

TEXUTIL
explained

Introduction

While typesetting a document, tables of contents, references and index entries are often to be included. The items of a table of contents is collected during the typesetting process, and therefore an additional typesetting pass is needed to get the right one at the front of the document. In a similar way, index entries are collected and need to be sorted before inclusion. And, forward references can only be resolved when known.

When processing a text, CONTeXt saves table and index entries, cross references, and some more information in a so called utility file. After a successful run, this file is converted into a format suitable for the next run. This conversion is done by `TeXUTIL`, a PERL script that comes with the CONTeXt distribution.

As a bonus, `TeXUTIL` can prepare documentation files, analyze graphics and filter log files. These modes of operation will be discussed later. I will not bother you with technical details on the file formats used. These can be found in the PERL file itself.

To fill you in on the backgrounds of the `TeXUTIL` logo as shown on this cover: many lines of data produced by CONTeXt are read in, interpreted, converted, sorted and or combined, and written back to file in a similar or adapted form: `TeXUTIL` is dealing with transformations.

A next version will integrate (simple) conversion from SGML to `TeX` as well as sorting of mapped encodings.

Running

references When you use `TeXEXEC`, running `TeXUTIL` is taken care of automatically. Otherwise, you need to call `TeXUTIL` manually:

```
texutil --references somefile
```

In this case, `TeXUTIL` loads the file `somefile.tui` sorts and converts the content, and writes a file called `somefile.tuo`. Additional switches may be passed; these will be discussed later.

documents The CONTeXt related source files often contain documentation. The documentation lines can be recognized by `%D` for text lines, `%M` for macro code needed for the documentation, and `%S B` and `%S E` for parts that do not make it into the documentation. In PERL scripts, the `%` is replaced by a `#`. So, when you take a look at source files, you now know what these symbols stand for.

```
texutil --documents somefile
```

This command converts a \TeX file with suffix `tex` into a file with suffix `ted`. A related switch `--sources` does the opposite: it removes the documentation lines and generates a source only file with suffix `tes`. More information can be found in `texutil.pl`.

figures Although PDFT\TeX can include graphics data, traditional \TeX depends on the DVI post-processor. Whatever method is used, \TeX has to reserve space and therefore has to know the dimensions of graphics. Determining the dimensions of for instance EPS graphics can be done by \TeX itself, but bitmap formats need to be analyzed by some other program, and that is where \TeXUTIL comes into view.

```
texutil --figures
```

When told to, \TeXUTIL starts searching for graphic files and collects some information on them in a file called `texutil.tuf`. When a graphic is included, and CONTEXT cannot sort out the dimensions itself, it will search for data in this file. When needed and enabled, CONTEXT will call \TeXUTIL to provide the information.

```
texutil --figures *.png *.tif
```

As demonstrated here, you can limit the inventory. This is useful when you want CONTEXT to typeset a directory of figures. This kind of \TeXEXEC features are not that special; they just take care of calling the right CONTEXT macro.

logfile \TeXUTIL is capable of concerns parsing the log file as produced by CONTEXT . The command:

```
texutil --logfile somefile
```

will report information on unknown references and overfull boxes: both being important issues to resolve when typesetting a file.

purge Although CONTEXT uses only one utility file, many scratch files can scatter your disk after a run: buffers, METAPOST graphics as well as files that communicate between CONTEXT and METAPOST , log files, etc.

```
texutil --purge
```

Afterwards, \TeXUTIL reports the number of bytes regained. Optionally, you can pass a file pattern, thereby limiting the cleansing.

Switches

Before we will explain the mode dependant switched, we will introduce some general ones.

`help` This switch results in some help information on the switches to use.

`silent` This switch suppresses all messages to the screen. It enables the use of `TeXUTIL` in ways like:

```
texutil --figures --silent --verbose somefile.png
```

This reports something like:

```
n=somefile.png t=png x=0bp y=0bp w=443bp h=591bp
```

This information can be piped into another program.

`verbose` Sometimes there is more to report than useful at first sight. This switch will therefore lead to more screen output.

