BIJLAGE J

Werkgroep 7: PC-zaken

TEX voor MS/PC-DOS PC's

Jos Winnink

September 21, 1990

Abstract

Searching for TEX implementations that were available for MS/PC-Dos PC's several *public domain* versions were discovered. These versions range from a system that only has TEX to a system that competes succesfully with commercially available implementations of TEX. Some indication of the speed of the TEX implementations is given.

Inleiding

Bij mijn poging een TEX implementatie voor de MS/PC-Dos PC's te krijgen bleken er meerdere *Public Domain* versies te bestaan.[1] De beschikbare versies liepen uiteen van één, die weinig meer biedt dan een (goede) implementatie van TEX tot één, die kan wedijveren met commerciëel verkrijgbare versies van TEX. Er werden ook een tweetal losse previewer's gevonden.

In dit artikel worden de gevonden implementaties beschreven. Tevens was dit *onderzoek* een gelegenheid om een vergelijking te maken tussen de snelheid van de verschillende TEX implementaties.

De volgende versies van TEX implementaties en *previewers* worden beschreven:

- DOSTEX
- DVIEW
- DVIVGA
- emTEX, 2 versies
- PUBTEX
- SbTEX, 2 versies
- TURBOTEX, een commerciëel verkrijgbare versie

DOSTEX

DOSTEX is een versie, die bestaat uit:

- AMS-TEX
- LATEX
- TEX 2.93
- previewer voor Hercules beeldschermen (DVI2HERC)
- driver voor Epson FX-80 matrix printers (DVIEPS)
- een beperkt aantal fonts, zowel PK als TFM files

Deze versie is niet erg snel. Het systeem is zeker niet foutvrij, zowel DVI2HERC als DVIEPS vertonen een aantal problemen, waardoor er bijvoorbeeld bij het afdrukken van teksten, die wiskundige formules bevatten, soms iets fout gaat. Ook lijkt er een probleem te zitten in de implementatie van TEX, die zich manifesteert als er mathematische formules in de tekst aanwezig zijn. (TEX blijft dan soms hangen)

Beschikbaarheid: SIMTEL20 archieven

DVIEW

Previewer gemaakt op het MIT door Steve Ward en Ricardo Jenez. Heeft als bijzonderheid dat er slechts gewerkt wordt met fontfiles op de grootte 480 (96 DPI). Is geschikt voor CGA/EGA/VGA/Toshiba T3100/Olivetti M24 schermen.

Beschikbaarheid: SIMTEL20 archieven

DVIVGA

Dit is een previewer speciaal ontwikkeld voor EGA/VGA schermen. Het voordeel is dat ook de source code beschikbaar is. Het systeem heeft veel fonts nodig in tegenstelling tot DVIEW. DVIVGA is niet overdreven snel.

Beschikbaarheid: SIMTEL20 archieven

emTEX

Op dit moment is emTEX het paradepaardje onder de *Public Domain* versies van TEX voor MS/PC-Dos PC's. Het kan naar mijn mening wedijveren met commerciële versies.

De meeste recente versie van emTEX bevat de volgende componenten:

[1] Een demo-versie van TEX, LATEX en DVIEW is samengesteld; zie ook diskette NTG25 en NTG26 (bijlage I).

1. BIBTEX
2. Drivers voor matrixprinters (9 en 24 pins), HP Laserjet, HP Deskjet en screenpreviewer (CGA/EGA/VGA/HERCULES ....), POSTSCRIPT nog niet beschikbaar maar er wordt aan gewerkt
3. LATEX
4. MAKEINDX
5. METAFONT 2.0
6. PICTEX
7. PKedit
8. TEX 3.0
9. TEXCAD
10. WEB

Enkele uitbreidingen die emTEX duidelijk onderscheiden van andere implementaties:

- TEXCAD, een programma waarmee interactief figuren kunnen worden gemaakt en waarvan het resultaat een file is met instrukties voor de LATEX-picture environment.
- PKedit, een programma waarmee PK files kunnen worden *geëdit*.
- het gebruik van zogenaamde *fontlibraries* waarbij PK files in een *library file* kunnen worden opgenomen. Het grote voordeel hiervan is dat fragmentatie van de harddisk voor een belangrijk deel wordt voorkomen. De fragmentatie wordt veroorzaakt door het feit dat een file in werkelijkheid op een schijf een geheel aantal allocatieeenheden inneemt. Op een PC is de minimale allocatieeenheid op een vaste schijf 2 Kbytes. Als we uitgaan van bijvoorbeeld 100 PK file's per vergrotingsmaatstaf en we hebben de vergrotingen `magstep 0` tot en met `magstep 5` beschikbaar dan praten we over zo'n 600 file's. Gemiddeld geeft elke file een verlies van $\frac{1}{2}$ allocatie-eenheid, zodat het gaat om een ongebruikte diskruimte van ongeveer 600 KByte. Hierbij is geen rekening gehouden met ruimte, die de directories in beslag nemen.

emTEX bevat tevens diverse specifieke files voor het duitse taalgebied, zoals afbreekpatronen e.d.

