TeXLive Collection
past and future

Abstract
Past and future of the TeXLive Collection is described.

Keywords
TeXLive Collection

Introduction

It must have been in the second half of the eighties that I obtained a copy of the TEXbook. It contained what appeared to me as fascinating magic. Then our company purchased microTEX, the software program ready to run on a personal computer. It came with a DVI viewer and a printer driver for a matrix printer. From there we moved on to a big PCTEX, Y&Y's DVIPSONE, BLUESKY's outline fonts, now all history.

A few years later we learned of the Dutch speaking TEX User Group NTG and, because we had run into some limitations of TEX —too small a hash— we tried EMTEX, which later became part of 4TEX. 4TEX was one of the first TEX distributions on CDROM, an integrated set of the most popular programs available in the TEX world. We depended on the yearly updates of 4TEX and later TEXLIVE, of which version 8 was released in 2003, until today.

Beginning with version 8 TEXLIVE has become the TEX Collection. It combines an out–of–the–box TEX system and the complete CTAN repository (Comprehensive TEX Archive Network: a snapshot of almost all that is available for TEX users). TEX systems started on floppy disks but soon filled CDROM's and now DVD's. An archive of a couple of hundred files grew into tens of thousands.

tree	directories	files	bytes
texmf	3.750	45.000	626 M
texmf-extra	115	1.500	66 M
bin	16	2.500	250 M
source	380	6.900	104 M

If the CTAN archive is included we have a grand total of 138.000 (unzipped even 420.000) files, organized in 10.000 directories, totaling 5.906.870.829 byte, or 6 GB.

With version 8 the organizers realized that comprehensive began to become incomprehensible. Even though the TDS, the TEX Directory Structure, had brought some order in grouping files they were still faced with the fact that old TEX systems had been replaced with new systems in a continuous process to adapt to changing operating systems, improved text editors and more sophisticated and generally available viewers and printers. Fundamental changes appeared necessary and are implemented in the TEX Collection 2004. This paper will focus on some of the most important of these changes.

The engine

Donald Knuth's TEX was the ground breaking program that could typeset and be a programming language at the same time. TEX as a typesetting engine has been adapted to handle larger size memory, extended with features, translated into other programming languages, like C, and with the coming of PDF, the Portable Document Format, is now capable of producing PDF output directly with PDF-ε-TEX. The most important change in the 2004 release is that PDF-ε-TEX has become the main TEX engine. PDF-ε-TEX incorporates all 'accepted' extensions with proven reliability, produces DVI output by default, PDF when commanded, and ε-TEX is in there once explicitly enabled. To trigger PDF output Context users just add as the first line in their text files:

```
% output=pdftex
```

Context is a monolithic and coherent package of macro definitions that use the programming abilities of almost any TEX to accomplish a large variety of easy to use special typesetting functions.

Other macro packages have often been associated with a specific TEX binary. In practice this lead to several combinations of so called format files holding the macro definitions and binaries.

For plain TEX the system call (on the command line) and the engine are the same.

system call	format	engine
tex	plain.fmt	tex
etex	etex.efmt	etex
pdftex	pdftex.fmt	pdftex
pdfetex	pdfetex.efmt	pdfetex

For Latex the system call matches not the engine but

the format name. Here the command that starts TEX and loads a format is just a shortcut to calling the engine with a specific format.

system call	format	engine
latex	latex.fmt	tex
pdflatex	pdflatex.efmt	pdftex

For Context each format is named after the user interface language, the language of commands, messages, keywords, and so forth. This must not be confused with the language of the document text to be typeset. Each interface can handle all document languages.

system call	format	engine interface
cont-cz	cont-cz.efmt	pdfetexczech
cont-de	cont-de.efmt	pdfetexgerman
cont-en	cont-en.efmt	pdfetexenglish
cont-it	cont-it.efmt	pdfetexitalian
cont-nl	cont-nl.efmt	pdfetexdutch
cont-ro	cont-ro.efmt	pdfetexromanian

Normally, however, Context is launched by TEXEXEC, a PERL script that automates many annoying user tasks.

So, what is the importance of the change to PDF-ε-TEX in the 2004 Collection? Very little for the user, the system calls are unchanged! For TEXLIVE system maintenance, however, the change means that the various different TEX binaries can be removed and replaced by a single TEX engine that combines them all: PDF-ε-TEX. Extensions like ε-TEX, pdfTEX, MLTEX and ENCTEX are no longer needed as separate entities.

system call	format	engine
tex	plain.fmt	pdfetex
etex	etex.efmt	pdfetex
pdftex	pdftex.fmt	pdfetex
pdfetex	pdfetex.efmt	pdfetex
latex	latex.fmt	pdfetex
pdflatex	pdflatex.efmt	pdfetex

Because of the growing dependency on this engine PDF-ε-TEX has rigourous quality assurance and DANTE, NTG, and TUG have decided to financially support its primary author Hàn Thế Thành to extend and improve the program.

A change such as this is not trivial since it must be certain that existing documents can be processed without change, and macro packages must still believe that the correct binary is available. Macro packages may use undocumented features and nasty tricks to determine what engine is present. Currently PDFTEX is extended to take care of this problem. The configuration file has gone, more extensive map file handling has been implemented, and extensions are being separated to allow for experimental versions (XPDFETEX).

