ISO

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION

ISO-IEC JTC1/SC2/WG2

Universal Coded Character Set
ISO DIS 10646

ISO-IEC JTC1/SC2/WG2
N 669

December 23, 1990

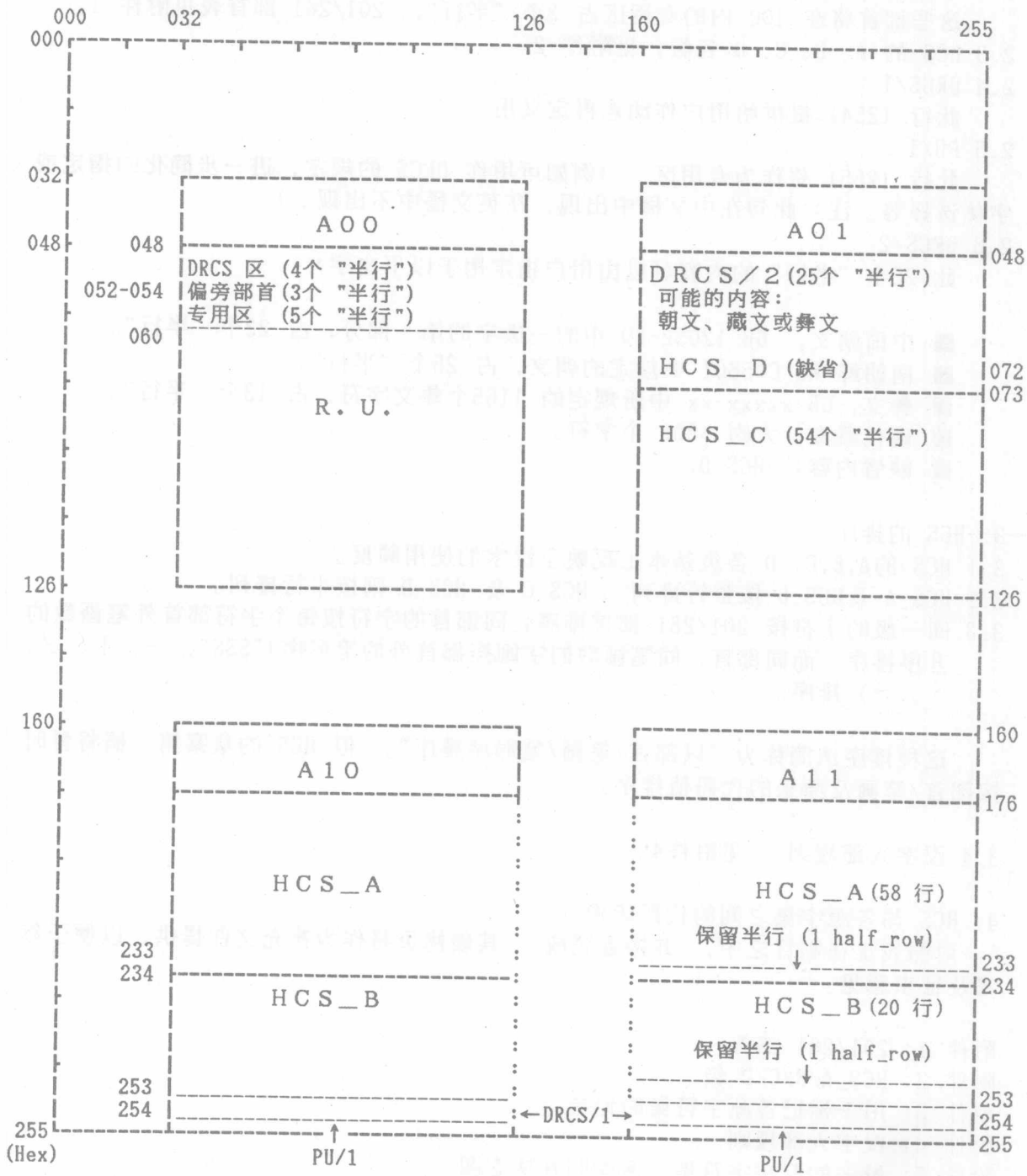
标题	:	GB 13000-90 (DP) 中的汉字集概貌
自	:	中国
要求	:	提供信息并请提意见

为跟踪并支持 ISO-IEC 10646 (UCS) 的制订，中国已经决定为新的字符集分配一个国家标准编号： GB 13000-90 (DP)。 该标准将包括汉字编码字符集 (HCS) 及一些在中国使用的其他文字， 以便满足中国海峡两岸、香港及澳门信息交换和处理时的需求 (希望亦能满足其他国家和地区的需求)。 HCS 的草案第一稿将于 1991 年 1月完成并发出征求意见。 以下是该国家标准的概貌。

1. GB 13000-90 (DP) 的总体结构

1.1 GB 13000-90 (DP) 将遵守 ISO-IEC DIS 10646 的编码结构， 并具体定义 DIS 10646 指定给中国用的各平面，即 032 组的 048 平面至 063 平面。

1.2 GB 13000-90 (DP) 中，基本中文平面 (BCP，048 平面) 的I01、I10 及I11 区将分为 4个主体部分，并在平面的底部外加 1行动态再定义区 (DRCS) 及 1行专用区 (PU)，其结构如下图所示。



General code table of Basic Chinese Plane
基本中文平面

2. GB 13000-90 (DP) 中 BCP 的内容

2.1 A00, A01, A10 及A11 将与 DIS 10646 中BMP 的相应内容相同，但可能会增加一些在中文环境下使用的特殊符号。

2.2 中国偏旁部首集。包含 201类/281个部首，与康熙部首相似但含有简化部首。这些部首将在 I00 内的专用区占 3个 "半行"，201/281 部首表见附件 1。

