Readiris™ Pro 12

用户指南

目录

版权 1

第 1 章 Readiris 介绍 3

节省时间，无需重新打字输入 3

Readiris 系列产品 6

第 2 章 安装 Readiris 9

系统要求 9

软件安装 9

卸载软件 11

软件注册 11

产品支持 12

第 3 章 入门 13

运行 Readiris 13

使用 OCR 向导 13

用户界面 15

更改用户界面语言 18

第 4 章 Readiris SmartTask 19

使用 Readiris SmartTask 19

第 5 章 扫描文档 23

选择选项 23

打开图像文件 24

扫描纸质文档 26

第 6 章 调整扫描文档 31

第 7 章 将文档保存为图像文件 35

第 8 章 文档识别窗口设定 37

文档识别窗口自动设定 37

文档识别窗口手动设定 39

使用识别窗口设定模板 43

第 9 章 识别文档 45

介绍 45

选择文档语言 46

定义文档属性 48

使用互动学习 50

使用字体库 52

第 10 章 格式化和保存文档 55

格式化文档 55

格式化文本文档 57

格式化基于表格的文档 60

创建 PDF 文档 64

选择 PDF 选项 65

PDF 文档的 iHQC 压缩 66

PDF 文档的转制 68

创建 XPS 文档 68

选择 XPS 选项 70

XPS 文档的 iHQC 压缩 71

选择图形选项 72

第 11 章 保存和载入设置 75

第 12 章 识别多页文档 77

打开和识别多个图像文件 77

扫描和识别多页文档 78

编辑多页文档 80

第 13 章 识别手写文字 83

第 14 章 识别条形码 87

索引 89

版权

ReadirisPro12-dgi-090312-01

版权所有© 1987-2009 I.R.I.S. 保留所有权利。

I.R.I.S. 拥有 Readiris 软件、联机帮助系统以及本出版物的版权。

本文档包含的信息归 I.R.I.S.
所有。其内容随时可能更改恕不另行通知，并且不代表 I.R.I.S
做出任何承诺。本文档描述的软件根据陈述本产品之使用条款的许可协议提供。本软件的使用或复制必须符合该协议中的条款。未经 I.R.I.S.
事先书面同意，不得复制、传播本出版物的任何部分，将它存储在检索系统或翻译为其他语言。

本用户指南使用虚构名称进行说明；如与实际的个人、公司或组织同名，纯属巧合。

商标

Readiris 徽标和 Readiris 是 Image Recognition Integrated Systems S.A.
的商标。
OCR、ICR 和条形码技术由 I.R.I.S 提供
AutoFormat 和 Linguistic 技术由 I.R.I.S 提供
BCR 和区域分析技术由 I.R.I.S 提供
iHQC 压缩技术由 I.R.I.S 提供
XML 解析器由 Apache 开发。本产品包含由 Apache Software Foundation
开发的软件。

本用户指南中提到的所有其他产品是其各自所有者的商标或注册商标。

第 1 章
READIRIS 介绍

节省时间，无需重新打字输入

恭喜您获得了
Readiris！本软件包必能大大地帮助您重新提取文本、表格、图形、条形码和手写文字。

尽管计算机效率高，但您必须先键入信息。如果您曾经重新输入过一份 15
页的报告或一个大的数字表格，您就知道这工作是多么枯燥费时。使用我们这一先进的 OCR
软件包，可自动将纸质文档或扫描的图像文件转换为可搜索和可编辑文字的文档以便存档和共享。有两种识别模式：一种模式确保最大的速度，另一种则保证最佳的 OCR 准确度。

扫描打印的或键入的文档，指示您想让 Readiris 识别的区域 - 或者让系统为您检测识别区域 -
执行字符识别并将文档导出至文字处理程序。一次即可从头到尾处理完包含许多页面的文档。让 Readiris
将您的纸质文档转换为可编辑的计算机文件，几次鼠标点击即胜过数小时的工作：最快可达到手动重新输入速度的 40 倍。

向导指导您顺利地完成操作 Readiris
所需的设置，让您快速简单地获得结果。或者使用 SmartTask
进一步加速进程。您可以将读取结果直接发送至文字处理程序

或电子数据表，将其存档为 PDF 或 XPS
文件等。要识别传真和转换 PDF 文档，请将其图像文件从 Windows 资源管理器拖放到 Readiris
应用程序窗口。或者通过上下文菜单迅速将图像发送至 Readiris。

Readiris
还识别表格式数据，将它们重新创建为电子数据表软件中的工作表或文字处理程序内的表格对象；让您的数字数据立即可用于进一步处理。

Readiris
采用了最先进的识别技术。独立于字体的文本识别与自学技术相得益彰。通过上下文和语言学分析，系统能够学习新的字符和单词。这意味着识别系统的 OCR
准确度将随着用户使用而提高。

Readiris 支持多达 128
种的语言：它支持所有美洲和欧洲语言，包括中欧、波罗的语和使用西里尔字母的语言，以及希腊语和土耳其语。另外，Readiris
可以读取阿拉伯语、波斯语和希伯来语文档以及四种亚洲语言 - 日语、简体中文、繁体中文和韩语。Readiris
甚至还能处理混合的字母：本软件可检测希腊语、使用西里尔字母的语言、阿拉伯语、希伯来语和亚洲语文档中出现的“西方”单词 -
许多无法誊写的专有名词、品牌名称等都是以西方符号书写的。

Readiris
在*识别过程中*而非识别之后使用语言学。因此，Readiris
以最高的准确度识别所有类型的文档，包括低质量的文档、传真和点阵式打印的文档。它漂亮地处理因扫描和复印不佳而包

含过亮或过暗字体形状的文档。连在一起的字符被分解，同时相分离的字符（例如点阵式符号）被重排。

此外，Readiris
有一个（可选的）用户验证功能。启用后，用户验证功能（"互动学习"）不仅标记识别系统不确定的字符，还可提高系统的准确度。Readiris
将记住您确认的所有解析，从而提高系统速度和置信度，使系统随着您的使用变得更加智能。这一强大的学习工具还使您能够训练 Readiris
学习特殊字符，例如数学符号和图像符号以处理变形的字体。

为了进一步提高您的效率，Readiris
不仅识别您的文本，还可以对其进行格式处理。Readiris
提供不同级别的格式化。当您使用“自动格式化”时，Readiris
重新创建扫描文档的精确复本：原始文档的文字、段落和页面格式保持不变。使用相似的字体，在识别过程中，源文档中使用的磅值、字体风格保持不变。分栏、文本块和图形的位置忠实原始文档。Readiris
甚至可以在识别文档中包含扫描页面的背景照片。由于 Readiris
可以轻松地支持灰阶和彩色扫描，您可以取回所有图形 - 无论它们是黑白照片还是彩色插图。如果文档包含表格，Readiris
将在真实的单元格中识别它们，并重新创建原始表格的单元格边框。

换言之，Readiris
允许您存档文档的真实副本，使之成为紧凑且可编辑的文本文件，而非扫描的图像。

扫描页面上出现的条形码也可被读取，手写文字也是如此，只要您写的是适当间隔的“正体字”。

Readiris 兼容
Twain，支持广泛的平板和馈纸型扫描仪、“一体机”设备或“MFP（多功能外设）”以及数码相机。间隔扫描使您能够在扫描仪未配备输稿器的情况下高效地扫描多页文档。

READIRIS 系列产品

下表给出了可用版本的概览：

Readiris Home 12
受限功能
25 种识别语言
支持 PDF、DCX、DJV、DJVU、JPG、JPEG、J2C、J2K、JP2、PNG、TIF、TIFF、BMP、PCX 图像
生成 PDF 图像-文本、DOCX、ODT、WordML、SpreadsheetML。HTM, XML, TXT, TIFF, RTF 输出

Readiris Pro 12	Readiris Corporate 12
基本功能	基本功能
128 基本功能	128 基本功能
支持 PDF、DCX、DJV、DJVU、JPG、JPEG、J2C、J2K、JP2、PNG、TIF、TIFF、BMP、PCX 图像	支持 PDF、DCX、DJV、DJVU、JPG、JPEG、J2C、J2K、JP2、PNG、TIF、TIFF、BMP、PCX 图像

生成四种类型的 PDF 文件、PDF-iHQC（I 级）、四种类型的 XPS 文件、XPS-iHQC（I 级）、DOCX、ODT、XLS、WordML、SpreadsheetML、RTF、HTM、XML、TXT、TIFF 等。	生成四种类型的 PDF 文件、PDF-iHQC（I-III 级）、PDF/A、四种类型的 XPS 文件、XPS-iHQC（I 级）、DOCX、ODT、XLS、WordML、SpreadsheetML、RTF、HTM、XML、TXT、TIFF 等。 大量识别 自动处理 文档索引 名片识别

Readiris Pro 12 Asian	**Readiris Corporate 12 Asian**
基本功能 128 基本功能 支持 PDF、DCX、DJV、DJVU、JPG、JPEG、J2C、J2K、JP2、PNG、TIF、TIFF、BMP、PCX 图像 生成四种类型的 PDF 文件、PDF-iHQC（I 级）、四种类型的 XPS 文件、XPS-iHQC（I 级）、DOCX、ODT、XLS、WordML、SpreadsheetML、RTF、HTM、XML、TXT、TIFF 等。 繁体中文和简体中文识别 日语识别 韩语识别	基本功能 128 基本功能 支持 PDF、DCX、DJV、DJVU、JPG、JPEG、J2C、J2K、JP2、PNG、TIF、TIFF、BMP、PCX 图像 生成四种类型的 PDF 文件、PDF-iHQC（I-III 级）、PDF/A、四种类型的 XPS 文件、XPS-iHQC（I 级）、DOCX、ODT、XLS、WordML、SpreadsheetML、RTF、HTM、XML、TXT、TIFF 等。 繁体中文和简体中文识别 日语识别 韩语识别 大量识别 自动处理

	文档索引 名片识别

Readiris Pro 12 Middle-East*	**Readiris Corporate 12 Middle-East***
基本功能	基本功能
128 基本功能	128 基本功能
支持 PDF、DCX、DJV、DJVU、JPG、JPEG、J2C、J2K、JP2、PNG、TIF、TIFF、BMP、PCX 图像	支持 PDF、DCX、DJV、DJVU、JPG、JPEG、J2C、J2K、JP2、PNG、TIF、TIFF、BMP、PCX 图像
生成四种类型的 PDF 文件、PDF-iHQC（I 级）、四种类型的 XPS 文件、XPS-iHQC（I 级）、DOCX、ODT、XLS、WordML、SpreadsheetML、RTF、HTM、XML、TXT、TIFF 等。	生成四种类型的 PDF 文件、PDF-iHQC（I-III 级）、PDF/A、四种类型的 XPS 文件、XPS-iHQC（I 级）、DOCX、ODT、XLS、WordML、SpreadsheetML、RTF、HTM、XML、TXT、TIFF 等。
阿拉伯语&波斯语识别	阿拉伯语&波斯语识别
希伯来语识别	希伯来语识别
	大量识别
	自动处理
	文档索引
	名片识别

