Case changing:
LaTeX reaches Unicodeland

23rd July 2022

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

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 mappings

Case changing: expl3 approach

- Implement all Unicode mappings
- Use \lccode and \uccode for core data ...

Case changing: expl3 approach

- Implement all Unicode mappings
- Use \lccode and \uccode for core data ...
- ... but support one-to-many changes

Case changing: expl3 approach

- Implement all Unicode mappings
- Use \lccode and \uccode for core data ...
- ... but support one-to-many changes
- Context-sensitive mappings, for example lowercasing sigma

Case changing: expl3 approach

- Implement all Unicode mappings
- Use \lccode and \uccode for core data ...
- ... but support one-to-many changes
- Context-sensitive mappings, for example lowercasing sigma
- Language-dependent mappings, *e.g.* Turkish dotless-i

Case changing: expl3 approach

- Implement all Unicode mappings
- Use \lccode and \uccode for core data ...
- ... but support one-to-many changes
- Context-sensitive mappings, for example lowercasing sigma
- Language-dependent mappings, *e.g.* Turkish dotless-i
- Implement expandably

Lessons from textcase

- Auto-escape math mode
- Handle \cite, \label, \ref, ...
- \NoCaseChange

Time to update . . .

- expl3 code now supports 8-bit input
- Wider use of \protected, particularly for actives
- Full Unicode range without needing extra support
- Language-dependent support a possible:
  probably for Fall '22

Something old, something new

Retained
- `\MakeUppercase`
- `\MakeLowercase`
- `\NoCaseChange`

Something old, something new

Retained
- \MakeUppercase
- \MakeLowercase
- \NoCaseChange

Added
- \MakeTitlecase
- \CaseSwitch
- \AddToNoCaseChangeList
- \DeclareCaseChangeEquivalent