ISO

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION

ISO-IEC JTC1/SC2/WG2

MULTI-OCTET CODED CHARACTER SET

ISO-IEC JTC1/SC2/WG2
N 647
September 1, 1990

Title	:	The arrangement of characters from China in ISO-IEC/DIS 10646 (中国字符在 ISO-IEC/DIS 10646 中的安排)
Source	:	Chinese national body
Action	:	To be considered by WG2 meeting in Munich

中国国家委员会注意并研究了 ISO-IEC/DIS 10646 自札幌会议以来的发展(WG2 N632, N633)。以下是反映中国国家立场的观点和要求。

1. 将 GB, JIS 和 KS表意字符从 032平面移出，并将 I01, I10 和 I11 作为未的统一的 CJK 字符的预留空间。中国支持札幌会议有关这个问题作出的结论 (WG2 N632)。

2. 汉字字符集

2.1 汉字数量较大， 除040平面外，尚需五个辅助平面，希望是 041-045 平面。为明确计，简称：

- 040 平面为基本中文平面 (BCP)
- 041-045 平面为中文辅助平面 (1st CSP－5th CSP)

2.2 汉字字符集 (Han Character Repetoire) 的内容

我国正在以下述汉字字符集/表为基础， 进行分析、比较、认同/甄别，逐步归并成一个统一的并集。

- 中国国家标准 GB 2312-80： 信息交换用汉字编码字符集基本集，简称 G0。
- GB 12345-90: 信息交换用汉字编码字符集 第一辅助集，简称 G1。
- GB 7589-87: 信息交换用汉字编码字符集，第二辅助集，简称 G2。
- 开发中的第三辅助集，约 7000 多字， 系 GB 7589-87 的非简化形式， 简称 G3。
- GB 7590-89:信息交换用汉字编码字符集 第四辅助集，简称 G5。
- 开发中的第五辅助集，约 7000 多字， 系 GB 7590-87 的非简化形式， 简称 G5。
- GB 8565-89 文本通信用编码字符集， 其中大多数已包含在 G1, G3, G5 之中，独特字仅 300个，暂称 g'。
- 《现代汉语通用字表》中罗列了的、而上述字符集未收入的少数汉字 (不足 100 字,暂称 g")。
- 在台湾使用的通用汉字标准交换码 CNS-11643
  - 第一字面 (简称 T1)，常用字 (5401个)
  - 第二字面 (简称 T2)，次常用字 (7650个)
  - 第十四字面 (简称 T14)，罕用字 (6319个)
  - 第十六字面 (简称 T16)，罕用字 (人名用字等)
- 其它罕用字
- 古汉字
- 外来汉字(The Han Characters originated from overseas)。

2.3 汉字字符集的编码规则

- 简体字与繁体字分别编码

简体字以《汉字简化字总表》为基础(2235 字)，并含少量(<500)在g'、g"收纳的表外简化字。

■ 平面分配：

- 040 平面：将争取收纳 G0,G1,G3,g',g", T1, T2 的全部汉字和 G5 中的部分字，还将收纳适用的外来汉字。
- 041 平面：放置在 040 平面中未能收入的汉字。
- 042-043 平面：留作扩展用。
- 044-045 平面：放置古汉字。

■ 排序

- 参照因素包括：部首、笔画、笔顺、 使用频度等。
- 细则待定

2.4 040 平面之专用区和动态再定义区

■ 第一专用区和动态再定义区 (P.U. & DRCS/1) 设定在 I10和 I11 的部，占2行，用于用户自定义字符或动态再定义字。

■ 第二专用区和动态再定义区 (P.U. & DRCS/2) 设定在 I01的上半部，约 40 个半行(Half Row)。

可以用于：

- 安放优先级低的汉字
- 亦可安放中国使用中、大字符集的少数民族文字， 如藏文，朝文，彝文等。以满足这些文字与汉字方便地并存共用的需求。

2.5 汉字字符集在 040平面的字量：近 18000。
详细字表将在晚些时候向 SC2/WG2 报告。

2.6 040 平面示意图：



3. 中国少数民族文字的编码安排：除 BMP/Axx 外，将放入 046 和 047平面

3.1 蒙文已在 BMP

3.2 维、哈、柯文亦作为 Arabic Script 的组成部分纳入 BMP 中

3.3 朝文

GB 12052-89 中之 Chinese Hangul 置入 047平面，（5000余字）

3.4 彝文

GB xxxx-xx 置入 047平面，（1165字）

3.5 藏文

字符集标准正在开发中，约 3000字。

3.6 其它少数民族文字

4. 除上述要求外，我们希望中文编码空间能扩大到更多的平面，如 32 个平面，并连续安排，以便放置繁多的中文字符。

（第四条系根据国家技术监督局标准司的意见增加）