PDF-ε-TEX, although quite universally useful, still lacks some features such as Unicode awareness. TEX engine development, therefore, must continue. Those on the Context mailing list may know Giuseppe Bilotta as an enthousiastic user and advocate of TEX. In 2003 Giuseppe published EOMEGA, an extended version of TEX that uses Unicode natively. His initiative evolved into the ALEPH project which aims at merging ε-TEX with OMEGA. This is because some Context users wanted to use OMEGA features. Latex is also moving towards ε-TEX, enhancing the importance of the ALEPH initiative.

Those who have become dependent on OMEGA may get attracted by ALEPH's image: stable realware thus giving it a good chance to become the default engine under the OMEGA based formats on TEXLIVE. Producing PDF output directly is not a feature but the DVIPDFMX converter can produce the same rich PDF output as PDF-ε-TEX does for Context users.

Latin Modern

What more is new on the TEXLIVE 2004? First of all, the Latin Modern fonts. This project was funded by user groups. The fonts are extended versions of Computer Modern, with additional characters covering all western languages. Latin Modern will replace the textual part of Computer Modern Roman. The fonts are already on TEXLIVE 2003, so you can play with them. For instance, `cmr10`, `aer10`, `plr10`, `csr10` as well as in the near future `vnr10` will be replaced by `lmr10`. This change is downwards compatible. It removes a lot of nearly duplicate files from TEXLIVE. If all works out well, users will not notice the font change. Of course, the original `cmr10` will still be present.

Currently extra instances are made with a few more glyphs, more kerning pairs. Visual improvements are made based on suggestions by Donald Knuth in his errata documents.

Font files

A more drastic change is that some files change place in the TDS tree. Until now the encoding (`enc`) and the fontmap (`map`) files were located under the DVIPS and PDFTEX paths:

```
texmf/dvips
texmf/dvips/config
texmf/dvips/config/whatever
texmf/pdftex
texmf/pdfte
texmf/pdftex/config/whatever
```

The configuration file `texmf.cnf` informs applications about where to find these encoding and fontmap files. A changed `texmf.cnf` assures that most applications and users will not encounter problems. The new locations are:

```
texmf/fonts/enc/whatever
texmf/fonts/map/whatever
texmf/fonts/lig/whatever
```

Note the new ligature path. It is used by for instance AFMTOPL. Some changes are already reflected in the current TEXLIVE version but probably go unnoticed because both old and new locations are supported.

If you install your own fonts you need to relocate your map files. Font metrics remain in their usual place and encoding files are seldomly made by users. Instead of relocating another option is to adapt the `texmf.cnf` file, but this would complicate future updating. It is better to not touch this file.

Scripts

Context includes some PERL scripts taking care of sorting the index, managing multiple runs and other chores. Initially, the number of scripts was small and they ended up in a dedicated Context directory.

Since then other macro packages also come with PERL scripts and Context added RUBY scripts leading to these paths:

```
texmf/context/perltk
texmf/context/ruby
```

TEXLIVE uses stubs in the binary path to launch such scripts. The stubs use KPSEWHICH to locate the main script file. For reasons of consistency, maintainance and robust locating, scripts now have their own root path, for Context:

```
texmf/scripts/context/perl
texmf/scripts/context/ruby
```

Companion files that do not fit in this directory structure remain where they are located presently. In practice users will not notice the changes because the stubs take care of things. Future versions of KPSEWHICH will provide more robust and convenient ways to locate such script files.

Beware: if you write your own scripts you should realize that call to KPSEWHICH have to be adapted, for instance:

```
kpsewhich -progname=context
  -format="other text files" texexec.pl
```

is now:

```
kpsewhich -progname=context
  -format="scripts" texexec.pl
```

More

AFM files will no longer be distributed in their compressed form (gzip). Engine dependent TEX source files end up in specific paths. Most common users will not notice because users of engine dependent sources have their own way of structuring the directory tree.

The KPSE file searching library and tools get a few more features. Next year's TEXLIVE will have a completely rewritten version of this library, one that opens some windows to the future such as automatic updating, remote processing, and fetching resources from zip archives.

Production

Getting TEXLIVE ready requires an enormous effort. Only a few macro collections are submitted in the right structure. Consequently, much scripting takes place to get the files where they belong in the tree. Interdependencies are not always made clear and maintainers of packages come and go. When the structure changes files need to be relocated. Bugs in binaries need to be solved. New features have to be tested first. Documentation needs to be updated. Frequently new CDROM images are constructed and tested, on all platforms. Thus the TEXLIVE mailing list is a busy one. Last year we even had a show–stopper. At press time it was discovered that 8 bit file output no longer worked.

Finally, the Collection has to be produced. The 2003 Collection was the first to be distributed on DVD. Even after TEXLIVE and CTAN were put on the DVDplenty of space was available, so extra's were added (in the `texmf-extra` path) and the next release will provide even more. The DVD is one of the first dual layer data DVD's. This meant producing special split ISO–images and proofing of the first DVD: the presses were actually stopped after the first copy for testing!

In 2003 and 2004 DANTE invited those involved in this monster performance to their main anual meeting, altogether some 15 contributors from all over the world. They discussed the present and the future of such distributions. I leave the reporting of that discussion to the chairman. Happy users of TEXLIVE, however, should recognize with gratitude that getting this job done is far from trivial and effortless. We all should treasure those who are making TEXLIVE happen year after year. You can find their names on the cover of the DVD and in the documentation.

Summary

When the next TEXLIVE shows up in your postbox, update and things will work as usual. If you have your own fonts installed, however, you need to relocate your personal mapfiles to `.../fonts/map`, and run `mktexlsr` to update your files database. Also, if your scripts use KPSEWHICH, check them.

Hans Hagen
pragma@wxs.nl