

# REA COLLEGE

auteur: Tom Bartels

versie: 11-4-2019

In deze les komen de volgende onderwerpen aan bod:

- .htaccess
- Homepage instellen
- Canonisatie bestanden
- HTTP status messages
- 301 redirects
- Error documents

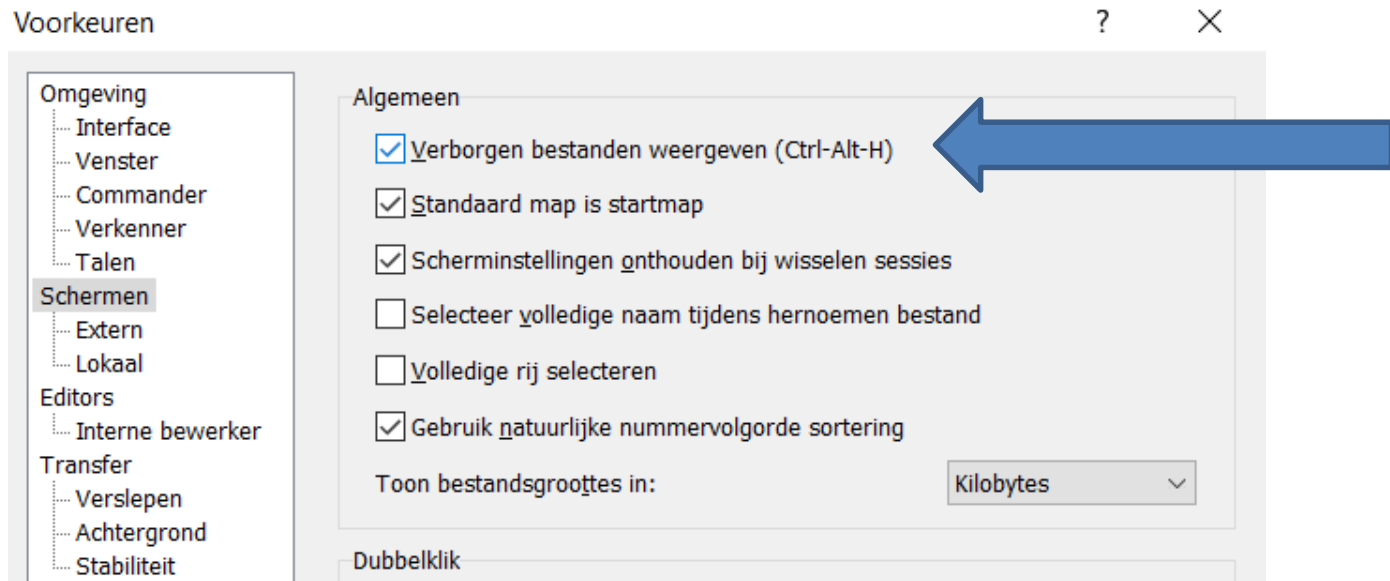
- Met een **.htaccess** bestand kun je *server side* bepaalde instellingen aanpassen.
- Een **.htaccess** bestand heeft lokaal dus geen effect.
  - Je hebt dus een (virtuele) server nodig.



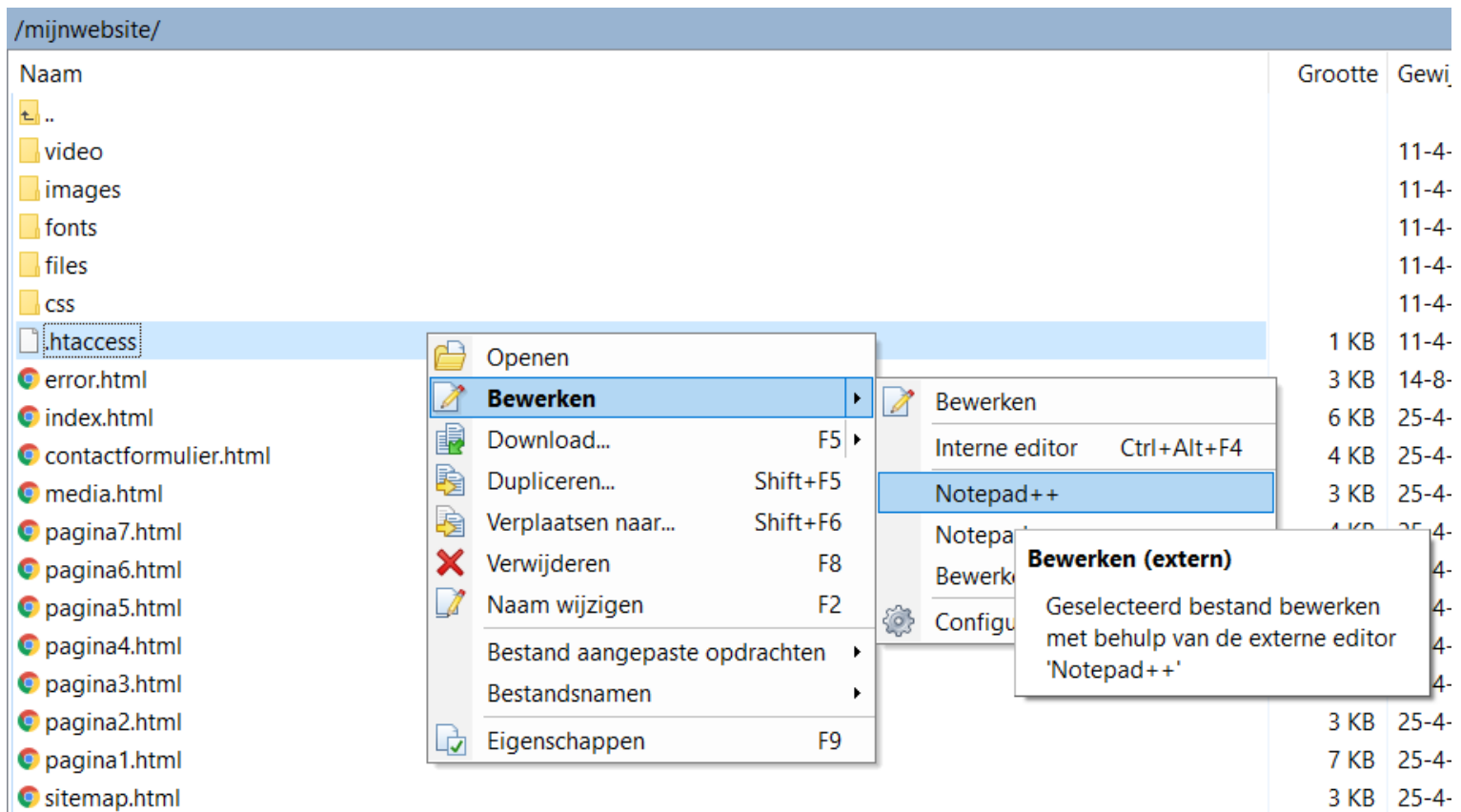
- Een aantal .htaccess configuraties heeft een direct of indirect effect op SEO.
- Enkele voorbeelden:
  - Nette errorpagina's maken.
  - Het instellen van canonieke URL's voor bestanden die geen echte pagina's zijn (zoals een .pdf).
  - Mensen automatisch doorverwijzen naar de juiste pagina.
  - SEO-vriendelijke URL's genereren.
  - .html of .php extensies verwijderen uit de URL.



