



### CMD 利用 ABBYY Recognition Server 提高工作效率

#### 客户概述



#### 客户

CMD Group

#### 办公地点

美国佐治亚州

#### 行业

建筑业

#### 网址

[www.cmdgroup.com](http://www.cmdgroup.com)

#### ABBYY解决方案合作伙伴



[www.Conarc.com](http://www.Conarc.com)

“我们的产品愿景是为客户提供可搜索文件，并在公司内部实现文档自动化处理，而ABBYY是实现这个愿景的关键所在。”

CMD数据策略副总裁Dan DuBois

建设管理数据（CMD）公司旨在通过网站为建筑公司提供海量过往、目前在建和筹划中的建设项目的信息。为此，公司需要收集繁杂的纸质和电子文档，并将其转换为可搜索的图纸和规范化文档，粗略统计，每年CMD需要处理大约3500万页此类文件。但由于CMD自身创建可搜索PDF文档的解决方案前期需要手动输入文档，而且无法处理大型建筑图纸。此外，它无法进行扩展，以适应在处理过往项目数据并转换为可搜索格式的同时，应对当前项目日益增长的文件量及工作量。为此，CMD试图寻找一种能够提高生产率、实现自动化并具有可扩展性的方法，最终它在ABBYY Recognition Server中找到了解决方案。

### 大规模转换以及超大文件的处理依赖于强大的可扩展OCR技术

北美各地的建筑公司在寻求有关市场的商业情报、寻找新的项目线索或寻找投标建议时，都会首先向CMD咨询所需信息。CMD致力于非住宅型商业建设项目，通过网站为用户提供一个拥有大量可搜索项目计划的数据库。“我们收集项目各个阶段的数据——从早期规划，直到签订合同为止。”CMD数据策略副总裁Dan DuBois解释说，“我们的研究人员从各种格式的资料，纸质文档和电子文档中获取数据，利用OCR技术将其转换为可检索的PDF文件，以便汇总到我们的数据库之中。”

正如DuBois所说，CMD每年需要转换海量的繁杂文档：“每年我们要处理大约10万份文件，总共大约3500万页，每页文件都可能包含文字、图表、图画和插图。”许多文件的物理尺寸也非常大。CMD高级软件开发人员Vic Mykulowycz称，公司先前的OCR技术不足以满足建筑行业的需求。“举例来说，它无法处理较大尺寸的文件，而我们需要处理大量44”X36”大小的建筑图纸。所以我们不得不在前期

## 挑战

通过实施能够处理大批施工文件的自动OCR服务,提高效率和数据质量。

## 解决方案

CMD采用ABBYY Recognition Server,自动采集和转换海量施工文件,包括非常复杂、格式超大的项目计划。

## 结果

- 信息的及时更新使得公司的服务质量得到提高
- 可以评估每个送货地区的业绩
- 单据不会变质或遗失
- 降低了时间与成本

“我们的文档转换过程已实现高度自动化,将人工干预降至最低。”

CMD高级软件开发人员  
Vic Mykulowycz

手工录入数据和准备数据。”另外,旧软件的准确度较低,需要进行大量的手工验证。Mykulowycz说:“我们需要通过自动化方案简化流程,减少人工干预。”

## 寻求大规模自动化OCR实施方案

对于CMD来说,新OCR解决方案需要针对所有文件——不管文本和图形如何组合——提供准确的全文搜索功能,这是至关重要的。有鉴于此,他们收到了各种颇具竞争力的解决方案提供商的试用版本,并开始进行一系列的测试。“我们采用多个解决方案处理相同的页面,” Mykulowycz说,“我们会仔细比对导出文本,查看其准确度,并计算每种解决方案的错误次数。结果,ABBYY Recognition Server更准确,速度更快,还有非常重要的一点,它可以处理大量的页面。”

鉴于需要进行转换的文档数量巨大,CMD需要对ABBYY Recognition Server进行严格的测试。因此,他们通过ABBYY合作伙伴——在OCR和文档管理解决方案领域非常知名的CONARC公司——获得了扩展许可证。

## 通过ABBYY Recognition Server寻找答案

“因为CMD的处理量巨大,” CONARC的高级软件开发人员Ben Holton说,“我们设法获得了远远超过常规页面处理量的试用许可证。”获得扩展许可证之后,CMD安装了11个新的物理服务器,开始使用它们运行ABBYY Recognition Server。

