

25 Years of T_EX and METAFONT

Looking Back and Looking Forward

Nelson H. F. Beebe

Department of Mathematics

University of Utah

Salt Lake City, UT 84112-0090

USA

Abstract

$\text{T}_{\text{E}}\text{X}$ has lasted longer than many other computer software technologies.

This talk reviews some of the history of $\text{T}_{\text{E}}\text{X}$ and METAFONT, how they have come to be used in practice, and what their impact has been on document markup, the Internet, and publishing.

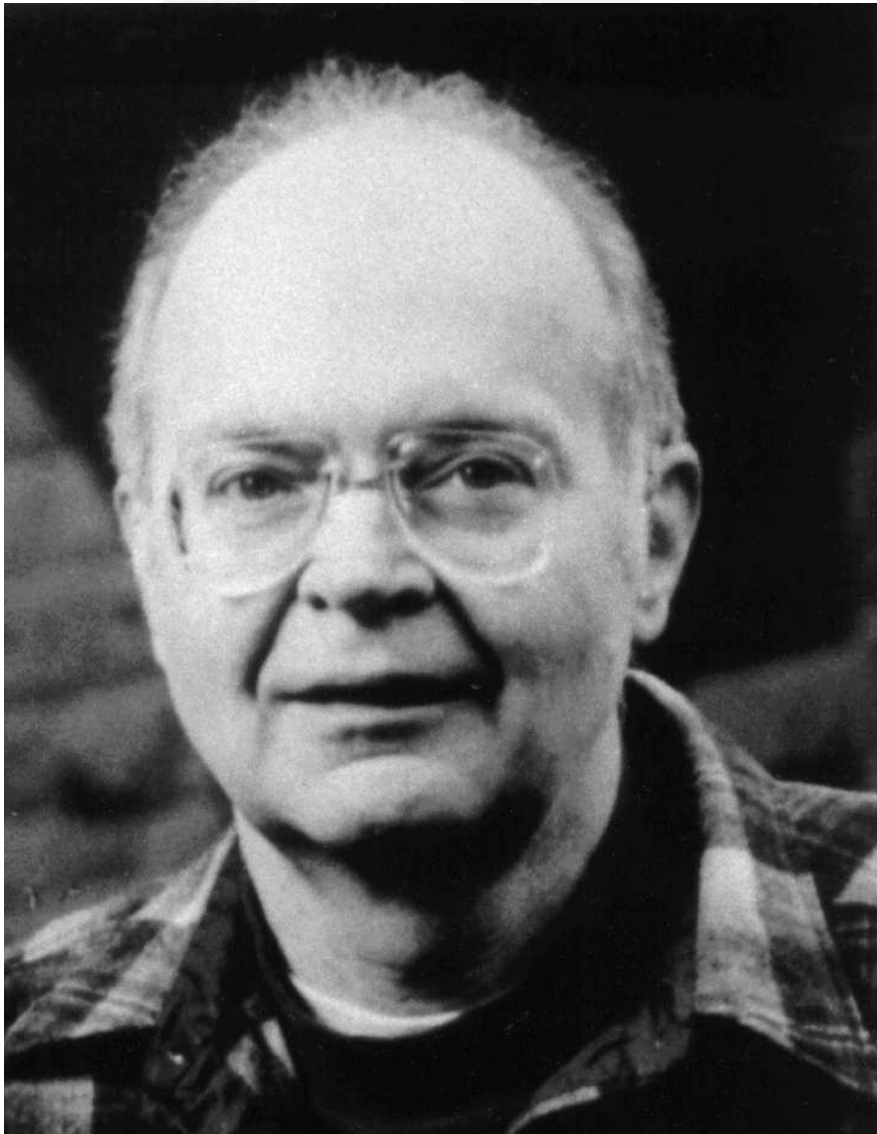
$\text{T}_{\text{E}}\text{X}$ has several design deficiencies that limit its use and its audience. We look at what $\text{T}_{\text{E}}\text{X}$ did right, and with 25 years of hindsight, what it did wrong.

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

Where are we?



Donald E. Knuth



2 Digital Typography



Too young to read books himself, Donald Knuth, 4, of 2961 N. 18th st. is nevertheless the youngest member of the public library's Ancient Order of Book Worms. His father reads storybooks to him. Wednesday he gave oral reports on his "reading" in the young people's room of the library. Some of the stories he reported on were "Country Bunny," "Pokey Bear," "When the Root Children Wake Up" and "Babar the King."

Slide 1.



Slide 2.

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.



Slide 3.

