Cahiers GUTenberg

TEX AND SGML

Kees VAN DER LAAN

Cahiers GUTenberg, nº 5 (1990), p. 74-81.

<http://cahiers.gutenberg.eu.org/fitem?id=CG_1990___5_74_0>

TEX and SGML*

C.G. VAN DER LAAN

Rekencentrum RUG, Landleven 1, 9700 AV, Groningen, The Netherlands. 050-633374/8080
e-mail: CGL@RC.RUG.NL, *(Formerly:* CGL@HGRRUG5)

1. Setting the scene

Lifecycle-phases of documents

- preparation
- distribution
- reading
- storing (Paper? Electronically? Optically?)
- other usage, reuse?

SGML supports the *complete Lifecycle*, where FUTURE usage of the document is not necessarily restricted to printing.

TEX supports formatting and electronical exchange.

2. What is SGML?

It stands for

Standard Generalized
Markup Language

For the definition see, [20]. An introduction is [8], and courseware is [11]. A Dutch chapter of the SGML Users Group exists.[1]

2.1. Purpose

To facilitate INFORMATION exchange

— *Then and There* —

via a description LANGUAGE, where information is packed in documents, containing, text, graphics, ...

2.2. *META* LANGUAGE

SGML is a *META* LANGUAGE which can be used to define an arbitrary number of markup languages in a standardized way.

2.3. Markup

Formerly: (typeset) MARKS in the margin (Marks are bound to a version; no 'data-integrity')

Presently: Marks are integrated with copy (Note: Discriminate copy from MARKUP! Data-integrity is preserved.)

Markup $\stackrel{\text{def}}{=}$ Term used to describe codes added to the electronically prepared document

2.4. Generalized

Formerly: (typeset)MARKS for *specific* 'here and now' printers

Presently: Marks are *generic* (Not specific to print/plot/photoset hardware)

Generalized $\stackrel{\text{def}}{=}$ Abstraction from the specific to the general to describe the structure of a document and to specify intent without regard for appearance

*Paper presented at: 2e SGML Holland users group seminar Amsterdam, GUTenberg'90 Toulouse, EuroTEX Cork 90.

[1] SGML-Holland Secretary: D. van Wijnen, Wolters Kluwer. P.O. Box 989, 3300AZ Dordrecht. 078-334933. e-mail: surf003@kub.nl
Pour la communauté francophone, s'adresser au Syndicat National de l'Édition (ndlr).

2.5. Standard

Formerly: no consensus on mark-up 'codes' (wordperfect, wordstar, applewrite, ...; Scribe, TEX, LATEX, ...)

Presently: SGML ISO standard

Standard $\stackrel{\text{def}}{=}$ It can be used to define an arbitrary number of markup languages in a *standardized* way.

Entails: general applicability,
longer lifetity,
improved reusability,
enhanced exchange possibilities.

2.6. Example markups

2.6.1. No markup

```
TeXAsystemforformattingtextTeX
andtheaccompanyingmacropackage
LaTeXprovidepowerfulmeans ...
```

2.6.2. Presentational markup

```
TeX:
A system for formatting text.

  TeX and its accompanying macro
package LaTeX provide powerful means of
formatting text to be output on either
  - a simple matrix printer,
  - a laser printer or
  - a photo typesetter.
```

Nice in this context is poetry, e.g., Alice's mousetail, [12], or DEK's favourite poem of Piet Hein, [24].

2.6.3. Procedural (LATEX) markup

```
\subsection{\TeX}
A system for formatting text.
\par
  \TeX\ and its accompanying macro
package \LaTeX\ provide powerful means of
formatting text to be output on either
\begin{itemize}
\item simple matrix printer,
\item a laser printer or
\item a photo typesetter.
\end{itemize}
```

2.6.4. Descriptive (SGML) markup

```
<h>&TeX;
<p>A system for formatting text.
<p>&TeX; and its accompanying macro
package &LaTeX; provide powerful means of
formatting text to be output on either
<li>
<it> simple matrix printer,
<it> a laser printer or
<it> a photo typesetter.
</li>
```

2.7. What is SGML not?

- No WYSIWYG (WYSIWY(A)G, ...) way of working
- Not a formatter, certainly not a standard formatter

3. What is TEX?

TEX is a formatter for 'making beautiful books', developed by Knuth, [23]. An introduction is given in [14].
LATEX, [29], is a macro collection for simplified use of TEX, in the *procedural* markup way. A Dutch TEX Users Group exists.[2] Courseware is [10].

