
M. Hazewinkel

Technical Word Processing

Inleiding

Een wetenschapper wordt dagelijks geconfronteerd met het probleem hoe zijn gedachten, resultaten en ideeën op papier te krijgen. En dat moet in een zodanige vorm dat het leesbaar is voor collega's. Vroeger ging dat per brief; tegenwoordig gaat nog altijd zeer veel via een informeel brieven en *preprint* circuit. Dat is energie- en tijdverspillend, want de mogelijkheden van *electronic mail* en, als ik het zo noemen mag in deze context, van *DeskTop Publishing* maken efficiëntere mogelijkheden vrij.

In deze schrijfproeve wil ik ingaan op wat een actief beoefenaar van de natuurwetenschappen nodig heeft aan *wordprocessing equipment* om redelijk efficiënt met zijn collega's te communiceren. En om te voldoen aan de eisen van de vele tijdschriften die zo langzamerhand *camera ready* manuscripten eisen.

In dit artikel komt af en toe een formule voor. Maar het is niet de bedoeling dat U die begrijpt. Waar het om gaat is dat U die bekijkt met een kritisch maar verder ongevoelend oog. Dus vanuit een typografisch en *text processing* standpunt, zonder U te bekommeren om de onderliggende structuren en betekenissen.

Achtergrond en motivatie

Oplagen van wetenschappelijke publicaties zijn de laatste jaren drastisch teruggelopen. Oorzaak: algemene financiële nood van universiteiten en bibliotheken en een profusie van boekenseries en tijdschriften gepubliceerd door wetenschappelijke uitgeverijen. Zelfs gerenommeerde series van gerenommeerde uitgeverijen zijn gereduceerd tot oplagen in de orde van een derde van wat ze zelfs maar tien jaar geleden gewend waren.

Andere oorzaak: het uitgeven van boeken op de min of meer traditionele manier door gevestigde uitgeverijen is een dure zaak geworden; althans vanuit het standpunt van de enthousiast die deze boeken graag zou kopen en lezen.

De niet deskundige amateur ziet hier mo-

gelijkheden in termen van *DeskTop Publishing*. Zo eenvoudig ligt dat op dit moment niet. De kosten van de vervaardiging van een boek (auteurs-honorarium, zetwerk, en drukwerk) zijn eigenlijk maar een relatief gering deel van de werkelijke kosten van een middelgrote tot aanzienlijke uitgeverij. Een veelvoud daarvan gaat zitten in acquisitie, management en verkoop. In zekere zin geeft dit nog meer potentiaal aan de *One Man DeskTop Publisher*. Echter, helemaal simpel ligt het niet. Geen van mijn wetenschappelijke vrienden, die zich in deze richting hebben trachten te manifesteren, heeft kans gezien er een echt succes van te maken; geen is er ook aan ten onder gegaan. En het is zeker dat de mogelijkheden op dit moment drastisch gunstiger zijn dan zelfs maar drie jaar geleden.

En er is potentiële ruimte, bijvoorbeeld in de markt voor gerichte teksten op het niveau van de eerste jaren van de universiteit. Het produceren van zulke boeken is vrij duur; aan de andere kant - omdat studenten het moeten betalen - kan de prijs niet echt hoog zijn. De, tot voor kort, enige manier om daar onderuit te komen was met behoorlijk grote oplagen te werken. Dat betekende weer dat het boek niet specifiek gericht kon zijn op de cursusbehoeften en interessen van één bepaalde opleiding of universiteit. Oplossing: maak een boek waarin iedere opleiding wel vindt wat nodig is. Dat maakt het boek dikker en duurder dan nodig is; dus grotere oplagen zijn nodig; dus ... ; en we zitten midden in een vicieuze spiraal.

Nog meer mogelijkheden voor relatief bescheiden *Technical Word Processing/Publishing* zie ik bij wetenschappers en wetenschappelijke instituten. De rol van het *preprint* circuit wordt snel belangrijker en daarbij komt steeds vaker de vraag van tijdschriften en *proceedings* publicaties om *camera ready copy*. Het aantal collega's dat zich een micro aanschaft om zelf in staat te zijn een acceptabel wetenschappelijk manuscript te produceren groeit exponentieel. In een zekere, maar verander-





de, zin gaan we weer terug naar een patroon van wetenschappelijke communicatie gebaseerd op de directe contacten van individu tot individu. Verschil: eens een brief van een persoon naar een andere; nu tientallen of meer copieën verstuurd door één individu naar al zijn belangrijke contacten. Om verschillende redenen vind ik dat een gunstige ontwikkeling.

