25 Years of TEX and METAFONT *Looking Back and Looking Forward*

Nelson H. F. Beebe

Department of Mathematics University of Utah Salt Lake City, UT 84112-0090 USA







T_EX has lasted longer than many other computer software technologies.

This talk reviews some of the history of T_EX and METAFONT, how they have come to be used in practice, and what their impact has been on document markup, the Internet, and publishing.

 T_EX has several design deficiencies that limit its use and its audience. We look at what T_EX did right, and with 25 years of hindsight, what it did wrong.

The talk closes with some observations about the challenges ahead for electronic representation of documents.



Where are we?







Donald E. Knuth





2 Digital Typography





Slide 1.

With me it has always been the opposite: I tend to err in the other direction. I often get so interested in Chapter 1 of the books that I'm reading or studying, I don't have much time to read the final chapters.

Once, when I was five years old, my parents let me take the streetcar to the downtown library by myself, and I was absolutely fascinated by the children's books. When I didn't come home on time, my parents were worried and phoned the library. One of the night staff went looking and found me in the stacks, reading happily—I had no idea that the library was closed and that everyone else had gone home! Even today my wife knows that when I go into a library, I'll probably come home late.

In fact, not only have I always loved books, I've also been in love with the individual *letters* in books. Here's a page from the first ABC alphabet book that I had when I was little [SLIDE 2]. Curiously, I marked each serif in the letters with a little x, and I counted the serifs: The letter K has 7 serifs. The letter P [SLIDE 3] has 4; the letter O [SLIDE 4] has none.



From this you can see that I like numbers as well as letters. By the time I became a professor at Stanford I had learned that my main talents were associated with computer programming, and I had begun to write



Donald E. Knuth's new book



DOCTOR FUN



Don Knuth finally sells out.



Donald E. Knuth and numerology



Don always enjoys finding patterns and connections in numbers. Here are some for the 25th anniversary of T_EX in 2003.

In 2003, 2 and 3 are the first 2 primes, and we have 2 zeros.

- $2003 = 11111010011_{2}$ $25 = \underbrace{11001_{2}}_{\text{mirror images}}$ $2003 = \text{prime} \quad \text{(first in this millenium; 304th prime)}$ $25 = 5^{2}$
 - = (third prime)^(oddest prime of all)



History, T_EX, and METAFONT



4000 BCE Egyptian invent papyrus from woven reeds

- 105 Ts'ai Lun invents bark/hemp/rags-based paper in China
- 1009 First paper mill in Europe, Xativa, Spain
- 1411 First paper mill in Germany
- 1452 Johannes Gutenberg invents movable type
- 1680 First paper mill in New World, in Culhuacan, Mexico
- 1690 First paper mill in English colonies, near Philadelphia
- 1798 Nicholas Robert invents first paper-making machine, in France

History, TEX, and METAFONT (cont.)

1850–1879 Paper from wood pulp perfected 1889–1900 Economical mass-produced paper **1940s** First digital computers 1968–1973 Niklaus Wirth invents Pascal language 1969–1970 Dennis Ritchie invents C language 1970s roff, script, runoff, document 1975–1978 eqn (B. W. Kernighan and L. Cherry) 1976 nroff and troff (J. Ossanna), 1978 bib and refer (M. Lesk)



- 1977–1978 classic T_EX and METAFONT in Sail (D. Knuth)
- **1978–1980** Scribe (B. Reid)
- 1979 tbl (M. Lesk)
- 1981 pic (B. W. Kernighan)
- 1982 ideal (C. Van Wyk)
- 1982 'final' TEX and METAFONT in Pascal
- 1983–1985 LATEX (L. Lamport)
- 1984 BIBT_EX (O. Patashnik)
- 1984 PostScript (Adobe Systems)

History, TEX, and METAFONT (cont.)

- 1986 grap (J. Bentley and B. W. Kernighan)
- **1989** 'new' T_EX and METAFONT (8-bit characters et al.)
- 1989–1991 HTML and HTTP at CERN (T. Berners-Lee)
- 1990 METAPOST (J. Hobby)
- 1991 World-Wide Web at CERN
- 1993 xmosaic browser (NCSA: M. Andreeson)
- 1993 PDF (Adobe Systems)
- **1994** $LAT_E X 2_{\mathcal{E}}$ (F. Mittelbach et al.)
- 1994 Ω (Y. Haralambous and J. Plaice) and Λ



- 1995–2000 WeBWork (University of Rochester)
- 1996 PDFT_EX (Hán Thế Thánh)
- 1997 eT_EX (P. Breitenlohner et al.)
- 1998 $\mathcal{N}_T \mathcal{S}$ (K. Skoupý)
- 2000 XMLT_EX (D. Carlisle)
- 2001 JadeT_EX (S. Rahtz)
- **2002** Donald Knuth celebrates $1,000,000^{\text{th}}_{2}$ birthday
- 2003 ant (ant is not T_EX: A. Blumensath) (OCaml: 24K lines)
- 2003 Nottingham font conversion project (D. Brailsford)





What we've accomplished



- **TUGboat and EPodd journals**







What we've accomplished (cont.)