(*) 没有 Mac 版本。

第 2 章
安装 READIRIS

系统要求

使用 Readiris 所需的**最低系统配置**：

- 使用 486 CPU 的英特尔计算机或兼容计算机。推荐使用奔腾 CPU 的计算机。
- 256 MB 内存。
- 120 MB 剩余磁盘空间。

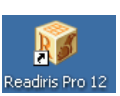
  （如果不安装样本文件，105 MB 的磁盘空间已足够）
- Windows Vista、Windows XP 或 Windows 2000 操作系统。

注意，一些扫描仪驱动程序在最新版本的 Windows 上可能无法工作。请参阅扫描仪随附的文档查找所支持的平台。

软件安装

如何安装 Readiris：

- 以管理员身份登录到
  Windows，或者确保您有必要的管理权限来安装软件。
- 将扫描仪连接至计算机并安装相应的软件。
  测试您的扫描仪。如果遇到问题，请联系您的扫描仪制造商。
- 插入 Readiris 光盘至光驱，按照屏幕上的说明安装软件。
- 单击 **Readiris** 开始安装（亦提供其他软件产品：Copernic Desktop Search Home Edition 和 Cardiris 4 LE）。
- 选择安装语言，然后单击**确定**。
- 接受许可协议条款。
- 软件提供**完整**和**自定义**安装方式。选择所需的选项，每次准备好进入下一屏幕时，单击**下一步**。
  默认情况下将安装所有字库和样本图像，以及电子用户手册和联机帮助。
- 单击**结束**完成安装。
  安装程序自动在 Windows **程序**菜单上创建子菜单 **I.R.I.S. 应用程序 - Readiris**。
  安装程序还在 Windows 桌面创建 Readiris
  应用程序的快捷方式。



- 重复安装过程以安装光盘上的任何其他软件。

卸载软件

卸载 Readiris 的正确方式只有一个：使用 Windows（卸载）安装向导。强烈建议您*不要*通过手动删除程序文件来卸载 Readiris 或其任何软件模块。

要卸载 Readiris：

- 关闭应用程序。
- 在**开始**菜单上，单击**控制面板**。
- 在**程序**图标下，单击**卸载程序**。
- 在列表中选择 **Readiris**，然后单击**卸载**按钮。
- 跟随屏幕指示操作。

软件注册

请记得注册您的 Readiris 许可证。注册后，您可以：

- 收到将来产品开发和相关 I.R.I.S. 产品的信息；
- 获得免费的产品支持；
- 获得 I.R.I.S. 产品优惠。

要进行注册：

使用**注册**菜单上的**注册向导**。跟随屏幕指示操作。

产品支持

一旦完成产品注册后，您即可享受 I.R.I.S.
提供的关于所有基本软件功能的产品支持。I.R.I.S. 联系方式：

欧洲：
support@irislink.com
电话：+32 10 45 13 64

美国：
support@irisusa.com
电话：+1 800 447 4744

亚太地区：
support@irislink.com
电话： +852 22646133

第 3 章
入门

运行 READIRIS

要运行 Readiris：

- 从 Windows **开始**菜单或双击桌面上的快捷方式启动 **Readiris**。

- 单击启动画面上的任意处以启动 Readiris。

  OCR 向导自动打开。

使用 OCR 向导

OCR 向导可让您快速设定操作 Readiris 所需的所有设置。

当启动 Readiris 时，单击启动画面上的任意处以启动 OCR 向导。

步骤 1

选择**图像来源**。
您可以使用**扫描仪**抓取图像或打开**图像文件**。
选择您想使用的**旋转**和**纠偏**选项。

> 更多信息请参阅**选择选项**部分。
>
> 要熟悉 Readiris，请使用软件提供的样本图像。
>
> 它们可在 Readiris 光盘和 Readiris
> 安装文件夹的子文件夹**样本**中找到。

单击**下一步**进入下一个步骤。

步骤 2

如果您选择了扫描仪，请单击**更改**按钮以选择扫描仪设置。

> 有关扫描仪设置的更多信息，请参阅**扫描纸质文档**部分。

单击**确定**以保存设置。
单击**下一步**进入下一个步骤。

步骤 3

单击**更改**按钮以更改文档语言。默认情况下，文档语言被设为美国英语。
在列表中选择所需的语言或语言组合，并单击**确定**。
使用滑块以设定所需的**速度-准确度**设置。

> 更多信息请参阅**选择文档语言**部分。

单击**确定**以保存设置。
单击**下一步**进入下一个步骤。

步骤 4

单击**更改**按钮以更改输出格式或目标应用程序。默认的目标应用程序是 Microsoft Word。
在**发送至**或**外部文件**列表中选择所需的输出格式或应用程序。
单击不同的选项卡，选择所需的选项。不能用于选定格式/应用程序的选项将灰显。

更多信息请参阅**格式化文档**部分。

单击**确定**以保存设置。
单击**下一步**进入下一个步骤。

步骤 5

单击**开始**以打开/扫描和识别文档。

用户界面

要浏览 Readiris 界面，请单击 **Readiris**
启动画面上的任意处，当 OCR 向导启动时单击**取消**。

空白的 Readiris 界面出现。

Readiris 界面包括：

- **SmartTask**（中间）

  SmartTask
  是一些预定义的命令，允许您按一个按钮即可使用最常用的 Readiris 功能。

  使用 SmartTask
  可扫描、识别和发送文档至您选择的目标应用程序或输出格式。

  SmartTask
  应用默认的设置，但可通过单击右键方便地进行配置以适合更特殊的需要。

- **主工具栏**（左工具栏）

  使用主工具栏命令和选项手动扫描和识别文档。
  OCR 向导给出了建议的操作顺序。

- **图像工具栏**（右工具栏）

  使用图像工具栏按钮以在 Readiris 界面中编辑文档。
  指向不同的按钮以显示其工具提示。

在 Readiris 打开或扫描文档后，界面会增加三个主区域：

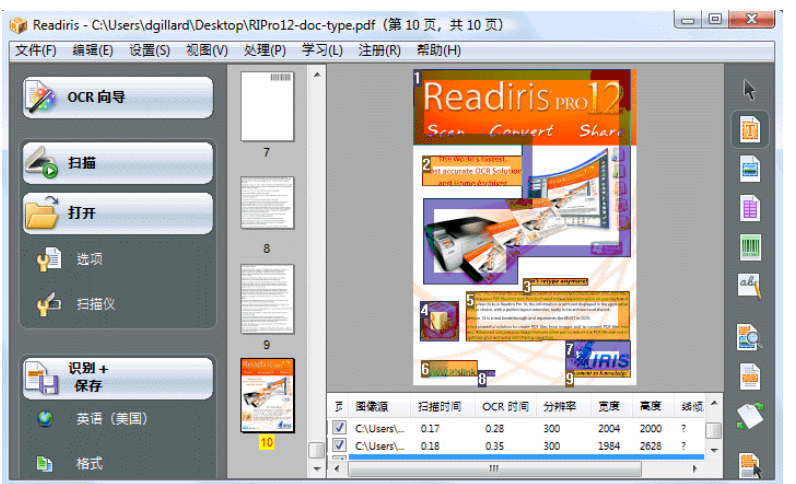
- **页面工具栏**（主工具栏右侧）

  页面工具栏显示页面缩略图，鼠标指向缩略图时会显示设置信息。

- **图像窗口**（中间）

- **文档面板**（底部）

  文档面板显示在 Readiris 中打开的文档的统计信息，例如扫描和 OCR 时间、分辨率、文档的宽度和高度等。



更改用户界面语言

Readiris 提供多种语言的用户界面。

要更改用户界面语言：

- 在**设置**菜单上，单击**用户界面语言**。
- 在**语言**列表中，选择所需语言，然后单击**确定**按钮确认。

注意：如果您选择了不正确的语言，请按 **Ctrl+U**。该操作将打开**语言**对话框，您可以在列表中选择另一语言。

第 4 章
READIRIS SMARTTASK

使用 READIRIS SMARTTASK

当启动 Readiris 时，单击 Readiris 启动画面上的任意处，在 OCR 向导启动时单击**取消**。Readiris SmartTasks 将显示。

SmartTask
是一些预定义的命令，允许您按一个按钮即可使用最常用的 Readiris 功能。

不同的 SmartTask 按钮允许您：

1. 扫描和识别文档并将其直接发送至 **Word** 以便进行文字处理；

Microsoft Word
是默认的目标应用程序。请参阅**格式化文本文档**部分以进一步了解其他可用的应用程序。

2. 扫描和识别文档并将其直接发送至 **OpenOffice** 进行文字处理；

OpenOffice.org Writer
是默认的目标应用程序。请参阅**格式化文本文档**部分以进一步了解其他可用的应用程序。

3. 扫描和识别表格并将其直接发送至 **Excel** 和其他电子数据表；

Microsoft Excel
是默认的目标应用程序。请参阅**格式化基于表格的文档**部分以了解更多信息。

4. 扫描和识别文档并将其**存档**为 **PDF** 文件；

Adobe Acrobat PDF 图像-文本是默认的输出格式。请参阅**创建 PDF 文档**部分以进一步了解其他可用的格式。

5. 扫描和识别文档并将其**存档**为 **XPS** 文件；

XPS 图像-文本是默认的输出格式。请参阅**创建 XPS 文档**部分以进一步了解其他可用的格式。

6. 扫描和识别文档并直接通过**电子邮件**发送它们。

默认情况下，文档将以 PDF 图像-文本格式由您默认的电子邮件应用程序发送。请参阅**格式化文档**部分以进一步了解其他可用的格式。

首次使用 Readiris 时，您必须配置 SmartTask。

要配置 SmartTask：

- 右键单击您想使用的 SmartTask。
- 选择**扫描仪**或**图像文件**作为图像来源。
    - 当使用**扫描仪**时，Readiris 将在您单击 SmartTask 时立即启动扫描仪。扫描的文档将在软件界面中显示、处理和保存。

