

Joseph's Adventures in Unicodeland

20th July 2015



Hardcopy versions of the Unicode Standard have been among the most crucial and most-heavily used reference books in my personal library for years. Unicode allows me to celebrate the fact that computer science is a vast worldwide collaboration. And Unicode is perhaps the best tool I know to help bring understanding between people of different cultures.

Donald E. Knuth

Exploring Unicodeland

- ✦ *The Unicode Standard*
- ✦ UnicodeData.txt
- ✦ EastAsianWidth.txt
- ✦ LineBreak.txt
- ✦ SpecialCasing.txt
- ✦ CaseFolding.txt
- ✦ ...

Setting up characters

- ✦ Collect all of the relevant data
- ✦ Text: `\catcode`, `\lccode`, `\uccode`
- ✦ Maths: `\Umathcode`
- ✦ XeTeX and LuaTeX in sync but ...
- ✦ ... include up `\XeTeXintercharclass` data
- ✦ Clear relationship between UCD and resulting set up

Unicode case terms

- Lowercasing** Convert all code points to their lowercase mapping if defined, otherwise leave alone; apply context and language rules
- Uppercasing** Convert all code points to their uppercase mapping if defined, otherwise leave alone; apply context and language rules
- Titlecasing** Convert the first letter to (almost always) uppercase and the rest to lowercase; some code points have a different upper- and titlecase mapping; complicated by language conventions
- Case folding** Remove case information for *non-text* uses: similar to lower casing but not identical; no context/language dependence

Case changing: expl3 approach

- ✦ Implement all Unicode case-related changes
- ✦ Context-sensitive mappings
- ✦ Language-dependent mappings
- ✦ (If possible) Avoid `\lccode` and `\uccode`
- ✦ (If possible) Implement expandably
- ✦ (If possible) Handle math mode
- ✦ (If possible) Escape mechanism

Discussion

- ✦ LaTeX-L List
- ✦ TeX-StackExchange chat
- ✦ joseph.wright@morningstar2.co.uk
- ✦ <https://github.com/latex3/>
- ✦ ...

