

ABBYY FlexiCapture 9.0

项目安装指南

亲爱的用户！

本指南对项目安装期间所执行的操作进行了描述。

如果要了解更多信息，请尝试以下方式：

- 使用其它帮助文件。使用程序菜单或单击“F1”，或单击“开始 > 程序 > ABBYY FlexiCapture 9.0 工作站 > 帮助”，打开帮助文件。
- 系统管理员指南：开始 > 程序 > ABBYY FlexiCapture 9.0 工作站 > 指南 > 系统管理员指南。
- 操作员指南：开始 > 程序 > ABBYY FlexiCapture 9.0 工作站 > 指南 > 操作员指南。
- 机器可读表单创建指南：开始 > 程序 > ABBYY FlexiCapture 9.0 工作站 > 指南 > 机器可读表单创建指南。
- 从 FlexiLayout Studio 9.0 和 FormDesigner 9.0 的菜单或按 F1 打开相应应用程序的帮助文件。也可以单击“开始 > 程序 > ABBYY FlexiCapture 9.0 工作站 > 帮助”来打开帮助文件。
- ABBYY FlexiLayout Studio 教程。开始 > 程序 > ABBYY FlexiCapture 9.0 工作站 > 指南 > 教程。

对于单机版，请用 ABBYY FlexiCapture 9.0 替换上述路径中的 ABBYY FlexiCapture 9.0 工作站。

希望您喜欢使用我们的产品！

目录

1. 介绍.....	4
1.1. 数据捕获的用途.....	4
1.2. 数据捕获自动化.....	4
1.3. ABBYY FlexiCapture 9.0 可以处理的文档.....	5
1.4. ABBYY FlexiCapture 9.0 工作站.....	5
2. 创建项目.....	6
3. 创建文档定义.....	6
3.1. 创建布局.....	7
3.1.1. 定义对象的几何标记.....	7
3.1.1.1. 字段组.....	8
3.1.1.2. 无标记的字段.....	9
3.1.1.3. 表格标记.....	9
3.1.1.4. 带有多个区域的字段.....	9
3.1.1.5. 带有多个实例的字段.....	10
3.1.1.6. 在识别时排除某个区域.....	10
3.1.1.7. 删除字段.....	11
3.1.2. 静态元素.....	11
3.1.2.1. 条形码规格.....	11
3.1.3. 字段属性.....	11
3.1.3.1. 常规字段属性.....	12
3.1.3.2. 数据类型.....	13
3.1.3.2.1. 文本输入字段的数据类型.....	13
3.1.3.2.2. 复选标记数据类型.....	16
3.1.3.2.3. 复选标记组数据类型.....	17
3.1.3.3. 字段识别属性.....	18
3.1.3.3.1. 文本输入字段识别属性.....	18
3.1.3.3.2. 复选标记和复选标记组识别属性.....	19
3.1.3.3.3. 条形码识别属性.....	20
3.1.3.3.4. 图像识别属性.....	20
3.1.3.4. 校验设置.....	21
3.1.3.5. 图像导出参数.....	21
3.1.3.6. 规则检查.....	22
3.1.3.7. 自定义操作/脚本编辑.....	23
3.1.4. 创建多页文档的文档定义.....	24
3.1.5. 创建包含附录页面的文档定义.....	27
3.1.6. 导出设置.....	27
3.1.6.1. 导出至文件.....	28
3.1.6.2. 导出至数据库.....	29
3.1.6.3. 导出至 SharePoint.....	29
3.1.6.4. 导出图像.....	30
3.1.6.5. 自定义导出（脚本）.....	30

3.1.7.	配置识别数据的显示.....	31
3.1.8.	测试文档定义.....	31
3.1.9.	编辑和发布文档定义.....	31
3.2.	创建 FlexiLayout	32
3.2.1.	分类器.....	32
3.3.	非结构化文档的规格	33
4.	配置项目属性	33
5.	配置批处理文件类型.....	34
5.1.	常规.....	35
5.2.	图像预处理.....	35
5.3.	识别.....	35
5.4.	完整性检查.....	36
5.5.	导出.....	36
5.6.	工作流.....	36
5.6.1.	单机.....	36
5.6.2.	分布式.....	36
6.	配置图像导入	37
7.	上传项目至应用程序服务器	38
8.	键盘快捷键.....	38
8.1.	主窗口.....	38
8.2.	文档定义编辑器窗口.....	40
8.3.	组校验窗口.....	42
8.4.	字段校验窗口.....	43

1. 介绍

1.1. 数据捕获的用途

现在的商业、生产和服务领域都会用到大量文档。申请表、调查问卷、发票、草稿以及其他文档，是任何一家公司的必备之物。现代信息技术突显了纸质文档的不足，且大部分数据被转换为电子形式来进行存储、分析和处理。

电子文档最耗时耗力的就是数据输入。只能凭借手工输入数据，输入少量信息尚可，但如果有大量文档，就勉为其难了。由于手工输入团队难以管理，并且更改的成本与从头开始处理的成本相当，因此在如此严苛的形势下，手工输入速度不能得到马上提高。

因此，手工输入并非最佳途径。另一种更简单也更有效的途径，就是使用像 ABBYY FlexiCapture 9.0 之类的自动数据捕获系统。

1.2. 数据捕获自动化

ABBYY FlexiCapture 9.0 是一款用于处理结构化、半结构化 (FlexiLayout) 和非结构化文档的数据捕获软件。

自动数据捕获由以下阶段组成：

- 首先使用文档扫描仪对一系列页面进行**扫描**；
- 扫描页面被自动组装成文档；
- 自动**识别**字符；
- 将不确定识别的字符发送给操作员**检查（校验）**，
- 确认的数据最终被**导出**至文件或数据库，且文档图像被保存至指定文件夹。这些图像可以采用图形格式或作为可搜索 PDF 文件保存。

ABBYY FlexiCapture 9.0 是一款高效的数据捕获自动解决方案，它可以轻松控制工作进度和质量。

1.3. *ABBYY FlexiCapture 9.0 可以处理的文档*

ABBYY FlexiCapture 9.0 是一种支持不同文档类型的数据捕获应用程序。

ABBYY FlexiCapture 9.0 可以处理的文档类型如下。

- **结构化文档**。包含专用数据字段且这些字段在各文档副本中的数量、位置和格式均保持不变的文档，即结构化文档。此类表单通常以印刷形式发行，并且要手工填写。要识别结构化表单和捕获表单数据，必须创建布局向本程序指示字段位置。布局在项目安装期间由 ABBYY FlexiCapture 9.0 创建。
- **半结构化文档**。各副本的数据字段在数量、位置和格式不同的文档，即半结构化文档或弹性文档。发票就属于这类文档，其原因是发票由不同公司开具，发票的项目和格式通常有所不同。所有发票都有账户号和支付金额，但这些项位于文档的不同部分。ABBYY FlexiCapture 9.0 使用 FlexiLayout 来识别半结构化文档和数据捕获。FlexiLayout 是在 ABBYY FlexiLayout Studio 中创建的。有关 ABBYY FlexiLayout Studio 的信息，请参阅其帮助文件。弹性文档和半结构化文档的处理，仅创建阶段和布局匹配阶段与结构化文档不同。
- **非结构化文档**。如果您要处理包含自由表单信息的非结构化文档，如合同、信函、订单、图表，ABBYY FlexiCapture 9.0 也是理想的解决方案。非结构化文档可以自动作为结构化文档或弹性文档的补充来识别，或者通过使用 FlexiLayout，然后导出至图像文件和可搜索 PDF 文件。索引字段可自动或手动地从非结构化文档捕获（使用 FlexiLayout）。典型的非结构化文档处理方案，就是将纸质存档转换为电子形式，捕获几个基于属性搜索所需的索引字段。

1.4. *ABBYY FlexiCapture 9.0 工作站*

根据安装类型（单机或分布式），系统将包含不同的工作站。如果使用的是单机版，项目将安装在管理员工作站上。如果使用的是分布式版本，项目将安装在项目安装工作站上，然后将会发送到应用程序服务器。

2. 创建项目

项目包含数据捕获所需的全部设置（文档定义、图像导入配置文件、程序设置）和要处理的文档。

多个文档被合并为**批处理文件**。批处理文件格式化类型取决于处理类型。例如，您可以合并创建日期相同的文档，或者合并来自一次扫描的一组文档中的文档。批处理文件识别期间使用的设置按批处理文件类型定义。您可以为最常用的处理设置创建多种批处理文件类型。在这种情况下，操作员只需选择所需的批处理文件类型。也可以为整个批处理文件指定设置。这些设置将对类型设为“默认”的项目应用。首先指定项目设置，然后在此基础上形成批处理文件类型（项目设置为该批处理文件的默认设置）。

