



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明有限公司。

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

景描述、<revisionDesc> 修订描述四个区段,其中只有第一个[<fileDesc>(文档描述)]是必要元素。

<fileDesc>可以包含七个子元素:<titleStmt>(题名描述)、<editionStmt>(版本描述)、<extent>(档案大小)、<publicationStmt>(出版描述)、<seriesStmt>(从编描述)、<notesStmt>(附注描述)及<sourceDesc>(来源描述)。其中,只有<titleStmt>(题名描述)、<publicationStmt>(出版描述)及<sourceDesc>(来源描述)是必要元素。因此最小的 TEI 标头只包含<fileDesc>。在这个<fileDesc>中,依次包含<titleStmt>、<publicationStmt>和<sourceDesc>。藏文 TEI 标记范例参见附录 A。

根据以上规范,tei 标记的整体架构可以表示如下:

```
<tei>
<teiheader>
<fileDesc>文档描述</fileDesc>
<encodingDesc>编码描述</encodingDesc>
<profileDesc>文件背景描述</profileDesc>
<revisionDesc>修订描述</revisionDesc>
</teiheader>
<text>
<front>正文前的信息描述</front>
<body>正文信息描述</body>
<back>正文后的信息描述
</back>
</text>
</tei>
```

5 藏文文献文本基本属性的标记

TEI 元数据标准规定了描述文本的书目信息所需要的标记,主要在 fileDesc 部分,有 6 个复合元素;在此 6 种元素的基础上,为了便于对语料搜索和统计,自定义了语料的标题、作者、来源、领域、体裁、语言层次、文类、语式、年代、译者、编者、本次、版本、出版、日期、语言、国家、性别、年代、创作时间、记录者等 18 项属性信息。

5.1 标题属性信息

文本标题信息用 TEI 标记为<title>…</title>。

示例:“གཏམ་རྒྱུད་གསེར་གྱི་ཐང་མ།”,可以标记为<title>གཏམ་རྒྱུད་གསེར་གྱི་ཐང་མ།</title>。

5.2 作者属性信息

作者信息用 TEI 标记为<author>…</author>,文本作者可以是个人或机构,如果无作者此项可以不填。

示例:“བོད་རང་སྤྱོད་ཕྱོད་པ་ཉང་ལུང་གཞུང་ལས་ཁང་།”或“དགེ་འདུན་ཚས་འཕེལ།”都可以标记为<author>བོད་རང་སྤྱོད་ཕྱོད་པ་ཉང་ལུང་གཞུང་ལས་ཁང་།</author>或<author>དགེ་འདུན་ཚས་འཕེལ།</author>。

5.3 来源属性信息

文本来源信息用 TEI 标记为<source>…</source>,来源可以是报纸、期刊、书籍等,若来源不详此栏可以不填。

示例:“ལེགས་བཤད་གསེར་ཕྱེང་།”、“བོད་ཕྱོད་པ་ཉིན་རེའི་ཚགས་པར།”、“སྤྱང་ཆར།”等可以标记为<source>

གསུང་ཚིག།</source>、<source>ཚགས་པར།</source>、<source>དུས་དེབ།</source>。

5.4 领域属性信息

文本领域信息用 TEI 标记为<field>…</field>。领域可以是文学、法律、科技、藏医等。

示例：“རྒྱ་མ་ཐའི་རྟོགས་བརྗོད།”、“དབང་གཉིས་ཐན་འཕྱར་གྱི་ངོ་བོ་དབྱེ་ཞིབ།”、“འབུམ་ཁུ་ཚུར།”等可以标记为<field>ཚིག་རིག།</field>、<field>ཚན་རྟུན།</field>、<field>གསལ་རིག།</field>。

5.5 体裁属性信息

文本体裁信息用 TEI 标记为<genre>…</genre>，体裁可以是文学、媒体、公文等。

示例：“དབང་པོ་ཅན་མའི་བརྗོད་པ།”、“འཆམ་མཐུན་སྤྱི་ཚགས་འཕྲུལ་སྐྱུ་གྱི་བྱེད་སྒོ་ཐོལ་བ།”、“སྐར་ཁྱིའི་ཡི་གེ”等可以标记为<genre>ཚིག་རིག།</genre>、<genre>གསར་འཕྱར།</genre>、<genre>གཞུང་ཡིག།</genre>。

5.6 语言层次属性信息

文本的语言层次信息用 TEI 标记为<language>…</language>，语言层次可以是单语、双语、多语种等。

示例 1：<language>བོད་ཡིག།</language>

示例 2：<language>汉文</language>

5.7 语式属性信息

文本的语式信息用 XML 标记为<style>…</style>，语式可以分为书面语和口语。

示例：“བོད་ཀྱི་ལ་གཞན་གཞིན་ལུ་འགྲུགས་པའི་ལྷགས་ཀྱི།”、“བྱིས་པ་བདེ་འཇུག།”，前者可以标记为<style>དག་ཚིག།</style>，后者可以标记为<style>ཡིག་ཚིག།</style>。

5.8 年代属性信息

文本年代信息用 TEI 标记为<date>…</date>，年代可以是出版日期、创作时间。

示例：“2010ལོའི་ཟླ10པར་དུ་བསྐྱུར།”、“1990ལོར་ལྷ་སར་བྲིས།”，可以标记为<date>2010.10</date>、<date>1990</date>。

