



中华人民共和国文化行业标准

WH/T 90—2020

汉文古籍文字认同描述规范

Unity description for Chinese character identification

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	I
1 范围	1
2 术语和定义	1
3 文字认同描述的基本原则	2
4 文字认同描述数据	2
参考文献	6

前 言

本标准按照 GB/T 1.1—2009 给出的规则起草。

请注意本文件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、汉王科技股份有限公司。

本标准主要起草人:王昭、陈红彦、谢冬荣、萨仁高娃、李国庆、潘慧敏、肖禹、张毅、白帆、杜立功、赵依澍、江世盛、孟晓静、王战波。

汉文古籍文字认同描述规范

1 范围

本标准规定了汉文古籍文字认同描述的元数据、文字认同规则描述以及文字认同实例描述的内容、结构及各要素的描述规则。

本标准适用于图书馆及相关机构开展汉文古籍数字化工作中对文字认同过程和结果进行描述。民国时期文献的文字认同可参考执行。

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

汉字 **Chinese character**

记录汉语的书写符号系统。汉字也被其他一些国家或民族用作为书写符号。

[GB/T 12200.1—90, 定义 4.1.3.6]

2.3

字体 **character style**

同一汉字(2.2)由于各种原因(历史演变、书写、印刷等)而形成的各种不同体式。

例：印刷体(宋体、仿宋体、黑体、楷体等)、手写体、古体、繁体、简体、正体、异体、俗体、讹体等。

[GB/T 12200.2—94, 定义 4.1.2.2]

2.4

文字 **script**

人类记录和传达语言的书写符号系统。

[GB/T 12200.1—90, 定义 4.1.2.7]

2.5

字音 **character pronunciation**

字的读音。汉字(2.2)中有的一字一音,有的一字多音。

[GB/T 12200.2—94, 定义 4.1.2.31]

2.6

语义 **semantics**

词或词组与它们的含义之间的关系。

[GB/T 12200.1—90, 定义 4.1.2.12]

2.7

文字认同 **Chinese character identification**

同一文字的不同字体转换为同一字体的过程。