

藏文信息处理研究进展*

Research on Tibetan Information Processing

高定国

GAO Dingguo

(西藏大学, 藏文信息技术研究中心, 西藏拉萨 850000)

(Tibetan Information Technology Research Center, Tibet University, Lhasa, Tibet, 850000, China)

摘要:藏文有着悠久的历史,是藏族人民交流思想的工具。1997年,藏文编码字符集国际、国家标准的制定作为藏文信息处理的开始,至今正好20年。这20年中藏文信息处理起步、发展,取得了较好的成绩。本文简要回顾了藏文信息处理中字、词、句、段、篇的特点、处理方法及取得的典型成果,也回顾了藏语资源建设和应用研究取得的成果,并对藏文信息处理未来的发展方向进行展望。希望能为迈入藏文信息处理的初学者展示一个藏文信息处理发展的脉络,提供一个参考。

关键词:藏文 信息处理 藏语资源

中图分类号: TP317.2 **文献标识码:** A **文章编号:** 1002-7378(2018)01-0001-11

Abstract: Tibetan has a long history and is a tool for Tibetan people to exchange ideas. In 1997, the establishment of the unicode and national standards of Tibetan-coded character sets was the beginning of Tibetan information processing, and it has been exactly 20 years. In the past 20 years, Tibetan information processing has started and developed, and good results have been achieved. This article briefly reviewed the characteristics, processing methods and typical achievements of syllable, words, sentences, paragraphs and articles in Tibetan information processing. It also reviewed the achievements of Tibetan language resources construction and application research. The future development of Tibetan information processing was also been prospected. I hope it can demonstrate the context of the development of Tibetan information processing and provide a reference for beginners entering Tibetan information processing.

Key words: Tibetan, information processing, Tibetan language resources

0 引言

藏文是一种具有1 000多年历史的拼音文字,是藏族人民交流思想的工具,是世界公认的成熟文字之一。信息时代,在党和国家领导人的高度重视下,北京、上海、西藏、甘肃、青海等地的一些院校及科研机构纷纷开展藏文信息处理研究,研制开发了较多的藏文信息处理系统,推动了藏文信息处理技术的发展,取得了较好的成绩。本文从藏文的字、词、句、段、篇的特点,信息处理方法及取得的典型成

收稿日期: 2018-01-06

作者简介: 高定国(1972—),男,教授,博士生导师,主要从事藏文信息处理研究, E-mail: gdg@utibet.edu.cn。

* 国家自然科学基金项目“跨语言社会舆情分析基础理论与关键技术研究”(61331013), 2015年度国家社会科学基金重大项目“《格萨尔》说唱语音的自动识别与格萨尔学的创新发展”(15ZDB111)和西藏大学珠峰学者人才发展支持计划项目资助。

果梳理近 20 年藏文信息处理发展的脉络。

1 藏文的字处理

藏字是由藏文字的构件组合而成的,其构件包括藏字的辅音字母、元音符号、藏文数字以及一些特殊符号。藏字虽由 30 个辅音字母和 4 个元音符号拼写组合而成,但字符不仅具有从左到右的横向排列,还具有从上到下的纵向叠加,从而构成二维的“平面文字”。在现代藏文语法中,对藏文字符构成藏字有很严格的约束。这些特征加大了计算机处理藏文字符的难度。

1.1 藏文字符属性统计

对藏文字符各种属性的研究从 1992 年开始,主要研究内容如下:用不同的样本统计藏字构件频度^[1-2]、藏文音节的构词、藏字字长、藏文音节的结构方式、藏文音节中字符位置^[3]、藏文叠加结构线性处理统计分析^[4]、书面藏语的熵值^[5]、字丁熵^[6]、音节的相对熵值和绝对熵值^[7]、藏文字符、部件、音节、词汇频度与通用度^[8]、现代藏字全集的个数、藏字的字长、藏字的结构方式、位置特征、字符频度以及所有现代藏字中的整基字丁^[9]、词汇的频度(频率)、累计频率以及通用度^[10]、藏文字的向量模型及构件特征^[11]等等。这些研究所用的样本并不相同,但研究结果很接近,为藏文信息处理技术的研究提供了详实的数据。

1.2 藏文字符输入技术

藏文字的输入是计算机处理藏文的基础和首要任务。藏文字的键盘输入技术的研究最早,接着国家标准 GB/T 17543—1998《信息技术 藏文编码字符集(基本集)键盘字母数字区的布局》^[12]被制定,藏文构件输入法、藏文的拉丁输入法、基字输入法、区位输入法^[13]、藏文音形输入法^[14]等相继被提出。然后藏汉西文混合输入和编辑的藏文处理系统——TCES^[15]、外挂式藏汉英混合处理系统^[13]、HT-藏文轻印刷系统^[14]、Sambhota、同元藏文字处理系统、班智达藏文字处理系统、藏大岗杰藏文输入系统等得以实现。之后北大方正、华光集团等开发了相当成熟的藏文激光照排系统,广泛应用于藏文报刊、书版印刷出版界。2007 年 1 月微软发布了新一代操作系统 Windows Vista,该系统自带藏文字处理系统,使得藏文的处理能力达到与英文和汉文相同的级别。近几年手持智能设备中也设计并实现了藏文的输入,苹果系统也自带藏文字符的处理,藏文的输入从计算机操作系统扩展到了 IOS、Android

系统上。最终,藏文字符的键盘输入按照藏文字符拼音性特点,以德沃拉克(Dvorak)原理把藏文字符构件布局在键盘上,以构件作为输入单位进行输入。

随着模式识别技术的发展,现代藏文识别^[16]的行切分法、藏文基本字符识别^[17]用投影法、基于字符轮廓信息的藏文笔段提取算法^[18]、外围轮廓笔划特征提取法^[19]、基于统计的 Markov 模型和藏文音节拼写规则^[20]、音节和词的 Markov 语言模型、词匹配的藏文识别后处理和统计与语法规则相结合等方法^[21]相继被提出,手写藏文识别系统^[22]得以实现。2003 年 11 月,清华大学电子工程系与西北民族大学合作研发的“多字体印刷藏文(混排汉英)文档识别系统”^[23],对藏文乌金印刷体的识别率较高,达到了应用水平。近几年藏文的识别研究除了藏文现代印刷体、传统雕版印刷体的字符识别,还拓展到藏文乌金、乌梅手写体的识别,也推出了部分系统。