- Belangrijk: **.htaccess** is de extensie van je bestand, niet de naam!
- Omdat het bestand geen naam heeft wordt het mogelijk standaard niet getoond in je verkenner.
- Zorg dus dat je de optie "extensies verbergen" uitzet in Windows.
- Ook in WinSCP zul je die optie moeten instellen.



- Zodra je je .htaccess bestand hebt aangemaakt kun je deze openen en bewerken in Notepad++.



- De syntaxis voor .htaccess bestanden is niet 'echt' een programmeertaal.
- Je geeft eigenlijk letterlijk je instellingen op.
- Met de hashtag ( # ) kunnen we commentaar plaatsen in onze .htaccess bestanden.

```
# Deze instelling doet iets.  
Instelling instelling
```

- Wanneer je een willekeurige directory op het internet bezoekt in je browser zal de server in eerste instantie zoeken naar een bestand met de naam "index.php".
- Wordt die niet gevonden, dan wordt er gezocht naar "index.html".
- Wordt die óók niet gevonden, dan krijg je een overzicht te zien van alle bestanden in een map – mits dat server side wordt toegelaten.



# REACOLLEGE

## Homepage aanpassen

---

- Met de opdracht **DirectoryIndex** kun je zelf bepalen wat de standaardpagina wordt van een bepaalde map.
- Zo zit je dus niet per se vast aan "index.php" of "index.html".

```
# Verandert de homepage van "index.php" naar "home.php"  
DirectoryIndex /kukomo/home.php
```

- In Les 5 – Webcrawlers hebben jullie geleerd dat **duplicate content** een groot probleem kan zijn voor je zoekmachineoptimalisatie.
- Om duplicate content te vermijden maken we gebruik van **canonieke URL's**.
  - We gebruiken een canonieke URL om aan zoekmachines te vertellen welke URL de standaard URL is die geïndexeerd dient te worden.

```
<!-- brochure.html -->
<head>
  <link rel="canonical" href="http://kukomo.nl/brochure.html">
  <meta name="robots" content="index, follow">
</head>
```

```
<!-- brochure-print.html -->
<head>
  <link rel="canonical" href="http://site.nl/brochure.html">
  <meta name="robots" content="noindex, follow">
</head>
```

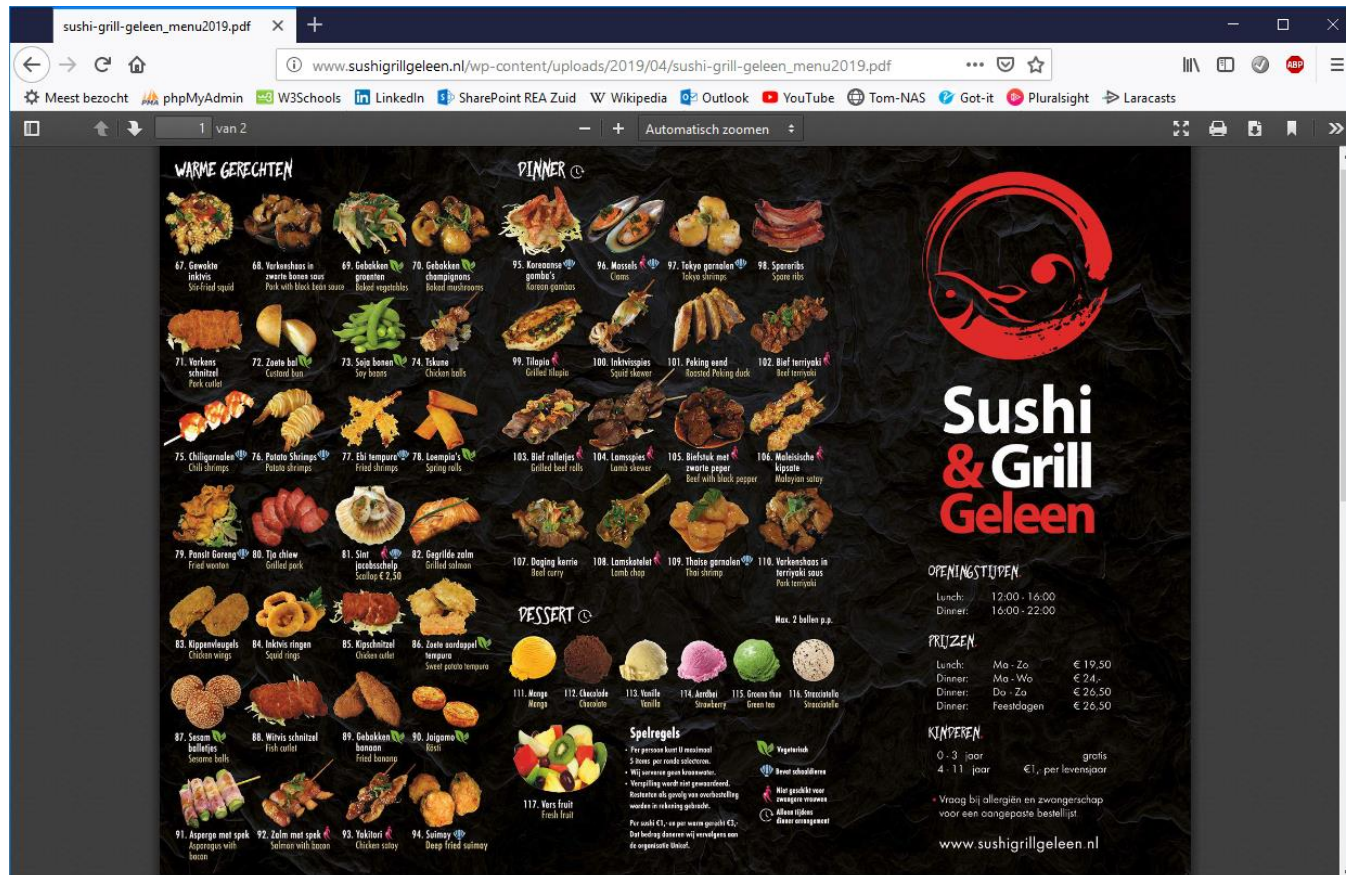
- Google is tegenwoordig ook in staat om – naast normale webpagina's – bepaalde bestandstypes te indexeren.
- Google kan overweg met onder andere:
  - Excel, PowerPoint en Word bestanden.
  - Tekst en rijke tekst (rtf) bestanden.
  - PDF bestanden.
  - Broncode van sommige programmeertalen.
- Zie hier een volledig overzicht:  
<https://support.google.com/webmasters/answer/35287>.

