titels worden gezocht, met `?by=author` naar auteurs.

`parse.php` Vergelijkbaar met `search.php`; `search.php` wordt gewoon vanuit de hoofd-pagina aangeroepen (via de oranje knoppen); `parse.php` als vanuit het door zoeken op titels gegenereerde resultaat wordt gezocht of elders rechtstreeks vanuit de webpagina's. En `parse2.php` is bedoeld voor de titels zonder plaatje. En `search2.php`?

`literatuur.php` Een aantal interessante titels over spannende boeken. Bevat foto van de voorkant van gids 43.

`links.php` Een aantal interessante links over spannende boeken. Bevat foto's van de voorkanten van de gidsen 12, 13, 15, 24, 25 en 30.

`afkortingen.php` De gebruikte afkortingen. Bevat foto van de voorkanten van de gidsen 38, 39 en 40 (en ruimte voor een paar volgende).

`help.php` Een hulp-pagina. Bevat foto's van de voorkanten van de gidsen 1 tot en met 16, en links naar foto's van alle voorkanten. Slechte vormgeving, overigens.

`haalop.php` Een pagina om de verschillende pdf-versies en de `csv`-versie op te halen, zie Hoofdstuk 3.3 en Hoofdstuk 3.5. Bevat foto's van de voorkanten van de gidsen 27 en 32.

`contact.php` De contactgegevens. Bevat foto's van de voorkanten van de gidsen 17 tot en met 24.

`browse.php` Een pagina die de blader-files aanstuurt. Bevat foto's van de voorkanten van de gidsen 12 tot en met 19 (14 twee maal), 22, 23, 25, 26, 27, 28 en 29.

`menu.inc.php` De include-file met de kopregels: alles boven de eerste horizontale balk.

`menu.mobile.inc.php` De include-file met de kopregels: alles boven de eerste horizontale balk — voor de mobiele versie, zie `mobiel.php`.

De `css` style-file `style.css` staat in de subdirectory `vnster/css/`. Er is ergens ook nog een file `android2.css`.

3.2 Verfilmingen

Bij een 2500-tal verfilmingen staat een link naar de [IMDb](http://www.imdb.com/title/tt0037055/), bijvoorbeeld:

```
1 <BR><A HREF="http://www.imdb.com/title/tt0037055/" TARGET="
  _parent">IMDb</A>
2 <em>in 1944 verfilmd door Jean Negulescu , met Peter Lorre en
3 Sydney Greenstreet</em>
```

Deze staan in tijdsvolgorde. Zie ook het voorbeeld in Hoofdstuk 2.1.

In de blader-files wordt, net als bij de wiki-pagina's (zie Hoofdstuk 2.2), bij het hoveren over een IMDb-link het eerste stukje van de betreffende webpagina getoond. Het systeem werkt analoog aan het wiki-gebeuren. Er wordt door `updatevn.pl` een script `imdbberbij.pl` (vergelijkbaar met `wikierbij.pl`) gedraaid in de map `vnster/`, dat gebruik maakt van de file `deimdb.txt` (vergelijkbaar met `dewikis.txt`; analoog is er ook een file `deimdbsextra`, in december 2021 met 42 regels), dat gemaakt wordt door het script `eersteimdb.pl` (vergelijkbaar met `eerstewiki.pl`; analoog `sommigeeersteimdb.pl`). Zie verder Hoofdstuk 2.2.

De uit de IMDb-pagina opgehaalde tekst staat tussen `<meta name="description" content="` en de eerstvolgende `"`. Nog te repareren: de verschillende karakter-coderingen.

3.3 Foto's

In principe staat bij iedere titel een foto van de voorkant. Deze zijn in `jpg`-formaat, en doorgaans hooguit 100 kB groot. Enkele oudere foto's zijn wat groter, evenals een paar bijzonder mooie scans die we van anderen kregen. Bij sommige boeken staan meerdere foto's, vaak in vrij willekeurige volgorde. Er wordt gepoogd alle foto's verschillende namen te geven (ze staan verspreid over allerlei mappen).

