

Title: Feedback on N2878R

Source: Ma Shijie

Status: Individual Contribution

Action: For IRG

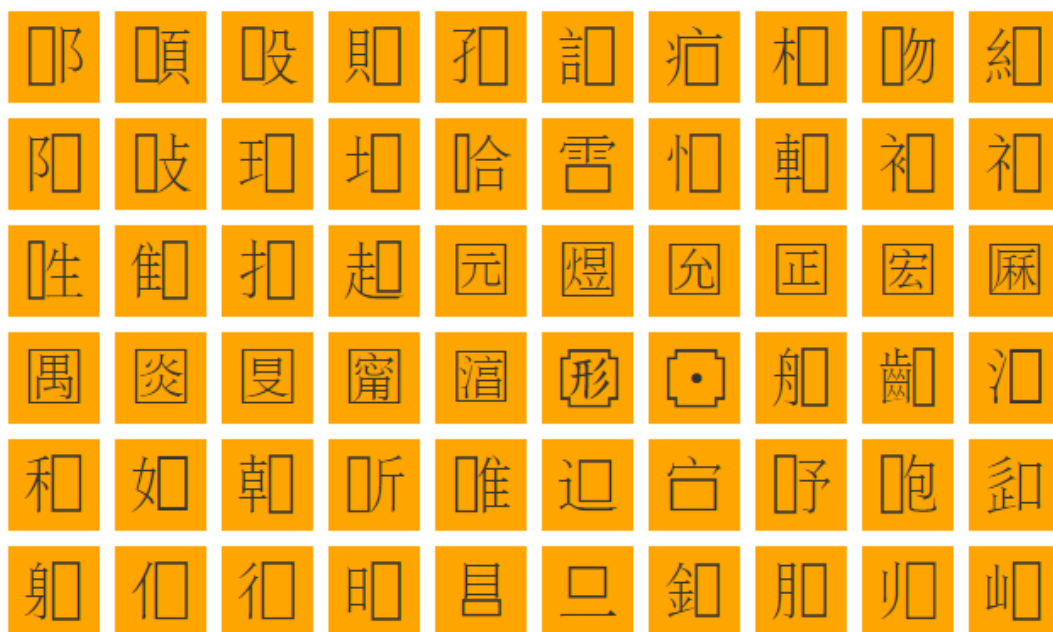
Opinions do not come from the same person, they may sometimes conflict, and we do not make any modifications. We respect the final outcome, whether it meets our needs or not.

1. New usage scenario: incomplete excavated documents.

華岳之西峙。若乃三德六行，列聖之所□也。舉厝必踐其域□□□略先□之所□也。□□息□□□□□□□□□□□□□□雲蜺以達□□□□與？然則高臺凌霄，始於覆□，□□維地，肇自涓流。是以平津筮仕，由賓王而佩

符授越遷夏商之鼎。艾綬銀章，弓旌先於髦俊；建社班瑞，光寵屬於勳庸。庶績所以咸凝，□□然□□□□授公□□□□□總管府長史□□□□公□□□□□□□□□□□□□□□雖連率總其方□，□□毗其政形，而滅沒

陝西古代文獻集成



籍合网-字符查询

2. To improve the quality of the first batch of components, each component must have 3 use cases in CJK(and Ext).

3. Comments on U+2FA20—U+2FC07

2FAAC  — 1.3 GCP-00145	Obviously it comes from 寶, but not 歲. Both of them are useful. Disunify. 8CD3  寶 寶 寶 寶 寶 寶 寶 頁 154.7 G1-3176 HB1-BBAB T1-6B33 J0-4950 K0-5E39 KP0-E0A2 V1-682F 6B72  歲 歲 歲 歲 歲 止 77.9 G1-4B6A HB1-B7B3 T1-6459 K0-6128 KP0-E5FB Refer to GCP-00129 and HCP-00564.
--	---

