

Overview of the T_EX historic archive

Ulrik Vieth

Vaihinger Straße 69

70567 Stuttgart

Germany

ulrik dot vieth (at) arcor dot de

Abstract

This paper presents an overview of the T_EX historic archive, an archive of historic T_EX distributions and packages hosted on the TUG FTP server. While the topic of this conference focuses on paths to the future, this archive aims to preserve snapshots of significant past and present releases of T_EX software, so they will remain available for reference and won't get lost when new versions are installed. Apart from explaining the motivation behind the project, this paper also presents an overview of what the archive currently contains and what is still missing.

1 Introduction

The history of T_EX goes back nearly 30 years now, starting from the earliest beginnings in 1977. In the case of T_EX, unlike most other software projects, the early history is pretty well documented in the literature [1, 2, 3], at least as far as the core system of the T_EX engine itself is concerned.

When it comes to related elements of the T_EX system, such as macro packages or fonts or systems like METAFONT or METAPOST, the documentation of the history isn't nearly as sophisticated as in the case of T_EX, and pretty soon one ends up piecing together the history from anecdotal evidence.

Probably the best way of preserving the history of a software project is by preserving the artifacts of the system, preferably the original sources and distribution packages of all significant releases.

While this may seem to be a well-established best practice in the age of project web sites featuring source code repositories and distribution archives, one has to keep in mind that such a level of project infrastructure is quite a recent achievement, far from standard just a few years ago.

2 Genesis of the T_EX historic archive

Looking back just one decade, establishing CTAN as a comprehensive archive network of all available T_EX software was a great achievement of the combined T_EX users groups, which has become a resource that is taken for granted now.

From the point of view of software archaeology, however, such an archive doesn't necessarily satisfy all the needs, since the different types of archives have different goals and different audiences.

While CTAN is a dynamic and ever-changing archive, which aims to carry only the latest versions of everything available, a historic archive is a static archive, which aims to preserve archival copies of various release versions of selected packages which may be of specific historic interest.

While CTAN aims to satisfy the needs of end users, making it easy to find the current version of a package without causing confusion, the audience of a historic archive is likely to consist of just a small group of expert users, who might be interested in browsing and comparing various different versions of a package with specific attention to small details such as original file dates and timestamps.

Based on these considerations, the idea of an archive of historic T_EX distributions and packages was conceived in the late 1990s in discussions between the author and some CTAN maintainers.

It was agreed that the historic archive should be kept separate from CTAN and it was decided to establish the archive on the TUG FTP server, which also served as a development server for some projects and already carried some interesting material.

3 Organization of the T_EX historic archive

As explained, the T_EX historic archive is intended as an archive of historic T_EX distributions, packages and files, which is hosted on the TUG FTP server at the following URL:

`ftp://ftp.tug.org/historic/`

The archive is maintained entirely manually, mostly by the author himself with a few contributions from others. It is updated infrequently on an occasional basis whenever new material becomes available that may be of significant historic interest.

The organization of the archive loosely follows the organization of CTAN, featuring top-level directory such as `fonts`, `macros`, `systems`, etc.

Below the top level, however, the structure is somewhat different. In some cases, it was sufficient to have a single directory for a package, containing multiple copies of distribution files that can be distinguished by version number. In other cases, it was more convenient to have individual subdirectories for each release of a distribution or package.

4 Contents of the \TeX historic archive

When the \TeX historic archive was first established, it was mostly based on the author's personal archive of material collected during the 1990s while serving as the local \TeX administrator on a Unix network at university and on Linux PCs at home.

Later, some additional material was added that was restored from old CTAN CDs or mirrored from various FTP servers located by search engines that were still carrying old material at that time.

In some cases the most recent versions of some packages were mirrored directly from CTAN or from distribution archives of project web sites.

The selection of material is obviously somewhat biased by personal preferences and interests, so the coverage of Unix-based distributions is quite extensive while the coverage of other platforms is largely missing. Similarly, the coverage of \LaTeX releases is very comprehensive, while $\text{Con}\TeX\text{t}$ is still missing. Fortunately, there exists an independent archive site of $\text{Con}\TeX\text{t}$ releases to fill the gap in this case.

In the following sections, we will look at some examples of distributions and packages preserved in the collection of the \TeX historic archive.

systems/knuth This directory contains the true original distribution files from Don Knuth, mirrored from Stanford, covering the 1995, 1998, and 2002 releases of \TeX and METAFONT.

In principle, the material is identical to what is found in `systems/knuth` on CTAN, except that those files are somewhat less reliable, because they are unpacked and rearranged without much regard to preserving timestamps or the original structure.

systems/unix This directory contains various releases of Unix \TeX distributions by Pierre MacKay. The coverage is complete from \TeX 3.0 to \TeX 3.141, but earlier releases are very rare.

The author was lucky to find a complete copy of \TeX 2.95 to mirror from an obscure FTP server during the late 1990s, but he was unable to locate a complete copy of \TeX 2.0 at that time.