3.1. Processing LATEX

'LATEX' is processed in three steps

'copy'
$\stackrel{editor}{\rightarrow}$ ASCII
$\stackrel{LaTeX}{\rightarrow}$ dvi-file
$\stackrel{driver}{\rightarrow}$ 'results'

The more steps the more cumbersome is correction handling

[2] NTG: Nederlandse TEX Gebruikersgroep. Secretary: G.J.H. van Nes, ENR Postbus 1, 1755ZG, Petten. 02246–4185. e-mail: **vannes@ecn.nl**.
Évidemment, pour la communauté francophone : GUTenberg, BP 21, F-78354 Jouy-En-Josas Cedex, e-mail : **gut@irisa.fr** *(ndlr).*

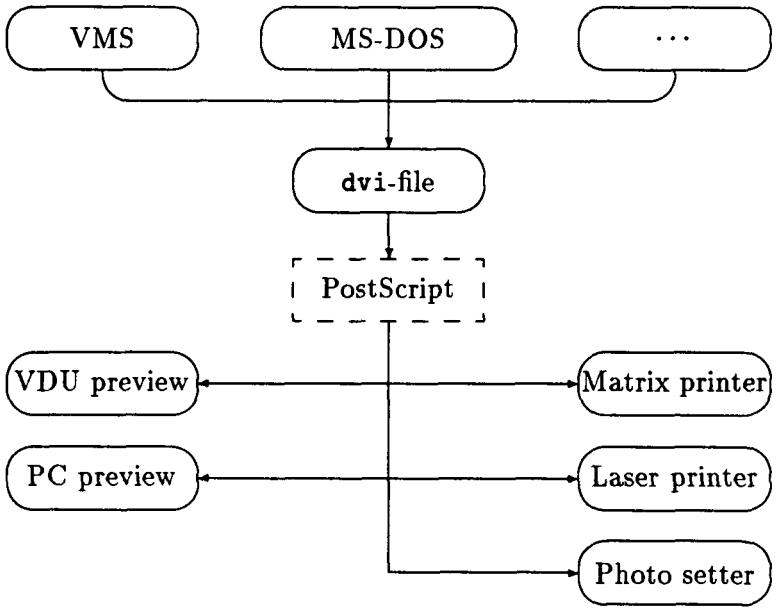
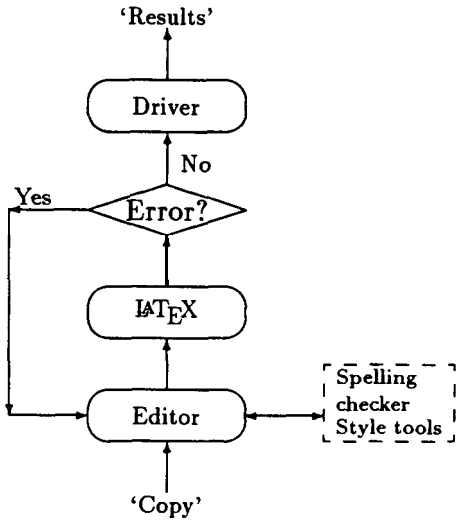


Figure 1: LaTeX's use



3.2. Availability

TeX is available on many computers under various operating systems with a variety of drivers for the VDU (previewing), printer (hardly any), and photo setter. So documents written in (La)TeX can be ported. Sending documents via e-mail is also generally possible except for the incorporated graphics. When graphics is part of the document TeX combined with Postscript is used within the TeX community. TeX is in the public domain. Drivers and in general added value by companies have to be paid for. See figure 1. See ads in [36].

4. Relationship: SGML, TeX and ...

The relationship of TeX, SGML and other applications is illustrated in the diagram in figure 2. The coupling — 'converters' — can be done in SGML, in TeX or via special 'compilers'. An integrated[3] implementation is Arbortext's The Publisher on a SUN.