	2FB2B 𠂔 八 12.5 GCP-00130 2FB2C 𠂔 八 12.5 HCP-00564
2FAB7 𠂔 𠂔 — 1.4 GCP-00089 HCP-00361	Delete. 2097D 本 本 + 24.3 GKX-0156.08 T5-215A
2FAC8 𠂔 𠂔 𠂔 — 1.7 GCP-00075 HCP-00574 UTC-00817	Disunify GCP and (HCP, UTC).
2FAD9 𠂔 𠂔 𠂔 2.2 GCP-00005 TCP-00117 HCP-00145	Disunify GCP and (TCP, HCP).
2FAE2 丸 丸 丸 、 3.2 GCP-00215 TCP-00136 HCP-00169	Disunify (GCP, TCP) and HCP.
2FAE6 𠂔 𠂔 J 4.1 GCP-00014 TCP-00068	Delete TCP. 2CF00 𠂔 J 4.1 JMJ-056854
2FAEA 𠂔 𠂔 𠂔 J 4.2 GCP-00055 TCP-01047 HCP-00156	Disunify GCP and (TCP, HCP).
2FB14 𠂔 + 8.5 GCP-00205	Obviously it comes from 亭, but not 高. Both of them are useful. Disunify. 4EAD 亭 亭 亭 亭 亭 亭 亭 + 8.7 G0-4D24 HB1-AB46 T1-4F68 J0-4462 K0-6F4D KP0-E8BB V1-4A61 9AD8 高 高 高 高 高 高 高 高 189.0 G0-385F HB1-B0AA T1-586D J0-3962 K0-4D54 KP0-D0A1 V1-6C56

	Refer to GCP-00120 and GCP-00124. 2FB11 𠂔 → 8.4 GCP-00121 2FB12 𠂕 → 8.4 GCP-00125
2FB15 𠂖 𠂗 → 8.8 GCP-00206 UK-30618	Disunify GCP and UK.
2FB1C 𠂘 儿 10.1 TCP-00172	Delete. 2BDA7 𠂙 𠂙 47.0 JK-08674 342C 𠂚 𠂛 𠂜 → 8.5 G5-334D T4-2337 JA4-2132
2FB46 𠂝 𠂞 𠂟 刀 18.1 GCP-00086 TCP-01048 HCP-00168	Disunify GCP and (TCP, HCP).
2FB63 𠂠 口 30.2 GCP-00105 2FB64 𠂡 口 30.2 HCP-00395	What is the difference?
2FB6E 𠂢 口 31.1 GCP-00189	Obviously it comes from 𠂣, but not 𠂤. Both of them are useful. Disunify.
2FB7B 𠂣 𠂤 夕 34.0 TCP-01034 HCP-00173	What is the difference? 5902 𠂥 𠂦 𠂧 𠂨 𠂩 𠂪 夕 34.0 G0-623A H-8BC1 T3-2139 J0-5469 K2-2B43 KP1-3C82

		Maybe need a GCP? Disunify.
2FB81 大 37.2 央 HCP-00394		What is the difference? 592E 大 37.2 央央央央央央央 G0-516B HB1-A5A1 T1-4641 J0-317B K0-6467 KP0-F7D9 V1-514B
2FB88 工 48.0 𠂇 GCP-00083 𠂈 TCP-00074 𠂉 HCP-00088 𠂊 UTC-03377		Change radical to 乙. After U+2FA7F. 2FAFF 乙 5.1 𠂋 GCP-00065 𠂌 TCP-00073 𠂍 HCP-00087 Agree to variant of 工, but is this its only use? Disunify (GCP, TCP, HCP) and UTC.
2FB8E 干 51.0 𠂎 TCP-01028		Where does it come from?
2FB9B 𠂏 58.3 𠂐 HCP-00512		Obviously it comes from 𠂑, but not 𠂒 and 𠂓. Both of them are useful. Disunify. 43CB 𠂑 129.7 𠂑𠂑𠂑 GKX-0972.01 T6-566D JA-257A 26BB5 𠂒 140.6 𠂒 TF-323E

	2EBC6 𪗇 齒 211.6 SAT-80762
2FBAE 𣎵 木 75.0 TCP-00186	Delete. 6729 𣎵 木 75.0 GK-682D 6761 𣎵 𣎵 𣎵 𣎵 𣎵 𣎵 木 75.3 G0-4C75 H-98E7 T3-264A J0-3E72 K2-3966 KP1-4A90
2FBB5 𣎵 𣎵 木 75.4 GCP-00152 HCP-00589	Delete. 2DA60 𣎵 木 75.4 JMJ-057715 68C4 𣎵 𣎵 𣎵 𣎵 𣎵 𣎵 木 75.8 G1-467A HB1-B1F3 T1-5B39 J0-347E K0-5125 KP0-D2CF
2FBB9 𣎵 木 75.6 UTC-00813	Obviously it comes from other source, but not G-source. Both of them are useful. Disunify. 207FE 𣎵 𣎵 刀 18.14 GKX-0145.11 T7-2C4E 207FF 𣎵 𣎵 刀 18.14 GKX-0145.12 T7-2C4D

