ISO/IEC JTC1/SC2/**WG2 N4246R2**
ISO/IEC JTC1/SC2/WG2/**IRG N1844R**
L2/12-095R

Universal Multiple-Octet Coded Character Set
International Organization for Standardization

Doc Type: Working Group Document
Title: Stabilizing CJK Compatibility Ideographs through the use of Standardized Variants
Author: Dr. Ken Lunde, Adobe Systems Incorporated
Source: The Unicode Consortium
Status: Individual Contribution
Action: For consideration by JTC1/SC2/WG2
Date: 2012-03-05 (revised; originally submitted on 2012-02-15)

Background

CJK Compatibility Ideographs, for which there are now 1,002 characters as of Unicode Version 6.1, are subject to normalization, and are thus considered unstable because the distinctions that they are intended to convey cannot be preserved, regardless of which of the four normalization forms is applied. When normalized, a CJK Compatibility Ideograph is reverted into its canonical equivalent, which is always a CJK Unified Ideograph. For example, the CJK Compatibility Ideographs U+FA5D (﩮) and U+FA5E (﩯), when normalized, are reverted to their shared canonical equivalent, specifically U+8279 (舛). Furthermore, given the broad extent to which text services interact in today's applications and OSes, it is not possible to guarantee that normalization will not be applied, except for completely closed environments.

In other words, a wide variety of products, protocols, and environments normalize text data on a regular basis, and this cannot be changed, so a solution for preserving the distinctions that are intended to be conveyed by CJK Compatibility Ideographs becomes necessary.

Recommendation

In order to preserve any distinctions that the 1,002 CJK Compatibility Ideographs were intended to convey, and to enable round-trip capability, it is recommended that 1,002 Standardized Variants be accepted into the standard, which would be equivalent to the CJK Compatibility Ideographs themselves, and which would be immune to the effects of normalization. Note that VS1 (U+FE00) through VS3 (U+FE02) are the VSes (*Variation Selectors*) that are used, and not the VSes in Plane 14, which are used by the IVD (*Ideographic Variation Database*).

It is also recommended that any further glyphic distictions for these 1,002 CJK Compatibility Ideographs be handled via the IVD.

This document includes an attachment that is a complete list of the proposed Standardized Variants for all 1,002 CJK Compatiblity Ideographs.

Migration Considerations For Existing Implementations

Implementations that wish to migrate from using CJK Compatibility Ideographs to these proposed 1,002 Standardized Variants need to consider both short- and long-term implications. For the short term, pre-processing filters could be used to convert CJK Compatibility Ideographs to their corresponding Standardized Variants, and for the long term, this conversion could be performed during transcoding. In order to display the appropriate glyph, affected fonts need to be updated to include the appropriate variation sequences in their Format 14 'cmap' subtable.

IME developers also need to be aware of these Standardized Variants if they intend to emit the desired results.

These migration considerations are not exhaustive, but rather they are intended to pinpoint known areas that require implementor attention.

Table of Supporting Data

The table that spans the following two pages provides details for 89 of the 1,002 CJK Compatibility Ideographs, specifically those that correspond to registered IVSes (*Ideographic Variation Sequences*).

The first two columns simply list these 89 CJK Compatibility Ideographs and their corresponding canonical equivalents. The "Adobe-Japan1" and "Hanyo-Denshi" columns indicate the registered IVSes and their sequence identifiers in the respective IVD collection. The "Standardized Variant (PROPOSED)" column, which has been shaded gray, includes a subset of the 1,002 proposed Standardized Variants. The "JIS X 0213," "Jinmei-yō Kanji," "IBM," and "ARIB STD-B24" columns indicate the Japanese sources for these 89 CJK Compatibility Ideographs.

(*Three pairs of cells have been shaded blue to denote CJK Compatibility Ideographs that correspond to more than one registered IVS within an IVD collection, with the darker one being the preferred IVS because it corresponds to a JIS X 0213 kanji. Two pairs of cells have been shaded yellow to denote pairs of CJK Compatibility Ideographs that simply share the same canonical equivalent.*)