[3] Ikons user interface, SGML layer, TeX layer, Postscript handling (optionally); with SGML, TeX and dvi files as intermediate results

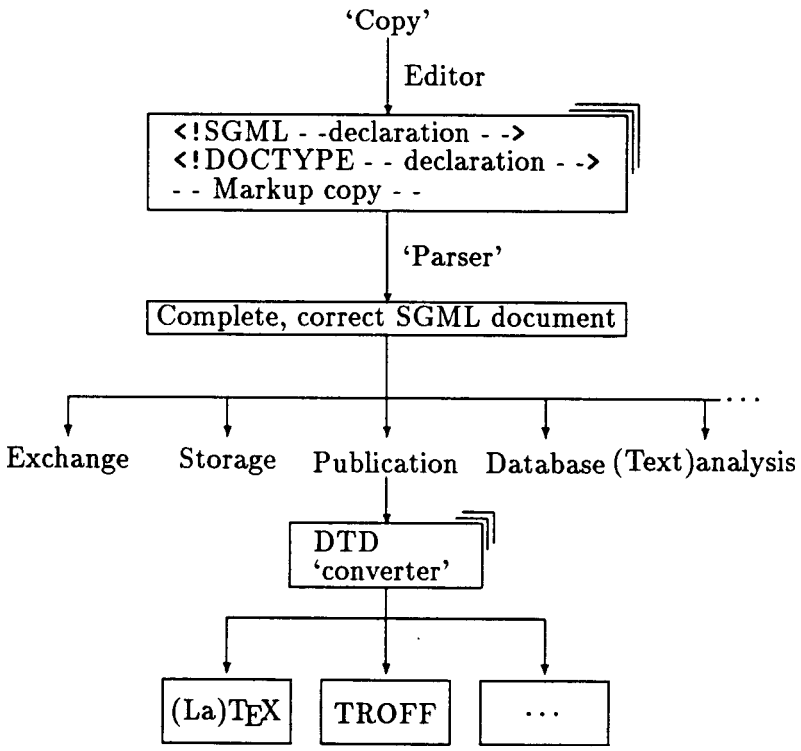


Figure 2: Relationship

5. Examples

5.1. Letter

5.1.1. Structure

- Background

Heading (Logo, address, phone, ...)
Footer (numbering, ...)

- Context (running heads next pages, ...)
- Reference
- Your reference
- Date
- Addressee (name, company, address, zip code)
- Beginning (Dear...)
- Contents
- End matter (Salutation, name, position)
- Additions (PS, enclosure, cc)

5.1.2. Letter result

Because a sample LaTeX letter could not be processed simultaneously in this context, the result is omitted. (Of course it could be pasted in, but that is not available electronically; it has been 'pasted into' the transparencies)

5.1.3. SGML markup

```
<!DOCTYPE letter PUBLIC
  -- DTD to be used  --
  "-//NTG//DTD Letter//EN">
<letter -- start-tag -->
<ref> CGL/Ba/B89-007
<yourref> MC/L1/L89-001
<date> 4 august 1989
<address> Malcolm Clark
   Imperial College Computer Centre
   Exhibition Road
   London SW7 2BP, England
%<email>
   janet: fps@uk.ac.ic.cc.vaxa
```

```
%</email>
<dear>Malcolm
<p> Thank you very much ...
..
<p> Some details about the course ...
..
<signed name=CGL>
</letter -- end-tag -->
```

5.1.4. LaTeX *specification*

```
\documentstyle[12pt]{letter}
  \address{% return address
          C. G. van der Laan \\
          \ldots}
  \signature{Kees}
\begin{document}
{\LARGE % This size just for
        % transparency
\begin{letter}{% address
          Malcolm Clark      \\
          \dots}
% no ref or your ref
% date is handled automatically
\opening{Dear Malcolm}
\par
Thank you very much \ldots
\begin{quote}
$\vdots$
\end{quote}
Some details about the course
\ldots
\begin{quote}
$\vdots$
\end{quote}
\closing{Best regards}
        % Handles signature
%ps, cc, enclosure all possible
\end{letter}
}
\end{document}
```

5.2. Bridge card deal

The LaTeX aspects have been published in [25]. An SGML elaboration has been done by Grootenhuis, [17].