What we've accomplished (cont.)



- many journals in mathematics, physics, and computer science use LATEX
- some major publishers use T_EX behind the scenes for book pages
- T_EX markup is *de facto* standard in mathematics, physics, and computer science



CTAN Archives



CTAN archives: 77K files, 6K directories, 76 mirrors







CTAN Archives



- UK CTAN mirror: 45M hits in 7 years (125K/week) from 1M hosts
- DANTE CTAN mirror: 150K/week
- Utah CTAN mirror: 5K/fortnight
- DEK's plain.tex: 1235 lines; manmac.tex: 715 lines
- CTAN *.{cls,ltx,sty,tex}: 2,796,695 lines
- DEK's cm/*.mf: 260 files
- CTAN * .mf: 6644 files



Document archives



 online article archives in physics (250,000+) and mathematics (29,000+) [Cornell, formerly Los Alamos]: about 1M connections/week (http://arxiv.org/) [Paul Ginsparg: 2002 MacArthur Fellow]





- CoRR (Computing Research Repository)
 (http://www.acm.org/repository/)
- Networked Computer Science Technical Reference Library (http://www.ncstrl.org/)



BIBT_EX bibliography archives



- DBLP: 187,494 [XML \longrightarrow BIBT_EX]
- MathUtah: 203,623 [in-house]
- TUG archive: 344,846 [21K accesses/month]
- Karlsruhe: 1,261,147 [300K accesses/month]
- American and European Mathematical Society databases provide BIBTEX output
- XML \leftrightarrow BIBTEX exchange



What did T_EX do right?



- open-source literate program
- small kernel of primitives specialized for typesetting (cf. PostScript)
- programming language makes it extensible
- DVI file frees it from output device dependency
- knows nothing about fonts beyond metrics (TFM files)
- GF/PK/TFM font files have open specification
- boxes and glue typesetting model



What did T_EX do right? (cont.)



- common cases are handled by compact markup:

 - \$...\$ for mathematics
 - for superscript
 - _ for subscript
 - empty line for paragraph break
- sequences of spaces are normally equivalent to a single space (cf: troff's significant spaces)
- typesetting based on exact arithmetic (32-bit integer: 1+1+14+16): identical results on all platforms
- \input allows dynamic loading of files



What did TEX do right? (cont.)



- \catcode adds great power, freeing T_EX from fixed meanings of characters (e.g., jadetex and xmltex)
- no \system call: no viruses! (but: \write still a danger)
- Iast definition holds (cf. SGML)
- T_EX is stable and reliable
- books illustrated by Duane Bibby



What did T_EX do wrong?



not based on rigorous grammar

cf. Donald E. Knuth, *On the translation of languages from left to right*, Information and Control, 8(6) 607–639 (1965): LR(k) parsing (reprinted in (38) *Great Papers in Computer Science*, IEEE 1996).

- programming language is arcane macro language, rather than true programming language with functions and procedures
- too much hard coded (better programming language could have removed many things from kernel)
- too many fixed-size objects (256 boxes, registers, ...)



What did T_EX do wrong? (cont.)



- too many global variables
- name collision: lack of function/procedure scoping and namespace control
- inadequate I/O (line-oriented with braces balanced, instead of being based on \getc and \putc)
- character set limits (but: hard to have forseen ISO10646 and Unicode in 1978, and *all* programming languages are suffering from the change)
- no input filters (ΩTP)
- no color state



What did T_EX do wrong? (cont.)



- no graphics (dot, vector, path fill primitives could have gone a *long way*) (cf. SIGGRAPH CORE 1979, PostScript 1984)
- one page at a time (need two for good book design)
- not general enough for all writing directions
- multicolumn output might have been a kernel primitive
- no DVI output pipe (cf. work by David Fuchs, BlueSky Research, and Jonathan Fine)
- no -safe option to sandbox T_EX into single directory (cf. ghostscript)



What did T_EX do wrong? (cont.)