CJK Compatibility Ideograph	Canonical Equivalent	Adobe-Japan1		Hanyo-Denshi		Standardized Variant (PROPOSED)	JIS X 0213	Jinmei-yō Kanji	IBM	ARIB STD-B24
		IVS	Identifier	IVS	Identifier					
U+F91D	U+6B04	<6B04,E0100>	CID+13392	<6B04,E0103>	JC8627	<6B04,FE00>	1-86-27	JINMEI		
U+F928	U+5ECA	<5ECA,E0101>	CID+20303	<5ECA,E0104>	JC8414	<5ECA,FE00>	1-84-14	JINMEI		
U+F929	U+6717	<6717,E0100>	CID+20305	<6717,E0106>	JC8546	<6717,FE00>	1-85-46	JINMEI		
U+F936	U+865C	<865C,E0100>	CID+13394	<865C,E0103>	JC9147	<865C,FE00>	1-91-47	JINMEI		
U+F970	U+6BBA	<6BBA,E0100>	CID+13344	<6BBA,E0103>	JC8641	<6BBA,FE00>	1-86-41			
U+F9D0	U+985E	<985E,E0100>	CID+13396	<985E,E0103>	JC9404	<985E,FE00>	1-94-04	JINMEI		
U+F9DC	U+9686	<9686,E0101>	CID+13393	<9686,E0106>	JC9361	<9686,FE00>	1-93-61			
U+FA10	U+585A	<585A,E0101>	CID+7746	<585A,E0105>	JC1555	<585A,FE00>	1-15-55			
U+FA10	U+585A	<585A,E0102>	CID+8422	<585A,E0106>	IB1603	<585A,FE00>			IBM	
U+FA12	U+6674	<6674,E0100>	CID+8481	<6674,E0103>	IB2015	<6674,FE00>			IBM	
U+FA15	U+51DE	<51DE,E0101>	CID+20307			<51DE,FE00>	1-87-58			
U+FA15	U+51DE	<51DE,E0104>	CID+8542			<51DE,FE00>			IBM	
U+FA16	U+732A	<732A,E0100>	CID+8548	<732A,E0103>	JC8779	<732A,FE00>	1-87-79	JINMEI	IBM	
U+FA17	U+76CA	<76CA,E0101>	CID+8571	<76CA,E0103>	JTFA17	<76CA,FE00>			IBM	
U+FA18	U+793C	<793C,E0101>	CID+8579	<793C,E0103>	IB2536	<793C,FE00>			IBM	
U+FA19	U+795E	<795E,E0100>	CID+8580	<795E,E0103>	JC8928	<795E,FE00>	1-89-28	JINMEI	IBM	
U+FA1A	U+7965	<7965,E0100>	CID+8581	<7965,E0103>	JC8929	<7965,FE00>	1-89-29	JINMEI	IBM	
U+FA1B	U+798F	<798F,E0101>	CID+8583	<798F,E0103>	JC8933	<798F,FE00>	1-89-33	JINMEI	IBM	
U+FA1C	U+9756	<9756,E0101>	CID+8587	<9756,E0103>	IB3208	<9756,FE00>			IBM	
U+FA1D	U+7CBE	<7CBE,E0100>	CID+8590	<7CBE,E0103>	JTFA1D	<7CBE,FE00>			IBM	
U+FA1E	U+7FBD	<7FBD,E0100>	CID+8599	<7FBD,E0103>	IB2730	<7FBD,FE00>			IBM	
U+FA20	U+8612	<8612,E0101>	CID+8612	<8612,E0107>	JTBAAD	<8612,FE00>			IBM	
U+FA20	U+8612	<8612,E0102>	CID+21073	<8612,E0105>	JD8724	<8612,FE00>	2-87-24			
U+FA22	U+8AF8	<8AF8,E0100>	CID+8622	<8AF8,E0103>	JC9214	<8AF8,FE00>	1-92-14	JINMEI	IBM	
U+FA25	U+9038	<9038,E0102>	CID+8633	<9038,E0108>	JTFA25S	<9038,FE00>			IBM	
U+FA26	U+90FD	<90FD,E0100>	CID+8636	<90FD,E0103>	JC9274	<90FD,FE00>	1-92-74	JINMEI	IBM	
U+FA2A	U+98EF	<98EF,E0100>	CID+8699	<98EF,E0104>	JTFA2A	<98EF,FE00>			IBM	
U+FA2B	U+98FC	<98FC,E0101>	CID+8700	<98FC,E0103>	JTFA2B	<98FC,FE00>			IBM	
U+FA2C	U+9928	<9928,E0101>	CID+8702	<9928,E0103>	IB0457	<9928,FE00>			IBM	
U+FA2D	U+9DB4	<9DB4,E0100>	CID+8715	<9DB4,E0103>	IB1173	<9DB4,FE00>			IBM	
U+FA30	U+4FAE	<4FAE,E0101>	CID+13382	<4FAE,E0103>	JC1424	<4FAE,FE00>	1-14-24	JINMEI		
U+FA31	U+50E7	<50E7,E0101>	CID+13360	<50E7,E0105>	JC1441	<50E7,FE00>	1-14-41	JINMEI		
U+FA32	U+514D	<514D,E0101>	CID+13389	<514D,E0104>	JC1448	<514D,FE00>	1-14-48			
U+FA33	U+52C9	<52C9,E0100>	CID+13385	<52C9,E0103>	JC1467	<52C9,FE00>	1-14-67	JINMEI		
U+FA34	U+52E4	<52E4,E0101>	CID+13338	<52E4,E0103>	JC1472	<52E4,FE00>	1-14-72	JINMEI		
U+FA35	U+5351	<5351,E0100>	CID+13378	<5351,E0103>	JC1478	<5351,FE00>	1-14-78	JINMEI		
U+FA36	U+559D	<559D,E0101>	CID+7651	<559D,E0103>	JC1512	<559D,FE00>	1-15-12			
U+FA37	U+5606	<5606,E0100>	CID+13366	<5606,E0103>	JC1515	<5606,FE00>	1-15-15	JINMEI		
U+FA38	U+5668	<5668,E0101>	CID+13333	<5668,E0103>	JC1522	<5668,FE00>	1-15-22	JINMEI		
U+FA39	U+5840	<5840,E0101>	CID+13384	<5840,E0103>	JC1558	<5840,FE00>	1-15-58			
U+FA3A	U+58A8	<58A8,E0100>	CID+13387	<58A8,E0103>	JC1562	<58A8,FE00>	1-15-62	JINMEI		
U+FA3B	U+5C64	<5C64,E0101>	CID+13361	<5C64,E0103>	JC4765	<5C64,FE00>	1-47-65	JINMEI		
U+FA3C	U+5C6E	<5C6E,E0100>	CID+16837	<5C6E,E0103>	JC4766	<5C6E,FE00>	1-47-66			
U+FA3D	U+6094	<6094,E0100>	CID+13326	<6094,E0103>	JC8448	<6094,FE00>	1-84-48	JINMEI		
U+FA3E	U+6168	<6168,E0101>	CID+13328	<6168,E0108>	JC8460	<6168,FE00>	1-84-60			
U+FA3F	U+618E	<618E,E0100>	CID+13363	<618E,E0103>	JC8462	<618E,FE00>	1-84-62	JINMEI		

