UNIVERSITY OF WATERLOO
COMPUTER SCIENCE DEPARTMENT

Computerized Typesetting of Technical Documents

Richard J. Beach

June 1977
CS-77-38

UNIVERSITY OF WATERLOO
COMPUTER SCIENCE DEPARTMENT

Computerized Typesetting of Technical Documents

Richard J. Beach

Department of Computer Science
University of Waterloo

June 1977

This paper was presented to the
8th Ontario Universities Computing Conference
May 30-31 1977, Waterloo, Ontario, Canada

Computerized Typsetting of Technical Documents

Richard J. Beach

Department of Computer Science
University of Waterloo
Waterloo, Ontario, Canada

Universities produce a great deal of technical material each year. Researchers are expected to contribute to scholarly journals and books, but within the university, reference material, course notes, computer software documentation, conference proceedings, etc. are also produced. For some time, computer systems have been used to edit, store, and produce these technical documents. This paper will describe the further step of producing "book quality" material from a computer system utilizing photocomposition output devices.

Computing centres have had the software facilities for editing and formatting text files for a number of years: programs such as ATS[1], QED[2], Wylbur[3], Runoff[4], ROFF[5], and SCRIPT[6], to name a few in use just at Waterloo. In fact, all large computer systems at Waterloo (IBM 360/75 and 370/158, Honeywell 66/60, DEC PDP-11/45 running UNIX, and even some minicomputers) support both text editing and document formatting.

A *text editor* provides facilities for entering and correcting documents. The user interacts with the editor via a computer terminal, and specifies functions such as append, insert, replace, and delete text in a variety of ways. Documents are stored in a *file system*, frequently in small chunks corresponding to a section or chapter of a large document. These files are read or written by the editor, and may be listed quickly on a high speed line printer. File systems are also supported by the operations staff with backup services.

A *text formatter* manipulates the appearance of a document, directed by commands within the text file to create paragraphs, centre titles, indent lists of points, justify right margins, and so on. Clever formatters can handle layout problems such as placing headings and footings on each page, positioning footnotes and tables, and providing multiple columns per page, etc. Better formatters allow you to use *macros* to simplify the entering of formatting commands. Skilled users can build macros for repetitive tasks or for standardizing the layout. For instance, a macro could be defined for a heading which is to be numbered sequentially automatically, always underlined, set off from the preceeding and trailing paragraphs by extra space. Then each heading in the document would always appear in the same format.

A *photocomposition output device* is a electomechanical device capable of producing master copies of documents using a photographic process. Characters are held on a glass disk or filmstrip in negative form and projected onto photographic paper by a strobe flash. The characters are sized by lenses, and positioned by mirrors onto the photo paper. The device is controlled by *teletypesetting codes* (TTS codes) usually punched onto paper tape, or possibly recorded on magnetic tape, or transferred by a communications device. The exposed photo paper is then developed and cut into pages for reproduction.

The use of line printers or even hardcopy terminals has resulted in a poor quality document in terms of appearance and legibility. Line printers were designed to go fast, and the print quality suffers; impact printers go out of adjustment, wavy lines appear, characters wear down and are no longer crisp or well defined, and reproduction exaggerates this poor definition even more. The IBM 1403 printer, with a special print train, and a one-time mylar ribbon is about the only satisfactory scheme; for instance the Mathematics Graduate office has issued a memo to all students to check with their supervisors about the quality of the line printer before it is used for printing their theses.

Furthermore, computer terminals and line printers print using monospace characters. Such characters are all of equal width, which differs from books or newspapers where the letter 'm' is 2 or 3 times wider than the letter 'i'. People who read computer-printed material for the first time, frequently notice the difference because one's overwhelming experience is with books and newspapers with proportional spacing.

Another problem faced by the Math Grad office is the increased tendency for mathematical equations to be linearized. The line printer or terminal restrict the way in which fractions and superscript expressions can be printed. This makes them come out looking like FORTRAN expressions, rather than the traditional mathematics two-dimensional presentation.

