

傣文信息技术研究进展*

Review of the Development of Information Technology in Tai Script

殷建民**

YIN Jianmin

(潍坊北大青鸟华光照排有限公司, 山东潍坊 261205)

(Weifang Beida Jade Bird Huaguang Information Technology Co., Ltd., Weifang, Shandong, 261205, China)

摘要:本文介绍了 20 世纪 90 年代以来傣文(含德宏傣文和西双版纳傣文)信息处理技术的发展,涉及傣文编码字符集、傣文字型、傣文键盘布局等基础标准的制订和傣文电子出版系统、傣文网站、傣文移动客户端、傣文识别系统、傣文分词系统、傣文语料库、傣文电子词典、傣文机器翻译等应用系统的开发。

关键词:傣文 德宏傣文 西双版纳傣文 信息技术 标准

中图分类号: TP391 **文献标识码:** A **文章编号:** 1002-7378(2018)01-0012-06

Abstract: This paper introduces the development of information processing technology in Tai script (Tai le and Tai Lue) since the 1990s, involving the formulation of encoding character set, font, keyboard layout standards and the development of electronic publishing systems, website, mobile client, OCR, word segmentation system, corpus, electronic dictionary, machine Translation system and so on.

Key words: Tai script, Tai Le, Tai Lue, information technology, standard

0 引言

傣语属汉藏语系壮侗语族壮傣语支,有多种不同形式的文字,主要包括德宏傣文(Tai Le)、西双版纳新傣文(New Tai Lue)和西双版纳老傣文(Old

Tai Lue, Tai Tham)。

傣文信息技术研究始于 20 世纪 90 年代初,主要包括傣文信息技术标准的制订与傣文信息化软件的研发。

1 傣文信息技术标准制订

1994 年,中国曾向 ISO/IEC JTC1/SC2/WG2 提出西双版纳老傣文编码标准提案(N1013 Motion on the coding of the Old Xishuang Banna Dai Writing Entering into BMP of ISO/IEC 10646)。2000 年,云南省民族语言委员会办公室组织专家修订完善了德宏傣文和西双版纳傣文的编码方案,并向国际标准化组织提交了关于德宏傣文和西双版纳新傣文编码方案的国际提案。

2001 年,德宏傣文编码字符集国际标准获得通过^[1],共收入 35 个字符,编码空间为 U1950~

收稿日期: 2018-01-06

作者简介: 殷建民(1962—),男,研究员,主要从事中文信息处理研究, E-mail: 110154413@qq.com。

* 国家新闻出版重大科技工程项目(XWCB-GDGC-FHCB/16, GAPP-ZDKJ-ZK/19),中央补助地方文化体育与传媒事业发展专项资金项目(西财教字[2009]35号,110号,184号),教育部、国家语委民族语言文字规范标准建设及信息化科研项目(MZ115-004, MZ115-067)和山东省信息产业发展专项资金项目(2008R04020)资助。

** 通信作者。

197F^[2]。

2005年,西双版纳新傣文编码字符集方案获得通过^[3],被收入ISO/IEC 10646-2003的第一补充集(Amendment 1)^[4],共收入80个字符(后补充为83个字符),编码空间为U1980~19DF。

2008年,西双版纳老傣文编码字符集方案获得通过^[5],被收入ISO/IEC 10646-2003的第五补充集(Amendment 5)^[6],共收入127个字符,编码空间为U1A20~1AAF。

德宏傣文编码字符集国际标准中包含了3个“可组合”的元音字母(U、OO、AUE),排版时需要对这3个元音字母与其后的韵尾字母或声调符号进行“动态组合”。具体处理方法如下:

(1)元音字母U、OO后面带有韵尾字母-P、-T、-K、-M、-N、-NG、-TSH时,后面的韵尾字母应“回退”并与前面的元音字母组合成一个整字。

(2)元音字母U、OO、AUE后面带有声调符号时,后面的声调符号应“回退”并与前面的元音字母组合成一个整字。

由于韵尾字母与声调符号的回退组合与前面的元音字母有关,仅仅使用TrueType字库中的“回退”属性无法解决这一问题,应使用OpenType字库中的Ligature功能。

