

Proposal to update “SpecialCasing”

Kent Karlsson
August 15, 2001

The current suggestion for updating SpecialCasing for version 3.1.1 is (like all older versions) inconsistent in its handling of the same original string when that string is turned into NFC versus NFD form (or unnormalised form for that matter). In particular, the 3.1.1d2 version (and all older ones) are inconsistent for the Turkish I-dot (highlighted below). To make this consistently handled requires change to the so-called normal case. L2/01-191R goes further and aims for total casing consistency between NFC and NFD, as well as avoiding i+dot (and j+dot) whenever case-mapping and the dot is not needed.

SpecialCasing-3.1.1d2.txt gives:

normal:

	tolower		toupper	
	NFC	NFD	NFC	NFD
i	i	i	I	I
I	i	i	I	I
i+dot	i+dot	i+dot	I	I
I-dot	i	i+dot	I-dot	I+dot
I-circ	i-circ	i+circ	I-circ	I+circ
i+dot+circ	i+dot+circ	i+dot+circ	I+dot+circ	I+dot+circ

Lithuanian:

	tolower		toupper	
	NFC	NFD	NFC	NFD
i	i	i	I	I
I	i	i	I	I
i+dot	i+dot	i+dot	I	I
I-dot	i	i+dot	I-dot	I+dot
I-circ	i+dot+circ	i+dot+circ	I-circ	I+circ
i+dot+circ	i+dot+circ	i+dot+circ	I+circ	I+circ

Turkish/Azeri:

	tolower		toupper	
	NFC	NFD	NFC	NFD
i	i	i	I-dot	I-dot
I	dotless-i	dotless-i	I	I
i+dot	i	i	I-dot	I+dot
I-dot	i	i+dot	I-dot	I+dot
I-circ	i-circ	i+circ	I-circ	I+circ
i+dot+circ	i+dot+circ	i+dot+circ	I+dot+circ	I+dot+circ

L2/01-191R gives:

normal:

	tolower		toupper	
	NFC	NFD	NFC	NFD
i	i	i	I	I
I	i	i	I	I
i+dot	i	i	I	I
I-dot	i	i	I-dot	I+dot
I-circ	i-circ	i+circ	I-circ	I+circ
i+dot+circ	i+dot+circ	i+dot+circ	I+dot+circ	I+dot+circ

Lithuanian:

	tolower		toupper	
	NFC	NFD	NFC	NFD
i	i	i	I	I
l	i	i	I	I
i+dot	i	i	I	I
l-dot	i	i	I-dot	I+dot
l-circ	i+dot+circ	i+dot+circ	I-circ	I+circ
i+dot+circ	i+dot+circ	i+dot+circ	I+circ	I+circ

Turkish/Azeri:

	tolower		toupper	
	NFC	NFD	NFC	NFD
i	i	i	I-dot	I-dot
l	dotless-i	dotless-i	I	I
i+dot	i	i	I-dot	I+dot
l-dot	i	i	I-dot	I+dot
l-circ	i-circ	i+circ	I-circ	I+circ
i+dot+circ	i+dot+circ	i+dot+circ	I+dot+circ	I+dot+circ