工作批处理文件用于文档处理。**测试批处理文件**在文档定义调试期间使用。这些批处理文件类型的区别，在于测试批处理文件使用本地（未发布）文档定义，而工作批处理文件则使用已发布的文档定义。您可从文档定义编辑器窗口直接访问“测试批处理文件”列表。

文档由一张或多张页面的图像（即单页文档和多页文档）以及从这些图像捕获的数据组成。

一个项目中可以包含多个文档定义。在这种情况下，将在一个项目中处理不同类型的文档。因此，不需对所处理的文档进行预排序。您可以将不同文档类型合并到一个流中。但是，如果文档流不交叉，则可以为它们创建独立的项目。

首先要创建且至少创建一个项目。

单击程序启动时打开的**打开项目**对话框中的**新建...**，或单击主菜单上的**文件 >新建项目...**，可创建新项目。选择要保存该项目的文件夹，并指定文件夹名称。

要向该项目添加批处理文件，请使用右键单击批处理文件列表时显示的快捷菜单的**新建批处理文件**命令。也可以在不创建批处理文件的情况下加载图像。如果属这种情况，程序将自动创建批处理文件。

在批处理文件名称上双击鼠标左键，可查看被添加至该批处理文件的文档。单击**项目 > 工作批处理文件**或按 **Ctrl+B**，可返回批处理文件列表。

3. 创建文档定义

创建文档定义是处理文档后影响所生成数据质量的主要项目设置阶段。要创建文档定义，执行如下操作：

- 创建布局（布局属性 — 在文档定义编辑器中创建，或通过加载 ABBYY FormDesigner 内置的表单，FlexiLayout – 通过加载 ABBYY FlexiLayout Studio 9.0 内置的文件）；
- 设置各字段的属性，并指定要在每个字段中查找的数据（可显著提高识别质量）。也可以指定要发送给操作员进行校验的字段。
- 设置字段值检查规则。这些规则有助于识别其值不符合要求的文档，例如字段值与特定数据库的值不一致。

- 设置数据导出方法。数据可以导出至文件、数据库、Microsoft SharePoint 或根据脚本程序导出。

在完成定义创建时，必须发布定义。然后，该定义即可用于文档处理。

创建和编辑文档定义的主要操作在**文档定义编辑器**窗口中执行，该窗口在创建新定义后打开。要打开文档定义编辑器，请单击**项目 > 文档定义...**，然后选择定义名称并单击**编辑...**。

要创建新定义，请单击**项目 > 文档定义...**，并在打开的对话框中单击**新建...**。将打开“文档定义创建向导”。在**创建新的文档定义**窗口中，可以指定文档定义的主要属性：名称、注释、语言和样式。然后，设置文本类型：从下拉列表选择 **ICR**（手印文本）或 **OCR**（机器打印的文本）。接下来将可以更改各个字段的文本类型，在此阶段选择的类型将用作默认类型。

下一步是加载或扫描作为文档定义原型的图像。（如果文档包含多页，暂时先加载第一页。要了解如何添加更多页面，请参考《创建多页文档的文档定义》）。您可以扫描页面（最好是未填写的页面）或从文件加载图像。如果要处理半结构化文档，请使用 **FlexiLayout**。选择**加载 FlexiLayout**，输入带扩展名 ***.afl** 的 **FlexiLayout** 文件的路径（该文件在 **ABBYY FlexiLayout Studio** 中创建）。如果创建的是弹性文档描述，该步就是最后一步。


如果创建的是固定文档描述，将继续执行下一步。选择要在图像上自动查找的字段类型。您可以设置复选标记和文本输入字段。如果使用带标记和矩形复选标记的文本字段，可以获得最佳自动搜索结果。如果输入字段没有标记，并且要对空白背景设置复选标记，建议采取手动创建。

如果表单有定位点（即所给定表单的静态元素，例如，用于识别和/或匹配文档定义的方块），那么还会在图像上查找和自动标记这些定位点。



3.1. 创建布局

3.1.1. 定义对象的几何标记

完成创建文档定义的所有步骤后，加载的图像将显示在文档定义编辑器窗口中。该图像包含最后一步选择的类型的字段和定位点。

以后可以通过单击 ，再单击要选择的元素区域某一位置，可自动选择对象。这样将自动识别元素类型和位置。

文档定义编辑器有一组非常方便访问的字段和静态元素标记工具：


- 字段区域模式 () 和
- 静态元素模式 ()。

要手动创建静态元素或字段，请单击工具栏上的相应按钮，然后单击并按住鼠标左键，围绕所需元素绘制一个矩形。也可以按住 **Shift** 并在对象区域中单击鼠标左键。在这种情况下，将自动定义字段或静态元素区域。

这是用来创建不同元素类型的工具列表：

字段:

 - 文本输入字段


 - 复选标记

 - 复选标记组

 - 条形码

 - 图片


 - 表格

 - (字段) 组

静态元素:

 - 定位点

 - 分隔符

 - 静态文本

 - 条形码

条形码可以是可识别的字段或静态元素。在根据条形码的用途选择模式时必须小心 — 如果要从条形码捕获信息，应采用字段区域模式标记区域；如果是用于定义识别和匹配，应采用静态元素模式标记区域。

创建的字段显示在**文档结构**窗口**字段**选项卡的列表中。匹配注释的名称或字段名称将作为默认值使用。在文档结构中选择字段名称并按 F2，可更改此字段名称。如果希望字段名称匹配注释，请选择该字段并在字段上面单击鼠标右键，从快捷菜单中单击**从图像获取名称...**。然后，围绕所需的文本绘制矩形。


可以按字段列表中的图标和框的颜色来识别字段类型。静态元素不会显示在该列表中。

可以复制（包括复制到其他文档节段）、删除、移动对象，以及调整对象大小。字段副本的名称将按数字进行编号。

您也可以一次选择几个对象。在选择时按住 Ctrl 即可选定多个对象。也可以使用  工具。

3.1.1.1. 字段组

您还可以将多个字段合并成组，来创建重复的字段组，或让文档结构更直观。例如，可以将城市、街道、门牌号合并为“居住地址”。然后，可复制“居住地址”字段组来创建“工作地址”字段。

要合并文档，请使用  工具。

如果文档中有重复的字段组，则可以为已创建的组创建多个实例。该组中的所有字段属性和规则集对该组的每个实例都有效。请参阅“带有多个实例的字段”。

您也可以复制字段组，但在此情况下，通过复制会创建另一个新的单独组。


3.1.1.2. 无标记的字段

有些字段在图像上没有对应的区域。无标记的字段在列表中标有红色星号。此类字段可用于存储所识别字段之值的计算结果。

无标记的字段具有其所属类型的所有属性特征：它们可以被发送给操作员校验，可以执行格式检查，可以导出字段值。

要创建无标记的字段，请执行以下操作：





1. 在文档定义编辑器中，单击**编辑 > 创建字段**，创建所需类型的字段。此字段将显示在列表中，并标有红色星号。在这种情况下，将创建文档结构字段，但会是在图像上创建区域。
2. 删除常规字段的标记。选择图像或列表上的字段，在快捷菜单上单击**删除区域**。该标记将被删除，字段将标上红色星号。

要为无标记字段创建区域，单击工具栏上的 。如果列表中有标星号的字段，程序将提示您为此类字段选择名称。如果先设置字段列表，而后再定义字段位置，可能就会发生这种情况。

3.1.1.3. 表格标记

ABBYY FlexiCapture 9.0 可以处理表格。**表格**字段可对该程序使用。

可以使用各色各样的工具在常规文档定义中进行表格标记。这些工具可用来绘制表格，以及定位行和列。

可使用  工具绘制表格区域。此区域不得包含表格标题。接下来，使用  工具标记表格单元以添加分隔符。要创建垂直分隔符，应将指针指向表格区域，将虚线分隔符拖至目标位置，然后单击鼠标左键将其放置到位。水平分隔符采用相同方法创建（只按住 Alt）。也可以启动分隔符自动搜索。要执行该操作，选择所创建的表格，并使用如下快捷菜单命令：**自动检测垂直分隔符**和**自动检测水平分隔符**。要删除分隔符，单击 。在添加所需数目的分隔符后，先单击  再单击列区域即可设置列。每列包含一类单元格：文本、复选标记、图形元素或条形码。在设置列时，程序会提示您选择列的类型。