5.2.1. LaTeX *result*

N/None	♠ J74 ♡ AJ ◇ QJT2 ♣ Q874	Deal: demo
♠ A3 ♡ K76 ◇ 963 ♣ KJ952	N W E S	♠ K86 ♡ T9542 ◇ 874 ♣ T3
	♠ QT952 ♡ Q83 ◇ AK5 ♣ A6	

5.2.2. *SGML markup*

```
<deal><vuln>N/None
      <comm>Deal: demo
<hand n>&spades;J74
        &hearts;AJ
        &diams;QJT2
        &clubs;Q874
<hand e>&spades;K86
        &hearts;T9542
        &diams;874
        &clubs;T3
<hand s>&spades;QT952
        &hearts;Q83
        &diams;AK5
        &clubs;A6
<hand w>&spades;A3
        &hearts;K76
        &diams;963
        &clubs;KJ952
</deal>
```

5.2.3. LaTeX *specification*

```
\crdima{N/None}{%
  \begin{minipage}[t]{\br}
     Deal:\\demo
  \end{minipage}}%
  {\hand{J74}{AJ}{QJT2}{Q874}}%N
  {\hand{K86}{T9542}{874}{T3}}%E
  {\hand{QT952}{Q83}{AK5}{A6}}%S
  {\hand{A3}{K76}{963}{KJ952}}%W
```

5.2.4. LaTeX *macros*

```
\newcommand{\hand}[4]{
 \begin{minipage}[t]{\br}
        %I chose \br=8em
 \begin{tabbing}
 %width of parbox equals:
```

```
 %min{\br, max{string #1, ...,
 %                 string #4}}
  \(\spadesuit\)  \= #1 \\
  \(\heartsuit\)  \> #2 \\
  \(\diamondsuit\)\> #3 \\
  \(\clubsuit\)   \> #4
 \end{tabbing}
 \end{minipage}       }%end \hand
%
\newsavebox{\NESW}
\savebox{\NESW}[4em]{%
\raisebox{-1.5\baselineskip}%
{\fbox{\small W
   \raisebox{2.6ex}{N}
   \hspace*{-1em}
   \raisebox{-2.6ex}{S}
   {E}
      }
}                    }%end \NESW
%
\newcommand{\crdima}[6]{%
\begin{tabular}[t]{lll}
 #1 & #3              & #2\\
 #6 & \usebox{\NESW} & #4\\
    & #5              &
\end{tabular}          }%end \crdima
```

5.2.5. SGML requirements

Declarations needed in DTD

```
<!ENTITY % ISOpub PUBLIC
  "ISO 8879-1986//ENTITIES Publishing//EN">
<!ELEMENT deal -- (vuln, comm?, hand*)>
<!ELEMENT (vuln|comm) - o CDATA>
<!ELEMENT hand - o (RCDATA, CDATA,
                    RCDATA, CDATA,
                    RCDATA, CDATA,
                    RCDATA, CDATA)>
<!ATTLIST hand nesw (n|e|s|w) #REQUIRED>
```

5.3. Some Math

5.3.1. LaTeX results

$$X \cap (A \cup B) = (X \cup A) \cap (X \cup B)$$

$$x \notin A \not\subset B$$

$$\|a(x+y)\| \leq |a|.(\|x\|+\|y\|)$$

$$\int \frac{1}{\sqrt{1+x^2}}\,dx = \log(1+\sqrt{1+x^2})$$

5.3.2. SGML markup

```
<fd>X&cap;(A&cup;B)=
     (X&cup;A)&cap;(X&cup;B)</fd>

<fd>x&nisin;A&nsub;B</fd>

<fd><fen d>a(x+y)<rp d></fen>&le;
  <fen>a<rp></fen>.(<fen d>x<rp d></fen>
  +<fen d>y<rp d></fen>)
</fd>

<fd><in><opd><fr>1</><rad>1+
  x<sup/2/</rad></fr>dx</in>=
  <rf/log/(1+<rad>1+x<sup/2/</rad>)
</fd>
```

Note. DTD used is an adapted version of AAP's DTD by D.C. Coleman, [26].