西双版纳新傣文的编码字符集与显现字符集是一致的,而老傣文的编码字符集与显现字符集相差较大,主要表现在两个方面:

(1)新老傣文的许多辅音字母都可以作为“韵尾”使用,但辅音字母的独立形式与其韵尾形式在显现时有明显的差别。新傣文中可以作为韵尾使用的辅音字母不是太多,在其编码字符集方案中,辅音字母的独立形式与其韵尾形式是分别编码的。而老傣文中可以作为韵尾使用的辅音字母非常多,为节省编码空间,辅音字母的韵尾形式没有单独编码,而是引入了一个控制符号SAKOT(U1A60)。当SAKOT之后紧跟一个辅音字母时,显现时需将SAKOT之后的辅音字母变为其韵尾形式,并且与SAKOT之前的其他字母组合成一个整字。

(2)新老傣文中都有一些组合字,新傣文中的组合字不多,在其编码字符集方案中,组合字单独编码。而老傣文中的组合字非常多,编码方案中没有为组合字单独编码,实现时需要采用“预组合”与“动态组合”技术^[7]。

2006年8月,经国家标准化委员会批准,

全国信息技术标准化技术委员会成立傣文信息技术标准国家工作组。该工作组成立以来,先后制订了12项傣文字型与键盘布局国家标准:

(1)GB 25901.1-2010,信息技术 通用多八位编码字符集 德宏傣文 32点阵字型 第1部分:伊香白体^[8];

(2)GB 25901.2-2010,信息技术 通用多八位编码字符集 德宏傣文 32点阵字型 第2部分:伊香黑体^[9];

(3)GB 25902.1-2010,信息技术 通用多八位编码字符集 西双版纳新傣文 32点阵字型 第1部分:赫罕白体^[10];

(4)GB 25902.2-2010,信息技术 通用多八位编码字符集 西双版纳新傣文 32点阵字型 第2部分:赫罕黑体^[11];

(5)GB 25902.3-2010,信息技术 通用多八位编码字符集 西双版纳新傣文 32点阵字型 第3部分:温暖菲白体^[12];

(6)GB 25902.4-2010,信息技术 通用多八位编码字符集 西双版纳新傣文 32点阵字型 第4部分:温暖菲黑体^[13];

(7)GB 25902.5-2014,信息技术 通用多八位编码字符集 西双版纳新傣文 32点阵字型 第5部分:法杭^[14];

(8)GB 25902.6-2014,信息技术 通用多八位编码字符集 西双版纳新傣文 32点阵字型 第6部分:温暖菲大黑体^[15];

(9)GB 25902.7-2014,信息技术 通用多八位编码字符集 西双版纳新傣文 32点阵字型 第7部分:勒达毕体^[16];

(10)GB/T 28175-2011,信息技术 通用多八位编码字符集 德宏傣文 通用键盘字母数字区布局^[17];

(11)GB/T 28176-2011,信息技术 通用多八位编码字符集 西双版纳新傣文 通用键盘字母数字区布局^[18];

(12)GB/T 32637-2016,信息技术 通用多八位编码字符集 西双版纳老傣文 通用键盘字母数字区布局^[19]。

2 傣文信息化处理软件研究进展

1991年,潍坊华光电子信息产业集团在云南民族出版社的协作下,成功研发“华光5型傣文书刊电子出版系统”,用于傣文图书的出版,但当时还

不能用来编排傣文报纸。差不多同时,北大方正也推出了方正傣文出版系统。

1996年,西双版纳报社与潍坊华光科技股份有限公司合作研发了“新傣文计算机组版系统”^[20-21]。该系统可用于新傣文书、报、刊的编排,但还无法处理传统的老傣文。1997年,该系统通过西双版纳州科技成果鉴定,并获得西双版纳州科技进步二等奖。