ML versus WYSIWYG.

Wat is het probleem? *Text processing* of *Word processing*, of hoe U het ook in het Nederlands wil noemen, is langzamerhand een goed ingeburgerde zaak. Er zijn vele goede pakketten die nauwelijks moeite hebben met het incorporeren van grafieken, illustraties, en wat dies meer zij. Een van de werkelijk pijnlijke problemen waar we mee zitten is formules. Om maar meteen met de deur in huis te vallen, zulke dingen als:

$$U_{m,n_1}^{j_1}(\rho_1, \rho_2, \delta_{n_1, n_2}) + \int_0^{\pi} \frac{\delta \rho_1 \rho_2^2}{3\pi^4} \sum_v \sum_m \sum_{\beta_j} (-1)^m \frac{U_{m,n_1}^{j_1,3,3}(\rho_1, \rho_2)}{\pi^2 \cdot k^2} \zeta_{1,m,n_1} \eta_v(\rho_1, \rho_2)$$

Op zich zijn dit gewoon twee dimensionale plaatjes en dus te maken met *MacPaint* of *MacDraw*. Echter slechts met grote moeite en zeker niet consistent in het geval er vele van dit soort dingen in een gegeven tekst verschijnen.

Een formule zoals de bovenstaande heeft zeer veel structuur en daarvan kan gebruik gemaakt worden voor het produceren ervan (intikken) met behulp van een *Formule Processor*. Ook kan zo een veel grotere afdruk-kwaliteit verkregen worden.

Globaal gesproken zijn er twee geheel verschillende systemen om de informatie, nodig voor een formule, in te voeren: *Markup Language* (ML) en *What You See Is What You Get* (WYSIWYG). Als voorbeeld is hieronder het invoerrijtje symbolen in *TEX* gegeven voor de eenvoudige formule die daaronder is afgebeeld.

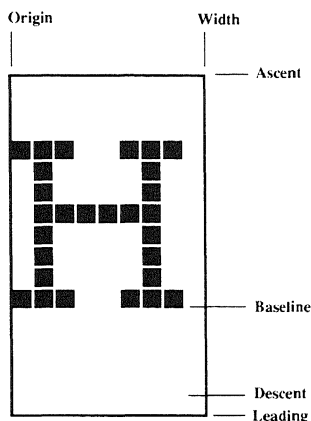
$$\int_0^{\pi/2} \sin^2(\theta) d\theta$$

$$\int_0^{\pi/2} \sin^2 \theta d\theta$$

De formule is gemaakt met een *WYSIWYG Formule Processor* (MacΣqn) en vereiste de volgende symbolenreeks: ⑈-i, 0, ⑈, ⑈-g, p, ⑈-t, /, 2, ⑈, s, i, n, ⑈--, ⑈, ⑈, 2, ⑈-g, q, ⑈-t, d, ⑈-g, q. Hierbij is ⑈-t: 'controle toets ingedrukt' houden en / tikken' en ⑈ betekent: 'de enter toets indrukken'. Dus een reeks van 22 toets-aanslagen voor de formule processor tegen 35 in het ML geval. Dat is echter niet het belangrijkste. Gewichtiger voor type-gemak is dat je bij het gebruik van een *formule processor* zoals deze ziet wat je doet, dat de cursor veelal automatisch gaat naar de plek waar het volgende symbool moet komen, en dat je meteen ziet of het fout of goed zal gaan. Vooral dat laatste is een probleem met ML systemen: het is heel moeilijk om aan een reeks symbolen direct te zien of het de juiste formule zal opleveren en om te zien welke correctie nodig zal zijn. Vele technische typistes en wetenschappers blijven hier permanent zeer grote moeite mee houden, vooral als -zoals vaak het geval is- niet dagelijks met zo'n systeem gewerkt wordt. Naar mijn smaak zijn dus ook sytemen gebaseerd op ML-input niet acceptabel. Dat wil beslist niet zeggen dat ML beschrijvingen van een formule niet bijzonder belangrijk en nuttig zullen blijven. Een zeer grote stap voorwaarts zou zijn een WYSIWYG systeem dat tevens uitgerust is met de mogelijkheid van de ingevoerde tekst de ML beschrijving te genereren. Dit laatste voor communicatie tussen (verschillende typen) computers, en om gebruik te kunnen maken van de zettechnisch fantastische mogelijkheden van zulke ML's als *TEX* en *TROFF/EQN*.

