

Michael van Hartskamp

Venuslaan 405, 5632 HM Eindhoven

michael@vanhartskamp.com

T_EX-rubriek

Het internet op

Moderne wetenschap kan niet meer zonder internet. Tijdschriften communiceren met auteurs via webpagina's, onderzoekers plaatsen hun werk op internet en in het onderwijs wordt steeds meer van het web gebruik gemaakt. Het internet is echter niet ontworpen om wiskunde weer te geven en het omzetten van T_EX in html gaat niet zonder problemen. Michael van Hartskamp, topoloog, beschrijft de op dit moment beschikbare technologie.

Het wereldwijde web is, zoals de meesten weten, in de afgelopen jaren exponentieel in omvang gegroeid. Het web staat voor toegankelijke documenten en bevat daarmee een onschatbare waarde aan kennis. Het is dan ook vrijwel noodzakelijk om wiskunde op het web te plaatsen ten einde de wiskunde die plek te geven die ze hoort in te nemen: de spin in het wetenschappelijke web.

De 'echte' webdocumenten zijn evenwel opgemaakt in HTML (de zogenaamde Hypertext Markup Language), een taal oorspronkelijk door fysici op het CERN in Zwitserland ontwikkeld en nu onder beheer en verdere ontwikkeling van de W₃C. W₃C staat voor World Wide Web Consortium, een groep mensen en bedrijven geïnitieerd door web-uitvinder Tim Berners Lee.

Ondanks de wetenschappelijke/fysische achtergrond zijn het web en HTML van huis uit niet geschikt voor het presenteren van wiskunde. Omdat het web echter dusdanig aantrekkelijk is als publicatiemedium zijn al jaren lang diverse lapmiddelen beschikbaar om toch papers en boeken op het web te zetten. Veel bestaande papers en boeken zijn geschreven in T_EX en nog altijd worden er veel in T_EX geschreven. Het ligt dus voor de hand om eerst naar T_EX te kijken en hoe je van daaruit naar HTML en op het web komt.

De makkelijkste manier is op een webserver een DVI, postscript of PDF-file te plaatsen met een link. De beeldkwaliteit is redelijk tot excellent en het printen levert normaal gesproken precies wat de auteur in gedachten had. Belangrijk nadeel is dat het speciale leessoftware vereist die niet iedereen standaard heeft geïnstalleerd. Om een extra blokkade op de toegankelijkheid te leggen, pakt menig auteur zijn document vervolgens in met een platformspecifieke compressietool. Het werkt allemaal wel, maar met het oorspronkelijke idee 'al bladerend komt men eenvoudig verder' heeft het weinig te maken.

Een andere, soortgelijke manier om wiskunde enigszins op het web te zetten is het gebruik van afbeeldingen. Topology Atlas, bijvoorbeeld, introduceerde eind jaren negentig een zogenaamd AtlasImage formaat. Iedere bladzijde werd omgezet in een redelijk leesbaar GIF-plaatje.

Deze methode wordt tegenwoordig nog slechts zelden gebruikt.

Meer in de geest van HTML zijn de L^AT_EX₂HTML en TtH (T_EX to HTML) vertalers. De twee vertalen L^AT_EX respectievelijk T_EX, al kan TtH ook veel L^AT_EX vertalen. De eerste gebruikt afbeeldingen voor iedere wiskundige formule. Dat ziet er dus meestal niet goed uit. TtH probeert formules te realiseren middels bestaande HTML concepten zoals tabellen. Géén van deze programma's is echter de ultieme oplossing. Het onderliggende probleem is namelijk dat HTML zelf ongeschikt is voor wiskunde, ook al kan men een heel eind komen met de ruime mogelijkheden die het huidige HTML biedt.

Gelukkig is ook binnen de W₃C bekend dat wiskunde eigen eisen stelt. Er is een *Math* werkgroep die zich al enige tijd bezighoudt met wiskunde voor gebruik op het web. Deze groep werkt aan MathML (de Math Markup Language). De huidige versie van de MathML specificatie is versie 2.0 uit 2001. Een tweede editie nadert de voltooiing.