      为了使用 SmartTask，您必须正确配置您的扫描仪。

      操作如下：

      - 单击主工具栏上的**扫描仪**按钮。
      - 单击**扫描仪型号**并从列表中选择您的扫描仪。

        如果列表中没有您的扫描仪，请选择 **Twain 其他型号**。

      - 在适用情况下，单击配置以选择 Twain 源。
      - 然后单击**确定**以保存设置。

        要了解扫描仪设置和扫描纸质文档的更多信息，请参阅**扫描纸质文档**部分。

    - 如果选择**图像文件**，单击 SmartTask，Readiris 将打开**输入**对话框让您选择要处理的图像文件。

      有关打开图像文件的更多信息，请参阅**打开图像文件**部分。

- 单击**配置**以更改输出**格式**及其**选项**。

注意，可用的输出格式和选项取决于选定的 SmartTask。

请参阅**格式化文档**部分以进一步了解可用的格式和选项。

- 完成配置 SmartTask
  后，使用主工具栏上的按钮以指定**语言**设置和图像增强**选项**，有可能还需要继续指定**扫描仪设置**。

要了解上述设置的的更多信息，请参阅**选择选项**、**扫描纸质文档**和 **选择文档语言**部分。

- 最后，单击 SmartTask 以进行使用。

Readiris 将自动完成整个识别过程。

第 5 章
扫描文档

选择选项

在扫描纸质文档或打开图像文件之前，您可以选择若干图像增强选项。启用后，在打开或扫描文档时将应用这些选项。

操作

- 单击主工具栏上的**选项**按钮以选择若干图像增强选项。

- Click单击**页面纠偏**以校正倾斜的扫描页面。
- 指向**旋转**按钮，确定您是想让 Readiris **自动旋转**，还是**向左旋转 90 度**、**向右旋转 90 度**或**旋转 180°**。

  注意，这两个选项会在一定程度上减缓扫描过程。仅在必要时才选择它们。

- 默认情况下已启用**页面分析**。

  这样，扫描的或打开的图像将被自动分离到窗口中。

您也可以使用图像工具栏上的识别窗口设定工具来修改页面分析结果或手动为文档设定识别窗口。更多信息请参阅**文档识别窗口手动设定**部分。

- 当完成所有这些设置（**扫描仪设置**、**选项**）后，单击**扫描**或**打开**按钮以扫描文档或打开图像文件。

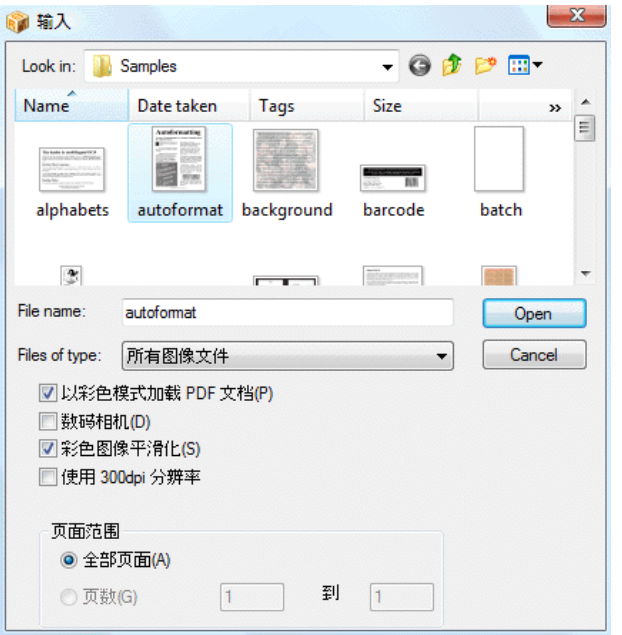
  注意，**设置**菜单上也提供上述选项。

打开图像文件

通过 Readiris，您可以处理用扫描仪扫描的纸质文档，也可以处理已有的不同格式的图像文件。

要打开已有的图像文件：

- 单击**打开**按钮以搜索图像文件。



提示：您也可以将图像文件**拖放**到 Readiris 图像窗口以打开它们。

提示：右键单击要打开的任何图像文件，指向**打开方式**并单击 **IOCR 应用程序**。Readiris 软件将打开并显示图像。

提示：当载入多页图像文件（TIFF 图像和 DCX 传真）和 PDF 文档时，您可以指定页面范围（例如，如果您只需要文档的某一章）。要指定页面范围，请单击主工具栏上的**打开**。在**页面范围**区域，选择**页面**并输入要载入的页面。另请参阅**打开和识别多个图像文件**。

提示：当处理 PDF 文档时，为了加快载入进程，请单击**打开**按钮并取消选择**以彩色模式加载 PDF 文档**。

- Readiris 支持以下图形格式：

图形格式	文件扩展名
Adobe Acrobat PDF	*.pdf
DCX 传真	*.dcx
DjVu 图像	*.djv, *.djvu
JPEG 图像	*.jpg, *.jpeg
JPEG 2000 图像	*.j2c, *.j2k, *.jp2
便携式网络图像	*.png
（压缩的多页）TIFF 图像 未压缩的以及 LZW、PackBits、Group 3、Group 4 和 JPEG 压缩	*.tif, *.tiff
Windows 位图	*.bmp
ZSoft Paintbrush 图像	*.pcx

- 选择所需的图像文件并单击**打开**。

注意：**输入**对话框的选项也适用于文档扫描，它在**扫描纸质文档**部分讨论。

扫描纸质文档

通过
Readiris，您可以处理用扫描仪扫描的纸质文档，也可以处理已有的不同格式的图像文件。

当处理纸质文档时，一旦您单击**扫描**按钮，Readiris
将启动您的扫描仪并在软件界面中显示扫描的文档。

要扫描文档：

- 单击**扫描仪**按钮以设置扫描仪。

注意，**扫描仪**对话框中的几个选项同样也存在于**打开**对话框中。

- 选择正确的扫描仪型号。

  如果列表中没有您的扫描仪，请选择 **Twain 其他型号**并单击**确定**。

- Readiris 几乎支持所有平板和馈纸型扫描仪、"一体机"设备（"MFP" - "多功能外设"）、数码相机和当前可用的扫描仪标准。Readiris 兼容 Twain。

  注意，**配置**按钮仅在适用的时候才可用。单击该按钮以选择 Twain 源。

- 选择扫描仪设置。

格式和分辨率

Readiris 支持广泛的纸张格式和分辨率。

注意，推荐使用 300 dpi 的扫描分辨率。

当识别亚洲文本或极小字体时使用 400 dpi 的分辨率。

颜色模式

Readiris
可以扫描文档并以**彩色**、**黑白**和**灰阶**模式打开图像文件。

对比度 - 亮度

使用滑块以确定合适的亮度和对比度设置以便获得最佳的扫描结果。

为 OCR 优化分辨率

选择**为 OCR 优化分辨率**以纠正细节过多（超过 600 dpi）的扫描图像的分辨率。Readiris 将降低分辨率。

注意，此选项不会提高细节过少的扫描图像的分辨率。

扫描多页文档

当扫描多页文档并使用配备输稿器的扫描仪时，请选择 **ADF** (自动输稿器)
选项。将要扫描的页面放入输稿器然后开始扫描。

或者在使用平板扫描仪时使用**间隔扫描：**选择**继续扫描前等待**并指定您想在几秒后让 Readiris 扫描另一页面。

更多信息请参阅**扫描和识别多页文档**部分。

重要：不能用于选定扫描仪的所有选项将灰显。

扫描横向图像

当扫描横向图像时，请选择**横向**选项。

自动曝光

一些扫描仪已默认选择了**自动曝光**选项。此选项自动调整对比度 和亮度设置。

扫描反转的图像

当扫描反转的黑白图像时，请选择**反转**选项。

使用数码相机

当使用数码相机作为扫描源时，请选择**数码相机**。Readiris 使用特殊的识别例程处理数码相机图像。

以数码相机作为扫描源的相关提示：

- 通过拍摄白色文档校准相机。
- 总是选择最高的图像分辨率。
- 启用相机的微距模式以进行近照。
- 只使用光学变焦而非数码变焦。
- 直接在文档上方拿住相机。避免以某一角度拍摄文档。
- 拍摄稳定的图像。必要时使用三脚架。
- 拍摄光面纸时避免使用闪光。
- 避免打开经压缩的相机图像。
- 根据周围环境（日光、灯光、霓虹灯）调整 Readiris 的亮度和对比度设置。
- 选择彩色或灰阶作为颜色模式。

使用 300dpi 分辨率

当处理分辨率不正确或未知的图像时，请选择**使用 300 dpi 分辨率**。处理图像时就好像图像的分辨率为 300 dpi。

数码相机图像的分辨率基本上都是未知的。

彩色图像平滑化

默认情况下已选择此选项，因为一些扫描仪需要图像平滑化以成功识别彩色和灰阶图像。

- 当完成所有这些设置（**扫描仪设置**、**选项**）后，请单击**扫描**以扫描文档。

注意：请留意线倾斜。超过 0.5° 的线倾斜将增加 OCR 错误的风险。

第 6 章
调整扫描文档

当打开或扫描极亮或极暗的灰阶和彩色图像时，为了获得满意的 OCR 结果，在执行识别之前可能需要对图像进行调整。

要调整图像：

- 打开或扫描彩色-灰阶文档。

  确保**扫描仪设置**是正确的。

  重要：**扫描仪设置**和**调整设置**看起来一样，但请注意，两套设置是在扫描/识别过程的不同阶段中使用的。

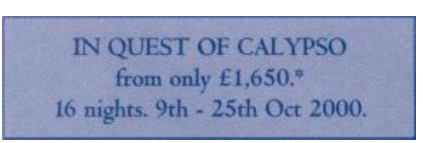
- 在**处理**菜单上，单击**调整图像**。

  Readiris 使用智能**二值化**例程将彩色-灰阶图像转换为用于执行 OCR 的黑白图像。

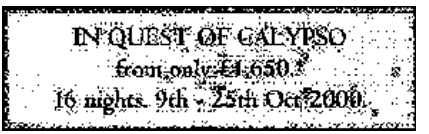
o 选择**彩色图像平滑化**使图像均匀化。

  通过减少密度差异，此选项使灰阶和彩色图像更均质化。结果，前景（文本）和背景（图片）之间的对比度更强。

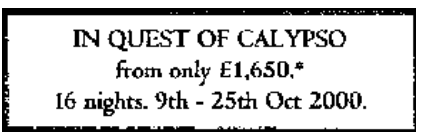
  注意：有时候平滑化是将文本从有色背景分离的唯一方法。



（原始图像）



（二值化的黑白图像）



（平滑化的图像）

- 使用滑块增加或减少**亮度**。

  亮度设置决定图像的整体亮度。当文字难以辨认时，请使用这些设置调暗或调亮图像。

  示例 1：调亮较暗的图像以去除页面背景。

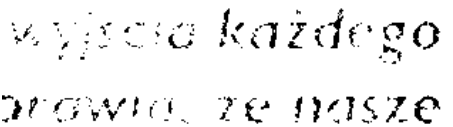
（彩色图像）

（二值化的图像。默认的二值化设置生成一个黑色图像）

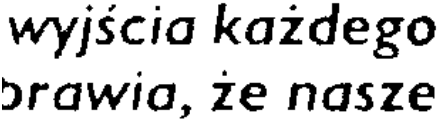
（调亮的图像生成令人满意的识别结果）

示例
2：当文字过亮以至于在二值化的图像中不显示时，调暗图像。

（彩色图像）



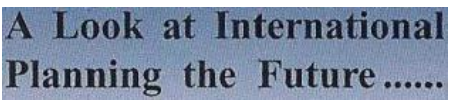
（二值化的图像。默认的亮度设置生成不完整的字符）



（调暗的图像生成令人满意的识别结果）

- 使用滑块增加或减少**对比度**。

  对比度设置决定图像较暗和较亮区域之间的对比度。使用这些设置使字符形状在有色背景中更为突出。



（彩色图像）

A Look at International
Planning the Future .....