2003年1月1日,西双版纳报社与潍坊北大青鸟华光科技股份有限公司合作研发的“西双版纳新老傣文计算机组版系统”正式发布^[22]。这是国内首套可编排传统傣文的电子出版系统。2003年4月12日,西双版纳报社使用“西双版纳新老傣文计算机组版系统”排出了《中国贝叶经》第1卷^[23],现已全部出版《中国贝叶经》全集100卷。

2004年以来,潍坊北大青鸟华光照排有限公司、新疆维吾尔自治区民族语言文字工作委员会和西双版纳报社等单位联合承担了教育部、国家语言文字工作委员会民族语言文字规范标准建设及信息化科研项目“基于ISO10646的维、哈、柯、傣文电子出版系统研发”^[24]。这是国内首套基于国际编码标准的傣文电子出版系统。2006年2月,该项目通过国家语言文字工作委员会结项鉴定,鉴定委员会一致认为该项目研究成果处于国际领先水平。2007年11月,该项目获得国家新闻出版行业最高科学技术奖——“王选科学技术一等奖”。

2006年以来,潍坊北大青鸟华光照排有限公司、西双版纳报社和北京书同文数字化技术有限公司等单位联合承担了教育部、国家语言文字工作委员会民族语言文字规范标准建设及信息化科研项目“中国贝叶文化资源库平台研究”^[25-26]。该项目主要研究中国贝叶文化资源库的关键技术与平台软件,形成资源库框架,并实现少量的“贝叶经”实体资源的数字化,为中国贝叶文化资源库的大规模建设摸索经验、做出示范、奠定基础。2014年6月,该项目通过国家语言文字工作委员会结项鉴定,鉴定委员会一致认为该项目研究成果处于国际领先水平。

中国科学院合肥智能机械研究所于2007年开展“汉傣双语农业智能系统开发平台”研制,并于2014年起开始研发傣汉机器翻译系统^[27]。该研究所通过在平台中集成统计机器翻译模块,并基于傣文的基础词典以及汉傣双语基础语料库(包括分词与相应的词性标注),实现汉傣双语语料库的对齐,在此基础上对相应的汉文农业智能系统知识进行自动翻译。

2008—2009年,西双版纳报社和潍坊北大青鸟华光照排有限公司联合建成了中国第一个傣文网站——西双版纳傣文新闻网站系统^[28-29],并开发了傣文移动采编系统、傣文数字报刊系统和傣文共享稿库系统。该系统于2010年通过云南省科技厅验收,鉴定结果为国际先进水平。2012年,该系统获得云南省科技进步三等奖、西双版纳州科技进步一等奖。

2010—2013年,教育部语言文字应用研究所等单位完成了国家社科基金项目“傣汉双语语料库建设及现代傣语词汇研究”^[30-31],构建了72万音节的傣文语料库,同时将中文分词方法应用到傣文中,结合傣文自身的特点,设计了一个基于音节序列标注的傣文分词系统,该分词系统达到了95.58%的综合评价。

2012年,西双版纳报社和潍坊北大青鸟华光照排有限公司联合编写了两部专著:(1)《傣文电脑实用技术》(傣、汉文对照版)^[32],该书主要包括傣文概况、电脑基础知识与基本操作、傣文输入法、傣文办公软件和傣文网站等内容。该成果获2014年度西双版纳州科技进步一等奖。(2)《傣文数字化技术》(傣、汉文对照版)^[33],该书收录了双方在各种学术刊物和学术会议上发表的傣文信息技术研究论文19篇。该成果获2015年度西双版纳州科技进步一等奖。

2013—2014年,西双版纳报社和潍坊北大青鸟华光照排有限公司联合研发了《西双版纳报》傣文手机APP客户端^[34],包括傣文全媒体资源库系统、傣文移动采编系统、傣文移动报刊制作系统、傣文移动客户端系统(Android、iPhone、iPad和Windows Phone)和傣文移动出版服务器系统。该系统获2016年度西双版纳州科技进步三等奖。