Beschikbaarheid:

via FTP:

rusmv1.rus.uni-stuttgart.de	[129.69.1.12]	(in: soft/tex/emtex)
terminator.cc.umich.edu	[35.1.33.8]	(in: msdos/text-mgmt/TeX/emtex)
eba.eb.ele.tue.nl	[131.155.2.25]	(in: pub/tex/emtex.new)
sol.cs.ruu.nl	[131.211.80.5]	(in: pub/tex/emtex)

grote hoeveelheden bij deze laatste alleen tussen 18.00 en 8.00

floppy disks:

Pieter Bison	BISON@KUNPV1.PSYCH.KUN.NL
ondergetekende	WINNINK@ECN.NL

GUTenberg versie

De GUTenberg diskettes bevatten:

1. BIBTEX
2. TFM-files voor de zogenaamde D-fonts, in afwachting van oplossing van het fontprobleem (TEX3.0 is immers 8-bit geworden) worden bijgevoegd. Deze D-fonts worden via het *virtual font* mechanisme op C-fonts afgebeeld.
3. Drivers voor matrixprinters (9 en 24 pins), HP Laserjet, HP Deskjet en POSTSCRIPT (DVIALW). Screenpreviewers voor CGA, EGA, VGA, HERCULES (CDVI).
4. JOVE (Jonathans Own Version of Emacs), zoals de naam zegt een implementatie van de EMACS editor voor de PC.
5. LATEX
6. TE, een eenvoudige editor vergelijkbaar met de editor van TURBO PASCAL.
7. TEX 3.0, gebruikt wordt SbTEX 30.

Hoewel dit een aardige distributie is moeten in elk geval van elders de PK-files worden verkregen. Omdat METAFONT niet is bijgevoegd kunnen geen eigen fonts worden gegenereerd. Voor opmerkingen over SbTEX zie aldaar.

Beschikbaarheid: groep van franstalige TEXgebruikers (GUTenberg)

PubliCTEX

Implementatie van Klaus Thull (zie o.a.: TUGboat, Vol. 10, #1, p.15-22). Met PubliCTEX had ik problemen met installeren. Vanwege het feit dat het een implementatie is van TEX 2.93 alsmede de verhalen in TUGboat dat het niet zo'n snelle versie zou zijn heb ik pogingen om deze versie aan de loop te krijgen gestaakt.

Overigens is het systeem redelijk geschikt als bron van allerlei TEX zaken, zo bevat het ondermeer de volgende componenten:

Implementatie:		**TEX versie**	**Relatieve snelheid**
MS/PC-DOS:			
	DOSTEX	2.93	0.13 – 0.29
	TurboTEX(1.1a)	2.92	0.13 – 0.27
	PubliCTEX	2.93	?
	SbTEX26	2.98	0.84
	SbTEX30	3.0	0.87
	emTEX	2.99[2g]	1.00
	emTEX (80286 versie)	2.99[2g]	1.04
	emTEX	3.0[3a]	1.14
	emTEX (80286 versie)	3.0[3a]	1.16
Overige:			
	VAX 750 (DEC distr. tape)	2.93	0.75
	SUN 3/60 (P. v. Campen tape)	2.99	3.92
	SUN/SPARC (P. v. Campen tape)	2.99	13.27

Table 1: Snelheidsvergelijking diverse TEX implementaties

1. *CHEMSTRUCT* van Dr. Michael Ramek
2. DVISELEC, een programma waarmee delen uit DVI-files kunnen worden geselecteerd.
3. EPIC, de Extended LATEX picture environment.
4. LATEX
5. METAFONT source's voor o.a. het lettertypes `OCRB` en `PUNK`
6. PICTEX
7. TGRIND, een programma dat listings van programma's in diverse programmeertlen *pretty* kan printen, door te functioneren als *preprocessor* voor TEX.
8. TEX 2.93
9. WEB

Beschikbaarheid: Heidelberg server

SbTEX

SbTEX is ontwikkeld door Prof. Wayne Sullivan in Dublin. Het is een relatief snelle versie maar is beperkt tot enkel TEX. Voor een werkend systeem is veel meer nodig, GUTenberg gebruikt SbTEX 30 als basis voor het door hen gedistribueerde TEX systeem.

Beschikbaarheid: SIMTEL20 archieven

TurboTEX

Over deze implementatie staat mij weinig informatie ter beschikking. Eigenlijk is TurboTEX enkel en alleen opgenomen vanwege de vergelijking in snelheid tussen de verschillende implementaties. Opmerkelijk is overigens dat de geteste versie van TurboTEX erg gevoelig bleek te zijn voor de hoeveelheid beschikbaar geheugen. Wanneer in plaats van 535 Kbyte *slechts* 490 Kbyte vrij geheugen ter beschikking stond dan nam de verwerkingstijd toe met een factor twee.

