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

Abstract

TEX has lasted longer than many other computer software technologies.

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

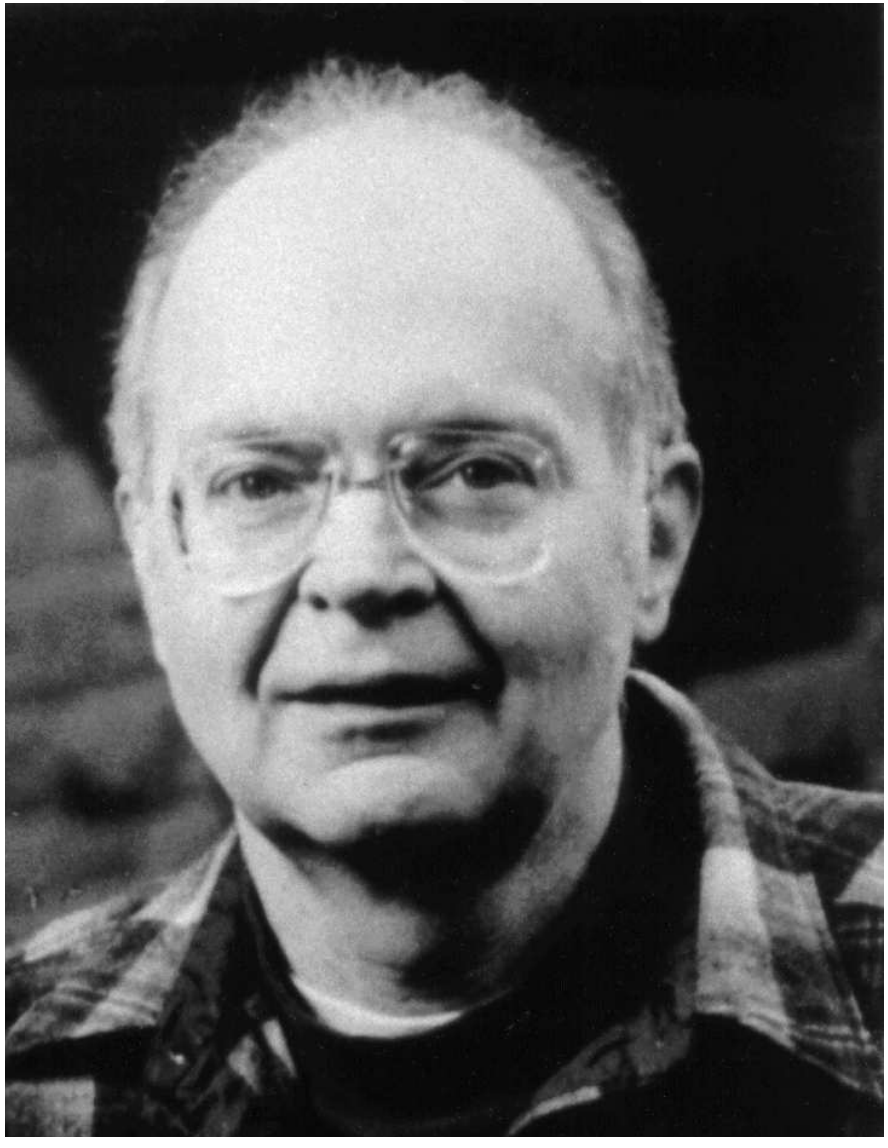
TEX has several design deficiencies that limit its use and its audience. We look at what TEX did right, and with 25 years of hindsight, what it did wrong.

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

Where are we?


Hawaii
The BIG Island
Upolu Pt.
Upolu Airport
Mookini Heiau
Kamehameha Birthsite Memorial
Mahukona Beach Park
Lapakahi State Historical Park
Kapaau
Hawi
Pololu Valley Lookout
Kohala Mountains
Kukuihaele
Honokaa
Paauilo
Kawaihae Harbor
Kawaihae
Hapuna Beach State Recreation Area
Waimea (Kamuela)
Waimea-Kohala Airport
Kohala Coast
Anaehoomalu Bay
Waikoloa
Laupahoehoe
Papaaloa
Ninole
Honomu
Akaka Falls State Park
Observatory
Hilo Bay
Coconut Island
Rainbow Falls
Hilo
Gen. Lyman Field (Hilo Airport)
Keaau
Kaloli Pt.
Kurtistown
Mountain View
Glenwood
Pahoa
Opihikao
Volcano
Thurston Lava Tube
Kilauea Caldera
Kilauea Military Camp
Hawaii Volcanoes National Park
Park Headquarters Kilauea Visitor Center Volcano Art Center Volcano House
Weather Observatory
Mauna Loa Elev. 13,680 ft.
Hualalai Elev. 8,271 ft.
Kalaoa
Keahole Airport
Makalawena Marsh
Kailua Kona
Hulihee Palace
Kahaluu
Keauhou
Honokohau
Mokuaikaua Church
Holualoa
Kainaliu
Kealakekua
Captain Cook
Capt. Cook Monument
Kealakekua Bay State Underwater Park
St. Benedict's Painted Church
Puuhonua o Honaunau National Historical Park
Milolii
Wood Valley
Pahala
Punaluu
Whittington Beach Park
Naalehu
Apua Pt.
Ka Lae (South Point)
Hawaii
270
250
240
19
27
190
11
200
180
130

Donald E. Knuth



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.

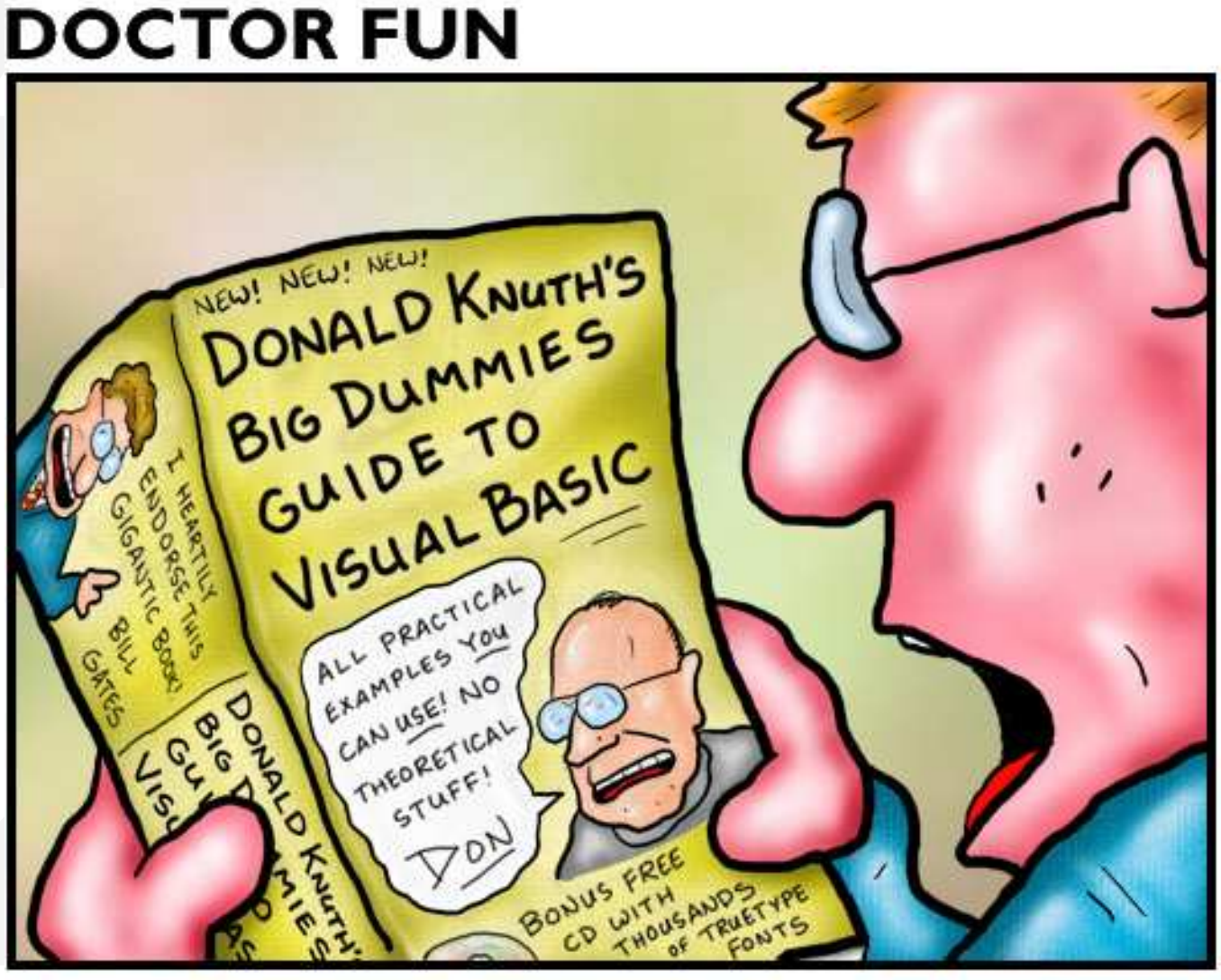
In fact, not only have I always loved books, I've also been in love with the individual *letters* in books. Here's a page from the first ABC alphabet book that I had when I was little [SLIDE 2]. Curiously, I marked each serif in the letters with a little x, and I counted the serifs: The letter K has 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