藏族的语言大的分为卫藏、安多、康巴 3 大方言,但同一种方言中各小方言也有很大的差别,这加大了藏文语音处理的难度。研究者提出了从文字上对藏文声母和韵母拆分的“字丁分解法”、安多方言夏河话音节音联结构^[24]、采用波形拼接技术构建藏语语音库^[25]、基于隐马尔可夫(HMM)的藏语文本信息预测重音标注^[26]、短时能量等短时分析方法在藏语语音处理中应用^[27]等方法,建立了藏语 13 个方言点的方音数据库^[28]。2016 年 11 月,西藏大学·讯飞语音及语言联合实验室发布了基于藏语合成、藏语识别、藏汉翻译等研究成果的藏语智能语音云平台,推出了三位一体藏语输入法,是全球首款基于蜂巢输入模型的全能藏语输入软件。

1.3 藏文字符的表示、存储、交换

在藏文编码的制定过程中,针对藏文字符在计算机中表示、存储与传输时涉及的藏文字符的编码问题,研究者提出过综合编码方案^[29]、《信息交换用藏文编码字符集 辅助集》^[30]的方案^[31]和《藏文编码字符集 构件集》^[32]等设想,这些方案也被用在一些自制系统的研发上。

1993 年,国家技术监督局、电子工业部、西藏自治区有关部门正式承担起草“信息交换用藏文编码国际标准”的工作。藏文编码研究人员历经 4 年藏文编码标准的制定工作,于 1997 年 7 月,在国际标准化组织(ISO)和国际电工委员会(IEC)第 33 届编码标准第二工作组(WC2)会议上,表决通过了藏文编码的国际标准 ISO/IEC10646—1997《藏文编码字符集 基本集》。1997 年 9 月,国家技术监督局批

准发布该标准为国家标准 GB 16959—1997《信息技术 信息交换用藏文编码字符集 基本集》^[33]。该国际、国家标准的制订是藏文信息处理的里程碑,标志着藏文信息处理正式走向世界,同时,藏文也成了国内第一个具有国际标准的少数民族文字。

藏文《基本集》以藏文字符构件为编码对象,由于当时的技术原因,无法只用藏文的《基本集》完成藏文字符“二维平面”的显示和打印。2003年全国信息技术标准化技术委员会、信息产业部电子工业标准化研究所组织来自西藏自治区藏语文工作委员会、西藏大学、西北民族大学、青海师范大学以及内地有关单位的专家,以“纵向叠加”的字符组合块“字丁”作为编码对象,制定了《信息技术 藏文编码字符集 扩充集 A》和《信息技术 藏文编码字符集 扩充集 B》推荐标准,把“二维平面”的藏文字符简化为“线性字符”来处理,并应用到一些系统中。直到2007年,微软推出的 Windows Vista 系统全面支持藏文 OpenType 布局表的执行后,藏文字符的处理仅基于《基本集》就能完全实现,藏文字符的编码也统一到国际、国家《基本集》标准上了。

1.4 藏文的输出、打印

藏文从左到右“线性”排列的同时具有从上至下的叠加,从而构成二维的“平面文字”,该特性增加了藏文字符处理的难度,字体也从点阵字体发展到 TrueType、OpenType 字体。GB/T 16960.1—1997《信息技术 藏文编码字符集(基本集) 24×48 点阵字型 第1部分:白体》^[34]《信息技术 藏文编码字符集(基本集及扩充集 A) 24×48 点阵字型 吾坚琼体》《信息技术 藏文编码字符集(基本集及扩充集 A) 24×48 点阵字型 朱匝体》《信息技术 藏文编码字符集(基本集及扩充集 A) 24×48 点阵字型 白徂体》^[35]等标准先后被制定,相应的点阵字库和 TrueType 字库也被推出。

基于 Unicode 标准编码,为了满足藏文既有横向拼写性,也有纵向拼写性的特点,研究者对 OpenType 技术在藏文字库设计中的应用进行研究。2005年后 OpenType 藏文字库的技术逐步成熟起来^[36],国内外相继推出 50 多种不同风格的藏文 OpenType 字体,并得到了广泛应用,满足了藏文字符处理的各种需求,这标志着藏文信息处理技术“字”的研究走向了成熟。

2 藏文词的处理

2.1 藏文的分词

藏文文本中每个字(音节)用“隔音符”隔开,用音节组成词,没有词标记,词具有黏着性。藏文具有中文类似的分词问题,比中文的分词还更复杂。随着藏文信息技术的发展,藏文的分词经历了字典匹配、规则、深度计算等过程。研究者提出了藏文分词的一些算法、歧义和未登录词的解决^[37]、一种基于格助词和连续特征(BCCF)的书面藏文自动分词方案^[38]、藏文自动分词中紧缩词的识别与还原^[39]、融合无监督特征的藏文分词^[40]等方法,实现了 SegT^[41]、班智达藏文自动分词^[42]、央金藏文分词^[43]等系统,并制定了《信息处理用藏文分词规范》^[44]国家标准。2017年中文信息学会民族语言文字信息处理专业委员会和计算语言学专业委员会共同举办了“少数民族语言分词技术评测(Minority Language Word Segmentation, MLWS 2017)”活动,其中评测了藏文新闻类文本自动分词技术^[45]。在评测中荣获第一名的西北民族大学李亚超团队的藏文分词系统的精确率达 93.14%,召回率达 92.17%,F 值达 92.66%^[46],说明该团队在藏文分词的研究方面取得了比较好的成绩,研发的系統基本达到应用水平。

2.2 藏文的词类划分

2003年开始不断有人提出藏文词类划分问题,先后提出了 26 个类的藏语词语分类体系^[47]、依据词的句法功能将藏语词类分为 14 个大类和 26 个基本类^[48],设计实现了基于 HMM 藏文词性标注系统^[49]等,现以 23 个大类 91 个细类制定了《信息处理用藏语词类标记集》^[46]国家标准。

2.3 藏文命名实体识别

藏族人无姓,一般用星宿、自然界的物体名、祝愿或吉祥词等有实意的词作为名,当出现一个类似于名的词时要判断是否是人名难度较大。近年来,研究人名自动识别的较多,有基于层次特征^[50]、条件随机场^[51]、统计与规则相结合^[52]、最大熵和条件随机场模型相融合^[53]等方法。研究指出,藏文人名识别的 F 值能达到 93%~95%。

2.4 其他藏文词的处理

近十年中藏文词的研究是一个热点,除了分词、词类划分外,还有围绕分词和词类划分研究藏文虚词的识别^[54]、格助词的识别^[55]、数词的识别^[56]、动词的黏着性及屈折性变化^[57]、动词语法属性^[58]、副

