



ABBYY FineReader Engine助力 拉脱维亚国家图书馆藏书数字化

“如果没有了图书馆，我们还有什么呢？我们没有过去，没有未来。”

——雷·布莱伯利

珍贵的文本在逐年朽坏

图书馆是通向知识和文化的大门，它能够塑造新的思路 and 观点，是社会创意和创新的源泉，能够确保祖先创造和积累的知识如实地保留下来，图书馆是国之宝藏。

拉脱维亚国家图书馆（NLL）中收藏了450万份纸质资料，包括特殊的收藏——珍贵的书籍、手稿、Letonica（即有关拉脱维亚和拉脱维亚人历史的书籍），波罗的海中央图书馆藏书、地图、曲谱、录音、图形文件、缩印和期刊。自1919年成立以来，NLL中保存的部分最古老的文献已经开始朽坏；与此同时，图书馆基金会收藏了大量宝贵的通俗文献。换句话说，要想为子孙后代保护这些资料，并使公众更容易查阅到这些资料——这个任务需要通过创建数字档案来完成。

大规模数字化带来新的机遇

互联网为人们创造了巨大的契机，让他们得以访问世界上最大的图书馆典藏文献。然而，NLL的大规模数字化还有待实现。该项目的第一阶段为：扫描和生成图像PDF文件，这还有待完善，因为它不能同时处理文字。

为了将材料转换成可搜索的格式，图书馆需要采用OCR技术。但这又有一个问题：很少有OCR解决方案能够提供优质的拉脱维亚文本识别功能，更不要说支持古拉脱维亚和欧洲字体了。然而，它们不久找到了解决方案，归档数字化的第二阶段引入了小规模ABBYY OCR技术应用试点项目。该项目是由内容转换专家（CCS）实施的。

CCS从2000年开始始终参与开发针对文化遗产界的软件解决方案。到了2003年，它们基于ABBYY FineReader Engine科技，针对结构数字化docWorks开发出了新的软件工具，并在随后将其用于NLL项目。

互联网为人们创造了巨大的契机，让他们得以访问世界上最大的图书馆典藏文献。然而，NLL的大规模数字化还有待实现。该项目的第一阶段为：扫描和生成图像PDF文件，这还有待完善，因为它不能同时处理文字。

为了将材料转换成可搜索的格式，图书馆需要采用OCR技术。但这又有一个问题：很少有OCR解决方案能够提供优质的拉脱维亚文本识别功能，更不要说支持古拉脱维亚和欧洲字体了。然而，它们不久找到了解决方案，归档数字化的第二阶段引入了小规模ABBYY OCR技术应用试点项目。该项目是由内容转换专家（CCS）实施的。



名称：
拉脱维亚国家图书馆

办公地点：
拉脱维亚里加

行业：
政府，教育

网站：
www.lnb.lv/en

ABBYY合作伙伴：



项目概述

挑战：
将拉脱维亚国家图书馆的文献转化成可搜索档案

解决方案：
实施基于ABBYY FineReader的解决方案

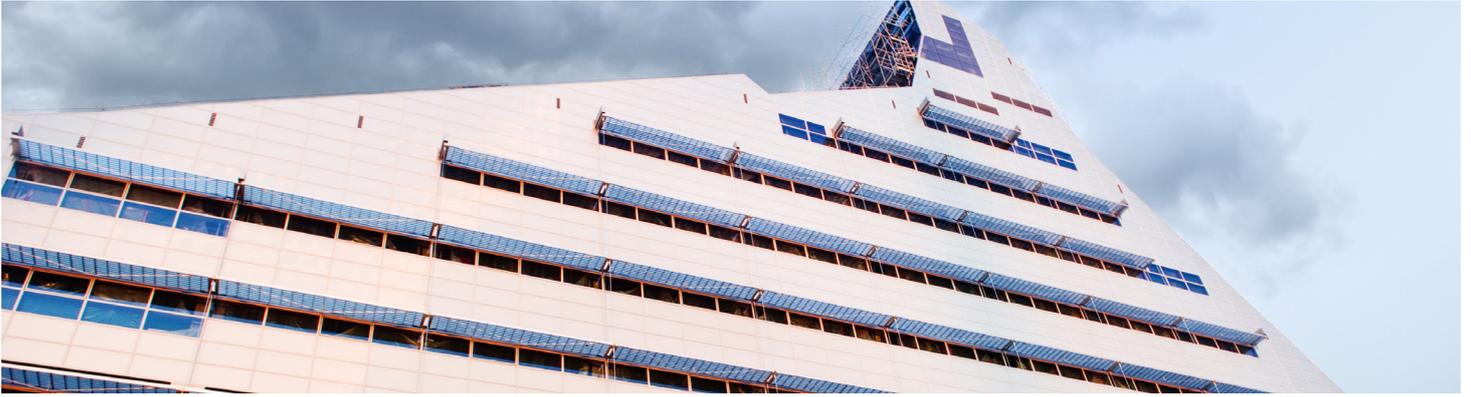
- 结果：**
- 在不到1年的时间内处理了400万页书籍和期刊（包含20种语言）；
 - 现在可在线访问图书馆资料