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 TEX in 2003.

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

$$
\begin{aligned}
2003 &= 11111010011_2 \\
25 &= \underbrace{11001_2}_{\text{mirror images}} \\
2003 &= \text{prime} \qquad \text{(first in this millenium; 304th prime)} \\
25 &= 5^2 \\
&= (\text{third prime})^{(\text{oddest prime of all})}
\end{aligned}
$$

History, TEX, 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)

History, TEX, and METAFONT (cont.)

1977–1978 classic TEX 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 BIBTEX (O. Patashnik)

1984 PostScript (Adobe Systems)

History, TEX, and METAFONT (cont.)

1986 grap (J. Bentley and B. W. Kernighan)

1989 'new' TEX 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 LATEX 2_ε (F. Mittelbach et al.)

1994 Ω (Y. Haralambous and J. Plaice) and Λ

History, TEX, and METAFONT (cont.)

1995–2000 WeBWork (University of Rochester)

1996 PDFTEX (Hán Thế Thánh)

1997 eTEX (P. Breitenlohner et al.)

1998 $\mathcal{NTS}$ (K. Skoupý)

2000 XMLTEX (D. Carlisle)

2001 JadeTEX (S. Rahtz)

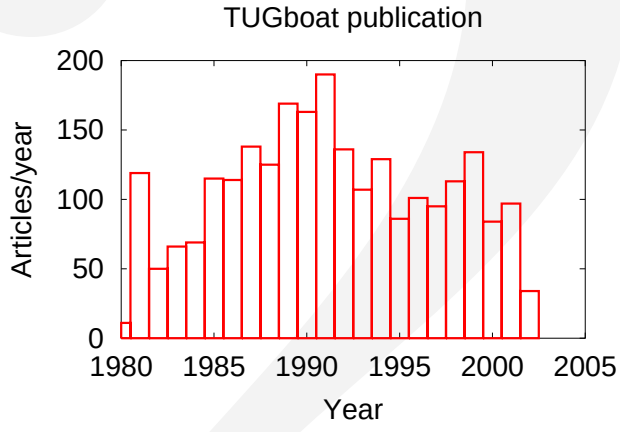
2002 Donald Knuth celebrates $1,000,000_2^{\text{th}}$ birthday

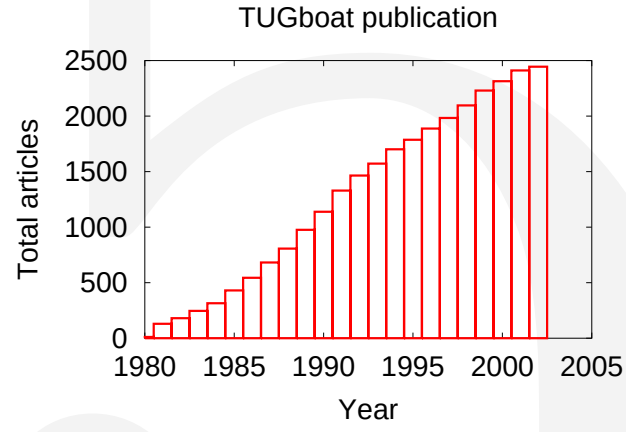
2003 ant (ant is not TEX: A. Blumensath) (OCaml: 24K lines)

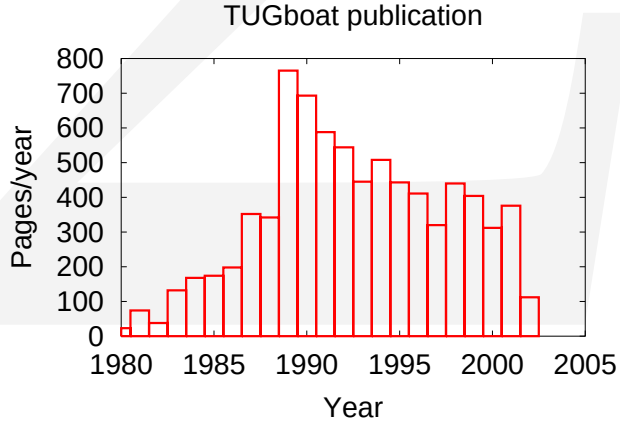
2003 Nottingham font conversion project (D. Brailsford)

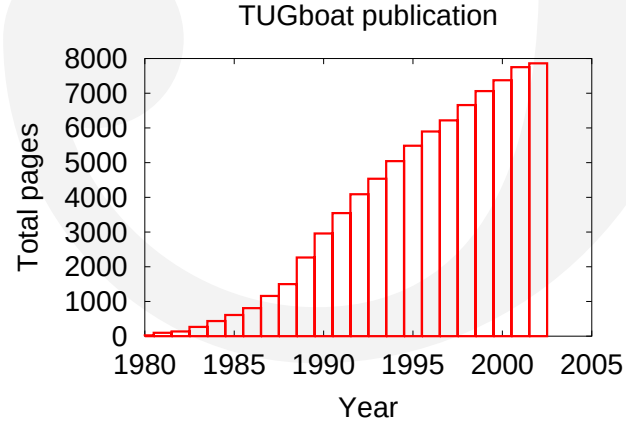
What we've accomplished

- 100+ books on TEX and METAFONT
- TUGboat and EPodd journals



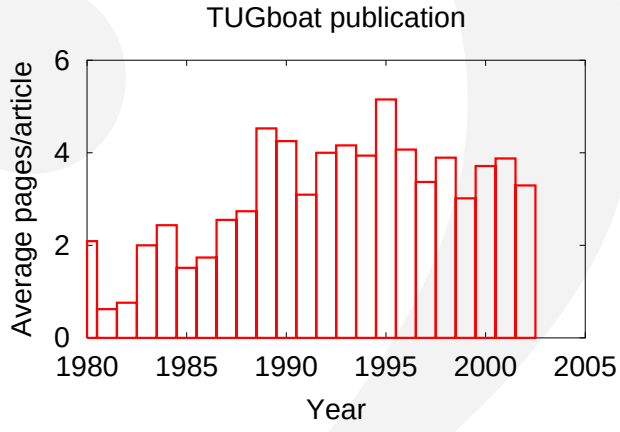


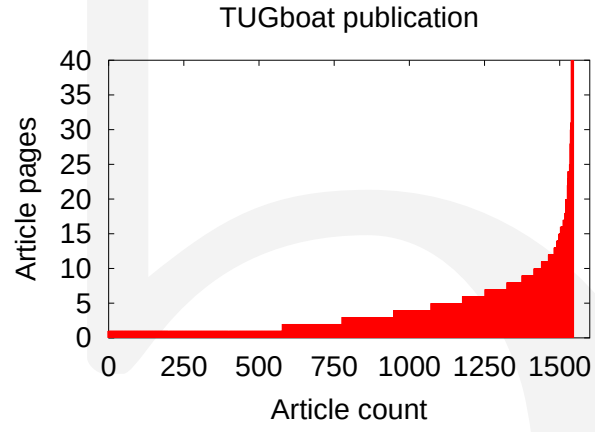




What we've accomplished (cont.)