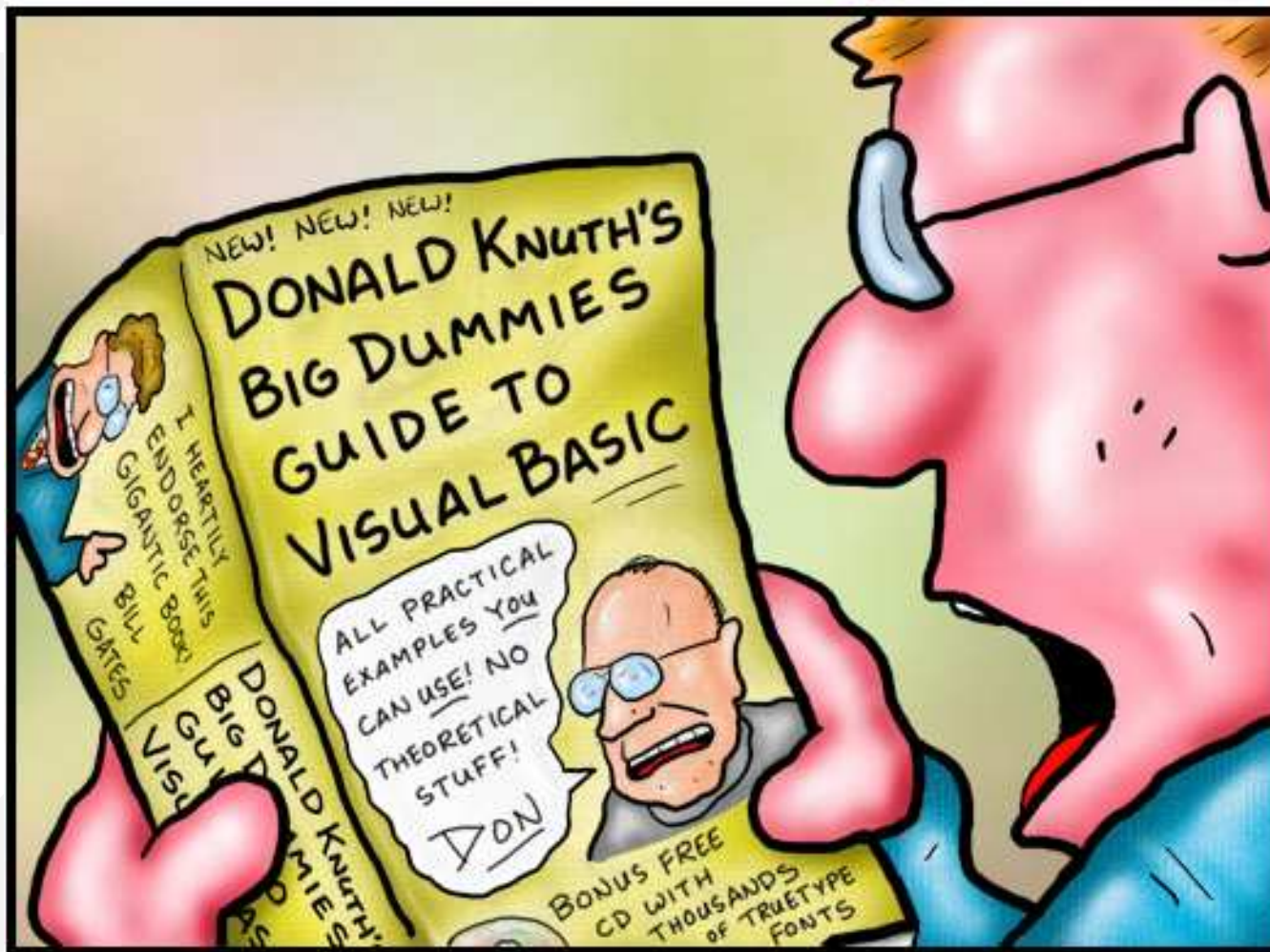


Slide 4.

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.

$$2003 = 11111010011_2$$

$$25 = \underbrace{11001}_2$$

mirror images

$$2003 = \text{prime} \quad (\text{first in this millenium; 304th prime})$$

$$25 = 5^2$$

$$= (\text{third prime})^{(\text{oddest prime of all})}$$

History, T_EX, and METAFONT



- 4000 BCE** Egyptians 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, T_EX, 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)



History, T_EX, and METAFONT (cont.)



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’ T_EX and METAFONT in Pascal

1983–1985 L^AT_EX (L. Lamport)

1984 BIBT_EX (O. Patashnik)

1984 PostScript (Adobe Systems)



History, T_EX, 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. Andreessen)
- 1993** PDF (Adobe Systems)
- 1994** L^AT_EX 2_ε (F. Mittelbach et al.)
- 1994** Ω (Y. Haralambous and J. Plaice) and Λ



History, T_EX, and METAFONT (cont.)



1995–2000 WeBWork (University of Rochester)

1996 PDFT_EX (Hán Thế Thánh)

1997 eT_EX (P. Breitenlohner et al.)