Since the Unix \TeX distribution used to be quite popular at times, it is quite likely that some copies of distribution tapes of older Unix \TeX releases may have survived in some basements or attics. However, finding a suitable tape drive and getting old tapes to read after 15 or 20 years could be quite difficult. On the other hand, unpacking the files once they are restored to disk should be no problem at all, since the TAR format has remained popular and is still widely used on Unix/Linux platforms today.

systems/web2c This directory contains various releases of the `web2c` distribution by Karl Berry and later Olaf Weber. The coverage is mostly complete from `web2c-5.0` of 1990 to `web2c-7.5.4` of 2005 and also includes many beta-test releases.

In recent years, maintenance of `web2c` was integrated with \TeX Live and has been moved to a Subversion repository on `svn.tug.org`. While such a source code repository provides even more fine-grained access to changes between releases, the lack of traditional distribution files unfortunately causes a gap in the file-based historic archive.

Moreover, unlike traditional text-based change log files, we are becoming more reliant on the technology of the Subversion repository to preserve the change history. We can only hope that the change history will be properly migrated and not blindly discarded in case the repository is ever moved to another system or platform.

systems/teTeX This directory contains various releases of the `teTeX` distribution by Thomas Esser. The coverage is mostly complete from `teTeX-0.2` of 1994 to `teTeX-3.0` of 2005 with the exception of beta-test releases, such as `teTeX-0.99`.

It is interesting to note how the distribution has grown from the original Linux-only release of `teTeX-0.2` which used to fit on 5 floppy disks to the later multi-platform Unix distributions featuring an ever-growing `texmf` tree of 30, 50 or 90 MB.

systems/metapost This directory features a collection of METAPOST releases by John Hobby and later Taco Hoekwater. The coverage is complete from version 0.62 of 1995 to the latest releases and also includes many beta-test releases.

All of the recent beta-test and production releases were mirrored directly from the project web sites at `sarovar.org` and `foundry.supelec.fr`, so in this case the flow of file-based distribution snapshots for archival purposes has fortunately remained intact after the development infrastructure was moved to a different platform.

systems/pdftex This directory features a collection of pdfTeX releases by Hàn Th^ế Thành and later Martin Schröder. Unfortunately, the coverage of the archive is currently only partly complete. It is fully complete only for the recent releases from versions 1.20 to 1.40, that could be mirrored directly from the project web site at sarovar.org, but most of the earlier versions are still missing.

The author has tried to fill in some of the gaps by adding versions of pdfTeX that were bundled with web2c releases, but the earliest versions don't go back any further than pdftex-0.11.

If anybody could provide additional copies of early versions of pdfTeX, such a contribution to the historic archive would be very welcome.

macros/latex209 This directory contains various releases of L^AT_EX 2.09 of 1991–92, including the final public release that can still be found on CTAN in the *obsolete* tree. It also contains several noteworthy add-on packages for L^AT_EX 2.09 such as NFSS and the so-called Mainz packages that later became part of the L^AT_EX 2_ε kernel or *tools* bundle.

It would be interesting to add original copies of L^AT_EX 2.09 of 1986 or even L^AT_EX 2.00 of 1984 for comparison, if such versions could be recovered.

macros/latex2e This directory features a collection of L^AT_EX 2_ε releases, starting from the earliest pre-test releases of 1993 until the current release. The coverage is very complete for most of the 1990s, but the author has failed to preserve snapshots of some of the more recent releases, which however could be recovered from CTAN CDs.

In addition to the official L^AT_EX 2_ε releases, the collection also includes some alpha-test releases from the author's personal collection that were originally circulated only to a limited group of testers.

Of course, the historic archive cannot possibly cover all the many contributed add-on packages and classes for L^AT_EX, so the coverage concentrates on the L^AT_EX base system and the required packages, but even those aren't fully complete for all releases. The archive usually includes the *graphics* and *tools* bundles, but sometimes lacks releases of *amslatex*, *babel*, or *psnfss*.

macros/context As mentioned before, the selection of material currently doesn't include much in the area of macro packages besides L^AT_EX.

In the case of ConT_EXt, there fortunately exists an independent project web site of ConT_EXt releases since 1997 at foundry.supelec.fr, which could be easily copied or mirrored, if desired.

fonts/cm This directory contains various releases of Computer Modern fonts of 1995, 1998, and 2002, extracted from the corresponding original sources in *systems/knuth*.

In principle, these files are redundant, but they are provided here for the convenience of font developers and researchers, who sometimes need to compare details of various releases of font sources.

It would be interesting to add additional earlier versions of CM fonts for comparison, possibly going back to the first releases of 1985 or even earlier to AM fonts, if such versions could be recovered.

fonts/dc-ec This directory contains various releases of DC and EC fonts by Norbert Schwarz and later Jörg Knappen between 1991 and 1997.