- Je kunt er dus voor kiezen om delen van je content aan te bieden in een andere format dan een traditionele website.
- Denk bijvoorbeeld aan:
  - Een brochure over een bedrijf als PDF bestand.
  - Een menukaart voor een restaurant als PDF bestand.
  - Een scriptie of werkstuk als Word bestand.
  - Een kostenbegroting als Excel bestand.
  - Een presentatie als PowerPoint bestand.

# REACOLLEGE

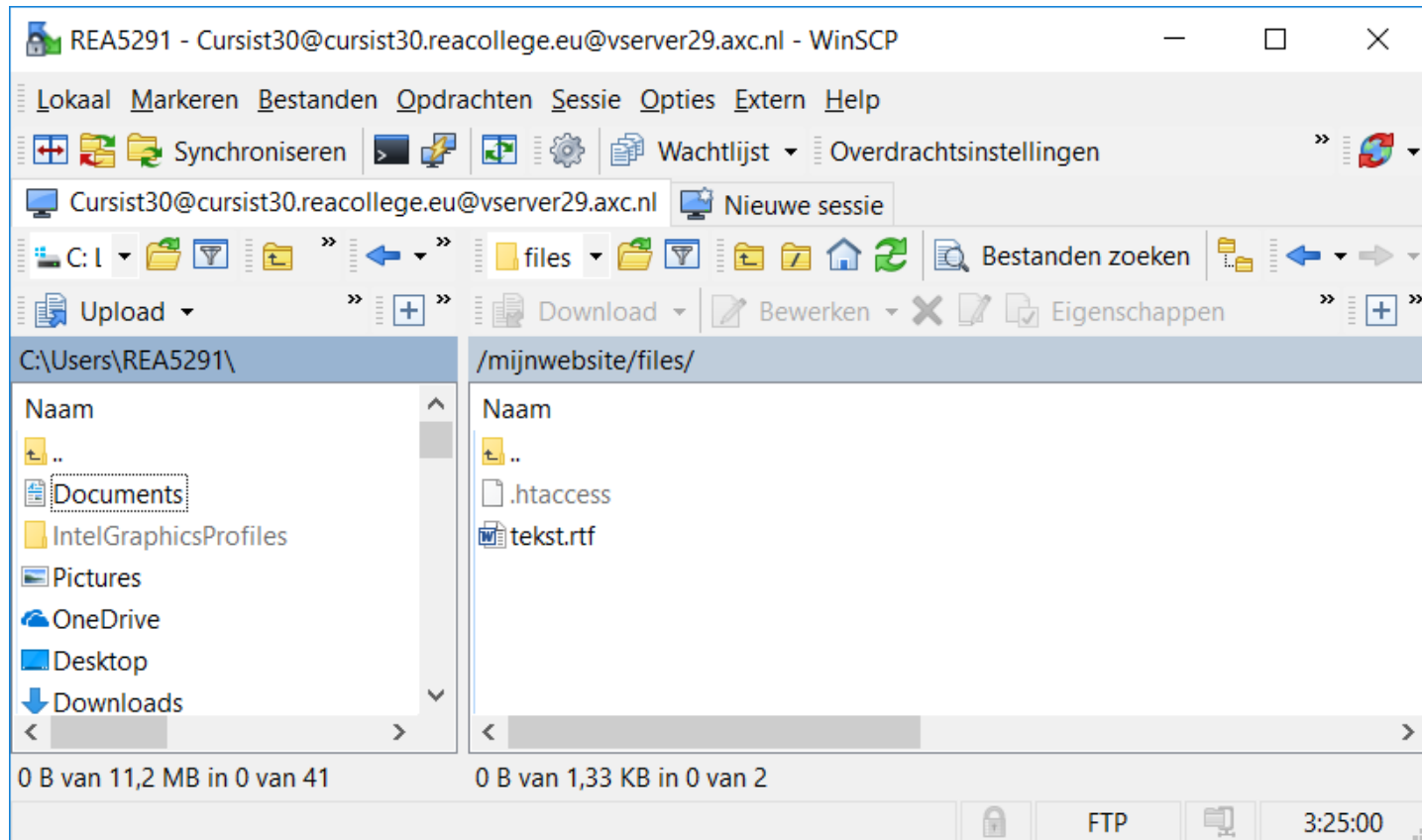
## Canonisatie bestanden

- Je kunt er dus voor kiezen om delen van je content aan te bieden in een andere format dan een traditionele website.