完成几何标记后，需要设置识别和校验属性以及各列的数据类型。列的属性设置方法与常规文档字段的属性设置方法相同。

3.1.1.4. 带有多个区域的字段

如果字段的区域由表单上的多个部分组成，例如跨多页的表格，那么可以创建包含多个区域的字段来处理此类数据。

同一字段中所有区域的值可进行合并，并作为一个字段导出。分行符在这里用作分隔符。

要创建有多个区域的字段，先创建一个字段区域，选中该区域并单击鼠标右键，在快捷菜单中单击**继续区域...**，然后选择该区域要继续的位置。重复此过程以达到所需的次数。

3.1.1.5. 带有多个实例的字段

文档可以包含重复的对象 — 即出现在多个实例中，并且描述的是相似对象的字段或字段组，例如有关员工或其子女、发票的相同信息。此类对象将创建有多个实例的字段。

除表格以外的任何字段，均可有多个实例，各区域的位置彼此相距较远，有可能是在不同页面上。字段实例具有相同的属性，那些包含多个实例的字段可导出至单独的数据库文件或表格。

在创建重复的字段组时，包含多个实例的字段十分有用：只要创建一个字段组，然后就可以创建所需数目的实例了。

要创建包含多个实例的字段，先要创建一个实例，然后选中该实例，在字段区域上单击右键，在快捷菜单上单击**新建实例...**。创建所需数目的区域实例数，并将其图像移至页面的目标位置。

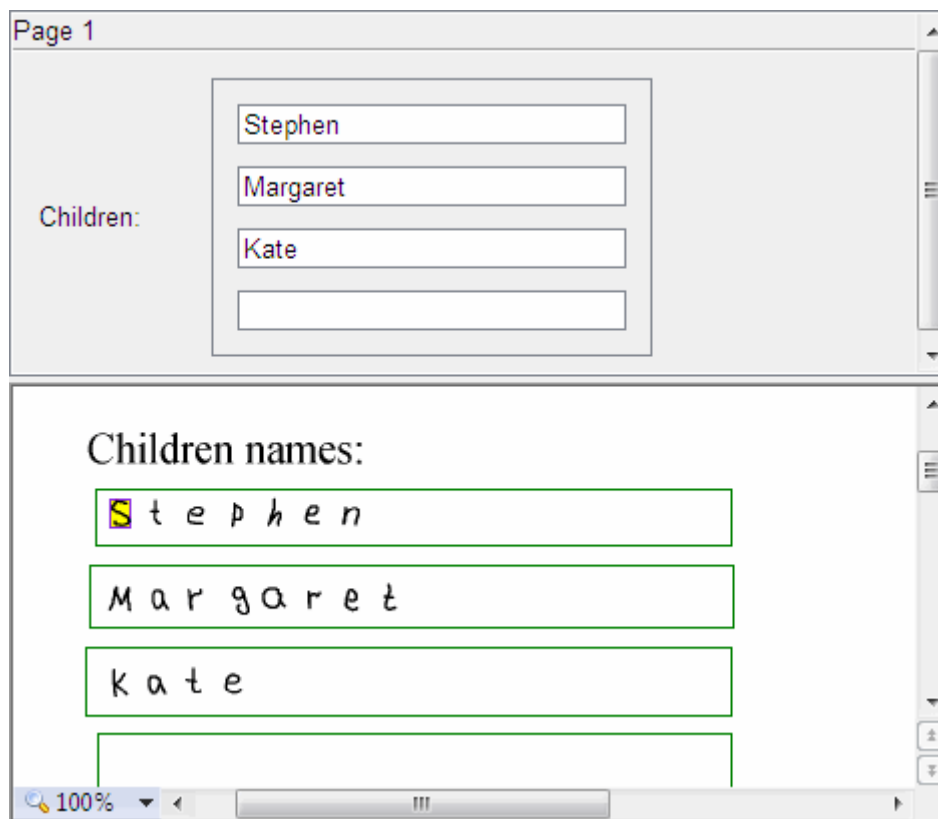


图 1. 使用包含多个实例的字段

3.1.1.6. 在识别时排除某个区域


在某些情况下，您可能需要将某个区域排除出识别范围，例如，排除妨碍字段数据捕获的带注释区域（图 2）。要将某个区域排除出识别范围，请单击  并用鼠标绘制要排除的区域。



图 2. 排除不可识别区域

3.1.1.7. 删除字段

要删除字段，选中该字段并按 **Delete** 键，或单击快捷菜单上的**删除按钮**。如果只删除标记，并将字段保留在文档结构中，请按 **Shift+Delete** 或单击字段快捷菜单上的**删除区域**。

3.1.2. 静态元素

静态元素就是不提供所要捕获的信息的对象。静态元素用于文档定义匹配和标识。定位点就是静态元素的一种类型。

单击 （将显示静态元素标记）切换到这种特殊模式，即可开始处理静态元素。

所有静态元素类型均可用于匹配文档定义，但仅当文档有标准定位点 — 黑方块形、十字形和直角形时，才能得到最佳结果。它们必须手动或自动地设为定位点类型的静态元素。必须在“元素属性”窗口的**常规**选项卡上指定静态元素的形状。

任何类型的静态元素均可用作标识符。通过分析标识符的位置和值，该程序可查找当前页面所属的文档。如果条形码用作标识符，可手动指定其值 — 这样有助于快速准确的识别页面。

要使用静态元素来进行文档定义匹配和/或识别，请选择“属性”对话框的**常规**选项卡中的相应选项。要打开任何元素的“属性”对话框，单击快捷菜单上的**属性...**。

在匹配文档定义（选择**用于匹配文档定义**）和/或进行文档定义识别（**用于识别文档定义**）时，可使用静态元素。

定位点。设置定位点类型（方块形、十字形、直角形或矩形）。对于标准定位点，选择**用于匹配文档定义**。只有在定位点位置对当前文档节段唯一时，才能选择**用于识别文档定义**。

静态文本。如果静态文本要用于识别，可输入文本值。只有在不能单纯通过位置来识别页面时，才需要输入文本值（例如，如果只有页面的标题文本不同，而这些标题的位置和大小相同）。

分隔符。使用分隔符来进行文档定义识别或匹配。

条形码。如果条形码要用作标识符，可以输入条形码值。在**识别**选项卡上，指定条形码类型、方向和图像处理选项。

3.1.2.1. 条形码规格

如果条形码用作标识符，那么它是定位点条形码，属于静态元素。它在处理静态元素时创建。此类条形码的“属性”对话框有两个选项卡：**常规**和**识别**。

如果某一条形码用于数据捕获，则该条形码就是一个字段。它在处理字段时创建。此类条形码的“属性”对话框包含“字段属性”对话框的所有选项卡：**常规**、**数据类型**、**识别**、**校验**和**规则**。此类条形码的值将进行识别，如果进行了设置，则将发送该值进行校验和导出。

3.1.3. 字段属性

正确设置字段属性非常重要，其原因是字段属性会影响字段识别质量，并且还要定义是否导出字段值以及将字段值发送给操作员进行校验。某些属性对数据识别极为重要。例如，应小心设置文

本字段标记属性，以便将扫描期间不会显现的标记从字符中间分离出来。这样就可以只识别文本，而不会识别不必要的元素。

正确的字段属性能产生更出色的识别结果，让操作员的工作变得更轻松。在设置字段属性时，文档校验可降至最低程度。

当创建特定类型的字段时，该字段将有默认属性。有些属性是取自文档定义（例如语言），而有些属性则是自动定义（例如标记类型）。可以在单击任何对象的快捷菜单上的**属性**时打开的对话框中更改字段属性。每种字段类型都有自己的特定属性。

3.1.3.1. 常规字段属性

每个字段的“属性”对话框都有**常规**（图 3）选项卡。**名称**和描述可在此选项卡上指定。您可将创建阶段自动指定给此字段的名称更改为最符合该字段的名称。**标题**是作为数据显示的字段名称。这种字段类型在该选项卡上显示为名称右侧的图标。

此外，还可以在**常规**选项卡上选择如下选项：

- **导出字段值** – 如果无需导出字段值，请清除此选项。例如，如果字段值要用于获取所计算字段的值，则必须执行此操作（请参阅[规则检查](#)），并且只需导出最终结果。
- **只读** – 选择此选项可禁止操作员进行任何更改。您可以为要（根据规则）自动计算值的字段选择此选项，例如，为要保存其它字段值之和或组合的字段选择此选项（请参阅[规则检查](#)）。
- **显示在数据表单中** – 如果在查看文档时不希望数据表单上显示字段，请清除此选项。在此类字段发生错误时，由于没有显示此类字段，因此不能在文档编辑器窗口中纠正错误。所以，建议指定不会导致此类字段发生任何错误的设置。
- **不得为空** – 如果为必填字段，请选择此选项。如果字段为空，将发生格式错误。
- **索引字段** – 如果要为字段编制索引，请选择此选项。在这种情况下，每个文档的字段值将显示在列表中，并且操作员可以进行排序，或通过此字段值来搜索文档。索引字段值也可以用于导出期间进行文件命名。