1998 $\mathcal{N}_{\mathcal{I}\mathcal{S}}$ (K. Skoupý)

2000 XMLT_EX (D. Carlisle)

2001 JadeT_EX (S. Rahtz)

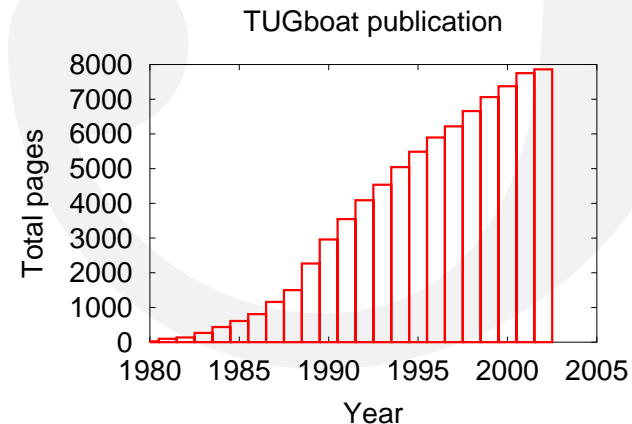
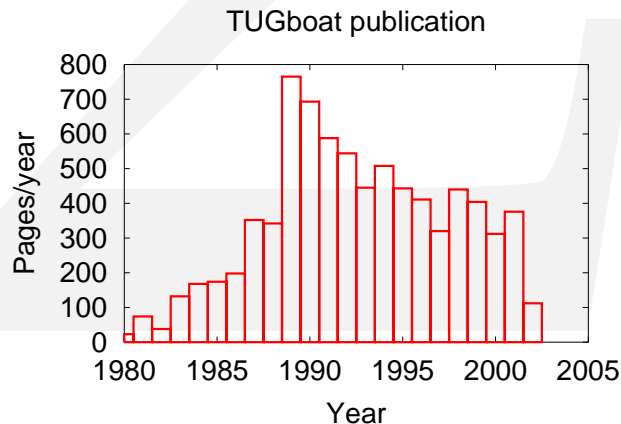
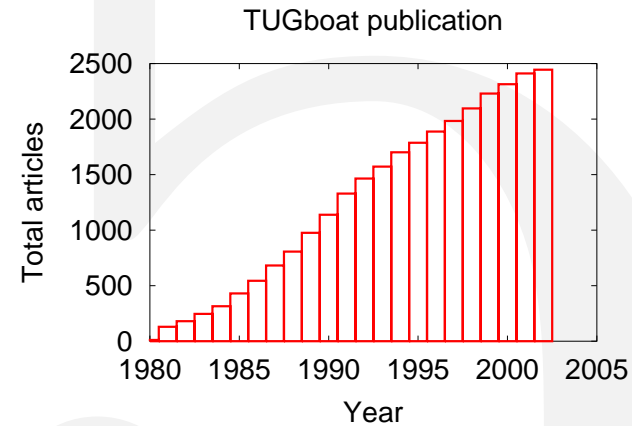
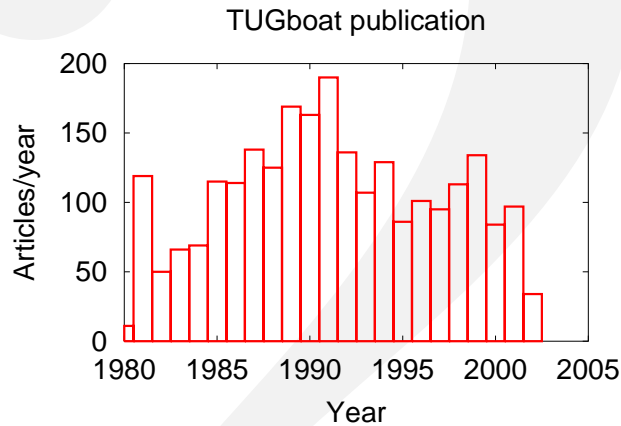
2002 Donald Knuth celebrates 1,000,000₂th birthday

2003 ant (ant is not T_EX: A. Blumensath) (OCaml: 24K lines)

2003 Nottingham font conversion project (D. Brailsford)

What we've accomplished

- 100+ books on T_EX and METAFONT
- TUGboat and EPOdd journals

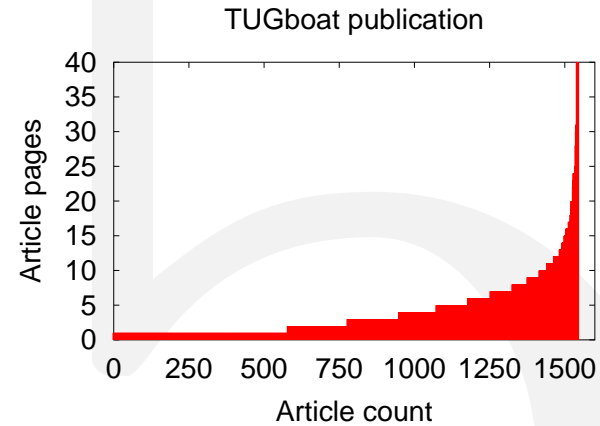
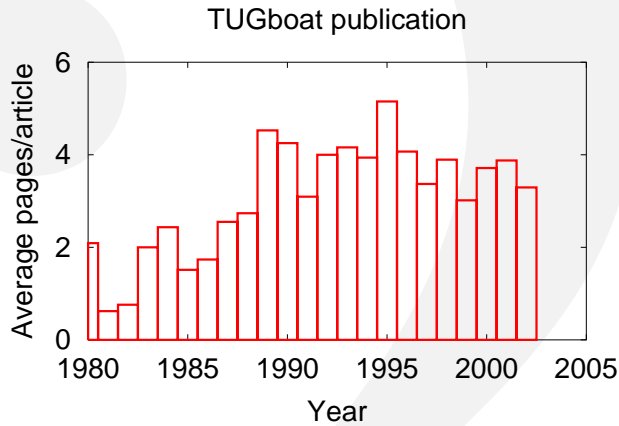




What we've accomplished (cont.)