词搭配^[59]、名词分类^[60]、数量词分类^[61]、动词分类^[62]、藻饰词语义^[63]等。

3 藏文句子的处理

藏文文本也没有显示的句子标记,为了达到计算机“理解”藏文文本,近年来,研究者从形式上研究了藏文句子自动断句方法^[64]、藏文句子边界识别方法^[65]、藏语单句的类型^[66]等;从语义上研究了藏语语义本体中的上下位关系模式匹配^[67]、基于信息处理的藏文框架语义关系^[68]、基于依存关系的藏文语义角色标准^[69]和藏文属格结构统计^[70]等。但由于藏文句子以逻辑、句意为主,从形式上断句较难,在语义理解的基础上再断句可能效果会更好。

4 藏文段落、篇章的处理

4.1 信息检索与提取

近几年中,藏文信息检索、信息提取等方面研究较多,也取得了较好的成绩,研究者提出了基于 Lucene 的藏文信息采集及检索^[71]、藏文网页倒排索引^[72]、基于奇异值分解的藏文 Web 不良信息检索^[73]、基于向量空间模型的藏文文本信息检索^[74]、语义词库管理的藏文语义检索^[75]等藏文信息检索方法,也提出了基于卡方统计量的藏文网页关键词提取^[76]、基于文本密度的藏文网页正文提取^[77]等藏文文本提取方法。特别是青海省海南藏族自治州藏文信息技术研究中心于 2016 年 8 月正式上线运行的“云藏”藏文搜索引擎,为藏文信息资源的检索

表 1 CWMT 中藏汉机器翻译评测指标对比

Table 1 Comparison of evaluation of Tibetan-Chinese machine translation in China Workshop on Machine Translation(CWMT)

评测名 Evaluation	系统名 System	BLEU5-SBP	BLEU5	BLEU6	NIST6	NIST7	GTM	mWER	mPER	ICT
CWMT2011	S6	0.58	0.610 8	0.572 7	10.546	10.605	0.844 5	0.347 1	0.226 6	0.632 7
CWMT2015	1best inhmm-system-combination	0.880 9	0.897 3	0.877 3	12.90	13.03	0.944 2	0.138 4	0.083 9	0.921 8

5 藏语资源建设

资源是计算机对藏文文本进行属性统计和机器学习材料,从 1999 年开始就有数据库和其他一些资源的建设,并先后建立一个 500 万音节字的初级平衡语料库^[85]、1.3 亿字的大型藏文语料库^[86]、一个 1.5 亿藏文字符的大型藏文平衡语料库^[87]。近几年为了机器学习,很多单位也建设了一定规模的双语平行语料、双语字典等等。

和提取提供了保障。

4.2 藏文文本检测与情感分析

研究者通过研究提出了藏文文本情感分析方法^[78]、基于 HTTP 协议的实时监控技术^[79]、基于藏文网络的舆情传播模型^[80]、基于多特征的藏文微博情感倾向性分析^[81]、深度学习算法在藏文情感分析中的应用等藏文文本监测与情感分析方法,也开发了部分测试系统。

4.3 藏文机器翻译研究

研究者于 1995 年开始研究机器翻译,并先后推出了汉藏科技机器翻译系统原型系统^[2]、“班智达汉藏公文机器翻译系统”^[82]等。随着计算机技术的发展和深度学习技术的应用,藏文机器翻译也取得了比较好的研究成果,近几年推出了面向行业的藏汉机器翻译系统及通用的藏汉机器翻译、汉藏机器翻译系统。2011 年第七届全国机器翻译研讨会 CWMT2011 中,中国科学院软件所、东北大学自然语言处理实验室、哈尔滨工业大学机器智能与翻译研究室、厦门大学、中国科学院计算技术研究所和中国科学院自动化研究所等 6 家单位参与了汉藏政府文献机器翻译评测^[83];2015 年第十一届全国机器翻译研讨会 CWMT2015 中,中国科学技术信息研究所(ISTIC)在开发集上测试了 6 个翻译系统,并用 4 个系统分别使用 1best 的翻译假设和 100best 的翻译假设进行系统融合^[84],评测指标如表 1 所示。可以看出,近 5 年来机器翻译各指标都有较大的提高。

6 应用研究

20 世纪 80 年代中期,研究者将 CCDOS.13 扩展为 TCDOS 来完成汉英文系统兼容藏文操作,并开发了可挂接在 WPS 下的藏文轻印刷系统——兰海藏文系统,藏文操作系统 ZWDOS 和 TCE 藏汉英文信息处理系统等。在信息产业部“藏文软件开发专项”的支持下,研究者研制了 Linux 藏文处理系统和 Linux/Windows 跨平台藏文办公套件^[88]、跨平台的藏文办公套件 TibetanOffice^[89]、中标普华 Of-

face 藏文版 V3.0^[89]、基于 Windows XP 藏文操作系统和基于 Linux 藏文操作系统等一系列支持藏文处理和藏文软件的系统。

研究者还提出了藏语计算机排序方案^[90]、藏语自动排序的规则^[91]、藏文排序的数学模型^[92]、引入藏文字符集扩展集 A 中的 UNICODE 编码为排序码^[93]、藏文字笔画编码排序^[94]等思想,开发了藏文拼写检查系统、藏文文本自动校对、藏文转码软件、藏文拉丁转写、藏文电子词典等。现在藏文不仅应用于 PC 机等计算上,还广泛应用在手持移动设备中^[95],得到 IOS、Android 等系统的支持。

7 展望

回顾藏文信息处理的历史,在藏文的字处理上取得了很好的成绩,也涉足到藏文词、句、段、篇的研究,但与其它语言文字的信息处理相比,目前藏文信息处理还存在许多不足之处,具体来说有以下 3 点:(1)藏文信息处理的方法和评价标准都是从中文、英语等的处理方法中借鉴过来的,对藏语本身的特点和规律的研究欠缺;(2)统计机器学习方法逐渐兴起并成为当前自然语言处理领域的主流方法,但藏文信息处理目前的很多研究还只是基于“规则”;(3)近年来基于神经网络的深度学习的方法备受推崇,但藏文信息方面不仅缺少“资源”,也缺少基于深度学习取得的成果。因此,若能在这些领域进行深入研究,将有利于藏文信息处理的长足发展。

参考文献:

- [1] 赵晨星. 藏文计算机键盘的国家标准的研究[C]//第三届中文信息处理国际会议论文集. 北京:中国中文信息学会,1992.
ZHAO C X. Research on the national standard of Tibetan computer keyboard[C]//Proceedings 1992 International Conference on Chinese Information Processing. Beijing:Chinese Information Processing Society of China,1992.
- [2] 陈玉忠,俞士汶. 藏文信息处理技术的研究现状与展望[J]. 中国藏学,2003(4):97-107.
CHEN Y Z, YU S W. Tibetan information processing: Past, present, and future[J]. China Tibetology, 2003(4):97-107.
- [3] 江狄,董颖红. 藏文信息处理属性统计研究[J]. 中文信息学报,1995,9(2):37-44.
JIANG D, DONG Y H. Research on property Tibetan characters as information processing [J]. Journal of Chinese Information Processing, 1995, 9(2):37-44.
- [4] 江狄,董颖红. 藏字叠加结构线性处理统计分析[J]. 中文信息,1994(4):44-46,54.
JIANG D, DONG Y H. Statistical analysis of linear processing of superposition structure of Tibetan characters[J]. Chinese Information Processing, 1994(4):44-46,54.
- [5] 江狄. 书面藏语的熵值及相关问题[C]//黄昌宁. 1998 中文信息处理国际会议论文集. 北京:清华大学出版社,1998.
JIANG D. An entropy value of classical Tibetan language and some other questions[C]//HUANG C N. International conference on Chinese information processing in 1998. Beijing: Tsinghua University Press, 1998.
- [6] 严海林,江获. 藏文大藏经信息熵研究[C]//中国少数民族多文种信息处理研究与进展. 呼和浩特,2004.
YAN H L, JIANG D. Study on information entropy of Tibetan Tripitaka[C]//Research and Development of Information Processing of China Ethnic Multilingual. Hohhot, 2004.
- [7] 王维兰,陈万军. 藏文字丁、音节频度及其信息熵[J]. 术语标准化与信息技术,2004(2):27-31.
WANG W L, CHEN W J. The frequency and information entropy of Tibetan character and syllable[J]. Terminology Standardization and Information Technology, 2004(2):27-31.
- [8] 卢亚军,马少平,张敏,等. 基于大型藏文语料库的藏文字符、部件、音节、词汇频度与通用度统计及其应用研究[J]. 西北民族大学学报:自然科学版,2003,24(2):32-42.
LU Y J, MA S P, ZHANG M, et al. Research of calculations of Tibetan characters, pieces, syllables, vocabulary and universal frequency and its applications[J]. Journal of Northwest University for Nationalities: Natural Science, 2003, 24(2):32-42.
- [9] 高定国,龚育昌. 现代藏字全集的属性统计研究[J]. 中文信息学报,2005,19(1):71-75.
GAO D G, GONG Y C. A statistically study on the qualities of all modern Tibetan character set[J]. Journal of Chinese Information Processing, 2005, 19(1):71-75.
- [10] 卢亚军,罗广. 藏文词汇通用度统计研究[J]. 图书与情报,2006(3):74-77.
LU Y J, LUO G. Study on the generality of Tibetan vocabulary with statistics[J]. Library & Information, 2006(3):74-77.
- [11] 才智杰,才让卓玛. 藏文字符的向量模型及构件特征分析[J]. 中文信息学报,2016,30(2):202-206.

- CAI Z J, CAI RANG Z M. Vector space models and component features analysis of Tibetan characters [J]. Journal of Chinese Information Processing, 2016, 30(2): 202-206.
- [12] 西藏自治区语文工作指导委员会办公室. 信息技术藏文编码字符集(基本集) 键盘字母数字区的布局: GB/T 17543—1998 [S]. 北京: 中国标准出版社, 1999.
Office of the Language Work Steering Committee of the Tibet Autonomous Region. Information technology—Keyboard layout of the alphanumeric zone for Tibetan coded character set (basic set): GB/T 17543—1998 [S]. Beijing: China Standard Press, 1999.
- [13] 彭寿全, 黄可, 万国根, 等. 外挂式藏汉英混合处理系统[J]. 中文信息学报, 1993, 8(2): 47-53.
PENG S Q, HUAGN K, WAN G G, et al. The Tibetan, Chinese and outside—Hanging system[J]. Journal of Chinese Information Processing, 1993, 8(2): 47-53.
- [14] 罗圣仪, 江虹, 张志芳. 藏文轻印刷系统[J]. 计算机世界, 1993(7): 73.
LUO S Y, JIANG H, ZHANG Z F. Tibetan light printing system[J]. PC World, 1993(7): 73.
- [15] 罗圣仪, 刘英杰, 刘璐. 一种藏汉西文混合输入和编辑的藏文处理系统[C]//1987 中文信息处理国际会议论文集. 北京: 中国中文信息学会, 1987.
LUO S Y, LIU Y J, LIU L. A Tibetan processing system with input and edit of Tibetan, Chinese and English mixed[C]//International conference on Chinese information processing in 1987. Beijing: Chinese Information Processing Society of China, 1987.
- [16] 王维兰. 现代藏文识别[C]//黄昌宁. 1998 中文信息处理国际会议论文集, 北京: 清华大学出版社, 1998.
WANG W L. Modern Tibetan recognition [C]//HUANG C N. International conference on Chinese information processing in 1998. Beijing: Tsinghua University Press, 1998.
- [17] 王维兰. 藏文基本字符识别算法研究[J]. 西北民族学院学报: 自然科学版, 1999, 20(3): 20-23, 51.
WANG W L. Algorithm study on feature extracting of Tibetan character recognition[J]. Journal of Northwest University for Nationalities: Natural Science, 1999, 20(3): 20-23, 51.
- [18] 王浩军, 赵南元, 邓钢铁. 一种现代藏文笔段提取算法[J]. 中文信息学报, 2001, 15(4): 41-46, 52.
WANG H J, ZHAO N Y, DENG G Y. A stroke segment extraction algorithm for Tibetan character recognition[J]. Journal of Chinese Information Processing, 2001, 15(4): 41-46, 52.
- [19] 普次仁. 多种印刷字体藏文字符的特征提取方法研究[J]. 西藏大学学报: 自然科学版, 2008, 23(1): 25-28.
PU T R. Research on extracting the character of Tibetan symbol at different printing shape letters[J]. Journal of Tibet University: Natural Science Edition, 2008, 23(1): 25-28.
- [20] 王维兰, 丁晓青, 戴玉刚. 藏文识别后处理研究[J]. 术语标准化与信息技术, 2002(2): 30-34.
WANG W L, DING X Q, DAI Y G. Study on post-processing of Tibet character recognition[J]. Terminology Standardization and Information Technology, 2002(2): 30-34.
- [21] 王维兰, 丁晓青, 祁坤钰. 藏文识别中相似字丁的区别研究[J]. 中文信息学报, 2002, 16(4): 60-65.
WANG W L, DING X Q, QI K Y. Study on similitude characters in Tibetan character recognition[J]. Journal of Chinese Information Processing, 2002, 16(4): 60-65.
- [22] 柳洪轶, 王维兰. 联机手写藏文识别中字丁规范化处理[J]. 计算机应用研究, 2006, 23(9): 179-181.
LIU H Y, WANG W L. Nonlinear shape normalization methods for on-line recognition of handwritten Tibetan characters[J]. Application Research of Computers, 2006, 23(9): 179-181.
- [23] 中国西藏信息中心. 多字体印刷藏文(混排汉英)文档识别系统成果简介[EB/OL]. [2009-07-07]. <http://www.tibetinform.com.cn/news/2003-11-19/N312003111995149.htm>.
Tibet Information Center, China. A brief introduction to the document recognition system of multi fonts printed Tibetan (mixed Chinese English)[EB/OL]. [2009-07-07]. <http://www.tibetinform.com.cn/news/2003-11-19/N312003111995149.htm>.
- [24] 杨阳蕊, 李永宏, 于洪志. 藏语安多方言的音联结构及统计分析[J]. 西北民族大学学报: 自然科学版, 2008, 29(2): 11-16.
YANG Y R, LI Y H, YU H Z. Juncture structure of and statistical analysis on Tibetan ando[J]. Journal of Northwest University for Nationalities: Natural Science, 2008, 29(2): 11-16.
- [25] 李永宏, 于洪志. 安多藏语语音合成语料库的设计[J]. 西北民族学院学报: 自然科学版, 2006, 27(1): 36-39.
LI Y H, YU H Z. Researches of voice database for Tibetan speech synthesis[J]. Journal of Northwest University for Nationalities: Natural Science, 2006, 27(1): 36-39.