- TUGboat article sizes



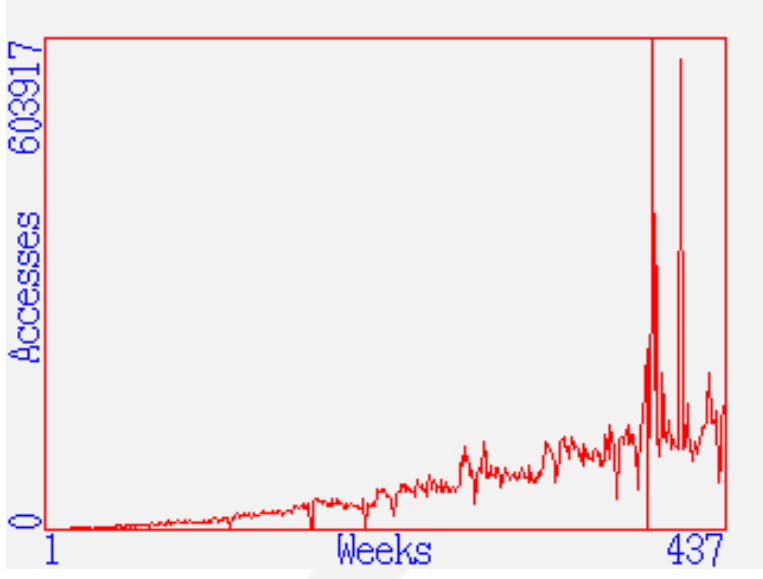


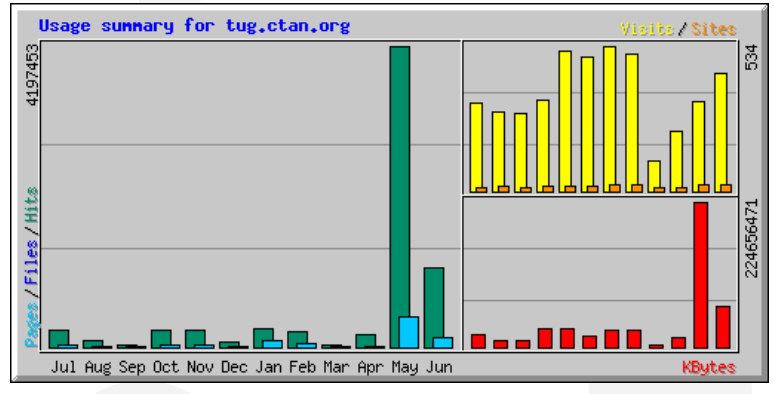
What we've accomplished (cont.)

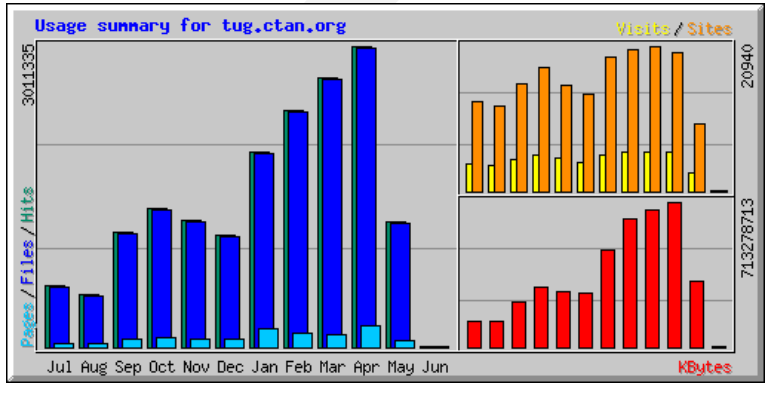
- many journals in mathematics, physics, and computer science use LaTeX
- some major publishers use TeX behind the scenes for book pages
- TeX markup is *de facto* standard in mathematics, physics, and computer science

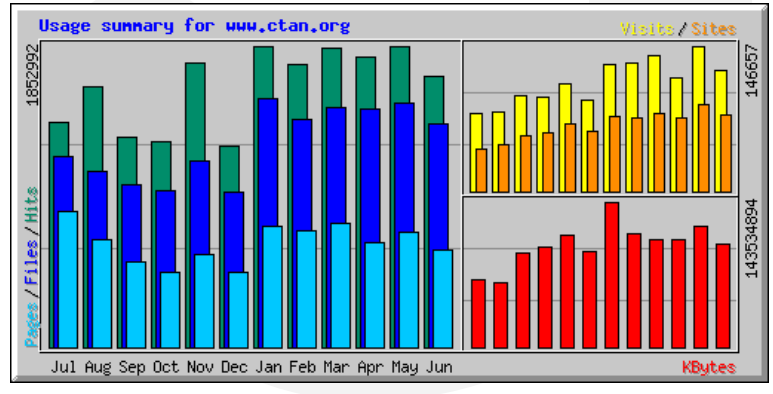
CTAN Archives

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







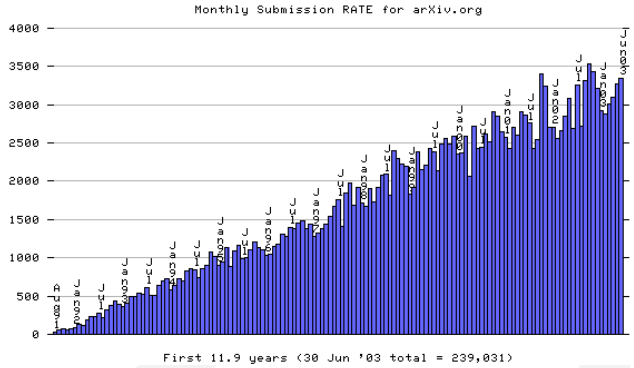


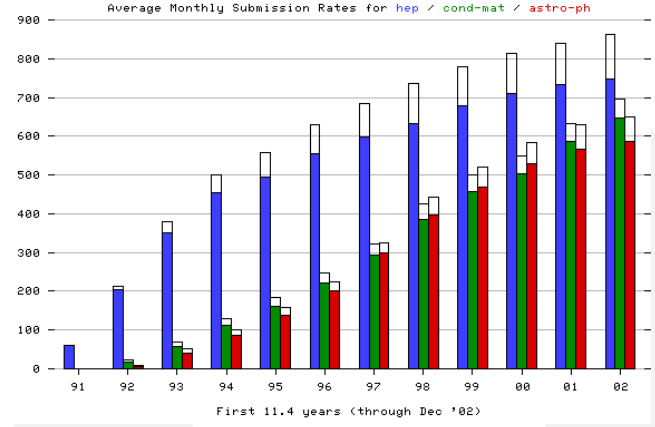
CTAN Archives

- 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.ncstrl.org/`)

BibTeX bibliography archives

- DBLP: 187,494 [XML $\longrightarrow$ BibTeX]
- 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 TEX 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 TEX 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 TEX do right? (cont.)

