

中央民族大学“985工程”中国少数民族语言文化教育与边疆史地研究创新基地文库
国家语言资源监测与研究中心少数民族语言分中心计算语言学系列丛书

中国少数民族语言文字 信息处理研究与发展

Researches and Advancements of Information Processing for
Chinese Minority Languages and Characters

主 编：戴 庆 厦 赵小兵

副主编：大罗桑朗杰 欧 珠



民族出版社

中国少数民族语言文字 信息处理研究与发展

Researches and Advancements of Information Processing For
Chinese Minority Languages and Characters

主 编：戴 庆 厦 赵小兵

副主编：大罗桑朗杰 欧 珠

民族出版社

图书在版编目 (C I P) 数据

中国少数民族语言文字信息处理研究与发展/戴庆厦, 赵小兵 主编. —北京: 民族出版社, 2010. 5

ISBN 978 - 7 - 105 - 10829 - 9

I. 中… II. ①戴… ②赵… III. 少数民族 - 民族语 - 语言信息处理学 - 研究 - 中国 IV. ①TP391

中国版本图书馆 CIP 数据核字 (2010) 第 098892 号

中国少数民族语言文字信息处理研究与发展

策划编辑: 欧光明

责任编辑: 冯 光 张宏宏

出版发行: 民族出版社

社 址: 北京市和平里北街 14 号 邮编 100013

电 话: 010 - 64228001 (编辑室)

010 - 64211734 (发行部)

网 址: <http://www.mzcbs.com>

印 刷: 文阁印刷厂

经 销: 各地新华书店

版 次: 2010 年 6 月第 1 版 2010 年 6 月北京第 1 次印刷

开 本: 787 毫米 × 1092 毫米 1/16

字 数: 573 千字

印 张: 25.5

定 价: 65.00 元

ISBN 978 - 7 - 105 - 10829 - 9/T · 22 (汉 13)

该书如有印装质量问题, 请与本社发行部联系退换。

为更好地提供中国少数民族语言计算语言学学科发展交流与合作平台，中央民族大学“985 工程”中国少数民族语言文化教育与边疆史地研究创新基地拟于 2010 年开始创办《民族语文信息学报》，欢迎国内外相关领域的专家与学者踊跃投稿。