2016年,云南大学完成了傣文印刷体识别系统^[35]、联机手写识别系统^[36-37]。其中,傣文印刷体识别系统从新傣文的构词特点出发提出了可行的新傣文字符切分方法,并采用了基于BP神经网络和卷积神经网络两类识别算法,最终识别率分别为98.03%和98.30%。傣文联机手写识别系统分别采用了基于SVM和基于随机森林的两种算法。此外,云南大学还完成了词典与统计相结合的傣文分词系统^[38]。该系统从网络上下载语料,对这些语料进行整理,然后构建词典,并使用MMSEG+CRF方法对这些傣语文本进行分词,其正确率达97.7%,召回率达95.6%,F1值为96.6,合成出的

语音具有较好的自然度。

同年,云南民族大学成功研发了傣汉文有声电子词典^[39],实现了傣汉对照互译、傣语真人朗读、傣语音标显示等常用功能,同时还支持对词库进行添加、修改、删除自定义操作,实现了良好的人机交互功能。

目前,西双版纳报社和潍坊北大青鸟华光照明有限公司正在联合研发“西双版纳傣文电子词典与机器翻译系统”^[40]。该系统将收入 15 万条傣汉对照词库与 30 万条傣汉双语对齐语料库,计划 2018 年 12 月完成。

3 展望

随着人工智能技术与自然语言处理技术的深度融合,傣文信息处理已进入智能时代。预计在未来的几年内,傣文机器翻译、语音识别和智能检校技术将取得重大突破。

参考文献:

- [1] 殷建民. 中国多民族文字编码标准研究[J]. 中国传媒科技, 2005(4): 23-26.
YIN J M. Study on the coding standard of Chinese multi-ethnic characters [J]. Science Technology for China's Mass Media, 2005(4): 23-26.
- [2] The Unicode Consortium. The unicode standard version 4.0[M]. Boston: Addison-Wesley, 2003.
- [3] 殷建民. 中国少数民族文字国际编码标准的进展与问题[C]. 第四届全国语言文字应用学术研讨会. 成都, 2005.
YIN J M. Progress and problems in the international coding standards of Chinese minority languages[C]. The Fourth National Symposium on the Application of Language. Chengdu, 2005.
- [4] FREYTAG A. Summary of repertoire for FDAM and FPDAM2 of ISO/IEC 10646 (ISO/IEC JTCl/SC2/WG2 N2924R)[EB/OL]. (2005-02-08)[2017-12-08]. <http://anubis.dkuug.dk/jtcl/sc2/wg2>.
- [5] 殷建民. 中国少数民族文字国际编码标准的最新进展[C]//戴庆厦,赵小兵. 中国少数民族语言文字信息处理研究与发展. 北京:民族出版社, 2010:3-9.
YIN J M. The latest progress in the international coding standards of Chinese minority languages[C]//DAI Q X, ZHAO X B. Researches and advancements of information processing for Chinese minority languages and characters. Beijing: The Ethnic Publishing House, 2010:3-9.
- [6] EVENSON M. Summary of repertoire for FDAM 5 of

ISO IEC 10646 - 2003 (ISO/IEC JTCl/SC2/WG2 N3465) [EB/OL]. (2008-04-24)[2017-12-12]. <http://anubis.dkuug.dk/jtcl/sc2/wg2>.