- arithmetic overflow not caught in addition (caught in multiplication) (cf. Java, C, C++, Pascal, Fortran, ...)
- 32-bit precision too limiting; 64-bit would have been much better (TeX arithmetic is already mostly in software anyway)
- could have had IEEE 754 floating-point arithmetic (1980–1985) in software for machine-independent computation
- conventional arithmetic expressions should have been standard, instead of \advance, \multiply, \divide (cf. Cobol)



What did METAFONT do right?



- open source literate program
- small kernel of primitives specialized for font design
- programming language makes it extensible
- 'Meta' fonts: families based on common programs with parameter variations
- font files (GF, PK, TFM) have open specification



What did METAFONT do wrong?



- conversion of outlines to bitmaps done too early: should have also supported output of outline fonts (but: could not foresee PostScript in 1978)
- inadequate I/O (worse than T_EX)



Future directions



- Unicode and ISO10646 character encoding:

Digital encoding of writing systems is a kludge. And boy, do we seem to be paying for the Unicode version of that kludge on the list this week. ;-)

> Kenneth Whistler on Unicode list Fri, 27 Jun 2003 13:08:17 -0700 (PDT)



Future directions: XML growth

Move afoot

XML traffic

to speed



Traffic jam

Because of its use in Web services applications, XML's network presence is predicted to grow rapidly.



Source: Network World, 23 June 2003, p. 1



Future directions: XML



T_FX XML (http://www.w3.org/TR/xexpr/) <define name="factorial" args="x"> <if> <It><x/>2</It> <X/> <multiply> <X/> <factorial> $F_0 = 1$ $F_n = nF_{n-1}$ (n > 0) <subtract><x/>1</subtract> </factorial> </multiply> </if> </define>





Books of note





na inti, i

D onald Knuth's influence in computer science ranges from the mathematical analysis of algorithms to the invention of literate programming. His award-winning textbooks have become classics that are often credited for shaping the field; his

scientific papers are widely referenced and stand as milestones of development over a wide range of topics. In the present volume, the third in a series of his collected works, Knuth explores the relationship between computers and typography. For more than a decade during the



Don Knuth in Mainz, Germany in 1987, admiring a replica of the printing press used by Johannes Gutenberg to print his 42-line Bible in the 1450s.

crucial formative years of the desktop publishing revolution, he directed a project that brought type designers, punch cutters, typographers, book historicans, and scholars to Stanford University and led to what some consider the golden age of digital typography. Knuth regards the present work as his legacy to that field, in celebration of an enormously exciting period of research. Many of the papers reprinted here introduced new technologies at the time they were first printed, and they can be published today only with the help of an astonishing collection of software that has been developed during the past twenty years. This is truly a work that only Knuth himself could have produced.

SLI LECTURE NOTES NUMBER 78

CSLI Publications STANFORD, CALIFORNIA

ISBN 1-57586-010-4



Books of note





0

SPRINGER PROFESSIONAL COMPUTING

Apostolos Syropoulos Antonis Tsolomitis Nick Sofroniou

Springer

Digital Typography Using LAT_EX



Www.springer-ny.com www.springer.de Desktop publishing/LaTeX Beginning/Intermediate Level

SPRINGER PROFESSIONAL COMPUTING

Digital Typography Using LATEX

The Tex and LaTeX typesetting systems have become the de facto standard for preparing complex technical documents in print and electronic form.

Digital Typography Using LaTeX collects together the most frequently required 'how to' information and tools and presents them in a single accessible volume. It provides the first detailed presentation of recent developments in multilingual typesetting using TeX that now make it easy for users to prepare documents in their own language and alphabet. In addition, the book troadly demonstrates how to make a single document that can generate multiple for mats: camera-ready copy for printing, World Wide Web pages, and portable electronic documents for distribution over the Internet.

Topics & Features

- An fall in one," comprehensive introduction to TeX and LaTeX, suitable for both those new to, and experienced with, the TeX software program.
- Provides detailed description of TeX's and LaTeX's new multilingual features, as well as how to combine languages into a single document
- Provides details for obtaining and installing popular versions of LaTeX and necessary utility programs and tools
- CD-ROM includes extensive utility programs, tools, and demonstrations presented in the book
- Covers areas of advanced typography, such as the addition of Postscript or TrueType fonts to a LaTeX installation, and how to work with graphic images
- Describes advanced typesetting systems, including Omega, Lambda, pdfLaTex, and epsilon-LaTeX

All scientists, engineers, and technical professionals who use LaTeX will find the book a rich source of methods and tools for their desktop publishing needs.