● TUGboat article sizes





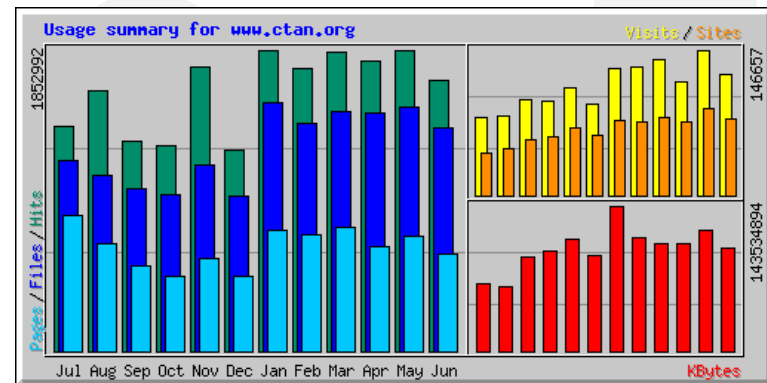
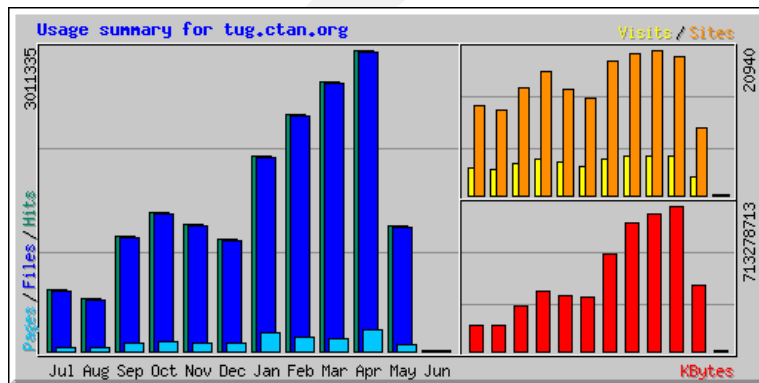
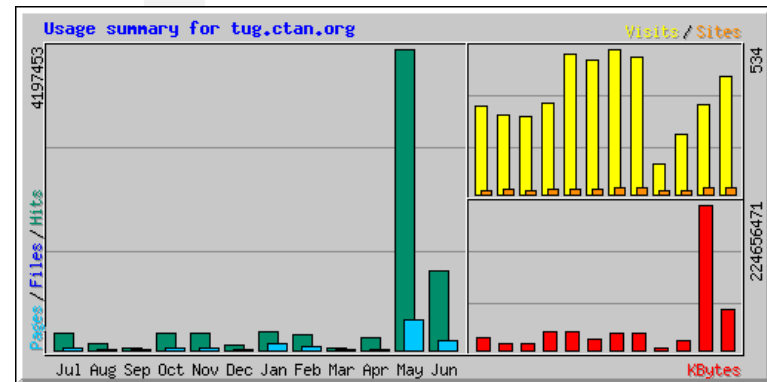
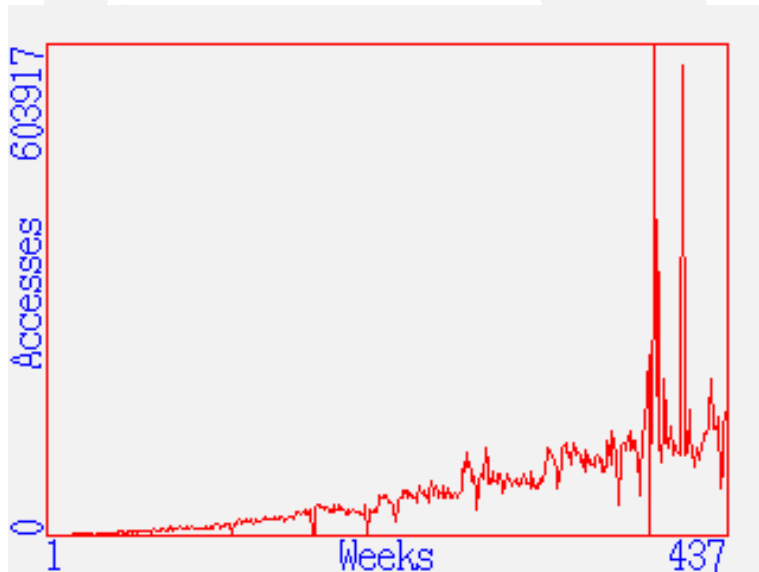
What we've accomplished (cont.)



