

Title: Proposal to add an RSIndex.txt data file to each release of the Unicode Standard

Author: Ken Lunde

Date: 2023-10-02

Proposal

This document proposes that a “plain text” representation of the Radical-Stroke Index file (PDF) be distributed with each subsequent release of the Unicode Standard, which may be useful for some developers. Its proposed location is the same as the PDF file, such as the following for Unicode Version 16.0:

<https://www.unicode.org/Public/16.0.0/charts/RSIndex.txt>

The proposed text and placement of its link on the landing page for a release is shown below in context and in bold:

- Interactive Han Radical-Stroke Index
- IICore Radical-Stroke Index (3.8 MB)
- Unihan Core 2020 Radical-Stroke Index (8.2 MB)
- Full Han Radical-Stroke Index (44 MB)
- **Full Han Radical-Stroke Index Collation Data**

The data file, with proposed header text and with content that corresponds to Unicode Version 15.1, is attached to this document as a PDF attachment. The data file specifies two tab-delimited fields: 1) a unique radical-stroke value pair whose values are separated by a period; and 2) one or more Unicode Scalar Values for Han ideographs in proper collation order.

Background

The Radical-Stroke Index file, *RSIndex.pdf*, was first introduced in Unicode Version 5.2 (2009), and was updated in Versions 6.0, 6.1, 8.0, 10.0, 11.0, 13.0, 14.0, 15.0, and 15.1. The app that is used to produce the PDF file, *RSIndex*, implements the sorting algorithm that is specified in [Section 2.1.2, Sorting Algorithm Used by the Radical-Stroke Indexes](#), of UAX #38. This app’s console output is used to update the *rs_public* table of the *unihan_public* MySQL database that runs on the Unicode website, which is referenced by the [Unihan Radical-Stroke Index](#) page.

I developed a Python script, *makeRSPublic.py*, which converts the raw *RSIndex* app console output into a ready-to-use MySQL data file, and includes a command-line option to alternatively output a “plain text” representation of the same data as proposed in this document.

IICoreRSIndex.txt & UnihanCore2020RSIndex.txt?

Of course, comparable data files could easily be produced based on the *IICoreRSIndex.pdf* and *UnihanCore2020RSIndex.pdf* files, but I recommend that we begin by deploying the data file for the full Radical-Stroke Index. If there is sufficient demand for these data files, we can add them at any time, because these PDF files are not tied to a specific release of the Unicode Standard.

That is all.