Authors of technical documents who use computer editing systems are faced with a dilema: either accept the low quality of computer printed materials, or send the manuscript to a printer for typesetting. The latter creates a new problem as the typesetter must rekeyboard the manuscript, thus creating new typographical errors and forcing more tedious proofreading. However, recent price changes and technology advances have brought the typesetting device to the computing centre. Devices are now priced as low as $6,000 and can be attached as a computer output device to provide in-house publishing of computer formatted documents. Software developments make it possible to format both straight text material and more exotic mathematical notation.

Typesetting Equipment at Waterloo

The first typesetter for technical printing at Waterloo was a Photon Econosetter purchased for research purposes in the Fall of 1974. It was a minimum system worth $10,000 and provided 4 fonts and 4 character sizes. It was initially driven by paper tape, although we quickly built an asynchronous communications interface to connect the typesetter with a terminal and the Honeywell 66/60. This interface allows us to run the typesetter over a telephone line if we choose, however it

is presently connected via a hardwired line. With the typesetter on-line, software development proceeded, the first product being PROFF[7], which gave us ROFF formatting features for the Photon typesetter. PROFF generates typesetting codes according to the formatting commands imbedded in the text file, and relies on the typesetter to perform the actual justification and hyphenation. Powerful macro processing capabilities in PROFF have permitted extensions such as mathematical equation typesetting to be investigated. The development of a new typesetting package is underway, one which will allow us to typeset running heads and feet, to compose more complex mathematical notation, to automatically generate table of contents and indices, and to produce multiple column output and many other layout designs.

The next equipment acquisition was a pair of old (circa 1970) Photon 532 typesetters donated to the Arts faculty by the Toronto Star newspaper. The typesetters were used by the Star for setting display advertising, and had the capacity for 32 type fonts and 23 character sizes. Unfortunately, they were paper tape driven, and came with only newspaper fonts. Plans are underway to modify the Econosetter's communications interface to put the 532s on-line, and we have discovered a collection of used 532 fonts containing European languages, Russian, Greek and Math symbols. Software development has been slower due to the use of paper tape and its frustrations, but we now have software to do most of the PROFF functions.

The University's Graphic Services department acquired a Mergenthaler VIP in August, 1976. It is a slightly faster machine, with 6 fonts, 12 character sizes and a higher price, $40,000. It uses paper tape only, and came with a stand-alone keyboard/editing station. Only standard text typefaces are presently available, although additional fonts can be purchased at reasonable costs. No computer output has yet been processed.

A Production Scenario

To help visualize the role of a computer system in the production of a technical report, let me describe the steps and procedures necessary to use the Photon Econosetter in Math.

1. Enter document.

The document is usually entered by the author or a typist on a computer terminal from a handwritten manuscript. Some authors compose their manuscripts at the terminal, in pieces, over long periods of time, such as writing the sections of a reference manual. The text is stored in a file, with large documents broken up into sections, each one in its own file.

2. Edit document for typographical errors.

Editing for typographical errors is usually done by the typist using a higher speed video (TV screen) terminal to check the document a screenful at a time. Corrections are made to the file and saved. Any text editor will do, as the typesetting commands appear simply as lines of text to be later interpreted by a formatter

program. Assistance may be provided by a "typo" program, which checks the spelling of words against an English dictionary and a file of jargon words.

3. Proof format.

Using a text formatter, the document is formatted and then printed on the line printer. Formatting macros simulate typesetter actions, such as simulating boldface by overstriking the characters 2 or 3 times, and italics by underlining.

Usually the editing and proofing steps are repeated as draft revisions take place, sometimes with major surgery being performed! Authors may wish to rearrange topics for emphasis or clarity, or information may become obsolete before it is printed and have to be replaced, especially with software documentation. Also, the authors find this method less resistive to change, permitting them to make these changes at will until the last moment, with the assurance that new drafts will not contain errors in unchanged material. The file system and powerful text editing features are central to providing this facility.

4. Typeset proof copy.

The document is then formatted for typesetting, and spooled to be typeset. The spooling may be to a file for later transfer via the Honeywell to Photon interface, or to paper tape for the Photon 532. The typesetter produces exposed photographic material that must be developed, dried, and proofread.

Many users have found proofreading typeset copy at this stage to be much more effective, primarily because the document is easier to read and errors appear more obvious. We would like to get a faster typesetter to satisfy the demand for proof copy. An alternative would be to use a graphics terminal or plotter to preview output, if the software to emulate the typesetter was written.