投稿地址：北京市海淀区中关村南大街 27 号中央民族大学理工楼 516

邮 编：100081

联系方式：010 - 68930880

邮 箱：mzxxxb@163. com

**《中国少数民族语言文字信息
处理研究与发展》编委会成员**

戴庆厦 赵小兵 王志娟
闫晓东 孙 媛 李叶青

第十二届全国少数民族语言文字信息 处理学术研讨会大会组织机构

大会主席：倪光南 院 士（中国工程院院士、中国中文信息学会理事长）

李宇明 司 长（教育部语言信息管理司）

俸 兰 司 长（国家民委教育科技司）

副主席：王铁琨 副司长（教育部语言信息管理司）

张京泽 副司长（国家民委教育科技司）

田福利 副主任（西藏发改委）

强俄巴·次央 副校长（西藏大学）

宋 敏 副校长（中央民族大学）

指导委员会主任：俞士汶 教 授（北京大学计算语言学研究所）

指导委员会委员：（按姓氏拼音排序）

陈 敏 处 长（教育部语言信息管理司规划协调处）

陈群秀 教 授（清华大学）

丁晓青 教 授（清华大学）

冯志伟 教 授（教育部语言应用研究所）

史忠植 研究员（中国科学院计算技术研究所）

李万瑛 副局长（中国民族语文翻译局）

刘 群 研究员（中国科学院计算技术研究所）

刘 挺 教 授（哈尔滨工业大学）

尼玛扎西 教 授（西藏大学现代教育技术中心）

宋 柔 教 授（北京语言大学）

孙茂松 教 授（清华大学）

孙 乐 副研究员（中国科学院软件研究所）

王守利 处 长（国家民委教育科技司科技处）

吴 健 研究员（中国科学院软件研究所）

殷建民 研究员（潍坊北大青鸟华光照排有限公司）

张 普 教 授（北京语言大学）

宗成庆 研究员（中国科学院自动化研究所）

程序委员会主席：戴庆厦 教 授（中央民族大学）

程序委员会副主席：大罗桑朗杰（藏） 教 授（西藏大学藏文信息技术研究中心）

嘎日迪（蒙古） 教 授（内蒙古师范大学）

吾守尔·斯拉木（维吾尔） 教 授（新疆大学）

程序委员会委员：（按姓氏拼音排序）

奥其尔（蒙古） 教 授（内蒙古大学计算机学院）

欧 珠（藏） 教 授（西藏大学工学院）

才让加（藏） 教 授（青海师范大学）

陈玉忠（藏） 研究员（中国标准化研究所）

崔荣一（朝鲜） 教 授（延边大学计算机学院）

刀福祥（傣） 编 审（云南省西双版纳报社）

古丽拉（哈萨克） 教 授（新疆大学计算机学院）

华沙宝（蒙古） 教 授（内蒙古大学蒙古语言研究所）

江 荻 研究员（中国社会科学院人类学与民族学研究所）

那顺乌日图（蒙古） 教 授（内蒙古大学蒙古语言研究所）

沙玛拉毅（彝） 教 授（西南民族大学）

吐尔根（维吾尔） 教 授（新疆大学计算机学院）

玄龙云（朝鲜） 高 工（朝鲜语信息学会）

于洪志 教 授（西北民族大学中国民族信息技术研究院）

玉索甫（维吾尔） 教 授（新疆师范大学数理信息学院电子研究所）

赵小兵（蒙古） 教 授（中央民族大学国家语言监测少数民族语言分中心）

组织委员会主席：大罗桑朗杰（藏） 教 授（西藏大学藏文信息技术研究中心）

组织委员会副主席：付 忠 秘书长（西藏自治区信息协会）

欧 珠（藏） 教 授（西藏大学工学院）

赵小兵（蒙古） 教 授（中央民族大学国家语言监测少数民族语言分中心）

委 员：（按姓氏拼音排序）

次仁罗布 讲 师（西藏大学藏文信息技术教育部工程研究中心）

高定国 讲 师（西藏大学藏文信息技术教育部工程研究中心）

格桑多吉 讲 师（西藏大学藏文信息技术教育部工程研究中心）

李叶青 副教授（中央民族大学国家语言监测少数民族语言分中心）

尼玛扎西 教 授（西藏大学现代教育技术中心）

邱莉榕 博 士（中央民族大学国家语言监测少数民族语言分中心）

孙 媛 博 士（中央民族大学国家语言监测少数民族语言分中心）

王志娟 副教授（中央民族大学国家语言监测少数民族语言分中心）

闫晓东 副教授（中央民族大学国家语言监测少数民族语言分中心）

周 建 副秘书长（西藏自治区信息协会）

珠 杰 讲 师（西藏大学藏文信息技术教育部工程研究中心）

序 言

语言是信息的主要载体,随着信息化时代的到来,语言处理技术已经日益成为全球信息化和我国社会经济发展的重要支撑技术。

中国是统一的多民族国家。在这块约 1000 万平方公里的大地上,分布着 56 个民族,其中主体民族——汉族,人口数为 115940 万人^①,占全国总人口的 91.59%,其他 55 个少数民族人口数为 10643 万人,占总人口数的 8.41%。新中国成立以来,汉族和少数民族群众相互取长补短,互通有无,形成了各民族团结合作、和谐发展、繁荣进步的大家庭。

今天,中国少数民族语言文字信息处理技术的研究工作在国家与地方政府的大力扶持下,经过近四十年发展,经历了从无到有、快速发展的历程。该领域研究的成果现在已深入到了社会的各个应用层面。主要包括:制定了一系列少数民族文字的编码字符集、键盘、字模等国家标准;开发了一批从 DOS 操作系统到 Windows、Linux 系统环境的应用软件系统;逐步建立和完善了民族语言文字的网站或网页;少数民族语言的信息化应用已经进入到手机、电子词典等个人电子产品领域,为广大少数民族群众带来了便利。另外,少数民族计算语言学研究还延伸到了文本自动校对、印刷体与手写体文字识别、语音识别与合成、网络跨语言搜索、民族语言文本海量计算、数据挖掘、机器翻译、语言自动分析、语言监测与信息安全等诸多领域。目前,无论是计算语言学的应用领域研究,还是计算语言学本体研究,都呈现出良好的发展态势,培养了一批专门从事少数民族语言的本学科领域研究的科研团队和技术骨干。

为了更好地推动中国少数民族语言信息化的发展进程和计算语言学的研究与应用,由中国中文信息学会组织,中国中文信息学会少数民族语言文字信息专委会主办,每两年举行一次“中国少数民族语言文字信息处理学术研讨会”。该研讨会的宗旨是构建学术和技术交流平台,促进少数民族语言信息化、标准化发展及民族语言信息化产业的转化与推广。2009 年 7 月 10—12 日,“第十二届中国少数民族语言文字信息处理学术研讨会”在西藏拉萨隆重召开,会议由中央民族大学、西藏大学和西藏自治区信息协会共同承办。来自教育部语言信息管理司、中国中文信息学会