In de data-files (zie Hoofdstuk 2.1) staat per plaatje een regel als

```
[<A K=kaft18/connellynachtploeg>KAFT</A>]
```

Let op de speciale tag “K”. Deze link gaat in werkelijkheid naar

```
https://www.liacs.leidenuniv.nl/~kosterstwa/kaft18/connellynachtploeg.jpg
```

De vertaling wordt gemaakt in het update-script, zie Hoofdstuk 4.2. In de blader-files, en ook na een zoekvraag via het zoek-script (zie Hoofdstuk 5.1), komt daardoor te staan (de drie regels vormen één regel, op de plek van de regelovergangen staat een spatie):

```
[<A HREF="https://www.liacs.leidenuniv.nl/~kosterstwa/kaft18/connellynachtploeg.jpg"
TARGET="_parent" onMouseOver="Tip('<img src=\'../kaft18/connellynachtploeg.jpg\'
height=\'90\' width=\'60\'>',BGCOLOR,'White')" onMouseOut="UnTip()">foto</A>]
```

Bovenin deze files staat (direct na de `<body>`-tag):

```
<script type="text/javascript" src="wz_tooltip.js"></script>
```

Dit zijn de tooltips van Walter Zorn, zie www.walterzorn.de. Als je nu bij het bladeren over een foto hovert, krijg je een klein plaatje van de betreffende voorkant te zien.

De laatste tijd is er per jaar één map (in de map `public.html` op de webserver) met foto's gemaakt, zoals `kaft19` en `kaft21` voor 2019 en 2021. In de jaren daarvoor is het wat onregelmatiger. De gebruikte mappen zijn: `kaft` (1711 files), `kaft2` (588 files), `kaft3` (131 files), `kaft4` (326 files), `kaft5` (1799 files), `kaft6` (1733 files), `kaft7` (1797 files), `kaft8` (1685 files), `kaft9` (3580 files), `kaft10` (1417 files), `kaft11` (686 files), `kaft12` (1204 files), `kaft13` (1107 files), `kaft14` (663 files), `kaft15` (821 files), `kaft16` (567 files), `kaft17` (670 files), `kaft18` (737 files), `kaft19` (931 files), `kaft20` (756 files), `kaft21` (1062 files), `kaft22` (842 files), `kaft23` (actueel, ≥ 760 files), `ambler` (Eric Ambler; 27 files), `gardner` (Erle Stanley Gardner; 106 files). De map `kaft` heeft nog wat sub-directories: `ac` (Agatha Christie; 293 files), `bob` (Bob Evers; 50 files), `umc` (UMC pockets etcetera; 436 files), `ivans` (Ivans; 98 files), `havank` (Havank; 46 files); en in `kaft2` zitten `multon` (Multon; 101 files) en `interpol` (Interpol; 204 files). Al deze laatste kleinere mappen zijn ietwat in vergetelheid geraakt; veel kaftjes die daarin hadden moeten staan, bevinden zich gewoon in de `kaft??`-mappen — voortaan te noemen *kaft-mappen*.

Er staan in de map `vnster/images/` foto's van de voorkanten van alle VN's, te zien op de verschillende PHP-pagina's: zie Hoofdstuk 3.1. In de map `vnster/` staan ook plaatjes van de

vlaggen van de voorkomende landen, in principe zowel in gif-formaat als in jpg-formaat; dat laatste is nodig voor L^AT_EX.

Verder is er een grote pdf-versie met alle foto's beschikbaar, zie Hoofdstuk 3.4. Bij het zoeken kan ook een pdf met alleen foto's gegenereerd worden, bijvoorbeeld "5-sterren-behang", zie Hoofdstuk 5.1.

Het script `checkjpg.pl` (zie ook Hoofdstuk 4.3) controleert of de foto's die genoemd worden in de data-files inderdaad in de kaft-mappen voorkomen.

Van circa 710 titels ontbreekt nog een foto. In `legenda.php` (zie Hoofdstuk 3.1) staat daarom:

```
Maar van <A HREF="parse2.php?by=title">circa 710 titels</A>
ontbreekt nog zo'n foto&nbsp;...
```

Het verouderde PHP-script `parse2.php` genereert alle titels zonder plaatje, in eenvoudig HTML-formaat. Het maakt gebruik van een verouderde versie `tezzt2.cgi` van het zoek-script `tezzt.cgi`, zie Hoofdstuk 5.1.