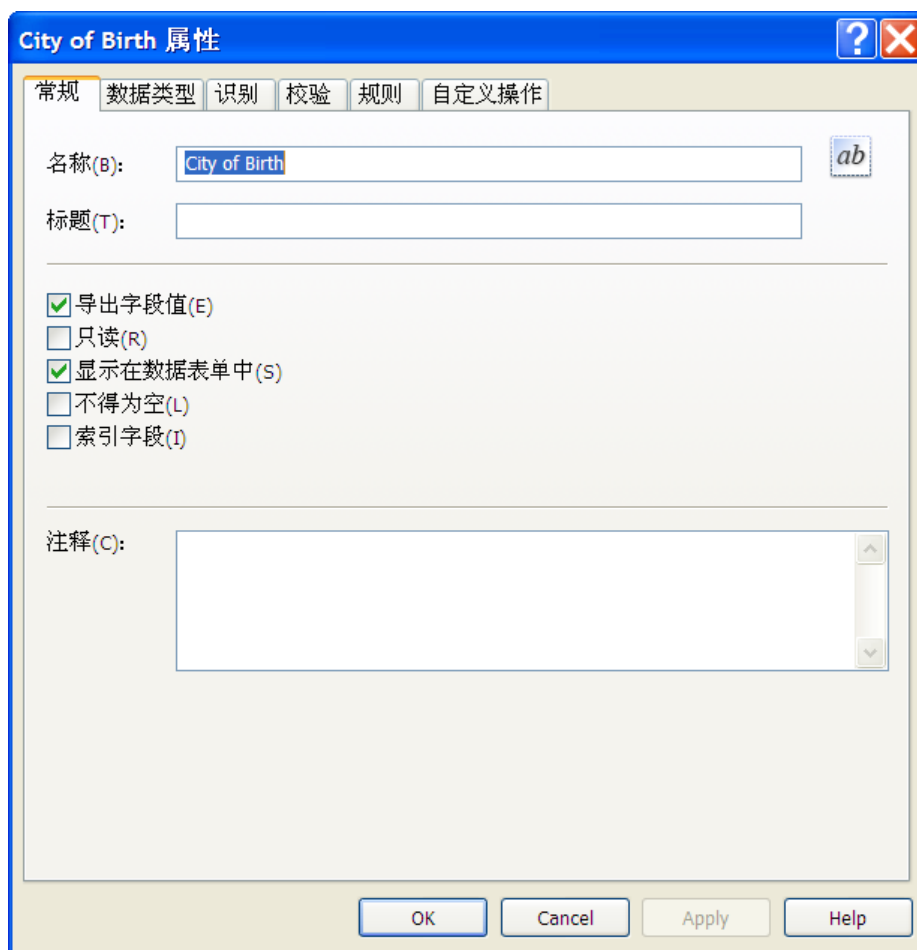


图 3. “字段属性”对话框“常规”选项卡

3.1.3.2. 数据类型

数据类型定义可能的字段值的集合及允许的字段格式。如果此字段中输入的值与指定的数据类型不一致，操作员将收到一条校验错误消息。文本数据类型通常有这些简单的有效值区域：日期、时间、地址、纳税人账号 (INN) 和总计。对于复选标记，数据类型表示的是选中或清除复选框时该字段所采用的值。

3.1.3.2.1. 文本输入字段的数据类型

必须正确设置文本输入字段的**数据类型**。此选项可告知程序要在字段中查找的数据类型 — 数字、字母、特定字符和日期等。该程序允许用户灵活设置数据类型。用户可轻松设置包含有常用类型的数据类型。此外，用户还可以根据所处理的任务创建自己的数据类型。

在设置数据类型时，可以配置系统来检查输入值的格式，还可以设置值的限制，例如最大最小值、日期期限、文本字段的最大字符数。

数据类型在“字段属性”对话框的**数据类型**选项卡上设置（图 4）。

从**内容**列表选择一种类别。在**详细信息**字段中，将显示属于此类别的一种数据类型（默认选择的类型或先前手动指定的类型）的描述。

如果选中了**将值作为文本处理**选项，包含任何内容的字段的值将作为文本进行处理并导出。字段格式检查则不会执行。

要更改识别**语言**或指定限制更多的数据类型，请单击描述右侧的**编辑...**按钮。

对于标准（或常规）类型（选择**内容设置**列表中的**常规**），可执行以下操作：

- 对于文本 – 选择多种识别语言（“...”按钮）。您可使用内置词典和/或自定义词典。然后，在字段识别过程中就会考虑词典中的信息。
- 对于数字 – 如果预计此值为整数，请选择**整数**。
- 对于日期 – 选择日期组成顺序，选择（不）显示单词中的月份，选择（不）显示时间和周中某日；
- 对于地址、名称和代码 – 使用自定义词典。

特殊类型（选择**内容设置**列表中的**特殊类型**）包含预定义数据类型，您可从中选择最适合的类型。在进行选择标记时，请注意对话框下方的说明。

如果此列表中没有符合您需求的类型，可以创建您自己的数据类型。

1. 要创建新数据类型，请在**数据类型**选项卡上选择**内容**列表的值。您可在列表中选择符合您的类型用途的任何值。得到的数据类型将存储在选定类别中，尽管类别本身并不影响这种新类型。
2. 单击**详细信息**字段右侧的**编辑...**按钮。在打开的对话框中，选择**内容设置**列表中的**特殊类型**。单击**新建...**。
3. 根据“新建数据类型向导”的指示操作。



图 4. “字段属性”对话框“数据类型”选项卡（文本输入字段）

可对任意类型的数据自动处理已识别的值。不必要的空格将会删除，大写和拼写将会得到更正。要启动自动处理，单击**替换字符**字段右侧的**编辑...**按钮。在打开的对话框中，指定所需的文本处理参数（图 5）。

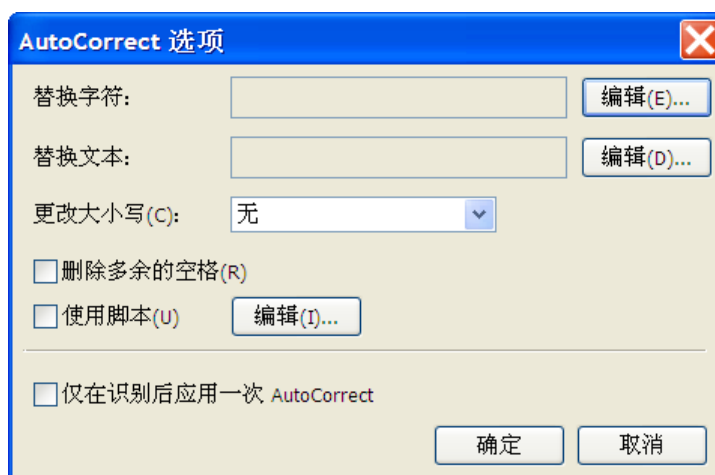


图 5. “自动替换设置”对话框

您还可以配置该程序，来检查识别的字段值是否符合设置条件（在这种情况下，如果是日期，本程序将检查此值是否在指定时期内；如果是文本字段，则检查此值是否有效、是否具备所需格式等）。要指定字段值的限制情况，请单击**检查值**字段右侧的**编辑...**按钮（图 6）。

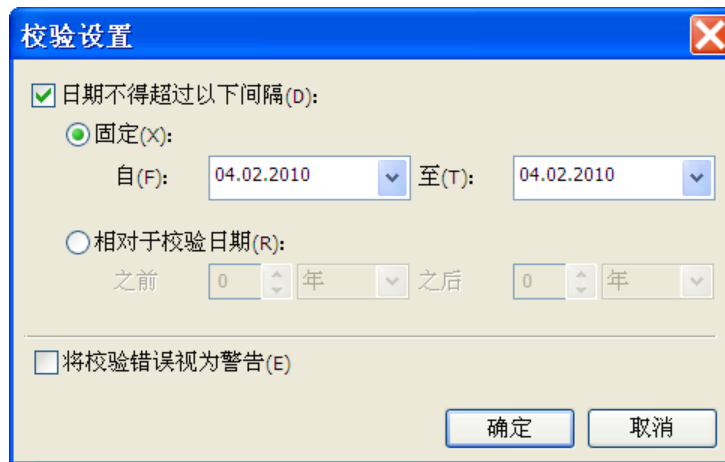


图 6. “检查值设置”对话框

3.1.3.2.2. 复选标记数据类型

对于复选标记，可在选中/清除复选标记时指定分配给字段的值。此操作在**数据类型**选项卡上执行（图 7）。

可分配给复选标记字段的值如下：

- 是 – 如果选中复选标记，否 – 如果清除复选标记；
- 1 – 如果选中复选标记，0 – 如果清除复选标记；
- 复选标记名称 – 如果选中复选标记，空行 – 如果清除复选标记；
- 也可选择相应选项来输入用户创建的值。