0





The Unicode Standard Version 3.0

The Unicode Consortium

Unicode

Characters for all the languages of the world The standard for the new millennium Required for XNUL and the Internet The basis for modern software standards and products The official way to implement ISO/IEC 10646 The key to global interoperability

The Unicode Standard, Version 3.0

The authoritative, technical guide to the creation of software for worldwide use. • Detailed specifications for Unicode:

Programming Languages

 * Structure, conformance, encoding forms, character properties, semantics, equivalence, combining characters, logical ordering, conversion, allocation, big/little endian usage, Korean syllable formation, control characters, case mappings, numeric values, mathematical properties, writing directions (Arabic, Japanese, English, and so on), character shaping (Arabic, Devanagari, Tamil, and so on)
 Expanded implementation guidelines by

Expanded implementation guidelines experts in global software design:

* Normalization, sorting and searching, case mapping, compression, language tagging, boundaries (characters, word, lines, and sentences), rendering of non-spacing marks, transcoding to other character sets, handling unknown characters, surrogate pairs, numbers, editing and selection, keyboard input, and more

• Comprehensive charts, references, glossary, and indexes:

* Codes, names, appearances, aliases, cross-references, equivalences, radicalstroke ideographic index, Shift-JIS index, and more

http://www.aw.com/cseng/ Cover design by Karin Hansen Cover photograph by Katie Noyes Cover concept by Joan Aliprand and Mark Davi

ADDISON-WESLEY Addison-Wesley is an imprint of Addison Wesley Longman, Ind

CD-ROM

- The comprehensive Unicode Character Database for:
 - * Character codes, names, properties, decompositions, upper-, lower-, and title cases, normalizations, shaping
- International, national, and vendor character mappings for:
 - * Western European, Japanese, Chinese, Korean, Greek, Russian, and others
 * Windows, Macintosh, Unix, and Linux
- Unicode Technical Reports that extend the standard for:
 - * Sorting, displaying, normalizing, linebreaking, compression, serialization, regular expressions, CR/LF, XML, case mappings, and more

The Unicode Consortium is a nonprofit organization founded to develop, extend, and promote use of the Unicode Standard. Members include companies and organizations on the vanguard of globalization technology; together they comprise a source of unrivaled internationalization expertise. Visit the Consortium's Web site: http://www.unicode.org

The principal authors and editors of *The Unicode Standard, Version 3.0*, are Joan Aliprand, Julie Allen, Joe Becker, Mark Davis, Michael Everson, Asmus Freytag, John Jenkins, Mike Ksar, Rick McGowan, Lisa Moore, Michel Suignard, and Ken Whistler.