我们需要看到该解决方案的实际处理能力有多大,” Mykulowycz回忆说,“以及在十几个服务器上运行核心许可时如何实现优化配置。紧接着,我们进行了广泛的测试。然后,在实施过程中,由于我们的内部软件和新的ABBYY解决方案之间存在技术差异,我们需要进行适当的变更,以适应ABBYY Recognition Server。”按照Mykulowycz的说法,CONARC和ABBYY迅速给予了支持。“他们作出了非常积极的反应。ABBYY甚至开发了特殊的一次性版本,以解决我们的问题。”

最初CMD购买了72核的ABBYY Recognition Server许可证——通过单一的服务器协调管理所有的分布式CPU处理。按照Mykulowycz的说法,结果“符合预期”。但不到1年,该公司决定为先前未转化的材料创建历史档案。正如Ben Holton所说,“这需要大大完善资源。此外,建筑行业的反弹意味着,我们终需使用这些资源来满足当前的建筑需求,还需要进行大幅扩展。”

## 进一步扩展:利用ABBYY Recognition Server

### 实现156核OCR解决方案

根据Mykulowycz的说法,新项目需要将一年半的历史数据转换为可搜索的PDF文件,这是一项艰巨的任务:“要处理3500多万页资料,同时要保证我们当前的生产负荷所要求的自动化OCR精度。”

为了完成历史数据的转换任务,CMD又购买了一个88核许可证,并专门用于该项目。转换任务如期完成;此时又出现了两个此前未考虑到的因素:“我们的业务是季节性的,” Mykulowycz解释说,“从二月到六月,我们都非常繁忙。我们目前的文档处理任务大多发生在这一时期。加上建筑行业回暖,待处理数据将日益增多,这意味着目前的配置不能实现我们所需的高处理量。”

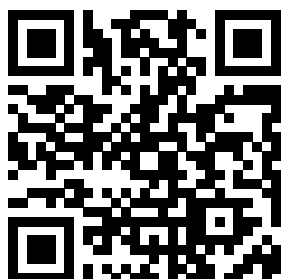
为了处理日益增多的图纸和说明书等文档,CMD将目前所有核心许可证集合到单个主机之下,从而获得了专门用于当前生产的解决方案。这个解决方案仅包含一个系统,该系统采用多个服务器,共包含156个联网内核。Mykulowycz证实说:“事实证明,利用单个服务管理器来处理所有项目工作的做法是成功的。”

## 关于泰比 (ABBYY)

ABBYY是一家领先的文档识别、数据捕获和语言技术与服务供应商。其主要产品包括ABBYY FineReader系列的光学字符识别 (OCR) 应用、ABBYY FlexiCapture 系列数据采集解决方案, 以及众多开发工具。ABBYY Language Services为企业客户提供全面的语言解决方案。来自世界各地的用纸密集型机构都在使用ABBYY解决方案, 对费时费力的任务进行自动化处理, 并简化业务流程。公司在德国、英国、法国、西班牙、乌克兰、塞浦路斯、俄罗斯、美国、加拿大、澳大利亚、日本和台湾设有办事处。欲了解更多信息, 请访问 [www.ABBYY.cn](http://www.ABBYY.cn)