5.9 译者属性信息

文本译者信息用 TEI 标记为<translator>…</translator>

示例：“བཀྲ་ཤིས་རྒྱལ་གྱིས་བསྐྱུར།”，可以标记为<translator>བཀྲ་ཤིས་རྒྱལ།</translator>。

5.10 编者属性信息

文本编者信息用 TEI 标记为<editor>…</editor>，编者可以分为主编、编者。

示例：“བསྐྱུར་གོས་གཙོ་སྤྱི་བྱས།”、“རྩོ་སྐྱོང་བསྐྱོགས།”，可以标记为<editor>བསྐྱུར་གོ།</editor>、<editor>རྩོ་སྐྱོང་།</editor>。

5.11 版本属性信息

文本版本信息用 TEI 标记为<edition>…</edition>，版本信息可以是不同的版本。

示例：“ཐེ་དགོ་པར།”、“ཙོ་ཞེ་པར།”，可以标记为<edition>ཐེ་དགོ་པར།</edition>、<edition>ཙོ་ཞེ་པར།</edition>。

5.12 地方属性信息

文本地方信息用 TEI 标记为<publiace>…</publiace>,地方信息可以是出版社所属的省市。

示例：“མི་རིགས་དབེ་སྐྱེན་ཁང་།”，可以标记为<publiace>མི་རིགས་དབེ་སྐྱེན་ཁང་།</publiace>。

5.13 出版属性信息

文本出版信息用 TEI 标记为<publisher>…</publisher>,出版信息是不同出版社的名称。

示例：“ཀྲུང་གོ་བོད་རིག་པ་དབེ་སྐྱེན་ཁང་།”，可以标记为<publisher>ཀྲུང་གོ་བོད་རིག་པ་དབེ་སྐྱེན་ཁང་།</publisher>。

5.14 文本语言信息

文本语言信息用 TEI 标记为<language>…</language>,语言信息是指文献的语言文种。

示例：“བོད་ཡིག”，可以标记为<language>བོད་ཡིག</language>。

5.15 国家属性信息

文本的国家信息用 TEI 标记为<nationality>…</nationality>,指不同国家的信息。

示例：“ཀྲུང་གོ”，可以标记为<nationality>ཀྲུང་གོ</nationality>。

5.16 作者性别属性信息

文本作者的信息用 TEI 标记为<author>…</author>,不同作者的性别属性。

示例：“ཚེ་དབང་ལྷ་མོས་བརྒྱུ་མཁོ།”，可以标记为<author>ཚེ་དབང་ལྷ་མོ།</author>。

5.17 时代属性信息

文本时代信息用 TEI 标记为<period>…</period>,文本时代信息可以分为古代、近代、现代、当代四个阶段。

示例：“དྲུང་ལྷ་མོས་བརྒྱུ་མཁོ།”，可以标记为<period>གནའ་རབས།</period>,表示文本的时代属性是古代。

5.18 记录者属性信息

文本记录者信息用 TEI 标记<recorder>…</recorder>,文本记录者可以是文本记录人员或文本整理人员。

示例：“ཡི་གེ་བཞེད་པེ་སྐྱེན་ཁང་།”，可以标记为<recorder>དགེ་སྐྱེན་ཁང་།</recorder>。

6 藏文文献文本结构信息的标记

6.1 文献结构

任何文件或专著都有其自身的结构,如:书信由发信人、收信人、信件本体等部分组成。公文由发文单位、收文单位、题目、文号、公文本体等部分组成;一般学术著作是由文本前的信息(出版说明等)、序、目录、正文、后记、文本后的信息(如:参考文献等)等组成。学术专著的结构如图 1 所示。

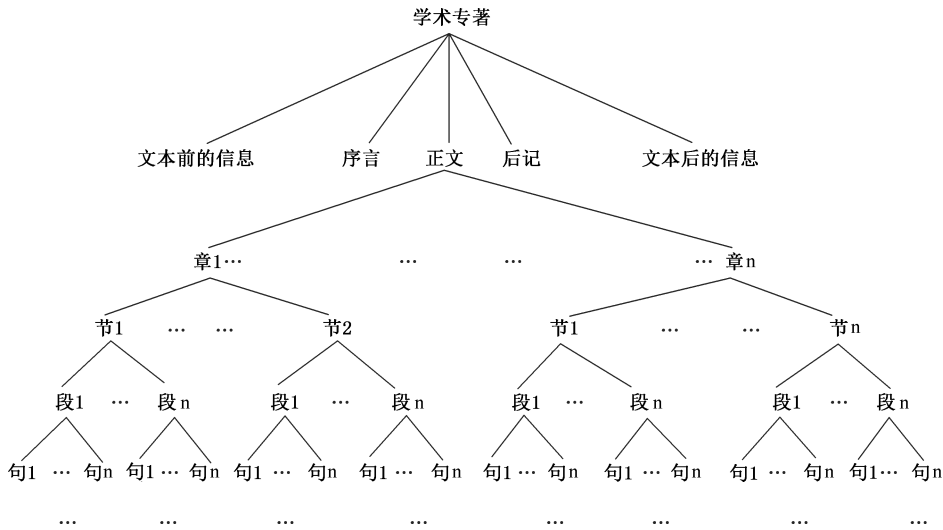


图 1 学术专著的结构树形示例图

树形图上的每一个非终极结点叫做“元素”，一个元素的子结点，叫做这个元素的“内容”。在学术著作的树形图结构中，“章”是“著作本体”的内容，“节”是“章”的内容，“段”是“节”的内容，“句”是“段”的内容，“词”是“句”的内容。树形图上的每一个终极结点没有内容，这些没有内容的终极结点，叫做“数据”。此树形图可以用来仿造书籍的传统结构。书籍的传统结构可依照阶层分解为许多单位。普遍来说，TEI 文件都符合这个简单的阶层模式。

