Catching up
PDF and HTML at the heart

Kees van der Laan
cgl@hetnet.nl

abstract

New hardware not only urged me to get TeX & Co running again, in a richer environment, but I had also to catch up with developments since. Most noteworthy in relation with TeX and documents, are the acceptance of the PDF exchange format and the HTML format, next to the realization of multi-media.

keywords

internet, HTML, multi-media, PDF, pdfTeX, PostScript, WWW

Introduction

When I complemented my (Mac) hardware with the PowerMac 5500/275 a complete new world opened up. For only a year or three I have TeX-ed and emailed happily on my Powerbook 150, next to an odd 5 years on my Mac Classic II. I have used the UNIX mail system for more than 11 years the same holds for TeX & Co... And see, ... the world has really changed of late.

The new emailers in the internet are much more convenient and more economical. The email is no longer handled at a remote machine via terminal access, or via telnet. But by a POP, Post Office Protocol, and a SMTP, Simple Mail Transfer Protocol, on whatever machines they are located. I can just collect and send my email with a snip of the fingers That fast. Off-line, email can be prepared, sorted, read and answered at leasure. No real need to have it on paper, one can increase the size of the font for easy reading and so on. Moreover, MIME (Multipurpose Internet Mail Exchange) allows fancy enclosures as attachments, to be sent encoded, that is economical, at will. Examples? Attached photos as (economical) jpeg format, PDF (Portable Document Format) documents uuencoded.

The latter brings us back to TeX & Co. For Knuth TeX & Co meant TeX and METAFONT, meaning tools for typesetting and creating types. Let us start from the accompanying earlier used, and TeX & Co biased, model. In this model TeX and METAFONT are at the heart, as oldest tools. But, Adobe has turned things upside down with their
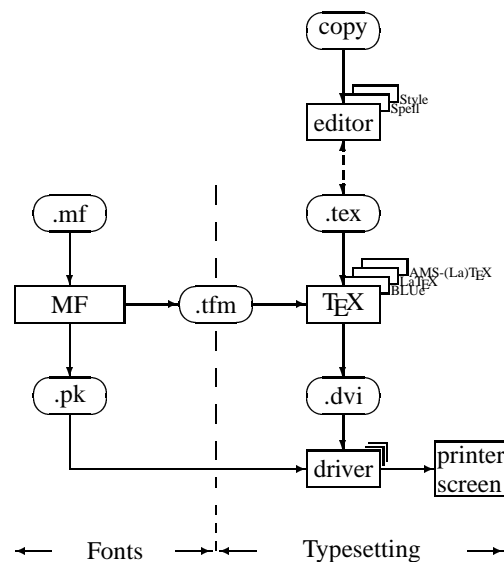
'PS on the clipboard'

and

'PDF as standard for document exchange.'

The latter not only passively to be read from, but also to be reused, modified or enriched with bookmarks, thumbnails, threading and hyperlinks

independently from the tool which created the document, and

independently from the platform or tool to be used

as long as they comply with PDF.

And then there is the internet, with the format HTML, its browsers[1] and its logical World-Wide Web, WWW for short. This stressed another aspect of multi-media documents. The location of (parts of) documents is no longer relevant, as long as they are connected by hyperlinks.

Something has really happened in the nineties.

In short, we have concentrated on books with their typographical traditions and concepts such as spread and so

---

[1] Netscape is a super, active HTML viewers and much more.

on. Now we really have entered the era of multi-media documents, with different, yet to be discovered rules for composing and (re)using.

In this Adobe with its PDF has earned the right of the first-born, the first prototype—PDF and Acrobat—which is catching on, IMHO, with all respect.

Adobe took the lead again.

For TₑX & Co the law of diminishing returns applies.

I have to work harder and really catch up. There is much more to do than before.

