

Universal Multiple-Octet Coded Character Set International Organization for Standardization

Doc Type: Working Group Document

Title: Proposal to correct 91 “kIRG_GSource” sources, remove 94 “kPseudoGB1” sources, and add “kGB12052” sources

Source: Ken Lunde, Adobe Systems Incorporated

Status: Corporate Full Member Contribution

Action: For consideration by the UTC

Date: 2010-09-30

Background

When GB/T 12345-90 was submitted to the CJK-JRG (the original name of the IRG) for compiling the URO (*Unified Repertoire & Ordering*), it included ideographs above and beyond the standard proper. 94 of these “additional” ideographs more properly map to another GB standard, specifically to GB 12052-89 (entitled 信息交换用朝鲜文字编码字符集; *Korean Character Coded Character Set for Information Interchange* in English). Unfortunately, the Unihan Database does not reflect these genuine sources. This document proposes a way to modify the Unihan Database to correctly reference these 94 ideographs’ proper GB sources. There is no practical benefit from continuing to neglect the proper sources, which only serves to confuse those who have access to the official GB/T 12345-90 standard, and not the version submitted for the URO.

Proposal Details

There are three parts to this proposal, described as follows:

- 91 of these 94 ideographs have “kIRG_GSource” sources that should be changed from GB/T 12345-90 to GB 12052-89 ones. I suggest the use of a “K-” prefix to distinguish GB 12052-89 sources. The remaining three ideographs have sources in other (and actual) GB standards, and should thus not be changed.
- 94 of the 153 “kPseudoGB1” sources that correspond to these 94 ideographs should simply be removed.
- A new “kGB12052” field should be added that specifies the GB 12052-89 sources for these 94 ideographs.

The table on page 2 of this document provides all of the necessary information for the 94 ideographs in question. The first column provides the current “kIRG_GSource” sources. 91 of these 94 sources (91 of 94) should simply be changed to reflect the correct GB 12052-89 sources, which is the hexadecimal value in the third column. The remaining three sources (3 of 94), which have been shaded in the table, should remain as-is. The second column are the 94 “kPseudoGB1” sources that should be removed. The third column are the sources for the 94 instances of the proposed “kGB12052” field, specifically the Row-Cell values in parentheses. The fourth column are the Unicode scalar values.

As evidence, I submit the relevant page from the GB 12052-89 standard that shows the 94 ideographs in Rows 71 and 72, along with the relevant page from the unofficial version of the GB/T 12345-90 standard that was submitted to the CJK-JRG that shows the same 94 ideographs, and in the same order, in Row 93. These are provided on pp 3 and 4 of this document.

The following table provides examples of all three changes being applied to the relevant Unihan Database fields:

	Original	Corrected
91 of 94	U+4E6B kIRG_GSource 1-7D21 U+4E6B kPseudoGB1 9301	U+4E6B kIRG_GSource K-6779 U+4E6B kGB12052 7189
3 of 94	U+9425 kIRG_GSource 5-762B U+9425 kGB5 8611 U+9425 kPseudoGB1 9345	U+9425 kIRG_GSource 5-762B U+9425 kGB5 8611 U+9425 kGB12052 7239