6.2 页数、页面与行数标记

页面标记是用来区分或记录藏文典籍类文献即梵夹本的阳页与阴页，用字母 A 表示阳页，字母 B 表示阴页，其标记位置在页码值前，页码用 page 表示，页码值应是数值，其后标记行数值，行用 line 表示，行数值也应是数值，一般梵夹本每页总行数为 5 或 7 行。

示例：标记举例如下：

```
<teiHeader>
<text>
<body>
<div1>
<head>འདུལ་བ་ཕྱན་ཆོགས་ཀྱི་གཞི།</head>
<milestone unit="line" n="1b. 1">༄༅། ཤེས་རབ་ཀྱི་པ་རོལ་ཏུ་ཕྱིན་པ་ཁྱིལ་བརྒྱད་སྟོང་པ། བཅ་པོ་བཞི་བརྒྱ་ཙུ་ལྔ་པ། གསོལ་པ།
བཅོམ་ལྷན་འདས་བྱང་རྒྱལ་སེམས་དཔལ་འཛིན་པོས་ཇི་ལྟར་ཤེས་རབ་ </milestone>
<milestone unit="line" n="1b. 2">ཀྱི་པ་རོལ་ཏུ་ཕྱིན་པ་ལ་སྟུང་པར་བབྱ། ཇི་ལྟར་ཤེས་རབ་ཀྱི་པ་རོལ་ཏུ་ཕྱིན་པ་མངོན་པར་བསྟུབ་
པར་བབྱ། ཇི་ལྟར་ཤེས་རབ་ཀྱི་པ་རོལ་ཏུ་ཕྱིན་པ་བསྐྱོན་པར་བབྱ།ལགས། བཅོམ་ལྷན་འདས་ཀྱིས་ </milestone>
<milestone unit="line" n="1b. 3">བཀའ་ལྷུ་ལ་། རབ་འབྱོར་གཟུགས་ཞི་བ་དང་། གཟུགས་ལ་མ་བརྟ་དང་། གཟུགས་གསོག་དང་།
གཟུགས་སྟང་པོ་མེད་པས་ཤེས་རབ་ཀྱི་པ་རོལ་ཏུ་ཕྱིན་པ་ལ་སྟུང་པར་བཟོ། །ཆོར་བ་དང་། </milestone>
<milestone unit="line" n="1b. 4">འདུ་ཤེས་དང་། འདུ་བྱེད་རྣམས་དང་། རྣམ་པར་ཤེས་པ་ཞི་བ་དང་། རྣམ་པར་ཤེས་པ་ལ་མ་བརྟ་
དང་། རྣམ་པར་ཤེས་པ་གསོག་དང་། རྣམ་པར་ཤེས་པ་སྟང་པོ་མེད་པས་ཤེས་རབ་ཀྱི་པ་རོལ་ </milestone>
```