Beschikbaarheid: commercieel pakket

Snelheidsvergelijking

Om enigszins een gevoel te krijgen voor de snelheid van de verschillende TEXimplementaties is besloten om hetzelfde document telkens door de verschillende implementaties te laten zetten. De resultaten zijn vermeld in tabel 1. Hierbij moet onmiddelijk worden opgemerkt dat het hier niet gaat om een uitgebreide en uitgebalanceerde *benchmark*, maar slechts om een test met de bedoeling om een indicatie van de verwerkingssnelheid van de verschillende implementaties te geven.

Als document is gekozen voor de documentatie van het macro pakket *CHEMSTRUCT* van Dr. Michael Ramek. De overwegingen om dit document te kiezen waren ondermeer:

- de omvang van in totaal 18 pagina's, zodat effecten van het laden van macro's, wegvallen ten opzichte van de benodigde CPU-tijd.
- het feit dat er ruimschoots gebruik gemaakt wordt van macro's om structuurformules van chemische verbindingen te zetten.
- het feit dat *CHEMSTRUCT* bedoeld is om te werken in combinatie met *PLAIN*, zodat eventuele geheugenproblemen, die soms optreden wanneer LATEX wordt gebruikt in combinatie met een ander macropakket, worden voorkomen.

Een enkele opmerking is op zijn plaats. De snelheid van emTEX 2.99[2g], die op 1.0 wordt gesteld, komt overeen met een verwerkingstijd van ongeveer 540 seconden op een TULIP COMPACT II, een XT achtige PC met een 10 Mhz V20.

De vergelijkingen met de VAX 750, de SUN 3/60 en de SUN/SPARC zijn toegevoegd om enigszins een indruk te geven van de prestaties van de PC implementaties ten opzichte van andere machines. Van een echte benchmark (wat dat dan ook moge zijn) is *geen* sprake.

Conclusie

Voor mijzelf is de conclusie inmiddels duidelijk. Nadat ik eerst geprobeerd heb om uit diverse distributies een goed werkende versie van TEX samen te stellen incusief drivers voor printers en beeldscherm alsmede programmatuur als BIBTEX, MKINDEX e.d., bleek het moeilijk te zijn om een compleet werkend systeem te verkrijgen. Dit werk is te vergelijken met het samenstellen van de GUTenberg diskettes.

emTEX daarentegen blijkt een complete implementatie te zijn inclusief METAFONT met een zeer goede coherentie tussen de verschillende delen van het systeem.

Alles overziende lijkt op dit moment emTEX de grote winnaar. Het pakket is compleet in die zin dat er weinig ontbreekt en op het gebied van snelheid is het ongeslagen. Of er commerciële versies beschikbaar zijn die beter voldoen is niet bekend, maar men moet van zeer goede huize komen om beter uit de bus te komen dan de *gratis* versie van emTEX.

emTEX aangevuld met enkele ontbrekende delen zoals AMS-TEX, PICTEX en de NTG style file's kan dan ook dienen als de NTG TEX3.0 distributieversie voor MS/PC-Dos PC's.

BIJLAGE K

Werkgroep 8: SGML-TEX Seminar

Terugblik op organisatie

Kees van der Laan

De voorbereiding kan gekenschetst worden door delegatie en het herhaald terugkoppelen met betrokkenen. Het begon met het aftasten van het idee een SGML-TEX-seminar te laten plaatsvinden. Dan de samenstelling van de commissie, de opzet, selectie van plaats van handeling, taakverdeling, tijdsplanning, etc. Met tenslotte de diverse terugkoppelingen bij veranderingen, c.q. verfijningen, en niet in het minst het uitvoeren van het vele werk.

De commissie bestond uit twee NTG-ers (Kees van der Laan, Theo Jurriens) en twee SGML-ers (Jan Maasdam, Jan Bleeker) als basis. Bij de uitwerking van de opzet zou blijken dat nog enige taken gedelegeerd moesten worden.
De opzet was een 1-daagse seminar te houden over SGML en TEX etc, met cursussen in de week ervoor en erna. Deze basiszaken werden in het najaar van '89 met de besturen en achterbannen besproken. Ook werden de benodigde zalen gereserveerd en principe-afspraken met de kantinedienst gemaakt in verband met lunches en koffie/thee pauzes. In december '89 kwam een opzet klaar m.b.t. cursussen, de locatie, de indeling van de conferentiedag. Ook werden twee raambegrotingen opgesteld: één voor het seminar en één (sjabloon) voor de cursussen. Dit geheel werd besproken met de besturen: NTG (december) en SGML Holland (februari). Elsevier en Samsom hebben toen aangeboden de deelnemers aan het seminar te onthalen op een receptie.
Op grond van de schatting van het aantal deelnemers werd de prijs vastgesteld voor deelname aan het seminar en voor deelname aan de diverse cursussen. Er werd onderscheid gemaakt in de prijs voor leden (ook van zusterverenigingen) en niet-leden. Voor cursussen werd tevens gelet op de internationaal gehanteerde prijsstellingen.) Met de voorgestelde taakverdeling en tijdsplanning werd, na her en der een kleine aanpassing, ingestemd.

