

Extra's

Extra's bij HTML en CSS - de basis, 3e editie

- ▶ Extra tekst: subscript en superscript
- ▶ Extra tekst: entiteiten

GEEN DOCTYPE

In de meeste browsers kun je de html-code van een webpagina bekijken (zoek naar iets als page source of toon bronversie in een menu dat bijvoorbeeld Weergave of View kan heten). Als u dat regelmatig doet, zult u zien dat er veel webpagina's zijn zonder een document type declaration vooraf. Het wereldwijde web zou in een klap een flink stuk kleiner worden als browsers deze pagina's niet meer zouden weergeven. Gelukkig (of helaas) zijn browsers zeer vergevingsgezind. Ze proberen er het beste van te maken en geven ook een incomplete webpagina weer. Sommige mensen vinden dat jammer omdat daardoor slordig werken niet wordt ontmoedigd.

Pagina's die voorzien zijn van een geldig doctype geeft een browser weer in 'standard-mode', volgens de norm die in de doctype is genoemd. Zonder geldig doctype schakelt een browser over naar de zogeheten 'quirks-mode', een eigenzinnige toestand waarin hij rekening houdt met fouten waarvan bekend is dat ze in oudere versies van browsers voorkwamen. Nadeel van deze toegelijke instelling is dat elke browser zijn eigen keuzes maakt. Als webbouwer kun je er dus niet van op aan hoe een pagina bij een websurfer zal worden weergegeven.

Subscript en superscript

In sporadische gevallen gebruiken we in een tekst subscript en superscript, bijvoorbeeld in een chemische formule of om vierkante meter aan te geven. In html bestaan daarvoor de elementen sub (voor subscript) en sup (voor superscript). Alles wat tussen _{en} staat geeft de browser in een kleinere lettergrootte lager gepositioneerd weer. Wat tussen ^{en} staat Hier is een voorbeeld met subscript en een voorbeeld met superscript:

```
<p>De formule voor ammoniak is NH<sub>4</sub>OH</p>
<p>Een vloer van 20 m<sup>2</sup></p>
```

De formule voor ammoniak is NH₄OH
Een vloer van 20 m²

Afbeelding x.1
Superscript en subscript in een tekst

Dit levert het resultaat op dat in afbeelding x-1 te zien is. Om de symbolen die boven en onder de regel staan goed tot hun recht te laten komen, is het meestal gewenst om de regelafstand (eigenschap line-height) voldoende groot te maken.

Nesten van elementen

Misschien is het nuttig hier te herhalen dat het toegestaan is om verschillende markeringen tegelijk te gebruiken. In de volgende alinea komt zowel het element `strong` voor als het element `sup`:

```
<p>Een vloer van <strong>20 m<sup>2</sup></strong>-</p>
```

In het weergegeven resultaat wordt de superscript 2 ook door het element `strong` versterkt omdat die zich binnen de twee `strong`-tags bevindt. Let er in dit soort gevallen op dat u eerst het binnenste element afsluit voordat u het buitenste afsluit. Elk element moet volledig binnen een ander element vallen.

Entiteiten

Speciale tekens, zoals ç, © en é, hebben oorspronkelijk niet op alle computersystemen dezelfde code, waardoor gemakkelijk een ander teken op iemands scherm kan verschijnen dan was bedoeld. Bij het uitwisselen van teksten tussen verschillende typen computers is dit veelvuldig een probleem. De laatste jaren is de ondersteuning van Unicode op gang gekomen, waarmee een deel van de problemen kunnen worden opgelost. Xhtml is gekoppeld aan Unicode. In het boek hebben we vanaf het begin in het stramien UTF-8 opgegeven als tekenverzameling. UTF is een van de varianten van Unicode. Oudere browsers en besturingssystemen waren nog niet ingesteld op Unicode en daarbij kunnen de bijzondere tekens verkeerd weergegeven worden.

Tekenverzameling

Een 'text only'-bestand bevat geen informatie over de opmaak van de tekst. Uitsluitend de binaire codes voor letters, cijfers, leestekens en codes voor tekstelementen als 'tab' en 'nieuwe regel'. Om de tekst te kunnen uitwisselen tussen verschillende computersystemen moet er een duidelijke afspraak bestaan over welke codering gebruikt wordt. Die coderingsafspraken noemt men de tekenverzameling (in het Engels: *character set*). In de tekenverzameling ligt vast welke tekens ermee kunnen worden aangegeven en welke volgorde wordt aangehouden. Binnen de computerwereld is het lang een rommeltje geweest met die tekenverzamelingen en eigenlijk is het dat nog steeds. Uiteindelijk is een internationale norm afgesproken (ISO-8859-1), maar Microsoft gebruikt een variant daarvan (Windows-1252) die net een beetje afwijkt. Een probleem vormen de niet-westerse alfabetten, zoals het Russisch, Grieks en Arabisch. Daarbij probeert men de tekens van de speciale alfabetten te koppelen aan het westerse alfabet (op de plaats van de a komt de Griekse alfa), maar dat is niet altijd een op een mogelijk. Nog gecompliceerder wordt het met talen die niet op een alfabet zijn gebaseerd, zoals het Chinees.