（默认的对比度设置生成不完整的字符）

A Look at International
Planning the Future .....

（提高对比度设置生成令人满意的识别结果）

- 使用滑块增加或减少**除斑**选项。

  除斑操作从黑白图像中除去小斑点。

- 单击**应用**预览结果。
- 如果结果令人满意，请单击**确定**。否则，请再次更改设置。
- 单击**识别 + 保存**以识别文档。

第 7 章
将文档保存为图像文件

您扫描的纸质文档无需马上进行 OCR
识别。可以将其保存为图像文件。

操作如下：

- 扫描文档。
- 在**文件**菜单上，单击**以图像方式保存整页或以图像方式保存所有页面**。
- 随后，打开保存的图像文件并执行识别。

仅保存图形

您也可以选择保存文档中不含文本的图形窗口。

操作如下：

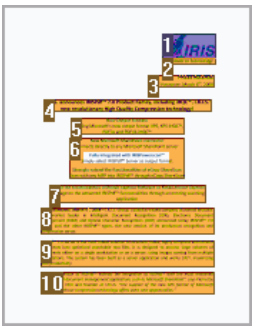
- 扫描或打开文档。
- 在**文件**菜单上，单击**保存图形**。
- 文档中的所有图形将被保存在一个文件中。

第 8 章
文档识别窗口设定

文档识别窗口自动设定

当扫描或打开文档时，Readiris
将自动应用**页面分析**将文档分割为不同的识别窗口。

> 默认情况下，**页面分析**选项已被选中。如果您不想进行自动页面分析，请单击**选项**按钮并禁用**页面分析**。
>
> 在自动页面分析后可手动修改页面分析结果。更多信息请参阅**文档识别窗口手动设定**部分。



窗口类型

Readiris
使用五种窗口类型：**文本块**、**图形区域**和**表格**、**条形码区域**和**手写区域**。

页面分析自动检测文本、图形和表格区域。条形码区域和手写区域则需要手动框选。

更多信息请参阅**文档识别窗口手动设定**部分。

每个窗口类型都有自己的颜色代码：文本窗口为橘黄色，图形窗口为紫色，表格窗口为粉红色。条形码区域为绿色，手写文字区域为蓝色。

窗口从上至下，从左至右排列。数字表示窗口的排列顺序。不过，您可以更改排列顺序和窗口类型。更多信息请参阅**文档识别窗口手动设定**部分。

忽略页面边框上的文本区域

当您的扫描仪在实际图像周围生成黑色边框时，页面分析会查找只有噪声的区域。

为了避免此情况，请单击**设置**菜单上的**不检测边框上的窗口**并再次扫描文档。

略过未选取区域

除了自动设定文档识别窗口，还可以使用**略过未选取区域**功能。当文档页面上只有一个特定区域需要进行 OCR
时，此功能很有用。

单击图像工具栏上相应的按钮选择**略过未选取区域**。

框选您想让 Readiris 识别的页面部分。

您选择的页面部分将被自动地分析。

软件将提示您是否不对文档每一页的相同未选取区域进行页面分析。

文档识别窗口手动设定

除了使用**页面分析**进行文档识别窗口自动设定之外，Readiris 还允许您为文档手动设定识别窗口。

当需要修改自动页面分析结果时，识别窗口手动设定就变得十分有用。当创建识别窗口设定模板时此功能也很有用。

有关识别窗口设定模板的更多信息，请参阅**使用识别窗口设定模板**部分。

注意，条形码和手写文字区域总是需要手动设定识别窗口。

操作

- 为了手动为文档进行识别窗口设定，首先单击**选项**按钮，然后取消选择**页面分析**。
- 通过单击**扫描**或**打开**按钮打开或扫描文档。

- 通过单击图像工具栏上对应的窗口类型按钮，选择正确的**窗口类型**。

Readiris
使用五种窗口类型：**文本块**、**图形区域**、**表格**、**条形码区域**和**手写区域**。

- 围绕您想设定识别窗口的文本块、图形、表格、条形码和手写区域画一个框。

  有关识别条形码和手写文字的更多信息，请分别参阅**识别条形码**和**识别手写文字**部分。

- 当完成文档的识别窗口设定时，单击**识别 + 保存**按钮以执行 OCR。

排列窗口

要更改窗口的排列顺序，请单击图像工具栏上的**排序**按钮，然后按所需顺序逐个单击窗口。

操作完成后，单击**识别 + 保存**按钮以执行 OCR。

软件将不会识别您没有单击的窗口。

绘制多边形

文档识别窗口的手动设定不限于矩形。通过合并多个矩形，您可以创建多边形窗口。当两个相同类型的窗口相交时，它们自动变成多边形。

自动页面分析

万一当前页面过于复杂而无法手动设定识别窗口，请单击图像工具栏上的**分析页面**按钮以自动设定页面的识别窗口。

注意，条形码和手写文字区域总是需要手动设定识别窗口。

更改窗口类型

要更改窗口类型，请右键单击窗口，指向**窗口**，再指向**类型**，然后单击所需的窗口类型。

您还可以同时更改若干窗口的窗口类型：

- 单击图像工具栏上的**指针**按钮。

- 在选择多个窗口的时候按住 **Shift** 键。
- 右键单击选定的窗口，指向**窗口**，再指向**类型**，然后单击所需的窗口类型。

修改窗口大小

单击您想修改的窗口。

将鼠标指针放在标记上（在窗口的各边和角落）。

单击标记并拖动鼠标以修改窗口大小。

移动窗口

选择您想移动的窗口。

单击窗口内部并拖动鼠标以修改窗口的位置。

识别特定的窗口

右键单击您想识别的窗口并选择**复制为文本**。

> 结果将被作为正文文本发送至剪贴板。这对于手写文字同样有效。
>
> 图形窗口和条形码窗口也可被复制到剪贴板。

删除窗口

选择您想删除的窗口或单击**编辑**菜单上的**全选**命令以选择所有窗口。

选择**编辑**菜单上的**剪切**或**清除**命令以剪切或删除窗口。

或

右键单击选定的窗口，指向**窗口**，然后单击**删除**。

删除小窗口

例如，一些文档或传真的页面上常常有"乱七八糟的"点，导致 Readiris 创建过多不含文本的窗口。

要删除所有小窗口，请单击**编辑**菜单上的**删除小窗口**命令。

此选项将删除所有小于 0.5 "的窗口并重新排列剩余的区域。

使用识别窗口设定模板

当识别许多版面设置类似的文档时，使用**识别窗口设定模板**而非自动页面分析可能很有用。那样，相同的窗口结构被应用至所有扫描的或打开的文档，使进程加速。

操作

- 通过使用图像工具栏按钮手动对文档首页进行识别窗口设定。

  更多信息请参阅**文档识别窗口手动设定**部分。

- 在**文件**菜单上，单击命令**保存版面设置**。
- 通过单击主工具栏上的**打开**或**扫描**按钮以打开或扫描文档的其他页面。
- 在**文件**菜单上，单击命令**加载版面设置**。
- 选择您保存的版面设置文件。
- 要将版面设置应用至**所有**打开的或扫描的页面，请在**版面设置文件**对话框中选择**将版面设置应用至所有页面**。
- 单击**打开**以加载版面设置文件。

  注意，当您添加文档至 Readiris 时，由于默认情况下启用了页面分析，因此必须再次加载版面设置文件。

略过未选取区域

除了使用识别窗口设定模板，您还可以使用选项**略过未选取区域**。那样，您可以限定页面上需要进行 OCR
识别的一个特定区域。 从而不识别 OCR 区域外的任何数据。

操作

- 通过单击图像工具栏上相应的按钮，选择**略过未选取区域**。

- 围绕您想让 Readiris 识别的页面部分画一个框。

  您选择的页面部分将被自动地分析。

  软件将提示您是否要略过该文档所有页面的相同未选取区域。

  要取消此功能，请单击图像工具栏上的**分析页面**按钮重新执行**页面分析**。

- 单击**识别 + 保存**以执行 OCR。

第 9 章
识别文档

介绍

Readiris 在识别阶段应用语言学来识别文档。因此，Readiris 可识别所有类型文档中的文本、表格、图形、条形码和手写文字。Readiris 甚至能够处理复杂的分栏文档、低品质文档、传真、点阵式打印稿、包含过亮或过暗字体形状的扫描/复印不佳的文档等。

Readiris 支持 128 种**语言**：它支持所有美洲和欧洲语言，包括中欧、波罗的语和使用西里尔字母的语言，以及希腊语和土耳其语。Readiris 可以随意地读取阿拉伯语、波斯语和希伯来语文档以及四种亚洲语言 - 日语、简体中文、繁体中文和韩语。Readiris 甚至还能处理混合的字母：本软件可检测希腊语、使用西里尔字母的语言、阿拉伯语、希伯来语和亚洲语文档中出现的“西方”单词 - 许多无法誊写的专有名词、品牌名称等都是以西方符号书写的。

Readiris 采用的是最先进的识别技术。独立于字体的文本识别与自学技术相得益彰。通过上下文和语言学分析，系统能够学习新的字符和单词。这意味着识别系统的 OCR 准确度将随着用户使用而提高。