Taakverdeling (in hoofdlijnen):

Commissie	Kees van der Laan, Theo Jurriens, Jan Maasdam en Jan Bleeker.
Financiele zaken	Koen Mulder.
Programmaboekje	Gerard van Nes.
PR-zaken	Jan Maasdam, Jan Bleeker.
Demos (Leveranciers)	Theo Jurriens.
Locale logistiek	Kees van der Laan.

Rond dezelfde tijd is ook het RC-RUG geïnformeerd over de mogelijke cursussen, en is overeengekomen nadere afspraken later te maken over:

- welke cursussen precies,
- het aantal deelnemers,
- de gewenste apparatuur,
- de installatie van de te gebruiken programmatuur, dit in nader en direct overleg met de docent.

(TEX etc. programmatuur was reeds operationeel op de VAX en een campuslicentie voor μ-TEX was in het bezit van RC-RUG. Sobemap zou benaderd worden voor het verkrijgen van het gebruiksrecht van hun parser door SGML Holland/MID).

Rond de jaarwisseling werd een 'Call for papers' uitgezonden naar de achterbannen en op de diverse electronische netten, TUGboat, e.d. Veel moeite is besteed aan het benaderen van de diverse (buitenlandse) sprekers.

De organisatie werd bemoeilijkt door ziekte van ondergetekende en doordat Jan Bleeker van werkkring veranderde. Jan Maasdam heeft contact met het RC-RUG ondernomen en zich van de doorgang en steun vergewist. Dieke van Wijnen nam de taken van Jan Bleeker over. Kees van der Laan zou thuis doorgaan met zijn organisatietaken.

Later toen het programma voor het seminar vaststond, is dit, inclusief vermelding van de te houden cursussen, verspreid aan de diverse achterbannen (ook hebben de diverse TEX-gebruikersgroepen in Frankrijk, Engeland en Duitsland hun leden geïnformeerd), verspreid over de diverse e-mail netwerken en is er een persbericht uitgegaan. Ook zijn er diverse tijdschriften benaderd voor plaatsing van een annonce c.q. redactioneel stukje. SGML Bulletin en TUGboat hebben de annonces geplaatst. In een nog later stadium heeft MID in Duitsland voor verspreiding van de diverse aankondigingen gezorgd.

Rond juni zijn er beslissingen genomen door de respectievelijke besturen, over het al dan niet doorgaan van enkele cursussen waarvoor weinig aanmeldingen waren binnengekomen. Wegens het geringe aantal aanmeldingen voor LATEX beginners, LATEX stijlen, en Metafont, zijn deze cursussen geannuleerd. TEX advanced zou (eventueel) gesubsidieerd worden door de NTG; ook werd het honorarium in overleg met de docent, bijgesteld. Om meer deelnemers te trekken, m.n. uit Duitsland, is de conferentietaal rond die tijd gewijzigd in Engels.
Al hoewel in principe hotelreserveringen door deelne-

mers, docenten en sprekers zelf geregeld zouden worden, zijn er enige bemiddelingen geweest, m.n. voor de Amerikaanse sprekers.
Op 28 augustus kon er niet van de standaard voorzieningen gebruik gemaakt worden en moesten er aparte regelingen getroffen worden.

Na het seminar is er een informeel diner gehouden voor de sprekers en de betrokkenen bij de organisatie ter afsluiting.

Enige cijfers

Aantal deelnemers seminar (incl. sprekers):
90 (5 Amerikanen, 2 Engelsen, 8 Duitsers, 1 Zwitser)
Deelnemers cursussen:

14	SGML orientatie	(28 aug)
9	SGML passief	(29–30 aug)
5	SGML actief	(3–4 sept)
16	TEX intermediate	(28-30 aug)
5	TEX advanced	(3–5 sept)

Opmerkingen

De toegang tot de VAX op 28 en 29 augustus ondervond enige problemen omdat de usernummers/passwords niet geaccepteerd werden. (De dag ervoor was onder *docenten*nummer/password de werking geverifieerd.) Ook was het niet altijd mogelijk uitvoer naar de laser printer te 'routen.' Eén van de TEX docenten heeft schriftelijk zijn ontevredenheid betuigd over de logistiek rondom zijn cursus.

Leveranciers demos

Arbortext, MID (Duitsland), TEXnology Inc., en de Universiteitsdrukkerij lieten resultaten zien verkregen met Varityper 600, en helaas was BYTE er niet vanwege problemen met de ATARI TT. Verder verzorgde Wristers/Scholten de boekenstand. Speciaal ter gelegenheid van deze bijeenkomst hadden NTG-ers een PD-MS-DOS tegen floppykosten beschikbaar gesteld.