Letters of symbolen en hun 'doosjes'

Voor ik verder ga moet ik een klein beetje vertellen over de basis voor ieder *word-processing* systeem, namelijk de *fonts* en de symbolen (letters, leestekens, cijfers) die daarin voorkomen. In een computer-context is een font een collectie van 256 gerelateerde symbolen. Bij de beschrijving van een font hoort een aantal gegevens zoals *ascent*, *descent*, *leading* en *baseline*. Die zijn hetzelfde voor ieder symbool in het font.



De *ascent* geeft aan hoe ver een symbool boven de basislijn mag komen, de *descent* hoe ver het er onder mag uitsteken; de *leading* geeft aan wat de afstand zal zijn tussen de laagst mogelijke symbool-pixels van een gegeven regel en de hoogst mogelijke van de regel daaronder. Elk symbool in een *font* heeft nog twee andere maten: *origin* en *width*. Samen met *ascent* en *leading* definiëren zij een rechthoek of doosje zoals hierboven geïllustreerd voor de letter 'H'. Een *word processing* programma werkt nu in eerste instantie met deze doosjes door ze horizontaal tegen elkaar aan te zetten met passende basislijn. De verticale afstand tussen twee regels wordt bepaald door het hoogste doosje van de onderste en het laagste van de bovenste van die twee regels.

Hiervan kan op allerlei manieren gebruik gemaakt worden. Wilt U bijvoorbeeld tussen twee paragrafen net iets meer of net iets minder dan een regel overslaan dan kan dat door naar de volgende regel te gaan, een spatie in te tikken van een iets groter of iets kleiner font en weer een regel naar beneden te gaan. Want ook de spatie zit in zijn eigen doosje. Door de basislijn, zeg, vijf eenheden naar beneden te schuiven ten opzichte van het symbool zelf krijgt U een font voor *super-scripts*, en de basislijn vier eenheden naar boven levert een font voor *subscripts*.

Extra eisen aan een TWP

Wat moet een *Technical Word Processor* nu nog meer kunnen dan een gewone? Laat ik proberen een eerste aanduiding te geven.

•1. Er moet een goede formule-verwerker in beschikbaar zijn, in eerste instantie voor het maken van '*display*' formules, diagrammen, matrices en dergelijke. Zulke formules moeten ook als formules bewaard en opgeslagen kunnen worden, voor eventueel latere verdere bewerking.

•2. Het moet goed mogelijk zijn relatief ongecompliceerde formules in de tekst zelf te produceren, die verder door het tekstverwerkend programma behandeld worden als een gewoon woord op één regel. Dit dient mogelijk te zijn tot een graad van gecompliceerdheid van ongeveer de hiernaast gegevens voorbeelden.

•3. Het moet mogelijk zijn een symbool of groep symbolen te selecteren en ze dan horizontaal en verticaal (een klein beetje) te verplaatsen.

•4. Het moet mogelijk zijn in een gegeven paragraaf tekst naast een illustratie te plaatsen. Dat wil zeggen dat een gebied afgezonderd wordt ten behoeve van een illustratie; de tekst van die alinea vloeit om dat gebied heen op de normale wijze van tekstverwerkers.

•5. De formule-verwerker heeft ingebouwde faciliteiten voor het maken van matrices, diagrammen.

•6. Het moet mogelijk zijn allerlei *macros* te definiëren.

•7. Het moet mogelijk blijven (zoals in *MacWrite*) om ad hoc even van font en/of van corpgrootte te veranderen.

•8. Al het bovenstaande moet mogelijk zijn zonder naar andere applicaties te moeten overschakelen.

•9. Er dient (apart) bij het pakket een applicatie te zijn voor het modifieren van fonts en voor het maken van nieuwe fonts en symbolen.

Wat er nu is en kan

De vraag rijst nu wat er zo al beschikbaar is en kan in termen van TWP's voor micro computers. Wat betreft de IBM PC *compatibles* zijn er nogal wat pakketten be-



$$\inf_{x,y \in A}$$

$$\int \frac{f(x) dx}{(x-a)}$$

$$x_{i_1}^p$$

$$\sqrt[3]{a - \sqrt{1-x^2}}$$

$$\overline{1 - \|\vec{x}\|_2^2}$$

$$\overrightarrow{AB}$$

$$\overleftarrow{AB}$$

Σ

schikbaar¹. Geen enkele echter waar ik iets meer van weet voldoet zelfs maar bij benadering aan al de boven geformuleerde desiderata.