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

CTAN Archives

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

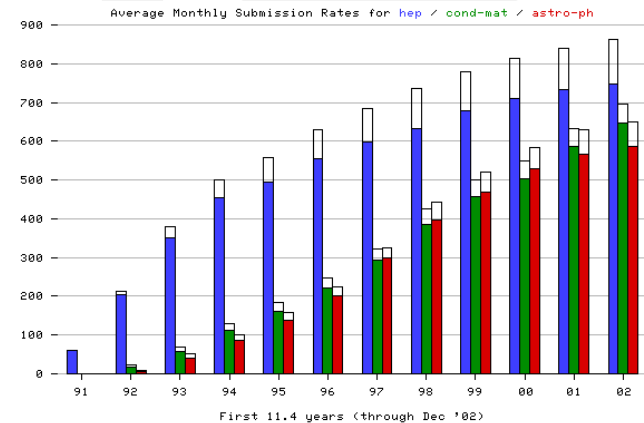
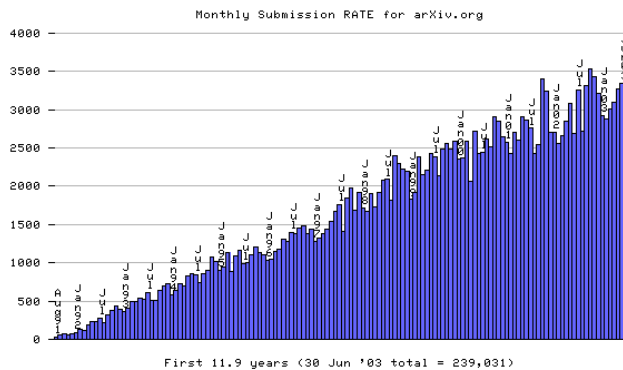


CTAN Archives

- 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.ncstr1.org/>)

BIB_TE_X bibliography archives

- DBLP: 187,494 [XML \longrightarrow BIB_TE_X]
- 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 BIB_TE_X output
- XML \leftrightarrow BIB_TE_X 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 grouping and arguments
 - `$. . . $` 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 T_EX 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)
- last 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 foreseen ISO10646 and Unicode in 1978, and *all* programming languages are suffering from the change)
- no input filters (Ω TP)
- 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

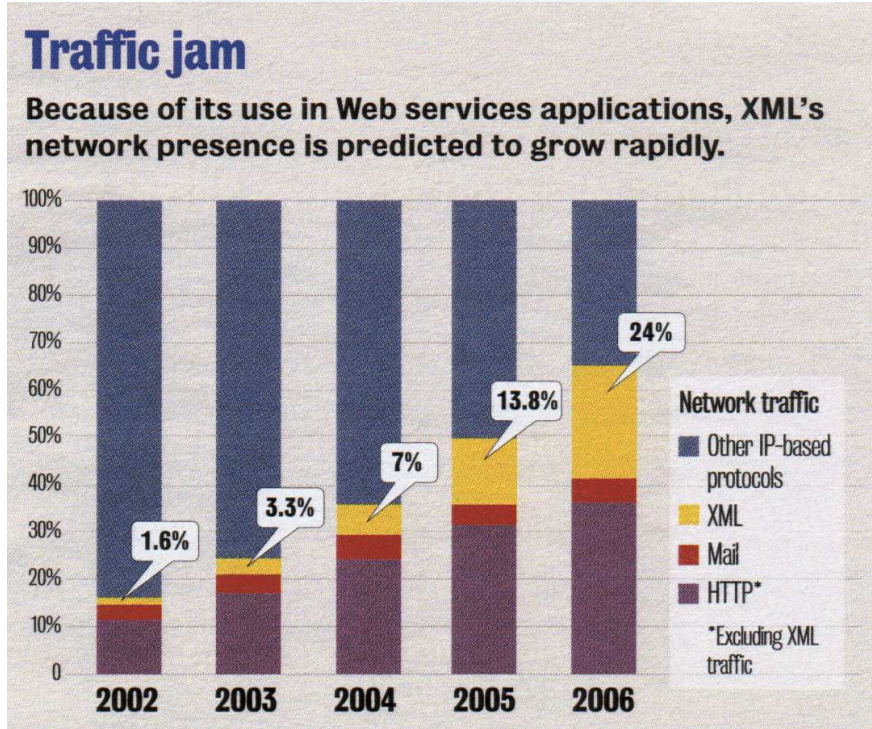
- XML, XML, XML, XML, XML, XML, XML, XML
- Unicode and ISO10646 character encoding:

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
to speed
XML traffic**



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



Future directions: XML

TEX

XML (<http://www.w3.org/TR/xexpr/>)

```

<define name="factorial" args="x">
  <if>
    <lt><x/>2</lt>
    <x/>
    <multiply>
      <x/>
      <factorial>
        <subtract><x/>1</subtract>
      </factorial>
    </multiply>
  </if>
</define>