Tot Slot

De financiële afwerking is nog niet rond en zal door en in overleg met de penningmeesters afgehandeld worden.

De (La)TEX bronteksten van annonces, begrotingen, planning, programmaboekje e.d. zijn op floppy beschikbaar voor het volgend organiserend committee.

BIJLAGE L

Working Group 13: 'Neerlandica'

Vergadering 18 april 1990

Nijmegen

Inleiding

Het begint traditie te worden dat werkgroep 13 enkele weken voor de halfjaarlijkse vergaderingen van de NTG bij elkaar komt. Het gebruikelijke gezelschap, Johannes Braams (PTT Research Neher Lab), Nico Poppelier (Elsevier Science Publishers) en Victor Eijkhout (Katholieke Universiteit Nijmegen, coördinator en notulist), was deze keer (eenmalig?) uitgebreid met David van Leeuwen (Rijksuniversiteit Leiden).

Werkgroep 13 mag reeds na een jaar terugzien op een roemrijk verleden. De vervangingsstijlen voor LaTeX, en de compleet nieuwe briefstijl zijn reeds een feit (hierover is gerapporteerd in TUGboat [1]); een eerste versie van de taaloptie `dutch` was al beschikbaar, en het werk aan de stijloptie `a4` is voltooid (rapportage hierover volgt in TUGboat [2]).

Het voornaamste programmapunt was deze keer de nieuwe versie van `dutch`, die deel uitmaakt van een groter project, `babel` genaamd; het werk van Johannes.

Taalspecifiek: Dutch

De taaloptie `dutch` bevatte reeds parameterwaarden voor taalafhankelijke delen van de standaardstijlen van LaTeX – dit omvat ook de `\date`. Sinds de vorige keer is daar een stuk bijgekomen dat de trema actief maakt; zie de notulen van de vorige vergadering.

Nico had hier grotendeels de macro's uit de stijloptie `german` – zie [3] – voor gebruikt. David had wat kritiek hierop, en al discussierend kwamen nog wat bezwaren aan het licht.

- De gebruikte benaming 'umlaut' is niet optimaal: 'trema' is beter.
- Nico tekent aan dat TUG aanraadt in publicaties over TeX de naamgeving uit de boeken over TeX en LaTeX te hanteren. In het TeX-boek wordt de Nederlandse trema aangeduid met 'dieresis or umlaut accent'. De term 'unlaut' is in zijn ogen dus niet misplaatst.
- De trema moet niet actief zijn in wiskunde-modus.
- Het is niet nodig uitgebreid te testen op welk teken de trema komt: er zijn uitsluitend enige tests nodig voor de speciale gevallen. In principe is een trema door `\discretionary{#1}{-}{\"#1}` te implementeren.

Nico of Johannes zal deze bezwaren verwerken.

Voor de Nederlandse 'aanhalingstekens openen' worden nu de macro's uit `german` gebruikt. Eigenlijk zou dit een aparte ligatuur moeten zijn.

Verder kan `dutch` nu vanuit Plain TeX gebruikt worden.

Taalonafhankelijk: Babel

In overleg met Hubert Partl is Johannes bezig geweest met een poging standaardisatie te bereiken van nationale aanpassingen aan de LaTeX stijlen.

Resultaat hiervan is een stijloptie `babel` die het gemeenschappelijk werk doet van alle nationale aanpassingen. Deze optie wordt automatisch geladen door de verschillende taalopties, en hierin worden de herdefinities van talige termen gedaan. Voorts bevat deze optie het mechanisme om nieuwe talen te kunnen definiëren.

Afzonderlijke taalopties hoeven nu enkel de parameterwaarden te zetten, en kunnen zelf zaken als de actieve trema – zowel in `german` als in `dutch`, maar onderling op enkele punten verschillend – regelen.

Het ziet er allemaal fraai uit en er was geen echte kritiek op. Kritiek geuit tijdens de vorige vergadering is allemaal verwerkt.

Overige punten

Over de optie `a4` is niet gesproken. Na het artikel van Johannes en Nico moet dat punt afgehandeld zijn. Ook de NTG stijlen zijn nauwelijks genoemd. Wel is de noodzaak uitgesproken in het algemeen goede gebruikersdocumentatie voor alle stijlen en opties te verzorgen.

Er komt een artikel door Johannes en Nico over het Babel-project in een nummer van TUGboat in 1990.

Er is, naar aanleiding van opmerkingen van David, gesproken over aangepaste fonts voor het Nederlands. David heeft enkele nieuwe ligaturen gemaakt, met name de 'fk' en de 'fij' ligatuur. Voor allebei geldt dat er vraagtekens te zetten zijn. Bij voorbeeld: de combinatie 'fk' bevat een lettergreepscheiding. Mag een ligatuur dan nog?

