

informatie is de combinatie van CD-ROM, thematisch maandblad en website die professionals in de informatievoorziening 'automatisch' laat meegroeien in de dynamische ICT-ontwikkelingen.

Hoofdredacteur a.i.: drs. Frank Noë,
f.noë@wkths.nl
Eindredacteur a.i.: Ampersand, redactie & productie, Martin Appelman

Uitgever: ten Hagen & Stam bv,
Postbus 34, 2501 AG Den Haag,
Nederland
telefoon redactie +31 (0)70 3045776
e-mail maandblad-informatie@wkths.nl
website www.informatie.nl
Uitgever: Sieds de Boer
Marketing: Stefan Koopmans

informatie is eigendom van het Nederlands Genootschap voor Informatica (NGI) en het Belgische Studiecentrum voor Automatische Informatieverwerking (SAI) en ten Hagen & Stam bv.

Klantinformatie: telefoon +31 (0)70 3046820, fax +31 (0)70 3045880,
e-mail info.ict@wkths.nl
Advertenties: hoofd exploitatie: Henk Jansen
verkoop leider: Rob de Kleijnen
klantenservice advertenties +31 (0)70 3045759
mediaorder: Sylvia Lekkerkerk
telefoon +31 (0)70 3046212

Abonnement en verzending: **informatie** verschijnt 10 x per jaar en 1 x per jaar volledig op CD-ROM. Nieuwe abonnementen kunnen schriftelijk of telefonisch worden opgegeven. Voor het veranderen van tenaamstelling en/of adres graag een gecorrigeerde adresband opsturen. Schriftelijke annulering van abonnementen is mogelijk tot uiterlijk twee maanden voor begin van het nieuwe abonnementsjaar.

• **Nederland:** ten Hagen & Stam bv, klantenservice ICT, Postbus 34, 2501 AG Den Haag, telefoon +31 (0)70 3046820. Abonnement 2001, incl. CD-ROM: f 275/€ 124,79 excl. BTW. Studenten: f 98/€ 44,47 excl. BTW. Losse nummers: f +0/€ 18,15.

• **België:** Kluwer, Santvoortbeeklaan 21-25, 2100 Deurne, telefoon 0800 30143, fax 0800 17529, e-mail: customer@kluwer.be
of via het SAI: prof. dr. J. Vandenbulcke, Bosstraat 41, 9830 St. Martens-Latem, telefoon +32 (0)9 2825530 (privé) of +32 (0)16 326877, fax +32 (0)9 2827944. Abonnement 2001, incl. CD-ROM: Bf 6295 incl. BTW en verzending. Leden van SAI ontvangen **informatie** aan gereduceerd tarief.

Ontwerp: Ivar van Bekkum
Opmaak en zetwerk: LINE UP
tekstproducties bv, Groningen
Druk: Salland de Lange, Deventer
Illustratie omslag: Ingrid Joustra

Artikelen uit **informatie** mogen alleen worden overgenomen, gekopieerd enz. na uitdrukkelijke schriftelijke toestemming van de uitgever.

ISSN 00199907

Uitgeversverbond
Groep uitgevers voor
vak en wetenschap

tenHagen&Stam
UITGEVER



16
XML
in vogelvlucht

Arjan Loeffen

20
XHTML

Steven Pemberton

24
Technische
oplossingen
op basis van XML

Gert van der Steen

30
e-Learning

Rob Koper

36 e.v.
Cases XML

Pieter van der Hijden en Karel Hagemans

46
Softwariemetriek

Joop van der Linden

50
Procesmatig
offreren

Rex Arendsen, Wim Schut & Johan Oldenkamp

Standaardisatie en het World Wide Web Consortium

Het web is ontwikkeld door Tim Berners-Lee en Robert Cailliau. Berners-Lee, een Engelsman, en Cailliau, een Nederlandstalige Belg, werkten bij CERN, op de grens tussen Frankrijk en Zwitserland. In een stoutmoedige daad noemde ze hun systeem, dat toen nog maar twee gebruikers had, het world wide web, en 'the rest is history'. Op een bepaalde manier was het systeem niet erg bijzonder, want er zaten geen nieuwe elementen in die niet al eerder in hypertextsystemen gebruikt waren. Een artikel over het systeem werd zelfs afgewezen voor een Hypertext Conferentie die begin jaren negentig gehouden werd.