```

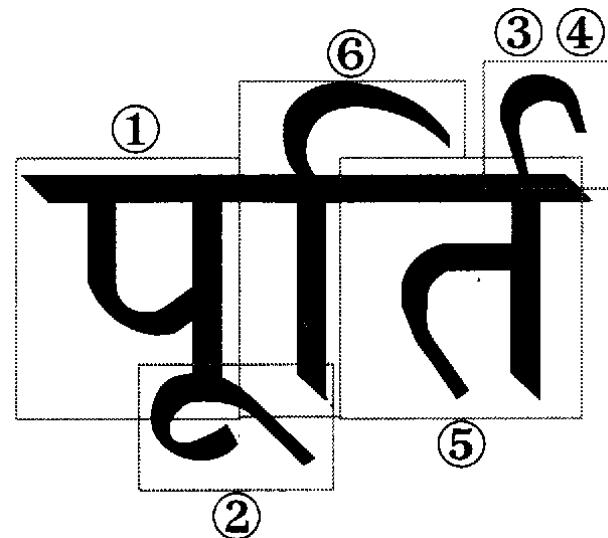
$$F_0 = 1$$

$$F_n = nF_{n-1} \quad (n > 0)$$

Future directions: Unicode

Typesetting some languages is hard!

پیچ و خم کاچ



Books of note

Digital Typography



Donald E. Knuth