Verder zijn het voornamelijk practische problemen van standaardisatie die wijder gebruik van deze ligaturen verhinderen. Een verder punt is dat David deze tekens in codes boven de 127 gestopt heeft, iets waar veel printeraansturingsprogramma's niet tegen kunnen. Dit zou opgelost kunnen worden door – wat David ook al gedaan heeft – enkele weinig gebruikte tekens uit de Computer Modern fonts als de 'æ' en 'œ' te vervangen door de 'fk' en 'fij' ligaturen. Met TEX 3.0 en de bijbehorende familie printeraansturingsprogramma's van Nelson Beebe moeten alle problemen hoe dan ook van de baan zijn.

Conclusie

Werkgroep 13 kan zichzelf nog niet opheffen, maar als de huidige projecten afgerond zijn zal er niet veel meer te doen zijn dan onderhoud plegen – althans, dat is wat de werkgroep voorziet. Victor is op deze toestand al een voorschot aan het nemen: hij heeft in het afgelopen half jaar enkel reparaties verricht aan de vervangingsstijlen.

Overigens worden de activiteiten van werkgroep 13 in dit tijdsgewricht wel interessanter (lees: lastiger) gemaakt door het feit dat zowel TEX als LATEX een revisie ondergaan. Misschien kan werkgroep 13 dus nog wel enige tijd vooruit.

References

[1] Johannes Braams, Victor Eijkhout en Nico Poppelier, *The development of national LATEX styles*, TUGboat 10 (1989) #3, 401–406.

[2] Johannes Braams en Nico Poppelier, *Babel, a multilingual style-option system for use with the standard document styles of LATEX*, te verschijnen in TUGboat vol. 11 (1990).

[3] Hubert Partl, *German TEX*, TUGboat 9 (1988) #1, 70–72.

BIJLAGE M

SGML-TEX conference
August 31, 1990
Groningen, The Netherlands

David Osborne

Buses and wierdness in Groningen

For their second full-day international meeting, the Nederlandstalige TEX Gebruikersgroep (NTG) organised, in conjunction with the Dutch SGML Users Group, a conference intended to focus interest on the use of TEX and SGML together. On August 31st, approximately 100 delegates from both 'camps' attended the day's events in Groningen, with what seemed a good balance between SGML-ers and TEXies.

TEX– and SGML streams

The only drawback was that most of the day was organised in two parallel streams of presentations and these tended to be a 'TEX stream' and an 'SGML stream'; consequently, delegates often chose to hear talks on the package with which they identified more closely – at least, this delegate did. So, one came away with a still hazy appreciation of the alternative approach to markup. I hope that this situation will improve in future and that staunch supporters of SGML and of TEX will each come to understand why their opposite numbers are so excited by *their* favourite method of tackling the markup problem. With that said, you may forgive me if the following account concentrates on TEX-related issues. Malcolm Clark, who was also at the conference has added some accounts (in square brackets) where he went to a different presentation.

Vendors

[There were some demonstrations from vendors, chiefly *The Publisher* from both TEXcel and MID Information, both running on Sun workstations. The local bookshop also had an impressive display of books for sale at the conference.]

Gerard Kempen

The day started with an introduction by Kees van der Laan, who, as President of NTG, had been instrumental in organising the meeting, with the assistance of members of the NTG and the SGML-Holland users group. In the opening talk, prior to the parting of the ways into the parallel sessions, Prof. Gerard Kempen, of the University of Nijmegen, discussed *Language technology and the future of text automation*. This was an account of some of the work in the Language Technology Project at Nijmegen, and focussed on methods for enhancing the transportability of text between systems including SGML, TEX and ODA/ODIF by separating content and form, logical structure and layout. The approach used employs object-oriented methods for parsing the syntax of sentences. Prof. Kempen described the use of 'Corrie', a module for checking and correcting Dutch spelling, which is based on the syntactic and morphological parsing of text. Other modules used in the project allow the inflection of known words and attempt to detect typos in unknown words, by analysing spelling and missing letters. Sentences are also scanned for idioms. In this way, parse trees are built which describe the hierarchical structure of sentences, storing not the words themselves but pointers to a standard lexicon. In the current implementation, it is possible to handle the rearrangement of subtrees graphically by means of a WIMP interface. It is hoped to develop a 'grammar spreadsheet' to allow the automatic propagation of word changes; for example, pluralising one word will result in appropriate changes in other parts of the sentence. The advantages of the approach taken in the project are claimed to be easier editing, enforcement of spelling standards, and applications including semi-automatic translation into other languages as well as speech, information retrieval, and hypertext.

