



中华人民共和国文化行业标准

WH/T 91—2020

汉文古籍集外字描述规范

Guidelines for description of Gaiji

2020-09-01 发布

2021-01-01 实施

中华人民共和国文化和旅游部 发布

目 次

前言	Ⅲ
1 范围	1
2 术语和定义	1
3 集外字描述基本原则	1
4 表意文字描述序列(IDS)语法规则	2
5 表意文字描述序列(IDS)的扩展	2
6 集外字拆分原则	3
7 集外字拆分流程	3
8 汉字集外字描述数据的结构	3
参考文献.....	4

前 言

本标准按照 GB/T 1.1—2009 给出的规则起草。

请注意本标准的某些内容可能涉及专利。本文件的发布机构不承担识别这些专利的责任。

本标准由中华人民共和国文化和旅游部提出。

本标准由全国图书馆标准化技术委员会(SAC/TC 389)归口。

本标准起草单位：国家图书馆、天津图书馆、北京汉王数字科技有限公司。

本标准主要起草人：白帆、陈红彦、张毅、王昭、杜立功、赵依澍、周升川、肖禹、谢冬荣、萨仁高娃、李国庆、李志峰、潘慧敏、江世盛、刘正珍、王晓健、王战波。

汉文古籍集外字描述规范

1 范围

本标准规范了汉文古籍集外字描述的基本原则、集外字拆分流程和描述数据结构。
本标准适用于汉文古籍数字化过程中对汉字集外字描述。

2 术语和定义

下列术语和定义适用于本文件。

2.1

古籍 ancient Chinese books

主要指 1911 年以前(含 1911 年)在中国书写或印刷的书籍。

[GB/T 3792.7—2008, 定义 3.1]

2.2

字符集 character set

按一定的规则确定的包含汉字及有关基本图形字符的有序集合。

2.3

集外字 Gaiji

特定的字符集以外的汉字。

注：Gaiji 为日语“外字”一词发音的英文转写。

2.4

表意文字描述符 ideographic description characters; IDC

描述文字各部件结构关系的符号,简称描述符。

2.5

表意文字描述序列 ideographic description sequence; IDS

表意文字描述符(2.4)和编码字符对表意文字字符作抽象描述的序列。

3 集外字描述基本原则

3.1 客观性

描述时应以保持字形原貌为基本前提。

3.2 唯一性

每一个被描述的集外字应具有唯一的描述方式。

3.3 可扩展性

在进行集外字描述时可依据项目具体需求进行扩展。