Splitting Up Unihan.txt

John H. Jenkins
28 January 2008

As we continue to add data to the Unihan database, the text version of the released data is becoming a little unwieldy. It is also causing a noticeable drain on Unicode's Web server as the entire document is downloaded with significant frequency.

The UTC may wish to consider splitting the document into pieces for future versions of the standard. Probably the easiest way to fracture it would be along the lines of general categories as specified in UAX 38. We would thus have Unihan-IRGSources.txt, Unihan-OtherMappings.txt, and so on.