Unicode

Met een theoretische capaciteit van meer dan een miljoen tekens moet Unicode de definitieve oplossing vormen. Unicode is een internationaal aanvaarde afspraak waarbij elk teken zijn eigen codenummer krijgt. Elk teken een getal en bij elk getal hoort nooit meer dan één teken. De kleine letter a heeft nummer 61, de alfa heeft nummer 945. De nummers kunnen op verschillende manieren genoteerd worden in een bestand, net zoals u het nummer 61 ook kan weergeven als 061 of als LXI zonder dat de waarde verandert. Over de verschillende manieren van noteren zijn ook weer afspraken over gemaakt. UTF-8 is een van de genormaliseerde notaties van Unicode. Het staat voor Unicode transformation format. De letters UTF komt u tegen in het stramien voor de webpagina's van het boek *HTML en CSS, de basis*.

Binnen html is het probleem met de codering van bijzondere tekens opgelost door niet de tekens zelf op te nemen in de html-tekst, maar een omschrijving van de tekens. Die omschrijvingen noemt men entiteiten. Een entiteit begint met een ampersand en eindigt met een puntkomma. Daartussen staat een afkorting van het symbool. Enkele voorbeelden:

■ ç	ç	c cedille
■ Ç	Ç	C cedille
■ ©	©	copyrightsymbool
■ €	€	euroteken
■ é	é	e met accent aigu (Engels: accent acute)
■ í	í	i met accent aigu (Engels: accent acute)
■ ë	ë	e met trema (Engels: umlaut)
■ à	à	a met accent grave
■ ñ	ñ	n met tilde

De entiteiten zijn 'case-sensitive': hoofdletters en kleine letters zijn niet identiek. Kijk maar naar de eerste twee voorbeelden in het rijtje hiervoor. Die verschillen op de tweede letters, bij de een een kleine letters, bij de ander een hoofdletter. Men heeft geprobeerd de afkorting zodanig te maken dat ze relatief makkelijk te onthouden zijn. Als je in het rijtje ziet hoe de e met een trema wordt gemaakt, dan is het niet moeilijk te verzinnen wat de entiteit is voor de letter o met trema: ö

Behalve als entiteit, kan een speciaal teken ook weergegeven worden als een getal in het decimale of het hexadecimale talstelsel. In plaats van € mag je ook € of € schrijven (maar waarom zou je?). Een overzicht van entiteiten vindt u onder meer op www.w3schools.com/tags/ref_entities.asp.

Op het toetsenbord en toch speciaal

Er zijn vier tekens die op de toetsen van het toetsenbord staan en dus eigenlijk niet thuishoren tussen de bijzondere tekens:

- < <
- > >
- & &
- “ "

Ze staan bij de entiteiten omdat ze bij gewoon gebruik in html-code voor verwarring kunnen zorgen. Het kleiner-dan-teken en het groter-dan-teken worden gebruikt bij tags, het ampersandteken & is het begin van een entiteit en het aanhalingsteken wordt binnen tags gebruikt om de waarden van attributen te omsluiten. Om binnen die aanhalingstekens een aanhalingsteken gebruiken, is de entiteitvorm nodig.

Vaste spatie

Er bestaan een paar entiteiten die geen bijbehorend symbool hebben, maar een andere functie vervullen. Het meest gebruikte voorbeeld daarvan is de ‘non-breaking space’ ` ` die een spatie laat zien, maar er tegelijk voor zorgt dat op dat punt niet naar een andere regel wordt gesprongen. Om bijvoorbeeld de woorden Van en Dale bij elkaar te houden op een regel, kunt u in html schrijven: `Van Dale`. Een regel wordt nu nooit afgebroken achter het woord Van.

Een andere nuttige toepassing van de vaste spatie is het euroteken vast te laten zitten aan het bijbehorende bedrag. Door te schrijven `€ 19,99` komt er een spatie tussen het eurosymbool en 19,99, maar zal nooit het euroteken in zijn eentje aan het eind van een regel komen te staan.

OVERZICHT SELECTORS

<code>lp, a {background: #eee; }</code>	zowel p-elementen als a-elementen krijgen een lichtgrijze achtergrond
<code>p a {background: #eee; }</code>	alleen a-elementen die zich binnen een p-element bevinden krijgen een lichtgrijze achtergrond
<code>p.a {background: #eee; }</code>	alleen elementen met de klasse a die zich binnen een p-element bevinden krijgen een lichtgrijze achtergrond
<code>.a {background: #eee; }</code>	alle elementen met de klasse a krijgen een lichtgrijze achtergrond
<code>p#a {background: #eee; }</code>	alleen het element met de id a die zich binnen een p-element bevindt krijgt een lichtgrijze achtergrond