- [26] 马明,刘华.基于隐马尔可夫过程的藏语语音预测初探[J].西北民族大学学报:自然科学版,2008,29(2):17-18.
MA M, LIU H. A preliminary study of Tibetan speech prediction based on Hidden Markov process [J]. Journal of Northwest University for Nationalities: Natural Science, 2008, 29(2): 17-18.
- [27] 李积逊,范武英.藏语语音的时域分析[J].福建电脑,2007(5):56-57.
LI J X, FAN W Y. The time domain analysis of Tibetan phonetics [J]. Fujian Computer, 2007(5): 56-57.
- [28] 李永宏,孔江平,于洪志.藏语文-音自动规则转换及其实现[J].清华大学学报:自然科学版,2008(S1):621-626.
LI Y H, KONG J P, YU H Z. The auto-transformation of Tibetan text to IPA by rules and its implementation [J]. Journal of Tsinghua University: Science and Technology, 2008(S1): 621-626.
- [29] 彭寿全,黄可,张义刚.藏文综合编码方案的研究与实现[J].中文信息学报,1996,10(4):32-39.
PENG S Q, HUANG K, ZHANG Y G. A research and implement of Tibetan comprehensive coding method [J]. Journal of Chinese Information Processing, 1996, 10(4): 32-39.
- [30] 于洪志.藏文内码扩展体系[J].中文信息学报,1999,13(1):50-58.
YU H Z. An extended Tibetan coded character system [J]. Journal of Chinese Information Processing, 1999, 13(1): 50-58.
- [31] 于洪志.计算机藏文编码性能指标分析[J].西北民族学院学报:自然科学版,1997,18(2):15-20.
YU H Z. Function indexes of Tibetan code in computer [J]. Journal of Northwest University for Nationalities: Natural Science, 1997, 18(2): 15-20.
- [32] 于洪志.藏文编码字符集构件集[J].西北民族学院学报:自然科学版,1998,19(1):11-16.
YU H Z. Code system of full Tibetan [J]. Journal of Northwest University for Nationalities: Natural Science, 1998, 19(1): 11-16.
- [33] 尼玛扎西,毛永刚,于洪志,等.信息技术 信息交换用藏文编码字符集 基本集:GB 16959—1997[G]//中国标准出版社.中国国家标准汇编 239.北京:中国标准出版社,1997.
NYIMA TRASHI, MAO Y G, YU H Z, et al. Information technology—Tibetan coded character sets for information interchange—Basic set; GB 16959—1997 [G]//China Standard Press. China national standard compilation 239. Beijing: China Standard Press, 1997.
- [34] 于洪志,熊涛,尼玛扎西,等.信息技术 藏文编号码字符集(基本集) 24×48点阵字型 第1部分:白体:GB/T 16960.1—1997[G]//中国标准出版社.中国国家标准汇编 239.北京:中国标准出版社,1997.
YU H Z, XIONG T, NYIMA TRASHI, et al. Information technology—Tibetan coded character set (basic set)—24×48 dots matrix font—Part 1; Bai Ti; GB/T 16960.1—1997 [G]//China Standard Press. China national standard compilation 239. Beijing: China Standard Press, 1997.
- [35] 全国信息技术标准化技术委员会.少数民族信息标准 [EB/OL]. [2007-07-07]. <http://www.nits.gov.cn/sc1002/tibetStandardAll.asp>.
National Technical Committee on Information Technology Standardization. Information standards for minority nationalities [EB/OL]. [2007-07-07]. <http://www.nits.gov.cn/sc1002/tibetStandardAll.asp>.
- [36] 高定国,龚育昌.设计 OpenType 藏文白体字库的技术研究[C]//第十届全国少数民族语言文字信息处理学术研讨会论文集.西宁:不详,2005.
GAO D G, GONG Y C. Research on design techniques of OpenType Tibetan BaiTi font [C]//Tenth academic symposium on information processing of national minority languages and characters. Xining: [s. n.], 2005.
- [37] 扎西次仁.一个人机互助的藏文分词和词登录系统的设计[C]//中国少数民族语言文字现代化文集.北京:民族出版社,1999.
TASHI TSERING. The design of a personal computer aided Tibetan word segmentation and word login system [C]//Modern anthology of Chinese minority language and language. Beijing: The Ethnic Publishing House, 1999.
- [38] 陈玉忠,李保利,俞士汶,等.基于格助词和接续特征的藏文自动分词方案[J].语言文字应用,2003(1):75-82.
CHEN Y Z, LI B L, YU S W, et al. A Tibetan segmentation scheme based on case-auxiliary words and continuous features [J]. Applied Linguistics, 2003(1): 75-82.
- [39] 才智杰.藏文自动分词系统中紧缩词的识别[J].中文信息学报,2009,23(1):35-37,43.
CAI Z J. Identification of abbreviated word in Tibetan word segmentation [J]. Journal of Chinese Information Processing, 2009, 23(1): 35-37, 43.
- [40] 李亚超,加羊吉,江静,等.融合无监督特征的藏文分词方法研究[J].中文信息学报,2017,31(2):71-75,