<milestone unit="line" n="1b.5">ཏུ་ཕྱིན་པ་ལ་སྤྱད་པར་བྱའོ། །གང་ཡང་འདི་སྐད་དུ་ངེ་ལྟར་ཤེས་རབ་ཀྱི་པ་རོལ་ཏུ་ཕྱིན་པ་མངོན་
 པར་བསྐྱབ་པར་བཀྱི་ཞེས་ཟེར་བ་ནི། རྣམ་མཁའ་ལྟར་སྟོང་པ་ཉིད་མངོན་པར་སྐྱབ་པས་ཤེས་རབ་ཀྱི་པ་ </milestone>
 </milestone><milestone unit="page" n="2a">
 <milestone unit="line" n="2a.1">རོལ་ཏུ་ཕྱིན་པ་མངོན་པར་བསྐྱབ་པར་བྱའོ། །གང་ཡང་འདི་སྐད་དུ་ངེ་ལྟར་ཤེས་རབ་ཀྱི་པ་རོལ་ཏུ་
 ཕྱིན་པ་བསྐྱོམ་པར་བཀྱི། ཞེས་ཟེར་བ་ནི་རྣམ་མཁའ་ལྟར་སྟོང་པ་ཉིད་སྐྱོམ་པས། ཤེས་རབ་ཀྱི་པ་རོལ་ཏུ་ཕྱིན་པ་བསྐྱོམ་པར་བྱའོ། །གསོལ་བ།
 བཅོམ་ </milestone>
 <milestone unit="line" n="2a.2">ལྡན་འདས་བྱང་རྒྱལ་སེམས་དཔའ་སེམས་དཔའ་ཆེན་པོས་ཤེས་རབ་ཀྱི་པ་རོལ་ཏུ་ཕྱིན་པ་ལ་ཡུན་
 རིང་པོ་ངེ་སྤྱིད་ཅིག་ཏུ་སྤྱད་ན། ཤེས་རབ་ཀྱི་པ་རོལ་ཏུ་ཕྱིན་པ་ལ་སྤྱད་པ་ལགས། བཅོམ་ལྡན་འདས་ཀྱིས་བཀའ་རྩལ་བ། རབ་འབྱོར་སེམས་དང་པོ་བསྐྱེ
 ད་པ་ནས་བཟུང་ </milestone>
 <milestone unit="line" n="2a.3">སྟེ། བྱང་རྒྱལ་གྱི་སྤྱིང་པོ་ལ་འདུག་གི་བར་དུ་ཤེས་རབ་ཀྱི་པ་རོལ་ཏུ་ཕྱིན་པ་ལ་སྤྱད་པར་བྱའོ། །དེ་
 ལྟར་མངོན་པར་བསྐྱབ་པར་བྱའོ། །དེ་ལྟར་བསྐྱོམ་པར་བྱའོ། །གསོལ་བ། བཅོམ་ལྡན་འདས་ཅི་སེམས་ཀྱི་རྒྱན་མ་ཆད་ཅིང་། མ་བཤམ་བར་རྩེས་སུ་འབྲེལ་བ་
 ས་ </milestone>
 <milestone unit="line" n="2a.4">སྤྱད་པར་བཀྱི་ལགས་སམ། བཀའ་རྩལ་བ། རབ་འབྱོར་ཡིད་ལ་བྱེད་པ་གཞན་དག་གི་གོ་སྐབས་མི་
 དབྱེ་བར་ཤེས་རབ་ཀྱི་པ་རོལ་ཏུ་ཕྱིན་པ་ལ་སྤྱད་པར་བྱའོ། །རྣམ་པ་ཐམས་ཅད་མཁྱེན་པ་ཉིད་ཡིད་ལ་བྱེད་པ་མི་དོར་བར་ཤེས་རབ་ཀྱི་པ་རོལ་ཏུ་ཕྱིན་པ་
 ལ་སྤྱད་པར་བྱ་ཞིང་། </milestone>
 <milestone unit="line" n="2a.5">ཤེས་རབ་ཀྱི་པ་རོལ་ཏུ་ཕྱིན་པ་མངོན་པར་བསྐྱབ་པར་བྱ། ཤེས་རབ་ཀྱི་པ་རོལ་ཏུ་ཕྱིན་པ་བསྐྱོམ་
 པར་བྱ་སྟེ་ཅི་ནས་ཀྱང་སེམས་དང་སེམས་ལས་བྱུང་བའི་ཆོས་རྣམས་མི་རྒྱ་བ་དེ་ལྟར་ཤེས་རབ་ཀྱི་པ་རོལ་ཏུ་ཕྱིན་པ་ལ་སྤྱད་པར་བྱའོ། །དེ་ལྟར་མངོན་
 པར་བསྐྱབ་པར་བྱའོ། ། </milestone>
 </div1>
 </body>
 </text>
 </teiHeader>

6.3 段落标记

段落标记用来标记文本中段落的开始边界和结束边界,文本段落标记用 TEI 标记为 <p>…</p>,该标记应有一个属性,属性的名称是 ID,ID 的值即为该段落的序号。规范规定,对于文中出现的标题、子标题、等均作为特殊的段落加以标记。

示例: 加入了段落标记的文本体部(含有 n 个段落)一般状如:

```

<body>
<p id="1">.....</p>
<p id="2">.....</p>
.....
<p id="n">.....</p>
</body>

```

6.4 句子标记

句子标记用来标记文本中句子的开始边界和结束边界。从逻辑的角度看,文本段落是由若干个句子组成的。在语料库中标记句子使用 TEI 标记 <s>…</s>,该标记应有一个 id 属性,属性值是句子的序号,需要注意的是,句子序号是相对于段落而言的,是段内句子编号。句子标记 <s> 应嵌套在段落标记

〈p〉内部,而不应在其内部出现〈p〉标记或者是和〈p〉标记发生交叉。

示例:在某个藏语文本第 20 个段落中出现了 4 个句子,标记情况如下:

```

<text>
<body>
<p id="20">
<a id="10" no="1">
<s id="1">
བོད་ཀྱི་བཙན་པོའི་མཚན་ལ་སྤྲོད་བཙན་ཞེས་ཟེར།</s></a>
<a id="10" no="1">
<s id="2">
ལྷ་བྱ་བར་ཡང་འབོད།</s></a>
<a id="10" no="1">
<s id="3">
བཙན་པོ་དེ་མཁྱེན་ཀྱི་ཆེ་ཞིང་། སྤྱི་ཅལ་ཆེ། སྐབས་འགར་ཏཱ་ལ་བཅེབ་སྟེ། གཡག་ཤོད་དང་རྒྱང་སོགས་ལ་རྡོན་བྱེད་དུ་ཡང་གཤེགས།</s></a>
<a id="10" no="1">
<s id="4">
རྒྱལ་ཁམས་མང་པོ་ཆབས་འོག་ཏུ་བརྩམས།</s></a></p>
</body>
</text>