De plaatjes van de "goede" boeken (zie Hoofdstuk 2.3) zijn speciaal. Deze (of preciezer gezegd steeds de eerste van de eventueel verscheidene plaatjes per boek) worden als willekeurig plaatje gebruikt bij de zoekpagina's en een enkele andere plek, zoals de mobiele pagina. Er wordt gekeken of bij het boek een citaat staat (de tag <BLOCKQUOTE>), wat bepaalt dat het een "goed" boek betreft. Het speciale script `maaktitels.pl` verzorgt een en ander: dit genereert de "array-regel" (zie straks) in een file `goedetitels.txt`, en het aantal "goede" titels: 324 (in mei 2023). In `zoeken.php`, `legenda.php`, `mobile.php` (deels 3×) en `search.php` (2×), zie Hoofdstuk 3.1, staat

```
1 <script type = "text/javascript" language = "JavaScript">
2 <!--
3 var titels = new Array ("Adler-Olsen", "kaft9/adlerfles", ...);
4 var nummer = Math.ceil (Math.random () * 324);
5 document.write ('');
6 //-->
7 </script>
```

Dit genereert bij iedere nieuwe aanroep een willekeurig plaatje. Merk op dat het getal 324 en de namenlijst in het array met de hand moeten worden vervangen (gebruik in `gvim` een commando als `:r goedetitels.txt` op de juiste plek) als er iets bij de goede boeken wijzigt, doorgaans medio juni (zie Hoofdstuk 4.1). Er moet goed gelet worden op auteurs met spaties in hun naam, of diacritische tekens. Het speciale script `maaktitels.pl` verzorgt zoals gezegd een en ander: dit genereert de "array-regel", en het aantal "goede" titels.

Op analoge manier worden de plaatjes van de voorkanten van de VN's gebruikt voor wisselende illustraties. Zo staat op de pagina `index.php`:

```

1 <script type="text/javascript" language="JavaScript">
2 <!--
3 var hetnummer = Math.ceil ( Math.random ( ) * 43 );
4 var jaar = hetnummer + 1979;
5 document.write ('');
6 //-->
7 </script>
8
9 

```

Het eerste stuk zorgt voor een wisselende voorkant, het tweede voor de voorkant van de meest recente gids. Merk op dat er momenteel 43 gidsen gebruikt worden, de huidige (de 44ste) wordt niet voor de wisselvoorkant ingezet. In de map `vnster/` zijn symbolische links gemaakt (zie boven voor het UNIX-commando `ln -s`) naar deze voorkanten: `1.jpg`, `2.jpg`, enzovoorts. Zie ook Hoofdstuk 3.1 voor de locatie van de plaatjes van de voorkanten.

3.4 De foto-versie

Met behulp van het Perl-script `f2h.pl` kan van de vijf data-files een bijzonder grote pdf-file `fotodetec.pdf` gemaakt worden, waarin alle foto's staan. De file is meer dan 800 MB groot, en bevat meer dan 900 pagina's.

Het bestand is op te halen van de pagina `haalop.php`, onder de naam

www.liacs.leidenuniv.nl/~kosterswa/vnster/fotodetec.pdf.

Hoe werkt het? De benodigde bestanden staan in de map `public.html`. In ieder geval nodig zijn de twee scripts `f2h.pl` en `verbeter.pl`, en uiteraard de vijf bovengenoemde files. Met behulp van PROMPT `./f2h.pl` wordt een L^AT_EX-file `fotodetec.tex` gemaakt. Om deze te “tex-en” zijn ook alle `jpg`-bestanden nodig. (Denk ook aan de plaatjes van de vlaggetjes.) Deze staan in allerlei mappen, zie Hoofdstuk 3.3. Het commando PROMPT `pdflatex fotodetec.tex` maakt de pdf-file. Om wat minder uitvoer te genereren, gebruik PROMPT `pdflatex -interaction=batchmode fotodetec.tex`. Het is verstandig om twee maal te “tex-en”, want dan is de inhoudsopgave ook correct. De grote file moet met de hand naar de map `vnster/` verplaatst worden, en aantal pagina's en filegrootte (in MB) moeten dan met de hand in `haalop.php` worden aangepast.

Het script `f2h.pl` gebruikt om onduidelijke redenen het Perl-script `verbeter.pl`, zie Hoofdstuk 4.2. (Het zet vermoedelijk de kaft-tags goed.)

