

作者：张双虎 来源：科学时报 发布时间：2008-5-11 22:45:30

小字号

中字号

大字号

于洪志藏文信息处理技术小组：为西藏架起信息天路

从今年5月开始，西藏自治区政府将在全区逐步推广使用藏文政府办公系统。同时，一系列新藏文软件得到广泛应用。中央电视台在5月1日的报道中称：目前藏文信息化处理已达到汉文信息化处理同等水平。这其中，西北民族大学教授于洪志带领的藏文信息处理技术研究小组功不可没。

藏族人民创造了许多宝贵的历史文化资源，要让它们得到保护，必须利用现代化的技术手段。西北民族大学教授、中国民族信息技术研究院院长于洪志说：“人类社会毕竟要进入信息化社会，如果假设藏族地区没有藏文软件，不能进入信息化社会，那么这个民族如何进步？如果不能普及电脑、不能够进行信息化交流，那么教育、科技、文化卫生、企业管理等等都不行，所以必须要有这个现代化的手段。”

1984年，计算机操作系统还是DOS（磁盘操作系统）的天下时，于洪志就开始研制藏文操作系统。后来，在国家自然科学基金等多方面的支持下，于洪志主持参与制定藏文编码国际标准和藏文编码、藏文字型、藏文键盘布局3项国家标准。1997年7月，藏文编码国际标准通过国际标准化组织（ISO）和国际电工委员会（IEC）审定，维护了我国标准化权益。藏文是我国第一个具有国际编码标准的少数民族文字，在国际社会引起强烈反响。

于洪志带领的课题组研制了世界上第一个藏文视窗平台、第一个藏文字处理软件、藏文识别软件等，还建立了世界上第一个藏文网站。1999年12月，世界首家藏文网站——同元藏文网站在西北民族学院建成。该网站突破了藏文系统支持的难题，现在，许多藏族学生也可以和世界各地的学生们一样，通过互联网来了解大千世界的各种信息。

藏文识别属于模式识别领域，因为藏文文稿有自己的特殊性，藏文识别研究和系统开发，只能部分借鉴中文和英文识别中一些成熟的适合于藏文识别的方法和技术，更多的是要根据藏文字、藏文句子、藏文版面自身的特殊性进行全面深入的研究。在此之前，藏文识别方面的论文很少，也都局限于理论探讨，系统开发方面基本上是空白。

藏文识别是重要而比较困难的模式识别问题之一，对藏文识别方法的进一步研究，不仅对模式识别的发展和人类认知的研究具有重要的理论意义，而且具有重要的实用价值。

2000年，清华大学电子工程系教授丁晓青领导的智能图文信息处理研究室和于洪志领导的中国民族信息技术研究院合作，致力于“多字体印刷藏文（混排汉英）文档识别系统”的研究。前者多年致力于东方文字及其与英文混排文档识别的研究，取得一系列重大成果。后者在藏文文字信息处理和系统开发方面有坚实的基础。

经过3年多的联合攻关，该课题组解决了藏文字符不等高、不等宽、相似形字符多的难题，实现了多字体现代藏文以及藏汉英混排文本的识别。该系统具有藏文文档图像输入及版面分析，印刷藏文与汉英混排文本识别，识别后文本编辑（包括文本图像与识别结果对照、识别候选字显示及选择、编辑插入、删除等）功能。鉴定委员会测试组经过测试，藏文白体、黑体等6种字体单字平均识别率达到99.83%，实际藏汉英混排文本的平均识别率达到97.28%以上。

“多字体印刷藏文（混排汉英）文档识别系统”首次解决了多字体印刷藏文文本的识别问题，以及首创实现藏汉英混排文本的识别系统，其整体性能达到了国际领先水平。该系统的成功研制，为藏文纸质文档转化为计算机可查询的电子文档提供了有力的工具，将为藏文文献整理、编辑等数字化工作发

挥重要的作用。因此，这项成果不仅在中文信息技术理论研究方面有重要的学术价值和广泛的应用前景，更具有久远的社会效益。

于洪志一直致力于藏文信息技术研究，2007年，她申请的“基于语音、嗓音和呼吸信号的藏语拉萨话韵律模型研究”又获得国家自然科学基金项目的资助。这项研究将为藏文信息处理和交换走向世界架起一条信息“天路”。

发E-mail给:



[打印](#) | [评论](#) | [论坛](#) | [博客](#)

相关新闻

综合搜索引擎网站市场群雄逐鹿
IPv6：究竟离我们还有多远
第17届国际万维网大会在京召开 万维网之父首次...
北大信息科学技术学院院长梅宏：甘做软件世界的守...
中国研制成功世界首创“多币种”纸币鉴定仪
山东科大虚拟人体漫游技术研究获新突破
AMD发布全球首批三核处理器
国际万维网大会首次在中国召开 将由北航主办

一周新闻排行

198所普通高校本科教学工作水平评估结果公布
中国卓越研究奖5月28日将在北京颁发 24篇论...
霍英东教育基金会第十一届高校青年教师基金资助名...
教育部公布08年具有招生资格的普通高校名单
美宇航员将登陆可能撞地球的小行星 直径仅40米
施一公：我被信仰追问，回国为什么是最好选择
杨振宁谈与丘成桐的分歧
武汉中国地大一优秀大学生校园内被割颈身亡