kIRG_GSource	kPseudoGB1	GB 12052-89	UCS	kIRG_GSource	kPseudoGB1	GB 12052-89	UCS
1-7D21	9301	6779 (7189)	U+4E6B	1-7D50	9348	684A (7242)	U+7A24
1-7D22	9302	677A (7190)	U+824D	1-7D51	9349	684B (7243)	U+5AA4
1-7D23	9303	677B (7191)	U+5DEA	1-7D52	9350	684C (7244)	U+680D
1-7D24	9304	677C (7192)	U+4E6C	1-7D53	9351	684D (7245)	U+6A74
1-7D25	9305	677D (7193)	U+8FF2	1-7D54	9352	684E (7246)	U+4E72
1-7D26	9306	677E (7194)	U+551F	1-7D55	9353	684F (7247)	U+4E7D
1-7D27	9307	6821 (7201)	U+848A	1-7D56	9354	6850 (7248)	U+6B0C
1-7D28	9308	6822 (7202)	U+5EE4	1-7D57	9355	6851 (7249)	U+785B
1-7D29	9309	6823 (7203)	U+53BC	1-7D58	9356	6852 (7250)	U+7873
1-7D2A	9310	6824 (7204)	U+603E	1-7D59	9357	6853 (7251)	U+7320
1-7D2B	9311	6825 (7205)	U+6AF7	1-7D5A	9358	6854 (7252)	U+5CBE
1-7D2C	9312	6826 (7206)	U+7A52	1-7D5B	9359	6855 (7253)	U+4E7C
1-7D2D	9313	6827 (7207)	U+83BB	1-7D5C	9360	6856 (7254)	U+9007
1-7D2E	9314	6828 (7208)	U+4FA4	1-7D5D	9361	6857 (7255)	U+55ED
1-7D2F	9315	6829 (7209)	U+7553	1-7D5E	9362	6858 (7256)	U+698B
1-7D30	9316	682A (7210)	U+7364	1-7D5F	9363	6859 (7257)	U+6A7B
1-7D31	9317	682B (7211)	U+729C	1-7D60	9364	685A (7258)	U+5DFC
1-7D32	9318	682C (7212)	U+4E6D	1-7D61	9365	685B (7259)	U+4E76
1-7D33	9319	682D (7213)	U+6729	1-7D62	9366	685C (7260)	U+55B8
1-7D34	9320	682E (7214)	U+6730	1-7D63	9367	685D (7261)	U+95CF
1-7D35	9321	682F (7215)	U+5788	1-7D64	9368	685E (7262)	U+54DB
1-7D36	9322	6830 (7216)	U+8968	1-7D65	9369	685F (7263)	U+4E37
1-7D37	9323	6831 (7217)	U+4EBD	1-7D66	9370	6860 (7264)	U+70BA
1-7D38	9324	6832 (7218)	U+7F56	1-7D67	9371	6861 (7265)	U+516F
1-7D39	9325	6833 (7219)	U+536A	1-7D68	9372	6862 (7266)	U+7233
1-7D3A	9326	6834 (7220)	U+56D5	1-7D69	9373	6863 (7267)	U+5655
1-7D3B	9327	6835 (7221)	U+4E87	1-7D6A	9374	6864 (7268)	U+4E64
1-7D3C	9328	6836 (7222)	U+65C0	1-7D6B	9375	6865 (7269)	U+4E4A
1-7D3D	9329	6837 (7223)	U+4E06	1-7D6C	9376	6866 (7270)	U+4E65
1-7D3E	9330	6838 (7224)	U+6927	1-7D6D	9377	6867 (7271)	U+593B
1-7D3F	9331	6839 (7225)	U+4E6E	1-7D6E	9378	6868 (7272)	U+56CD
1-7D40	9332	683A (7226)	U+78D7	1-7D6F	9379	6869 (7273)	U+9064
1-7D41	9333	683B (7227)	U+6D4C	1-7D70	9380	686A (7274)	U+4E5B
1-7D42	9334	683C (7228)	U+693A	3-314B	9381	686B (7275)	U+503B
1-7D43	9335	683D (7229)	U+5DEC	1-7D72	9382	686C (7276)	U+7F98
1-7D44	9336	683E (7230)	U+5DED	1-7D73	9383	686D (7277)	U+4ED2
1-7D45	9337	683F (7231)	U+517A	1-7D74	9384	686E (7278)	U+4E7B
1-7D46	9338	6840 (7232)	U+698C	1-7D75	9385	686F (7279)	U+6B15
1-7D47	9339	6841 (7233)	U+58ED	1-7D76	9386	6870 (7280)	U+65D5
1-7D48	9340	6842 (7234)	U+5381	1-7D77	9387	6871 (7281)	U+7B7D
1-7D49	9341	6843 (7235)	U+6A75	5-4A3A	9388	6872 (7282)	U+6637
1-7D4A	9342	6844 (7236)	U+8644	1-7D79	9389	6873 (7283)	U+97B0
1-7D4B	9343	6845 (7237)	U+95AA	1-7D7A	9390	6874 (7284)	U+4E6F
1-7D4C	9344	6846 (7238)	U+7E07	1-7D7B	9391	6875 (7285)	U+4E41
5-762B	9345	6847 (7239)	U+9425	1-7D7C	9392	6876 (7286)	U+5391
1-7D4E	9346	6848 (7240)	U+87A6	1-7D7D	9393	6877 (7287)	U+4EAA
1-7D4F	9347	6849 (7241)	U+4E7A	1-7D7E	9394	6878 (7288)	U+591E

33区	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19
01/19	21	船	昌	夏	达	蕊	莲	鹿	余	识	榎	喬	蕊	傍	查	徵	徵	夏	木
3	李	岱	澍	今	罗	嘴	𠂇	旅	丁	榆	空	浅	蕊	桺	梁	梁	𠂇	𢈵	韦
10/59	管	撤	躉	闔	闔	鑄	𧆸	稜	𠂇	柱	櫟	𡗻	𠂇	𠂇	𠂇	𠂇	𠂇	達	邁
60/19	迦	真	楫	楫	昆	𠂇	𠂇	𠂇	𠂇	𠂇	𠂇	𠂇	𠂇	𠂇	𠂇	𠂇	𠂇	詔	邏
80/94	一	𠂇	𦥑	𦥑	𦥑	𦥑	𦥑	闔	𠂇	𠂇	𠂇	𠂇	𠂇	𠂇	𠂇	𠂇	𠂇	𠂇	邏
	1010	1011	1012	1013	1014	1015	1016	1017	1018	1019	1020	1021	1022	1023	1024	1025	1026	1027	1028

34区	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19
01/19	1E21	1E22	1E23	1E24	1E25	1E26	1E27	1E28	1E29	1E2A	1E2B	1E2C	1E2D	1E2E	1E2F	1E30	1E31	1E32	1E33
20/39	1E34	1E35	1E36	1E37	1E38	1E39	1E3A	1E3B	1E3C	1E3D	1E3E	1E3F	1E40	1E41	1E42	1E43	1E44	1E45	1E47
40/59	1E48	1E49	1E4A	1E4B	1E4C	1E4D	1E4E	1E4F	1E50	1E51	1E52	1E53	1E54	1E55	1E56	1E57	1E58	1E59	1E5A
60/19	1E5C	1E5D	1E5E	1E5F	1E60	1E61	1E62	1E63	1E64	1E65	1E66	1E67	1E68	1E69	1E6A	1E6B	1E6C	1E6D	1E6F
80/94	1E70	1E71	1E72	1E73	1E74	1E75	1E76	1E77	1E78	1E79	1E7A	1E7B	1E7C	1E7D	1E7E				

G3 17-43

= G5 42-26

Remove

木