Corrections at the typeset proof stage arise because the document is subjected to closer scrutiny, or because the formatting and layout decisions need adjustment. For example, it is poor practice to have only a single line or two on the last page, so the text is reworked to get them onto the last few pages. Changes at this stage involve using the text editor and may be substantial enough to require line printer proof copy again. The cycle continues until a final typeset copy is satisfactory.

5. Cut and paste.

Photographic output from the typesetter must be cut and mounted onto pages for reproduction. Frequently the formatter has already provided page headings and footings, so this is simply a separation process; however, many composition decisions about the placement of figures, footnotes or page breaks are preferred to be made manually. This is because automatic page layout, when there are as many degrees of freedom as with a typesetter, often involve either too much software or too many reruns, or both. And recalling that we have a slow typesetter, many users prefer to get parts of the document right each pass, and then to cut and paste the results together.

6. Reproduction.

The pages are now available for printing, most frequently by photo-offset, but depending on the number of copies and quality desired, other means may be used.

From personal experience, the whole process, including proofreading and author's revisions, can take less than an 8 hour day for a 6 page report, or up to 3 months for a 130 page book. Certainly time is greatly saved by using the file system to store the document and by revising only small files containing the appropriate section.

Typesetting Features for Technical Reports

This section will describe some of the problems in technical publishing that have been solved with our typesetting software, PROFF. Since a complete rewrite of the software is underway, it is possible to comment with hindsight as to the ways to simplify the process and the user interface.

1. Mathematics

Simple inclusion of math symbols is no problem; this can be accomplished by including math symbols on the typesetter, and selecting the appropriate characters as required. Therefore, examples like $x \geqslant \alpha$ are easy. Superscripts and subscripts can be included as superior and inferior figures. However, the provision of superscripts and subscripts that are not simply numbers requires both a character size change and a temporary movement of the typesetter paper. Thus the example,

$$y = e^{2x+\ln x} = e^{2x}e^{\ln x} = xe^{2x}$$

is a bit more trouble. Building up summation or integral formulae involves more intricate size changes and typesetter movement, plus the ability to back up and overstrike characters. Thus the examples:

$$\sum_{k=1}^{\infty} \mu_k \quad \text{or} \quad \int_{\alpha}^{\beta} \sin x \, dx$$

are a bit more fun. These examples demonstrate that we can typeset some mathematics, and have built the macros to make the job easier. In fact, the macro call needed to make the integral was

```
↑(int ↑(alpha) ↑(beta)) sin↑(i x) ↑(i dx)
```

where the ↑ character serves as a macro marker, causing whatever is inside the following parentheses to be treated as a macro. The *int* macro takes 2 arguments: the from-limit and the to-limit; the *i* macro causes its argument to be typeset in italics; and the 2 macros *alpha* and *beta* cause the appropriate Greek letter to be typeset.

2. Computer programs.

Publishers always have trouble representing computer programs accurately in text books because they do not use the monospace (equal width) characters used by terminals and line printers. Also, they do not realize the significance of blanks in

computer programs, and often put in variable amounts of space causing some confusion. Quotes around character strings are often changed to opening and closing quotation marks, when they should be left as apostrophes.

PROFF has a *program mode* that tries to accomodate these difficulties as best as possible, and the following examples are evidence:

FORTRAN:

```
      REAL X
      INTEGER I
      DO 123 I=1,10
      X = 3.14159 / FLOAT( I )
  123 PRINT, 'SIN( ', X, ') IS ', SIN(X)
      STOP
      END
```

B:

```
main( argc, argv)    /*  demonstration program  */
   {  auto i;
      for ( i=0; i < argc; i++)
          printf( "argument %d is %s*n", i, argv[i]);
      }
```

APL:

```
      ∇ DRILL;I;J;K
[1]   I←?10;'+';J←?10;' = ?'
[2]   ←3×'Q'≠K←□
[3]   ←⍳K=I+J
[4]   ←2,⍴□←'SORRY. TRY AGAIN.'
[5]   ∇
```

3. Overhead slides.

Some professors have used our typesetter's crisp image and larger point sizes to create overhead slides for talks and presentations. The slides for this talk were prepared this way, using the 18-point Times Roman Boldface font, and then copying the output via a Xerox machine (to get carbon content) and then using 3M Thermofax transparency film. Tables and point-form formats are the most-frequently-used formatting functions, as well as centered titles.

