Further adventures in Unicodeland

16th July 2023
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
Previously . . .

- Implement case mappings for Unicode engines
Previously . . .

- Implement case mappings for Unicode engines
- . . . then partial support for 8-bit engines
Previously . . .

- Implement case mappings for Unicode engines
- . . . then partial support for 8-bit engines
- Context-sensitive mappings, for example lowercasing sigma
- Language-dependent mappings, e.g. Turkish dotless-i
Previously . . .

- Implement case mappings for Unicode engines
- . . . then partial support for 8-bit engines
- Context-sensitive mappings, for example lowercasing sigma
- Language-dependent mappings, e.g. Turkish dotless-i
- Auto-escape math mode
- Handle $\texttt{cite}$, $\texttt{label}$, $\texttt{ref}$, . . .
- $\texttt{NoCaseChange}$
Further improvements

- Automatic locale switching: \texttt{\textbackslash BCPdata}
Further improvements

- Automatic locale switching: \texttt{BCPdata}
- Full Unicode support for pdfTeX
  (using two-stage tables)
Further improvements

- Automatic locale switching: \texttt{BCPdata}
- Full Unicode support for pdfTeX
  (using two-stage tables)
- Tailoring of codepoints: \texttt{DeclareUppercaseMapping}, etc.
Further improvements

- Automatic locale switching: \texttt{\textbackslash BCPdata}
- Full Unicode support for pdfTeX
  (using two-stage tables)
- Tailoring of codepoints: \texttt{\textbackslash DeclareUppercaseMapping}, etc.
- Handle full Greek requirements for pdfTeX
Further improvements

- Automatic locale switching: \BCPdata
- Full Unicode support for pdfTeX
  (using two-stage tables)
- Tailoring of codepoints: \DeclareUppercaseMapping, etc.
- Handle full Greek requirements for pdfTeX
- ... refining details not covered properly in CLDR
Further improvements

- Automatic locale switching: \BCPdata
- Full Unicode support for pdfTeX
  (using two-stage tables)
- Tailoring of codepoints: \DeclareUppercaseMapping, etc.
- Handle full Greek requirements for pdfTeX
- ... refining details not covered properly in CLDR
- ... LICR input being looked at by Günter Milde
Further improvements

– Automatic locale switching: \BCPdata
– Full Unicode support for pdfTeX (using two-stage tables)
– Tailoring of codepoints: \DeclareUppercaseMapping, etc.
– Handle full Greek requirements for pdfTeX
– … refining details not covered properly in CLDR
– … LICR input being looked at by Günter Milde
– Low-level Unicode general category support (\codepoint_to_category:n)
Ongoing

- True titlecasing (and sentence casing):
  may need some more complex word division
Ongoing

- True titlecasing (and sentence casing):
  may need some more complex word division
- Better detection of non-letters
Ongoing

- True titlecasing (and sentence casing):
  may need some more complex word division
- Better detection of non-letters
- Character token generation using general category
Graphemes

- Think Spinal Tap:
  nine letters, one space
Graphemes

- Think Spinal Tap:
  nine letters, one space
- A grapheme is a human-perceived character
Graphemes

- Think Spinal Tap:
  nine letters, one space
- A *grapheme* is a human-perceived character
- Fully implemented for mapping
Graphemes

- Think Spinal Tap:
  nine letters, one space
- A grapheme is a human-perceived character
- Fully implemented for mapping
- Likely next is `\text_item:nn`
  (or `\text_char:nn` or `\text_grapheme:nn` or ...)
Thanks

- Javier Bezos
- Jürgen Spitzmüller
- Peter Kankowski
- Günter Milde