Wat de Macintosh betreft is er een soort van paradox aan de hand. Een echt WYSIWYG TWP pakket is er eigenlijk nog niet. Toch vinden vele wiskundigen en andere wetenschappers het niet moeilijk op de Macintosh hun researchteksten te vervaardigen met veelal heel redelijke resultaten. De Macintosh is dan ook hard op weg om in deze wereld misschien wel de populairste machine te worden.

Ik schrijf dit artikel met behulp van *MacAuthor* (versie 1.2) als tekstverwerker, *MacSigma* (versie 1.9) voor de 'display' formules, en met gebruik van een aantal speciale of aangepaste fonts gemaakt met *FONTastic* (versie 2.6). Echter vrijwel alles wat hier ingetikt staat had ook gekund met *MacWrite* in plaats van *MacAuthor*.

Er zijn echter nog een aantal verdere zeer waardevolle mogelijkheden in *MacAuthor* die ons weer een flinke stap verder brengen op weg naar een echt goed WYSIWYG TWP pakket. Een ervan is de 're-terugtoets'. Deze wist niet het voorgaande symbool uit en maakt het zo mogelijk symbolen boven elkaar te plaatsen zoals in een binomiaal coëfficiënt. Het tweede symbool van zo'n combinatie wordt automatisch gecentreerd ten opzichte van het eerste. Dat is vaak prettig en handig, maar niet altijd. Samen met de mogelijkheid tot zeven niveaus van *superscript* en *subscript* geeft dit grote, zij het moeizame, mogelijkheden voor het intikken van formules in de tekst zelf. (Desideratum 2) Om dit met *MacWrite* te bereiken zijn allerlei speciale fonts nodig zoals een *superscript* font met de doosjes van elk symbool één eenheid breed gemaakt en precies links van het symbool zelf geplaatst. Deze noem ik voor het gemak nul-breedte fonts. Het lijkt zelfs dat met een 'niet destructieve terugtoets' en de mogelijkheden tot *super-* en *subscripts* de noodzaak voor al die speciale font-trucs verdwenen is. *MacAuthor*

heeft ook de mogelijkheid 'kerning' te doen (het naar elkaar toe spatiëren van lettercombinaties). In combinatie met de speciale 'breedte nul' fonts geeft dit de mogelijkheid bijvoorbeeld de pijltjes boven een vector te plaatsen waar men wil. Een tweede goede reden om de font-trucs te blijven gebruiken.

Laat ik even het lijstje van desiderata voor een TWP zoals hierboven geformuleerd doornemen voor het pakket *MacAuthor* / *MacSigma* / *FONTastic*.

•1. *MacSigma* is een zeer elegante creatie met prachtige mogelijkheden. De *editing* faciliteiten zijn nog te beperkt en er is eigenlijk geen adequate mogelijkheid formules op te slaan voor later hergebruik in licht veranderde vorm. Er is natuurlijk nog een aantal verbeteringen te bedenken. Het zou bijvoorbeeld prettig zijn om indices en exponenten ook apart te kunnen behandelen naast de mogelijkheid *supersub* en voor intersectie en vereniging zouden wellicht de zelfde mogelijkheden moeten bestaan als voor som en product. Er kleven ook aan de versie die ik op dit moment gebruik (1.9) nog wat smetten, vooral de bijzonder zware en lelijke haken en accolades die verschijnen bij het aanroepen van de *group facility*. Ik hoorde dat deze smet in versie 2.0 verdwenen is. Al met al is *MacSigma* hard op weg een hele adequate formule processor te worden.

•2. Van de zeven voorbeelden die hiernaast nog eens te zien zijn, zijn de eerste, derde en zesde redelijk te doen met behulp van de font-trucs, niet destructieve terugtoets en super- en subscripts; de nadere vier zijn vrijwel hopeloos, speciaal de nummers twee, vier en vijf. Het is niet mogelijk een *MacAuthor* frame zich precies zo te laten gedragen als een 'symbooldoosje' van een font.

•3. Horizontaal verschuiven van symbolen is mogelijk (*kerning*); verticaal zijn er geen mogelijkheden.

•4. Dit is simpelweg niet mogelijk. Volgens de press release van *Microsoft* over *Word 3.0* heeft deze tekstverwerker wel deze mogelijkheid.