此外，Readiris
还有（可选的）用户验证功能。启用后，用户验证功能（**互动学习**）不仅标记识别系统不确定的字符，还可提高系统的准确度。识别过程中将临时记住您确认的所有解析，从而提高系统速度和置信度，使系统随着您的使用变得更加智能。这一强大的学习工具还允许您训练 Readiris
学习诸如数学符号和图像符号之类的特殊字符，以及处理变形的字体。

互动学习结果还可以永久地存储在**字体库**中以便将来使用。

选择文档语言

Readiris 支持 128 种语言的 OCR
识别。它支持所有美洲和欧洲语言，包括中欧、使用西里尔字母的语言和波罗的语，以及希腊语和土耳其语。

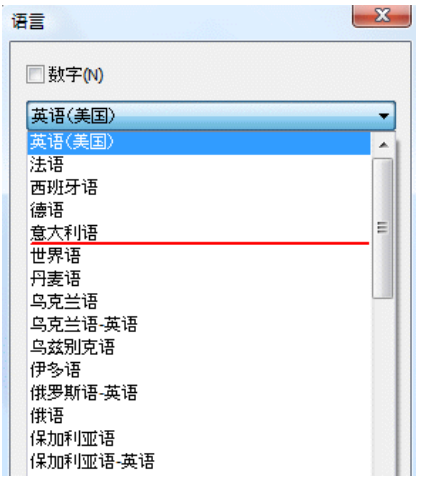
Readiris Pro Asian 和 Readiris Corporate Asian
还可以识别使用日语、简体中文、繁体中文和韩语的文档。

Readiris Pro Middle-East 和 Readiris Corporate Middle-East
还可以识别使用阿拉伯语、波斯语和希伯来语的文档。

Readiris 支持 25 种语言的 OCR 识别。 为了让 Readiris
识别文档，用户必须指定文档语言。

操作如下：

单击主工具栏上的**地球**按钮并在下拉列表中选择所需的语言。



最近选择的 5 种语言被移到语言列表的顶部。

重要：当处理亚洲语、希伯来语和阿拉伯语文档时，请在执行页面分析*之前*选择文档语言。这些文档将使用特定的页面分析例程。

也可以将识别限制为**数字**字符集以最好地识别表格和数字。那样，Readiris 将只识别数字 0-9 和下列符号：

+ 加号	* 星号	/ 斜杠	% 百分号	, 逗号
. 句号	(左圆括号	) 右圆括号	- 连字号	= 等号
$ 美元符号	£ 英镑符号	€ 欧元符号	¥ 日元符号	

要激活数字模式，请在**语言**对话框中选择**数字**。

识别含有混合语言的文档

Readiris 还允许您使用混合的字符集。通过该方式，Readiris 在句子中间自动切换语言，识别"非英语"语言中出现的英语单词（专有名词等）。

单击主工具栏上的**地球**按钮并在语言下拉列表中选择所需的语言组合。

注意：当处理亚洲语或希伯来语文档时，将自动使用混合的字符集。

速度与准确度

请在 OCR 速度和 OCR 准确度之间选择正确的折衷。

此折衷可用于拉丁、西里尔和希腊字母。

提示：当图像质量相当差时，首选准确度而非速度。

定义文档属性

除了文档语言，诸如**字体类型**和**字符间距**之类的其他文档属性在识别过程中也起着重要的作用。

字体类型

Readiris 区分"常规"和点阵式打印的文档。（9
针类型的）点阵式符号由独立分开的点组成。

需要使用特别的分割和识别技术识别点阵式文档，并且需要激活。

Far out in the uncharted back

要选择字体类型：

- 在**设置**菜单上，将鼠标指向**字体类型**。
- 默认情况下，字体类型被设为**自动**。

  那样，Readiris 可识别 "25 针"或
  "NLQ"（仿信函质量）的点阵或其他"正常"的打印。

- 如果只想识别点阵式打印的文档，请单击**点阵式**。

  Readiris 将识别所谓的"草稿"或 "9 针" 点阵式打印文档。

字符间距

字符间距是字体间每英寸的字符数量。字符间距可以是**等宽的**（这种情况下，所有字符宽度相同），或者**非等宽的**（这种情况下，字符有不同的宽度）。

要选择字符间距：

- 在**设置**菜单上，将鼠标指向**字符间距**。
- 默认情况下，字符间距被设为**自动**。

- 如果字体的所有字符宽度相同，请单击**等宽**。老式打字机文档常常是这种情况。
- 如果字体的字符宽度不同，请单击**非等宽**。报纸、杂志和书籍中几乎所有字体都是非等宽的。

重要：这些文档属性不适用于亚洲语、希伯来语或阿拉伯语文档。

使用互动学习

Readiris
提供互动学习功能。通过**互动学习**，您可以在字体和字符形状方面训练识别系统，必要时纠正 OCR
结果。在互动学习过程中，识别系统不确定的任何字符将在预览窗口中显示，同时显示其原单词和提议的解析。

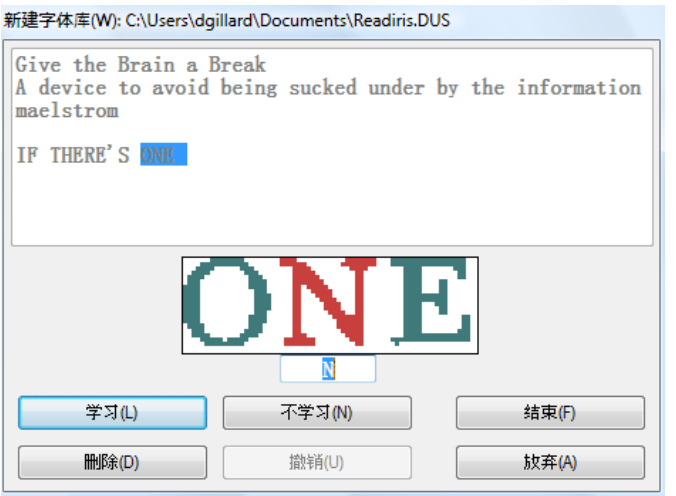
互动学习可大大地提高识别系统的准确度，在识别变形的、有污损的表单时特别有用。通过互动学习功能，还可在最初无法识别的特殊符号（例如数学、科学符号和图像符号）方面训练 Readiris。

要启用互动学习：

- 在**学习**菜单上，单击**互动学习**。
- 定义必要的设置并单击**识别 + 保存**按钮以识别文档。

- 识别结束后，Readiris 进入互动学习阶段。

  识别系统不确定的字符将显示出来。



如果结果是正确的：

- 单击**学习**按钮，将结果作为确定的字符保存下来。

  在识别期间，学习结果将临时存储在计算机内存中。在对文档的其余部分进行 OCR 时，Readiris 将不再显示学到的字符。

  当对新的文档进行 OCR 时，学习结果被删除。

  要永久保存学习结果，请使用字体库。更多信息请参阅**使用字体库**部分。

- 单击**结束**以保存软件提供的所有解析。

如果结果是错误的：

- 输入正确的字符并单击**学习**按钮。

或

- 单击**不学习**按钮将结果作为不确定的字符保存下来。

  对于如果学习后会与其他字符相混淆的损坏字符，请使用此命令。例如，数字 1 和字母 I，它们在许多字体中形状完全相同。

- 单击**删除**以将字符从输出结果中删除。

  使用此按钮以避免文档噪音出现在输出文件中。

- 单击**撤销**以更正错误。

  Readiris 保留最后的 32 个操作记录。

- 单击**放弃**以放弃互动学习。

  这将删除所有学习结果。下次您单击**识别 + 保存**时，互动学习将再次启动。

使用字体库

当扫描多个相同类型、字体质量和印刷质量的文档时，您可能不想每次都重复学习过程。因此，使用字体库很有用。字体库包含在**互动学习**过程中学到的字体信息，可以实质性地提高识别结果。

注意，字体库限制为 500
个形状。建议您为具体的应用创建单独的字体库。

要创建新的字体库：

- 在**学习**菜单上，单击命令**新建字体库**。

默认情况下，Readiris 将打开字体库 **Readiris.dus**。更改文件名并单击**保存**以将其保存为新的字体库。

- 单击**学习**菜单上的**互动学习**以激活此功能。
- 单击**识别 + 保存**识别文档。
- Readiris 进入互动学习阶段。使用对话框按钮将字符保存在字体库中。

要使用现有的字体库：

- 在**学习**菜单上，单击**字体库**。
- 选择要使用的字体库并单击**打开**。
- 在**学习**菜单上，单击**附加字体库或读取字体库**。

当选择**附加字体库**时，务必启用**互动学习**。Readiris 将识别字体库中存储的字符形状并使用互动学习，使您能够将新信息存储在字体库中。

当选择**读取字体库**时，Readiris 将识别字体库中存储的字符形状，但不会向字体库中增加新内容，即使**互动学习**已启用。注意，使用**互动学习**进行检查仍然很有用，若有需要，还可以纠正字体库中没有存储的识别结果。

注意：当字体库模式**新建字体库**启用时，请勿单击**学习**菜单上的**字体库**并打开已有的字体库。否则，将删除已有字体库的内容。

- 单击**识别 + 保存**识别文档。

第 10 章
格式化和保存文档

格式化文档

在 Readiris 中进行了 OCR
的文档可保存为各种输出格式。Readiris 可将 OCR 结果保存为 Adobe Acrobat PDF 文件、Microsoft XPS 文件、Word、RTF 和 OpenDocument 文本文件、HTML 和 XML 文件、WordML 和 SpreadsheetML 工作表以及 Ansi 和 Unicode 文本文件。

除此之外，Readiris 可将结果直接导出至目标应用程序，例如 Microsoft Word 和 Excel、Adobe Reader、Microsoft XPS Viewer、主要的 Web 浏览器和电子邮件软件等。

> 注意：也可以在基于 Web 的办公应用程序
> AjaxWrite、ThinkFree、Zoho Writer 和 Google Writely（可打开 RTF 和 HTML 文档）中打开由 Readiris 创建的 RTF
> 文件。当使用这些应用程序之一时，务必选择版面选项**保留文字和段落格式**。

操作

- 单击主工具栏上的**格式**按钮选择输出模式。

  Readiris

  - 会把文档发送至将自动打开的应用程序，或

o 将文档保存为外部文件。

选项**通过电子邮件发送**会创建新的电子邮件并将识别的文档作为电子邮件附件插入。

- 单击不同的选项卡以选择您想应用的设置。

  不能用于所选输出格式的设置将灰显。

  最常用的输出文件格式及其选项将在下文讨论。

- 单击**识别 + 保存**执行识别并保存文档。

  可以多次导出 OCR 结果而无需重复识别。再次单击**格式**按钮并更改文本格式和格式化选项。然后再次单击**识别 + 保存**。

  为便于搜索和整理，Readiris 允许您定义 PDF、**XPS**、**Word**、RTF、WordML、SpreadsheetML 和 HTML 输出的文档属性。要定义文档属性，请单击**文件**菜单上的**文档属性**。注意，在**输出文件**对话框中也可以访问文档属性选项。该对话框在您单击**识别 + 保存**时打开。

