Suggested updates for the Soft_Dotted property

Kent Karlsson
2003-06-05

Introduction

The list of "Soft_Dotted" characters in Unicode 4.0 is the following:

```
0069..006A   ; Soft_Dotted # L&   [2] LATIN SMALL LETTER I..LATIN SMALL LETTER J
012F         ; Soft_Dotted # L&       LATIN SMALL LETTER I WITH OGONEK
0268         ; Soft_Dotted # L&       LATIN SMALL LETTER I WITH STROKE
0456         ; Soft_Dotted # L&       CYRILLIC SMALL LETTER BYELORUSSIAN-UKRAINIAN I
0458         ; Soft_Dotted # L&       CYRILLIC SMALL LETTER JE
1D62         ; Soft_Dotted # L&       LATIN SUBSCRIPT SMALL LETTER I
1E2D         ; Soft_Dotted # L&       LATIN SMALL LETTER I WITH TILDE BELOW
1ECB         ; Soft_Dotted # L&       LATIN SMALL LETTER I WITH DOT BELOW
2071         ; Soft_Dotted # L&       SUPERSCRIPT LATIN SMALL LETTER I
```

There are however, some 'i'-like characters that do not have the Soft_Dotted property in Unicode 4.0. The characters in the following sections are therefore suggested to get the Soft_Dotted property. Recall that the Soft_Dotted property means that the dot intrinsic in glyphs for the character is lost when a combining above character is applied to it. To make these characters hard-dotted, a combining dot above must be applied. This is used for some Lithuanian letters.

Dutch small ij ligature (definite)

```
0133; Soft_Dotted # LATIN SMALL LIGATURE IJ
```

This is a compatibility letter, both in the sense that it has a compatibility mapping and is taken from a "legacy" character encoding. It is, however, not necessarily a character that should not be used. Even though in most cases it is sufficient to use ordinary i and j in sequence to write the Dutch ij, in some cases it may still be best to use the ij ligature character, for best spacing, titlecasing, and vertical layout. The ij can in some cases be acute accented, in which case both of the dots should be replaced by acute accents. The representation of this is quite straightforward if ordinary i and j are used. If, however, the ij ligature is used, the ij ligature must first of all be soft-dotted. Applying just ordinary combining acute accent to produce an accent on each part of the ligature. It may be more logical to apply combining double acute accent. A (small) typographic problem is to align the two acute accents over the constituent letter bodies. It is not certain that a grave accent may be applicable to the ij, but if it is, the story is similar. At least one Dutch dictionary (Ter Laan, Niew Groninger Wordenboek, 1929) uses a macron over the ij. If coded as separate i and j, one would use 035E;COMBINING DOUBLE MACRON in-between. If the ij ligature is used, a combining macron should be applied after the ligature character to get a macron that goes over both the dotless constituent letter bodies.

In each of these cases, the ij ligature must be soft-dotted.

Phonetic letters (uncertain)

```
029D; Soft_Dotted # LATIN SMALL LETTER J WITH CROSSED-TAIL
02B2; Soft_Dotted # MODIFIER LETTER SMALL J
```

Phonetic letters often get combining characters applied to them. Whether these characters should loose their intrinsic dot when a combining character above is applied to them, I don't know for certain, but it seems plausible.

Greek small letter (uncertain)

```
03F3; Soft_Dotted # GREEK LETTER YOT
```

This is a Greek character that looks just like a j. Whether this character should loose its intrinsic dot when a combining character above is applied to it, I don't know for certain, but it seems plausible.

"Mathematical" small letters (definite)

Among the "mathematical" letters on plane 01 (and plane 00), there are no dotless i not any dotless j (which would correspond to the glyphs called via the command \i and \j respectively in TeX source documents. It is thus apparent that the ordinary dotted i-s and j-s in plane 01 is to be used also when diacritics above are applied (as they often are in math expressions). These base characters must then loose their intrinsic dots. Therefore, the following letters must have the Soft_Dotted property:

```
2148; Soft_Dotted # DOUBLE-STRUCK ITALIC SMALL I
2149; Soft_Dotted # DOUBLE-STRUCK ITALIC SMALL J
1D422; Soft_Dotted # MATHEMATICAL BOLD SMALL I
1D423; Soft_Dotted # MATHEMATICAL BOLD SMALL J
1D456; Soft_Dotted # MATHEMATICAL ITALIC SMALL I
1D457; Soft_Dotted # MATHEMATICAL ITALIC SMALL J
1D48A; Soft_Dotted # MATHEMATICAL BOLD ITALIC SMALL I
1D48B; Soft_Dotted # MATHEMATICAL BOLD ITALIC SMALL J
1D48A; Soft_Dotted # MATHEMATICAL BOLD ITALIC SMALL I
1D48B; Soft_Dotted # MATHEMATICAL BOLD ITALIC SMALL J
1D4BE; Soft_Dotted # MATHEMATICAL SCRIPT SMALL I
1D4BF; Soft_Dotted # MATHEMATICAL SCRIPT SMALL J
1D4F2; Soft_Dotted # MATHEMATICAL BOLD SCRIPT SMALL I
1D4F3; Soft_Dotted # MATHEMATICAL BOLD SCRIPT SMALL J
1D526; Soft_Dotted # MATHEMATICAL FRAKTUR SMALL I
1D527; Soft_Dotted # MATHEMATICAL FRAKTUR SMALL J
1D55A; Soft_Dotted # MATHEMATICAL DOUBLE-STRUCK SMALL I
1D55B; Soft_Dotted # MATHEMATICAL DOUBLE-STRUCK SMALL J
1D58E; Soft_Dotted # MATHEMATICAL BOLD FRAKTUR SMALL I
1D58F; Soft_Dotted # MATHEMATICAL BOLD FRAKTUR SMALL J
1D5C2; Soft_Dotted # MATHEMATICAL SANS-SERIF SMALL I
1D5C3; Soft_Dotted # MATHEMATICAL SANS-SERIF SMALL J
1D5F6; Soft_Dotted # MATHEMATICAL SANS-SERIF BOLD SMALL I
1D5F7; Soft_Dotted # MATHEMATICAL SANS-SERIF BOLD SMALL J
1D62A; Soft_Dotted # MATHEMATICAL SANS-SERIF ITALIC SMALL I
1D62B; Soft_Dotted # MATHEMATICAL SANS-SERIF ITALIC SMALL J
1D65E; Soft_Dotted # MATHEMATICAL SANS-SERIF BOLD ITALIC SMALL I
1D65F; Soft_Dotted # MATHEMATICAL SANS-SERIF BOLD ITALIC SMALL J
1D692; Soft_Dotted # MATHEMATICAL MONOSPACE SMALL I
1D693; Soft_Dotted # MATHEMATICAL MONOSPACE SMALL J
```

Note that U+0131, LATIN SMALL LETTER DOTLESS I, is not a mathematical character, and should not be used as a mathematical variable, except when using a word of Turkic origin, with a dotless i, as a variable name (usually not single-letter). Similarly, a dotless j (used in landsmålsalfabetet, a phonetic alphabet not yet covered by Unicode, should not be used in a mathematical expression. They do *not* correspond to \i and \j in TeX.