3.5 De csv-versie

Met behulp van het script `html2csv.pl` kan van de vijf data-files een csv-file `detec.csv` gemaakt worden. De file is ruim 2 MB groot, en heeft precies evenveel regels als er titels zijn. De file wordt twee maal per jaar gemaakt: eind juni en eind december. Het bestand is op te halen van `haalop.php` onder de naam

www.liacs.leidenuniv.nl/~kosterswa/vnster/detec.csv.

Elke regel bevat elf velden, gescheiden door punt-komma's: auteur, land, originele titel, titel van de vertaling, jaartal, genre, hoofdpersoon, sterren, series, prijzen en aantal verfilmingen. De afzonderlijke velden staan tussen dubbele aanhalingstekens. Accenten worden weggelaten. Hoe werkt het? De benodigde bestanden staan in de map `public_html/` op de webserver. Gebruikt worden allereerst de vijf bovengenoemde data-files. Het script `html2csv.pl` in de submap `cgi-bin/` produceert de file `detec.csv`. Met behulp van het UNIX-commando `diff` kan het verschil met de vorige versie bepaald worden, en deze file (doorgaans `diff???.txt`, met ?? het volgorde-nummer, geheten; momenteel lopend van 1 tot en met 35) wordt ook opgeslagen in `vnster/` en is van de bovengenoemde website op te halen, via `haalop.php`. Vermoedelijk zijn niet alle diacritische tekens even goed afgewerkt, en er kunnen nog wat andere vreemde foutjes in zitten.

3.6 De app

Met behulp van de “App Inventor” van MIT (zie www.appinventor.mit.edu; inloggen via Walter's gmail-account; compileren met “Build” naar een app) is snel een app “MobiëlVNster” in elkaar geknutseld. De betreffende file heet `MobiëlVNster.apk`, en is op te halen vanaf `app.php`.

In feite is de app een eenvoudige browser, die direct opstart in

www.liacs.leidenuniv.nl/~kosterswa/vnster/mobiel.php

Merk op dat `mobiel.php` en `mobile.php` hetzelfde zijn, de eerste is een symbolische link naar de tweede (met `ln -s`, zie Hoofdstuk 3.3).

3.7 Collectie

Voor mensen die zelf een collectie boeken willen aanleggen, is bij `collectie.php` te lezen hoe dat op een eenvoudige manier kan. De bedoeling is een text-file te maken met daarin nummers die corresponderen met titels. Een voorbeeldfile `vijfsterren.txt` staat in de map `vnster/` en bevat alle titels met vijf sterren. Verder bevat `favo.txt` wat favorieten van Walter, en `honderd.txt` zijn keuze van 97 goede boeken. Bij `collectie.php` is gedetailleerd te zien hoe de spelregels zijn.

Het nummer van een een boek is in de data-files eenvoudig te vinden in de ``-tag, bijvoorbeeld `<li id="N00084">` bij auteur Paul Adam in Hoofdstuk 2.1: nummer 00084 of 84; er wordt dus een speciale identifier gebruikt. Deze nummers worden enkele keren per jaar

met behulp van het script `nummer.pl` in de data-files aangebracht, waarbij oude nummers onveranderd blijven en nieuwe worden toegevoegd. Tot die tijd hebben nieuwe titels als tag ``, op enkele plekken als 0 behandeld. Een aantal nummers (88 stuks in augustus 2022) is in de loop der tijden weer verwijderd, maar wordt niet opnieuw gebruikt. Dit staat geadmineerd in de file `vnster/nummers.txt`, toegankelijk via `collectie.php`. Deze file wordt met de hand bijgehouden.

4 Onderhoud

In dit hoofdstuk wordt technisch onderhoud omschreven.

4.1 Juni

De belangrijkste actie vindt ieder jaar plaats in juni, met het uitkomen van de nieuwe VN-gids. In de data-files moeten dan alle nieuwe titels worden ingevoerd, een flinke klus. Talrijke auteurs worden weggepromoveerd uit de categorie “niet in VN”. Daarna kan de update-stap uit Hoofdstuk 4.2 steeds worden gebruikt. Uiteraard moet `index.php` als eerste worden aangepast.

Waarschijnlijk is kort hiervoor de nummering, zie Hoofdstuk 3.7, bijgewerkt. Dit heeft als voordeel dat de nieuwe titels uniek te herkennen zijn aan hun ``-tag zonder “id”.