注：如果复选标记组合在一起，则这些复选标记将共享为整组创建的共同属性。复选标记值也在该组属性中创建。

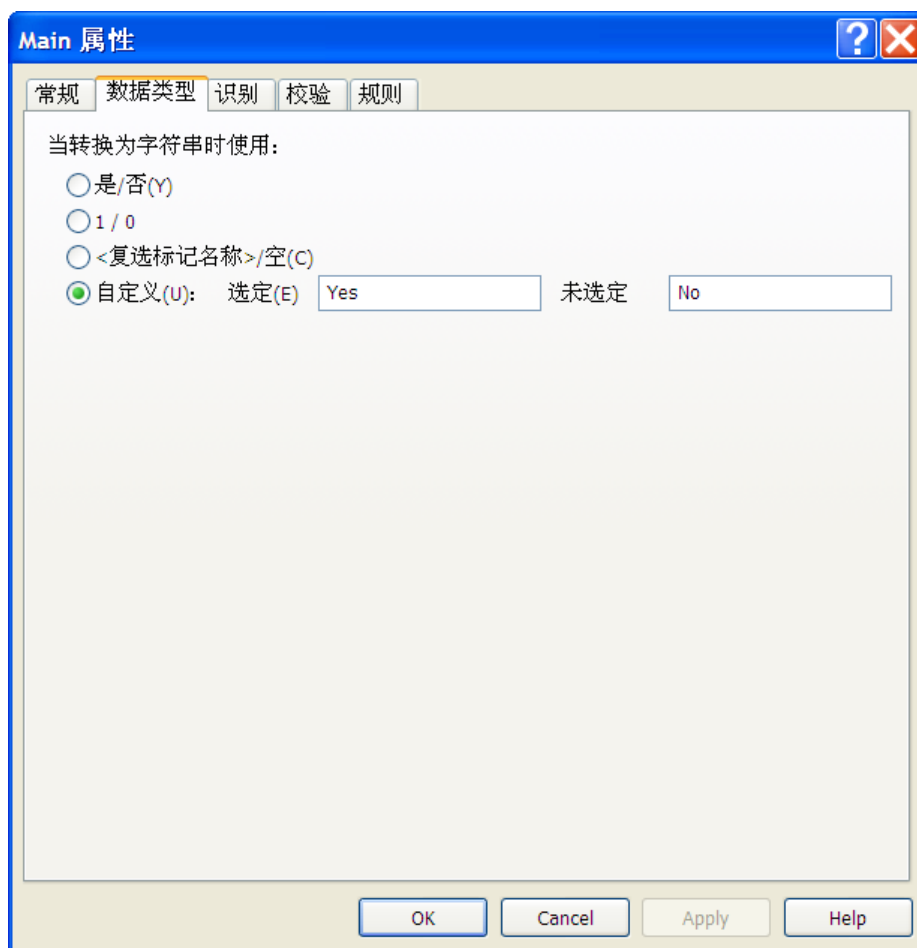


图 7. “字段属性”对话框，“数据类型”选项卡
(组中无复选标记)

3.1.3.2.3. 复选标记组数据类型

在“复选标记组属性”对话框中的**数据类型**选项卡上，显示有该组中复选标记的名称列表（图 8）。

如果希望组中至少有一个复选标记，请清除**允许空选**选项。

如果要允许选择组中的多个复选标记，请选中**允许多选**。

您可以对未选中任何复选标记，或者选中了多个复选标记时的导出值进行定义。选择<空选>或<多选>，单击**编辑...**。在打开的对话框中，在**导出值**字段输入一个值。如果未设置导出值，在未选中任何复选标记时将导出一个空行。如果选中了多个复选标记，则将导出由逗号分隔的值。

如果选中了**将校验错误视为警告**，将显示警告，不显示错误消息。

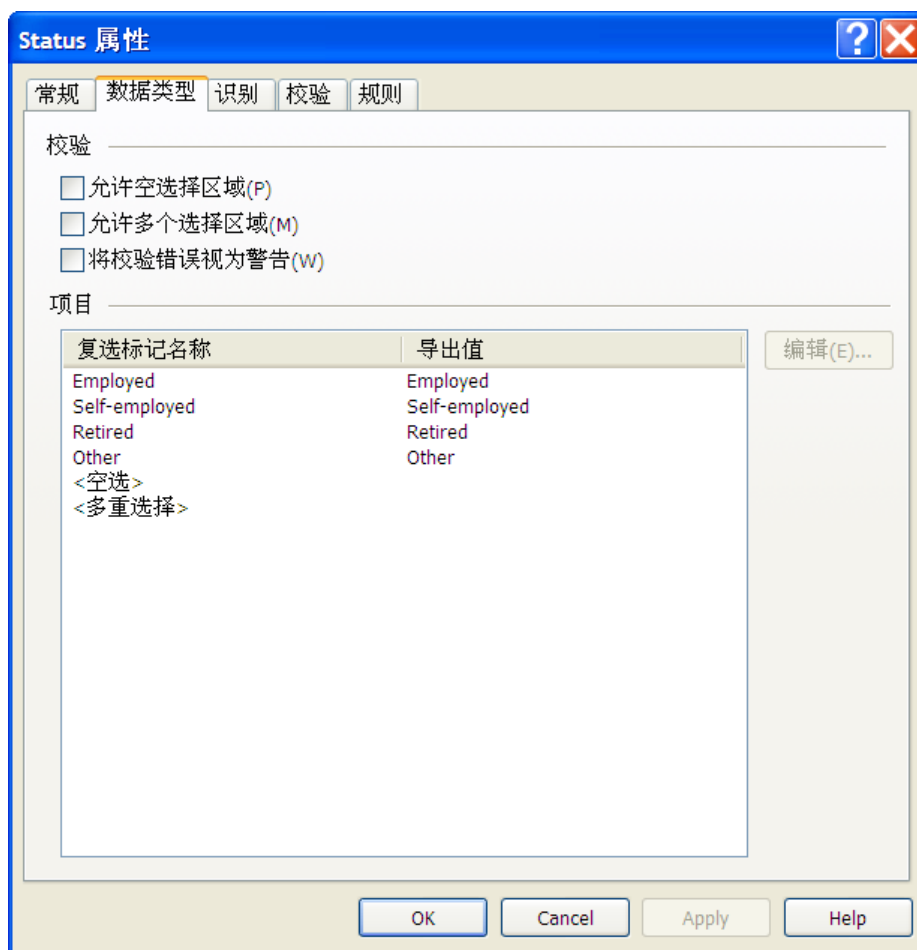


图 8. “字段属性”对话框，“数据类型”选项卡（复选标记组）

3.1.3.3. 字段识别属性

ABBYY FlexiCapture 9.0 允许为每个字段指定识别设置。在“字段属性”对话框的**识别**选项卡上正确定义字段属性将提高识别质量，错误可能性可降至最低。字段类型不同，属性也不同。按字段列示的识别属性如下。

3.1.3.3.1. 文本输入字段识别属性

如果不想识别该字段，或者如果将由操作员手动输入值，请选择**不要识别（将手动输入图像字段中的密钥）**。如果字段值因字符间距较小而不可识别，那么必须执行这项操作。在这种情况下，您不必指定其它识别属性，其原因是不会识别这类字段。在检查字段值期间，程序将提示操作员输入字段值。

现在选择文本类型：**ICR（手印文本）**或**OCR（打印文本）**。如果选择了**OCR**，请从下拉列表选择打印机类型（印刷打印机、点矩阵打印机和打字机等）。要设置多种文本类型或使用模板，请选择**高级**并单击**修改...**

从示例的下拉列表中选择**标记类型**。如果标记在扫描期间消失，建议选择适合固定间距文本（**灰色矩形**）的类型。如果标记没有消失，并且被分割成多个字符单元格，请指定字符单元格数（可自动定义）。无标记字段使用的是**简单类型**（通常用于印刷打印机上印制的文本）。

