

ISO/IEC JTC1/SC2/WG2/IRG N 2189

Date: 2017-06-15

ISO/IEC JTC 1/SC 2/WG 2/
Ideographic Rapporteur Group

Source / Contribution Identifier: Japan

Meeting: IRG#48 @ Bundang

Title: Japan Activity Report (June 2017)

Keywords: (none)

Status: For your information

Short Description: This paper summarizes activities of Japanese National Body

Proposed Conclusion / Requested Action: for review

Arguments / Text of Contribution: See attachments.

Japan Activity Report (June 2017)

1. Introduction

Japan SC2 is organized under the IPSJ/ITSCJ (Information Processing Society of Japan / Information Technology Standards Commission of Japan). The committee is responsible for all JTC1/SC2 activities, and monthly meets face-to-face. We have 18 members, 2 observers and 6 experts from IT companies and research organizations.

2. Recent Activities

2.1. Attendance at the international meetings

Ten experts attended the IRG #47 Matsuyama meeting in October.

2.2. IRG Collection 2015 Review

Reviewed the assigned part of IRG Collection 2015 version 3.0. As previous review for version 2.0, some issues on the evidences are raised. Also Japan proposes one more round of review to guarantee that all IRG members are assigned to review all characters.

2.3. Preparation of IVD Registration of Moji_Joho Collection and UNC

We have been making the proposal of new IVD collection which is the second registration from Moji-Joho Kiban Project. Most of glyphs in this collection are the glyph variants of CJK Extension F in ISO/IEC 10646 Edition 5, which is waiting for the publication.

Also, we are going to propose 4 Urgently Needed Characters which had been once concluded to be unified with the characters on CJK Extension F, but it turned out that they should be disunified.

2.4. Tool development to review proposed CJK Unified Ideograph

A software tool is under development in Japan. This tool, taking machine learning approach, helps identifying proposed CJK ideographic characters which could be unified with existing CJK ideographic character in ISO/IEC 10646 standard. We have made some feasibility tests using the glyphs on Moji-Joho Kiban project (see attached for the output samples). It would contribute to reduce the workload of IRG review in the future. Note that we need to clear the intellectual property issue, because the font resource of all CJK unified ideographs is required for this tool.

(end of document)

About the output samples:

The feasibility tests compare the glyph collection of IRG WS2015 and that of Moji_Joho Kiban project.

- The most left column show an IRG WS2015 glyph, and the right columns show the glyphs of Moji_Joho Kiban project
- Moji_Joho Kiban glyphs are identified by JMJ ID (thus, multiple JMJ glyphs might exist for single UCS codepoint), but each column shows the corresponding UCS codepoint, instead of JMJ ID.
- Each column shows the similarity score (0~1). The JMJ glyph with higher score is ordered to the left.

WS2015 glyphs	1st	2nd	3rd
02864	U+40B3	U+235B5	U+36D7
	0.995586	0.000687	0.00052
礎	礎	礎	礎

IRG WS2015 sequence for the searched glyph

UCS for found Moji_Joho glyph

Similarity score

IRG WS2015 glyph to be searched

Moji_Joho Kiban glyph with the highest score

Moji_Joho Kiban glyph with the 2nd highest score

existing Moji_Joho glyphs similar to IRG WS2015 submissions (sample)

