



中华人民共和国国家标准

GB/T 31219.2—2014

图书馆馆藏资源数字化加工规范 第2部分：文本资源

Specification of library collections digitization—
Part 2: Text resources

2014-09-30 发布

2015-01-01 实施

中华人民共和国国家质量监督检验检疫总局
中国国家标准化管理委员会 发布

目 次

前言	I
1 范围	1
2 规范性引用文件	1
3 术语和定义	1
4 加工级别及内容编码	2
5 加工准备	3
6 资源采集与处理	3
7 元数据加工	4
8 命名规则	6
9 质量管理	6
参考文献	7

前 言

GB/T 31219《图书馆馆藏资源数字化加工规范》分为五个部分：

- 第 1 部分：总则；
- 第 2 部分：文本资源；
- 第 3 部分：图像资源；
- 第 4 部分：音频资源；
- 第 5 部分：视频资源。

本部分为 GB/T 31219 的第 2 部分。

本部分按照 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。

本部分由中华人民共和国文化部提出。

本部分由全国图书馆标准化技术委员会(SAC/TC 389)归口。

本部分起草单位：国家图书馆、首都图书馆、北京大学图书馆、中国科学院文献情报中心、上海图书馆上海科学技术情报研究所、浙江大学图书馆、汉王科技股份有限公司、北京方正阿帕比技术有限公司。

本部分起草人：李晓明、龙伟、赵四友、朱云、陈建新、王炜、张春红、刘秀文、张建勇、周静怡、徐强、黄晨、李明敬、魏丕。

图书馆馆藏资源数字化加工规范

第2部分:文本资源

1 范围

GB/T 31219 的本部分规定了图书馆文本资源数字化加工遵循的技术标准。

本部分适用于以文字为主要表达形式,可存在少量图表的文本文献(不包括古籍善本、手稿等特殊文献)的数字化加工。

注:数字化加工对象可以是一般印刷型文献,也可以是印刷型文献经过数字转换后的图像文件。

本部分适用于图书馆文本资源数字化加工,其他文献信息机构的文本资源数字化加工也可参照使用。

2 规范性引用文件

下列文件对于本文件的应用是必不可少的。凡是注日期的引用文件,仅注日期的版本适用于本文件。凡是不注日期的引用文件,其最新版本(包括所有的修改单)适用于本文件。

GB 2312 信息交换用汉字编码字符集 基本集

GB/T 4894—2009 信息与文献 术语

GB 13000 信息技术 通用多八位编码字符集(UCS)

GB 18030 信息技术 中文编码字符集

GB/T 25100—2010 信息与文献 都柏林核心元数据元素集

ISO/IEC 10646 信息技术 通用多八位编码字符集(UCS)[Information technology—Universal Multiple-Octet Coded Character Set (UCS)]

3 术语和定义

下列术语和定义适用于本文件。

3.1

文献 document

在文献工作过程中作为一个单位处理的记录信息或实物对象。

[GB/T 4894—2009,定义 4.1.2.2]

3.2

文本 text

以字符、符号、词、短语、段落、句子、表格或者其他字符排列形成的数据,用于表达意义,其解释基本上取决于读者对于某种自然语言或者人工语言的知识。

[GB/T 4894—2009,定义 4.1.1.2.4]

3.3

图像 image

用各种观测系统以不同形式和手段观测客观世界而获得的,可以直接或间接作用于人眼进而产生视知觉的实体。