Er wordt gebruik gemaakt van de unieke letter voor het betreffende jaar, bijvoorbeeld `R` voor 2023. Er zal dus na het jaar `Z` (in 2032) een klein technisch probleempje ontstaan — maar dat zien we dan wel weer. Op enkele plaatsen moet deze letter worden ingevoerd: bij de jaarmenu’s in `search.php` (2×!) en `search2.php` (?) bijvoorbeeld, en in `f2h.pl` en `html2tex.pl`, en in het zoek-script (bij de variabele `$laatstejaar`, en waar deze letter door een getal moet worden vervangen), en in `update.pl`. En evenzo bij “recent” in de jaren-optie op `search.php`, `search2.php` en `mobile.php`.

Bij `zoeken.php` staan drie voorbeeld-aanroepen naar PDF-pagina’s; de aantallen pagina’s veranderen doorgaans zodra er weer meer gesterde boeken bij komen.

Er moet een foto van de voorkant in de map `vnster/images/` worden toegevoegd, zie Hoofdstuk 3.3 voor de benodigde acties. (Let ook op `f2h.pl`.) En er moet een plaatje worden gemaakt op (bijvoorbeeld) `afkortingen.php`, waar nu alleen de voorkanten van 2018, 2019 en 2020 staan. En op `help.php` staan alle voorkanten — als link.

Verder moeten de nieuwe vijf-sterren titels een citaat krijgen. En ook moeten, zoals omschreven in Hoofdstuk 3.3, enkele aanpassingen gemaakt worden voor de foto’s hiervan. Verder moet hun nummer worden toegevoegd in de voorbeeldfile `vijfster.txt` in de map `vnster/`, zie Hoofdstuk 3.7.

Het percentage “gesterd” moet worden bijgewerkt. In `f2h.pl` en `html2tex.pl` staan in de tekst nog meer zaken die moeten worden geupdated, zoals verschijningsdatum van VN en enkele getallen.

4.2 Klein onderhoud

Als er iets aangepast is aan de data-files, moet het Perl-script `updatevn.pl` gedraaid worden. Dit script verzorgt nogal wat. Zo worden de 26 nieuwe blader-files gegenereerd (met vlaggetjes, sterretjes en dergelijke; de “KAFT”-regels worden omgezet naar de juiste links, met tooltips; allerlei interne links worden goed gezet (bijvoorbeeld de Keating-links, maar ook interne verwijzingen naar andere auteurs); genres worden voluit geschreven; VN-jaren worden gegenereerd uit de letter-coderingen), evenals nieuwe `php`-files waarin aantal titels en auteurs bijgewerkt is (er wordt gebruik gemaakt van een simpele `grep "<li"` en `wc`). Ook wordt er via het script `maakeen.pl` één licht opgeschoonde file `eenvnsteraz.html` gegenereerd, met alle informatie uit de vijf data-files, die vervolgens door het zoek-script (zie Hoofdstuk 5.1) gebruikt wordt. Elke titel staat daarin volledig op één regel. Het eenvoudige script `verbeter.pl` wordt hierna aangeroepen om onduidelijke redenen.

Het Perl-script `html2tex.pl` maakt een \LaTeX -file `detec.tex` met vrijwel alle informatie; doe hierna `pdflatex detec` aan. Zo wordt een nette file `detec.pdf` gemaakt, die te krijgen is via `haalop.php` als

www.liacs.leidenuniv.nl/~kosterwa/vnster/detec.pdf

De file is momenteel circa 6 MB groot, en telt omstreeks 570 pagina's. Sinds januari 2021 moet dit script op een eigen machine draaien, en de file `detec.pdf` met hand naar `vnster` gekopieerd worden.

Eens in de zoveel tijd moeten de wiki-files en de imdb-files worden bijgewerkt, zie respectievelijk Hoofdstuk 2.2 en Hoofdstuk 3.2.

4.3 Controle

Er zijn enkele scripts die wat controleren. Zo kijkt `checkbr.pl` of er regels zijn die op `
` eindigen terwijl er geen (zinvolle) regel na komt. En het script `stats.pl` genereert enkele statistieken, zie Figuur 1.

Het script `checkniet.pl` kijkt of er auteurs in VN staan die geen genre hebben, en of de “niet in VN” vermelding juist is.

