

# 汉字编码标准化工作机制创新探讨

Discussion on Chinese Character Coding Standardization Mechanism Innovation

■ 中国电子技术标准化研究院 黄姗姗

**摘要** 我国汉字编码标准化工作已历经四十余年的发展，是一项为信息技术产业提供基础支撑的持续性工作。随着新时期信息技术的高速发展和数字化服务的普及应用，传统的汉字编码标准化工作机制逐渐与新时期信息化需求的不匹配。通过分析当前汉字编码标准化工作机制的优势和弊端，探讨适应现阶段及未来汉字编码需求的工作机制创新路径和方法。

**关键词** 汉字编码 标准化 中文编码字符集 GB 18030 工作机制

**Abstract:** As a continuous work that provides basic support for the information technology industry, the standardization of Chinese character encoding has gone through more than 40 years of development. With the rapid development of information technology and the widespread application of digital services in the new era, the traditional working mechanism of Chinese character encoding gradually shows a mismatch with the needs of the new era. This article analyzes the advantages and disadvantages of the current standardization mechanism for Chinese character coding, and explores innovative paths and methods for adapting to the current and future needs of Chinese character encoding.

**Keywords:** Chinese character coding; standardization; Chinese coded character set; GB 18030; working mechanism

## 1 引言

中文信息处理产业是富有我国文化和语言特色的产业，汉字的信息处理是其中最重要的组成部分。汉字的编码是汉字信息处理的基础，相关标准化工作自“748工程”启动至今，已历经四十余年的发展。我国的汉字编码标准及规范先后经历了从最基础的GB/T 2312—1980《信息交换用汉字编码字符集 基本集》，到指导性技术文件GBK《汉字内码扩展规范》，再到GB 18030《信息技术 中文编码字符集》的演变。收录的汉字由最初的6 763个增长到目前的88 115个，为汉字信息处理技术的大规模应用奠定了坚实的基础。在这一过程中，也形成了较为稳定的汉字编码标准化工作组织和工作机制。

随着互联网、大数据、人工智能等信息技术的快速发展，数字化、智能化服务广泛应用，对汉字编码标准化工作提出了更高的要求。传统的工作机制在新时期的需求面前逐渐呈现出弊端，例如工作

周期长，难以满足高时效性的用字需求；新增汉字来源渠道单一，行业覆盖面不足等。因此有必要研究建立更加适应当前以及未来发展需求的创新工作机制。

## 2 当前汉字编码标准化工作机制

### 2.1 工作组织

我国汉字编码标准的制修订工作由全国信息技术标准化技术委员会(SAC/TC28,简称“全国信标委”)下属的字符集与编码分技术委员会(SAC/TC28/SC2,简称“全国信标委SC2”)组织开展。

全国信标委SC2由国家标准化管理委员会和工业和信息化部共同管理，其印章由国家标准化管理委员会颁发，接受国家标准化管理委员会和工业和信息化部的工作指导，在专业上接受全国信标委的指导。

全国信标委SC2在汉字编码标准化方面的职

责包括：组织并指导汉字编码相关国家标准的研究和制、修订工作，负责相关的技术审查工作，承办有关的国际标准化文件并提出国际标准化活动建议等。

## 2.2 工作程序

### 2.2.1 总体情况

为确保我国国家标准与国际标准的兼容性，自1990年代至今，我国汉字编码工作一直采用“国际标准先行、国家标准同步”的模式。计划新增的汉字首先提交国际标准 ISO/IEC 10646《信息技术 通用编码字符集(UCS)》，待国际标准批准收录后等同转化为我国国家标准 GB/T 13000《信息技术 通用多人位编码字符集(UCS)》。同时，这些新增汉字也收录进我国自主研发的强制性国家标准 GB 18030《信息技术 中文编码字符集》的更新版本。

### 2.2.2 国际标准化工作程序

ISO/IEC 10646 是对全球所有文种进行统一编码的国际标准，由 ISO/IEC JTC1/SC2 负责修订维护。其中，汉字编码工作设立了 IRG(表意文字工作组)，负责审批来自中国、日本、韩国、越南等国家的新增汉字编码申请。全国信标委 SC2 组织业界专家开展专业技术性工作，形成新增汉字编码申请，以中国国家提案的形式提交到 IRG，再经过 IRG 多轮严格的技术审查，提交 ISO/IEC JTC1/SC2 批准，在国际标准 ISO/IEC 10646 中赋予正式编码。新增汉字编码的审核周期一般为4年。