MathML maakt voor gebruik van wiskunde een belangrijk onderscheid tussen presentatie en content. Omdat de presentatievariant het meest aansluit bij wat we kennen zullen we deze eerst beschouwen. Eigenlijk is een voorbeeld het meest geschikt om in met MathML kennis te maken.

```
<mrow>
  <mi>x</mi> <mo>=</mo>
  <mfrac>
    <mrow>
      <mrow>
        <mo>-</mo> <mi>b</mi>
      </mrow>
      <mo>&PlusMinus;</mo>
    <msqrt>
      <mrow>
        <msup>
          <mi>b</mi> <mn>2</mn>
        </msup>
        <mo>-</mo>
      </mrow>
      <mn>4</mn> <mo>&InvisibleTimes;</mo>
      <mi>a</mi> <mo>&InvisibleTimes;</mo>
      <mi>c</mi>
    </mrow>
  </mrow>
```

```

</msqrt>
</mrow>
<mrow>
  <mn>2</mn> <mo>&InvisibleTimes;</mo>
  <mi>a</mi>
</mrow>
</mfrac>
</mrow>

```

Waarschijnlijk valt het de lezer onmiddellijk op dat MathML erg breed-sprakig is. Dit voorbeeld komt overigens uit de MathML-specificatie en het is de welbekende *abc*-formule.

Hier blijkt ook een belangrijk verschil met \TeX . In plain \TeX is de formule:

```
$$x = \{ -b \pm \sqrt{b^2-4ac} \over 2a }$$
```

en in \LaTeX

```
\[x = \frac{ -b \pm \sqrt{b^2-4ac} }{ 2a } \]
```

\TeX was (en is) bedoeld om door mensen te worden gebruikt. De formule invoer bij \TeX is aanmerkelijk korter en ook natuurlijker. Na de eerste syntactische barrières te hebben overwonnen, blijkt in \TeX een formule vrijwel zo te worden ingevoerd als een wiskundige hem zou voorlezen. Dit is bij het \TeX -voorbeeld met `\over` nog natuurlijker dan met \LaTeX 's `\frac`. (Dat \LaTeX toch `\frac` gebruikt heeft te maken met \TeX -interne problemen; met `\frac` is het meteen duidelijk dat de volgende twee termen kleiner moeten worden gezet, bij `\over` blijkt dit pas na de eerste formule.)

Waar \TeX impliciet de *a*, *b*, *c* en *x* in *math italics* zet omdat het variabelen zijn, de +, - en = herkent als *math operators* en cijfers 2 en 4 als *math numbers*, moet dit in MathML expliciet worden gecodeerd door respectievelijk `<mi>`, `<mo>` en `<mn>` om de *tag* te openen, en `</mi>`, `</mo>` en `</mn>` om deze weer te sluiten. Van de tien tekens zijn er dus negen markup.

En er zijn nog meer tags nodig. Natuurlijk zijn er bijzondere symbolen. Deze worden in de bekende HTML vorm gepresenteerd met een `&` gevolgd door de naam van het symbool en een `;`. De `⁢` is natuurlijk een heel goede manier om overhead te genereren. De betekenis van `<msqrt>` en `<mfrac>` spreekt voor zich. Interessanter is `<mrow>` het wordt gebruikt om termen samen te stellen uit losse onderdelen.

Het spreekt voor zich dat mensen dergelijke markup niet willen schrijven noch lezen. Bij MathML is men er van uitgegaan dat de gebruiker niet zelf de formule zal intikken maar gebruik zal maken van speciale conversiesoftware. Dit formaat is juist bedoeld om door machines te worden gelezen en geschreven. Bij de ontwikkeling van MathML is dan ook speciaal gelet op de mogelijkheid van vertalers, en de bewering is dat het heel goed mogelijk is om een vertaler te bouwen die van \TeX naar MathML vertaalt of omgekeerd. Helaas heb ik een echt goede vertaler nog niet gevonden.