TₑX and PostScript

Jacko[2] in one of his lectures stressed the 'octopussy' model:

> EPS at the heart

as the central exchange format for documents, where for example TₑX & Co and Adobe Illustrator could cooperate happily. In order to let this work he developed mf2eps, because MetaPost was not yet in the public domain. He followed in the footsteps of Adobe, and how right he was. But, in the meantime, the concept of hypertext got a boost when the internet took off, with virtually all computers connected into a huge web, the WWW. Adobe picked it up and developed the successor to Postscript, aimed at not only page description, but also at handling hyperlinks, or more generally, at handling and (re)using multi-media documents. In doing so, known deficiences with respect to PostScript, such as the size of the PS files and the (non)availability of fonts, were taken care of.
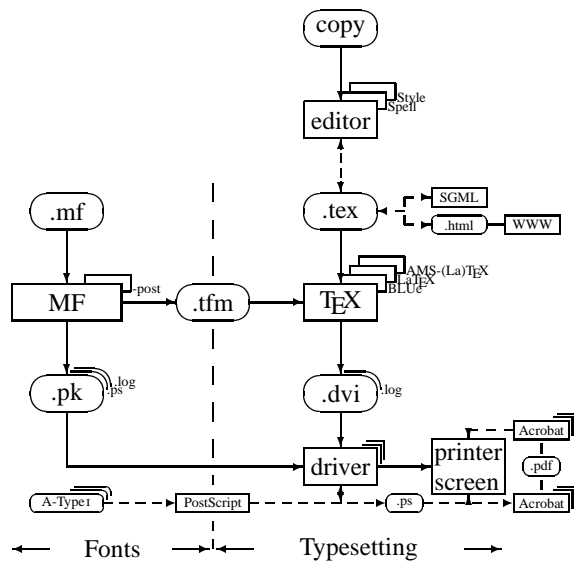
SGML and TₑX

In the eighties SGML (Standardized General Markup Language) ánd TₑX were the buzzwords in the TₑX community.

> SGML as frontend and TₑX the backend formatter

Publishers believe in this model for several reasons. I think that SGML and TₑX are alike, more two of a kind modulo some syntactic sugar, IMHO with all respect. In practical BLUe work there was no much need to go for SGML. Moreover their editors and viewers were not for free, to put it euphemistically. We also faced the problem of the Document Type Definition confusion, and the practical problems with fonts.

That HTML (HyperText Markup Language) took off with the internet. That it used an SGML-biased syntax is a mere coincidence, IMHO, with all respect.

These developments can be added to the-biased-by-TₑX model as follows.

Let us look more closely at HTML and PDF and what these formats and their accompanying tools have brought us.

PDF and TₑX

This is a short-term future of TₑX, if we want TₑX & Co to have a niche in multi-media land.

I believe that the potential of PDF is enormous, and that it can't be overestimated.

I consider the chain

$$.tex \xrightarrow{TeX} .dvi \xrightarrow{dvi2ps} .ps \xrightarrow{ps2pdf} .pdf$$

flexible, maybe it is too flexible. On the other hand little experience with pdfTₑX

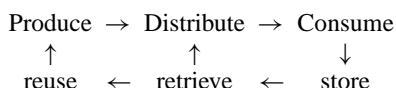$$.tex \xrightarrow{pdfTeX} .pdf$$

was heartwarming.

Adobe's PDFWriter to create .pdf files from without any complying application is powerful, and definitely a way to get things done by other manufacturers.

From the above we might distill that it is important to find the right balance in spending time on the tools involved. TₑX is just one of the many.

With respect to TₑX there is a weakness: the advantage of TₑX as creator is thrown away with the advantage of

---

2 Bogusław Jackowski

abstracting from the creating tools. Not only should there be tools to start with (all)TeX and end with PDF files, but also to go back from PDF to .tex, that is ASCII with TeX markup. I'm afraid that is too cumbersome, and not realistic. Now already we include EPS files for the graphics, for example. My model of the biological invariant of use and reuse does no longer hold for (all)TeX as such. TeX & Co must be seen with respect to cooperating tools around a common file format, the PDF. My idea of the life-cycle of multi-media documents does apply with respect to PDF, with other tools as plug-ins.

$$\begin{array}{ccccc} \text{Produce} & \rightarrow & \text{Distribute} & \rightarrow & \text{Consume} \\ \uparrow & & \uparrow & & \downarrow \\ \text{reuse} & \leftarrow & \text{retrieve} & \leftarrow & \text{store} \end{array}$$

What will reasonable emerge is that a 'federation' of tools, all contributing with their strong sides to the one-and-only PDF file, be it TeX's quality of typesetting math, be it PostScript's suitability toward graphics, or whatever spreadsheet program towards spreadsheets, or whatever database tool towards database reporting, or the emerging and captivating sounds and movies. The latter two are still in their infancy with respect to muti-media, though I have watched baffling demos already.

