Universal Multiple-Octet Coded Character Set
International Organization for Standardization

Doc Type: Working Group Document
Title: Information in support of N4234 (L2/12-087) to demonstrate extensive use of PUA in common IDS data
Author: Dr. Ken Lunde (小林劍), Adobe Systems Incorporated
Status: Individual Contribution
Action: For consideration by JTC1/SC2/WG2
Date: 2012-02-14

A typical IDS (*Ideographic Description Sequence*), as used by the IRG, looks like "⿰僉刂" (<U+2FF0,U+50C9,U+5202>), which represents 劍 (U+528D). The table below shows that among the first 256 characters in the URO, meaning U+4E00 through U+4EFF, the IDSes for 32 (1-in-8) include one or more private use characters as components.*

UCS		IDSes (private use characters, when used, are enclosed in angled brackets and in U+ notation)		
U+4E03	七	⿻**<U+F696>**乚		
U+4E07	万	⿱一**<U+F506>**		
U+4E0C	丌	⿱一**<U+F56D>**		
U+4E0D	不	⿱一**<U+F1EE>**		
U+4E0E	与	⿹**<U+F537>**一 *[GTKV]*	⿻**<U+F537>**一 *[J]*	
U+4E1A	业	⿱**<U+F1BA>**一		
U+4E1D	丝	⿱⿰**<U+F3BC><U+F3BC>**一		
U+4E21	両	⿱一**<U+F150>**		
U+4E24	两	⿱一**<U+F137>**		
U+4E27	丧	⿱**<U+F462><U+F5DB>**		
U+4E2E	丮	⿹乁**<U+F569>**		
U+4E33	丳	⿻吕**<U+F56D>**		
U+4E34	临	⿰〢⿱**<U+F3FA><U+F556>**		
U+4E46	乆	⿰**<U+F400>**乀		
U+4E4C	乌	⿹**<U+F41C>**一		
U+4E54	乔	⿱夭**<U+F56D>**		
U+4E55	乕	⿸厂**<U+F42B>**		
U+4E80	亀	⿱⺈**<U+F3FB>**		
U+4E89	争	⿱⺈**<U+F576>**		
U+4E99	亙	⿳一**<U+F3B9>**一		
U+4E9E	亞	⿳一**<U+F342>**一		
U+4EA3	亣	⿱丶丌	⿱亠**<U+F56D>**	
U+4EA5	亥	⿳亠**<U+F400>**人		
U+4EA6	亦	⿱亠**<U+F5CA>**		
U+4EA7	产	⿱**<U+F526>**厂		
U+4EAD	亭	⿱**<U+F584>**丁		
U+4EAE	亮	⿱**<U+F584>**几 *[G]*	⿱**<U+F584>**儿 *[TJK]*	
U+4EB3	亳	⿱**<U+F584>**乇		
U+4EB4	亴	⿳**<U+F584>**土九 *[GK]*	⿳**<U+F584><U+F569>**九 *[T]*	
U+4ECB	介	⿱𠆢**<U+F56D>**		
U+4EE4	令	⿱亽マ *[G]*	⿱亼マ *[TV]*	⿱亼**<U+F45D>** *[JK]*
U+4EFA	仺	⿱𠆢**<U+F421>**		

* https://github.com/kawabata/kanji-database-ids/blob/master/ids.txt

The list below indicates how many characters in each CJK Unified Ideograph block use private use characters in their IDSes, based on the current version of the IDS database:

URO: 480
Extension A: 107
Extension B: 1,972
Extension C: 55
Extension D: 14

That is all.