对于要查找特定大小写的字母，可选择**字母大小写**。如果该字段既可填入小写字母也可填入大写字母，请选择**自动**。

选择水平或垂直的**文本方向**。

对于一行字段，请选择**一行**。对于始终包含一个单词的字段（即无空格），选择**一个单词**。

指定图像处理属性。例如，通过调整要删除的“垃圾”的大小，对图像进行去斑。也可以反转图像或移除纹理。

注：如果选择了**图像去斑**选项，将默认为自动定义“垃圾”大小。要自定义大小，请选择**仅清除指定大小的垃圾**并指定此大小。

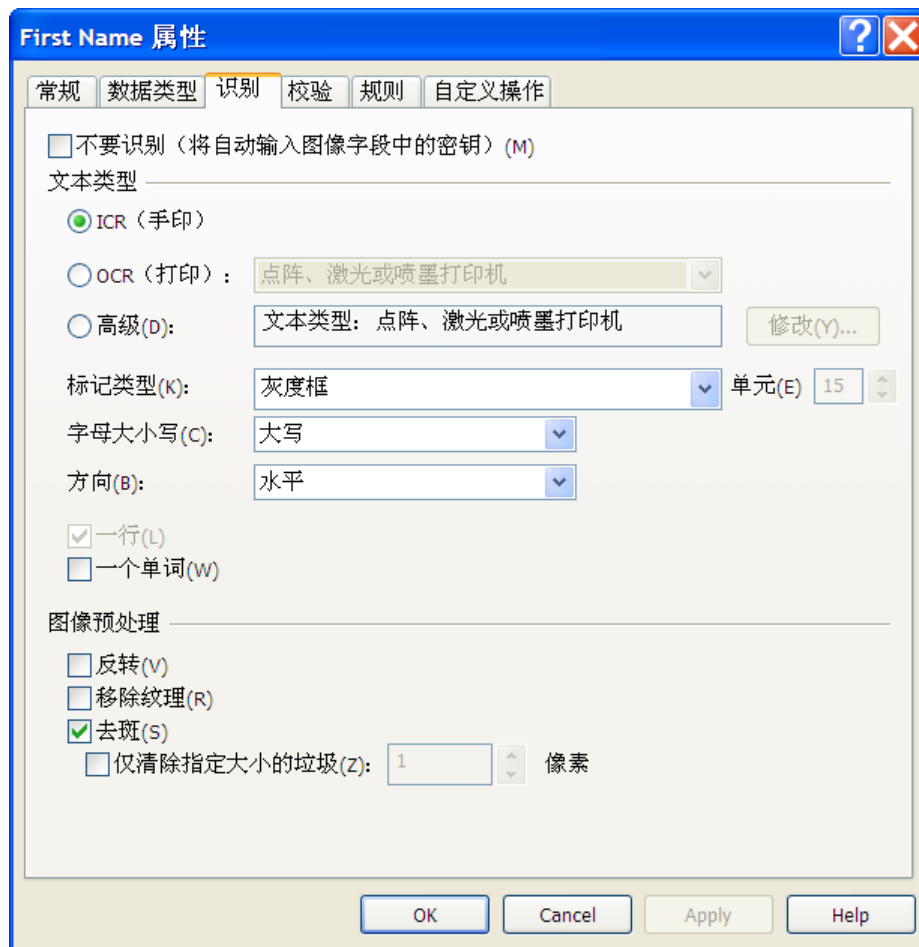


图 9. “字段属性”对话框，“识别”选项卡（文本输入字段）

3.1.3.3.2. 复选标记和复选标记组识别属性

要正确识别复选标记，请使用下拉列表指定**复选标记类型**。如果复选标记在方块中，选择**方块形**；如果要对没有任何框的白色背景设置复选标记（或者如果此框在扫描期间消失），选择**无框**。若选择**自动**，程序可自动定义复选标记类型。在这种情况下，复选标记必须为空，其原因是该程序将对所处理文档的此区域图像与文档定义所依据的表单中的图像进行比较，来定义是否有复选标记。

您可以允许纠正某些复选标记 – 如果复选标记的标记错误，则进行标记的人可以遮盖整个复选标记。被完全遮盖住的复选标记将被视为无效复选标记。但如果选择了**自动**类型，那么将不允许纠正。

也可以对复选标记来配置图像处理，与对文本字段的配置一样。

如果复选标记组合在一起，这些复选标记将共享相同属性。识别属性可类似地进行定义，但属于针对整组复选标记组进行定义。

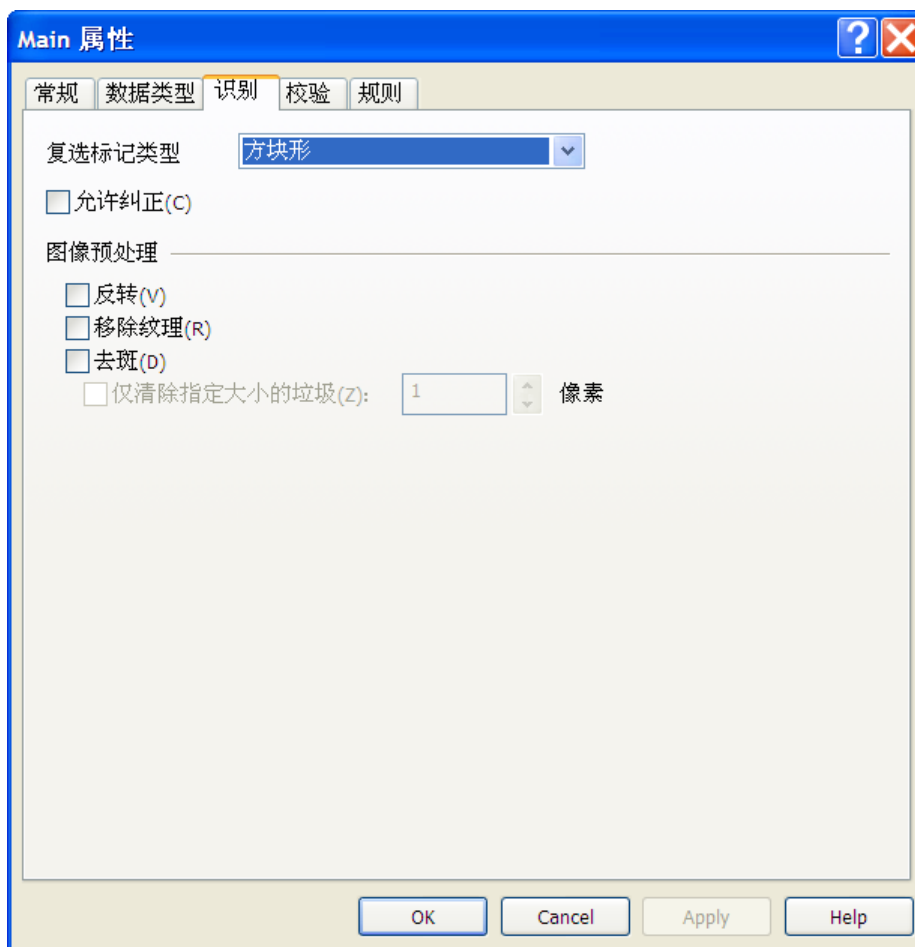


图 10. “字段属性”对话框，“识别”选项卡（复选标记字段）

3.1.3.3.3. 条形码识别属性

字段条形码的属性与静态元素条形码的属性相同。对于字段条形码，应指定条形码类型、方向和图像去斑设置。唯一的区别就是字段值可由操作员输入。要允许此操作，请选择**不要识别**（将手动输入图像字段中的密钥）。

3.1.3.3.4. 图像识别属性

如果要将图像区域排除出识别范围，请在“图像字段属性”窗口的**识别**选项卡上选择**排除识别**。（此选项仅适用“图像”字段类型，并且是用于与 FormReader 6.5 模板兼容。）

3.1.3.4. 校验设置

校验就是由操作员对识别的数据进行检查。在文档定义创建期间，您可以在“字段属性”对话框的**校验**选项卡上配置校验设置（图 11）。不确定识别的字符将被程序突出显示，并发送给操作员进行检查。但是，也可以这样配置校验：即便没有不确定字符，该字段也将发送给操作员；反之亦然，即便有不确定识别的字符，也不发送给操作员进行检查。如果要绝对确保字段值没有错误，请选择**全部（必须校验）**。