  注意，当将多页文件保存为**外部文件**时，您可以在 Readiris 中为每个页面创建单独的输出文件，或者将同一文档的所有页面保存至一个输出文件。只需单击**输出文件**对话框中的对应选项：分别是**每页创建一个文件**和**每文档创建一个文件**。

但是请注意，仅当将文档保存为外部文件而非在目标应用程序中打开文档时，**每页创建一个文件**和**每文档创建一个文件**选项才可用。

格式化文本文档

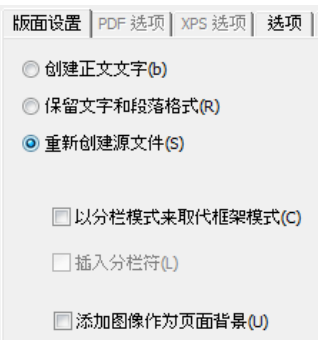
通过 Readiris，您可以生成几种类型的基于文本的输出格式。Readiris 还提供 Word、WordML、RTF、文本和 OpenDocument 文本输出。

要生成基于文本的输出文件：

- 单击主工具栏上的**格式**按钮并在**发送至**或**外部文件**下拉列表中选择所需的输出格式。
- 视您选择的文本格式而定，若干格式化选项可用。

不能用于选定文本格式的所有选项将灰显。

版面选项



- 选项**创建正文文字**可避免让 Readiris 进行文本格式化。

  Readiris 生成连续的文本。

- 选项**保留文字和段落格式**介于正文文字和自动格式化之间。

  在识别过程中字体类型、大小和字型风格得到保持。

  每个区块的制表符和对齐被重建。

  文本块和分栏不会被重建；段落与段落会互相连接。

  表格被正确取回。

- 选项**重新创建源文档**重新创建原始文档的复制本。

  Readiris 生成源文档的真实副本而不再是扫描图像。

  Readiris 还创建链接至电子邮件地址和网站的超级链接。

o 选项**以分栏模式来取代框架模式**创建分栏的文档。

  分栏的文本比包含多个框架的文档更易于编辑。文本从一个栏自然地连接到下一个栏。

  注意：当系统无法检测源文档的栏时，此格式化模式使用框架作为替代。

o 选项**插入分栏符**在每个分栏的末尾插入分栏符。

  您编辑、添加或删除的任何文本仍然保留在其栏内，文本不会自动跨过分栏符。

  提示：当您有分栏的正文文字时，请禁用此选项。将确保文字从一个栏自然地连接到下一个栏。

o 选项**添加图像作为页面背景**将扫描的图像作为页面背景放在识别文字底下。

不过，此选项会显著增加输出文件的大小。

格式 **PDF 文本-图像**为 PDF 文件提供同样的结果。

选项选项卡上的**保留背景色**选项提供较不严格的、更紧凑的替代。

常规选项

- 单击**选项**选项卡以选择常规选项：

o 选项**将多行合成段落**可进行自动的段落检测。

Readiris
将识别的文本自动换行，直至新的段落开始，并"重新连接"行末带有连字符号的单词。

o 选项**包含图像**将在自动格式化的文件中包含图形。

这对于创建文档的真实副本至关重要。

使用**图形**选项卡上的图形选项以确定输出文件中存储的图形的色彩模式和分辨率。

o 选项**保留文字颜色**在识别过程中保留文字的原始颜色。

o 选项**保留背景色**在识别过程中保留页面背景的专色。

在输出文件中每个段落创建统一的背景色。

纸张大小

视您选择的格式而定，您可以指定期望的纸张大小：

- 单击**纸张大小**选项卡，使用箭头按钮应用和排除纸张大小。

- Readiris
  将以指定的顺序逐个检查启用的纸张大小，然后使用第一个足以容纳扫描文档的纸张大小。

格式化基于表格的文档

通过
Readiris，您可以将表格数据输出至电子数据表文字处理程序以及 Web
浏览器：表格被逐格地在工作表中重新构建，并可作为表格对象插入文字处理程序文件中。

Readiris 可识别网格化和非网格化的表格。

Performance test optical media	
CD-ROM Digital Versatile Disk	Average time (n
CD-ROM 32x speed	60
DVD	58
Tested on 333 MHz Pentium II	

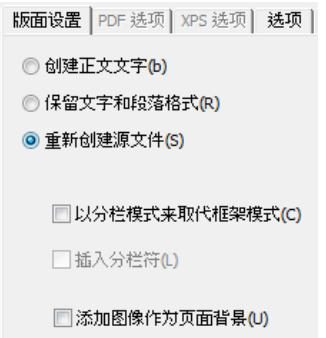
（网格化）

123.985	69	31
987.745	129	2
19	149	91
287.410	49	52

（非网格化）

要生成基于表格的文档：

- 单击主工具栏上的**格式**按钮并在**发送至**或**外部文件**下拉列表中选择所需的输出格式。
- 选择所需的版面选项。

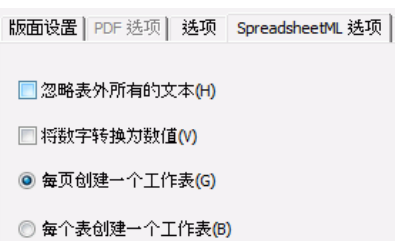


有关格式化选项的更多信息，请参阅**格式化文档**部分。

SpreadsheetML 选项

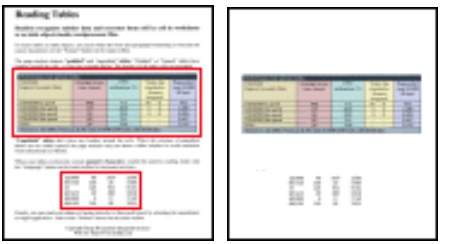
当选择 **Microsoft Excel 2002/2003** 作为目标应用程序时，特定的 **SpreadsheetML** 选项可用。单击选项卡 **SpreadsheetML 选项**以显示选项：

注意，选择此格式时版面选项**重新创建源文件**变得不可用。



- 选项**忽略表外所有的文本**将保存表格并忽略所有其他识别结果。

表格内的所有数据将被取回，表格外的数据则不会。



- **将数字转换为数值**选项将把识别的数字编码为数值。

  结果，您可以对那些单元格进行算术运算。（所有表格中的）文本单元格将保留文本。

  注意，只有表格内的数字被编码为数值。

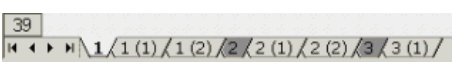
  Excel 专门对被编码为数值的数据执行算术运算。

- 选项**每页创建一个工作表**保证每扫描页面创建一个工作表。

  如果页面包含表格和文本，则全部都将被放在同一工作表上。

- 选项**每个表格创建一个工作表**将每个表格放在一个单独的工作表中，将识别的文字（表格外的）放在另一工作表中。

  如果识别的文档包含**数页**，您将看到每页都将重复该结构。



常规选项

单击**选项**选项卡以选择常规选项：

- 选项**将多行合成段落**可进行自动的段落检测。

  Readiris
  将识别的文字自动换行，直至新的段落开始，并"重新连接"行末带有连字符号的单词。

- 选项**保留背景色**重新创建每个单元格的背景色。

	A	B	C
1	Performance optical media		
2	CD-ROM	Average access	CPU
3	Digital Versatile Disk	time (msec)	utilization (%)
4	CD-ROM 24x speed	80	58.2
5	CD-ROM 32x speed	60	72.1
6	DVD	58	78.9
7	Tested on 333 MHz Pentium II with 64 MB RAM and 4 GB HD		

纸张大小

视您选择的格式而定，您可以指定期望的纸张大小：

- 单击**纸张大小**选项卡，使用箭头按钮应用和排除纸张大小。

- Readiris
  将以指定的顺序逐个检查启用的纸张大小，然后使用第一个足以容纳扫描文档的纸张大小。

创建 PDF 文档

Readiris 生成四种类型的 **PDF** 文件：**文本**、**文本-图像**、**图像-文本**和**图像**。

要生成 PDF 文件：

单击主工具栏上的**格式**按钮并在**发送至**或**外部文件**下拉列表中选择所需的 PDF 类型：

PDF 图像

如果选择 PDF **图像**，Readiris 生成仅包含图像的 PDF 文档而不执行 OCR。

PDF 图像-文本

如果选择 PDF **图像-文本**，Readiris 将识别文字，并创建包含页面图像和所识别文字的可搜索 PDF 文件。

页面图像覆盖在文字之上。

通过此格式，您始终能够看到原始文档（就如扫描的那样），同时您能够搜索并复制-粘贴隐藏在图像之下的已识别文字。因此，此格式有助于**进行存档**。

PDF 文本

如果选择 PDF **文本**，Readiris 将识别文字并创建可搜索的 PDF 文件。

这些单层 PDF 文件中不包含页面图像。

PDF 文本-图像

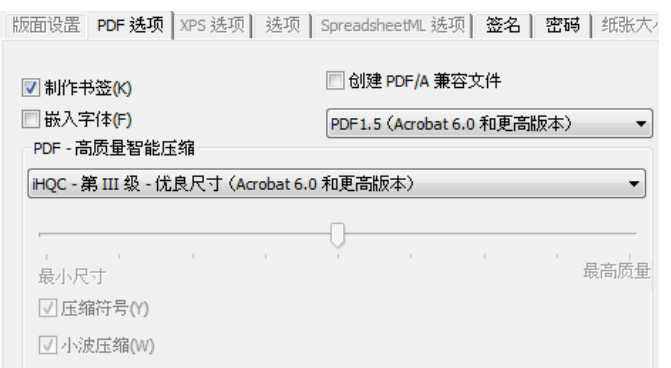
如果选择 PDF **文本-图像**，Readiris
将识别文字，并创建包含页面图像和所识别文字的可搜索 PDF 文件。

文本覆盖在页面图像之上。

选择 PDF 选项

要选择 PDF 选项：

- 单击主工具栏上的**格式**按钮并在**发送至**或**外部文件**下拉列表中选择所需的 PDF 类型。
- 视您选择的 **PDF 类型**而定，若干选项可用。单击 **PDF 选项**选项卡以访问这些选项：



