

NAME

ltx2unitxt – convert LaTeX source fragment to plain (Unicode) text or simple html

SYNOPSIS

ltx2unitxt [*-c CONFIG*] [*-o OUTPUT*] [*--html*] [...] [*INFILE*]...

DESCRIPTION

Convert the LaTeX source in INFILE (or standard input) to plain text using Unicode code points for accents and other special characters; or, optionally, output HTML with simple translations for font changes and url commands.

Common accent sequences, special characters, and simple markup commands are translated, but there is no attempt at completeness. Math, tables, figures, sectioning, etc., are not handled in any way, and mostly left in their TeX form in the output. The translations assume standard LaTeX meanings for characters and control sequences; macros in the input are not considered.

The input can be a fragment of text, not a full document, as the purpose of this script was to handle bibliography entries and abstracts (for the ltx2crossrefxml script that is part of the crossrefware package). Patches to extend this script are welcome. It uses the LaTeX::ToUnicode Perl library for the conversion; see its documentation for details.

Conversion is currently done line by line, so TeX constructs that cross multiple lines are not handled properly. If it turns out to be useful, conversion could be done by paragraph instead.

The config file is read as a Perl source file. It can define a function ‘LaTeX_ToUnicode_convert_hook()’ which will be called early; the value it returns (which must be a string) will then be subject to the standard conversion.

For an example of using this script and associated code, see the TUGboat processing at <https://github.com/TeXUsersGroup/tugboat/tree/trunk/capsules/crossref>.

OPTIONS

- c, --config=FILE**
read (Perl) config FILE for a hook, as explained above
- e, --entities**
output entities &#xNNNN; instead of literal characters
- g, --german**
handle some features of the german package
- h, --html**
output simplistic HTML instead of plain text
- o, --output=FILE**
output to FILE instead of stdout
- v, --verbose**
be verbose
- V, --version**
output version information and exit
- , --help**
display this help and exit

Options can be abbreviated unambiguously, and start with either **-** or **--**.

Dev sources, bug tracker: <https://github.com/borisveytsman/bibtexperllibs>

Releases: <https://ctan.org/pkg/bibtexperllibs>

ltx2unitxt (bibtexperllibs) 0.51 Copyright 2023 Karl Berry. This is free software: you can redistribute it and/or modify it under the same terms as Perl itself.