- 85.
- LI Y C, JIA Y J, JIANG J, et al. Study on fusion of unsupervised features for Tibetan word segmentation [J]. Journal of Chinese Information Processing, 2017, 31(2):71-75,85.
- [41] 刘汇丹, 诺明花, 赵维纳, 等. SegT: 一个实用的藏文分词系统[J]. 中文信息学报, 2012, 26(1):97-103.
- LIU H D, NUO M H, ZHAO W N, et al. SegT: A practical Tibetan word segmentation system[J]. Journal of Chinese Information Processing, 2012, 26(1):97-103.
- [42] 才智杰, 班智达藏文自动分词系统的设计与实现[J]. 青海师范大学民族师范学院学报, 2010, 21(2):75-77.
- CAI Z J. Design and implementation of Banzhida Tibetan word segmentation system[J]. Journal of Minorities Teachers College of Qinghai Teachers University, 2010, 21(2):75-77.
- [43] 史晓东, 卢亚军. 央金藏文分词系统[J]. 中文信息学报, 2011, 25(4):54-56.
- SHI X D, LU Y J. A Tibetan segmentation system—Yangjin[J]. Journal of Chinese Information Processing, 2011, 25(4):54-56.
- [44] 中国电子技术标准化研究院, 西藏大学, 等. 信息处理用藏文分词规范(送审稿)[Z]. 2017.
- China Institute of Electronic Technology Standardization, Tibet University, et al. The specification for Tibetan participle in information processing (manuscript)[Z]. 2017.
- [45] MLWS2017 组委会. 民族语言自动分词评测 MLWS2017[EB/OL]. [2017-05-09]. <http://nmlr.muc.edu.cn/huiyixinxi/2017/07-19/305.html>.
- MLWS2017 Organizing Committee. Evaluation of minority language word segmentation (MLWS2017) [EB/OL]. [2017-05-09]. <http://nmlr.muc.edu.cn/huiyixinxi/2017/07-19/305.html>.
- [46] 高定国. MLWS2017 中藏文分词评测的技术报告 [J]. 高原科学研究, 2017, 1(1):89-97.
- GAO D G. Technical report on evaluation of Tibetan words in MLWS 2017[J]. Plateau Science Research, 2017, 1(1):89-97.
- [47] 陈玉忠. 信息处理用现代藏语词语的分类方案[C]//第十届全国少数民族语言文字处理学术研讨会论文集. 西宁: 不详, 2005.
- CHEN Y Z. Contemporary Tibetan language word classification scheme for information processing[C]//The 10th academic symposium on information processing of national minority languages and characters. Xining:[s. n.], 2005.
- [48] 多拉, 扎西加, 欧珠, 等. 信息处理用藏文词类及标记集规范[C]//第十一届全国少数民族语言文字处理学术研讨会论文集, 西双版纳: 不详, 2007.
- DOLHA, ZHAXIJA, OUZHU, et al. The parts-of-speech and tagging set standards of Tibetan information process[C]//The 11th academic symposium on information processing of national minority languages and characters. Xishuangbanna:[s. n.], 2007.
- [49] 扎西多杰, 安见才让. 基于 HMM 藏文词性标注的研究与实现[J]. 计算机光盘软件与应用, 2012(12):100-101.
- TASHI D J, ANJIAN C R. Research and implementation of Tibetan word based on HMM [J]. Computer CD Software and Application, 2012(12):100-101.
- [50] 刘飞飞, 王志娟. 基于层次特征的藏文人名识别研究[J/OL]. 计算机应用研究, 2017, 35(9). (2017-08-28) [2017-11-20]. <http://www.aocmag.com/article/02-2018-09-033.html>.
- LIU F F, WANG Z J. Research on recognition of Tibetan names based on hierarchical features[J/OL]. Application Research of Computers, 2017, 35(9). (2017-08-28) [2017-11-20]. <http://www.aocmag.com/article/02-2018-09-033.html>.
- [51] 珠杰, 李天瑞, 刘胜久. 基于条件随机场的藏文人名识别技术研究[J]. 南京大学学报: 自然科学版, 2016, 52(2):289-299.
- ZHU J, LI T R, LIU S J. Research on Tibetan name recognition technology under CRF [J]. Journal of Nanjing University: Natural Science, 2016, 52(2):289-299.
- [52] 窦嵘, 加羊吉, 黄伟. 统计与规则相结合的藏文人名自动识别研究[J]. 长春工程学院学报: 自然科学版, 2010(2):113-115.
- DOU R, JIA Y J, HUANG W. Automatic recognition of Tibetan name with the combination of statistics and regular [J]. Journal of Changchun Institute of Technology: Natural Science Edition, 2010(2):113-115.
- [53] 加羊吉, 李亚超, 宗成庆, 等. 最大熵和条件随机场模型相融合的藏文人名识别[J]. 中文信息学报, 2014, 28(1):107-112.
- JIA Y J, LI Y C, ZONG C Q, et al. A hybrid approach to Tibetan person name identification by maximum entropy model and conditional random fields[J]. Journal of Chinese Information Processing, 2014, 28(1):107-112.
- [54] 高定国, 扎西加, 赵栋材. 计算机识别藏语虚词的方法