De werkelijke kracht van het web lag in zijn integrerende rol. Veel mensen hadden informatie beschikbaar op FTP-servers en het web zorgde ervoor dat deze informatie snel, op een gebruikersvriendelijke en op een elegante manier beschikbaar kwam via een webpagina. De sterke punten waren universaliteit en de aandacht voor de gebruiker. Het werd al snel duidelijk dat het webproject een ander onderkomen

XHTML

Het world wide web is in het begin van de jaren negentig ontwikkeld als hypertextsysteem voor natuurkundige projecten. De angst dat het web binnen de kortste keren zou bestaan uit een vergaarbak van incompatibele documentenformaten leidde tot de oprichting van het World Wide Web Consortium.

Steven Pemberton

dan CERN moest krijgen. Cailliau bleef bij CERN werken, maar Berners-Lee vertrok en zette het World Wide Web Consortium (W3C) op, een consortium van bedrijven en onderzoeksinstituten, gebaseerd op het succesvolle voorbeeld van het X Windows Consortium. Al snel was W3C gevestigd in de Verenigde Staten, Frankrijk en Japan.

De angst dat het web zou vervuilen door incompatibele documentformaten leidde tot het opzetten van het W3C. Netscape had op eigen houtje al eigen elementen aan HTML toegevoegd. Microsoft deed hetzelfde. Het plan was dus om een autoriteit te creëren die voor consensus zou zorgen op het gebied van formaten en protocollen, met als doel dat software van verschillende makers nu en in de toekomst samen zou kunnen blijven werken. De eerste W3C-activiteit was Casca-

ding Style Sheets (CSS). Een belangrijke eerste stap, aangezien de wijzigingen die Netscape en Microsoft inmiddels aan HTML aangebracht hadden, ervoor gezorgd hadden dat HTML verder af kwam te staan van zijn basis: een taal om structuur te bieden. Na deze eerste activiteit heeft het W3C een enorme vlucht genomen. Het is inmiddels uitgegroeid tot meer dan 20 verschillende deelactiviteiten en meer dan 500 bedrijven zijn lid.

XHTML

Een van de W3C-activiteiten betreft HTML. De eerste versie van HTML was onderdeel van het oorspronkelijke web van Berners-Lee en Cailliau. Deze HTML-versie was vrij simpel: er konden geen beelden in weergegeven worden en het werd alleen gebruikt om natuurkundige rapporten te beschrijven. Het werd ontwikkeld op basis van SGML, een

Samenvatting

Het World Wide Web Consortium (W3C) is bedoeld als internetautoriteit op het gebied van formaten en protocollen. Het kent een uitgebreide standaardisatieprocedure. Binnen W3C is onder andere de HTML-werkgroep actief. Die werkgroep is verantwoordelijk voor XHTML, de taal die een brug slaat tussen HTML en XML.

Het standaardisatieproces

W3C activiteiten worden uitgevoerd door werkgroepen.

Working groups bestaan voornamelijk uit werknemers van bedrijven die lid zijn, maar er is ook ruimte voor *invited experts* die gespecialiseerde kennis bezitten over het betreffende onderwerp.

- Eerst wordt een programma van eisen geïdentificeerd dat wordt aangeboden aan de andere werkgroepen voor commentaar.

- De volgende fase is de *Working Draft* die uiteindelijk leidt tot een *Public Working Draft* die buiten W3C beschikbaar wordt. Een *Public Working Draft* wordt regelmatig heruitgegeven op basis van opmerkingen van buiten W3C.

- Dit leidt uiteindelijk via een aantal fasen tot de *Recommendation*, wat betekent dat de standaard klaar is voor publiek gebruik.

- Er zijn twee speciale working groups die als taak hebben alle W3C-documenten te controleren: de *Web Accessibility Initiative (WAI)* en *Internationalization (I18N)* groep.