### 2.2.3 国内标准化工作程序

我国的国家标准 GB/T 13000 等同采用国际标准 ISO/IEC 10646。在 ISO/IEC 10646 发布更新版本后，全国信标委 SC2 组织标准起草组开展 GB/T 13000 的修订预研，进行技术内容审查、翻译、校对等工作，形成标准草案提交国家标准委申请立项。国家标准委下达标准修订计划后，标准依次形成征求意见稿、送审稿和报批稿提交，最终由国家标准化管理委员会批准发布。推荐性国家标准自申请立项到发布的工作周期一般为2~3年。

强制性国家标准 GB 18030 在字汇上也与国际标

准 ISO/IEC 10646 协同更新，每次修订收录的字汇与 ISO/IEC 10646 已发布版本保持一致。强制性国家标准修订同样要经历立项申请、下达计划、征求意见稿、送审稿、报批稿以及批准发布等阶段。由于可能涉及对外通报等流程，强制性国家标准的制、修订工作周期通常比推荐性国家标准更长。

## 2.3 专业技术工作

对汉字进行编码的专业技术工作包括搜集、查证、查重、认同、赋码等一系列过程，需要语言学、文字学、历史文献学、信息处理以及标准化等大量交叉学科知识的积累和运用，通常需要多名专家协同工作。

早期各国专家对汉字的搜集整理和标准化工作进展顺利，被编码的汉字数量在2005年就已达到了7万多字，覆盖绝大多数我国社会的实际在用字。接下来的工作进度则逐渐趋于缓慢，因为尚未被编码的汉字大多是非常罕见的生僻字，具有搜集难、查证难的特点，有时需要各国专家开展大量考证工作，反复讨论审议，才能形成一致结论，为一个汉字分配合理的编码。

## 3 当前工作机制优劣分析

### 3.1 优点

统一管理、跟从国际标准，可以避免重复编码问题。

从国际标准化工作程序来看，我国在参与国际标准制修订工作时由全国信标委 SC2 统一归口管理，组织国内专业力量形成中国国家提案，与日本、韩国、越南等国家以及中国香港、中国台湾和中国澳门等地区的新增汉字编码提案一起提交到国际标准 ISO/IEC 10646。国际标准化组织对各国/地区提案进行统一管理，经过多轮审查、投票，最终为新增汉字赋予标准码位。这一过程可以在较大程度上避免汉字在不同国家/地区发生重复编码问题。

我国当前在汉字编码国家标准制修订方面采用的跟随国际标准的模式，也能够保障国家标准与国

际标准的协调性，较大程度上避免不同标准之间出现一字两码 / 一字多码等问题。

### 3.2 弊端

#### 3.2.1 标准制修订周期长，难以及时满足应用需求

同样出于避免一字多码问题的考虑，国际标准化组织对新增汉字的审查非常严格，有些字会被认为来源证据不足，或跟已收录的某个字是认同的，所以不能赋予编码。一批新增汉字从第一次提交至 IRG，一般要经过 4 年左右的立项、起草、审查、投票才能获得标准编码。

标准的修订周期长，但一些行业的应用场景对新增汉字数字化的效率需求高，比如居民身份证制证、中国台湾同胞回乡证制证等应用场景，需要在较短时间内对没有编码的生僻汉字实现数字化输入、输出和处理。在标准制定速度不能满足信息化需求的情况下，应用单位可能采用给生僻汉字分配临时码位的方式来处理。但这种做法会使相关信息在进行跨系统、跨机构传输交换时，产生编码不互通等新问题，间接造成实际应用中的一字两码 / 一字多码等问题。

#### 3.2.2 渠道来源单一，行业覆盖不全面

全国信标委 SC2 收到的新增汉字编码申请主要来自教育部和出版印刷等行业有关单位，覆盖古籍、文献、出版、科教等专业领域。其他行业单位虽然也存在生僻汉字的数字化需求，但大多没有形成向标准化组织提交待编码汉字的机制，例如公安、民政、社保等涉及人名、地名用字的行业。标准化组织不了解这些行业的具体需求，也无从获取相应的待编码汉字信息。

此外，大众也缺少向标准化组织及各类专业机构提交新增汉字编码申请或有关信息的渠道。

## 4 关于工作机制的创新路径

### 4.1 用标准修改单收录新增汉字，提高国家标准更新速度

我国的强制性国家标准 GB 18030 要跟进国际标

准 ISO/IEC 10646 的更新，最主要的部分是新增的编码汉字。对这些新增汉字，采用国家标准修改单的形式来收录，从标准化工作程序上可以较大程度地缩短时间，让国家标准保证与国际标准衔接的时效性。

### 4.2 对接重点行业，急用字先纳入国家标准

上文提到，部分机构采用给生僻汉字分配临时编码的方式来满足信息处理需求，因此部分信息系统里的实际在用字尚未获得标准编码，间接产生一字两码等问题，应尽快对这些实际在用字进行编码。全国信标委 SC2 应识别具有权威来源的实际在用字，例如公安人口信息用字。与有关单位建立对接机制，对此类汉字采用急用先行的原则，收录于国家标准 GB 18030，率先满足我国的汉字信息处理需求，再以国家标准引领国际标准的更新。