- `\catcode` adds great power, freeing TEX 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)
- TEX is *stable* and *reliable*
- books illustrated by Duane Bibby

What did TEX 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 TEX 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 TEX 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 TEX into single directory (cf. `ghostscript`)

What did TEX 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 TEX)

Future directions

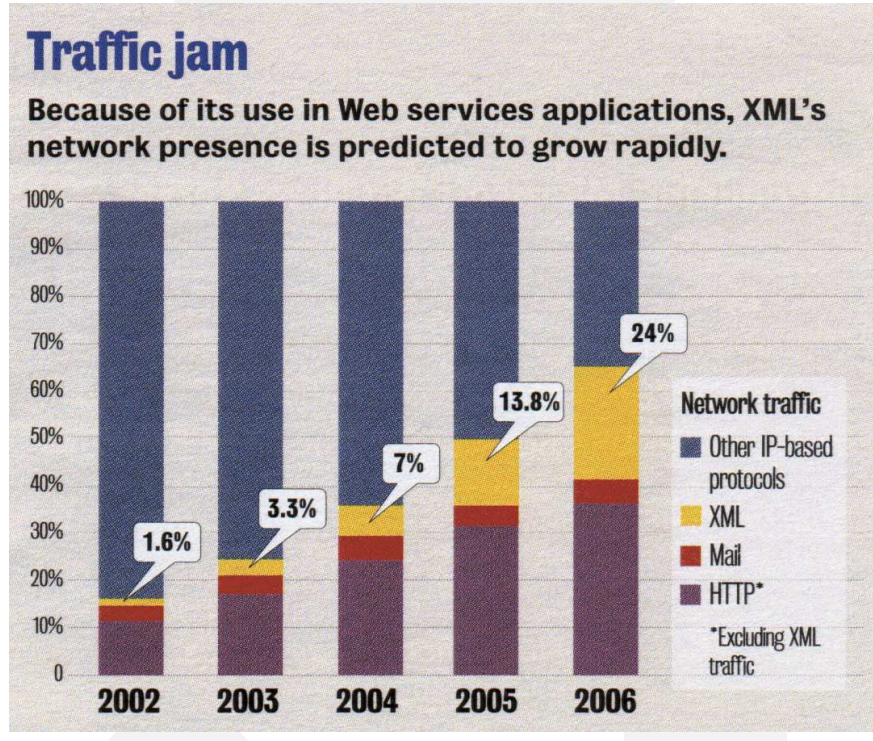
- XML, XML, XML, XML, XML, XML, XML, XML
  [cf. NPR *Morning Edition* story 3 July 2003]
- Unicode and ISO10646 character encoding:

  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

$F_0 = 1$

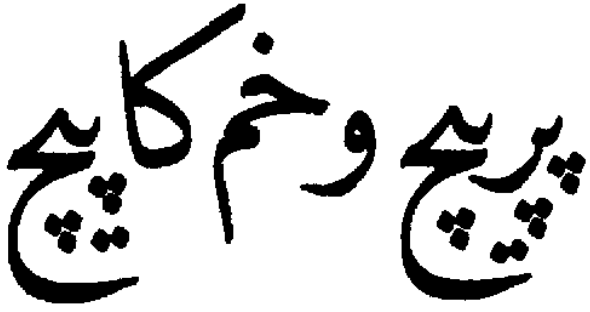
$F_n = nF_{n-1} \quad (n > 0)$

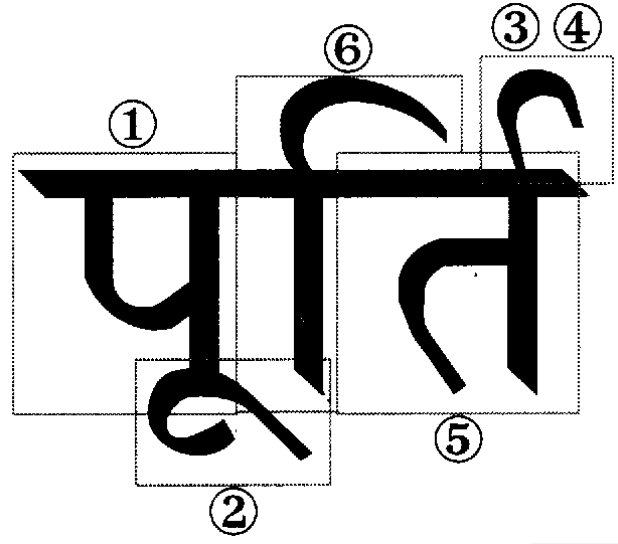
XML (`http://www.w3.org/TR/xexpr/`)