^① 2001 年 3 月 28 日,中华人民共和国国家统计局《2000 年第五次全国人口普查主要数据公报》发布。

的领导以及国内外知名专家和少数民族地区的学者汇聚一堂,就少数民族语言文字信息处理的研究发展进行了充分地交流和探讨。

会议期间,来自北京语言大学的宋柔教授、清华大学的马少平教授、哈尔滨工业大学的赵铁军教授、西藏大学的欧珠教授、中科院计算所的刘群教授、厦门大学的史晓东教授、中科院自动化所的宗成庆教授以及中央民族大学国家语言资源监测与研究中心少数民族语言分中心的赵小兵教授分别做了精彩的报告。会议还开展了分组专题讨论、成果展示及优秀论文的评选等多项学术活动。

本次会议的圆满举行对促进中国少数民族语言联盟体系的建设,建立中国少数民族语言资源的“统一规划、统一建设、统一标准、统一平台、资源共享”的机制具有积极的推动作用,对中国少数民族语言文字信息处理技术的交流与发展具有深远的影响。我们以本届会议的研究成果为基础,择优收录研究论文 55 篇,内容涵盖了中国少数民族语言信息处理基础理论与标准化研究、语料库与知识库构建、支撑平台与条件建设、信息化处理技术及网络应用等几大研究领域。这些成果反映了目前我国少数民族语言信息化技术的研究与发展程度。我们相信,在大家的共同努力下,不但可以提升少数民族地区的科技、文教领域的民众综合素质,加速民族地区的经济发展,而且还将为缩小民族之间、内地与边疆之间的差距,增强我国的综合国力发挥重要作用。少数民族语言信息化事业的明天必将更加灿烂辉煌!

在此,向一直关注和支持少数民族语言文字信息化事业的教育部语言信息管理司、国家民委教育科技司、中国中文信息学会表示感谢!向承办本次会议的西藏大学、中央民族大学、西藏自治区信息协会、西藏藏语言文字工作委员会等单位表示感谢!

中国中文信息学会少数民族语言文字信息专委会

戴庆厦

赵小兵

2009 年 7 月 20 日

致“第十二届中国少数民族语言信息处理学术研讨会” 的贺信

出席致“第十二届中国少数民族语言信息处理学术研讨会”的各位专家和朋友：

首先热烈祝贺“第十二届中国少数民族语言信息处理学术研讨会”在拉萨隆重召开。

感谢大会组织者给了我十分崇高的荣誉，让我担任指导委员会主任，并邀请我在大会上作特邀报告。由于年事已高，最终还是放弃了这样一个与大家交流、向大家学习的难得机会。敬请原谅。

最近两年，我分别参加了在西双版纳、呼和浩特、合肥召开的有关民族语言信息处理的学术会议，亲身感受到了我国民族语言信息处理事业的蓬勃发展，作为一名长期从事语言信息处理研究的科研人员，备感兴奋和欣慰。

大家经常谈论社会信息化的话题。社会信息化的方方面面不胜枚举，实现自然语言的计算机自动处理是其关键技术之一。自然语言处理的最高境界是自然语言理解，这是一个引人入胜又困难重重的研究课题。我曾看到一篇谈论“生命进化的十大奇迹”的文章，其中，人类的大脑和语言分别列为第3项和第4项。大脑常常被看作进化过程中的最高成就，它赋予了人类一些高级特征，例如语言、智慧、意识。语言是进化的终极发明。在把人类与动物区别开来的特征中，语言处于核心地位。我们的祖先如何实现了语言从无到有的飞跃，这也许是科学史上最大的谜。语言系统是动物进化到人的两大标志之一（另一个标志就是大家常说的人类会制造工具）。由于对大脑、对语言的认知机制尚不甚了了，目前还不能在计算机中建立语言模型，对语言进行模拟，其结果能够让人满意。对语言的研究也并非只是信息处理的需要，自从人类有了语言，就有了对语言的研究。无论计算机的使用与否，语言研究到目前还没能取得突破性的进展，根本原因可能就在于此。还有一点，人们研究任何一个问题，都是借助语言这个工具进行思考的，当研究语言时，语言既是对象，又是工具，界限无法划清，犹如“不识庐山真面目，只缘身在此山中”，这可能也是原因之一。