- [7] 殷建民. 基于 ISO10646 的中国少数民族文字排版技术研究[C]//嘎日迪,吾守尔·斯拉木,德熙嘉措. 民族语言文字信息技术研究——第十一届全国民族语言文字信息学术研讨会论文集. 北京:西苑出版社, 2007:133-139.
YIN J M. Research on Chinese minority typesetting technology based on ISO10646[C]//GARIDI, WUSHUR SILAMUE, DICHIJATSU. Research on information technology of national language and characters—The eleventh national symposium on national language and language information. Beijing: Xiyuan Publishing House, 2007:133-139.
- [8] 克炳珍,和丽峰,陈壮,等. 信息技术 通用多八位编码字符集 德宏傣文 32 点阵字型 第 1 部分:伊香白体: GB 25901.1—2010[S]. 北京:中国标准出版社, 2011.
KE B Z, HE L F, CHEN Z, et al. Information technology—Universal multiple-octet coded character set—Dehong Dai 32 dot matrix font—Part 1: Yixiang Bai Ti: GB 25901.1—2010[S]. Beijing: China Standard Press, 2011.
- [9] 克炳珍,和丽峰,陈壮,等. 信息技术 通用多八位编码字符集 德宏傣文 32 点阵字型 第 2 部分:伊香黑体: GB 25901.2—2010[S]. 北京:中国标准出版社, 2011.
KE B Z, HE L F, CHEN Z, et al. Information technology—Universal multiple-octet coded character set—Dehong Dai 32 dot matrix font—Part 2: Yixiang Hei Ti: GB 25901.2—2010[S]. Beijing: China Standard Press, 2011.
- [10] 岩化,和丽峰,陈壮,等. 信息技术 通用多八位编码字符集 西双版纳新傣文 32 点阵字型 第 1 部分:赫罕白体: GB 25902.1—2010[S]. 北京:中国标准出版社, 2011.
YAN H, HE L F, CHEN Z, et al. Information technology—Universal multiple-octet coded character set—New Xishuang Banna Dai 32 dot matrix font—Part 1: Hehan Bai Ti: GB 25902.1—2010[S]. Beijing: China Standard Press, 2011.
- [11] 岩化,和丽峰,陈壮,等. 信息技术 通用多八位编码字符集 西双版纳新傣文 32 点阵字型 第 2 部分:赫罕黑体: GB 25902.2—2010[S]. 北京:中国标准出版社, 2011.
YAN H, HE L F, CHEN Z, et al. Information technology—Universal multiple-octet coded character set—New Xishuang Banna Dai 32 dot matrix font—Part 2: Hehan Hei Ti: GB 25902.2—2010[S]. Bei-

