

ABBYY® FineReader XIX



使用ABBYY打开225年历史事迹

弗劳恩霍夫媒体通信研究所(IMK)使用ABBYY FineReader XIX打开了新苏黎世报的档案。

多亏了ABBYY技术，位于德国圣奥古斯丁的弗劳恩霍夫媒体通信研究所(IMK)才能处理整个新苏黎世报(NZZ)的数字化档案，IMK正使用ABBYY FineReader XIX和ABBYY FineReader Engine软件开发工具包，NZZ档案包含200万页各种格式和字体(比如罗马和哥特式字体)的文本，使用ABBYY FineReader XIX，这些文本好比开启全文搜索模式，FineReader XIX是一款光学字符识别(OCR)软件，还可以处理古老的欧洲语言和哥特式打印体。

“这个项目在很多层面上都具有挑战，除了实际规模巨大，文档质量常常很差，还有哥特式和罗马字体的使用问题”，IMK项目经理Stefan Eickeler博士说道，“我们必须制定特殊的解决方案用于一些函数，至于文本识别则使用ABBYY FineReader XIX。该程序具有高度识别精确度，能够处理哥特式打印体，在软件开发工具包的帮助下，能够轻松适应并整合到现有应用程序中。”

文档位于索引卷和缩微胶卷上一大约有1500卷35毫米厚的胶卷。缩微胶卷文档(文本识别的起点)的质量总是改变，照片图像捕捉常常变得扭曲，影像数据随后转换成图像文件用于数字化，IMK开发了自己的软件用于这种过程，几乎消除了扭曲和模糊现象，图像文件构成了材料，供ABBYY FineReader XIX用来识别文本。

ABBYY FineReader XIX结合了著名的ABBYY FineReader 7.0和识别古老欧洲语言与哥特式打印体的所有功能，通过该程序，用户可以扫描、阅读和数字化哥特式打印格式的文档，无需先培训使用该系统，IMK专家们已经将FineReader Engine SDK和FineReader XIX结合到他们的整体解决方案中，在20台电脑上运行，该方案随后获取FineReader的识别结果，将每一页生成XML文件，这些页面的段落标题上包含元数据，或者包含单词的其他排版功能，每个页面要求大约4兆字节数据，整个数字化档案数据库将占10兆兆字节。

“我们很高兴与Fraunhofer研究所合作，”ABBYY欧洲公司总裁Jupp Stoepetie说道，“这个项目展示了我们OCR技术的发展，也因为我们的技术，IMK的数字化项目在技术和经济上实现了可能。几年前基于成本原因，数字化NZZ档案的第一个规划被拒绝了，但是通过使用FineReader XIX，整个过程实现了自动化，并且产生了更高的成本效益，要不是选择识别罗马和哥特式打印体，这个项目也不可能成功。”

关于弗劳恩霍夫媒体通信研究所 IMK
弗劳恩霍夫媒体通信研究所(IMK)是数字媒体技术领域涉及商业、文化、教育和公共部门的创新和发展伙伴，IMK是弗劳恩霍夫公司Iuk组的一名成员，17个弗劳恩霍夫研究所组成的小组，在信息和通信技术(IuK)领域进行研究和开发。

www.fraunhofer.de

关于新苏黎世报

新苏黎世报可以回顾225多年的传统事迹，首期期刊发表于1780年1月12日，今天，NZZ AG已是一家现代化媒体公司，出版质量书刊并跻身于世界领先的瑞士公司。

www.nzz.ch

关于 ABBYY

ABBYY是一家领先的开发商，提供文档识别、文档转换、数据捕获和语言技术。

ABBYY的产品包括：**FineReader**和**PDF Transformer** - 文档转换的终端用户应用程序；**识别服务器** - 一种基于服务器的OCR和PDF转换解决方案；

FlexiCapture - 处理表格、半结构化和非结构化文档的数据捕获程序；

FineReader Engine SDKs - 提供全方位的ABBYY识别技术；以及**Lingvo** - 一款字典软件。