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

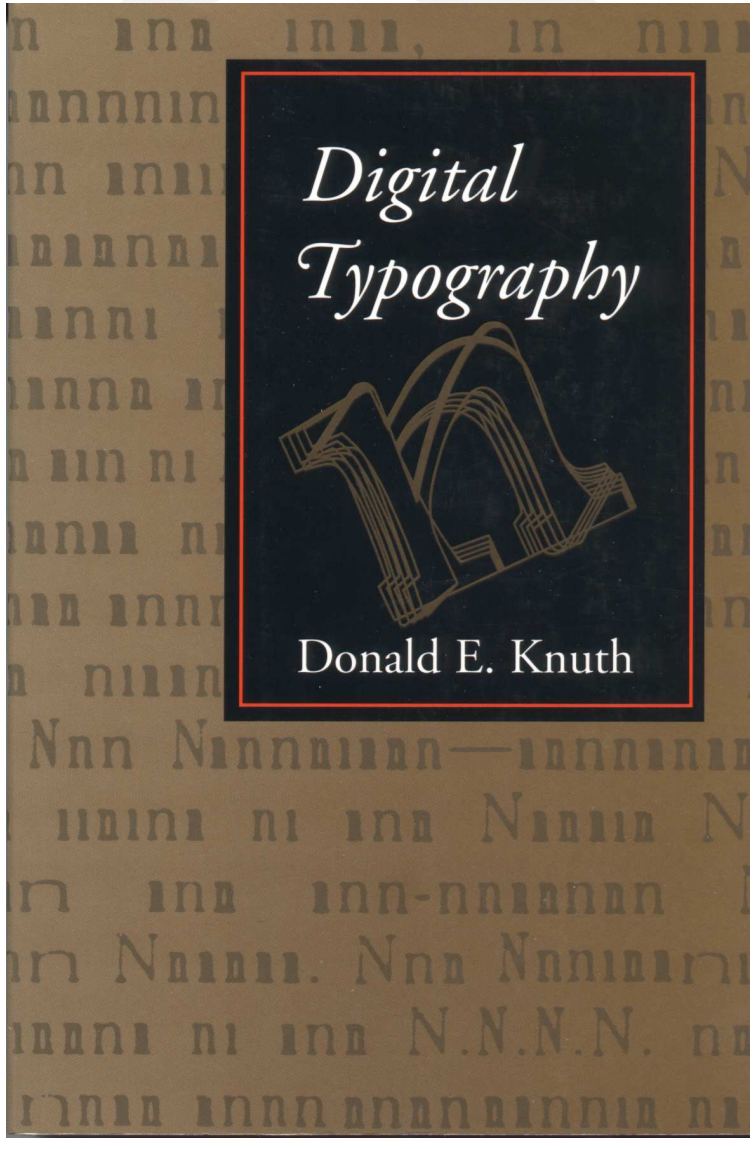
Future directions: Unicode

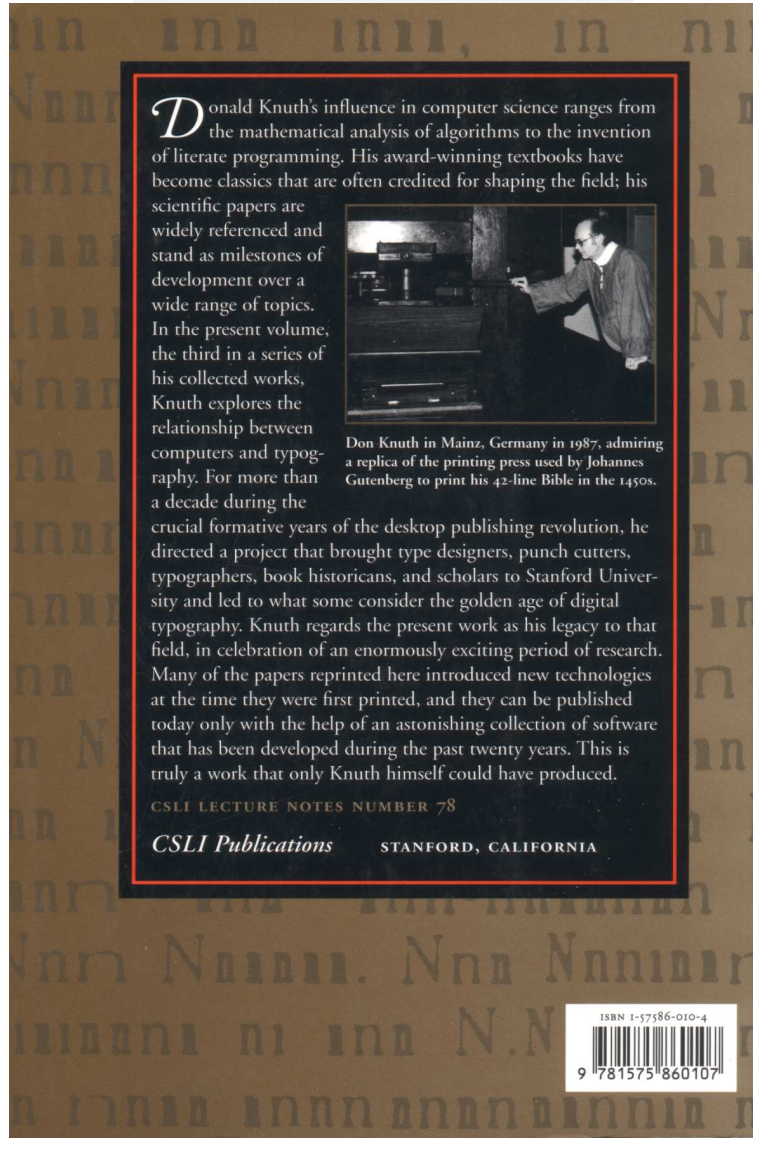
Typesetting some languages is hard!