- jing;China Standard Press,2011.
- [12] 岩化,和丽峰,陈壮,等.信息技术通用多八位编码字符集 西双版纳新傣文 32点阵字型 第3部分:温暖菲白体:GB 25902.3—2010[S].北京:中国标准出版社,2011.
YAN H, HE L F, CHEN Z, et al. Information technology—Universal multiple - octet coded character set—New Xishuang Banna Dai 32 dot matrix font—Part 3: Wenuanfei Bai Ti; GB 25902.3—2010[S]. Beijing:China Standard Press,2011.
- [13] 岩温胆,和丽峰,陈壮,等.信息技术通用多八位编码字符集 西双版纳新傣文 32点阵字型 第4部分:温暖菲黑体:GB 25902.4—2010[S].北京:中国标准出版社,2011.
YAN W D, HE L F, CHEN Z, et al. Information technology—Universal multiple - octet coded character set—New Xishuang Banna Dai 32 dot matrix font—Part 4: Wenuanfei Hei Ti; GB 25902.4—2010[S]. Beijing:China Standard Press,2011.
- [14] 岩温胆,和丽峰,刀福祥,等.信息技术通用多八位编码字符集 西双版纳新傣文 32点阵字型 第5部分:法杭:GB 25902.5—2014[S].北京:中国标准出版社,2014.
YAN W D, HE L F, DAO F X, et al. Information technology—Universal multiple-octet coded character set—New Xishuang Banna Dai 32 dot matrix font—Part 5: Fahang Ti; GB 25902.5—2014[S]. Beijing: China Standard Press,2014.
- [15] 刀福祥,岩温胆,和丽峰,等.信息技术通用多八位编码字符集 西双版纳新傣文 32点阵字型 第6部分:温暖菲大黑体:GB 25902.6—2014[S].北京:中国标准出版社,2014.
DAO F X, YAN W D, HE L F, et al. Information technology—Universal multiple-octet coded character set—New Xishuang Banna Dai 32 dot matrix font—Part 6: Wenuanfeidahei Ti; GB 25902.6—2014[S]. Beijing:China Standard Press,2014.
- [16] 岩温胆,和丽峰,陈壮,等.信息技术通用多八位编码字符集 西双版纳新傣文 32点阵字型 第7部分:勒达毕体:GB 25902.7—2014[S].北京:中国标准出版社,2014.
YAN W D, HE L F, CHEN Z, et al. Information technology—Universal multiple - octet coded character set—New Xishuang Banna Dai 32 dot matrix font—Part 7: Ledabi Ti; GB 25902.7—2014[S]. Beijing: China Standard Press,2014.
- [17] 唐英敏,和丽峰,宿忠民,等.信息技术通用多八位编码字符集 德宏傣文通用键盘字母数字区布局:GB/T 28175—2011[S].北京:中国标准出版社,2012.
TANG Y M, HE L F, SU Z M, et al. Information technology—Universal multiple-octet coded character set—Universal keyboard layout of the alphanumeric zone for Dehong Dai;GB/T 28175—2011[S]. Beijing: China Standard Press,2012.
- [18] 和丽峰,唐英敏,宿忠民,等.信息技术通用多八位编码字符集 西双版纳新傣文通用键盘字母数字区布局:GB/T 28176—2011[S].北京:中国标准出版社,2012.
HE L F, TANG Y M, SU Z M, et al. Information technology—Universal multiple-octet coded character set—Universal keyboard layout of the alphanumeric zone for New Xishuang Banna Dai; GB/T 28176—2011[S]. Beijing:China Standard Press,2012.
- [19] 殷建民,王康龙,黄姗姗,等.信息技术通用多八位编码字符集 西双版纳老傣文通用键盘字母数字区布局:GB/T 32637—2016[S].北京:中国标准出版社,2016.
YIN J M, YU K L, HUANG S S, et al. Information technology—Universal multiple-octet coded character set—Universal keyboard layout of the alphanumeric zone for Old Xishuang Banna Dai;GB/T 32637—2016[S]. Beijing:China Standard Press,2016.
- [20] 殷建民.少数民族文字电子出版系统的回顾与展望[C]//2004中国少数民族语言信息技术与语言资源建设学术研讨会论文集.[不详;不详],2004.
YIN J M. Review and prospect of electronic publishing system for minority languages[C]//The Academic Symposium on the Construction of Chinese Minority Language Information Technology and Language Resource Library in 2004. [S.l:s. n.],2004.
- [21] 殷建民.中国少数民族文字电子出版系统的研发实践与思考[C]//中国新闻技术工作者联合会四届一次理事会暨学术年会论文集.[不详;不详],2005:187-191.
YIN J M. The research and development of the electronic publishing system of Chinese minority languages[C]//One of the Four Council and Academic Annual Meeting of the Chinese Federation of Journalists and Technicians. [S.l:s. n.],2005:187-191.
- [22] 刀福祥,殷建民.西双版纳傣文信息技术研究[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,2006:165-171.
DAO F X, YIN J M. Research on information technology of Tai Lue in Xishuangbanna[C]//The Tenth Session of the Academic Symposium on the Information Processing of the Language and Character of the