摘自NLL网站某期刊

“拉脱维亚国家图书馆参与了大型数字化项目，旨在处理大约400万页的历史书籍和期刊，并将它们上线。ABBYY FineReader Engine是这个项目的重要组成部分，它提供了精度极高的OCR结果。项目中大多数文本处理的准确度接近100%。因此，用户既可以使用优质的OCR结果，也可以在期刊门户进行全文搜索：
www.periodika.lv”

Joachim Bauer
CCS集团 docWorks负责人



CCS从2000年开始始终参与开发针对文化遗产界的软件解决方案。到了2003年，它们基于ABBYY FineReader Engine科技，针对结构数字化docWorks开发出了新的软件工具，并在随后将其用于NLL项目。

ABBYY精湛的识别技术

项目初期，NLL挑选了受到损坏以后必须以数字形式加以“挽救”的资料，或者在读者中人气颇高或被认为具有重大历史价值的资料进行数字化。初期项目大致规模包括250万页期刊（大约相当于1000种全套期刊）和150万页书籍（大约相当于7000本书）。

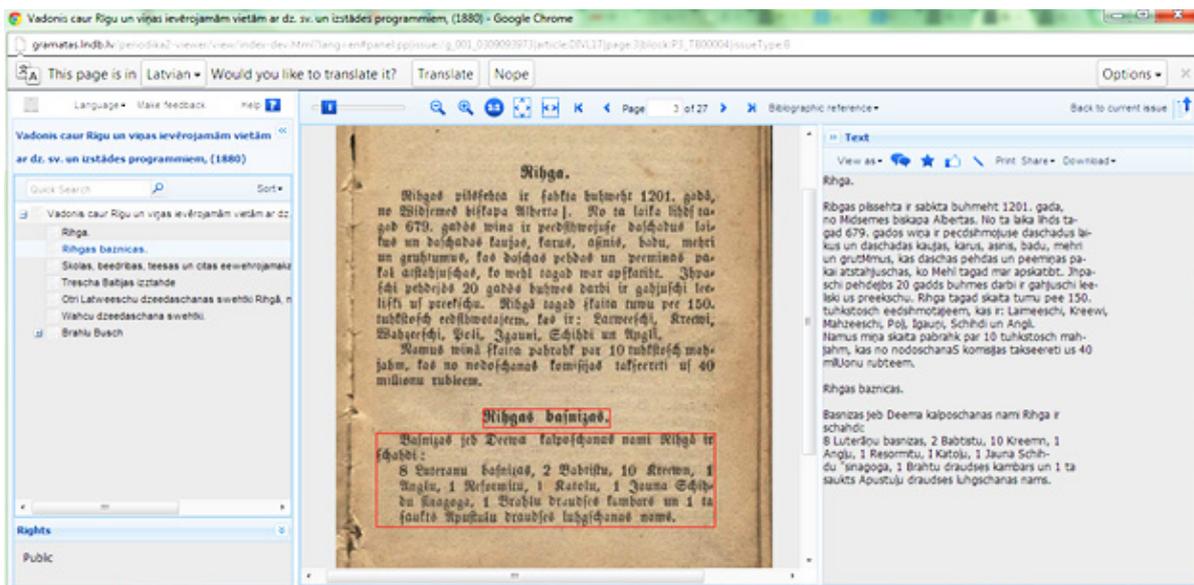
ABBYY FineReader Engine是CCS docWorks解决方案的组成部分，用来对多达20种不同语言的历史文本进行光学字符识别。由于ABBYY FineReader Engine对于拉脱维亚和俄罗斯文本的支持程序几近完美，准确率高达100%。因此成为该项目选择OCR提供商时最为关键的因素。