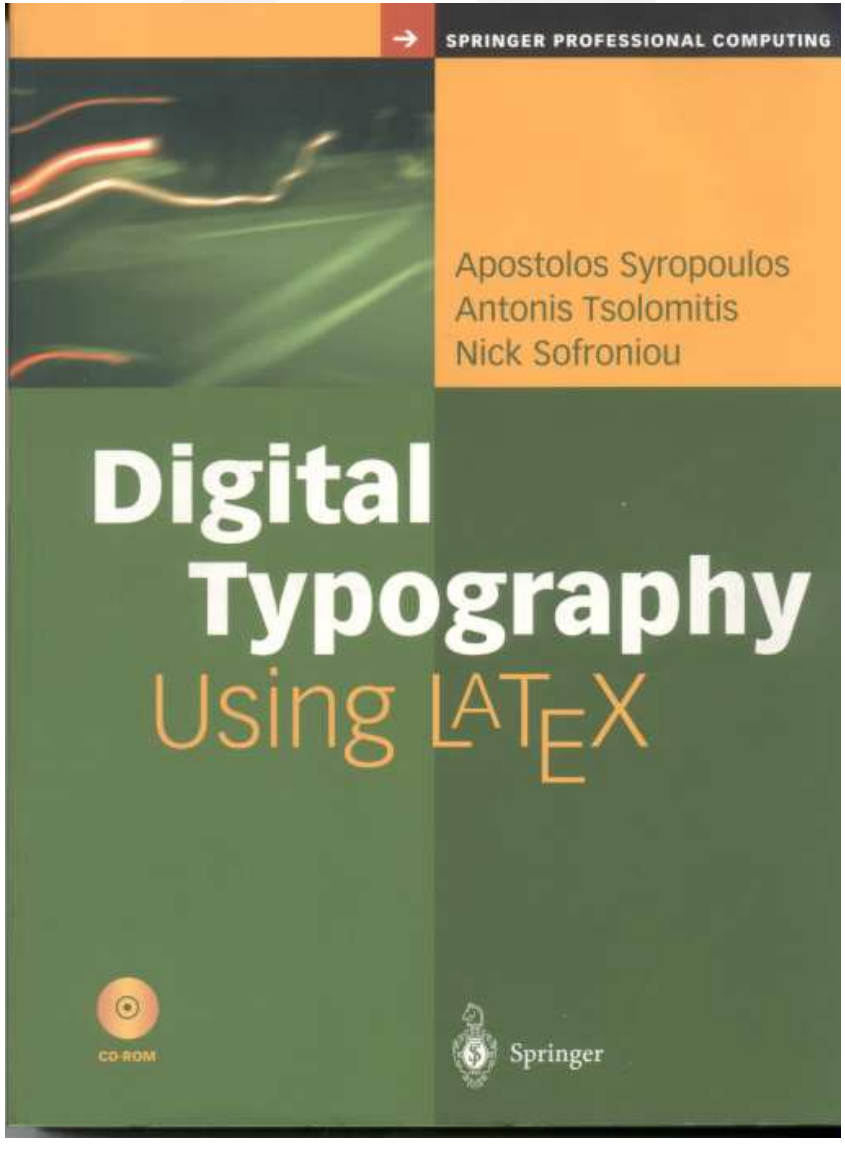


Books of note

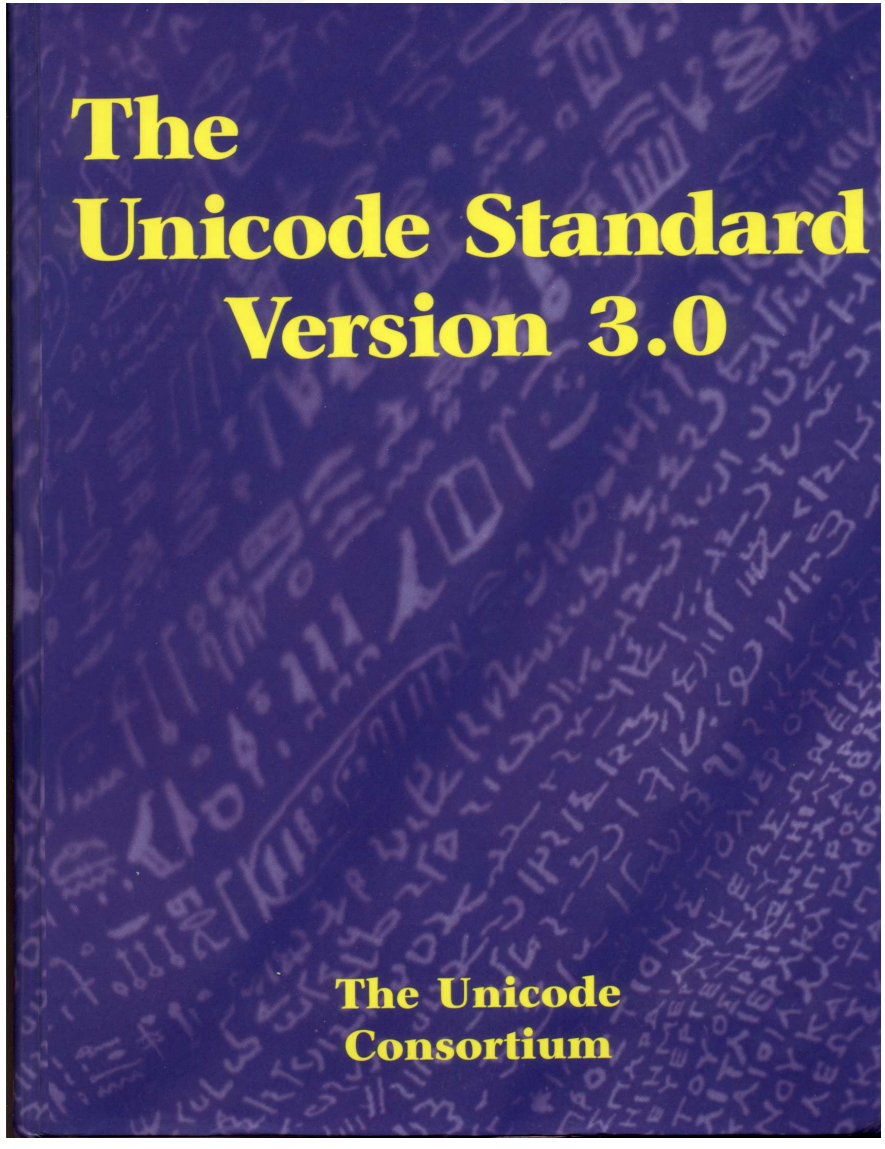




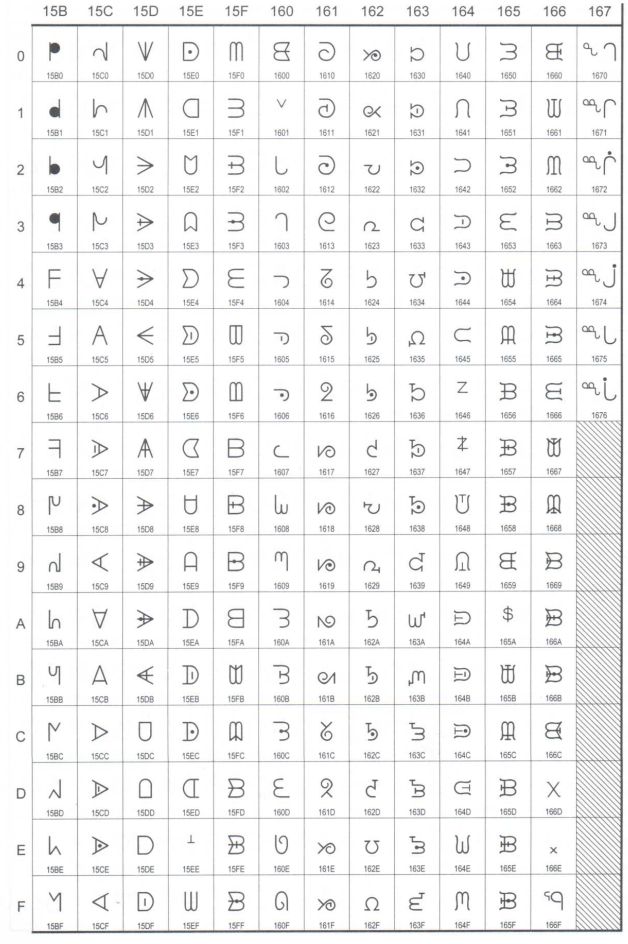
Books of note



Books of note (cont.)



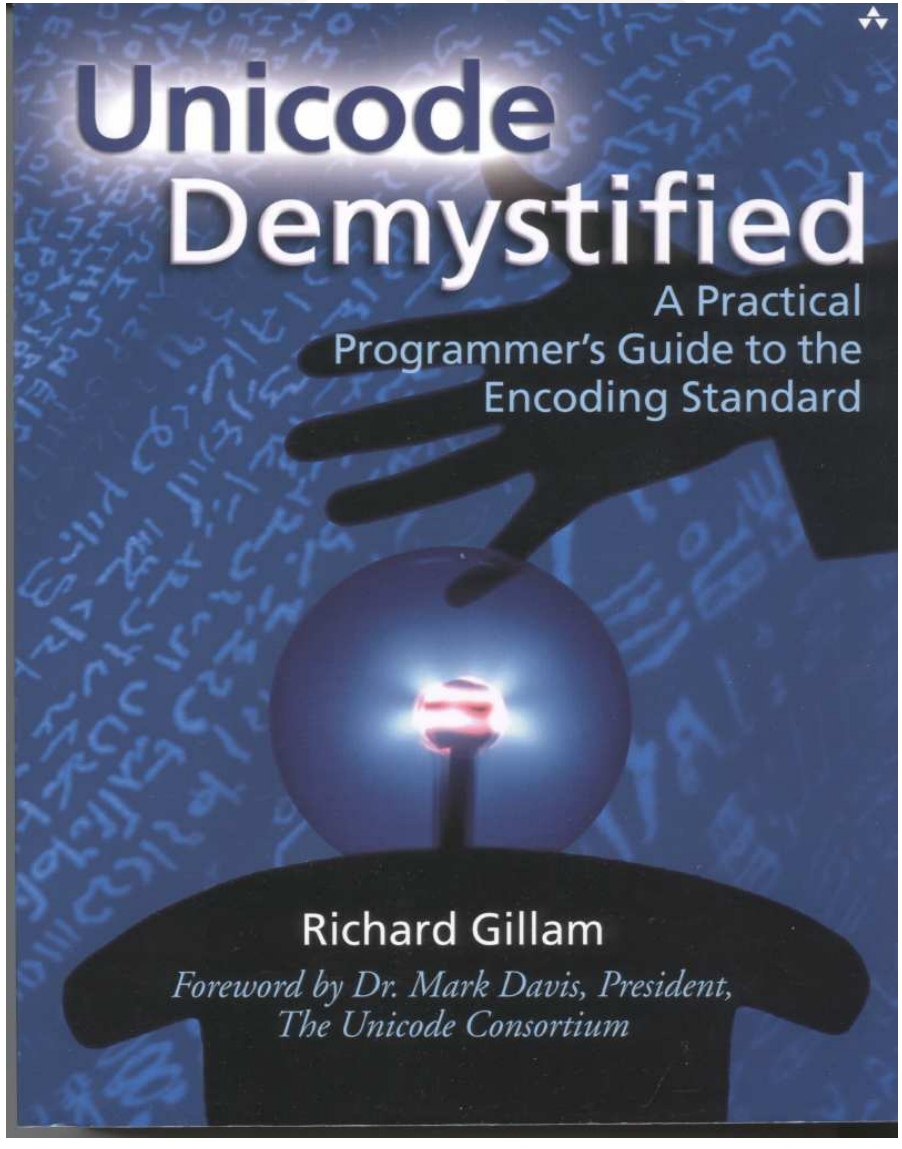
Books of note (cont.)