- National Minorities. Xining, 2006: 165-171.
- [23] 殷建民, 刀福祥, 岩温胆, 等. 传统傣文与“贝叶经”的计算机排版[J]. 中国传媒科技, 2003(3): 51-52.
YIN J M, DAO F X, YAN W D, et al. The traditional Dai script and “Pattra” computer typesetting[J]. Science Technology for China’s Mass Media, 2003(3): 51-52.
- [24] 殷建民, 吕建春, 张羽宏. 中国少数民族文字电子出版系统 20 年[C]//国家民族事务委员会文化宣传司. 构建多语和谐的社会语言生活: 民族语文国际学术研讨会论文集. 北京: 民族出版社, 2009: 517-524.
YIN J M, LV J C, ZHANG Y H. 20 years of electronic publishing system for Chinese minority languages [C]//Cultural Propaganda Department of the National Committee for National Affairs. The construction of multilingual harmonious social language life: International symposium on national language. Beijing: The Ethnic Publishing House, 2009: 517-524.
- [25] 殷建民, 刀福祥. 中国贝叶文化数字博物馆初探[C]//中国新闻技术工作者联合会学术年会四届一次理事会暨学术年会论文集. [不详: 不详], 2005: 153-156.
YIN J M, DAO F X. A probe into the Digital Museum of China’s Pattra-leaf culture[C]//One of the Four Council and Academic Annual Meeting of the Chinese Federation of Journalists and Technicians. [S. l.: s. n.], 2005: 153-156.
- [26] 殷建民, 刀福祥, 张轴材. 西双版纳傣文“贝页经”资源库技术研究[C]//曹右琦, 孙茂松. 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. 北京: 清华大学出版社, 2006: 248-251.
YIN J M, DAO F X, ZHANG Z C. Study on resource database of Tai Lue Pattra-leaf Buddhist scriptures in Xishuangbanna[C]//CAO Y Q, SUN M S. Advance in Chinese information processing—The 25th anniversary academic conference of Chinese information society. Beijing: Tsinghua University Press, 2006: 248-251.
- [27] 胡泽林, 高翔, 李森, 等. 汉民双语农业智能系统开发平台与关键技术研究[C]. 第十六届全国少数民族语言文字信息处理学术研讨会. 桂林, 2017.
HU Z L, GAO Y, LI M, et al. Research on the development platform and key technologies of Han bilingual agriculture intelligent system[C]. The Sixteenth Session of the Academic Symposium on the Information Processing of the Language and Character of the National Minorities. Guilin, 2017.
- [28] 刀福祥, 殷建民. 努力缩小“数字化鸿沟”——关于傣文数字化报纸网络研发的思考[J]. 中国地市报人, 2009(3): 36-38.
DAO F X, YIN J M. Efforts to narrow the “digital divide”—Thinking on the research and development of digital newspaper network in Dai script [J]. China Prefecture and Town, 2009(3): 36-38.
- [29] 殷建民, 刀福祥, 唐金宝, 等. 西双版纳傣文新闻网站与数字报刊技术研究[J]. 中文信息学报, 2011, 25(4): 11-15, 44.
YIN J M, DAO F X, TANG J B, et al. Study on the news site and digital newspaper for Tai Lue[J]. Journal of Chinese Information Processing, 2011, 25(4): 11-15, 44.
- [30] 戴洪亮, 高庭丽. 西双版纳傣文语料库的建设[C]. 第七届全国语言文字应用学术研讨会. 湘潭, 2011.
DAI H L, GAO T L. The construction of Tai Lue corpus in Xishuangbanna [C]. The Seventh National Symposium on the Application of Language. Xiangtan, 2011.
- [31] 高庭丽, 陶建华, 戴洪亮, 等. 傣文自动分词系统的设计与实现[J]. 中文信息学报, 2013, 27(6): 187-192.
GAO T L, TAO J H, DAI H L, et al. Daiwen word segmentation system design and implementation[J]. Journal of Chinese Information Processing, 2013, 27(6): 187-192.
- [32] 刀福祥, 殷建民, 玉康龙, 等. 傣文电脑实用技术[M]. 昆明: 云南科技出版社, 2012.
DAO F X, YIN J M, YU K L, et al. Practical computer technology of Dai script [M]. Kunming: Yunnan Science and Technology Press, 2012.
- [33] 刀福祥, 殷建民, 玉康龙, 等. 傣文数字化研究[M]. 昆明: 云南科技出版社, 2012.
DAO F X, YIN J M, YU K L, et al. Research on digitalization of Dai script [M]. Kunming: Yunnan Science and Technology Press, 2012.
- [34] 殷建民, 玉康龙, 胡世波, 等. 西双版纳傣文移动出版技术研究[C]. 第十四届全国少数民族语言文字信息处理学术研讨会. 兰州, 2013.
YIN J M, YU K L, HU S B, et al. Research on the mobile publishing technology of Tai Lue [C]. The Fourteenth Session of the Academic Symposium on the Information Processing of the Language and Character of the National Minorities. Lanzhou, 2013.
- [35] 李登峰. 印刷体新傣泐文字符识别方法研究[D]. 昆明: 云南大学, 2016.
LI D F. Printed new Tai Lue character recognition technology study [D]. Kunming: Yunnan University, 2016.

