DR 18-0002 — SML: Sorting CJK ideographic characters based on Japanese phonetics

Status: Further Consideration Required

Subject: SML: Sorting CJK ideographic characters based on Japanese phonetics

Qualifier: Request for clarification

Submitter: MURATA Makoto Organization: JISC

Contact Information: eb2m-mrt@asahi-net.or.jp

Submitter’s Defect Number: None

Supporting Document(s): none

Date Circulated by Secretariat: 2018-02-05

Deadline for Response from Editor: 2018-04-05

IS 29500 Reference(s): §18.3.1.92, “sortState (Sort State)”

Related DR(s): DR 17-0009

Nature of the Defect:

CJK ideographic characters in Japanese are converted from Hiragana characters, which are typed by users. Excel does not discard initially-typed Hiragana characters; it keeps them using rPh elements (see §18.4.6, Part 1) and uses them for sorting cells containing the converted CJK ideographic characters. This behavior should be documented.



Solution Proposed by the Submitter:

None

Schema Change(s) Needed:

No

**Editor’s Response:**

**2018-06-05/08 London Meeting:**

As of WD1, Part 1, §18.18.73, “ST\_SortMethod (Sort Method)” contains, “The details of the sort options are implementation-defined.”

We agreed to add a note with examples that suggests how an implementation might take advantage of this. For example, in Japanese Excel, a kanji string might be entered exactly as that, it might be entered as kana, which is then converted to kanji. And identical kanji strings might be the result of converting different kana strings. Separately, the same kana strings can result in different kanji strings. In these cases, the sort might be done on the initially types kana rather than the kanji.

Murata-san provided XML examples for various scenarios, which will be added to the note.

Rex will own this issue.

**2018-06-24 Rex Jaeschke:**

I propose the following note:

**Part 1: 18.18.73 ST\_SortMethod (Sort Method), p. 2486.**

Sort method. Chinese Simplified, Chinese Traditional, and Japanese support alternate sort methods (multiple sort options are available). All other languages support only 1 sort option. In that case, the value pinYin is used. The details of the sort options are implementation-defined.

[Note: When writing in Japanese, a string might be entered as kanji, or it might be entered as kana that is then converted to kanji. Also, under certain circumstances, the conversion of different kana strings might result in identical kanji strings, and identical kana strings might be converted to different kanji strings. When kana-to-kanji conversion is applied, the sort might be done on the initial kana rather than on the resulting kanji. end note]

This simple type's contents are a restriction of the W3C XML Schema string datatype.

**…**

Changes to Part 1: N Part 2: N Part 3: N Part 4: N