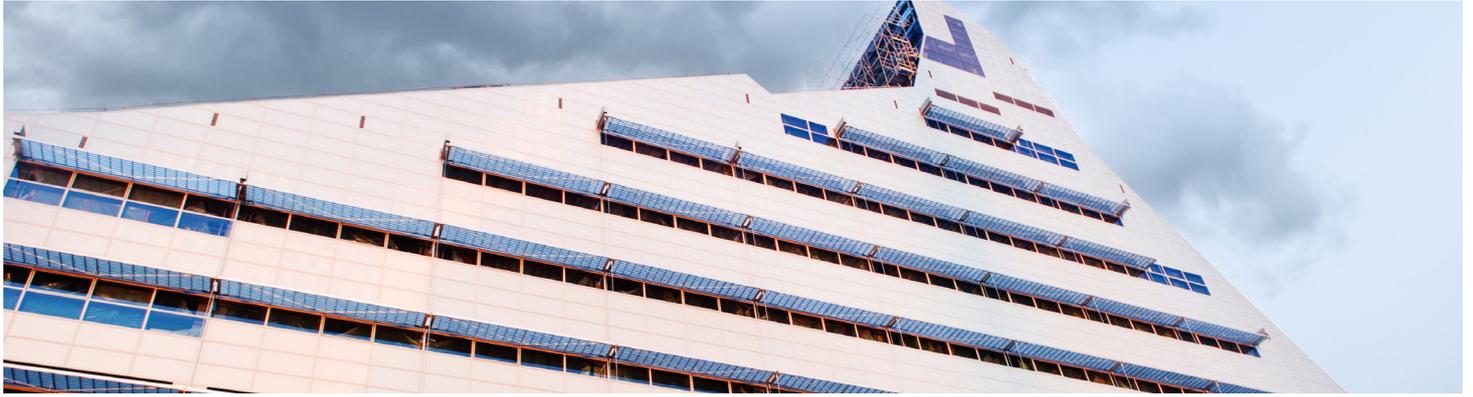
应当指出的是，NLL需要识别的文本中包含着若干罕见的哥特字体，这些字体目前已很少使用，多数最先进的光学字符识别解决方案都不提供支持。然而，ABBYY FineReader Engine技术能够轻松处理这些带有特殊装饰设计的Antiqua 和Fraktur系列字体。

宝藏亮相

NLL花了1年多的时间，处理完了400万页的古籍和现代期刊。怀着致力于崇高目标的热情，在项目高峰期，60个工作人员每天实行三班倒，每班工作8小时。

NLL将文档处理完以后，导出为各种格式（PDF，JPEG，XML），然后导入到期刊门户网站www.periodika.lv。科学家、研究员、教授、学生和广大民众可以通过网络访问这些资料。为了保护版权，大部分材料都只能通过拉脱维亚图书馆的网络访问，不过1941年之前出版的所有期刊都可以无限制访问，公共领域图书（即版权过期的图书）也可以供所有互联网用户访问。





关于NLL

拉脱维亚国家图书馆成立于1919年，如今，它是文化部监管下的全国性文化机构。NLL的使命是：自由而创造性地运用拉脱维亚文化和科学遗产，以便促进教育、研究、知识发展和生活质量。为了交流专业经验和知识，NLL与拉脱维亚和全世界的许多图书馆、组织和机构展开了合作。

更多信息请见 www.lnb.lv/en

关于泰比 (ABBYY)

泰比 (ABBYY) 是一个文档识别、数据捕获及语言技术和服务领域的世界领先供应商。公司设有专门负责南美、亚洲及非洲 (包括中东及波罗的海国家) 市场的泰比3A办公室。自2004年开始，泰比3A执行团队在上述地区不断拓展和发展合作伙伴。泰比产品被应用于诸多大型项目，如：科威特全国人口普查、立陶宛税收稽查系统、越南政府门户网站、智利总统竞选等等。更多信息请访问 www.ABBYY.cn

关于CSS

在过去35年中，内容转换专家 (CCS) 为大型数字化项目和媒体监督开发出了智能化集成软件和服务解决方案。他们的客户包括世界各地最负盛名的图书馆、大型出版商、全球性公司、内容提供商、联邦各部和媒体监测机构。CCS的宗旨是：为他们的客户和业务合作伙伴提供最佳的转换技术和服务，将印刷材料转化为可搜索、可使用的数字化文档，供用户访问。就图书馆来说，它们也有利于为将来保护文化遗产。

更多信息请见 www.content-conversion.com



ABBYY 3A
Asia, Baltic, Middle East, South America, Africa
P.O. Box #32, Moscow, 127273, Russia
Tel: +7 495 7833700
Fax: +7 495 7832663
sales_3a@abby.com
www.ABBYY.cn