

ICS 35.240.01  
L 70



# 中华人民共和国国家标准

GB/T 36338—2018

---

## 信息处理用藏文文献文本信息标记规范

Specification on Tibetan literature text markup for information processing

2018-06-07 发布

2019-01-01 实施

---

国家市场监督管理总局  
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 藏文文献的标记框架 .....	1
5 藏文文献文本基本属性的标记 .....	2
6 藏文文献文本结构信息的标记 .....	4
7 藏文文献文本标记规范 .....	8
附录 A (资料性附录) 《更敦群培文集》的 TEI 标记范例 .....	18

## 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。

本标准由全国信息技术标准化技术委员会(SAC/TC 28)提出并归口。

本标准起草单位:中国电子技术标准化研究院、西北民族大学、西藏自治区藏语文工作委员会办公室、西藏大学、青海师范大学、西藏自治区工业与信息化厅、中国标准化研究院、中国科学院软件研究所、青海民族大学、中国藏学研究中心、民族出版社、北京北大方正电子有限公司、潍坊北大青鸟华光照排有限公司。

本标准主要起草人:多拉、扎西加、熊涛、尼玛才让、杨本加、多杰卓玛、索南才让、冷本扎西、刘汇丹、仁青卓么、项见措、格桑多吉、黄鹤鸣、才让加。

# 信息处理用藏文文献文本信息标记规范

## 1 范围

本标准规定了藏文数字文献的文本信息标记。  
本标准适用于藏文语料库建设及藏文各种文献的数字化标记。

## 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/T 14814—1993 信息处理 文本和办公系统 标准通用置标语言(SGML)

## 3 术语和定义

下列术语和定义适用于本文件。

### 3.1

**藏文文献 Tibetan literature**

由藏文(包括古藏文及梵文转写)构成的各种文献。

### 3.2

**文本置标语言 text encoding initiative; TEI**

一种定义电子文档结构和描述其内容的国际标准语言。

注 1: 目前在信息处理界电子图书馆相关计划中普遍使用的置标语言,适用于对电子形式的全文的编码和描述。

注 2: TEI 元数据标准同时也规定了可供数据交换的标准编码格式,采用 SGML 文件格式。

### 3.3

**班智达 Bantita**

梵语,意为学识渊博的学者,专指参与经典翻译的印度等地学者。

注: 班智达和译师分为藏文文献厘定前与厘定后两种。

## 4 藏文文献的标记框架

TEI 适用于对电子形式的全文的编码和描述。TEI 元数据标准同时也规定了可供数据交换的标准编码格式,使用 GB/T 14814—1993 规定的 SGML 作为其编码语言。TEI 格式具有很大限度的灵活性、综合性、可扩展性,能支持对各种类型或特征的文档进行编码。TEI 元数据标记可以对语料库的文本属性信息与文本结构信息进行规范的标记。

TEI 文档,也是一个 SGML 文档,一般有四个部分:teiHeader(题名),front(文本前的信息),body(正文),back(文本后的信息)。TeiHeader:对电子文本对象的描述。front:对正文前的信息描述,包含位于文件最前端的项目(标头、题名页、前言、献词等。)。body:对正文信息的描述,单篇文章的整体部分,不包含正文前及正文后信息。back:对正文后的信息描述,包含附录等。

TEI 标头的第一层可以包含<fileDesc> 文档描述、<encodingDesc> 编码描述、<profileDesc>文件背