In addition, the collection also includes some test releases from the author's personal collection that were never widely distributed.

fonts/modes This directory features a collection of various releases of *modes.mf* by Pierre MacKay and Karl Berry since 1991, providing a repository of METAFONT mode definitions for various devices.

fonts/fontname This directory contains various releases of the *fontname* document by Karl Berry since 1992, which not only lists the assignment of font names for several suppliers, but also documents several important font encodings.

fonts/utilities/fontinst This directory holds a collection of *fontinst* releases by Alan Jeffrey, Sebastian Rahtz and later Lars Hellström.

The coverage is relatively comprehensive since *fontinst-1.800* of 1998, but only a few copies of earlier releases since 1994 could be recovered from various sources such as CTAN CDs.

Some versions of *fontinst* between 1995 and 1998 never existed as official releases on CTAN and were only available as unofficial releases as part of other font tools, so there remains a gap in the archive that can only be partially closed.

fonts/psfonts/tools This archive complements the collection of *fontinst* releases and contains a collection of font tools that were used between 1995 and 1998 to generate font metrics for *psnfss*, using a combination of Perl scripts and unofficial *fontinst* releases by Sebastian Rahtz.

In recent years, the use of these tools has been discontinued in favor of standard *fontinst-1.9xx* after maintenance of *psnfss* and *psnfss-sources* has been taken over by Walter Schmidt.

fonts/lm Among the most recent additions to the T_EX historic archive are the Latin Modern fonts by the GUST e-foundry.

Unfortunately, the coverage is still fairly incomplete at present. While the development of LM fonts has started in 2002, the author has started taking archival snapshots only recently, when the development was already well advanced, so he missed most of the early development versions.

In order to provide a comprehensive reference for font researchers, it would be interesting to add copies of earlier releases for comparison, if those still exist somewhere. Once again, such a contribution to the archive would be very welcome.

fonts/tex-gyre Finally, the T_EX Gyre fonts by the GUST e-foundry have also been added to the T_EX historic archive, as soon as the first releases have become available.

As a lesson learned from Latin Modern fonts, where taking archival snapshots was started too late, the author has concluded that archiving releases of important projects should better be started as early as possible before anything can get lost when the next release is installed over the previous one.

5 Summary and conclusions

As was explained in the introduction, the historic archive was established for a very specific purpose: provide a collection of archival copies of important T_EX distributions, packages or fonts.

The archive started out in the late 1990s with a small personal collection and has steadily grown over the years, adding new material mirrored from CTAN or project web sites as well as old material restored from earlier CTAN CDs or mirrored from obscure FTP servers still carrying old material.

While the archive covers a lot of ground in many areas, there are still a number of gaps that remain to be filled, not only for older material, but also for some of the more recent development projects.

As was already emphasized, contributions to the archive are very welcome, especially in the areas where gaps remain to be filled, such as in the cases of pdfT_EX and the Latin Modern fonts.

Besides the areas already covered in the archive, there are certainly other areas that have been somewhat neglected so far, but equally well deserve to be archived. Some examples include additional macro

packages besides L^AT_EX, such as ConT_EXt, and additional distributions for other platforms besides Unix/Linux such as emT_EX or fpT_EX. Once again, contributions to the archive are welcome, if you have anything to offer in these areas.

6 Additional resources of historic material

Finally, concerning the early history of T_EX and METAFONT before 1990 there remains the resource of the Stanford SAILDART archive located at

<http://www.saildart.org/>

The SAILDART site hosts an archive of the SAIL computer system at Stanford, restored from backup tapes, which happens to be the very machine used by Don Knuth from the 1970s until 1990, on which T_EX and METAFONT were first developed.

At present the SAILDART archive only allows access to the system areas [TEX,SYS] and [MF,SYS], whereas the private areas [TEX,DEK] and [MF,DEK] remain restricted.

Nevertheless, even without access to the private areas, this archive site provides a wealth of historic material from between 1980 and 1990 that remains largely untapped.

Among the material that may be found here are numerous releases of T_EX and METAFONT sources, numerous versions of CM and AM fonts as well as several early L^AT_EX releases. In addition, one might even find some more exotic stuff here, such as macros for TEX78 or font sources for MF79.

So far, only a few examples have been copied from SAILDART to the T_EX historic archive, such as a snapshot of L^AT_EX 2.0 for T_EX 1.0, but there is certainly a lot more that could be added.

We can conclude that even after 30 years since 1977, the history of T_EX remains an interesting topic of research, considering all those new developments currently going on, as well as all the old material that is left to be rediscovered.

References

- [1] Donald E. Knuth. The Errors of T_EX. *Software — Practice and Experience*, 19:607–681, 1989.
- [2] Donald E. Knuth. *Literate Programming*. CSLI Publications, Stanford, CA, USA, 1992.
- [3] Donald E. Knuth. *Digital Typography*. CSLI Publications, Stanford, CA, USA, 1999.