WS2015 glyphs	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	11th	12th	13th	14th	15th	16th	17th	18th	19th	20th
02864	U+40B3 0.995586	U+235B5 0.000687	U+36D7 0.000528	U+40D8 0.000316	U+21BE8 0.000219	U+4FB3 0.000171	U+47F6 0.000143	U+7802 0.000116	U+2558C 0.000103	U+7898 0.000091	U+25491 0.000078	U+2544F 0.000077	U+23A0E 0.000076	U+5511 0.000063	U+37C7 0.000062	U+787F 0.000062	U+40EF 0.000053	U+254D7 0.000053	U+254AA 0.000050	U+213E9 0.000049
	硃	硃	硃	硃	硃	硃	硃	硃	硃	硃	硃	硃	硃	硃	硃	硃	硃	硃	硃	硃
01738	U+235B4 0.963459	U+235F4 0.019654	U+3B9A 0.002524	U+20A00 0.001369	U+20A07 0.001120	U+21649 0.000648	U+22EDB 0.000615	U+2156D 0.000510	U+210A2 0.000297	U+4183 0.000226	U+7980 0.000192	U+25EC6 0.000170	U+28F97 0.000149	U+258DA 0.000133	U+765B 0.000133	U+2163B 0.000122	U+2164F 0.000120	U+21687 0.000118	U+2192F 0.000117	U+6AFD 0.000099
	卓	卓	卓	卓	卓	卓	卓	卓	卓	卓	卓	卓	卓	卓	卓	卓	卓	卓	卓	卓
02112	U+23FD5 0.934391	U+240F3 0.018723	U+23F49 0.006996	U+7033 0.005418	U+2DCEf 0.005354	U+700C 0.005319	U+2DCE9 0.003319	U+203F0 0.002979	U+240A2 0.001104	U+6EFD 0.000740	U+6FFF 0.000367	U+50F2 0.000315	U+6F59 0.000267	U+9F4B 0.000256	U+240F0 0.000202	U+240F9 0.000196	U+22898 0.000188	U+7007 0.000175	U+6F09 0.000172	U+23F1F 0.000168
	瀉	瀉	瀉	瀉	瀉	瀉	瀉	瀉	瀉	瀉	瀉	瀉	瀉	瀉	瀉	瀉	瀉	瀉	瀉	瀉
01985	U+5F66 0.905242	U+2248A 0.005342	U+5F65 0.003611	U+7523 0.003589	U+23CBE 0.002505	U+2092E 0.001732	U+2E4D7 0.001604	U+23CEB 0.001491	U+6DEB 0.001249	U+82A6 0.001006	U+34CC 0.000990	U+7522 0.000878	U+5E8B 0.000705	U+2194B 0.000669	U+2D08E 0.000596	U+2195E 0.000433	U+6D3C 0.000418	U+25A55 0.000413	U+205C4 0.000408	U+21D82 0.000370
	彦	彦	彦	彦	彦	彦	彦	彦	彦	彦	彦	彦	彦	彦	彦	彦	彦	彦	彦	彦
02771	U+2524F 0.898388	U+2524F 0.098162	U+23259 0.001475	U+7FF0 0.000257	U+28ED1 0.000133	U+7FF0 0.000125	U+28336 0.000125	U+8792 0.000027	U+28335 0.000025	U+8D11 0.000023	U+2A4D3 0.000021	U+28301 0.000021	U+29434 0.000018	U+2D0FF 0.000018	U+229E2 0.000017	U+28888 0.000015	U+2823B 0.000014	U+3CA6 0.000014	U+27402 0.000010	U+2A142 0.000010
	翰	翰	翰	翰	翰	翰	翰	翰	翰	翰	翰	翰	翰	翰	翰	翰	翰	翰	翰	翰
01944	U+3CBC 0.896806	U+6C76 0.006041	U+6FAC 0.001973	U+6C76 0.001966	U+6C77 0.000963	U+3CC4 0.000788	U+2844E 0.000622	U+6CE4 0.000416	U+6C4A 0.000241	U+6C44 0.000224	U+23C8B 0.000204	U+23D3D 0.000185	U+3D51 0.000159	U+82C2 0.000145	U+3CC2 0.000137	U+4E49 0.000133	U+23D12 0.000116	U+6FFD 0.000114	U+6CA1 0.000103	U+21B5E 0.000102
	汙	汙	汙	汙	汙	汙	汙	汙	汙	汙	汙	汙	汙	汙	汙	汙	汙	汙	汙	汙
04379	U+91C3 0.806163	U+2890D 0.049352	U+28901 0.038984	U+91C0 0.022502	U+288F7 0.007014	U+288EC 0.006660	U+91C5 0.005286	U+2890A 0.004909	U+27956 0.002873	U+77D6 0.002733	U+91B2 0.002435	U+24177 0.001912	U+9A6A 0.001881	U+288FC 0.001811	U+288FF 0.001186	U+288D9 0.001067	U+28909 0.000950	U+56C7 0.000791	U+283FD 0.000781	U+2660D 0.000750
	醞	醞	醞	醞	醞	醞	醞	醞	醞	醞	醞	醞	醞	醞	醞	醞	醞	醞	醞	醞
02117	U+58CD 0.763467	U+912D 0.015894	U+2413D 0.011551	U+240A5 0.005771	U+807B 0.004245	U+39A3 0.003617	U+6E31 0.002893	U+703F 0.002250	U+3C58 0.002020	U+8593 0.001899	U+29ACA 0.001881	U+20187 0.001501	U+27606 0.001445	U+276F6 0.001360	U+240F2 0.001348	U+22443 0.001258	U+7054 0.001256	U+27603 0.001256	U+6173 0.001241	U+4A1F 0.001205
	漣	漣	漣	漣	漣	漣	漣	漣	漣	漣	漣	漣	漣	漣	漣	漣	漣	漣	漣	漣
04642	U+962A 0.663867	U+28E1A 0.078715	U+9659 0.031976	U+28E1D 0.018805	U+966C 0.014210	U+9642 0.012336	U+53A5 0.011152	U+5438 0.011026	U+20A71 0.008298	U+22EEE 0.007811	U+22F31 0.007750	U+20A58 0.005518	U+20BD8 0.004691	U+963A 0.004246	U+28E30 0.003936	U+6604 0.003504	U+2D190 0.003245	U+230DE 0.003042	U+20CD3 0.002423	U+28E1F 0.002365
	阪	阪	阪	阪	阪	阪	阪	阪	阪	阪	阪	阪	阪	阪	阪	阪	阪	阪	阪	阪
01947	U+6CB7 0.246931	U+6C73 0.226567	U+3CEC 0.006913	U+6DDA 0.005345	U+4F16 0.004706	U+6C70 0.002306	U+8328 0.002188	U+6CE6 0.002001	U+6CE2 0.001711	U+6D3E 0.001619	U+6D6A 0.001304	U+6D3E 0.000907	U+6CB7 0.000829	U+6CF6 0.000799	U+6D64 0.000760	U+6E60 0.000739	U+6C72 0.000683	U+224C9 0.000541	U+83E0 0.000482	U+3D40 0.000447
	波	波	波	波	波	波	波	波	波	波	波	波	波	波	波	波	波	波	波	波

