

NAME

`bibtexu` – UTF-8 Big BibTeX

SYNOPSIS

`bibtexu` [*options*] *aux-file*

DESCRIPTION

BibTeXu is the Unicode-compliant version of BibTeX. It is largely based on Niel Kempson's BibTeX8, and it provides a better support for UTF-8 by integrating ICU library. Therefore, **BibTeXu** no longer requires the Codepage and Sort order ("CS") file; instead, the method of sorting and case-changing can be controlled via command-line options.

OPTIONS

- ? --help**
display some brief help text.
- d --debug TYPE**
report debugging information. TYPE is one or more of all, csf, io, mem, misc, search.
- s --statistics**
report internal statistics.
- t --trace**
report execution tracing.
- v --version**
report BibTeX version.
- l --language LANG**
use language LANG to convert strings to low case. This argument is passed to ICU library.
- o --location LANG**
use language LANG for sorting. This argument is passed to ICU library.
- B --big**
set large BibTeX capacity.
- H --huge**
set huge BibTeX capacity.
- W --wolfgang**
set really huge BibTeX capacity for Wolfgang.
- M --min_crossrefs ##**
set min_crossrefs to ##.
- mstrings ##**
allow ## unique strings.

UNICODE SUPPORT

BibTeXu supports extended features to handle Unicode characters. Several built-in functions in bibliography styles are enhanced as follows.

- &** Pops the top two (integer) literals and pushes their bitwise AND.
- |** Pops the top two (integer) literals and pushes their bitwise OR.

add.period\$

Pops the top (string) literal, adds a ‘.’ to it if the last non‘}’ character isn’t a ‘.’, ‘?’ or ‘!’ or a Unicode punctuation mark and pushes this resulting string. The mark may be U+203C, U+203D, U+2047, U+2048, U+2049, U+3002, U+FF01, U+FF0E or U+FF1F.

chr.to.int\$

Pops the top (string) literal, makes sure it’s a multibyte string of a single Unicode code point, converts it to the corresponding Unicode scalar value (integer), and pushes this integer.

int.to.chr\$

Pops the top (integer) literal, interpreted as the Unicode scalar value of a single code point, converts it to the corresponding single character multibyte string, and pushes this string.

num.names\$, format.name\$

The function is the same as original BibTeX but an Ideographic/Fullwidth Comma (U+3001, U+FF0C) in addition to an " and " string is accepted as a separator between persons and Ideographic Space (U+3000) in addition to a space " " is accepted as a separator between a family name and a given name.

substring\$, text.length\$, text.prefix\$

The function is the same as original BibTeX but the unit of operand numbers is Unicode code point.

change.case\$

The function is the same as original BibTeX but letters of non-english Latin, Greek and Cyrillic are supported.

width\$ The function is the same as original BibTeX but letters of Latin-1 and Latin Extended-A and CJK characters are supported.

is.cjk.str\$

Pops the top (string) literal, set flag bits to an integer if CJK characters are found in the string, and pushes the resulting integer, otherwise pushes 0. Flags 0x001, 0x002, 0x004, 0x008 and 0x800 are corresponding to Hanzi (Kanji, Hanja), Kana, Hangul, Bopomofo and other CJK characters, respectively. For example, an integer 0x003 will be pushed if Hanzi and Kana characters are found in a popped string literal.

is.kanji.str\$

Same as **is.cjk.str\$** for compatibility with (u)pBibTeX.

SEE ALSO

More detailed description of **BibTeXu** is available at \$TEXMFDIST/doc/bibtexu/README.

AUTHORS

BibTeXu was written by Yannis Haralambous and his students. It is maintained as part of TeX Live.

This manpage was written for TeX Live.