```

6.5 词汇标记

词汇标记用来标记文本中词汇的开始边界和结束边界。无论是汉语文本还是藏语文本,一个句子都是由若干个词组成的。在语料库中标记词汇使用 TEI 标记〈w〉…〈/w〉来加以标记,规定了分词和词性标记规范。标记有一个名为 pos 的属性,用来记录词性该信息。

示例:词汇标记示例:

```

ཤར་ཕྱོགས་རི་བོ་ཅེ་ནས་དཀར་གསལ་རྒྱ་བ་ཤར་བྱང་།
<body>
<p id="1">
<a id="1" no="1">
<s id="1">
<title><w pos="f">ཤར་ཕྱོགས</w>
<w pos="n">རི་བོ</w>
<w pos="n">ཅེ</w>
<w pos="p">ནས</w>
<w pos="a">དཀར་གསལ</w>
<w pos="n">རྒྱ་བ</w>
<w pos="v">ཤར</w>
<w pos="vu">བྱང</w>
</title></s></a></p>
<p id="2">
.....
</body>

```

7 藏文文献文本标记规范

7.1 藏文文献文本标记项名称

本规范规定的藏文文献文本标记项分大类 39 项,小类 55 项。小类中除作者、译者、厘定审校 3 项外,实际小类标记项为 51 项,如表 1 所示。

表 1 藏文文献文本标记项及藏汉译名表

序列号	大类序号	标记项	藏文译名	汉文译名	字符长度(字符)
1	1	letter	ཀ་ཐོ།	字母序号	20
2	2	letter_no	ཀ་ཐོའི་གྲངས།	字母数号	20
3	3	reel	བཅ་པོ།	卷	20
4	4	category	རིགས་ཁོངས།	类别	30
5	5	case	ཆས་ཆན།	函	30
6	6	languages	ཡིག་རིགས།	文种	20
7	7	title	ཁ་བྱང་།	题名	200
8	8	short_name	བསྟུན་མིང་།	缩名	100
9	9	decade	བརྒྱམས་ལོ།	年代	20
10	10	place_of_creation	ཚུལ་ལུ་ལ།	创作地点	50
11	11	author	ཚུལ་པ་པོ།	作者	此为大类
12	11-1	full_name	མཚན་དངོས།	全名	100
13	11-2	another_name	མཚན་གཞན།	别名	100
14	11-3	about_the_author	ཚུལ་པ་པོའི་ལོ་རྒྱུས།	作者简介	5 000
15	12	translator	སྒྱུར་པ་པོ།	译者	此为大类
16	12-1	t_bantita	བཞི་ཏ་མ་མ།	前班智达	100
17	12-2	t_about_the_bantita	བཞི་ཏ་མ་མའི་ལོ་རྒྱུས།	前班智达简历	5 000
18	12-3	t_lotsawa	ལོ་ལྷ་བ་མ་མ།	前译师	100

表 1 (续)

序号	大类序号	标记项	藏文译名	汉文译名	字符长度(字符)
19	12-4	t_about_the_lotsawa	ལོ་ལྷ་པ་ལྷ་མའི་ལོ་རྒྱུས།	前译师简历	5 000
20	12-5	year_of_translation	སྐྱུར་དུས།	翻译年代	20
21	12-6	place_of_translation	སྐྱུར་ཡུལ།	翻译地点	100
22	12-7	transcribed	ཡིག་མཁན།	抄录者	50
23	13	text_determining	སྐད་གསར་བཅད།	厘定审校	此为大类
24	13-1	d_bantita	བརྟེན་ཉ་ཕྱི་མ།	后班智达	100
25	13-2	d_about_the_bantita	བརྟེན་ཉ་ཕྱི་མའི་ལོ་རྒྱུས།	后班智达简历	5 000
26	13-3	d_lotsawa	ལོ་ལྷ་པ་ཕྱི་མ།	后译师	100
27	13-4	d_about_the_lotsawa	ལོ་ལྷ་པ་ཕྱི་མའི་ལོ་རྒྱུས།	后译师简历	5 000
28	13-5	revise_year	ཁྱུས་པའི་དུས།	审校年代	100
29	13-6	Revise place	ཁྱུས་པའི་ཡུལ།	审校地点	100
30	14	Tibetan priest	གཏེར་སྟོན།	掘藏师	100
31	15	teach orally	སྟན་བརྒྱུད།	耳传	100
32	16	Applicant	སྐུལ་མཁན།	请求著述者	100
33	17	Funders	སྦྱིན་བདག།	资助者	100
34	18	Engraving	བར་བཞོ་མཁན།	雕版	100
35	19	version_species	དཔེ་གཞི།	版种	70
36	20	version_of_the_type	དཔེ་དབྱིབས།	版型	30
37	21	paper	ཤོག་བྲ།	纸张	30
38	22	number_of_sheets	ཐེབ་གྲངས།	张数	30
39	23	format_cm	ཐེབ་ཚད་cm	规格 cm	20
40	24	number_of_rows	ཡིག་ཐེང་།	行数	20
41	25	thang_ga	དབུ་ཞབས་སྒྱ།	唐卡	100