5.3.3. LaTeX specification

```
X\cap(A\cup B) =(X\cup A)\cap(X\cup B)

x \notin A \not\subset B

\|a(x+y)\| \leq |a|.(\|x\|+\|y\|)

\int\!\frac{1}{\sqrt{1+x^2}}\,dx
=
\log(1+\sqrt{1+x^2})
```

6. Developments

A survey is given in [8].

6.1. Usage

- DOD (Automated Technical Order System)
- European Communities (FORmalised EXchange of Electronic Documents; office official publications)
- Publishers (AAP, British Library, KNUB-(Elsevier, Kluwer, ...), ...)
- Her Majesty's Stationary Office (legal text)
- HP Technical documentation
- Oxford University Press (abridged forms, database applications)
- McGraw Hill Encyclopedia of Science and technology (CD-ROM)

• SGML Users Group (chapters in various countries)
• ...

6.2. Plans

• DOD (Computer-aided Acquisition and Logistic Support)
Object: To produce an integrated system in which information is held electronically, and which interfaces to CAD/CAM systems, electronic publishing systems and databases and those operated by the many defense contractors who supply the department, so that it will be possible to receive, distribute and use technical information in digital form.

6.3. Local work in progress

• Elsvier's experiment, [9]
• Examples tabular matter (LaTeX and SGML)
• Coupling SGML to LaTeX
• ...

Acknowledgements

This article is an article representation of a presentation prepared via TRSPAR.STY the author's modification of REPORT.STY. Although the structure is such that the TRSPAR copy can be processed by any other style, the file needed some adaptation. E.g. some more text here and there, removing \Large from within the description labels, adaptation of the minipage size, and omitting \Large in the literature list. The latter is used by the author to supply the full literature list on the handouts of the transparencies while attention is focussed on the enlarged items on the transparency.

Most SGML codings are tentative, only the original SGML codings of mathematics have been parsed, [26]. No coupling of SGML to LaTeX has been done yet by the author.

References

[1] Adobe Systems Incorporated (1985): Postscript Language Reference Manual. Addison-Wesley.

[2] Adobe Systems Incorporated (1985): Postscript Language Tutorial and Cookbook. Addison-Wesley.

[3] Association of American Publishers (1987): Standard for electronic manuscript preparation and markup. AAP inc. [4]

[4] Association of American Publishers (198?): Author's guide to electronic manuscript preparation and markup. AAP inc.

[5] Association of American Publishers (198?): Markup of Reference manual on electronic manuscript preparation and markup. AAP inc.

[6] Association of American Publishers (1986): Markup of tabular material. AAP Inc.

[7] Association of American Publishers (1986): Markup of Mathematical Formulas. AAP Inc.

[8] Barron, D. (1989): Why use SGML? Electronic publishing, 2,1, 3–24.

[9] Bleeker, J. (1989): Electronische verzending, bewerking en opslag van wetenschappelijke artikelen. In: SGML de consequenties. De eerste Nederlandse SGML conferentie. SGML User's Group Holland. (Dutch)

[10] Bruin, R. de, C.G. van der Laan, J.R. Luyten, H.F. Vogt (1989): Publiceren met LaTeX. CWI-syllabus 19. (Dutch, but the modular set-up, exercises and answers are useful for a broader audience.)

[11] Bryan, M. (1988): SGML, an Author's Guide to the Standard Generalized Markup Language. Addison-Wesley.

[12] Clark, M. (1989): TeX and/or SGML. Proceedings EuroTeX89. Karlsruhe. (Context sensitivity as a tool for checking input correctness is stressed; an example of how to do this within TeX is given.)

[13] Coombs, J.H., A.H. Renear, S.J. DeRose (1987): Markup systems and the future of scholarly text processing. Comm. ACM, 30, 11, 933–947.

[4] Association of American Publishers, 2005 Massachusetts Avenue, NW. Washington, DC 20036, Phone: (202) 232-3335

[14] Doob, M. (1989): A gentle introduction to TEX. A manual for selfstudy. Also as RC-RUG rapport 25.