CJK Compatibility Ideograph	Canonical Equivalent	Adobe-Japan1 IVS	Adobe-Japan1 Identifier	Hanyo-Denshi IVS	Hanyo-Denshi Identifier	Standardized Variant (PROPOSED)	JIS X 0213	Jinmei-yō Kanji	IBM	ARIB STD-B24
U+FA40	U+61F2	<61F2,E0103>	CID+21072	<61F2,E0105>	JC8465	<61F2,FE00>	1-84-65	JINMEI		
U+FA41	U+654F	<654F,E0100>	CID+13381	<654F,E0103>	JC8508	<654F,FE00>	1-85-08	JINMEI		
U+FA42	U+65E2	<65E2,E0100>	CID+13334	<65E2,E0104>	JC8511	<65E2,FE00>	1-85-11			
U+FA43	U+6691	<6691,E0101>	CID+13352	<6691,E0103>	JC8535	<6691,FE00>	1-85-35	JINMEI		
U+FA44	U+6885	<6885,E0100>	CID+13375	<6885,E0103>	JC8569	<6885,FE00>	1-85-69	JINMEI		
U+FA45	U+6D77	<6D77,E0100>	CID+13327	<6D77,E0103>	JC8673	<6D77,FE00>	1-86-73	JINMEI		
U+FA46	U+6E1A	<6E1A,E0100>	CID+7700	<6E1A,E0103>	JC8687	<6E1A,FE00>	1-86-87	JINMEI		
U+FA47	U+6F22	<6F22,E0101>	CID+13332	<6F22,E0103>	JC8705	<6F22,FE00>	1-87-05	JINMEI		
U+FA48	U+716E	<716E,E0100>	CID+13347	<716E,E0103>	JC8753	<716E,FE00>	1-87-53	JINMEI		
U+FA49	U+722B	<722B,E0101>	CID+15398	<722B,E0102>	JD8009	<722B,FE00>	2-80-09			
U+FA4A	U+7422	<7422,E0100>	CID+7732	<7422,E0105>	JC8805	<7422,FE00>	1-88-05	JINMEI		
U+FA4B	U+7891	<7891,E0100>	CID+13379	<7891,E0103>	JC8907	<7891,FE00>	1-89-07	JINMEI		
U+FA4C	U+793E	<793E,E0101>	CID+13348	<793E,E0103>	JC8919	<793E,FE00>	1-89-19	JINMEI		
U+FA4D	U+7949	<7949,E0101>	CID+13345	<7949,E0103>	JC8920	<7949,FE00>	1-89-20	JINMEI		
U+FA4E	U+7948	<7948,E0100>	CID+13335	<7948,E0103>	JC8923	<7948,FE00>	1-89-23	JINMEI		
U+FA4F	U+7950	<7950,E0100>	CID+13391	<7950,E0103>	JC8924	<7950,FE00>	1-89-24	JINMEI		
U+FA50	U+7956	<7956,E0101>	CID+13359	<7956,E0103>	JC8925	<7956,FE00>	1-89-25	JINMEI		
U+FA51	U+795D	<795D,E0100>	CID+13351	<795D,E0103>	JC8927	<795D,FE00>	1-89-27	JINMEI		
U+FA52	U+798D	<798D,E0100>	CID+13325	<798D,E0103>	JC8931	<798D,FE00>	1-89-31	JINMEI		
U+FA53	U+798E	<798E,E0101>	CID+13371	<798E,E0103>	JC8932	<798E,FE00>	1-89-32	JINMEI		
U+FA54	U+7A40	<7A40,E0100>	CID+13343	<7A40,E0103>	JC8945	<7A40,FE00>	1-89-45	JINMEI		
U+FA55	U+7A81	<7A81,E0101>	CID+13373	<7A81,E0103>	JC8949	<7A81,FE00>	1-89-49	JINMEI		