表 1（续）

序号号	大类序号	标记项	藏文译名	汉文译名	字符长度（字符）
42	26	ink	ལྷག་ཚ།	墨	20
43	27	specially_marked	དམིགས་བསལ་རྟགས།	特殊记号	100
44	28	version	བར་གཞི།	版本	20
45	29	publication	དཔེ་སྐྱུན་སྒྲོར།	出版情况	100
46	30	collection	ཉར་ཚགས།	馆藏	500
47	31	damaged_case	ཆ་ཚང་མིན།	残损情况	30
48	32	keyword	ཐ་སྟན་གཙོ་བོ།	关键词	200
49	33	abstract	ནང་དོན་གནད་བསྟུན།	内容提要	5 000
50	34	text	མ་ཡིག།	正文	不限
51	35	various_versions	བར་གཞི་ཁག།	各种版本	500
52	36	version_comparision	བར་གཞི་གཤིབ་བསྟུན།	版本比较	不限
53	37	entry_people	ཡིག་གཏགས།	录入	100
54	38	proofreader	ཁྱེས་དག་བ།	校对	100
55	39	remark	ཟུར་མཆན།	备注	5 000

7.2 藏文文献文本标记项名称定义

7.2.1 字母序号

- 指用藏文字母或字丁组合而成的部、卷标记字母顺序号。
- 注 1：不加元音的 30 个字母以及它们分别与四个元音组合而成的字母序号总计有 150 个。
- 注 2：不加元音的藏文字母及梵文等字母还可以构成更多字母组合。
- 注 3：在各学科名称之前加上这样的字母标号，可代替书名、易于记忆、不必直接登记书名，能防止遗失等特征。

7.2.2 字母数号

指某一函经典共有多少个字母构成，如，འདུལ་བ་གཞི་ཀའག་等构成完整经典的字母总数。

7.2.3 卷

指一个大纸卷的容量。

注：在藏文中以4句为一颂，散文以20字为偈，9 600字作为一卷。如藏文大藏经目录里清楚地记载了各部类的卷数和偈颂数，这是为了防止佛学家和译师们翻译经典时会有增减。

7.2.4 类别

指文献内容的传统学科分类。

注：按照藏文文献分类方法，《大藏经》可分为甘珠尔和丹珠尔两大类。国外学者根据文献所论述的内容，又分为显宗和密宗。显宗类若加以细分，有律、经、论三部分；密宗可细分为无上瑜伽部、瑜伽部、行部、事部四部分。这样，加上密宗总续根本八大部类分法是甘珠尔和丹珠尔的共同分类法。在这根本的八大分类之上，加上大五明里除了内明的另外四明，再加上小五明，一共分为19个根本大类。在此基础上，各类又分为若干分类。

7.2.5 函

函是指不同主题的一个完整经典，它有时包括在“部”里面，每个“部”里包含多个“函”，有时一个“部”即是一个“函”。

7.2.6 文种

指藏文或梵文。像《心咒集》全部为梵文，但一般文献基本都为藏文，只是偶尔夹杂梵文，这部分不作为记录对象。

7.2.7 题名

指文献全名称。

7.2.8 缩名

指文献缩略名。

7.2.9 年代

指文献撰写年代。

7.2.10 创作地点

指文献撰写地点。

7.2.11 作者

泛指学者。

注：许多藏族学者有不同的名字，例如：作者自己经常使用的署名；作者学习声明后上师给起的通用名称；作者自己记载不同教派的各种佛教著作中的名称；其学生们根据各自的意愿表示赞扬和崇拜的称呼；秘密名称等。

7.2.11.1 全名

指填写作者经常使用的署名。

7.2.11.2 别名

指填写作者常用的别名。

7.2.11.3 作者简介

指作者简介。

7.2.12 前译者

泛指文献的译者。

7.2.12.1 前班智达

指藏文厘定前参加文献翻译工作的外来学者。

7.2.12.2 前班智达简历

指藏文厘定前参加文献翻译工作的学者简介。

7.2.12.3 前译师

指藏文厘定前参加文献翻译工作的翻译师,或第一译者。

7.2.12.4 前译师简历

指藏文厘定前参加文献翻译工作的翻译师简历,即第一译者简介。

7.2.12.5 翻译年代

指第一次翻译工作时间。

7.2.12.6 翻译地点

指第一次翻译的地点。

7.2.12.7 抄录者

指文献抄录者。

7.2.13 厘定审校

指藏文厘定规范之后的审校。

注:据藏文史籍记载,藏文在历史上曾进行过三次较大规模的厘定规范,分别是8世纪中叶~9世纪初叶,9世纪中叶,11世纪初叶。厘定审校项主要记录文献成书后进行的第二次大规模语言厘定的信息。