That Adobe also provides more and more modern navigating and enriching tools only supports their PDF approach.

Blue Sky Research has in their TeXtures 2.0 also picked up the hyperlink idea under the buzzword: synchronicity, meaning dynamic links between any place in the source window to the typeset window and vice versa, eventually controlled by the log window.

Granularity
Some 5 years ago Adobe in their Acrobat Exchange considered pages or screens as the right amount of granularity. The building blocks of PDF format. They could be refreshed, shuffled, linked and so on. More recently on the internet pages I found that frames, as part of nested pages/screens form the right granularity and can be of any size, so even one character.

New and completely beyond TeX are digital photos or videos, which by an amateur can be processed by Adobe's PhotoDeLuxe.

Sounds can be processed, saved and sent too. As yet I have no experience with it, nor with digital video.

TeX and HTML
HTML stands for HyperText Markup Language and came along with the internet, well its logical layer www. The homepages on internet are written in HTML and can be viewed by the ubiquitous and for free browsers like

Netscape and Internet Explorer.

New PCs or Macs come nowadays with tools like Pagemill to build Webpages, especially when subscribing to commercial providers. For transforming (All)TeX into HTML tools are available not in the least Gurari's TeX4ht, as reported by Erik Frambach in MAPS 20, spring 1998.

Erik communicated separately that HTML provide links automatically, that search enginges can look into HTML code, and that downloading can be done in parts. Only those frames one is interested in, instead of the bulky PDF file.

Daydreaming

Would it be feasible to think of MAPS as a multi-media PDF bulletin or as an HTML hypertext? Could we deliver PDF submissions instead of the stubborn and harnessing LaTeX & Co, or just supply our HTML hypertexts on our homepages and let the MAPS editors compose a symphony out of the hyperlinks in their editorial comments? No longer printing and dissemination hazzle, but it will be simply available via NTG's Web page to be downloaded on demand? Or even more liberal with only links from without on NTG's Webpage to the real material, whereever it may be?

In the small with respect to BLUe
I got with respect to BLUeTeX feedback from Russia to parameterize over the fonts. Maybe, I should allow to easily change the CMR fonts for their PS variants. In doing so, my PS 4/600 Laser printer with 2MB RAM would no longer complain. And maybe, BLUe will be better geared toward PDF use, because we can easily adjust for the fonts given the aim. Next to that some loose ends should be handled better, not in the least to launch it as a real format, and have a clear idea about what can be scaled and what not.

Some time ago I wrote my convertor assistant which transformed

$$.bluetex \xrightarrow{TeX} .latex$$

to alleviate the retyping of commands in submitting BLUe scripts to MAPS.

Maybe I should consider

$$.bluetex \xrightarrow{TeX} .html$$

Of course, I'm familiar with transformations and know that we usually are talking about different things, for and after the transformation. For example, the concept of nested frames is not in BLUeTeX. But for the endleaves, where we like to open up documents for direct access via the internet, this tranformation might be useful. On the other

hand a TEXie can process a .tex file easily, although generally several source files, or databases as in BLUeTEX's case, are necessary.

Acknowledgements

Conclusions

Looking back, I think I did a good investment in familiarizing myself with TEX and PostScript, and having had a look into SGML. For the future Adobe's Acrobat suite and their PDF is a must, nor can I afford it to leave HTML alone, especially when I'll come with my WWW homepage, some day.

I'm only too happy that I did not suboptimize in using TEX & Co, did know when to stop.

It seems that with respect to TEX & Co the law of diminishing returns applies.

Netscape and Internet explorer, Outlook Express, Alpha integrated with CMacTEX on my Mac are musts, and maybe I'll have time to really delve into MPW (Mac Programmar's Workbench), or Knuth's literate programming, which has been lingering on my desk for so long already. In between some nitty-gritties about plug-ins and convertors have to be worked through, and back-ups of my 'desks' secured.

BLUeTEX will be adapted, and the PWT (Publishing with TEX) guide will be renewed as an hypertext, in PDF and in HTML, hopefully. Maybe, all made available from my homepage to start with.

In short there is more to be done than before. . .

My case rests. Have fun, and all the best.