- In een aantal situaties wil je je content ook in meerdere formaten beschikbaar stellen.
- Je zou je kunnen voorstellen dat je je cv hebt gedigitaliseerd als HTML pagina, maar deze ook beschikbaar stelt in .doc en/of .pdf formaat.
- Ook al zijn het verschillende bestandstypes... in de ogen van Google ontstaat er dan toch duplicate content.
- Normaal zou je dat oplossen door dezelfde canonieke URL te plaatsen in alle verschillende bestanden.
  - Het is echter niet mogelijk om een HTML `<link>` element toe te voegen aan een .doc of .pdf tekstbestand.

- Je kunt dit probleem oplossen door een .htaccess bestand te plaatsen in dezelfde map als waarin het desbetreffende bestand zich bevindt.



- Je kunt dit probleem oplossen door een .htaccess bestand te plaatsen in dezelfde map als waarin het desbetreffende bestand zich bevindt.
- Met "HTML-achtige" syntaxis kun je een **<Files>** element aanmaken.
  - Let op: dit is hoofdlettergevoelig.

```
# Het bestand tekst.rtf bevat dezelfde inhoud als de HTML  
# pagina waarin er naar gelinkt wordt.
```

**<Files>**

**</Files>**



- Binnen dit element geef je de naam op van het bestand dat je wil aanspreken.

```
# Het bestand tekst.rtf bevat dezelfde inhoud als de HTML  
# pagina waarin er naar gelinkt wordt.
```

```
<Files tekst.rtf>
```

```
</Files>
```

- Binnen dit element geef je de naam op van het bestand dat je wil aanspreken.
- Vervolgens kun je met de **Header add Link** functie een canonieke URL koppelen aan dit bestand.

```
# Het bestand tekst.rtf bevat dezelfde inhoud als de HTML  
# pagina waarin er naar gelinkt wordt.
```

```
<Files tekst.rtf>  
    Header add Link '<http://www.mijnwebsite.nl/  
    files/tekst.rtf>; rel="canonical" '  
</Files>
```

- Webcrawlers kunnen nu al voordat ze het bestand bekijken zien dat er mogelijk sprake is van duplicate content.
  - Ze zullen dan automatisch worden doorverwezen naar de officiële, canonieke locatie.

```
# Het bestand tekst.rtf bevat dezelfde inhoud als de HTML  
# pagina waarin er naar gelinkt wordt.
```

```
<Files tekst.rtf>  
  Header add Link '<http://www.mijnwebsite.nl/  
  files/tekst.rtf>; rel="canonical"'  
</Files>
```

- Stel je voor dat je een pagina een andere bestandsnaam of URL geeft, maar wel nog steeds wil dat de oude URL (tijdelijk) beschikbaar blijft.
  - Bijvoorbeeld omdat je weet dat je klanten die URL gebookmarkt hebben.

- Je kunt een dan het commando **Redirect 301** gebruiken in je .htaccess bestand om een automatische doorverwijzingen te regelen.

```
# Gebruikers omleiden naar nieuwe pagina  
Redirect 301
```

- Je kunt een dan het commando **Redirect 301** gebruiken in je .htaccess bestand om een automatische doorverwijzingen te regelen.
  - Eerst noem je de oude URL.

```
# Gebruikers omleiden naar nieuwe pagina  
Redirect 301 /keukenapparatuur.php
```

- Je kunt een dan het commando **Redirect 301** gebruiken in je .htaccess bestand om een automatische doorverwijzingen te regelen.
  - Eerst noem je de oude URL.
  - Vervolgens noem je de nieuwe URL.

```
# Gebruikers omleiden naar nieuwe pagina  
Redirect 301 /keukenapparatuur.php /keukens.php
```

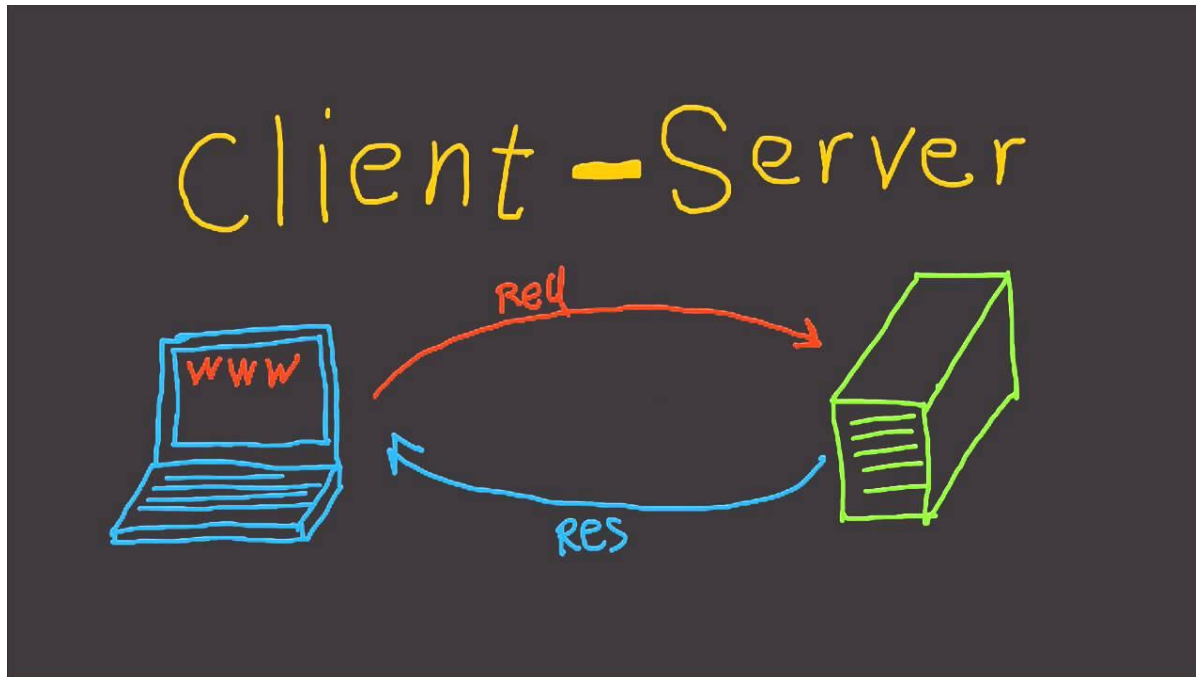
- Je kunt een dan het commando **Redirect 301** gebruiken in je .htaccess bestand om een automatische doorverwijzingen te regelen.
  - Eerst noem je de oude URL.
  - Vervolgens noem je de nieuwe URL.
- Zorg er wel voor dat je bij beide URL's uitgaat van een volledig pad; vanaf de *root* van de webserver – anders zal het mogelijk niet werken.

```
# Gebruikers omleiden naar nieuwe pagina
```

```
Redirect 301 /pad/keukenapparatuur.php /pad/keukens.php
```



