

Title: Proposal to add "kHDZRadBreak" to Unihan.txt  
 Contributor: Richard Cook  
 Date: 20040422:15:38:00  
 Action: For UTC consideration

This proposal seeks to add new field data to Unihan.txt. This new field, "kHDZRadBreak", provides information on the 200 部首 Radical section breaks occurring in the major lexical print source 《漢語大字典》 <Hanyu Da Zidian> (HDZ == "Unihan.txt:kHanYu"). It identifies the first character classified under each of the 200 HDZ Radicals using the kHanYu mapping, and additionally maps this character to one of the 214 康熙 KangXi Radicals (U+2F00..U+2FD5). Note that the set of 200 HDZ Radicals is a subset of the traditional set of 214 KangXi Radicals. This subset is differently ordered than the KangXi (KX) set of 214, and omits 14 "useless" radicals, conflating these 14 with other radicals.

This new data is important because it permits determination of the HDZ radical assignment for all 56,097 HDZ characters on the basis of existing kHanYu mappings. Note that this radical assignment information is useful because it may in some cases be different from kRSUnicode and kRSKangXi assignments. It can also be used for determining kHanYu radical/stroke information (which Unihan.txt at present lacks). Such data would allow for systematic improvements in the printed Unicode Radical Stroke chart.

The formatting and content of the proposed new field is exemplified by the following excerpted lines (KX data is in UTF-8, followed redundantly with bracketed USV):

# USV	Field Name	KX[USV]:kHanYu Mapping
U+4E00	kHDZRadBreak	一[U+2F00]:10001.010
U+4E28	kHDZRadBreak	丨[U+2F01]:10028.020
U+4E3F	kHDZRadBreak	丿[U+2F03]:10031.020
U+4E36	kHDZRadBreak	丶[U+2F02]:10042.080
U+4E59	kHDZRadBreak	乙[U+2F04]:10047.040
...	(192 lines omitted)	...
U+9F8D	kHDZRadBreak	龍[U+2FD3]:74803.010
U+9FA0	kHDZRadBreak	龠[U+2FD5]:74807.010
U+9F9C	kHDZRadBreak	龜[U+2FD4]:74809.010

The full "kHDZRadBreak" data set (200 lines total) is available here:

<<http://www.unicode.org/~rscook/text/kHDZRadBreak.u8>>

Similar data is in preparation for the other major lexical mapping sources, based on the specific editions used in the Unihan.txt mappings.