Joop van Gent

As the first talk in the TEX stream, Joop van Gent, of the Institute of Language Technology and AI, Tilburg, spoke on *Two Faces of Text*, describing document retrieval and TEX. He described an approach to the structure of published documents which considered two types of structure: logical structure and typography. While automatic methods can be used for scanning the logical structure of a document, human readers tend to use typographic cues in the typefaces and layout of a document to locate information. It is difficult to imitate this search strategy in computer systems since it requires pattern recognition on the typeset page image. Only the typesetting software can 'know' the relationship between logical parts of a document and the typeset output. Work is underway to link these by storing the positions of logical page-elements. At Tilburg, this has been done by adapting TEX to produce a logical structure tree and predicates in a format which can later be searched to locate parts of a document on the typeset page. The documents must be parsed while processing a search query, but the query language, EL/DR, can be interpreted or compiled by Prolog, allowing the use of natural language queries. Some knowledge of the document structure, as nested lists or tree structures with well-defined properties, is used to make basic operations trivial.

Malcolm Clark

The next speaker was Malcolm Clark, who addressed *Exchanging documents: a TEXnical perspective*, briefly reviewing document 'views' – typographic versus logical structure; content, which may not be orthogonal to the structure; and revisable versus non-revisable documents – and the requirements for document exchange, either within or between groups, perhaps over heterogeous networks. In the TEX community, the latter need is met to a greater or lesser extent by the use of electronic mail as a transmission method, and Malcolm discussed some of the problems which can arise. With the new version of TEX, providing support for 8-bit character sets, source files for documents must be somehow encoded into 7-bit data for transmission using *email*, and Malcolm advocated the provision of a 7-to-8-bit filter, written in WEB, as a standard tool in the TEX toolbox. This would enable the exchange of TEX source files, representing revisable documents, and DVI files, when compressed and encoded to ASCII, as non-revisable documents.

Since Malcolm had not used all of his allotted time, he was persuaded by Kees van der Laan briefly to discuss the use of graphics in TEX, to fill the remaining minutes before lunch.

Frank Mittelbach

Frank Mittelbach had been scheduled to conduct a 'work lunch' session on the new features in LATEX 3.0, but his talk was instead re-arranged into a third mini-stream at the start of the afternoon session. I chose to attend that, since it is more relevant to my work, thereby regretfully missing Amy Hendrickson's talk on TEX macro techniques, which I look forward to hearing at the Cork meeting. With Frank's talk over-running, I missed most of Victor Eijkhout's discussion of *The document style designer as separate entity*, but heard Johannes Braams' description of *The Dutch national LATEXeffort*. This included a discussion of replacements for the `article` and `report` document styles to produce documents with a more European appearance than the LATEX book. A new contribution from the Netherlands in this area is 'Babel', a multilingual document style for the standard LATEX styles. This replaces the language-specific headings, hyphenation patterns and explicit hyphenation points with support for more than one language in a document and allows easy switching between them. Usable with plain TEX or LATEX, each language can have a separate option file. When used with LATEX, the language can be selected on the `\documentstyle` command.

Sake Hogeveen

Sake Hogeveen, *Aspects of Scientific Publishing*, is an astronomer who also works with some publishing houses. He claimed to be presenting an introductory course in TEX in five minutes and a LATEX introduction in six! This seemed a bold claim which I could not resist. Basically he tried to argue through the markup route from genuine' copy editor mark up through to TEX typographic markup and then to structural markup. No great conceptual leaps there, but since it was addressed to an audience made up mainly of SGML people – that is, suits and ties – it had a useful function. One of the things to be learned fromn this meeting was the almost complete ignorance of the TEX world for SGML, and the equal ignorance of the SGML world for TEX and LATEX. Neither position is helped by our inability to speak to one another at the same level. Sad. Sake contended that good typography supports structure. Few could argue with that. He also seemed to be advocating that we should be prepared to get our fingers dirty by changing style files. This seems like a bad idea to me. LATEX style files are tortuous to 'amend': it often seems to me that this is a good thing. The average document producer is not a document designer, and his or her results can be appalling.]

Victor Eijkhout & Andries Lenstra

Victor Eijkhout & Andries Lenstra described a system, written in TEX which allowed a 'document style designer' to develop styles, without having to know TEX. They use a fairly free syntax to define constructs, placing the designer in a position between format writer and document producer. It was not at all clear to me whether this existed as a real product, or whether it was a set of stubs, full of good intentions. But even as a prototype it introduced many ideas which could be fruitful.]

Barbara Beeton

A fascinating trip down memory lane was then given by Barbara Beeton, who described ten years of *TUGboat production: TEX, LATEX and paste-up*, with a look at the pitfalls and rewards of editing TUG's journal, interspersed with sideways glances at fragments of TEX hagiography. A rich goldmine for devising a TEX trivia quiz!

Manfred Krüger

The streams were finally drawn together for a talk by Manfred Krüger of MID, who discussed *SGML and*

TEX: Two core modules of information processing. He highlighted the misunderstandings that SGML-ers and TEXies have of each other's worlds, claiming that the question, 'Which is better?', is not a useful comparision, since SGML 'does nothing', yet TEX is 'a solution for everything.' Publishing requires the use of SGML *and* TEX. A slightly heated discussion between speaker and audience resulted when it was suggested that the coupling of SGML marked-up documents with the new LATEX could be achieved by making LATEX macros read the SGML reference syntax, since this could allow an interface for the specification of processing instructions in a structured and standard way (cf. the results of the DAPHNE project).