existing Moji_Joho glyphs similar to IRG WS2015 submissions (sample)

WS2015 glyphs	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	11th	12th	13th	14th	15th	16th	17th	18th	19th	20th
02019	U+7172	U+23D9A	U+6840	U+209F6	U+23E7A	U+2781D	U+213C8	U+3A2A	U+23EEF	U+67CB	U+21C74	U+2ACB3	U+980C	U+262AE	U+3A0E	U+2628B	U+57A0	U+24857	U+2DC41	U+22B3D
	0.172422	0.029023	0.027588	0.013263	0.010088	0.008062	0.006535	0.005461	0.004952	0.003965	0.003550	0.003483	0.002919	0.002323	0.002321	0.002320	0.002009	0.001817	0.001782	0.001605
𧇗	𧇘	𧇙	𧇚	𧇛	𧇜	𧇝	𧇞	𧇟	𧇠	𧇡	𧇢	𧇣	𧇤	𧇥	𧇦	𧇧	𧇨	𧇩	𧇪	𧇫
02113	U+6F64	U+6F64	U+24075	U+23FAC	U+6FF6	U+648B	U+6F1D	U+2B79C	U+6F96	U+23FCC	U+6F6C	U+6E61	U+24061	U+240B6	U+6F63	U+6F0D	U+8B5A	U+6EB2	U+6F6D	U+3D4D
	0.168608	0.017404	0.001969	0.000773	0.000653	0.000221	0.000221	0.000215	0.000109	0.000097	0.000096	0.000084	0.000081	0.000079	0.000060	0.000051	0.000046	0.000044	0.000042	0.000041
潤	潤	潤	潤	潤	潤	潤	潤	潤	潤	潤	潤	潤	潤	潤	潤	潤	潤	潤	潤	潤
02757	U+7764	U+405C	U+25279	U+280FF	U+401F	U+25240	U+2698A	U+7747	U+2D9A0	U+8CD7	U+28CA1	U+77A1	U+7765	U+47F8	U+2338C	U+25505	U+35D7	U+805B	U+4073	U+25A8E
	0.151682	0.071478	0.046565	0.024087	0.019994	0.018360	0.015115	0.013357	0.012544	0.011808	0.011602	0.009748	0.009082	0.008581	0.007203	0.004976	0.004493	0.004276	0.004154	0.003887
睥	睥	睥	睥	睥	睥	睥	睥	睥	睥	睥	睥	睥	睥	睥	睥	睥	睥	睥	睥	睥
02195	U+2425D	U+71D6	U+713B	U+20C72	U+714A	U+24213	U+712A	U+710A	U+711E	U+7128	U+6114	U+242BB	U+71A0	U+22B1E	U+7106	U+70DC	U+2424A	U+3DD0	U+71BA	U+241BC
	0.130632	0.095489	0.091138	0.063251	0.059385	0.047831	0.025474	0.017852	0.015530	0.012777	0.010750	0.009588	0.008100	0.007280	0.007032	0.006075	0.005801	0.005655	0.004813	0.004361
焗	焗	焗	焗	焗	焗	焗	焗	焗	焗	焗	焗	焗	焗	焗	焗	焗	焗	焗	焗	焗
02391	U+6B24	U+23886	U+24761	U+6B24	U+72BE	U+20836	U+39CB	U+3E5C	U+2476F	U+27CB7	U+2476E	U+24746	U+3783	U+5798	U+27CB6	U+271A6	U+3EB4	U+72C0	U+7D0E	U+2475C
	0.118814	0.112794	0.104056	0.051278	0.022486	0.020367	0.017334	0.011603	0.008075	0.007941	0.005436	0.005323	0.003982	0.003585	0.003458	0.003397	0.003383	0.003354	0.003312	0.002977