U+FA56	U+7BC0	<7BC0,E0101>	CID+13358	<7BC0,E0105>	JC8968	<7BC0,FE00>	1-89-68	JINMEI		
U+FA57	U+7DF4	<7DF4,E0100>	CID+13399	<7DF4,E0103>	JC9014	<7DF4,FE01>	1-90-14	JINMEI		
U+FA58	U+7E09	<7E09,E0101>	CID+18366	<7E09,E0103>	JD8448	<7E09,FE00>	2-84-48			
U+FA59	U+7E41	<7E41,E0101>	CID+13376	<7E41,E0103>	JC9019	<7E41,FE00>	1-90-19	JINMEI		
U+FA5A	U+7F72	<7F72,E0100>	CID+13353	<7F72,E0103>	JC9026	<7F72,FE00>	1-90-26	JINMEI		
U+FA5B	U+8005	<8005,E0101>	CID+13349	<8005,E0103>	JC9036	<8005,FE00>	1-90-36	JINMEI		
U+FA5C	U+81ED	<81ED,E0101>	CID+13350	<81ED,E0103>	JC9056	<81ED,FE00>	1-90-56	JINMEI		
U+FA5D	U+8279	<8279,E0101>	CID+14199	<8279,E0103>	JD8584	<8279,FE00>	2-85-84			
U+FA5E	U+8279	<8279,E0102>	CID+14198	<8279,E0104>	JD8585	<8279,FE01>	2-85-85			
U+FA5F	U+8457	<8457,E0101>	CID+13367	<8457,E0104>	JC9107	<8457,FE00>	1-91-07	JINMEI		
U+FA60	U+8910	<8910,E0101>	CID+13331	<8910,E0103>	JC9179	<8910,FE00>	1-91-79			
U+FA61	U+8996	<8996,E0101>	CID+13346	<8996,E0103>	JC9189	<8996,FE00>	1-91-89	JINMEI		
U+FA62	U+8B01	<8B01,E0100>	CID+13321	<8B01,E0103>	JC9215	<8B01,FE00>	1-92-15	JINMEI		
U+FA63	U+8B39	<8B39,E0100>	CID+13339	<8B39,E0104>	JC9216	<8B39,FE00>	1-92-16	JINMEI		
U+FA64	U+8CD3	<8CD3,E0100>	CID+13380	<8CD3,E0103>	JC9224	<8CD3,FE00>	1-92-24	JINMEI		
U+FA65	U+8D08	<8D08,E0101>	CID+13364	<8D08,E0103>	JC9229	<8D08,FE00>	1-92-29	JINMEI		
U+FA66	U+8FB6	<8FB6,E0101>	CID+15403	<8FB6,E0102>	JD8973	<8FB6,FE00>	2-89-73			
U+FA67	U+9038	<9038,E0101>	CID+13320	<9038,E0107>	JC9257	<9038,FE01>	1-92-57	JINMEI		
U+FA68	U+96E3	<96E3,E0100>	CID+13374	<96E3,E0104>	JC9367	<96E3,FE00>	1-93-67	JINMEI		
U+FA69	U+97FF	<97FF,E0101>	CID+13337	<97FF,E0108>	JC9386	<97FF,FE00>	1-93-86	JINMEI		
U+FA6A	U+983B	<983B,E0100>	CID+7788	<983B,E0103>	JC9391	<983B,FE00>	1-93-91			
U+FA6B	U+6075	<6075,E0101>	CID+13740	<6075,E0103>	JTB16C	<6075,FE00>				ARIB
U+FA6C	U+242EE	<242EE,E0101>	CID+14281	<242EE,E0103>	JTB538	<242EE,FE00>				ARIB
U+FA6D	U+8218	<8218,E0101>	CID+13695	<8218,E0103>	IB1067	<8218,FE00>				ARIB