Richard Southall

At the conclusion of this talk, the faint-hearted departed for the bar and the cocktail party. However, those with a thirst for knowledge who stayed were rewarded by a discussion of *Presentation rules and rules for composition in the formatting of complex text* by Richard Southall, now of Xerox EuroPARC. Richard began by demonstrating that both SGML and LATEX have missed aspects of the highest quality typesetting which can be achieved manually. While generalised markup can define 'content objects', that the presentation rules define the layout is only *sometimes* true. He then took examples from entries in *Math Reviews*, published by the AMS, having been recently involved in a complete redesign of the layout and typography. The markup in the MR entries is almost exhaustive, yet the entries before the redesign project show 'wierdnesses' in the typography since presentation rules are being applied to the whole entry, both in the body of the review proper and in the bibliographic details. Former colleagues from Reading University made a structural analysis of the whole journal, while Richard concentrated on the structure of individual entries. He found examples of poor spacing in authors' and journal names in the bibliographic references, leaving large spaces ('buses'... large enough to take a London omnibus) which produce barely perceptible interruptions when a reader scans the line. Since these entries frequently contain abbreviations and punctuation, he offered quotations from typographer's handbooks by Tchischold, DeVinne and others illustrating the presentation rules for punctuated text. These seemed to indicate that TEX's so-called 'french spacing' should be applied in such circumstances, since traditional spacing between sentences is that of normal word spacing. Lastly, the concept of 'compositional environments', in which a set of presentation rules apply, was discussed. In electronic formatting systems, these seem to have first been supported in *Scribe*, which provided delimited areas for such rules. In TEX, there are two clear compositional environments (modes): inline and display math. In other environments, implied or procedural rules are used, such as the spacing in quoted quotes, for example. The problems in the first TEX version of *Math Reviews* were mainly due to the application of the same presentation rules over all the content objects in each entry. Once appropriate compositional environments were introduced, it was interesting to see from the finished examples that the entries were more readable, yet occupied less space. Following a short question and answer session, and with our thirst for knowledge, if not beer, now slaked, we retired to the bar to join the end-of-conference festivities.

(Note added by editor)
Course along the meeting:

- SGML orientation (1 day)
- SGML passive (2 days)
- SGML active (2 days)
- TEX intermediate (3 days)
- TEX advanced (3 days)

- Arbortext, and Krüger MID
- TEXnology Inc.
- Bookstand Scholten/Wristers
- Groningen University Press showed products made by Varityper 600.

Program TEX & SGML Conference

9:00	Welcome, registration, coffee
10:00	Opening
10:15	•**A1**• Language technology and the future of text automation *Gerard Kempen (KUN)*
11:00	Coffee

	Room ZG15	Room ZG8
11:30	•**T1**• Two faces of Text[1] *Joop van Gent (KUB)*	•**S1**• SGML at the European Patent Office *Terry Rowlay (EPO)*
12:15	•**T2**• Exchanging documents: a TEXnical perspective *Malcolm Clark (ICRF)*	•**S2**• Aspects of a marriage: SGML and TEX - an example from real life *Tibor Tscheke (Univ. Wurzburg)*

13:00	'Work' lunch •**T7**• Towards LATEX 3.0[1] *Moderator: Frank Mittelbach & Rainer Schöpf*	Lunch

	Room ZG15	Room ZG8
14:00	•**T3**• Getting TEXnical: Insights into TEX Macro Writing Techniques[1] *Amy Hendrickson (TEXnology Inc.)*	•**S3**• Aspects of Scientific Publishing *Sake Hoogeveen (UvA)*
14:45	•**T4a**• The document style designer as separate entity[1] *Victor Eijkhout, Andries Lenstra (KUN)*	•**S4**• SGML Applications in the Building and Construction Industry *Hin Oey (TNO)*
15:15	•**T4b**• The Duth national LATEX effort[1] *Johannes Braams (PTT Research Neher Lab.)*	

15:45	Tea

	Room ZG15	Room ZG8
16:00	•**T5**• TUGboat production: TEX, LATEX and paste-up[1] *Barbara Beeton (TUG/AMS)*	•**S5**• SGML and TEX at Elsevier Science Publishers[1] *Jeroen Soutberg (ESP)*

16:45	•**A2**• SGML and TEX: Two core modules in information processing *Manfred Krüger (MID)*

	Room ZG15
17:30	•**T6**• Presentation rules and rules of compo- sition in the formatting of complex text *Richard Southall (Rank Xerox)*

17:30	Cocktail party

All day: Vendor/Demonstration booths

[1] Conference material included in this report