关于语言的上述认识是普遍的，不分语种的，汉语是这样，藏语、蒙古语、维吾尔语、哈萨克语、朝鲜语等也是如此。每种语言都有丰富的词汇、灵活的句法结

构以及合适的语义表现形式。每个民族都运用自己的语言创造了辉煌的物质文明和精神文明。但人类为什么会有这么多种不同的语言，这可能又是一个难解的谜。

存在不同的语言是客观事实，但不同语言的存在又不是静止不变的，有语言，就有语言的融合和分化。人们特别注意到，有的语言是强势语言，在发展，在扩张；有的语言，使用的人在减少，甚至濒临消失。而一种语言的消失即意味着一种文明的消失。物种多样性的重要已经为人们所认识，文明以及承载文明的语言之多样性同样重要。

要保护一种语言，最有效的办法是让以这种语言为母语的人坚持使用它。不同语言之间及其他他们所承载的文明之间需要交流，翻译是不同文明之间交流的桥梁。不过，翻译的成本很高。如何降低翻译的成本，提高翻译的水平，则使语言信息处理技术有了用武之地。把翻译的任务交给机器，通过这种翻译机器，人们都使用自己的母语，却可以方便地进行交流，这或许是母语中心主义者的梦想。当然将梦想变成现实，需要付出艰苦的努力。自然语言理解需要几代人甚至十几代人的努力。这期间，懂得多种语言的专家会做出重要的贡献。与会的民族语言专家显然有明显的优势，你们都精通多种语言，比我们这些只会汉语的人更容易理解不同语言的个性以及所有语言的共性，又掌握信息处理的技术，所以大家应该很好地团结协作，相信在各位专家的努力下，我国民族语言信息处理事业一定能得到更快的发展，我国民族地区的社会信息化进程进一步加快，为人类文明的进步作出贡献。

与会的民族语言专家中，有很多人是我相交多年的老朋友，你们的真诚、热情、豪放经常让我感动。这次失去了与你们欢聚的好机会，深感遗憾。你们到北京的机会多，欢迎你们常到北大计算语言学研究所来进行交流。向大家通报一个消息，计算语言学教育部重点实验室已经在北京大学开始建设，王厚峰教授任主任，李宇明教授、王洪君教授任学术委员会主任。希望这个平台能为计算语言学领域的交流提供更多的机会和便利。

谢谢大家！祝会议取得圆满成功！

俞士汶

2009年7月6日于北京

目 录

少数民族语言信息技术研究现状

中国少数民族文字国际编码标准的最新进展	殷建民 (3)
维吾尔文媒体用语调查方案初探	玉素甫·艾白都拉 潘伟民 买买提艾力 (10)
有关少数民族语言信息化相关项目调研	闫晓东 赵小兵 阿亚拉古 (15)
有关少数民族语言信息化获奖、标准、专利情况调研	李叶青 闫晓东 赵小兵 (22)
藏语网站的现状	王志娟 覃其文 (26)
回顾与思索	嘎日迪 (32)

少数民族语言信息处理技术标准和规范方面的研究

基于 Unicode 的八思巴文信息处理的研究	敖其尔 陈建东 包春梅 (41)
计算机通用键盘键位属性研究——以藏文计算机键盘布局设计为例	卢亚军 (49)
Microsoft Himalaya 的 OpenType 布局表分析	高定国 关白 (58)
新疆民族文字信息技术字形标准研究	亚森·伊明 佟加·庆夫 (65)
哈萨克文语料库 Xml 格式标注规范初探	吴守用 古丽拉·阿东别克 (71)
国际通用语言研究 (精要概述)	岳元富 (78)

资源库建设研究

现代汉语普通话数字化样本库的设计与建设

- 侯敏 邹煜 滕永林 何伟 张佳 (83)
- 蒙古语语音语料库的研究和建设..... 红梅 嘎日迪 (91)
- 计算机在汉阿语料库建设中的运用问题 希夏姆 (95)
- 《蒙古语同形词信息词典》管理平台的设计和实现 哈斯 (114)
- 基于 C/S 模式的汉、英、维、哈、柯文名词术语管理系统的设计与实现
..... 亚森·伊明 阿布力米提·库尔班 (120)
- 达日罕谜语系统的设计与实现 呼日乐吐什 长青 那顺乌日图 (125)