- Standaardisatie, zoals onder andere door de *International Standardisation Organisation (ISO)* gedaan wordt, houdt zich bezig met het codificeren van bestaande praktijken. Binnen W3C daarentegen bevatten de *Recommendations* (de W3C-standaarden) vaak een nieuw ontwerp en nieuwe onderzoeksresultaten. Het woord standaardisering is dus niet helemaal terecht en het kan veel beter beschreven worden als een open softwaredesignproces.

markup metataal die voornamelijk wordt gebruikt om de structuur van documenten te beschrijven en niet de manier waarop documenten er (grafisch) moeten uitzien.

Toen het web een grote vlucht nam, begonnen Microsoft en Netscape elementen toe te voegen die betrekking hadden op de manier waarop het document er (grafisch) uitziet (zoals de gehekelde <BLINK>-tag). Daarom startte W3C activiteiten om HTML terug te brengen tot haar basis: de structuur. Met het gebruik van de *Cascading Style Sheets* zou de grafische weergave beschreven worden.

De behoefte om op een simpele manier nieuwe elementen aan HTML toe te voegen bleef echter aanwezig. Hoewel SGML gebruikt kon worden om nieuwe talen te definiëren, was deze taal zeer moeilijk te implementeren. Als antwoord hierop kwam W3C met een simplificatie van SGML in de vorm van XML. Het werd makkelijker om een eigen taal op het web te definiëren zonder dat er iedere keer een nieuwe browser ontwikkeld moest worden. Verder zorgde CSS ervoor dat de nieuwe taal zichtbaar werd in de browser.

De vraag die resteert is of het bestaan van XML, met al zijn mogelijkheden tot het definiëren van eigen, nieuwe talen, het voortbestaan van HTML rechtvaardigt. Deze vraag werd gesteld op een tweedaagse workshop in San Francisco in mei 1998. Het antwoord was een duidelijk 'ja': HTML is een zeer bekende en bruikbare lingua franca

op het web en hij heeft belangrijke semantiek, impliciet en expliciet, die onder andere gebruikt wordt door zoekmachines. De uitslag van de workshop was dat HTML gherdefinieerd zou gaan worden in XML, opgeschoond en gemodulariseerd, met nieuwe functionaliteit zoals verbeterde formulieren, event handling en markup om Aziatische talen te ondersteunen.

De nieuwe HTML-werkgroep formuleerde een werkplan om dit te bereiken. De eerste stap was de uitgave van een gecorrigeerde vorm van HTML 4, die sinds hij de *Recommendation*status bereikt had een aantal wijzigingen moest ondergaan. Het resultaat was XHTML 1.0, een herformulering van HTML 4 in XML. In theorie zou deze stap vrij makkelijk moeten zijn geweest, maar in de praktijk bleek XML een nieuwe technologie waarvan nog niet alle aspecten uitgekristalliseerd waren. Ook waren de verwachtingen hoog gespannen omdat het de eerste echt grote XML-applicatie was. Uiteindelijk moesten verschillende punten doorgestuurd worden naar autoriteiten zoals de XML-werkgroep voordat de definitieve versie van XHTML 1.0 opgeleverd kon worden.

Hoewel SGML en XML van elkaar verschillen zijn de verschillen dusdanig klein dat XHTML 1.0 zodanig ontwerpen kon worden dat met behulp van technische kronkels in bestaande browsers, een webpagina geschreven in XHTML 1.0 daar toch getoond kon worden. Als je simpele richtlijnen volgt, is het mogelijk om



documenten in zowel op XML gebaseerde browsers als in oude HTML-browsers te tonen. Dit is dé brug naar XML: iedereen kan nu XML gaan gebruiken omdat de sites zichtbaar blijven in de meeste browsers.

De volgende stap was het definiëren van de modularisatie van XHTML. Een probleem op het web, in het bijzonder voor aanbieders van content, is het voortbestaan van varianten van HTML. CHTML is in Japan bijvoorbeeld gedefinieerd voor mobiele telefoons. Modularisatie is een techniek die gebruikmaakt van DTD's (en binnenkort ook van XML-schema's) waar we geparameteriseerde subtalen, of modules, van XHTML definiëren, bijvoorbeeld voor tabellen. Door het invullen van de parameters wordt een nieuwe taal gecreëerd op basis van de gebruikte modules. Ons doel is om de taalontwerper de vrijheid te geven varianten van XHTML te definiëren (in de vorm van subsets of supersets) met de zekerheid dat als ze een module gebruiken ze er zeker van zijn wat toegestaan is. Met andere woorden, alle XHTML-varianten die gebruikmaken van de tabelmodule gebruiken dezelfde definitie van een tabel. Er blijven weliswaar verschillende talen bestaan maar de verschillen zijn beter hanteerbaar.