- Iedere keer wanneer je een pagina bezoekt of herlaadt of een formulier verstuurt, maak je opnieuw verbinding met de webserver.
- Achter de schermen wordt er dan een verzoek ingediend bij de server.
- De webserver beoordeelt steeds opnieuw of hij aan je verzoek kan voldoen.



- Om aan te geven of het verzoek gelukt is of niet, geeft de server altijd een bericht terug.
- Deze berichten noemen we **HTTP status codes**.
- We maken een onderscheid in 5 groepen codes:
  - 1xx – voor informatie
  - 2xx – voor succes
  - 3xx – voor doorverwijzingen
  - 4xx – voor client-side foutmeldingen
  - 5xx – voor server-side foutmeldingen
- Lijst met alle mogelijkheden: <https://httpstatuses.com/>

- **200** ontstaat wanneer het verzoek gelukt is zonder problemen. Dit is in de meeste gevallen de standaardsituatie.



200  
OK

- **301** ontstaat wanneer een bron *permanent* verplaatst is en alle verwijzingen hiernaar de nieuwe URI moeten gebruiken.



# 301

Moved Permanently

- **400** ontstaat wanneer de server niet begrijpt wat de client wil doen, bijvoorbeeld door een verkeerde syntaxis.



# 400

Bad Request

- **401** ontstaat wanneer een deel van de site is afgeschermd door een gebruikersnaam en wachtwoord, maar de client deze niet heeft ingevuld.



**401**  
Unauthorized



- **403** ontstaat wanneer je geen toegang hebt tot een bepaalde map of bestand, bijvoorbeeld omdat deze is afgeschermd d.m.v. .htaccess.



# 403

Forbidden

- **404** ontstaat wanneer de server de pagina die de client wil bezoeken niet kan vinden (denk aan een typfout of gebroken link).



# 404

Not Found



- **405** ontstaat wanneer je een verzoek uitvoert met een verkeerde method dan is toegestaan (bijvoorbeeld GET terwijl POST verwacht wordt).



# 405

Method Not Allowed

- **410** ontstaat wanneer een bron bewust is verwijderd is en alle verwijzingen hier naar verwijderd dienen te worden.



# 410

Gone

- **429** ontstaat wanneer er te veel verzoeken bij de server zijn ingediend in een te korte tijd.



# 429

Too Many Requests

- **500** ontstaat wanneer er een intern technisch probleem op de server zelf is, bijvoorbeeld door een fatale fout in de PHP code of het .htaccess bestand.



500

Internal Server Error

- **502** ontstaat wanneer er een communicatieprobleem plaatsvindt tussen twee servers onderling, bijvoorbeeld wanneer er een proxy gebruikt is.



# 502

Bad Gateway



- **503** ontstaat wanneer de server tijdelijk niet beschikbaar is, bijvoorbeeld wanneer deze gecrasht is of er onderhoud plaatsvindt.



- **504** ontstaat wanneer een communicatieprobleem tussen twee servers heeft geleid tot een time-out, bijvoorbeeld wanneer een server traag of down is.



# 504

Gateway Timeout

- **508** ontstaat wanneer er een loop wordt gevonden in het verzoek waardoor een aantal processen oneindig door blijven gaan.



**508**  
Loop Detected



# REACOLLEGE HTTP status codes

---

- Wanneer alles goed is gegaan, of je doorverwezen wordt, krijg je de status code niet te zien.
- Wanneer er iets fout gaat (bij de 4xx of de 5xx range) krijg je doorgaans wel een foutmelding op je scherm.



- Dergelijke standaard foutmeldingen zijn echter zeer slecht voor je SEO.
- Google volgt immers letterlijk de hyperlinks op je site om van pagina naar pagina te springen.
- Wanneer de Googlebot op een "lelijke" error pagina komt, zonder een verwijzing terug naar de website zelf, bereikt hij dus een dead end.
  - Het proces van pagina's indexeren houdt dan op!
- Tevens kunnen gebruikers afgeschrikt worden door dergelijke foutmeldingen en daardoor je site verlaten.
- Je loopt dan dus het risico om verkeer en/of conversie mis te lopen.

# REACOLLEGE HTTP status codes

---

- Helaas zijn 4xx en 5xx niet altijd te voorkomen.
- Denk aan de 404 foutmelding – als een gebruiker een spelfout maakt in de URL bereikt hij een pagina die niet bestaat. Daar kun jij niets aan doen.