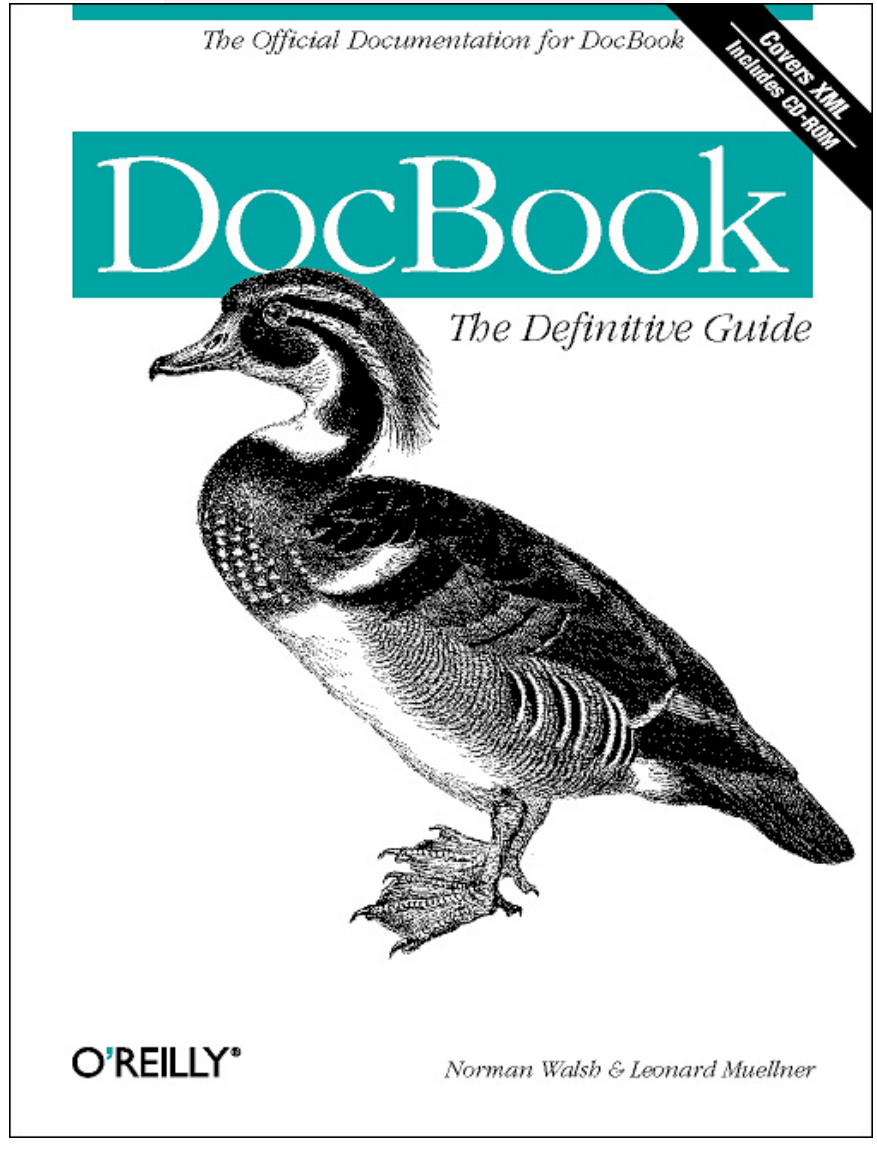


Syllables

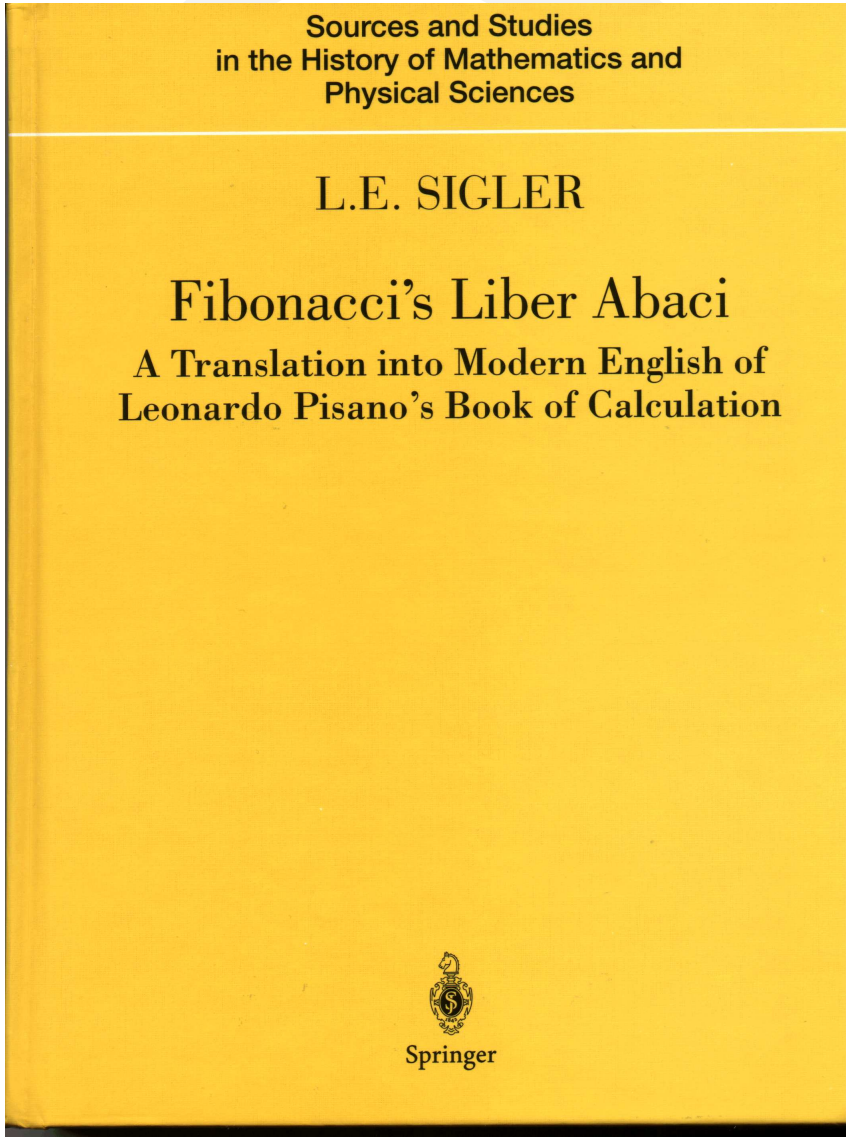
1401 ᐁ CANADIAN SYLLABICS E
1402 ᐂ CANADIAN SYLLABICS AAI
1403 ᐃ CANADIAN SYLLABICS I
1404 ᐄ CANADIAN SYLLABICS II
1405 ᐅ CANADIAN SYLLABICS O
1406 ᐆ CANADIAN SYLLABICS OO
1407 ᐇ CANADIAN SYLLABICS Y-CREE OO
1408 ᐈ CANADIAN SYLLABICS CARRIER EE
1409 ᐉ CANADIAN SYLLABICS CARRIER I
140A ᐊ CANADIAN SYLLABICS A
140B ᐋ CANADIAN SYLLABICS AA
140C ᐌ CANADIAN SYLLABICS WE
140D ᐍ CANADIAN SYLLABICS WEST-CREE WE
140E ᐎ CANADIAN SYLLABICS WI
140F ᐏ CANADIAN SYLLABICS WEST-CREE WI
1410 ᐐ CANADIAN SYLLABICS WII
1411 ᐑ CANADIAN SYLLABICS WEST-CREE WII
1412 ᐒ CANADIAN SYLLABICS WO
1413 ᐓ CANADIAN SYLLABICS WEST-CREE WO
1414 ᐔ CANADIAN SYLLABICS WOO
1415 ᐕ CANADIAN SYLLABICS WEST-CREE WOO
1416 ᐖ CANADIAN SYLLABICS NASKAPI WOO
1417 ᐗ CANADIAN SYLLABICS WA
1418 ᐘ CANADIAN SYLLABICS WEST-CREE WA
1419 ᐙ CANADIAN SYLLABICS WAA
141A ᐚ CANADIAN SYLLABICS WEST-CREE WAA
141B ᐛ CANADIAN SYLLABICS NASKAPI WAA
141C ᐜ CANADIAN SYLLABICS AI
141D ᐝ CANADIAN SYLLABICS Y-CREE W
141E ᐞ CANADIAN SYLLABICS GLOTTAL STOP
141F ᐟ CANADIAN SYLLABICS FINAL ACUTE
1420 ᐠ CANADIAN SYLLABICS FINAL GRAVE
1421 ᐡ CANADIAN SYLLABICS FINAL BOTTOM HALF RING
1422 ᐢ CANADIAN SYLLABICS FINAL TOP HALF RING
1423 ᐣ CANADIAN SYLLABICS FINAL RIGHT HALF RING
1424 ᐤ CANADIAN SYLLABICS FINAL RING
1425 ᐥ CANADIAN SYLLABICS FINAL DOUBLE ACUTE
1426 ᐦ CANADIAN SYLLABICS FINAL DOUBLE SHORT VERTICAL STROKES
1427 ᐧ CANADIAN SYLLABICS FINAL MIDDLE DOT
1428 ᐨ CANADIAN SYLLABICS FINAL SHORT HORIZONTAL STROKE
1429 ᐩ CANADIAN SYLLABICS FINAL PLUS
142A ᐪ CANADIAN SYLLABICS FINAL DOWN TACK
142B ᐫ CANADIAN SYLLABICS EN
142C ᐬ CANADIAN SYLLABICS IN
142D ᐭ CANADIAN SYLLABICS ON
142E ᐮ CANADIAN SYLLABICS AN
142F ᐯ CANADIAN SYLLABICS PE
1430 ᐰ CANADIAN SYLLABICS PAAI
1431 ᐱ CANADIAN SYLLABICS PI
1432 ᐲ CANADIAN SYLLABICS PII
1433 ᐳ CANADIAN SYLLABICS PO
1434 ᐴ CANADIAN SYLLABICS POO
1435 ᐵ CANADIAN SYLLABICS Y-CREE POO
1436 ᐶ CANADIAN SYLLABICS CARRIER HEE
1437 ᐷ CANADIAN SYLLABICS CARRIER HI
1438 ᐸ CANADIAN SYLLABICS PA
1439 ᐹ CANADIAN SYLLABICS PAA
143A ᐺ CANADIAN SYLLABICS PWE
143B ᐻ CANADIAN SYLLABICS WEST-CREE PWE
143C ᐼ CANADIAN SYLLABICS PWI
143D ᐽ CANADIAN SYLLABICS WEST-CREE PWI
143E ᐾ CANADIAN SYLLABICS PWII
143F ᐿ CANADIAN SYLLABICS WEST-CREE PWII
1440 ᑀ CANADIAN SYLLABICS PWO
1441 ᑁ CANADIAN SYLLABICS WEST-CREE PWO
1442 ᑂ CANADIAN SYLLABICS PWOO
1443 ᑃ CANADIAN SYLLABICS WEST-CREE PWOO
1444 ᑄ CANADIAN SYLLABICS PWA
1445 ᑅ CANADIAN SYLLABICS WEST-CREE PWA
1446 ᑆ CANADIAN SYLLABICS PWAA
1447 ᑇ CANADIAN SYLLABICS WEST-CREE PWAA
1448 ᑈ CANADIAN SYLLABICS Y-CREE PWAA
1449 ᑉ CANADIAN SYLLABICS P
144A ᑊ CANADIAN SYLLABICS WEST-CREE P
144B ᑋ CANADIAN SYLLABICS CARRIER H
144C ᑌ CANADIAN SYLLABICS TE
144D ᑍ CANADIAN SYLLABICS TAAI
144E ᑎ CANADIAN SYLLABICS TI
144F ᑏ CANADIAN SYLLABICS TII
1450 ᑐ CANADIAN SYLLABICS TO
1451 ᑑ CANADIAN SYLLABICS TOO
1452 ᑒ CANADIAN SYLLABICS Y-CREE TOO
1453 ᑓ CANADIAN SYLLABICS CARRIER DEE
1454 ᑔ CANADIAN SYLLABICS CARRIER DI
1455 ᑕ CANADIAN SYLLABICS TA