Donald 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 historians, 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.

CSLI LECTURE NOTES NUMBER 78

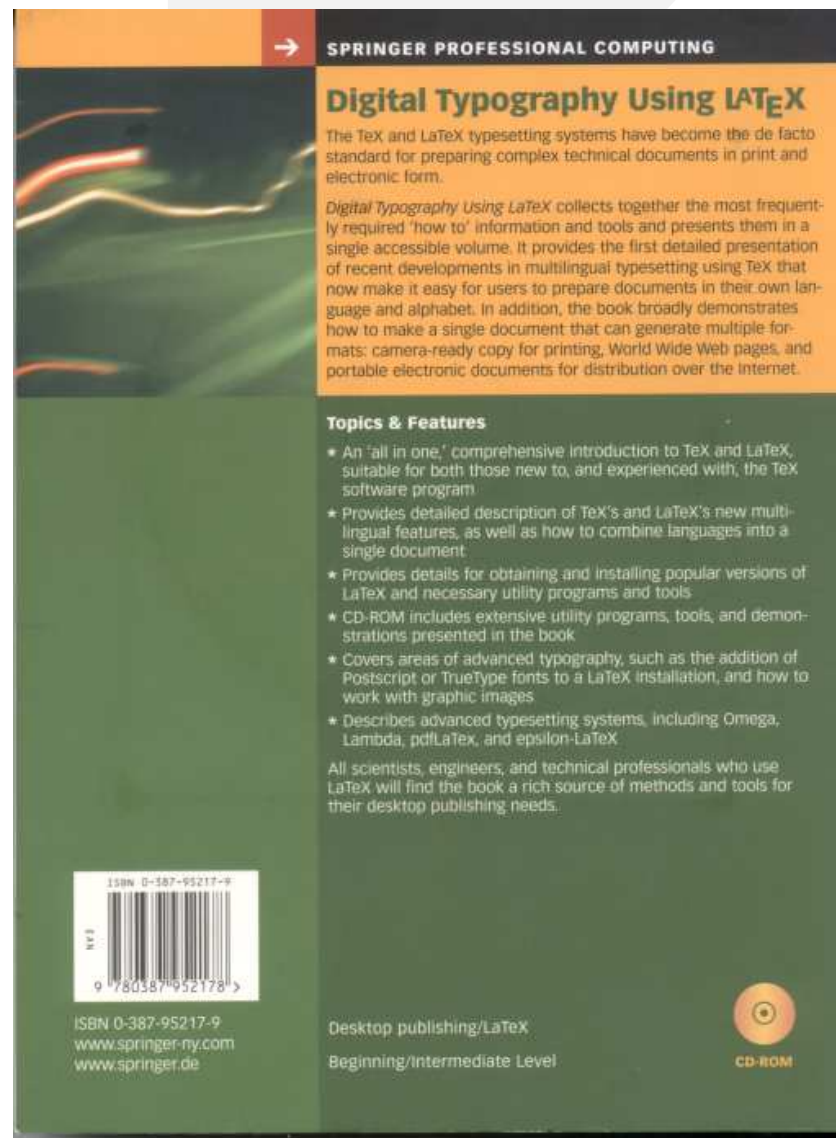
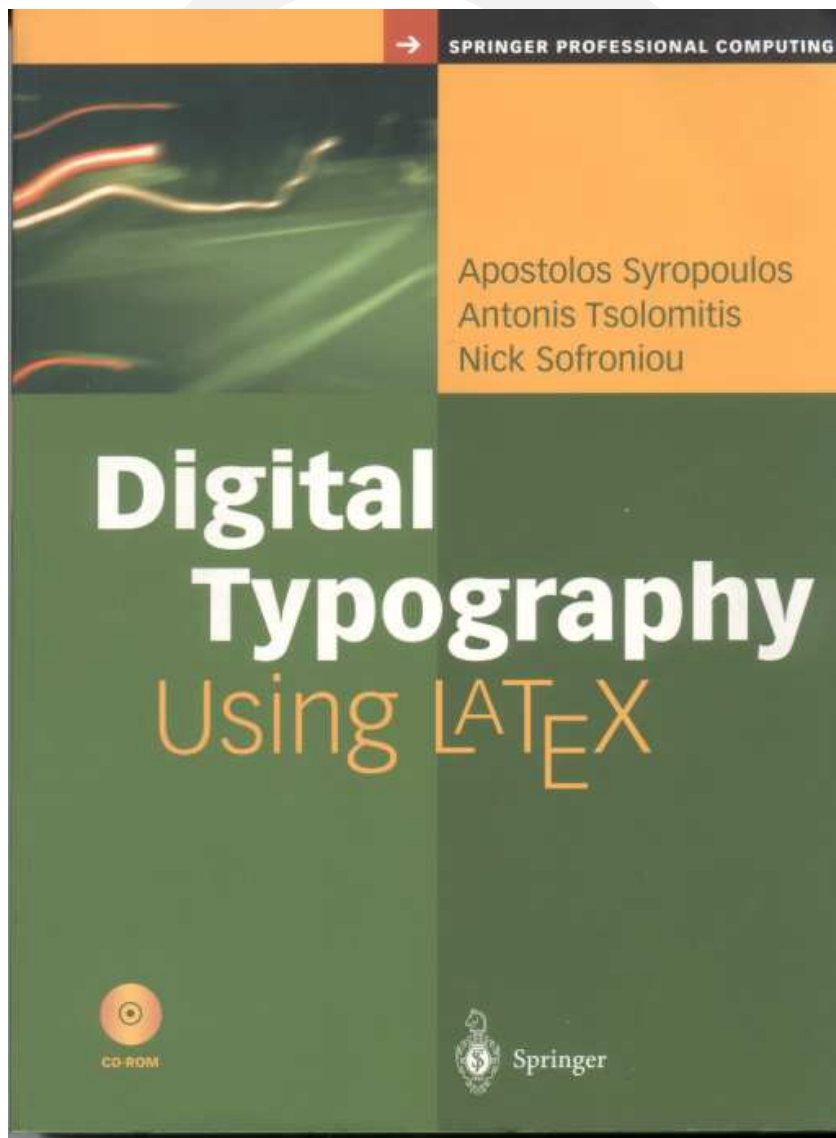
CSLI Publications