4. Reports and manuals.

The largest volume of technical typesetting work has been done on the Photon Econosetter, and consists mainly of Computer Science research reports, and Math Faculty Computing Facility manuals. In the academic year 1976-1977, approximately a dozen technical reports, two theses, one book, and four manuals were produced. Most of this work relied heavily on the use of macros to simplify the entering task. Professors and secretaries alike could prepare a straight-forward report with commands like: .title, .author, .report, .chapter, .section, .para, .list, .item, etc. It was found that the prerequisite level of typesetting knowledge was thus

greatly reduced, and that mnemonic names were more convenient for speed typists. Also, longer names were easier to learn and remember, and easier to proofread because they were more readable than abbreviations.

Cautions

The use of typesetting technology can be accomplished for a small capital investment and little previous experience. However, let me describe some of the pitfalls that you might encounter:

1. Aesthetics

Better quality output tends to generate better quality criticism. Now that you have book quality computer generated output, the style is questioned. You should seek the guidance of co-operative graphics designers to make the most effective use of your equipment. The selection of type styles, type sizes, paragraph styles, headings and footings, word and letter spacing, are all specifications given to the commercial typesetter by the graphics designer, who has incorporated much experience and many aesthetic, visual impact, and communication considerations into the design. Please seek the advice of a graphics designer, to at least establish a "house style" for typeset work.

2. Software.

These same aesthetic considerations can make the computer formatting software extremely complicated. Typesetting is much more than just aligning the right margin and hyphenating words. Using multiple columns, providing running headings and footings, avoiding one-line widows on pages, formatting columnar tables, providing vertical justification, and so on, are difficult algorithms. The typesetter requires that you control many more degrees of freedom than a line printer, and the aesthetics are often in conflict with what is easy to program. This is a prime reason why cutting and pasting pages together is still so widely practiced.

3. People.

The volume of work which can be handled by computer text editing systems almost always implies the involvement of non-programmer, non-technical staff. Secretaries are certainly trainable and can become very proficient, but they are not at all tolerant of inconsiderate computer programs or terminals. For example, the Psychology department at Waterloo has not been successful in retraining their secretaries from ATS to Wylbur and SCRIPT, due to both software and terminal complaints. On the other hand, Environmental Studies has had almost 100% success in going from ATS to CMS SCRIPT.

Computer terminal keyboards frequently are not equivalent to those of an electric typewriter. Either their layout resembles a Teletype model 33, or the keyboard contains an imposing array of extra function keys. Speed typists will not be able to perform well initially, unless the keyboard they use matches those they trained and practised with. Also, many electronic keyboards have very poor tactile feedback, quite different from an office typewriter. Often the tactile sensation is too

soft, or too hard, or a beeper is used for each key struck, which distracts the typist. Another distraction is the delayed typing noise made by buffered terminals, such as the Diablo or DECwriter terminals, which throws off their typing rythym.

Speed typists become irritated if they must wait at the end of each line for permission from a computer to proceed, especially if the keyboard mechanically locks. Full duplex communications, allowing type-ahead, is mandantory for such typists, so that they can type rythmicly, at speed, letting the computer catch up to them instead of them waiting.

The filing system used by the typist must be "fail safe" and forgiving. Accidental destruction of file contents must be avoided, perhaps requiring confirmation. Files, once stored, must be recoverable, both in the event of software failure, or file system device failure. An archive facility makes the long term storage of documents realistic, since the space occupied by text files can grow quite dramatically as the service enjoys success.

Now consider the state of computer software which these non-technical people must use. Most text editors have been designed by programmers, for programmers, without much human engineering concern. The programs are inconsistent, frequently impolite, and often very terse or cryptic. Cosmetic changes to the error messages and command analyzers can remedy some frustration, but new designs may be the only solution.