创建书签

选项**创建书签**可为 Adobe Acrobat PDF
文件中的每个文本块、图形和表格创建书签。

嵌入字体

选择选项**嵌入字体**以在 Adobe Acrobat PDF 文件中嵌入字体。

嵌入字体可以防止字体替换，不论读者的计算机配置如何，都能确保读者看到以其原始字体呈现的文字。

嵌入字体会在某种程度上增加识别文档的文件大小。

PDF 文档的 iHQC 压缩

除了四种类型的"常规" PDF 输出，Readiris 还提供经 iHQC
压缩的 PDF 输出：**图像-文本**和**图像**类型的 PDF 文档可以通过 iHQC 方式高度压缩却不降低图像质量。

iHQC stands foriHQC 是指 **intelligent High-Quality Compression**（高质量智能压缩）， 它是 I.R.I.S.
专有的高效压缩技术。iHQC 与图像的关系就像 MP3 与音乐和 DivX 与电影的关系。

要生成经 iHQC 压缩的 PDF 文件：

- 单击主工具栏上的**格式**按钮并选择两种输出模式之一。
- 在**发送至**或者**外部文件**列表中，选择您所要的 PDF
  类型：**PDF 图像-文本**或者 **PDF 图像**。

- 在选项卡 **PDF 选项**上，选择所需的压缩级别。

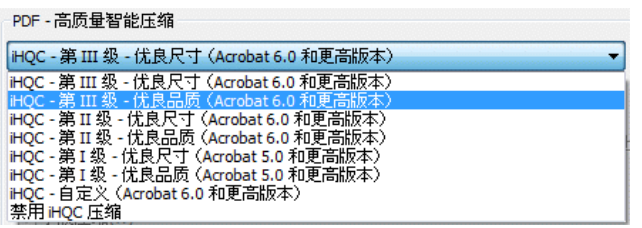
  Readiris Pro 支持**第 I 级 - 优良尺寸**和**第 I 级 - 优良品质**压缩。

  Readiris Corporate 还支持**第 II 级和第 III 级优良尺寸**和**优良品质**压缩以及**自定义**压缩。

  在第 II 级压缩中，选项**压缩符号**被自动启用以紧凑地压缩文本。

  在第 III 级压缩中，选项**小波压缩**也被自动启用以紧凑地压缩图形。

  如果选择了**自定义**压缩，您可以单独地启用或禁用这些选项。您也可以使用滑块来定义**优良尺寸 - 优良品质**率。



- 单击**识别 + 保存**识别文档并将其保存为经 iHQC 压缩的 PDF 文件。

PDF 文档的转制

除了生成 PDF 文件，Readiris 还能转制 PDF 文件：Readiris 将**图像** PDF 转为**文字** PDF
或所支持的任何其他文本格式，并**解除锁定**只读的 PDF 内容。

警告：Readiris 不能打开受用户密码保护的 PDF 文档。

操作

- 单击主工具栏上的**打开**按钮，选择您想让 Readiris 转制的 PDF 文件。

  在对话框的**页面范围**区域，选择**页面**并指定您想让 Readiris 打开的页面。

- 单击对话框中的**打开**按钮，选择所需的 PDF 文件。
- 单击主工具栏上的**格式**按钮，选择所需的 PDF 类型。

  有关 PDF 类型的更多信息，请参阅 **创建 PDF 文档**部分。

- 单击**识别 + 保存**按钮以转制文档。

创建 XPS 文档

Readiris 生成四种类型的 XPS 文件：**文本**、**文本-图像**、**图像-文本**和**图像**。

XPS 代表 XML
纸张规范，是由微软公司开发的固定版面的格式。

要生成 XPS 文件：

单击主工具栏上的**格式**按钮并在**发送至**或**外部文件**下拉列表中选择所需的 XPS 类型：

XPS 图像

如果选择 XPS **图像**，Readiris 生成仅包含图像的 XPS
文档而不执行 OCR。

XPS 图像-文本

如果选择 XPS **图像-文本**，Readiris
将识别文字，并创建包含页面图像和所识别文字的可搜索 XPS 文件。

页面图像覆盖在文字之上。

通过此格式，您始终能够看到原始文档（就如扫描的那样），同时能够搜索并复制-
粘贴隐藏在图像之下的已识别文字。因此，此格式有助于**进行存档**。

XPS 文本

如果选择 XPS **文本**，Readiris 将识别文字并创建可搜索的 XPS 文件。

这些单层 XPS 文件中不包含页面图像。

XPS 文本-图像

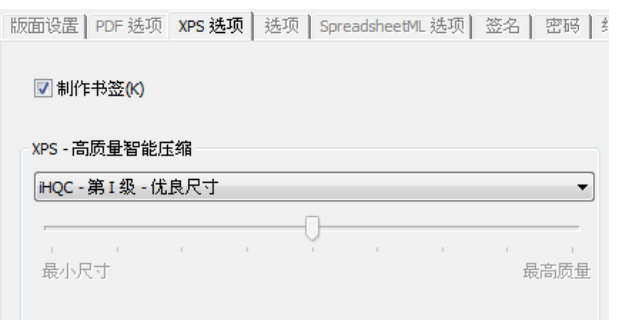
如果选择 XPS **文本-图像**，Readiris
将识别文字，并创建包含页面图像和所识别文字的可搜索 XPS 文件。

文本覆盖在页面图像之上。

选择 XPS 选项

要选择 XPS 选项：

- 单击主工具栏上的**格式**按钮并在**发送至**或**外部文件**下拉列表中选择所需的 XPS 类型。
- 视您选择的 **XPS 类型**而定，若干选项可用。单击 **XPS 选项**选项卡以访问这些选项：



创建书签

选项**创建书签**可为 Microsoft XPS
文件中的每个文本块、图形和表格创建书签。

XPS 文档的 iHQC 压缩

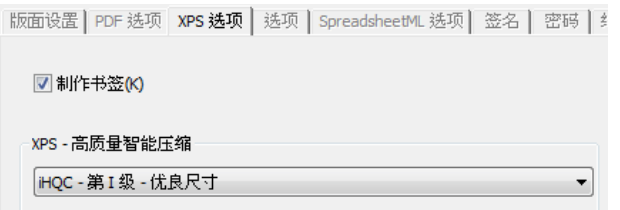
除了四种类型的"常规" XPS 输出，Readiris 还提供经 iHQC 压缩的 XPS 输出。**图像-文本**和**图像**类型的 XPS 文档可以通过 iHQC 方式高度压缩。

iHQC 是指 **intelligent High-Quality Compression（高质量智能压缩）**，它是 I.R.I.S. 专有的高效压缩技术。iHQC 与图像的关系就像 MP3 与音乐和 DivX 与电影的关系。

要生成经 iHQC 压缩的 XPS 文件：

- 单击主工具栏上的**格式**按钮并选择两种输出模式之一。
- 在**发送至**或者**外部文件**列表中，选择您所要的 XPS 类型：**XPS 图像-文本**或 **XPS 图像**。
- 在 **XPS 选项** 选项卡上，选择所需的压缩级别。

  Readiris 支持**第 I 级优良尺寸**和**第 I 级优良品质**压缩。

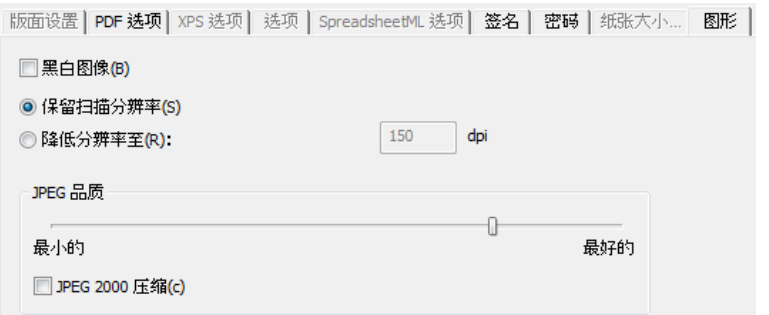


选择图形选项

依您选择的输出格式和目标应用程序而定，高级图形选项可能可用。图形选项可用于更改图像质量和分辨率。

要访问图形选项：

- 单击主工具栏上的**格式**按钮并在**发送至**或**外部文件**下拉列表中选择所需的输出格式。
- 单击**图形**选项卡以显示选项。



颜色模式

默认情况下，Readiris 以**彩色**模式保存图形。

选择**黑白**，则将以黑白模式保存图形。

分辨率

默认情况下，Readiris **保留扫描分辨率**。

您也可以选择**减少分辨率**。

提示：当把文档保存为 HTML
文件以便在网站上发布时，请将分辨率降低至 70 dpi（屏幕分辨率）。

JPEG 品质

PDF、XPS、Word 和 RTF 文档中存储的图形将以 JPEG 格式保存。

使用滑块调节 JPEG 品质。

JPEG 2000 压缩

当以 PDF 或 XPS 格式保存文件时，Readiris 可以将 JPEG 2000 压缩应用至那些文件中存储的彩色-灰阶图像。

JPEG 2000 是最新的、更紧凑版本的 JPEG 标准。

选择选项 **JPEG 2000 压缩**以进行应用。

第 11 章
保存和载入设置

您在 Readiris
中指定的任何设置都被自动保存，在您关闭应用程序后，将来仍能继续使用。要恢复出厂设置，请单击**文件**菜单上的**恢复出厂设置**命令。

当扫描全都需要不同设置的不同组的文档时，为每组保存单独的**设置文件**很有用。

操作

- 选择您要用于某一文档组的设置。
- 在**文件**菜单上，单击命令**保存设置**。
- 以后扫描或打开同一组的文档时，单击**文件**菜单上的**加载设置**命令。
- 选择正确的设置文件并单击**打开**按钮。
- 单击**识别 + 保存**以使用正确的设置识别文档。

注意：可通过**文件**菜单上的**信息**命令概览您所选的最重要的设置。

第 12 章
识别多页文档

打开和识别多个图像文件

Readiris 可一次处理多个图像文件。

要打开多个图像文件：

- 单击主工具栏上的**打开**。

- 选择要打开的图像文件：
  - 选择第一个图像文件并按住 **Ctrl** 键选择其他图像；或
  - 单击第一个图像并按住 **Shift** 键选择最后一个图像，从而选择连续范围的图像文件。

注意：当打开一个包含多个页面的文件时，例如一个多页 TIFF 文件或 PDF 文件，您可以指定页面范围。在对话框的**页面范围**区域，选择**页面**并指定要识别的页面。