STANFORD, CALIFORNIA

ISBN 1-57186-010-4



Books of note



Books of note (cont.)



The Unicode Standard Version 3.0

The Unicode Consortium

Programming Languages

Unicode

*Characters for all the languages of the world
The standard for the new millennium
Required for XML 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:**
 - * 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 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, radical-stroke ideographic index, Shift-JIS index, and more

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.

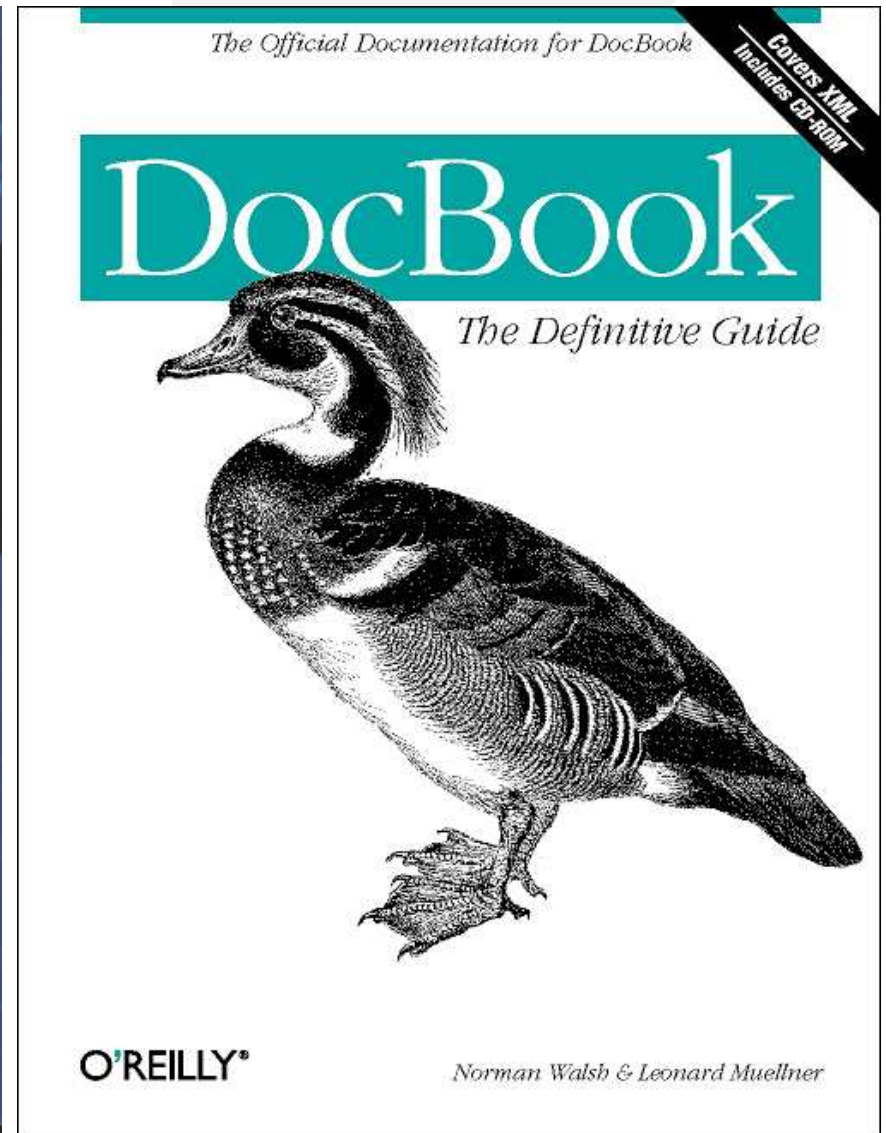
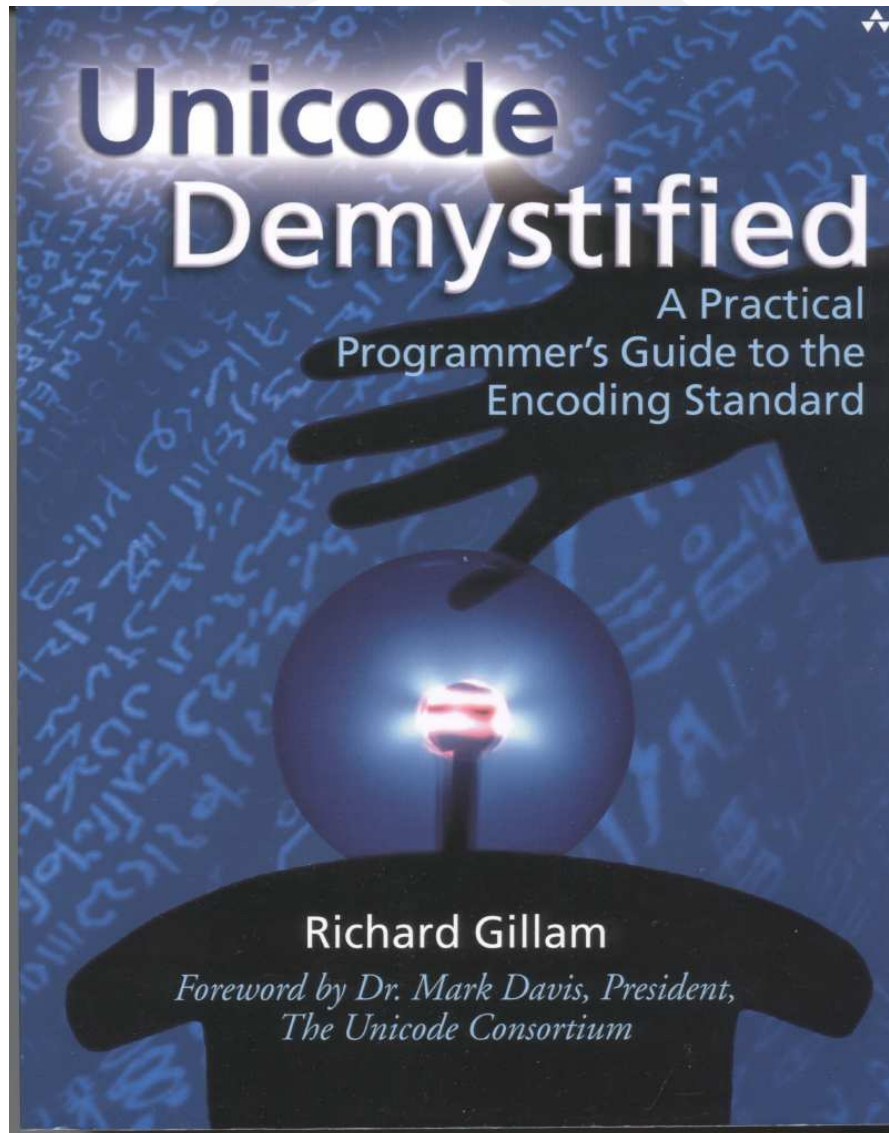
<http://www.aw.com/eseng/>
Cover design by Karin Hansen
Cover photograph by Katie Noyes
Cover concept by Joan Aliprand and Mark Davis



★ ADDISON-WESLEY
Addison-Wesley is an imprint
of Addison Wesley Longman, Inc.



Books of note (cont.)



Books of note (cont.)

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



Springer

First published in 1202, Fibonacci's *Liber Abaci* was one of the most important books on mathematics in the Middle Ages, introducing Arabic numerals and methods throughout Europe.

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