Een belangrijke vraag is natuurlijk: zijn er ook browsers die MathML kunnen renderen? Een eerste probleem hierbij zijn de lettertypen die de wiskundige symbolen bevatten. Hier wordt aan gewerkt door onder

andere de AMS. Voorlopig kan men vaak \TeX -fonts of Mathematica fonts gebruiken.

Een belangrijker probleem is de browser zelf. Microsoft Internet Explorer bijvoorbeeld, één van de meest gebruikte browsers, heeft geen eigen ondersteuning voor MathML [1]. Er zijn echter plugins: MathPlayer van Design Science [2] en TechExplorer van IBM die zowel MathML als \TeX aankan [3]. De Mozilla Browser [4] daarentegen ondersteunt MathML zonder dat speciale plugins noodzakelijk zijn.

Hiermee is het laatste woord over MathML evenwel nog niet gezegd. Binnen de HTML, SGML, XML-gemeenschap van Markup-talen heeft men zich gerealiseerd dat opmaak en structuur twee verschillende problemen zijn. In de XML-wereld, XML is de extensible Markup Language, ligt de nadruk veel meer op structuur. Nadat de eerste stap naar structuur was gezet, gaat men nu naar betekenis. De droom van onderzoekers in de webtechnologie is het zogenaamde *semantic web*, een web waar de pagina's een wel gedefinieerde betekenis hebben zodat ze daadwerkelijk 'begrepen' kunnen worden door mens én machine.

MathML is feitelijk óók XML. De content variant, de tweede vorm van MathML, gaat over betekenis en structuur. Omdat MathML zo breed-sprakig is, hebben we hier een klein stukje uitgeknipt:


```

<apply>
  <times/>
  <cn>4</cn> <ci>a</ci> <ci>c</ci>
</apply>

```

We beschouwen de term *4ac*. We zien nu de nieuwe *tag* `<apply>`. Binnen deze tag staat nu een toe te passen operator `<times/>` die ook meteen wordt afgesloten (vandaar de `/` op het eind) gevolgd door de drie termen `<cn>4</cn>`, `<ci>a</ci>` en `<ci>c</ci>`. De termen zijn ook nu weer gecategoriseerd in *content number* en *content italics*. De `</apply>` sluit de toepassing van *times* op de tussenliggende termen af. Naast *times* zijn er meer dan honderd van dit soort operatoren gedefinieerd. De bedoeling van de content MathML is dan ook het overdragen van de betekenis van de formule, en niet enkel de vorm en weergave.

Vaak zullen beide vormen van MathML naast elkaar gebruikt worden. Deze tweede vorm van MathML maakt MathML echter goed geschikt als uitwisselingsformaat tussen wiskundige applicaties zoals bijvoorbeeld Maple en Mathematica. Deze applicaties kunnen vaak MathML importeren en exporteren. Het open formaat betekent dat men ook vrij is om van verwerkingsprogrammatuur te veranderen terwijl bestaande 'formules' kunnen worden meegenomen.

MathML [5] is een goed initiatief om wiskunde terug te brengen als spin in het web. Helaas is de meest gebruikte browser niet in staat om rechtstreeks met MathML om te gaan, en vermoedelijk blijft dat zo. De breed-sprakigheid van MathML maakt het schrijven en lezen hiervan, behalve op kleine schaal, ongeschikt voor mensen. Hulp-middelen zijn daarom een vereiste, zodat wiskundigen hun formules kunnen invoeren zoals ze zijn gewend. De grondige scheiding tussen vorm en inhoud maakt nieuwe uitwisselingen van wiskundige data tussen software mogelijk en opent daarmee wellicht wegen naar interessante nieuwe wiskundige toepassingen van de PC. 

Referenties

- 1 Zie voor meer informatie bijvoorbeeld <http://msdn.microsoft.com/library/en-us/dnwebteam/html/webteam10012002.asp>
- 2 <http://www.dessci.com>

- 3 <http://www-3.ibm.com/software/network/techexplorer>
- 4 <http://www.mozilla.org>

- 5 Meer informatie over MathML is te vinden op <http://www.w3c.org/Math>