𧇬	𧇭	𧇮	𧇯	𧇰	𧇱	𧇲	𧇳	𧇴	𧇵	𧇶	𧇷	𧇸	𧇹	𧇺	𧇻	𧇼	𧇽	𧇾	𧇿	𧈀
04660	U+965B	U+965E	U+251AF	U+2E953	U+2131B	U+2E94F	U+2465E	U+28ECA	U+2435B	U+49C6	U+245C7	U+26648	U+28EFC	U+2B776	U+28EA6	U+3B74	U+9656	U+43F6	U+965F	U+205F2
	0.110124	0.044554	0.014873	0.005854	0.004629	0.004626	0.003864	0.003546	0.003007	0.002849	0.002774	0.001320	0.001233	0.001165	0.001019	0.001014	0.000876	0.000762	0.000761	0.000759
陞	陞	陞	陞	陞	陞	陞	陞	陞	陞	陞	陞	陞	陞	陞	陞	陞	陞	陞	陞	陞
02046	U+6F04	U+3D29	U+6E8F	U+2DC91	U+9032	U+6ED7	U+9032	U+6F50	U+6E55	U+23F62	U+23F43	U+23D95	U+6E8F	U+6E5D	U+6ED9	U+6F01	U+6EA0	U+6DEE	U+23D9F	U+23FD4
	0.109526	0.104511	0.077859	0.062441	0.057467	0.043146	0.042205	0.036959	0.025554	0.019849	0.019769	0.019452	0.016086	0.013112	0.009762	0.009554	0.008614	0.008311	0.007794	0.006910
淮	淮	淮	淮	淮	淮	淮	淮	淮	淮	淮	淮	淮	淮	淮	淮	淮	淮	淮	淮	淮
02020	U+6E13	U+6DEB	U+23E4E	U+6E40	U+3CE5	U+6D6B	U+6D6E	U+6DCF	U+2DC4F	U+23E32	U+6EAA	U+23D7D	U+6ED8	U+2D6F3	U+6EAA	U+6DCF	U+6F60	U+6FA4	U+3D31	U+6D6E
	0.085476	0.015272	0.015084	0.001204	0.001111	0.001096	0.001077	0.000985	0.000836	0.000745	0.000678	0.000671	0.000527	0.000334	0.000309	0.000260	0.000255	0.000250	0.000242	0.000229
溪	溪	溪	溪	溪	溪	溪	溪	溪	溪	溪	溪	溪	溪	溪	溪	溪	溪	溪	溪	溪
04791	U+97F6	U+97F6	U+207CF	U+2BACC	U+4AA8	U+2798E	U+29419	U+5F6E	U+2D9A8	U+20758	U+52CF	U+2087C	U+999A	U+8A40	U+26975	U+2078E	U+27CB8	U+238E9	U+29425	U+4AA9
	0.083353	0.022688	0.019199	0.017480	0.009425	0.009073	0.008793	0.008079	0.006327	0.005855	0.005834	0.004590	0.004512	0.004405	0.004096	0.003845	0.003814	0.003750	0.003656	0.003516
韶	韶	韶	韶	韶	韶	韶	韶	韶	韶	韶	韶	韶	韶	韶	韶	韶	韶	韶	韶	韶
04715	U+96DB	U+28FE5	U+76BA	U+29023	U+24CFC	U+4A04	U+29043	U+2E8D9	U+93E3	U+275D2	U+96E2	U+23B16	U+2CBF2	U+2A62E	U+96E2	U+2A6BA	U+2A62C	U+29037	U+238EF	U+4D86
	0.031366	0.008822	0.007713	0.004172	0.003887	0.001777	0.001668	0.001561	0.001483	0.001205	0.001095	0.000875	0.000803	0.000783	0.000684	0.000532	0.000530	0.000528	0.000462	0.000445
雜	雜	雜	雜	雜	雜	雜	雜	雜	雜	雜	雜	雜	雜	雜	雜	雜	雜	雜	雜	雜