`interface=languagecode` If you want to let `TeXUTIL` talk back in your own language, use this switch. Currently `en`, `de` and `nl` are supported.

Installation

Installation is not complicated. When you have PERL running, `TeXUTIL` should run too. On MS WINDOWS you can copy `runperl.exe` into `texutil.exe` to directly run `TeXUTIL`. This program is part of the FPT_EX distribution.

References

The slightly misleading switch `--references` concerns more than handling cross references. Actually these, as well as list entries and two pass data, are simply passed on to the output file. Index entries and sorted list entries however are sorted and redundant index entries are removed.

`ij` Sort `IJ` as `Y`. This is a typical dutch oriented switch.

`high` This switch converts high standard MSDOS ASCII values into `TeX` commands. This switch is provided for backward compatibility reasons.

quotes Normally quotes are used as boundary characters for strings. This switch forces `TEXUTIL` to handle quotes differently.

Documents

Documentation is part of the source files. The `--documents` option lets `TEXUTIL` enter a filter mode. Because typesetting of documentation is supported by `TEXEXEC`, users seldom have to deal with this option directly.

type=suffix Normally `TEXUTIL` will try to sort out the right filetype, but when more instances of a file (with different suffixes) are present, it makes sense to tell `TEXUTIL` explicitly which filetype is to be converted.

outputfile=filename This file can be used to force `TEXUTIL` in using a different name for the resulting output.

sources By default, documentation files are generated, but when this switch is passed, source code only files are produced.

setups There used to be a time when syntax definitions were part of the source files. Currently these are collected in one file, but maybe some day they will make it to the source files again. This option collects all definitions in a file `texutil.tus`.

infos The first few years of `CONTEXT` all macro definitions were collected into a few files. These files also contained explanations of the commands, that could be used by help systems, as implemented in `TEXEDIT`. While documenting the sources, this information is removed, but the feature is still there. The help information is recognized by `%I` directives and is collected in files with the suffix `tud`.

templates This option is related to the previous one and concerns editor templates. Again we're dealing with features that might become obsolete.

Figures

As mentioned, `TEXUTIL` can analyze graphics (PNG, JPG, TIF, EPS, PDF and MPS). The next few switches deal with a different topic: conversion from EPS to PDF.

epspage In preparation for conversion to PDF, EPS files should be prepared in such a way that the bounding box is similar to the paper dimensions. This option prepares the files and also removes some redundant (rubbish) information. To prevent repetitive adaption the files are marked as being adapted.

epstopdf When issued, this option calls GHOSTSCRIPT to convert the EPS graphic into a PDF one.

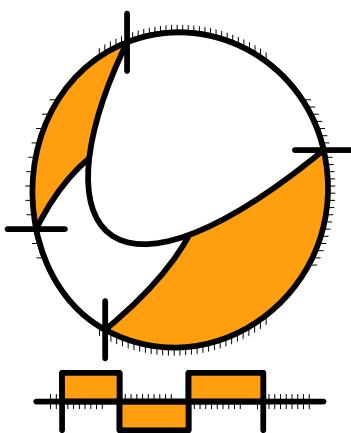
LogFile

The faster machines get, the less obvious are messages. Filtering the log files therefore makes sense. The **--logfile** option serves this purpose.

box Issuing this switch lets **T_EXUTIL** report overfull boxes, both horizontal and vertical. Alternatively you can use **--vbox** and **--hbox**.

criterium=number When reporting overfull boxes, small deviations, in the order of magnitude of a few points, are seldom disturbing. By setting a criterium **T_EXUTIL** will only report those overfull boxes that exceed the criterium. The number represents points.

unknown A missing reference or an unknown figure is not what we expect to be reflected in the final document. This option makes sure we don't miss them



PRAGMA

Advanced Document Engineering | Ridderstraat 27 | 8061GH Hasselt NL
tel: +31 (0)38 477 53 69 | email: pragma@wxs.nl | internet: www.pragma-ade.com