[15] Genussa, P.L. (1987): Document Preparation Method of the United States Aire Force Automated Technical Order System. SGML users' group. Bulletin 2, 1.

[16] Grootenhuis, J. (priv. comm.): Koppelen LATEX aan SGML.

[17] Grootenhuis, J. (priv. comm.): Kaartverdelingen en biedverloop bij bridgen —Een SGML tutorial. (Dutch)

[18] Guittet, C. (1986): FORMEX: une mise en practique des normes internationales. SGML users' group. Bulletin, 1, 2.

[19] Herwijnen, E. van (1988): Electronic submission of Physics articles to publishers. De 1e Nederlandse SGML conferentie. SGML: De Consequenties. (Also submitted to: EP-ODD (Electronic Publishing Originating, Dissemination and Design). In the context of this paper the discussion of SGML related to TEX is relevant.)

[20] ISO8879 Information Processing — Text and Office Systems — Standard Generalized Markup Language (SGML). 1986-10-15.

[21] ISO8613 Office Document Architecture (ODA) and Interchange Format. International Organization for Standardization. 1988.

[22] ISO/TR9573 Information Processing — SGML support facilities — Techniques for using SGML. 1988-09-12.

[23] Knuth, D.E. (1984): The TEXbook. Addison-Wesley.

[24] Knuth, D.E. (1989): The Errors of TEX. Softw. prac. exp. 19, 7. 607-685.

[25] Laan, C.G. van der (1989): Typesetting Bridge via LATEX. TUGboat, 10, 113-116.

[26] Laan, C.G. van der, D.C. Coleman, J.R. Luyten (1989): SGML-LATEX. 1. Mathematical Formulas. RC-RUG report 24.

[27] Laan, C.G. van der, D.C. Coleman, e.a. (in preparation): SGML-LATEX. 2. Tabular Matter. RC-RUG report.

[28] Laan, C.G. van der (in preparation): An introduction to SGML.

[29] Lamport, L. (1985): LATEX a Document Preparation System. Addison-Wesley.

[30] Price, L. (1987): SGML and TEX. TUGboat, 8, 2, 221-225. (The context sensitivity of SGML is emphasized.)

[31] MARK-IT (1989): SGML Parser, version 2. Sobemap NV, Place du Champ de Mars 5, Bte 40, 1050 Bruxelles.

[32] Smith, J.M. (1987): The standard generalized markup language (SGML): Guidelines for editors and publishers. British National Bibliography Research Fund Report 26. ISBN 0 7123 3111 5.

[33] Smith, J.M. (1987): The standard generalized markup language (SGML): Guidelines for authors. British National Bibliography Research Fund Report 27. ISBN 0 7123 3112 3.

[34] Smith, J.M., R. Stutely (1988): SGML - the users' guide to ISO8879. Ellis Horwood/Halsted Press. ISBN 0 7458 0221 4 / 0 470 21126 1.

[35] SGML Users' Group Newsletters. Editorial address: Pindar Infotek, 2 Grosvenor Road, Wallington, Surrey SM6 0ER, UK.

[36] TUGboat. Quaterly of TEX Users Group. Editorial address: TEX Users Group. P.O. Box 9506. Providence RI 02940. email: TUGboat@Math.AMS.com.

[37] Warmer, J., S. van Egmond (to appear in EP-ODD): The implementation of the Amsterdam SGML Parser.

[38] Wittbecker, A. (1989): TEX enslaved. Proceedings TUG89. Stanford. (Advantages and disadvantages of TEX-formatter with SGML 'front-end' are discussed, related to DEC's VAX Document.)

Addentum Quelques références bibliographiques publiées en français (ndlr).

Références bibliographiques

[39] Borceux, F. (1989) : LATEX, *la perfection dans le traitement de texte*, édition Ciaco, Bruxelles.

[40] Seroul, R. (1989) : *Le petit livre de TEX*, Inter-Éditions, Paris.

[41] Vignaud, D. (1989) : *L'édition structurée dans les documents, SGML applications à l'édition française*, Éditions du Cercle de la Librairie, Paris.