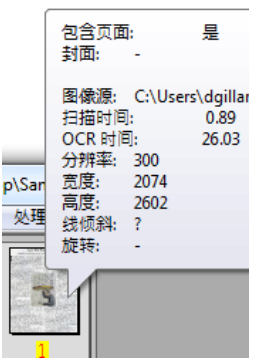
- 单击**打开**按钮打开图像。

注意，您也可以将图像文件从 Windows 资源管理器拖放到 Readiris 图像窗口来打开它们。

- 页面工具栏将显示打开的图像文件。

提示：将鼠标指针放在页面缩略图上可显示每页的设置信息。

页面工具栏可被用于编辑多页文档。更多信息请参阅**编辑多页文档**部分。



- 确定识别设置并单击**识别 + 保存**以执行识别。
- 如果您想在当前文档的基础上打开或扫描更多图像，请单击主工具栏上的**扫描**或**打开**按钮。

您将被提示是否要删除当前的文档。单击**是**将删除当前的文档并开始新的文档，单击**否**将在当前文档的基础上增加其他扫描。

扫描和识别多页文档

Readiris 的设计可处理包含多个页面的文档。

Readiris Home 和 Readiris Pro 最多可处理 50 页的文档。

Readiris Corporate 可处理不限页数的文档。

要在 Readiris
中扫描多页文档，当使用馈纸型扫描仪时可以使用**自动输稿器**功能，当使用平板扫描仪时可使用**间隔扫描**功能。

用输稿器（馈纸型扫描仪）扫描多页文档

- 单击主工具栏上的**扫描仪**按钮，选择 **ADF**（自动输稿器）选项。
- 将页面放入扫描仪的输稿器中然后单击**扫描**开始扫描。
- 单击**识别 + 保存**以识别文档。

用间隔扫描功能（平板扫描仪）扫描多页文档

- 单击主工具栏上的**扫描仪**按钮。
- 选择**继续扫描前等待**并使用箭头按钮指定间隔时间。

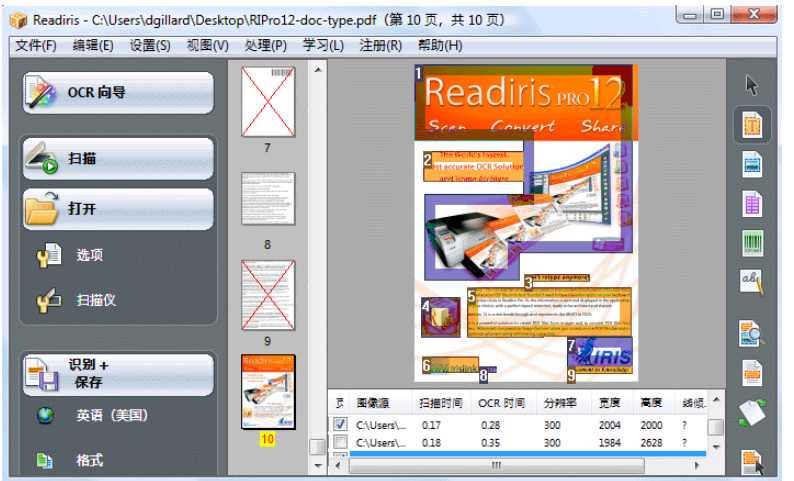
扫描仪将在指定秒数后自动扫描另一页面，无需您每次单击**扫描**按钮。

单击间隔扫描对话框中的**放弃**按钮或按键盘的 **ESC** 键将结束自动扫描。

单击间隔扫描对话框中的**暂停**按钮或按键盘的空格键将暂停计算扫描间隔。准备好继续扫描时请单击**继续**。

编辑多页文档

当在 Readiris
中打开或扫描多页文档时，页面工具栏会显示它们的缩略图。页面工具栏中的缩略图可用于编辑多页文档。



在文档内移动页面：

- 右键单击要移动的页面，单击**选择页面**。
- 将页面拖到想要的位置。
- 或者右键单击页面并单击**向上移动页面**或**向下移动页面**。

删除页面：

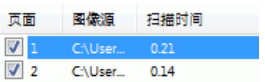
- 右键单击要删除的页面，单击**删除页面**。
- 或者选择页面并按键盘上的**删除**键。

排除识别页面：

- 右键单击要排除的页面，单击**“排除页面”**。
- 或者清除它在文档面板中的页码框。

被排除的页面在页面工具栏中被打叉删除。

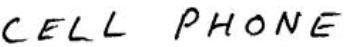
当打印扫描图像和保存扫描至多页图像文件时，被排除的页面将被忽略。



提示：**编辑**菜单上的命令**包含所有页面**和**排除所有页面**被同时应用至所有页面。

第 13 章
识别手写文字

除了键入的文本、表格、图形和条形码，Readiris
还可识别**手写的**文字。手写文字包括相互隔开的正体字。



需要采用高度专业化的
ICR（智能字符识别）软件来识别手写字符。

要识别手写文字：

- 单击图像工具栏上的手写文字按钮。

- 围绕手写文字画一个框。
- 单击主工具栏上的**识别 + 保存**按钮。

将识别包括手写文字在内的整个文档。

注意：右键单击手写文字区域并单击 **复制为文本**，这样只识别手写文字并将其内容发送至剪贴板。

识别的符号

手写文字识别仅限于拉丁字母，支持数字 (0-9)、大写字母 (A-Z) 以及逗号、句号、加号和连字号这几种标点符号。

不支持重音符号、变音符号和其他特殊字符。

注释

- Readiris 支持**手写文字**，但不支持**草写文字**。

  更多信息请参阅**文字手写规则**部分。

- 大写字母在识别后被替换为小写字母，除非它们出现在句首。
- 文档属性语言、字体类型和字符间距不适用于手写文字。
- 互动学习也不适用。ICR
  技术是以一百万个以上的书写样本为基础的。

文字手写规则

要让 Readiris 识别手写文字，必须考虑若干规则：

- 书写正规、适当间隔的字符。

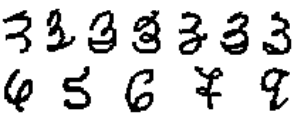
注意字母 A、G 和 Q 的写法。

- 使用足够粗的圆珠笔。

  黑笔的效果比蓝笔好。

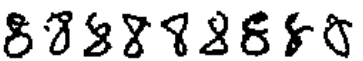
  不要使用铅笔。

- 不要过于风格化。



  过度风格化的字符会增加 OCR 错误的风险。

- 不要让该闭合的笔画打开，该打开的却闭合。



- 避免字符断裂。

- 避免折返。

  折返会降低图像质量和手写符号的清晰度。

  被完全删去的字符将不会被识别。

- 正确书写。

  数字 1 的书写可以采用"英语化"和"欧洲"风格。

  1 有无下划线均可。

水平下划线不一定要接触字体形状的其他部分。

提示：当获得的结果不理想时，请使用 I.R.I.S.
书写表格并调整您的书写风格。空白的 I.R.I.S.
书写表格可作为整页模板使用，您可在模板上以正确的大小正确填写正体字。此表格可在 Readiris 光盘和 Readiris 安装文件夹中找到。

第 14 章
识别条形码

条形码读取介绍

除了 128 种语言的光学字符识别，Readiris 也可识别条形码。

本软件支持所有普遍的条形码符号表示法：Codabar、Code 128、Code 39、Code 39 extended、Code 39 HIBC、Code 93、Datalogic 2 of 5、Discrete 2 of 5、EAN-13、EAN-8、Interleaved 2 of 5、MSI Pharmaceutical、MSI-Plessey、Kodak patch code、PDF-417、PostNet、UCC-128、UPC-A 和 UPC-E。

注意，必须是激光打印或喷墨打印的条形码才能使 Readiris 执行 OCR
识别。不支持点阵式打印的条形码，因为此类条形码没有形成足够的对比，其分辨率多数限制为 60 dpi。

手动条形码读取

- 确定您想让 Readiris 识别的条形码：

- 在**设置**菜单上，单击**条形码**。
- 确定您想让 Readiris 识别的条形码。
- 确定是让 Readiris 验证校验数位还是移除校验数位。

- 单击图像工具栏上的条形码按钮，并框选文档中的条形码区域。

- 单击主工具栏上的**识别 + 保存**按钮。

将识别包括条形码内容在内的整个文档。

注意：右键单击条形码区域并单击**复制为数据**，将其内容复制到剪贴板上。

索引

A

ADF 79

E

Excel 输出 55

H

HTML 输出 55

J

JPEG 2000 压缩 73

O

OpenDocument 输出 55

P

PDF iHQC 输出 66

PDF 文档的转制 68

PDF 选项 65

PDF 输出 55, 64

R

RTF 输出 55

S

SmartTask 16, 19

SpreadsheetML 输出 55

U

Unicode 输出 55

W

Word 输出 55

WordML 输出 55

X

XML 输出 55

XPS iHQC 输出 71

XPS 选项 70

XPS 输出 55, 68

三划

工作表 60

四划

支持的图像格式 25

中东版本 4, 8, 46
手写文字 83, 84
互动学习 50
反转图像 29
分辨率 27
文本文档 57
文档面板 17
文档属性 48, 55

五划

电子数据表文档 60
用户界面 15
用户界面语言 18
主工具栏 16
出厂设置 75
对比度 28, 33
纠偏 23

六划

扫描仪设置 26
亚洲版本 4, 7, 46
亚洲语文档 4, 7, 46
页 80
删除 80
选择 80
排除 80
移动 80
灰阶图像.....28
页面分析.....23, 37
页面纠偏.....23
自动输稿器.....79
向导.....13
多页文档.....77, 78
产品支持.....12
字体库.....52
字体类型.....49
字符间距.....49
安装.....9
设置文件.....75

七划

更改用户界面.....18
希伯来语文档.....4, 8, 46
条形码.....87
系统要求.....9

间隔扫描 79
识别窗口手动设定 39
识别窗口自动设定 37
识别窗口设定模板 43
阿拉伯语文档 4, 8, 46
纸张大小 59

八划

表格 60
表格单元格的背景颜色..... 63
图形选项 72
图像工具栏 17
版面文件 43
版面选项 57
注册 11
线倾斜 30

九划

点阵 49
背景颜色 58
卸载 Readiris..... 11
选项 23
重建源文档 58
保存设置.....75
亮度.....28, 32
恢复出厂设置.....75
语言.....46
除斑.....34

十划

载入设置.....75
速度与准确度.....48
准确度与速度.....48
调整扫描文档.....31
通过电子邮件发送文档.....56

十一划

彩色图像.....28, 31
彩色图像平滑化.....30, 31
旋转.....23
混合语言.....48

十二划

黑白图像.....28
编辑多页文档.....80

十三划

输出格式.....55

数字 47

数码相机 29

十五划

颜色模式.....28