

- RSS订阅
- 网站地图
- 发布信息

繁體版 | English

[首页](#) [印刷新闻](#) [印刷技术](#) [印刷展会](#) [印刷词典](#) [院校社团](#) [印刷企业](#) [印刷杂志](#) [电子书库](#)

您现在的位置: [首页](#) > [印艺新闻](#) > [行业动态](#) > [我国传统蒙古文印刷体文字识别软件研究取得重大进展](#)

我国传统蒙古文印刷体文字识别软件研究取得重大进展

更新日期: 2016-11-30

我国传统蒙古文印刷体文字识别软件(OCR)研究取得重大进展。据其研发者乌日力嘎博士向新华网内蒙古频道独家披露,目前OCR软件对蒙古文白体的识别率已达到99%以上,对铅印版的识别率也达到95%以上。

随着互联网的高速发展,网络已经成为人们共享信息的一个大平台。“在信息共享的大平台之中,蒙古文信息的传播因为数字化处理的识别限制,变得极为繁琐和无奈。”毕业于内蒙古大学的乌日力嘎博士直言不讳。

“传统蒙古文有大量的图书、报刊,但其中很多都没有实现数字化,”乌日力嘎博士介绍说,高效数字化需要OCR(Optical Character Recognition)软件自动识别扫描或拍照的图片,将图片中的文字转换成文本。

据了解,本世纪初伊始,就陆续有科研机构、相关公司尝试开发传统蒙古文自动识别软件,但因为识别准确率不高、操作繁琐等诸多原因,使用者至今未能等到可广泛使用和可信赖的产品。乌日力嘎博士却凭借自己的执着追求和一己之力,用不到两年时间完成了这一传统蒙古文印刷体OCR。

“读博士期间研究机器翻译,机器翻译需要大量的语言资源,但是蒙古文的资源特别少,我们在建立语言数据库时需要大量的人力、时间和经费,”乌日力嘎博士说:“从那个时候起,我就萌发一个念头,自己是否也试着研发一款高精度的蒙古文自动识别系统。”

2015年夏天博士毕业之后,乌日力嘎开始全身心地投入到传统蒙古文OCR软件的开发之中,她参考了国内外大量学术文献,并使用深度学习和语言模型等高新技术,在短时间内成功研发出了蒙古文高精度自动识别软件。据介绍,这一软件的优势是既能识别扫描图片,又能识别拍照的图片,并且将打字和校对的功能融为一体。

“上世纪90年代以前的书本都是铅印版本,但目前还没有其他可以把铅印版转换成文本的软件,而乌日力嘎博士研究的软件优势是能识别铅印版,并且有承前启后的作用,”内蒙古大学教授那顺乌日图认为,乌日力嘎研究的软件理论上有所创新,这一软件结合语言规则和统计方法,再利用深度学习和神经网络等新技术来完成的。因为以前的OCR软件是基于规则或者是基于统计的方法来实现的,他们最终效果都不理想。

乌日力嘎最后表示,希望她所开发的这款软件能为广大的蒙古语言文字工作者带来更多的工作上的便利。

分享至:

发表评论

电子邮件地址不会被公开。必填项已用*标注

<input type="text"/>	姓名 *
<input type="text"/>	电子邮件 *
<input type="text"/>	站点

[发表评论](#)