

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 \TeX ies.

\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 \TeX cel 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 \TeX technical 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 heterogeneous 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 \LaTeX effort*. 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 from 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 *TUG-boat 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 \TeX ies have of each other's worlds, claiming that the question, 'Which is better?', is not a useful comparison, 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 Tschold, 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
- \TeX nology Inc.
- Bookstand Scholten/Wristers
- Groningen University Press showed products made by Varsityper 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 <i>Joop van Gent (KUB)</i>	●S1● SGML at the European Patent Office <i>Terry Rowlay (EPO)</i>
12:15	●T2● Exchanging documents: a \TeX nical perspective <i>Malcolm Clark (ICRF)</i>	●S2● Aspects of a marriage: SGML and \TeX - an example from real life <i>Tibor Tscheke (Univ. Wurzburg)</i>

- 13:00 ‘Work’lunch
●T7● Towards \LaTeX 3.0¹
Moderator: Frank Mittelbach & Rainer Schöpf
- Lunch

	Room ZG15	Room ZG8
14:00	●T3● Getting \TeX nical: Insights into \TeX Macro Writing Techniques ¹ <i>Amy Hendrickson (TeXnology Inc.)</i>	●S3● Aspects of Scientific Publishing <i>Sake Hoogveen (UvA)</i>
14:45	●T4a● The document style designer as separate entity ¹ <i>Victor Eijkhout, Andries Lenstra (KUN)</i>	●S4● SGML Applications in the Building and Construction Industry <i>Hin Oey (TNO)</i>
15:15	●T4b● The Dutch national \LaTeX effort ¹ <i>Johannes Braams (PTT Research Neher Lab.)</i>	

- 15:45 Tea

	Room ZG15	Room ZG8
16:00	●T5● TUGboat production: \TeX , \LaTeX and paste-up ¹ <i>Barbara Beeton (TUG/AMS)</i>	●S5● SGML and \TeX at Elsevier Science Publishers ¹ <i>Jeroen Southberg (ESP)</i>

- 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 composition in the formatting of complex text <i>Richard Southall (Rank Xerox)</i>

- 17:30 Cocktail party

All day: Vendor/Demonstration booths

¹ Conference material included in this report