•5. Matrices en diagrammen worden

$\inf_{x,y \in A}$

$\int \frac{f(x) dx}{(x-a)}$

$x_{i_1}^p$

$\sqrt[3]{a - \sqrt{1-x^2}}$

$\frac{1}{1 - \|x\|_2^2}$

\overrightarrow{AB}

\overleftarrow{AB}

¹ Voor een recent overzicht kan een evaluatie door de PC Technical Group of the IBM PC Users Group of the Boston Computer Society dienen. Deze is verschenen in de Notices of the American Mathematical Society van Januari van dit jaar.



door *MacΣqn* bepaald goed behandeld. De beperkte omvang die (de input van) een *MacΣqn* formule mag hebben (212 symbolen, spaties etc. meegerekend) veroorzaakt hier nogal eens problemen.

•6. *Styles*, *Headings* en *Highlight Styles* in *MacAuthor* bieden hier prettige mogelijkheden. Maar meer zou aanwezig moeten kunnen zijn.

•7. Dit is een stuk lastiger in *MacAuthor* dan het zou moeten zijn. *Richard Palais* merkt in zijn kolommen in de *Notices* van de *American Mathematical Society* op dat dit aspect alleen al wel eens *MacAuthor* zou kunnen beletten een werkelijk populaire tekstverwerker bij wetenschappers te worden. Zo ver wil ik niet gaan, maar ik beschouw het wel als een handicap.

•9. Geen enkel probleem. *FONTastic* werkt uiterst goed en plezierig. Officieel is *FONTastic* alleen geschikt om fonts te modificeren en te maken voor de *ImageWriter*. Voor fonts voor de *LaserWriter* is er een corresponderend stuk gereedschap genaamd *Fontographer*.

Samenvattend kan wellicht gezegd worden dat de combinatie *MacAuthor* / *MacΣqn* / *FONTastic*/*Fontographer* hard op weg is een volwaardig WYSIWYG TWP pakket te worden, maar dat het dat nog lang niet is. Wel is het naar mijn mening een aantrekkelijke combinatie. Ik weet op dit moment geen betere. Er blijft echter nog zeer veel te doen en te ontwikkelen; ook om de programmatuur betrouwbaarder, sneller en robuuster te maken.

Universaliteit

Het totale aantal symbolen dat op een phototypesetter in een bepaalde grootte beschikbaar is voor wiskundig zetwerk is zo'n 1400. En toch blijkt met grote regelmaat dat er bepaalde gewenste symbolen weer eens niet zijn. Het moet dus mogelijk blijven in een TWP pakket 'ad hoc' symbolen erbij te maken. Evenzo zullen er altijd formule-types

blijven opduiken die een gegeven formule processor niet of nauwelijks aankan, en er dienen dus adequate mogelijkheden te zijn dingen aan te passen, te veranderen en toe te voegen. Dit is ook de filosofie die ten grondslag ligt aan een interactieve, grammatica gestuurde formule verwerker die op dit moment door de groep van *Hans van Vliet* bij ons op het *CWI* wordt gemaakt. Het is een heel stuk eenvoudiger iets in de sturende grammatica te wijzigen of er een mogelijke constructie aan toe te voegen dan het is om in het tekstverwerkende programma zelf veranderingen aan te brengen. Zo komt aanpasbaarheid van deze programma's door de gebruiker weer een flinke stap dichterbij.

Euromath

Wetenschappers communiceren veel met elkaar. Ze wisselen *preprints* uit, schrijven samen artikelen en sturen elkaar brieven. Deze documenten tenderen behoorlijk vol te zitten met formules, diagrammen en dergelijke. *Electronic mail* zou ideaal zijn voor de communicatie op allerlei wetenschapsgebieden. Maar dan moeten wel de formules begrijpelijk mee gestuurd kunnen worden. Dit is een verre van triviaal probleem en een hoofdaspect van een project dat *Euromath* heet. Dit project is er op gericht de circa 10000 wiskundigen van Europa te voorzien van een geïntegreerd pakket dat mathematische tekstverwerking, elektronische communicatie en database faciliteiten omvat. Dit zal worden ontwikkeld voor zowel micro's (met volgens de huidige planning in eerste instantie de Macintosh; later ook andere PC's) als voor veel grotere machines.

Er is veel in ontwikkeling op dit moment in de wereld van *Technical Word Processing* en in onmiddellijke samenhang daarmee *DeskTop Publishing*, iets waar elke wetenschapper die zelf zijn *preprint* fabriceert -en dat worden er steeds meer- eigenlijk aan doet. De komende jaren beloven bijzonder interessant te worden.