2.3 HCS 的 A、B、C、D 各级，见附件 2。

2.4 DRCS/1

此行 (254) 提供给用户作动态再定义用。

2.5 PU/1

此行 (255) 将作为专用区。(例如可用作 DRCS 的规定，进一步简化的指定或字体选择等。注：此句在中文稿中出现，在英文稿中不出现。)

2.6 DRCS/2

此 25个 "半行" 的内容可以由用户指定用于以下文字：

- 中国朝文，GB 12052-89 中的一级字的第一部分，占 23个 "半行"。
- 南朝鲜 KS C 5601 中规定的朝文，占 25个 "半行"。
- 彝文，GB xxxxx-xx 中所规定的 1165个彝文字符，占 13个 "半行"。
- 现代藏文，大约 1000 个字符。
- 缺省内容：HCS_D。

3. HCS 的排序

3.1 HCS 的A,B,C, D 各级基本上反映了汉字的使用频度。

3.2 HCS_A 及HCS_B 按整行排列，HCS_C 及 HCS_D 则按半行排列。

3.3 同一级的字符按 201/281 部首排序；同部首的字符按每个字符部首外笔画数的升序排序；而同部首、同笔画数的字则按部首外的笔形序("SSS"，一、丨、丿、丶、㇀）排序。

这种排序法简称为 "以部首/笔画/笔形序排序"。但 HCS 的草案第一稿将暂时按部首/笔画及原来的代码值排序。

3.4 汉字入部规则。见附件4。

4. HCS 与各源字集之间的代码转换

对照表正在制订之中，并接近完成。其硬拷贝将作为补充文件提供，以便于参考及征求意见。

附件 1: 201/281 部首

附件 2: HCS_A/B/C/D 级

附件 3: 用于标记各源字符集的记号

附件 4: 汉字入部规则

附件 5: 统一的汉字字符集：两岸归并状态图

ANNEX 1:

201/281 部首表

| | | |

附件 2:　　HCS A/B/C/D 各级

		1	2	3	4
标　准　号		GB 13000.01	GB 13000.02	GB 13000.03	GB 13000.04
缩　　写		HCS_A	HCS_B	HCS_C	HCS_D
字　符　数		11,000	3,800	5,100	2,400
覆　盖　行	自	176	234	073	048
(或半行)	止	233	253	126	072
对各源字符集的覆盖率 *	G0	100%			
	G1	100%			
	G3	< 2% → …… … → 66%			… → 100%
	G5	< 1% → …… … → 50%			
	G7	100%			
	G8	100%			
	T1	100%			
	T2	>22% → ……→ 85%		… → 100%	
	Te	> 8% → …… … → 50%			
	J0	100%			
	J1	…… → 100%			
	K0	100%			
	K1	未知	未知	未知	未知
注　解		保留半行	保留半行		动态定义 DRCS/2

* 参见附件 3。附件 3罗列并解释了此处所使用的符号。

附件 3：用于标记各源字符集的记号

	字符集标准	参考字形	字符个数	注　解
G0	GB 2312-80	24×24 宋体 GB 5007	6763	
G1	GB 12345-90	GB 5007 风格仿真	2177	不包括与 G0 相同的字
G3	GB 7589-87 非简化形式	GB 5007 风格仿真	7327	
G5	GB 7590-87 非简化形式	GB 5007 风格仿真	7039	
G7	现代汉语通用字表	GB 5007 风格仿真	41	G0，1，3，5，8 未包括的字
G8	GB 8565-89 文本通讯用	类似 GB 5007 GB 11459	289	G0，1，3，5 未包括的字
T1	CNS 11643/第一字面	资策会字型	5401	
T2	CNS 11643/第二字面	资策会字型	7650	
Te	CNS 11643/第14字面	资策会字型	6319	含大约 70个重见字
J0	JIS x 0208-90	JIS 风格仿真	6353	
J1	JIS x 0212-90	JIS 风格仿真	5801	
K0	KS C 5601-87	KS C 风格仿真	4888	含大约 200个重见字
K1	KS C xxxx_xx			

附件 4：汉字入部规则

(1) 对于左/右结构的字符，当其左，右部分均为部首时，取左不取右。

(2) 对于上/下结构的字符，当其上，下部分均为部首时，取上不取下。

(3) 对于内/外结构的字符，当其内，外部分均为部首时，取外不取内。

(4) 对于左/右、上/下均等结构的字符，并且在字符的左/右和上/下部分中均没有部首时，按下列顺序取部首：左上角、右上角、左下角然后右下角。若四个角均不成部首，则取单笔画作为部首。

ANNEX5:

Consolidated Chinese Character Set

Date 12/26/90 Time 09:50:05 G_T Merging Status

T \ G	G0	G1	G3	G5	G7	G8	T_Uniq	SUM
	6763	2177	7237	7039/7036	41	289	4550	G / T
	G0T1	G1T1	G3T1	G5T1	G7T1	G8T1	T1T1	
T1 5401	3362	1597	111	32	0	31		5133
5401	3362	1597	111	32	0	31	268	5401
	G0T2	G1T2	G3T2	G5T2	G7T2	G8T2	T2T2	
T2 7650	1048	472	3438	1605	0	6		6569
7649	1048	472	3438	1605	0	6	1080	7649
	G0TE	G1TE	G3TE	G5TE	G7TE	G8TE	TETE	
Te 6319	404	75	936	1611	0	21		3047
6249	404	75	936	1611	0	21	3202	6249
G_Uniq	G0G0	G1G1	G3G3	G5G5	G7G7	G8G8		
8794	1949	33	2752	3788	41	231		

G + T (Total) = 42842

G_Total = 23543 G_Overlap= 14749 63%

T_Total = 19299 T_Overlap= 14749 76%