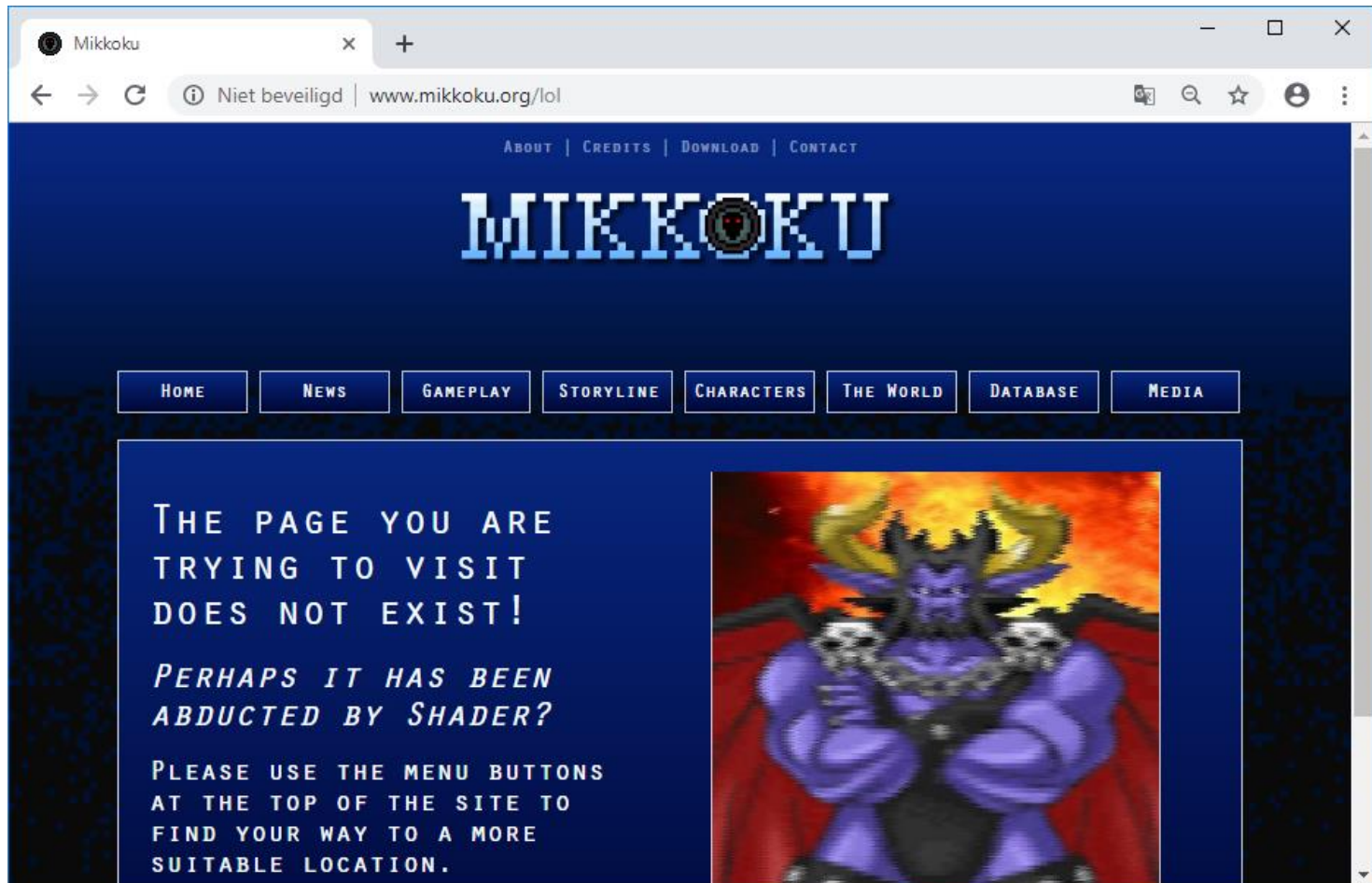
- Om deze problemen op te lossen kun je je eigen errorpagina's maken in plaats van de standaard foutmeldingen.
  - We noemen dat **error documents**.

404  
It looks like you found a  
Dead Link



- Wanneer je errorpagina overeenkomt de layout van de rest van de website (inclusief het menu), of simpelweg een hyperlink naar de 'echte' website, kunnen zowel mensen als robots weer eenvoudig verder gaan met navigeren.
- Wanneer gebruikers een positief gevoel overhouden bij je zelfgemaakte errorpagina's, bijvoorbeeld omdat ze hebben moeten lachen, zullen ze ook minder snel afhaken.

- Enkele voorbeelden van *custom* 404 error documents:

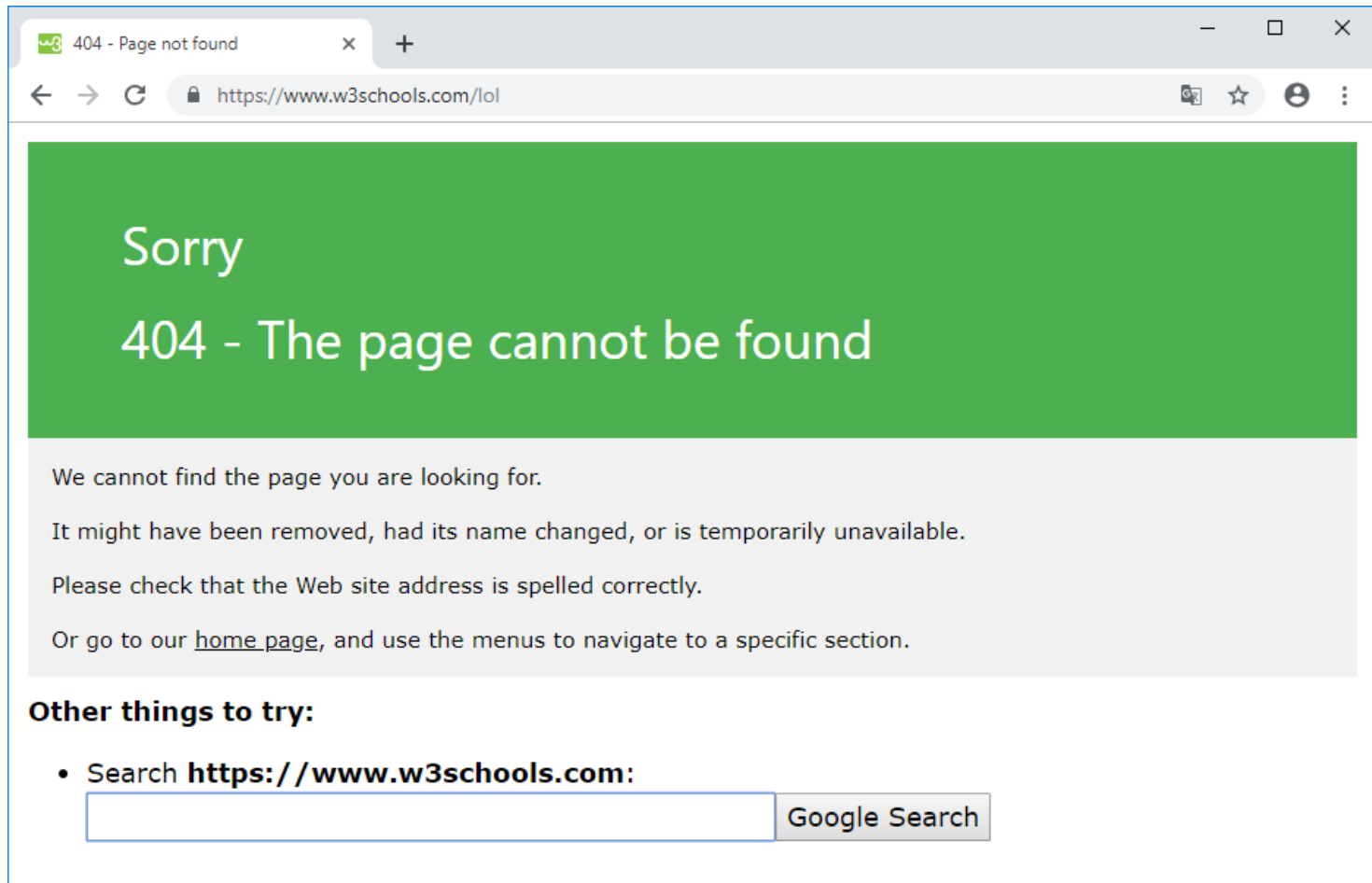


# REACOLLEGE

## Error documents

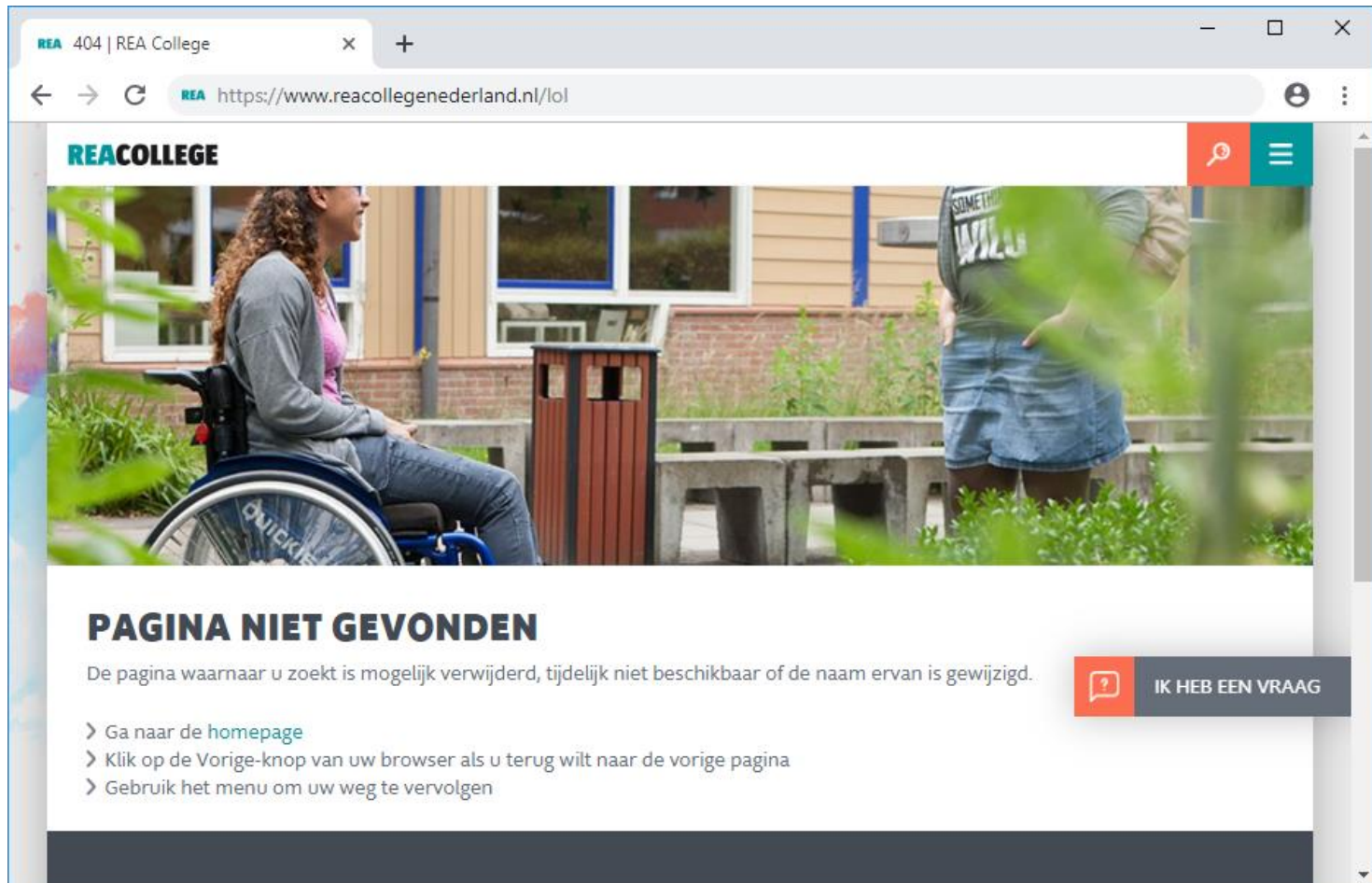
---

- Enkele voorbeelden van *custom* 404 error documents:



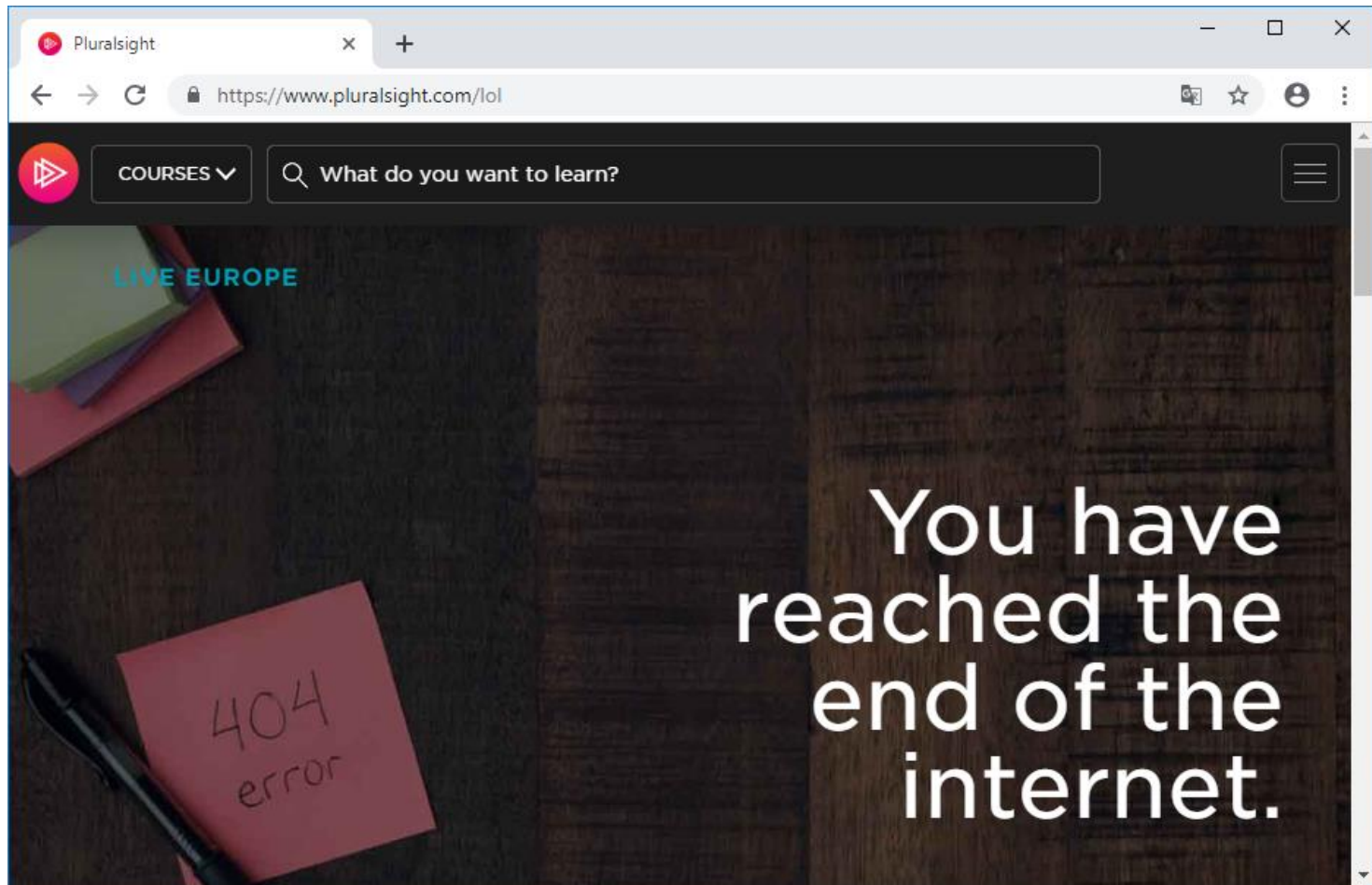
# REACOLLEGE Error documents

- Enkele voorbeelden van *custom* 404 error documents:





- Enkele voorbeelden van *custom* 404 error documents:



- Behalve de **404** error wordt het aanbevolen om ook de **400**, **401**, **403** en **500** foutmeldingen van een custom error document te voorzien.



Access Denied/Forbidden

The page or resource you were trying to reach is absolutely  
forbidden for some reason



Search helpdesk ...

Search Now

[← Back to Dashboard](#)

- Behalve de **404** error wordt het aanbevolen om ook de **400**, **401**, **403** en **500** foutmeldingen van een custom error document te voorzien.



- Zodra je je error documents hebt gemaakt, kun je hier naar verwijzen d.m.v. het commando **ErrorDocument** in je .htaccess bestand.

```
# Verzoek dat de server niet begrijpt
ErrorDocument 400 /kukomo/400.html
# Authenticatie ontbreekt
ErrorDocument 401 /kukomo/401.html
# Verboden toegang
ErrorDocument 403 /kukomo/403.html
# Pagina niet gevonden
ErrorDocument 404 /kukomo/404.html
# Interne server fout
ErrorDocument 500 /kukomo/500.html
```