	208B6 力 19.14 勦 勦 GKX-0149.36 T7-2C4F 26306 网 122.11 罌 罌 GKX-0949.27 T5-4F63 270E1 艸 140.17 藻 藻 GKX-1070.16 T7-5453 27B44 言 149.14 譟 譟 GKX-1184.16 T7-5534 283CB 車 159.14 輶 輶 GKX-1249.27 T7-554C
2FBC4 火 86.2 𤇗 𤇗 GCP-00128 SAT-10292	Delete. 2419B 火 86.3 𤇗 K4-0069 Refer to 与. 与 康熙字典 4E0E — 1.3 与与与与与与 G0-536B HB2-C94F T2-212F J0-4D3F K2-2123 V1-4A2A

2FBE3 耒 127.0 耒 TCP-00386 耒 HCP-00482 耒 UTC-00819	Disunify (TCP, HCP) and UTC.
2FBE8 肉 130.0 夕 TCP-00383	Delete. 31F00 肉 130.0 夕 SAT-05603 7136 火 86.8 然 然 然 然 然 然 然 G0-483B HB1-B54D T1-6053 J0-4133 K0-6654 KP0-F8EF V1-5E2C
2FBED 臼 134.1 申 GCP-00073 𠂔 HCP-00553	Delete GCP. 26954 臼 134.1 申 𠂔 GHZ-53037.03 T6-2D73
2FA20 一 1.4 E GCA-Y0028	Component for 𠂔, 𠂔, 𠂔 and 𠂔. 20CE0 𠂔 30.7 𠂔 TF-2E49 21563 𠂔 35.5 𠂔 𠂔 GKX-0245.07 T6-2A53 21FF4 𠂔 47.5 𠂔 GHZ-21097.17 32A00 𠂔 79.7 𠂔 SAT-05865

2FA92

臣 131.3

𡈼
GCA-Y0142

Component for 監, 覽, 鑒,
𡈼, 𡈼, 𡈼, 𡈼, 𡈼, 𡈼 and 𡈼.

76E3

Ⅲ 108.9

監 監 監 監 監 監 𡈼
G1-3C60 HB1-BACA T1-6971 J0-3446 K0-4A78 KP0-CDF9 V1-6035

89A7

見 147.9

覽 覽 覽 覽 覽 覽
GE-4C51 H-8FCD T3-5266 J0-4D77 K2-5F64 KP1-74FB

9373

金 167.9

鑒 鑒 鑒 鑒 鑒 鑒
GE-5155 H-90AF T3-5663 J14-7B34 K2-6966 KP1-80D6

473F

豆 151.9

𡈼
T3-5271

2353F

木 75.9

𡈼

GKX-0539.14

𡈼

T5-3F68

𡈼

KP1-4C33

25AA1

立 117.9

𡈼

TF-4A6A

260B3

糸 120.9

𡈼

GHZ-53423.12

𡈼

VN-260B3

27A7E

言 149.9

𡈼

GKX-1174.03

𡈼

T7-3353

	28352 車 159.9  GKX-1246.24  T7-3435
	29235 面 176.9  GKX-1384.26  T5-6571  KP1-8517

4. According to position and shape, we can get 4 我 and 4 鳥, are they all necessary?

9D5D 鳥 196.7  G1-366C  HB1-C35A  T1-7826  J0-7321  K0-643D  KP0-F7B0  V1-6C74	9D5E 鳥 196.7  GE-5671  H-9DD8  T3-5A22  J0-7322  K1-662C  KP1-8E9E
4CD8 鳥 196.7  GKX-1491.01  T3-597C  KP1-8E9F	4CD7 鳥 196.7  GKX-1491.22  T3-5A21

5. Lost components from GF 3001 and GF 0014

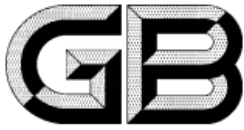
Although two national standards have been cited, there are still omissions.

 负贝	57 26 ㄣ 角争免 GF 3001 ㄣ 295 255 负 (fù) 字头 负危急陷 GF 0014 2FAE9 丿 4.1 TCP-00062 I cannot understand that it has TCP, but not GCP.
 页贝	121 72 一 页面 GF 3001 一 398 346 夏 (xià) 字头 夏嘎厦