通过选择相应选项，可对所包含的字符启用组校验和字段校验。在组校验中，仅校验复选标记和数字。如果字段可能包含其它字符（字母），选择**字段校验**。

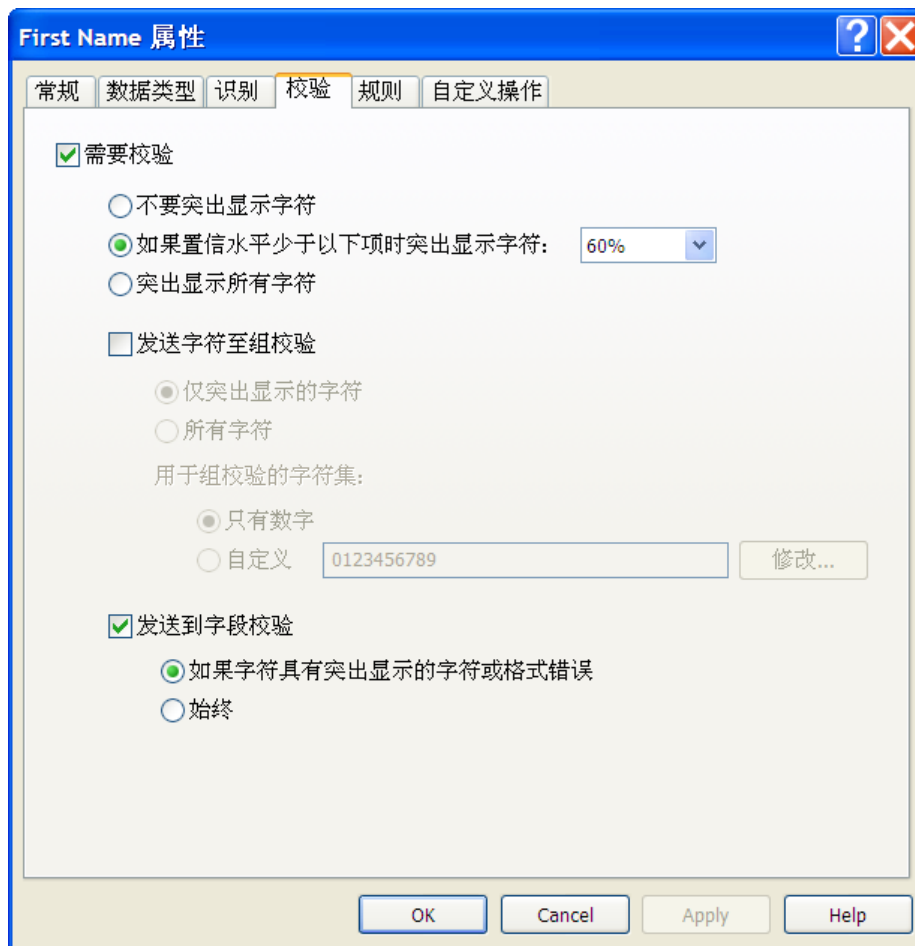


图 11. “字段属性”对话框，“校验”选项卡

3.1.3.5. 图像导出参数

在**图像**字段属性中，可指定导出参数，例如导出文件类型、质量、颜色类型和分辨率。要执行此操作，请转至“字段属性”对话框（在字段的快捷菜单上，单击**属性...**）的**导出**选项卡。可以配置参数如下：

- **文件类型**（TIFF、JPEG、BMP、JPEG2000、PCX packbits 和 PNG）。
- **质量**。对于 TIFF、JPEG 和 JPEG2000 文件，可以选择导出文件质量（最佳、高、正常和低）。

- **颜色类型** – 彩色、灰度和中间色（黑白）。

要指定所导出图像的分辨率，请选择**将分辨率更改为**，并输入分辨率或从列表中选择预定义的分辨率。

3.1.3.6. 规则检查

要自动检查识别数据，需要有**规则**。通过运用规则和数据类型，可以实施数据限制，例如，定义字段值必须满足的条件。否则，包含此类字段的页面将进行标记并标以相应的消息。规则主要用于文档完整性检查。也可以利用规则来处理识别的数据 — 组合多个字段值，或将识别的值替换为数据库中的值。

规则检查的优先级较低，排在自动纠正值、检查数据是否符合字段格式以及数据校验后面（检查日期是否在指定期限内，检查数字是否在指定间隔内等）。

ABBYY FlexiCapture 9.0 允许指定以下规则：

- **检查总和** – 检查多个字段值之和。然后，此和会与指定的数字或其他字段的值进行比较。例如，如果文档描述了员工的工资、奖金，并且包含有一个总收入字段，那么可以设置规则来检查工资和奖金之和是否与总费用一致。将发出一条警告消息。
- **比较字段** – 比较多个字段值（包括来自不同节段的字段之值）。如果您的文档中有多个字段的值必须相等，可使用此规则。如果值不相等，将发出规则检查错误。
- **数据库检查** – 通过将输入的值与特定的值进行比较，来检查所输入的值。
- **合并字段** – 合并多个字段值。可以用点、空格或其它分隔符来分隔字段值。例如，从用点分隔的已识别字段（日、月、年）编译日期，比将日期作为一个整体来识别更为方便。合并结果可以存储在文档定义的任何字段中。无标记字段可以很方便地用于这种用途（请参阅无标记的字段）。
- **总图数** – 总字数（俄语） – 将数字总和与用字书写的同一总和进行对比（只限于俄语）。
- **脚本** – 用户借助于脚本来描述限制。脚本语言在本应用程序的帮助文件中作了详细描述。由于未指定规则检查顺序，因此两个脚本规则不能同时使用同一个字段。只在只读模式下才对任意数目的规则可用。

规则在“字段属性”对话框的**规则**选项卡上指定（图 12）。规则可以影响一个或多个字段的值。

可以指定规则安全性（选择错误或警告）。如果发生错误，规则将标记为红色；如果发出警告，则标记为黄色。

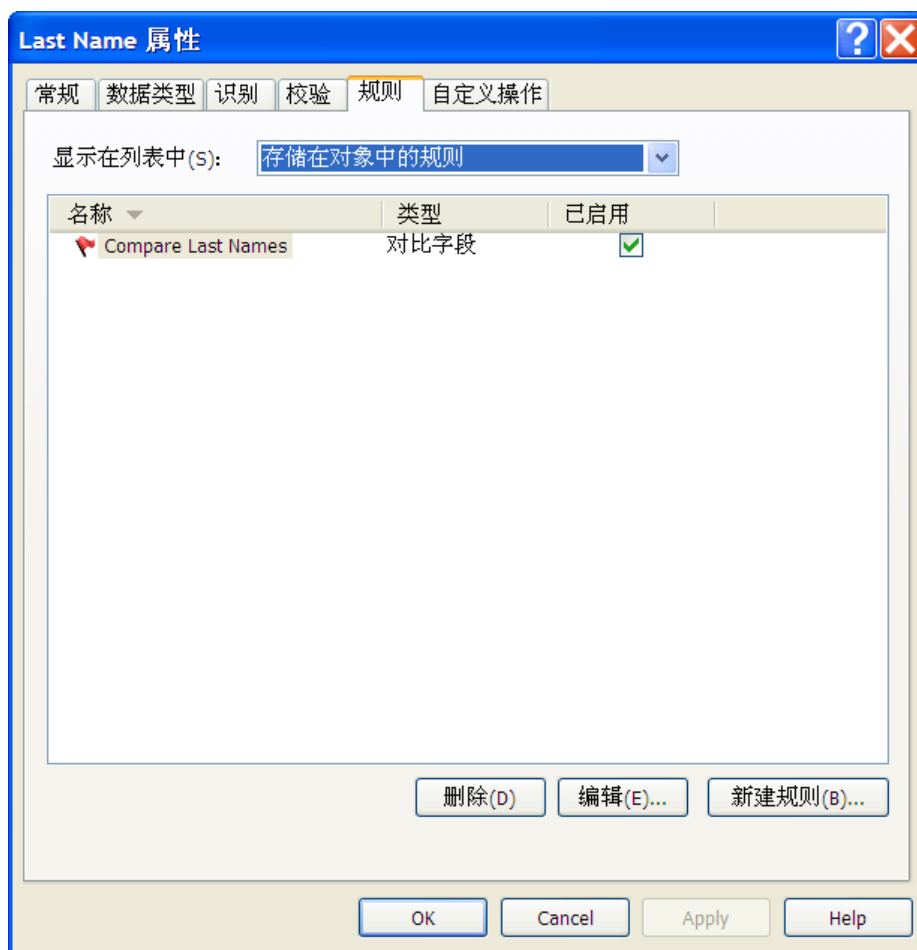


图 12. “字段属性”对话框，“规则”选项卡

3.1.3.7. 自定义操作/脚本编辑

可以创建自定义脚本。与自动触发的脚本操作相反，用户定义的自定义操作是在字段校验期间，或在文档编辑器中查看文档时由用户触发。单击字段值右边的 ... 可触发脚本。脚本语言在本应用程序的帮助文件中作了描述。

自定义操作在**自定义操作**选项卡上配置（图 13）。要添加或删除自定义脚本可用的字段，请分别使用**添加...**和**删除**按钮。要打开脚本编辑器，单击**编辑...**。要启用自定义操作/脚本，选择**启用自定义操作**。