## 关于CMD

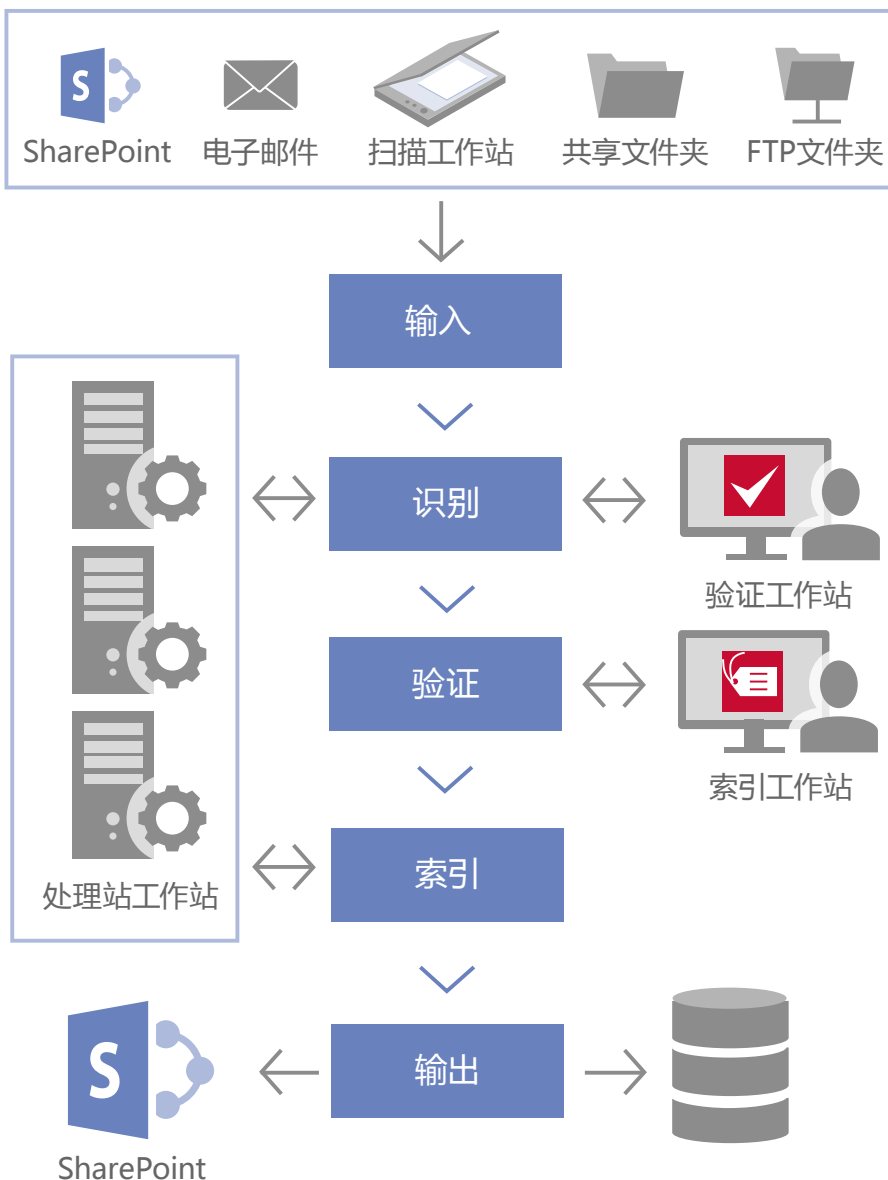
建设管理数据 (CMD) 是北美的建筑信息领先供应商。CMD 的多元化创新产品和服务组合包括: 国家、区域和地方项目线索; 营销解决方案; 以及为美国和加拿大建筑专业人士提供深度洞察的市场情报。



## 结果

在实施CMD基于ABBYY Recognition Server的解决方案7个月之后, 公司完成了上述存档转换项目, 主生产系统的处理能力也提升了。“目前, 该系统处理文件的速度较以前快得多,” Mykulowycz说, “我们的文档转换流程已实现高度自动化, 将人工干预降到最低程度。另一个好处是, ABBYY Recognition Server提供了单词在页面中的坐标位置。当我们搜索文字时, 找到的文字周围会出现红色框——这为用户凸显了搜索结果。”

正如Dan DuBois 证实的那样, ABBYY Recognition Server帮助CMD实现了长期目标: “ABBYY对于我们实现产品愿景和蓝图, 无疑起到了重要作用。无论是在为客户提供可搜索文件的能力方面, 还是实现文档自动化内部处理方面, 都是如此。”



### ABBYY 3A

Asia, Baltic, Middle East, South America,  
Africa  
P.O. Box #32zzz, Moscow, 127273, Russia  
Tel: +7 495 7833700  
Fax: +7 495 7832663  
[sales\\_3a@abby.com](mailto:sales_3a@abby.com)  
[www.ABBYY.cn](http://www.ABBYY.cn)

# ABBYY®

© 2015 ABBYY Software House. ABBYY, the ABBYY Logo, FineReader, ABBYY FineReader, FlexiCapture, Lingvo and PDF Transformer are either registered trademarks or trademarks of ABBYY Software Ltd. Other product and company names mentioned herein may be trademarks and/or registered trademarks of their respective owners and are hereby recognized. Part #7408ch

[www.ABBYY.cn](http://www.ABBYY.cn)