ISBN 0-201-61633-5

\$49.95 US \$74.95 CANADA





					Unine	eu Ca	nadia	III AU	origir	al Sy	llabic	S		1
	15	B 1	ISC	15D	15E	15F	160	161	162	163	164	165	166	167
Ī				117		0		5				_	~	0.0
0		1	\sim	V	0		R	\odot	\sim	5	U	З	F	Ψ. I
	15B	0 1	15C0	15D0	15E0	15F0	1600	1610	1620	1630	1640	1650	1660	1670
			-	٨		-	V	-		10	0	-)T(00.0
1			$n \mid$	/1\	U	\supset		0	0X	D)(Э	ω	4
	15B	1 1	15C1	15D1	15E1	15F1	1601	1611	1621	1631	1641	1651	1661	1671
	1		. 1	~	M	-	1	2		6	5	5	m	a c
2			9	\rightarrow	0	Ð	U	0	U	\odot	~	3)1(U
	15B	2 1	15C2	15D2	15E2	15F2	1602	1612	1622	1632	1642	1652	1662	1673
			N .	~	0		0	0		~	5	~	5	90
3	٦			7	M	\supset		e	2	G	1	5	D	0
	15B	3 1	15C3	15D3	15E3	15F3	1603	1613	1623	1633	1643	1653	1663	1673
			\vee	\geq	D	\subseteq	-	Z	h	75	5	ж	E	9
4			A	7	D	\subset)	O	5	0	2	ω	Ð	Ű
	15B	4 1	15C4	15D4	15E4	15F4	1604	1614	1624	1634	1644	1654	1664	1674
_		r	Δ	4	5	Ш	-	2	h	0	C	m	P	99.1
5		1 /	\square		D	W	U	0	9	22	5	ж	2	0
	158	5 1	1505	15D6	15E5	15F5	1605	1615	1625	1635	1645	1655	1665	1675
~				₩/	5.)	m	-	9	6	Ъ	Z	R	J	a.i.
б				¥		ш		~	9	N			5	
-	158	0 1	1500	1506	1555	1560	1000	1010	1020	1630	1040	1050	1000	10/0
7	1 7	1	ìD	A	7	R	C	10	6	Ъ	Z	Ŧ	M	
'				/11	1007	1657	1007	1617	1627	1627	1647	1657	1667	
	108	r 1	1567	1507	IJEI	1967	1007	1017	1021	1037	1047	1037	1007	
8	μ	:	·>	\rightarrow	H	R	lu lu	10	51	5	17	B	£	
0	400		1500	1600	1000	1659	1608	1619	1629	1638	1648	1658	1668	
	130		1500	1300	1020	1010	1000	1010	1020	1000	1010	1010	1440	******
9	0		\checkmark	\Rightarrow	Ω	B	η	10	0.	d	n	B	B	
5	10		1600	1600	1650	1650	1600	1610	1620	1639	1649	1659	1669	
	138	9 1	1909	1303	102.9	101-9	1005	1013	1023	1035	1040	1000	1000	<u>HHH</u>
Δ	6	1	\forall	\rightarrow	D	B	3	NO	5	Цľ	Ð	\$	B	
	158	A 1	1504	1504	15554	15FA	1604	1614	162A	163A	164A	165A	166A	
	1000		100m	100/1	1021	10111	10011							<i>()))))</i>
в	Ч		Δ	€	D	M	B	Q1	5	m	Ē	Æ	B	
	15B	B 1	15CB	15DB	15EB	15FB	1608	161B	162B	163B	164B	165B	166B	
С		()	\triangleright		D	L M	3	8	5	Э	Ð	Æ	B	
	15B	C 1	15CC	15DC	15EC	15FC	160C	161C	162C	163C	164C	165C	166C	
ľ				~	-	_	-	-	_	-		-		
D]	\triangleright	\square	D	B	3	2	5	В	E	B	X	
	158	D 1	15CD	15DD	15ED	15FD	160D	161D	162D	163D	164D	165D	166D	
				-		-	10			-	2.4	~		
Е	h		\triangleright	D	1	B	0	XO	U	З	W	B	×	
	15B	E 1	15CE	15DE	15EE	15FE	160E	161E	162E	163E	164E	165E	166E	
			,		m	5	0			-	~	~	50	
F	Ý	<	\triangleleft	U	W	B	6	XO	Ω	Ξ)'(Ĕ	Y	
	15B	F 1	15CF	15DF	15EF	15FF	160F	161F	162F	163F	164F	165F	166F	
460												The U	nicode	Standard





Unicode Demystified A Practical

Programmer's Guide to the Encoding Standard

Richard Gillam

Foreword by Dr. Mark Davis, President, The Unicode Consortium







Sources and Studies in the History of Mathematics and **Physical Sciences**

L.E. SIGLER

Fibonacci's Liber Abaci

A Translation into Modern English of Leonardo Pisano's Book of Calculation



Its author, Leonardo Pisano, known today as Fibonacci, was a citizen of Pisa, an active maritime power, with trading outposts on the Barbary Coast and other points in the Muslim Empire. As a youth Fibonacci was instructed in mathematics in one of these outposts; he continued his study of mathematics while traveling extensively on business and developed contacts with scientists throughout the Mediterranean world. A member of the academic court around Emperor Frederick II, Leonardo saw clearly the advantages for both commerce and scholarship of the Hindu positional number system and the algebraic methods developed by al-Khwarizmi and other Muslim scientists.

Though it is now mostly known for introducing the Hindu number system and the algorithms of arithmetic that children now learn in grade school, Liber Abaci is much more: It is an encyclopaedia of thirteenth-century mathematics, both theoretical and practical. It develops the tools rigorously, establishes them with Euclidean proofs, and then shows how to apply them to all kinds of situations in business and trade, including conversion of measures and currency, allocations of profit, computation of interest, and alloying of currencies. It is rigorous mathematics, well applied, and vividly described.

As the first translation into a modern language of the Liber Abaci, this book will be of interest not only to historians of science, but to all mathematicians and mathematics teachers interested in the origins of their methods.



ISBN 0-387-95419-8 www.springer-ny.com





That's all folks



\drinkfill \relax \bye