语音、词法、句法、语法、语义分析、分词标注等研究

- 蒙古语元音的标准音测试方法研究 孟和吉雅 白音门德 (133)
- 拉萨话文读辅音韵尾的声学分析 央金卓嘎 扎西加 (138)
- 藏汉英电子词典硬件产品中藏语语音的实现方法 洛藏 (145)
- 维吾尔语音节音高的聚类分析 古丽娜尔·艾力 艾斯卡尔·艾木都拉 (148)
- 基于有声媒体语音语料的主持人口语韵律词重音分析
..... 邹煜 何伟 侯敏 滕永林 朱维彬 (152)
- 藏英虚词的对比 东智布加 (159)
- 蒙古文词干还原系统设计和研究中的若干问题 那日松 淑琴 (165)
- 基于 LR 算法的维吾尔语 KP 短语的句法分析方法
..... 阿力木江·托乎提 米娜瓦尔·吾买尔 (171)
- 藏文 n 元语法模型系统 谢文森 吴建 (183)
- 信息处理用藏语虚词位格助词的语法功能研究 索南才让 (193)
- 基于聚类的中文词汇链构建算法的研究 练睿婷 史晓东 (202)
- 基于语义相似度的信息类 Web 服务组合 邱莉榕 (212)
- 斯拉夫蒙古文音节切分方法初探 包权 (219)
- 藏文自动分词及其应用研究 才让卓玛 (223)
- 藏语自动分词方案的设计 孙媛 罗桑强巴 杨锐 赵小兵 (228)
- 藏语交集型歧义字段切分方法研究 孙媛 罗桑强巴 杨锐 赵小兵 (238)
- 基于维特比算法单字估价值的中文自动分词 李成城 李艳秋 (244)

机器翻译、文字识别、文本分类、信息抽取等研究

- 班智达汉藏机器翻译系统简述 才让加 (251)
- 藏文文字识别系统中的数字图像预处理方法研究 刘 芳 欧 珠 (257)
- 藏文手写识别中相似字区分的研究 孙 嫣 吴 健 (265)
- 维吾尔文联机手写识别的预处理和特征提取
..... 阿力木江·亚森 哈力木拉提·买买提 (275)
- 一种基于正例的汉语意见型主观性文本分类方法 姚天昉 张 鑫 (284)
- 基于规则的 WEB 文本信息抽取平台设计与实现
..... 朱 义 邢春晓 黄梦醒 (294)

面向信息处理的网络技术研究

- Heritrix 网页抓取器的分析与扩展 珠 杰 顿珠次仁 (303)
- 关于少数民族网络语料下载中的一个技术问题及解决方法 张新平 (309)
- 一种支持 QoS 评价的 Web 服务推荐模型 李叶青 邱莉榕 (313)

应用系统

- 一种面向汉字教学研究的字形描述方法及工具 林 民 宋 柔 (321)
- MHK (三级) 口语考试计算机评分的效度研究 张 洁 (329)
- 蒙古文控制字符对编辑操作的影响
..... 呼斯勒 白双成 巴图赛恒 苏雅拉图 (337)
- 蒙古文在线输入法研究与实现 胡其图 六 月 巴图赛恒 (342)
- 一种蒙古文单词拼写校正的方法 敖登巴拉 巩 政 (346)
- 基于蒙古文拉丁转写联合方案的自动转写系统 斯·劳格劳 华沙宝 (351)
- 蒙古文网站内容管理系统研究 白双成 (357)
- 浅谈开放式蒙古语网上百科全书的实现及意义 孟高娃 (366)
- 蒙医药信息化中的蒙古文信息处理 阿茹娜 呼斯勒 巴图赛恒 孟高娃 (374)
- 维吾尔、哈萨克、柯尔克孜文数字报纸的实现 张东升 (378)
- 锡伯文、满文传媒出版技术研究 佟加·庆夫 亚森·伊明 (384)

CONTENTS

Status of information processing for minority language

- The Latest Development on International Coding Standard of Chinese Minority Script
..... Yin Jianmin (3)
- A Survey on Uighur language for the media program
..... Yusup · Abadulla Pan Weimin Mamatili (10)
- Research on minority language informatization projects
..... Yan Xiao dong Zhao Xiao bing Ayalagu (15)
- Research on minority language informatization Awards、Standards、Patents
..... Li Ye qing Yan Xiao dong Zhao Xiaobing (22)
- The Present Situation of Websites in Tibetan
..... Wang Zhijuan Qin Qiwen (26)
- Review and reflection
..... Ga Ridi (32)