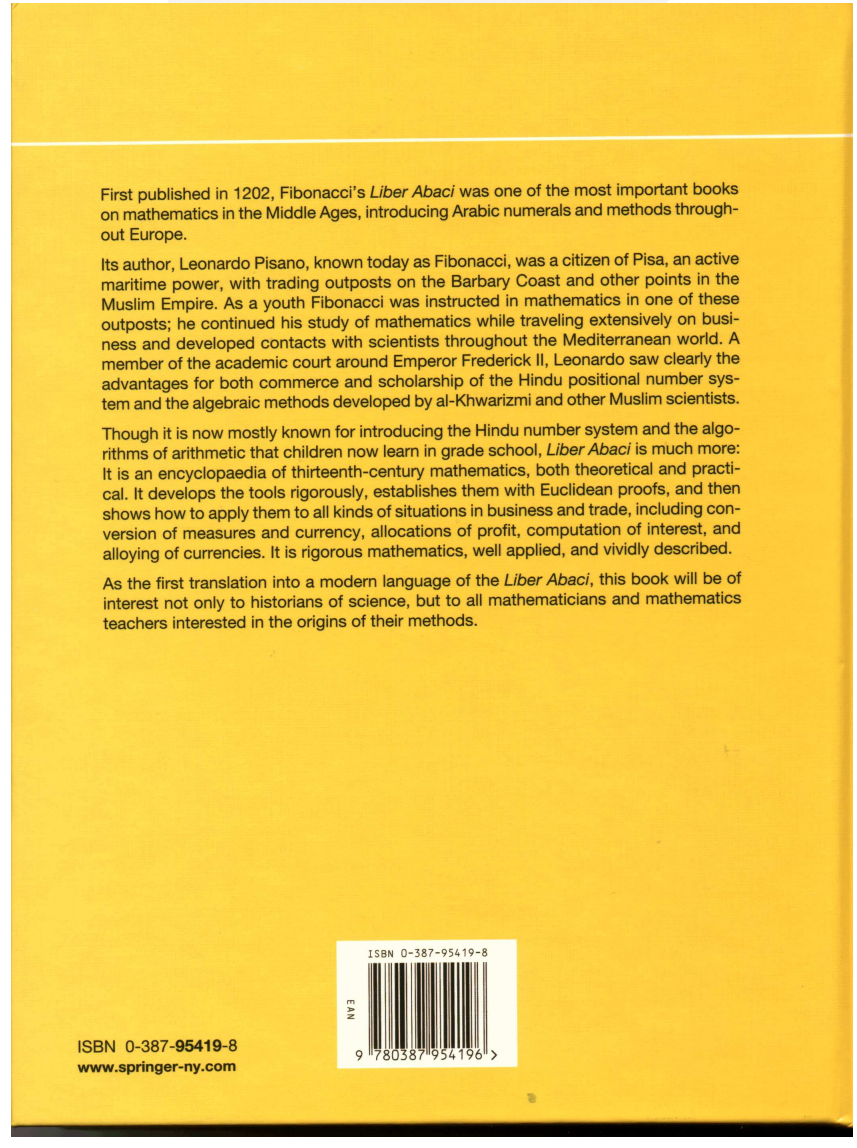
Books of note (cont.)





Books of note (cont.)





That's all folks

```
\drinkfill
\relax
\bye
```