4. Equipment.

Cheaper typesetters are invariably slow; 12 characters per second has been the average speed over a year's operation of our on-line typsetter. Complicated mathematical notation can take up to an hour per page. Mathematics and foreign languages require additional fonts, and many typesetters do not permit a large number of type styles simultaneously; for example, the Econosetter allows only 4, the VIP 6. Furthermore, it takes 3 of those 4 fonts to hold Times Roman, Boldface, and Italics, leaving only 1 font for exotic typefaces. Another restriction affects the mixing of typefaces: glass disks or filmstrips contain several fonts, thereby limiting the mixing possiblities if you can only mount 1 or 2 such disks.

Developing Trends

Standalone Systems.

Microprocessor computers, floppy disks and cheaper typesetters are being combined into a standalone editing, storage and typesetting system. One example, the Compugraphic Editwriter 7500 costs about $16,000 and has a CRT display, floppy disk, an 8 font, 12 size typesetter controlled by 2 microprocessors. Connections via communications lines to host computers and to word processing systems are being developed. This would be an intelligent terminal!

Computer Terminals

The Diablo print mechanism provides a breakthrough in quality hardcopy terminals that go fast. Ink jet technology may appear in hard copy terminals soon too,

giving both speed and image quality. ASCII line support is needed for most new terminals, and these faster terminals require special modems or dedicated communications ports.

Full duplex operation, parity selection, line speed etc. are increasingly provided by switches on these terminals, so that you may configure the terminal as needed.

Video terminals are improving, both in terms of display quality and size, but also in keyboards. These terminal keyboards have had notoriously bad layouts and feel for speed typists.

Graphics terminals are less expensive now (under $5,000 for a Tektronix storage tube terminal) and can provide typesetting preview and editing. A high line speed is necessary to provide reasonable response time for image requests.

Digital Typesetters

These devices combine line printer speed with typeset quality. They are currently more expensive, being in the $40-100,000 range (and up, up, up) but provide tremendous flexibility. The fonts are stored digitally and recreated by either a high resolution CRT or laser beam. It is much easier to mix fonts from a stored digital library, or even to augment your library with special symbols, accents, logos, etc. The output media is currently still photographic paper, but experiments with laser typesetters to produce direct-view media or exposed printing plates is well underway. The former eliminates photographic processing, and the latter eliminates page makeup and plate production completely, a tremendous time and cost saver. With the speed advantage, and a cheaper direct-view output media, this would provide convenient proof production. However, to fully utilize the speed, and eliminate manual paste-up, the software support must handle more complicated layout and composition functions. This will require more computing power, and more complicated software in the future.

Graphics

Digital typesetters and improved computer graphics software packages will be combined to allow the direct insertion of charts and figures into technical publications. In fact, laser typesetters will shortly be coupled to laser half-tone scanners, making possible the insertion of photographs and pictures. Otherwise, charts and figures that can be represented as line drawings could be drawn by the typesetter using a modified graphics package. Thus programs that generate charts could direct their output to be included in a typeset document using a digital typesetter.

Conclusion

Computer generated typesetting has reached an economic level where it is relatively inexpensive and straight-forward to use. It is feasible for in-house publishing, such as that done by universities, to make use of computing centres for document preparation and editing for typesetting, or to acquire a stand-alone editing/typesetting system. The savings to the authors come in the form of reduced

input costs when the document is typeset. Editorial control over the manuscript is retained until the last moment, permitting the author to revise and improve the document with confidence that errors will not appear in proofread portions of the file.

Advances in digital typesetters and computer software will increase the flexibility for typesetting technical material. It is a fascinating marriage of computing with the graphic arts, and promises many immediate and potential rewards.

References

[1] ATS/OS Terminal Operations Manual, Internatinal Business Machines, White Plains, N.Y., H20-0589-02.

[2] QED Reference Manual, Mathematics Faculty Computing Facility, University of Waterloo, 1977.

[3] WYLBUR Manual, Computing Centre, University of Waterloo, 1975.

[4] RUNOFF Text Formatter, "Time Sharing Text Editor", Honeywell Information Systems, Waltham, Mass., Order Number DD18.

[5] ROFF Text Formatter, Mathematics Faculty Computing Facility, University of Waterloo, 1977.

[6] SCRIPT Text Formatter, Computing Centre, University of Waterloo, 1977

[7] Beach, R.J., PROFF Text Formatter, Deptartment of Computer Science, University of Waterloo, CS-76-08, 1976.