- 研究[J]. 中文信息学报, 2014, 28(1): 113-117.
- GAO D G, TASHIGYAL, ZHAO D C. Research on automatic indentfication of Tibetan function word[J]. Journal of Chinese Information Processing, 2014, 28(1): 113-117.
- [55] 朋毛吉, 安见才让. 藏文从格助词“ལམས”的识别算法研究[J]. 信息通信, 2017(9): 97-98.
- PENG M J, ANJIAN C R. Research on recognition algorithm of Tibetan language from the lattice “ལམས” [J]. Information & Communications, 2017(9): 97-98.
- [56] 孙萌, 华却才让, 刘凯, 等. 藏文数词识别与翻译[J]. 北京大学学报: 自然科学版, 2013(1): 75-80.
- SUN M, HUA Q C R, LIU K, et al. Tibetan number identification and translation [J]. Acta Scientiarum Naturalium Universitatis Pekinensis, 2013(1): 75-80.
- [57] 索南尖措. 藏语动词的黏着性及屈折性变化研究[J]. 西藏大学学报: 自然科学版, 2013(4): 70-75.
- SUONAN J C. Study on the changes in adhesive and inflectional of Tibetan verbs[J]. Journal of Tibet University: Natural Science Edition, 2013(4): 70-75.
- [58] 仁青卓么, 祁坤钰. 藏语动词语法属性计量研究[J]. 西北民族大学学报: 自然科学版, 2014, 35(2): 37-42.
- REN C Z M, QI K Y. Study on the grammatical attribute of Tibetan verbs[J]. Journal of Northwest University for Nationalities: Natural Science, 2014, 35(2): 37-42.
- [59] 道杰本. 藏文副词的语法功能属性库研究[D]. 兰州: 西北民族大学, 2012.
- DAO J B. The gramartical functions of the Tibetan adberbs[D]. Lanzhou: Northwest University for Nationalities, 2012.
- [60] 祁坤钰. 基于语料库的藏语名词分类与统计研究[J]. 西北民族大学学报: 自然科学版, 2012, 33(3): 44-49.
- QI K Y. Study on the classification and statistics of Tibetan nouns based on corpus[J]. Journal of Northwest University for Nationalities: Natural Science, 2012, 33(3): 44-49.
- [61] 德吉梅朵. 面向藏语语料库建设的数量词的分类研究[J]. 剑南文学: 经典教苑, 2013(1): 228-229.
- DEJI M D. Study on the classification of quantitative words for the construction of Tibetan corpus[J]. Jian Nan Literature, 2013(1): 228-229.
- [62] 索南尖措, 高定国. 信息处理用藏文动词的分类[J]. 西藏大学学报: 自然科学版, 2011, 26(6): 113-117.
- SUONAN J C, GAO D G. Classification of Tibetan language verbs in information processing[J]. Journal of Tibet University: Natural Science Edition, 2011, 26(6): 113-117.
- [63] 张同玲, 多杰卓玛. 藻饰词语义网络的构建研究[J]. 电脑开发与应用, 2011, 24(7): 25-27.
- ZHANG T L, DUOJIE Z M. The constructional research on semantic network of Zaoshi words [J]. Computer Development & Applications, 2011, 24(7): 25-27.
- [64] 徐涛, 加羊吉, 于洪志. 统计与规则相结合的藏文句子自动断句方法[J]. 云南大学学报: 自然科学版, 2012, 34(6): 653-657.
- XU T, JIA Y J, YU H Z. An approach of automatic segmentation for Tibetan sentence based on rules and statistics[J]. Journal of Yunnan University: Natural Sciences, 2012, 34(6): 653-657.
- [65] 李响, 才藏太, 姜文斌, 等. 最大熵和规则相结合的藏文句子边界识别方法[J]. 中文信息学报, 2011, 25(4): 39-44.
- LI X, CAI Z T, JIANG W B, et al. A maximum entropy and rules approach to identifying Tibetan sentence boundaries[J]. Journal of Chinese Information Processing, 2011, 25(4): 39-44.
- [66] 张同玲, 祁坤钰. 浅谈藏语单句的类型[J]. 科技信息, 2011(17): 191, 199.
- ZHANG T L, QI K Y. The type of Tibetan sentences [J]. Science & Technology Information, 2011(17): 191, 199.
- [67] 邱莉榕, 翁彧, 赵小兵. 藏文语义本体中的上下位关系模式匹配算法[J]. 中文信息学报, 2011, 25(4): 45-49.
- QIU L R, WENG Y, ZHAO X B. Acquisition method of hyponymy concepts based on patterns in Tibetan semantic ontology[J]. Journal of Chinese Information Processing, 2011, 25(4): 45-49.
- [68] 多杰卓玛. 基于信息处理的藏文框架语义关系研究[J]. 西北民族大学学报: 自然科学版, 2010, 31(1): 16-19.
- DUOJIE Z M. Analyses of semantic relations between Tibetan framework of semantic knowledge[J]. Journal of Northwest University for Nationalities: Natural Science, 2010, 31(1): 16-19.
- [69] 祁坤钰. 基于依存关系的藏文语义角色标注研究[J]. 西北民族大学学报: 哲学社会科学版, 2014(1): 139-143.
- QI K Y. Research on semantic role tagging of Tibetan language based on dependency relationship [J]. Journal of Northwest University for Nationalities: Philosophy and Social Science, 2014(1): 139-143.
- [70] 丁海兰, 祁坤钰. 基于依存句法的藏文属格结构统计研究[J]. 西北民族大学学报: 自然科学版, 2016, 37

- (2):32-36.
- DING H L, QI K Y. Study on Tibetan genitive statistics based on dependency parsing [J]. Journal of Northwest University for Nationalities: Natural Science, 2016, 37(2):32-36.
- [71] 蒋明原, 孔令德. 基于 Lucene 的藏文信息采集及检索系统研究[J]. 电脑开发与应用, 2011, 24(2):34-37.
- JIANG M Y, KONG L D. A study on Lucene-based Tibetan information collection and retrieval system [J]. Computer Development & Applications, 2011, 24(2):34-37.
- [72] 扎西拉旦, 安见才让. 基于 XML 文档的藏文网页倒排索引的研究与实现[J]. 软件工程, 2017, 20(6):12-14.
- ZHAXI L D, ANJIAN C R. Research and implementation of inverted index of Tibetan web pages based on XML documents[J]. Software Engineer, 2017, 20(6):12-14.
- [73] 普措才仁, 蔡光波. 基于奇异值分解的藏文 Web 不良信息检索算法研究[J]. 西北民族大学学报: 自然科学版, 2015, 36(4):23-27.
- PUCUO C R, CAI G B. Research on retrieval algorithm of Tibetan Web bad information based on singular value decomposition[J]. Journal of Northwest University for Nationalities: Natural Science, 2015, 36(4):23-27.
- [74] 才华. 向量空间模型在藏文文本信息检索中的应用[J]. 西藏大学学报: 自然科学版, 2011, 26(6):118-121.
- CAI H. Application of vector space model in Tibetan text information retrieval[J]. Journal of Tibet University: Natural Science Edition, 2011, 26(6):118-121.
- [75] 高红梅, 魏西峰, 王崧华, 等. 语义词库关联的藏文 Web 语义检索系统研究与实现[J]. 西藏大学学报: 自然科学版, 2015, 30(5):90-95.
- GAO H M, WEI X F, WANG S H, et al. Research on searching Tibetan Web' semantic related to semantic association thesaurus[J]. Journal of Tibet University: Natural Science Edition, 2015, 30(5):90-95.
- [76] 徐涛, 蓝传琦. 基于卡方统计量的藏文新闻网页关键词提取方法[J]. 电脑知识与技术, 2017, 13(26):171-173.
- XU T, LAN C Q. Keyword extraction method of Tibetan news web pages based on Chi square statistics [J]. Computer Knowledge and Technology, 2017, 13(26):171-173.
- [77] 洛松求培, 安见才让. 基于文本密度的藏文网页正文提取方法[J]. 计算机时代, 2017(8):46-47.
- LUOSONG Q P, ANJIAN C R. A content extraction method of Tibetan web based on text density [J]. Computer Era, 2017(8):46-47.
- [78] 李苗苗. 藏文文本情感分析方法研究[D]. 拉萨: 西藏大学, 2017.
- LI M M. Study on the method of emotion analysis in Tibetan text[D]. Lasa: Tibet University, 2017.
- [79] 余长春. 基于 HTTP 协议面向藏文文本的实时监测技术研究[D]. 拉萨: 西藏大学, 2016.
- YU C C. Research on real-time monitoring technology for Tibetan text based on HTTP protocol[D]. Lasa: Tibet University, 2016.
- [80] 邓竞伟, 邓凯英, 李永生, 等. 基于藏文网络的舆情传播模型[J]. 计算机系统应用, 2013(3):209-211.
- DENG J W, DENG K Y, LI Y S, et al. Opinion spreading models on Tibetan networks[J]. Computer Systems & Applications, 2013(3):209-211.
- [81] 江涛, 袁斌, 于洪志, 等. 基于多特征的藏文微博情感倾向性分析[J]. 中文信息学报, 2017, 31(3):163-169.
- JIANG T, YUAN B, YU H Z, et al. Multi-feature based sentiment analysis of Tibetan microblogs [J]. Journal of Chinese Information Processing, 2017, 31(3):163-169.
- [82] 才藏太, 华关加. 班智达汉藏公文翻译系统中基于二分法的句法分析方法研究[J]. 中文信息学报, 2005, 19(6):7-12.
- CAI Z T, HUA G J. Research of Banzhida Chinese-Tibetan document translation system based on the dichotomy of syntax analysis[J]. Journal of Chinese Information Processing, 2005, 19(6):7-12.
- [83] 熊伟, 王震, 于新, 等. ISCAS 机器翻译系统与评测技术报告[C]//机器翻译研究进展——第七届全国机器翻译研讨会论文集. 厦门: 不详, 2011.
- XIONG W, WANG Z, YU X, et al. Machine translation system and evaluation technical report for ISCAS [C]//Progress in machine translation research—The seventh national machine translation symposium. Xiamen: [s. n.], 2011.
- [84] 何彦青, 孟令恩, 丁亮, 等. 2015 全国机器翻译研讨会 ISTIC 评测报告[J]. 情报工程, 2016, 2(5):58-66.
- HE Y Q, MENG L E, DING L, et al. Evaluation technical report of ISTIC from CWMT' 2015 [J]. Technology Intelligence Engineering, 2016, 2(5):58-66.
- [85] 周季文, 江荻. 藏语计算机统计用语料抽样文本筛选 [C]//中国少数民族语言文字现代化文集. 北京: 民族出版社, 1999.
- ZHOU J W, JIANG D. Selection of text sampling text