De HTML-werkgroep heeft inmiddels twee gemodulariseerde versies van XHTML geproduceerd: XHTML Basic, ontworpen voor mobiele apparaten als telefoons (en binnen drie maanden na release al te krij-

gen op twee handsets) en XHTML 1.1, een opgeschoonde versie van XHTML 1.0 Strict, met de toevoeging van de Ruby markup, een markup om Aziatische talen te ondersteunen. Het idee van modularisatie is inmiddels overgenomen door een aantal groepen binnen en buiten W3C. Een aantal talen, zoals HTML+SMIL (for multimedia markup), HTML+Math (voor wiskunde), die gebruikmaken van modularisatie, staat op het punt van verschijnen. Het WAP-forum heeft inmiddels aangekondigd dat ze van plan is om XHTML Basic in WAP2 te gebruiken.

De toekomst

De HTML-werkgroep bevindt zich nu in de laatste fase van zijn geplande ontwikkeling: het toevoegen van nieuwe functionaliteiten. Een daarvan bestaat uit formulieren. Formulieren zijn essentieel voor e-commerce en de ervaring van de afgelopen jaren heeft duidelijk gemaakt op welke punten formulieren verbeterd kunnen worden. Daarom is een nieuwe werkgroep opgericht die verantwoordelijk is voor de nieuwe forms module. Een ander onderwerp dat opgepakt moet worden is 'event handling' (zoals bijvoorbeeld de mouseover events). Momenteel is event handling 'ingebakken' in HTML met als gevolg dat het toevoegen van nieuwe events alleen mogelijk is door een herdefinitie van de taal zelf. Een nieuwe module moet dus zorgen voor een uitgebreid 'event model' dat het mogelijk maakt om nieuwe event types toe te voegen (bijvoorbeeld voor apparaten als telefoons en webtelevisies die in combinatie met het web gebruikt worden).

Als laatste staat het aanpassen van tekstmarkup in HTML op het programma. Dit is bedoeld om het web meer toegankelijk te maken voor mensen met een handicap (dit impliceert ook mensen die in een

auto zitten en die geen beeldscherm kunnen lezen).

Wat betekent XHTML voor bestaande websites?

In vergelijking met HTML heeft XML voordelen die nu pas duidelijk beginnen te worden. XML, en dus ook XHTML, is nauwkeuriger gedefinieerd dan HTML. Dit heeft tot gevolg dat hij makkelijker te verwerken is en dat er minder verschillen zijn tussen browsers. Door XHTML 1.0 te gebruiken is er de verzekering dat gebruikers met oudere browsers de content kunnen zien. Ook kunnen nieuwe ontwikkelingen op het gebied van XML direct geïntegreerd worden in de website. Het gebruik van XHTML in combinatie met CSS (meer dan 95% van de pc-gebruikers heeft een browser die CSS aankan) in plaats van verouderde tags als en <CENTER>, zorgt ervoor dat de content op de website beter geschikt is om te bekijken met apparaten als mobiele telefoons en PDA's.

Meer informatie

De belangrijkste bron van informatie is de W3C-site: www.w3.org en de Nederlandse poot: www.w3c.nl. Informatie over XHTML is te vinden op: www.w3.org/Markup. Indien je deelnemer bent via een aangesloten bedrijf, kun je de activiteiten van de werkgroep volgen via: www.w3.org/Markup/Group.

Steven Pemberton

is onderzoeker bij het Centrum voor Wiskunde en Informatica (CWI) te Amsterdam. Hij is voorzitter van de W3C HTML Working Group en voorzitter van de W3C Forms Working Group. E-mail: steven.pemberton@cwi.nl. www.cwi.nl/~steven/

11