	GF 0014									
欠人	<table><tr><td>165</td><td>97</td><td>ㄣ</td><td>欠丁</td></tr></table> GF 3001 <table><tr><td>ㄣ</td><td>89</td><td>76</td><td>尔 (ěr) 字头</td><td>尔弥你您</td></tr></table> GF 0014 2FAE9 丿 4.1 ㄣ TCP-00062 I cannot understand that it has TCP, but not GCP.	165	97	ㄣ	欠丁	ㄣ	89	76	尔 (ěr) 字头	尔弥你您
165	97	ㄣ	欠丁							
ㄣ	89	76	尔 (ěr) 字头	尔弥你您						
虎几	<table><tr><td>181</td><td>107</td><td>虍</td><td>虎虍盧</td></tr></table> GF 3001 <table><tr><td>虍</td><td>145</td><td>126</td><td>虎 (hǔ) 字头</td><td>虎虚彪滹</td></tr></table> GF 0014	181	107	虍	虎虍盧	虍	145	126	虎 (hǔ) 字头	虎虚彪滹
181	107	虍	虎虍盧							
虍	145	126	虎 (hǔ) 字头	虎虚彪滹						
兩从	<table><tr><td>368</td><td>230</td><td>𠂔</td><td>兩爾𠂔</td></tr></table> GF 3001 2FC03 雨 173.4 𠂔 𠂔 TCP-01083 HCP-00247 I cannot understand that it has TCP and HCP, but not GCP.	368	230	𠂔	兩爾𠂔					
368	230	𠂔	兩爾𠂔							

6. Lost components from GB/T 25741

A national standard has been completely omitted.



中华人民共和国国家标准

GB/T 25741—2010

信息技术 汉字编码字符集
汉字部首序和笔顺序

Information technology—Chinese ideograms coded character set—
Radical order and stroke order of Chinese ideograms

前言

本标准的附录 A、附录 B 和附录 C 是规范性附录。
本标准由中华人民共和国信息产业部提出。
本标准由全国信息技术标准化技术委员会归口。
本标准起草单位：中国电子技术标准化研究所、中华书局、中国人民解放军二炮装备研究院第四研究所。
本标准起草人：魏励、张力伟、王立建、王永平、周济萍、何正安、陈壮、代红。

GB/T 25741

 止	49 辵 (辵、𨓐) 2FBFE 辵 162.0  TCP-01121 Disunify.
---	--

𧰨祗氏	100 示（𠄎、𠄎）
𧰨雷 雷	107 𦉳（网、𦉳、𦉳、 𦉳、𦉳、𦉳、𦉳、𦉳、 𦉳）
𧰨旱干	107 𦉳（网、𦉳、𦉳、 𦉳、𦉳、𦉳、𦉳、𦉳、 𦉳）
𧰨崑山	107 𦉳（网、𦉳、𦉳、 𦉳、𦉳、𦉳、𦉳、𦉳、 𦉳）
𧰨杲木	107 𦉳（网、𦉳、𦉳、 𦉳、𦉳、𦉳、𦉳、𦉳、 𦉳）
𧰨罔亡	107 𦉳（网、𦉳、𦉳、 𦉳、𦉳、𦉳、𦉳、𦉳、 𦉳）
𧰨耒	122 耒（耒）
𧰨龜	164 龟（龜、龜）

𠃵角	165 角 (角)
𠃵辣束	167 辛 (𠃵)
𠃵麻彡	193 麻 (麻)
𠃵膏日	191 高 (高、𠃵)
𠃵鬼亼	184 鬼 (鬼)
𠃵𩺰子	177 鱼 (鱼、𩺰)

7. Lost components from National Library of China and Zhonghua Book Company

·汉字属性字典

汉字属性字典规范要求处理的汉字范围为GB 18030-2005 (UNICODE、ISO 10646-2003) 所包括的全部汉字。汉字属性字典规范的基本内容是汉字字型标准化、汉字标准发音、字型特征 (包括汉字总笔画数量、汉字起笔至末笔笔形值、部首笔画数量、部首序号、部首外起笔至末笔笔形值、异体字数量、异体字字型等)、各种编码 (包括四角号码、输入编码、其他汉字字符集编码等) 以及构词和使用频度等。