Standards and guidelines of information processing for minority language

- The research on information processing of Phage ~ pa script based on Unicode
..... Ochir Chen Jiandong Bao Chunmei (41)
- A Study of Standard Keyboard Attribution; A Tibetan Keyboard Layout
..... Lu Yajun (49)
- Analyze the OpenType Layout of Microsoft Himalaya
..... Gao Dingguo Guan bai (58)
- Study on the Font Standards for the Technology of Xinjiang National Information Exchange technology
..... Yasin · Imin Tongjia · Qingfu (65)
- Exploring To The Xml Markup Guidelines Of the Kazakh Corpus
..... Wu Shou yong Gulila · Altenbek Feng Jinghua (71)
- An international common language research (Essentials Overview)
..... Fu Yueyuan (78)

Researches of Corpora building

- Construction and Analysis of Chinese Mandarin Digital Multi – modal Corpus
 Hou Min Zou Yu Teng Yonglin He Wei Zhang Jia (83)
- The Research and Construction of Mongolian Speech Corpus
 Hong Mei Ga Ridi (91)
- The Problems of Computational treating and constructing Chinese and Arabic Corpus
 Hesham El – Malky (95)
- The Design and Implementation of “Mongolian Homograph Words Information Dictionary” Management Platform
 Hasi (114)
- Design and Implementation of Chinese, English, Uyghur, Kazak, Kirghiz terminology management system Based on C/S model
 Yasin · Imin Ablimit · Kurban (120)
- “Darhan riddle system” Design and Implementation
 Nasvnvrtv Hureltux ChangQing (125)

Researches of Speech, Morphology, Syntax, Semantic analysis, Segmentation & tagging, etc

- Research of Accessory Testing Method of Mongolian Vowels Standard Speech
 Monghjaya Bayarmend (133)
- Acoustic analysis of reading – Lhasa’ s consonantal ending of a final
 Yang Jin Zhuo Ga TashiGyal (138)
- A Solution to Actualize the Tibetan Speech in Tibetan – Chinese – English Electronic Dictionary Hardware
 Luo Zang (145)
- Uighur language Pitch syllable cluster analysis
 Gulnar · Aili Askar · hamdulla (148)
- Statistical Analysis on Prosodic Word Stresses Based on Broadcast Media Speech Corpus
 Zou Yu He Wei Hou Min Teng Yonglin Zhu Weibin (152)
- The comparison between Tibetan and English Function word
 Dondropga (159)
- The Several issues on design and research to system of Mongolian stemming
 Narisong Shuqin (165)
- A Syntactic Analysis of KP phrase in Uyghur Based on LR
 Alimjan Tohti Minawar · Omar (171)
- Tibetan n – gram Model System
 Xie Wensen Wu Jian (183)

A Study, through Information Management, about the Grammatical Function of Auxiliary Word Formation in the “Empty Words” of Tibetan Language	Suonancairang (193)
Research on Computing Chinese Lexical Chains with Clustering	Lian Ruiting Shi Xiaodong (202)
Information Web Service Composition based on Semantic similarity	Qiu Lirong (212)
Initial Research of The Divide Method of Slavic Mongolian Syllable	Bao Quan (219)
A Study on Tibetan Word Segmentation and it’ s Applications	Cai rang zhuo ma (223)
Design of a Tibetan Automatic Segmentation Schem	Sun Yuan Luo Sangqiangba Yang Rui Zhao Xiaobing (228)
Study of Segmentation Strategy on Tibetan Crossing Ambiguous Words	Sun Yuan Luo Sangqiangba Yang Rui Zhao Xiaobing (238)
Segmentation of Chinese Automatic Based on Viterbi of word evaluation	Li Chengcheng Li Yanqiu (244)

Researches of Machine translation, Character recognition, Text classification, Information extraction, etc

BZD – CTMT summarize	Cai Rang jia (251)
The Research on The Preprocessing of Digital Images in Tibetan Character Recognition System	Liu Fang Ngodrup (257)
Study on Similitude Characters of Tibetan Handwriting Recognition	Sun Yan Wu Jian (265)
The Preprocessing and Feature Extraction of Online Uyghur Handwriting Recognition.	Alimjian. Yasin Halmurat Muhammed (275)
A Classification Approach based on Positive Cases for Chinese Opinioned Subjective Texts	Yao Tianfang Zhang Xin (284)
Design and Implementation of a Platform for Web Information Extracting Based on Rules	Zhu Yi Chunxiao Huang Mengxing (294)