7.2.13.1 后班智达

指藏文厘定后参加文献翻译工作的外来学者。

7.2.13.2 后班智达简历

指藏文厘定后参加文献翻译工作的学者简介。

7.2.13.3 后译师

指藏文厘定后参加文献翻译工作的译师,第二译者。

7.2.13.4 后译师简历

指藏文厘定后参加文献翻译工作的译师简历,即第二译者简介。

7.2.13.5 审校年代

指该文献文本厘定审校工作的年代或时间。

7.2.13.6 审校地点

指厘定审校工作的地点。

7.2.14 掘藏师

指从某处找寻挖掘到该文献的第一人。

7.2.15 耳传

以口耳相传的方式将该文献的口碑资料保存和传承的人。

7.2.16 请求著述者

根据文献创作传统,须有一名或多名请愿者向作者请求著书立说,请求著述者指该请愿者。

7.2.17 资助者

指将作品付梓印刷提供赞助的人,古籍文献多为木刻板本,资助者也指赞助刻板的施主。

7.2.18 雕版

指木刻板本的雕版或刻板人。

7.2.19 版种

指该书是否是从梵文和汉文等文字译成藏文的,以及该书的底本和来源的详细记载。

7.2.20 版型

藏文典籍总体上可以分为书写本和印本两大类。

注 1: 书写本每部版本形态都具有样式迥异、风格别样的特点,每一个书写本的书写人、书写地点、书写时间、书写字体、纸张墨料、书写底本,都有所差异甚至根本不同。

注 2: 印本除了木刻本以外,有铅印本、复印本、影印本、现代印本等不同版本类型的区别。木刻本在同一木版可以印若干次复本,这些复本的内容和版框形式基本相同。

7.2.21 纸张

指藏文书籍用的是什么纸张材料。

7.2.22 张数

指藏文书籍的张数。

7.2.23 规格

指藏文书籍的抄本和印版纸面的大小,以 cm 为计算单位。

7.2.24 行数

指藏文书籍每页的行数值。

7.2.25 唐卡

指藏文书籍插图(唐卡)。

7.2.26 墨

指墨汁、金汁、银汁等。

7.2.27 特殊记号

指书籍上的特殊记号。

7.2.28 版本

指藏文典籍的版本。

注：如对《大藏经》而言，有拉萨版、北京版、德格版等十几种版本，一般典籍也因寺院印经院不一样而版本不同。

7.2.29 出版情况

指书籍是否出版。

注：出版主要反映解放后的文献整理信息。

7.2.30 馆藏

指登记书籍馆藏地点。

7.2.31 残损情况

指登记书籍是否完整。

7.2.32 关键词

指藏文书籍的关键词，不能少于三个。

7.2.33 内容提要

指具体文献的内容提要。

7.2.34 原文

电子文献的源文本，一般为扫描件，将源文件扫描后对应到相应录入文本页。

7.2.35 各种版本

指藏文文献的各种版本，如藏文《大藏经》各种版本的图片(扫描)。

7.2.36 版本比较

相同典籍不同版本的简要比较。是为收录其他散失版本的信息。

7.2.37 录入

指电子文档录入人员姓名。

7.2.38 校对

指电子文档校对人员姓名。

7.2.39 备注

需要做说明的其他事宜内容记录。

7.3 藏文文献文本标记模式

<teiheader>

<filedesc>

<titlestmt>

<letter>ཀ་ཤོ།(字母序号)</letter>

<letter_no>ཀ་ཤོའི་གྲངས།(字母数号)</letter_no>

<reel>བཅ་པོ།(卷)</reel>

<category>རིགས་ཁོངས།(类别)</category>

<case>ཚན་པ།(函)</case>

<languages>ཡིག་རིགས།(文种)</languages>

<title>ཁ་བྱང་།(题名)</title>

<short_name>བསྐྱུ་མིང་།(缩名)</short_name>

<decade>བརྒྱམས་ལོ།(年代)</decade>

<place_of_creation>ཚུལ་ཡུལ།(创作地点)</place_of_creation>

<author>ཚུལ་པ་པོ།(作者)</author>

<full_name>མཚན་དངོས།(全名)</full_name>

<another_name>མཚན་གཞན།(别名)</another_name>

<about_the_author>ཚུལ་པ་པོའི་ལོ་རྒྱུ།(作者简介)</about_the_author>

<translator>སྒྱུར་པ་པོ།(前译者)</translator>

<t_bantita>བརྟེན་ཏུ་ཐུག།(前班智达)</t_bantita>

<t_about_the_bantita>བརྟེན་ཏུ་ཐུག་མའི་ལོ་རྒྱུ།(前班智达简历)</t_about_the_bantita>

<t_lotsawa>ལོ་ལྡན་པ་པོ།(前译师)</t_lotsawa>

<t_about_the_lotsawa>ལོ་ལྡན་པ་པོའི་ལོ་རྒྱུ།(前译师简历)</t_about_th_lotsawa>

<year_of_translation>སྒྱུར་དུས།(翻译年代)</year_of_translation>

<place_of_translation>སྒྱུར་ཡུལ།(翻译地点)</place_of_translation>

<transcribed>ཡིག་མཐུན།(抄录者)</transcribed>

<text_determining>ཡིག་མཐུན།(厘定审校)</text_determining>

<d_bantita>བརྟེན་ཏུ་བྱེ།(后班智达)</d_bantita>

<d_about_the_bantita>པནྟི་ཏ་ཕྱི་མའི་ལོ་རྒྱུས།(后班智达简历)</d_about_the_bantita>
 <d_lotsawa>ལོ་ལྡན་པ་ཕྱི་མ།(后译师)</d_lotsawa>
 <d_about_the_lotsawa>ལོ་ལྡན་པ་ཕྱི་མའི་ལོ་རྒྱུས།(后译师简历)</d_about_the_lotsawa>