另外，其他行业部门在制定汉字相关的标准、规范和管理办法等文件时，也建议从现阶段的汉字数字化需求出发，充分考虑各类文件在技术内容方面与国家标准的协调性。

### 4.3 拓宽完善新增汉字的提交渠道

新增汉字编码的提交渠道需要进一步拓宽和完善，全国信标委 SC2 一是要主动与相关用字行业、机构建立联系；二是要广泛宣传和指导，提供公开的申请渠道，吸引有需求的机构来主动对接；三是要鼓励有意愿、有能力的机构以多种方式面向大众收集新增汉字需求。

## 5 工作展望

数字时代，汉字编码是信息化重要基础，而标准化是汉字编码发挥作用的重要途径。因此，汉字编码标准化工作是信息技术产业发展和各行业信息化发展必不可少的技术支撑。希望汉字编码标准化工作在原有的工作基础上探索形成更加适应现阶段语言文字信息化、规范化、标准化发展需求的创新工作机制，持续助力国家通用语言文字工作高质量发展。 (下转第 33 页)

续表 3

序号	检测方法	GB/T 39560 系列	GB/T 26125—2011 对应章节	主要技术差异分析	RoHS 符合性检测结果影响分析
9	邻苯二甲酸酯类 (聚合物)	GB/T 39560.8—2021	无	新增方法, 针对邻苯二甲酸酯类有害物质提出相应的测试方法	基本没影响

## 4 展望

今年是实现我国工业绿色发展“十四五”规划承前启后的关键年, 新时代新征程对工业绿色制造提出了更高要求。伴随电器电子产品的快速发展及向其他工业门类的高速渗透, 电器电子行业有害物质管控工作已经成为一项重要的基础性工作, 是电器电子产品绿色环保质量提升的重要环节, 也是促进电器电子行业绿色高质量发展的关键点。近 20 年来, RoHS 标准化工作已经成为支撑行业有害物质管控工作发展的重要基础, 作为我国 RoHS 管理工作框架中承前启后的重要环节, 对提高我国 RoHS 管控的规范性和科学性、推动电器电子产品绿色环保发展具有重要意义。立足当下, 为落实《“十四五”工业绿色发展规划》中有关“无害化、减量化、再利用和资源化”等工业绿色发展新要求, RoHS 工作组将进一步引导我国电器电子行业实现有毒有害物质替代与无害化的绿色发展, 同时也大力支撑国内相关企业积极应对欧美等发达国家的 RoHS 及 REACH 等指令法规的“绿色壁垒”, 保障我国电子制造“走

出去”发展战略的顺利实施。[15]

## 参考文献

- [1] 张军华. 电子电气产品有害物质限制使用政策法规体系综述 [J]. 信息技术与标准化, 2019(12): 6-8.
- [2] 程涛, 邢卫兵, 消融, 等. 原子荧光光谱法测定电子电气产品中的六价铬和测定方法标准化研究 [J]. 分析仪器, 2012(1): 97-101.
- [3] 高坚, 高亚欣, 赵俊莎, 等. 高效液相色谱法测定电子电气产品中四种增塑剂 [J]. 信息技术与标准化, 2019(10): 34-38.
- [4] 高坚. 电子电气产品 RoHS 符合性检测中的问题与应对 [J]. 信息技术与标准化, 2019(12): 21-26.
- [5] 高坚. RoHS 测试领域 X 荧光光谱仪的校准与性能验证 [J]. 信息技术与标准化, 2020(8): 28-32.
- [6] 李晓娟, 王赳强. 电子电气产品有害物质检测方法标准化研究 [J]. 信息技术与标准化, 2019(12): 12-15.

(收稿日期: 2023-06-25)

(上接第 26 页)

## 参考文献

- [1] 工信部. GB 18030—2022《信息技术 中文编码字符集》强制性国家标准发布 [J]. 信息技术与标准化, 2022(8): 4.
- [2] 周正正, 周晓文. 汉字编码问题的回顾与展望 [J]. 中国文字研究, 2022(1): 178-184.

- [3] 代红. GB 18030 中文编码字符集标准实施要求探讨 [J]. 信息技术与标准化, 2020(6): 14-17.
- [4] 尉迟治平. 再论中文汉字字符集 [J]. 语言研究, 2020, 40(1): 78-89.
- [5] 许寿椿. 编码字符集中子集的完整性 [J]. 中文信息学报, 1991(4): 56-62.

(收稿日期: 2023-04-25)