- technology standardization[J]. Journal of Chinese Information Processing, 2016, 30(3): 85-89.
- [12] 朴明姬, 崔荣一. 多语种文本图像中的文字语种辨识方法的研究[J]. 中文信息学报, 2017, 31(2): 220-225.
PIAO M J, CUI R Y. An approach to script identification in image with multi-lingual texts[J]. Journal of Chinese Information Processing, 2017, 31(2): 220-225.
- [13] 刘双君, 金小峰, 崔荣一. 基于基频的朝鲜语方言辨识方法的研究[J]. 中文信息学报, 2017, 31(2): 55-60, 70.
LIU S J, JIN X F, CUI R Y. Research on Korean dialect identification based on pitch feature[J]. Journal of Chinese Information Processing, 2017, 31(2): 55-60, 70.
- [14] 卢星华. 韩国语听力教材研究——基于韩国语准口语语料库中的代词计量特征分析[J]. 东疆学刊, 2017, 34(3): 62-67, 112.
LU X H. Research on Korean listening textbooks—The analysis of the characteristics of pronouns based on the speech corpus of Korean language [J]. Dongjiang Journal, 2017, 34(3): 62-67, 112.

(责任编辑: 陆 雁)

(上接第 17 页 Continue from page 17)

- [36] 钟卿. 基于 SVM 的联机手写新傣文字符识别[D]. 昆明: 云南大学, 2016.
ZHONG Q. Online New Tai Lue handwritten character recognition based on the SVM[D]. Kunming: Yunnan University, 2016.
- [37] 陈瑞新. 基于随机森林的联机手写新傣文字符识别技术研究[实现][D]. 昆明: 云南大学, 2016.
CHEN R X. Online handwritten character recognition of New Tai Lue based on random forest [D]. Kunming: Yunnan University, 2016.
- [38] 李慧. 词典与统计相结合的傣文分词方法与实现[D]. 昆明: 云南大学, 2016.
LI H. Dai language segmentation based on dictionary and statistics [D]. Kunming: Yunnan University, 2016.
- [39] 胡刚, 王嘉梅, 李炳泽, 等. 傣泐文-汉文互译有声电子词典[J]. 计算机系统应用, 2016, 25(7): 8-16.
HU G, WANG J M, LI B Z, et al. Daile Wen-Chinese translation audible electronic dictionary[J]. Computer Systems & Applications, 2016, 25(7): 8-16.
- [40] 殷建民, 玉康龙, 岩香, 等. 西双版纳傣文电子词典及辅助翻译技术研究[C]. 第十五届全国少数民族语言文字信息处理学术研讨会. 延边, 2015.
YIN J M, YU K L, YAN X, et al. Research on electronic dictionary and auxiliary translation technology of Tai Lue[C]. The Fifteenth Session of the Academic Symposium on the Information Processing of the Language and Character of the National Minorities. Yanbian, 2015.

(责任编辑: 陆 雁)