Het script `uniek.pl` kijkt welke foto's meer dan één maal voorkomen, en zet dubbele namen in de file `dubbele.txt`. Dit zijn omnibussen, etc. In `dedubbele.txt` staat de meest recente versie. Tevens wordt het aantal foto's geteld.

Ook het script `checkjpg.pl` moet genoemd worden, zie Hoofdstuk 3.3.

4.4 Data en data

Hier wordt bijgehouden wanneer welke files voor het laatst zijn geupdated, zie Tabel 1.

```
27177 101591 1087130 vnsterac.html
19437 70645 759659 vnsterdg.html
25136 89947 951443 vnsterhl.html
24909 89266 956882 vnstermr.html
26052 93608 982417 vnstersz.html
122711 445057 4737531 total

Aantal regels:          122711
Aantal auteurs:        5992 en ZIE: 67
Aantal niet in VN:     2597
Aantal titels:         22000
Aantal verhalen:       3207
Aantal citaten:        324
Verfilmingen:          2552
Kaftjes:                27472
Aantal hoofdpersonen: 2902 waarvan overig: 628
Aantal thuislinks:    2477
Aantal wikis:          3006
Jaartallen:            21838
Geen jaartallen:      156
Geen jaartal nodig:   6
Met genre:             4626
Zonder genre:          1366
```

Figuur 1: Uitvoer van `stats.pl` op 23 november 2023.

5 De scripts

In dit hoofdstuk worden de verschillende Perl-scripts genoemd, hun locatie, en de hoofdstukken waar ze worden uitgelegd, zie Tabel 2.

De meeste scripts staan dus in de map `cgi-bin/` op de webserver; deze bevat ook nog enkele oudere versies onder iets aangepaste namen.

5.1 Het zoek-script

Het belangrijkste Perl-script is het *zoek-script* `tezzt.cgi`. Het heeft momenteel (mei 2022) 1822 regels, en 76.010 bytes, en is helaas vrij ingewikkeld. Binnen `parse2.php` wordt overigens nog een sterk verouderde versie `tezzt2.cgi` gebruikt, zie Hoofdstuk 3.3. (Ook in `search2.php`, maar waar dat PHP-script voor dient?)

Waar en hoe wordt het zoek-script aangeroepen? Momenteel wordt het gebruikt in `search.php`,

naam	laatste update	Hoofdstuk
de vijf data-files 22040 titels	8 december 2023	2
dewikis.txt	24 juli 2023	2.2
dewikisextra.txt	24 juli 2023	2.2
deimdb.txt	19 april 2021	3.2
deimdbsextra.txt	3 oktober 2023	3.2
detec.pdf 22040 titels, 577 pagina's, 7 MB	8 december 2023	4.2
fotodetec.pdf 22000 titels, 940 pagina's, 868 MB	26 november 2023	3.4
detec.csv 22040 titels	12 december 2023	3.5
nummering 22000 titels	23 november 2023	3.7
de app	1 juli 2021	3.6
teztt.cgi	28 mei 2023	5.1
dedubbele.txt 466 titels	8 november 2023	4.3

Tabel 1: Updaten van files.

`mobiel.php` en `parse.php`. In bijvoorbeeld `mobiel.php` staat op één regel de JavaScript-aanroep

```
$file = "https://www.liacs.leidenuniv.nl/~kosterswa/cgi-bin/teztt.cgi?"
. implode("&", $a);
```

Dit is de aanroep die het vaakst voorkomt. In `search.php` staat naast deze ook nog een versie met `... $a) . "&metfile=$randfile";`

De belangrijkste parameters van het zoek-script staan in Tabel 3. De meeste hiervan worden bij de zoek-pagina `search.php` ingesteld.

Een voorbeeld-aanroep is

```
https://www.liacs.leidenuniv.nl/~kosterswa/teztt.cgi?genre=H&land=GB
```

Dit genereert Engelse historische thrillers.

Bij het zoeken wordt ook een link “oeuvre” gemaakt, die leidt naar alle titels van de betreffende auteur — met veel informatie erbij. Er staat dan (uiteraard op één regel) zoiets als

```
<A HREF="parse.php?by=title&heldjanee=j&metfilms=j&citaten=j&extra=j
&verhalen=j&plaatjes=j&zoeken=Ambler,+Eric">oeuvre</A>
```

Let op het gebruik van het PHP-script `parse.php`.