 <revise_year>ཁྱུས་པའི་དུས།(审校年代)</revise_year>
 <Revise place>ཁྱུས་པའི་ཡུལ།(审校地点)</Revise place>
 <Tibetan priest>གཏེར་སྟོན།(掘藏师)</Tibetan priest>
 <teach orally>སྟན་བརྒྱུད།(耳传)</teach orally>
 <Applicant>སྐུལ་མཁན།(请求著述者)</Applicant>
 <Funders>བྱིན་བདག།(资助者)</Funders>
 <Engraving>པར་བསྐྱོ་མཁན།(雕版)</Engraving>
 <thang_ga>དབུ་ཞབས་སྒྲ།(唐卡)</thang_ga>
 <ink>སྐག་ཇ།(墨)</ink>
 <pecially_marked>དམིགས་བསལ་རྟགས།(特殊记号)</pecially_marked>
 <keyword>ཐ་སྟན་གཙོ་བོ།(关键词)</keyword>
 <abstract>ནང་དོན་གནད་བསྡུས།(内容提要)</abstract>
 <text>མ་ཡིག།(正文)</text>
 </titlestmt>
 <publicationstmt>
 <version_species>དཔེ་གཞི།(版种)</versions_pecies>
 <version_of_the_type>དཔེ་དབྱིབས།(版型)</version_of_the_type>
 <paper>ཤོག་བྲ།(纸张)</paper>
 <number_of_sheets>ཐེབ་གྲངས།(张数)</number_of_sheets>
 <format_cm>ཐེབ་ཆད།cm(规格)cm</format_cm>
 <number_of_rows>ཡིག་ཐེང་།(行数)</number_of_rows>
 <version>པར་གཞི།(版本)</version>
 <publication>དཔེ་སྐྱུན་སྒྲོར།(出版情况)</publication>
 <collection>ཏྲ་ཆ་གས།(馆藏)</collection>
 <damaged_case>ཆ་ཆད་མེན།(残损情况)</damaged_case>
 <various_versions>པར་གཞི་ཁག།(各种版本)</various_versions>
 <version_comparision>པར་གཞི་གཤེབ་བསྡུར།(版本比较)</version_comparision>
 </publicationstmt>
 </filedesc>

<encodingdesc>
 <entry_people>ཡིག་གཏགས།(录入)</entry_people>
 <proofreader>ཁྱེད་དག་སྤྱི།(校对)</proofreader>
 <remark>བྱས་མཆན།(备注)</remark>
 </encodingdesc>
 </teiheader>

附录 A

(资料性附录)

《更敦群培文集》的 TEI 标记范例

A.1 说明

下面以《更敦群培文集》为范例,详细描述具体的标记方法。本范例描述了电子文本的标题、作者、来源、语式、载体、语言、谁负责建立这个数字文本、出版日期、修订、编码、来源等方面做了详细的描述,这对藏文文献标记提供了详实的参考依据。

A.2 《更敦群培文集》的 TEI 标记结果

```

<teiHeader>
<fileDesc>
<titleStmt>
<title>མཁས་དབང་དགེ་འདུན་ཚེས་འཕེལ་གྱི་གསུང་ཚམ་ཕྱོགས་བསྒྲིགས།</title>
<author>དགེ་འདུན་ཚེས་འཕེལ་གྱིས་མཛད།</author>
<respStmt><resp>ལྷོག་རྩལ་བཟོ་མཁན།</resp>
<name>བཟ་ཤིས།</name></respStmt>
<language>བོད་ཡིག</language>
<period>དེང་རབས།</period>
<style>ཡིག་ཚུ།</style>
<field>ལྷ་རྩལ།</field>
</titleStmt>
<publicationStmt>
<distributor>སི་ཁྲོན་མི་རིགས་དཔེ་སླུང་ཁང་།</distributor>
<pubplace>སི་ཁྲོན།</pubplace>
<date>1988. 1</date>
<idno type="ISBN">7-5407-0045-9/I. 12</idno>
<address><p>ཁྲིན་ཏུ་ཡན་ཏོ་མང་ལམ་ཨང་ལ།</p></address>
<availability><p>པར་དབང་དཔར་སླུང་ཁང་ལ་དབང་བས། བསྐྱར་བཤུས་བརྒྱབ་མི་ཆོག།</p></availability>
</publicationStmt>
<sourceDesc>
<bibl><p>2010. 1. 1 ཁྲིན་དགེ་འདུན་ཚེས་འཕེལ་ཐེབས་ཚུ་ཆོགས་པའི་ལྷོག་རྩལ་པར་གཞི་མཐོ་སྒྲིང་གནང་།</p></bibl>
</sourceDesc>

```

</fileDesc>
<encodingDesc>
<normalization>
<p>མ་ཡིག་སྒྲུབ་ཀྱིན་སྒྲིག་ཨང་རྒྱལ་སྤྱིའི་ཚད་ཅན་སྒྲིག་ཨང་unicodeཅུ་བརྒྱུད་བ།</p>
<p>དེབ་ཀྱི་ནང་གི་དམིགས་བསལ་བཅད་རྟགས་དག་རང་སོར་བཞག་ཡོད།</p>
</normalization>
</encodingDesc>
<revisionDesc>
<change date="2011-03-25" who="EMB"><p>བྲུ་ཤིས་ཀྱིས་བཟོ་བཅོས་བརྒྱབ།</p></change>
</revisionDesc>
</teiHeader>