委托研制单位: 中华书局

预计完成时间: 2008年9月

国家数字图书馆工程

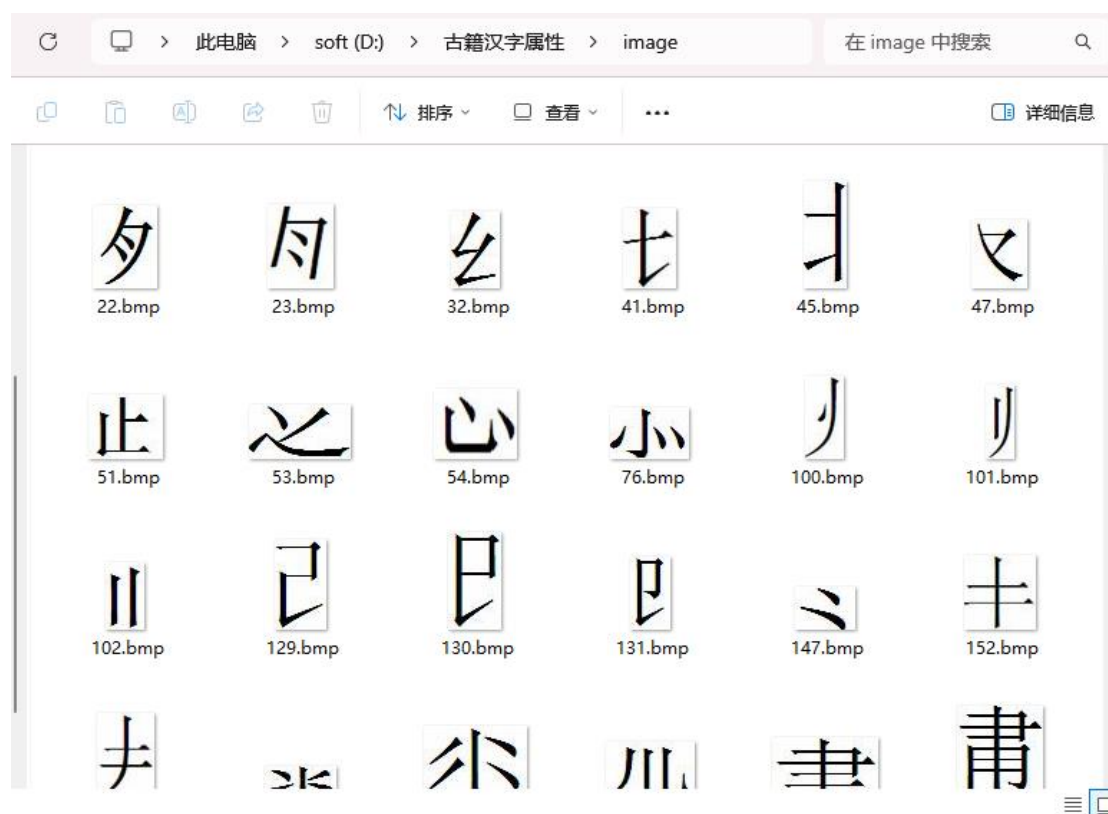
在国家图书馆子项目组的协助下,研制单位调研了国家图书馆在各类型数字资源建设中、汉字信息服务与管理的特点与特色,有针对性地调整相关规范,使其更好地符合国家图书馆数字资源建设的实际需要。研制单位于2008年4月起分批提交研制成果,与国家图书馆子项目组多次沟通,对成果进行了反复修改,最终形成了古籍用字 (包括生僻字、避讳字) 属性字典规范和应用指南。于2008年11月5日

通过项目组验收,2008 年 11 月 19 日通过馆内专家验收,2008 年 12 月 10 日至 12 月 23 日进行网上公开质询,2009 年 1 月通过业界专家验收,至此该项目全部完成。

本书是国家图书馆古籍用字(包括生僻字、避讳字)属性字典规范的研究成果集成。

在规范的研制过程中,得到了国家图书馆张志清、汪东波、申晓娟、陈宏彦、史睿、汪桂海等专家、同仁的大力支持;得到中华书局张忱石、刘宗汉,北京师范大学李国英,教育部语言文字应用研究所王晓明,北京大学图书馆沈乃文,社科院语言研究所郭小武,首都师范大学文学院叶培贵,工业和信息化部电子信息标准化研究所陈壮,中国科学院软件研究所吴建,北京羿创文源信息技术公司曾冬初,微软公司尹江红,北大方正集团唐英敏等专家的许多帮助,在此致以诚挚的谢意。

张力伟,翟喜奎主编《古籍用字(包括生僻字、避讳字)属性字典规范和应用指南》



古籍汉字属性 Demo 文件目录

叶十	𠂇
鬼𠂇	鬼
练纟	东
𠂇巳	𠂇

8. Question

Unicode used to encode the same shape into different blocks and is currently doing so.

Encode Radical, Stroke into CJK.

 31C0	<u>00001</u> 𠂇
 2E95	<u>01386</u> 𠂇 𠂇 𠂇 𠂇

Encode Radical into Component.

 2EBE	2FB76  140.0 TCP-00104 HCP-00126
---	--

After encoding the Component, if we find new evidence that

it can be used as CJK, should it be encoded once into CJK?

END.