Voor het in Hoofdstuk 3.3 genoemde “5-sterren-behang” wordt het script in `zoeken.php` als volgt aangeroepen:

naam	locatie	hoofdstuk
checkbr.pl	thuis	4.3
checkjpg.pl	cgi-bin	3.3
checkniet.pl	thuis	4.3
eersteimdb.pl	locaal/Leiden	3.2
eerstewiki.pl	locaal/Leiden	2.2
f2h.pl	locaal/Leiden	3.3
html2csv.pl	cgi-bin	3.5
html2tex.pl	cgi-bin/thuis	4.2
imdberbij.pl	vnster	3.2
maakeen.pl	cgi-bin	4.2
maaktitels.pl	vnster	3.3
nummer.pl	thuis	3.7
stats.pl	thuis	4.3
tezzt.cgi	cgi-bin	5.1
uniek.pl	thuis	4.3
updatevn.pl	cgi-bin	4.2
verbeter.pl	cgi-bin	4.2
wikierbij.pl	vnster	2.2

Tabel 2: De verschillende scripts.

```
<A HREF="parse.php?by=title&sterren=5&alleen=j">
10-bij-10 vijf-sterren-behang</A>
```

Dat hier overigens `&` staat betekent waarschijnlijk niets.

6 De files

In dit hoofdstuk worden de verschillende files —afgezien van de PHP-scripts (zie Hoofdstuk 3.1) en de Perl-scripts (zie Hoofdstuk 5) — beschreven. Zie Tabel 4.

Walter A. Kusters, Den Haag, 13 december 2023

naam	omschrijving
by	OF name OF title
auteurs	OF j OF “niks” (alleen auteurs?)
heldjanees	OF j OF “niks” (hoofdpersonen erbij?)
verhalen	OF j OF “niks” (verhalen erbij?)
zoeken	een zoekstring
jaren	welke jaren? bijvoorbeeld “ 1960-9”
land	welk land? bijvoorbeeld NL
genre	welke genre? bijvoorbeeld PT
sterren	hoeveel sterren? bijvoorbeeld 45
plaatjes	OF j OF “niks” (plaatjes erbij?)
plagr	OF klein OF middel OF groot (grootte van de plaatjes, in onbruik geraakt)
extra	OF j OF “niks” (extra auteur-informatie erbij?)
alleen	OF j OF “niks” (alleen plaatjes, en PDF)
serie	welke serie? bijvoorbeeld ZB
sleutel	OF n of “niks” (gesorteerd? n betekent op nummer!!)
metfilms	OF j OF “niks” (verfilmingen erbij?)
film	OF j OF “niks” (alleen verfilmde titels?)
citaten	OF j OF “niks” (citaten erbij?)
uitso	OF PDF OF “niks” (PDF- of HTML-uitvoer?)
startjaar	beginjaar voorkomens in VN (bijvoorbeeld v voor 2001)
eindjaar	idem, eindjaar (beide bij zoeken op auteur)
minsterren	minimaal totaal aantal sterren (bij zoeken op auteur)
mintitels	minimaal aantal titels (bij zoeken op auteur)
nummers	OF j OF “niks” (met boeknummers erbij?)
juistwel	OF n OF “niks” (“inverse” titels bij collectie?)
metfile	OF “niks” OF de naam van de eigen file
alleen	OF j of “niks” (alleen plaatjes, en PDF)
hebbes	euuh ... iets met vinkjes

Tabel 3: De verschillende parameters voor het zoek-script `tezst.cgi`. Doorgaans betekent j “ja” en n betekent “nee”.

naam	locatie	hoofdstuk(ken)
a.html (blader-files) (idem b...z)	vnster	2, 4.2
deimdb.txt	vnster	3.2
detec.csv	vnster	3.5
detec.pdf	vnster	4.2
dewikis.txt	vnster	2.2
diff...txt	vnster	3.5
eenvnsteraz.html	public_html	4.2
fotodetec.pdf	vnster	3.4
MobielVNster.apk	vnster	3.6
nummers.txt	vnster	3.7
vijssterren.txt	vnster	3.7
dedubbele.txt	vnster	4.3
favo.txt	vnster	3.7
honderd.txt	vnster	3.7
vnsterac.html (data-files) (idem dg/hl/mr/sz)	public_html	2

Tabel 4: De verschillende files.