- for Tibetan language computer statistics[C]//Modern anthology of Chinese minority language and language. Beijing: The Ethnic Publishing House, 1999.
- [86] 卢亚军, 马少平, 张敏, 等. 基于大型藏文语料库的藏文字符、部件、音节、词汇频度与通用度统计及其应用研究[J]. 西北民族大学学报: 自然科学版, 2003, 24(2): 32-42.
- LU Y J, MA S P, ZHANG M, et al. Researches of calculations of Tibetan characters, pieces, syllables, vocabulary and universal frequency and its applications [J]. Journal of Northwest University for Nationalities: Natural Science, 2003, 24(2): 32-42.
- [87] 西藏大学. 教育部、国家语委民族语言文字规范标准建设与信息化项目“大型藏文基础语料库建设”(MZ115-039)成果简介[R]. [S. l.: s. n.], 2013.
- Tibet University. The achievement introduction of ministry of education, the National Language Committee of national language standards construction and informatization project—“The large basic Tibetan corpus construction” (MZ115-039) [R]. [S. l.: s. n.], 2013.
- [88] 西藏自治区藏语文工作委员会办公室.《基于 LINUX 的跨平台藏文信息处理系统》在京通过验收[EB/OL]. [2007-07-07]. <http://xz6.2000y.net/mb/1/ReadNews.asp?NewsID=273429>.
- Office of the Tibetan Language Working Committee of the Tibet Autonomous Region. “LINUX based cross platform Tibetan information processing system” is accepted in Beijing[EB/OL]. [2007-07-07]. <http://xz6.2000y.net/mb/1/ReadNews.asp?NewsID=273429>.
- [89] 信息产业部电子信息产品管理司. 藏文软件开发专项介绍[J]. 信息技术与标准化, 2007(8): 2-3.
- Department of Electronic Information Products Management Department of the Ministry of Information Industry. Special introduction to the development of Tibetan software [J]. Information Technology & Standardization, 2007(8): 2-3.
- [90] 江荻. 藏语文本信息处理的历程与进展[C]//中文信息处理前沿进展——中国中文信息学会二十五周年学术会议论文集. 北京: 不详, 2006.
- JIANG D. The process and progress of text information processing in Tibetan language [C]//The advancement of Chinese information processing frontiers—The 25th anniversary academic conference of Chinese information society. Beijing: [s. n.], 2006.
- [91] 扎西次仁. 藏文的排序规则及其计算机自动排序的实现[J]. 中国藏学, 1999(4): 128-135.
- TASHI TSERING. The sorting rules of Tibetan language and the realization of computer automatic sorting[J]. China Tibetology, 1999(4): 128-135.
- [92] 江荻, 康才峻. 书面藏语排序的数学模型及算法[J]. 计算机学报, 2004, 27(4): 524-529.
- JIANG D, KANG C J. The sorting mathematical model and algorithm of written Tibetan language[J]. Chinese Journal of Computers, 2004, 27(4): 524-529.
- [93] 黄鹤鸣, 赵晨星. 引入排序码实现藏文字符的排序[J]. 计算机技术与发展, 2008, 18(10): 68-70, 74.
- HUANG H M, ZHAO C X. Introducing sort code to realize Tibetan characters' sort[J]. Computer Technology and Development, 2008, 18(10): 68-70, 74.
- [94] 刘城, 黄鹤鸣, 李继文. 藏文字笔画编码排序的设想[J]. 计算机系统应用, 2013(5): 18-20.
- LIU C, HUANG H M, LI J W. Tibetan strokes of computer codes sorting[J]. Computer Systems & Applications, 2013(5): 18-20.
- [95] 尼玛扎西, 李志蜀, 群诺, 等. 一种在移动电话上实现藏文处理的方法[J]. 四川大学学报: 工程科学版, 2009, 41(1): 162-167.
- NYIMA TRASHI, LI Z S, QUN N, et al. A method of implementing Tibetan processing for mobile phone [J]. Journal of Sichuan University: Engineering Science Edition, 2009, 41(1): 162-167.

(责任编辑: 陆雁)