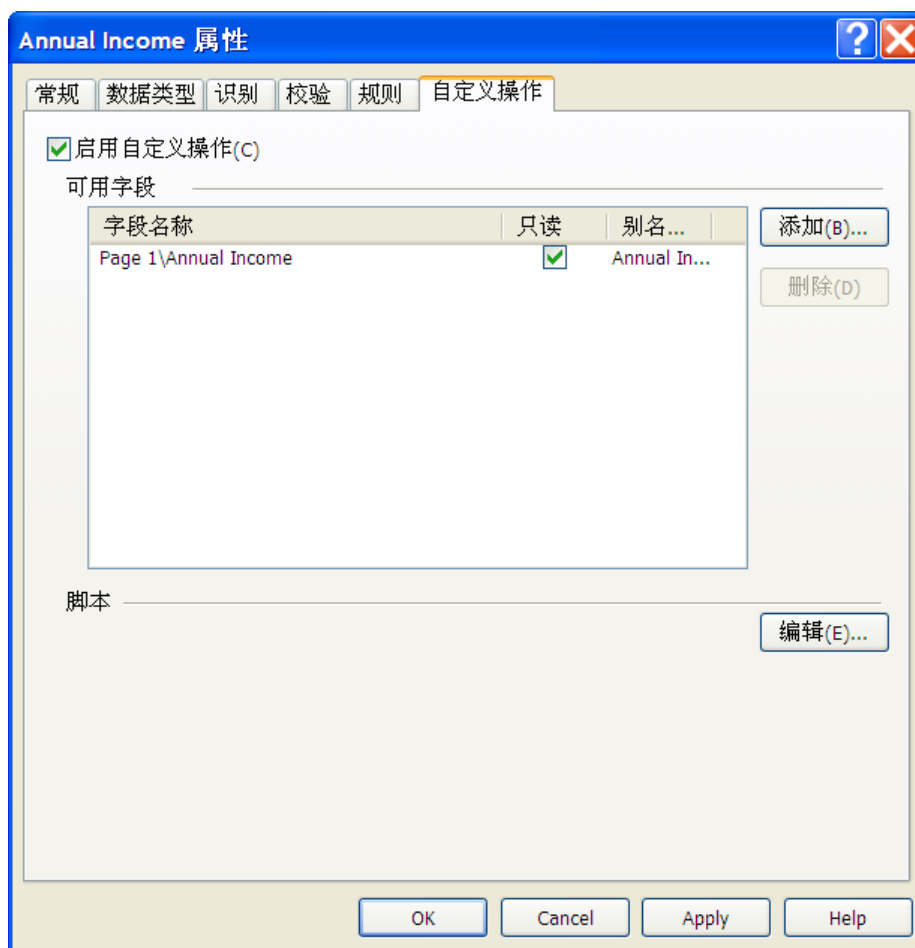


图 13. “字段属性”对话框，“自定义操作”选项卡

3.1.4. 创建多页文档的文档定义

ABBYY FlexiCapture 9.0 允许创建多页文档定义。一个定义可以由任意数目的 **节段** 组成，每个节段可以包含一个或多个 **页面**。可对多页文档定义指定节段顺序、数目和文档组装规则。

使用文档定义创建向导可以生成简单的文档定义（无节段）。您可以手动添加页面。在添加页面时，页面图像将显示在文档定义编辑器窗口的图像区域。在此页面上创建的字段将被添加至常规字段列表。在这种情况下，文档将由一个包含多个页面的节段组成。

如果文档包含跨多页的区域、重复区域或占用多页的表格，那么可以向一个节段添加多个页面。加载 FlexiLayout 可以采用相同方法（请参阅 [创建 FlexiLayout](#)）。FlexiLayout 的所有字段均将属于同一个文档定义。

您也可以向文档定义添加节段。最简单的情况是一个节段包含一页。这对于页面上设置的字段为独立字段的情况十分有用。这种方法相当透明，您可以轻而易举地查看字段所属的节段，并设置您自己的文档组装结构。例如，假设您的文档有 3 页，第 1 页和第 3 页只出现一次，而第 2 页出现 2 至 5 次。那么您必须为每页创建一个独立节段，然后设置文档结构，例如节段顺序和重复次数。

有一种更为复杂的情况，即文档定义包括多个节段，每个节段包括多页。例如，它可以是由非弹性节段和多页弹性节段组成的文档定义，或者描述由双面页面组成的、可能重复特定次数的这类

文档的文档定义。假设您要创建的文档定义包含一个标题页（页面 A），多个双面页面（页面顺序 B,C,B,C...）。在这种情况下，创建包含一个页面 A 的节段 A，创建多个包含两个页面（页面 B 和 C）的 BC 节段。该文档结构应是这种结构。节段 A 有一次重复，随后的 BC 节段重复 3 至 7 次。页面也可以在节段内重复。应广泛使用多次重复，以避免导致文档错误组装。

对于由多个节段组成的文档定义，除了设置页面顺序和重复次数外，还可以启用**关键字段**检查。在每个文档节段中，必须定义不同节段中有匹配的值的**关键字段**。例如，您可以将填写文档的人名设为**关键字段**。在这种情况下，必须填写每页上的名称。您还可以使用某些类型的标识编号。

在批处理文件处理期间，程序在检查组装规则时将尝试把连续页面组装到文档中。另外，还将校验**关键字段**值。如果**关键字段**值与同一个文档中各页面上的不匹配，将显示一条警告消息。如果在扫描期间混淆了文档页面，就有可能发生这种情况。在这种情况下，需要更改页面顺序，方可符合文档组装规则。

要向节段添加页面，请在文档定义编辑器菜单上单击**文档定义 > 添加页面...**，或单击此图像快捷菜单上的**添加页面...**。加载空页面的图像，并选择要在此页上自动标记的对象类型。

要添加新节段，请在文档定义编辑器菜单上单击**文档定义 > 添加文档节段...**。屏幕上将打开**创建新的节段布局**窗口，在这里可以指定所需的全部参数。根据屏幕指示操作：

- 指定节段名称；
- 加载空页面图像（扫描或从文件加载）；
- 根据需要加载 FlexiLayout；
- 选择要自动标记的对象类型。

新的节段及其字段将显示在文档结构窗口中。

要查看创建文档定义时生成的文档结构，请单击文档定义编辑器窗口右侧**文档结构**窗口中的**页面**选项卡。屏幕上将显示文档结构和页面缩略图。在这个选项卡上，可以更改文档节段数。要执行此操作，请在节段名称右侧的方块中输入新值（数字或范围）。您也可以添加附录页面，更改节段的页序，使用拖放法将页面从一个节段移到其它节段。

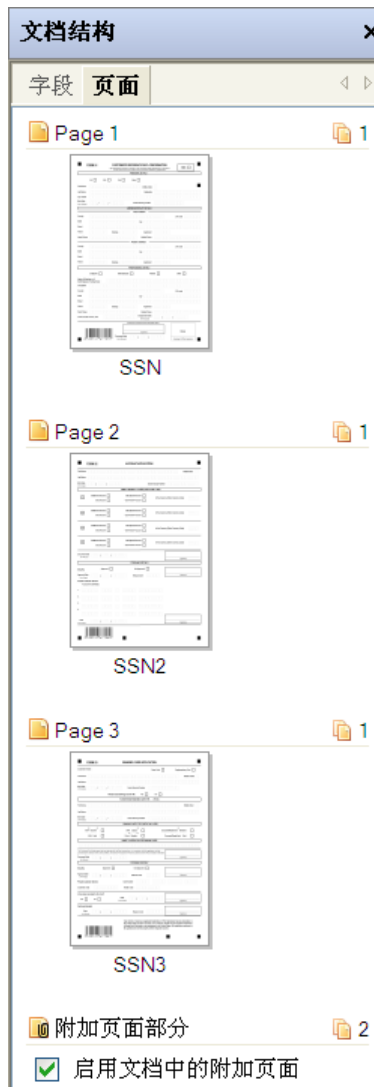


图 14. 文档结构，缩略图视图模式

您也可以在文档定义编辑器窗口单击**文档定义 > 文档定义属性...**，指定节段重复顺序和次数。

在**组装**选项卡上（图 15），指定文档中的节段最大/最小重复次数（默认次数为 1）。

如果要检查关键字段值，选择**检查关键字段的均等性**，然后为每个节段选择一个关键字段。通过双击左键即可选中。

在某些情况下，可能需要禁用节段的顺序检查，例如，它对于文档组装并不重要时，就可以禁用顺序检查。要执行此操作，选择**禁用节段顺序检查**。但仍会检查该文档中所示的所有节段，不过不会检查节段顺序。

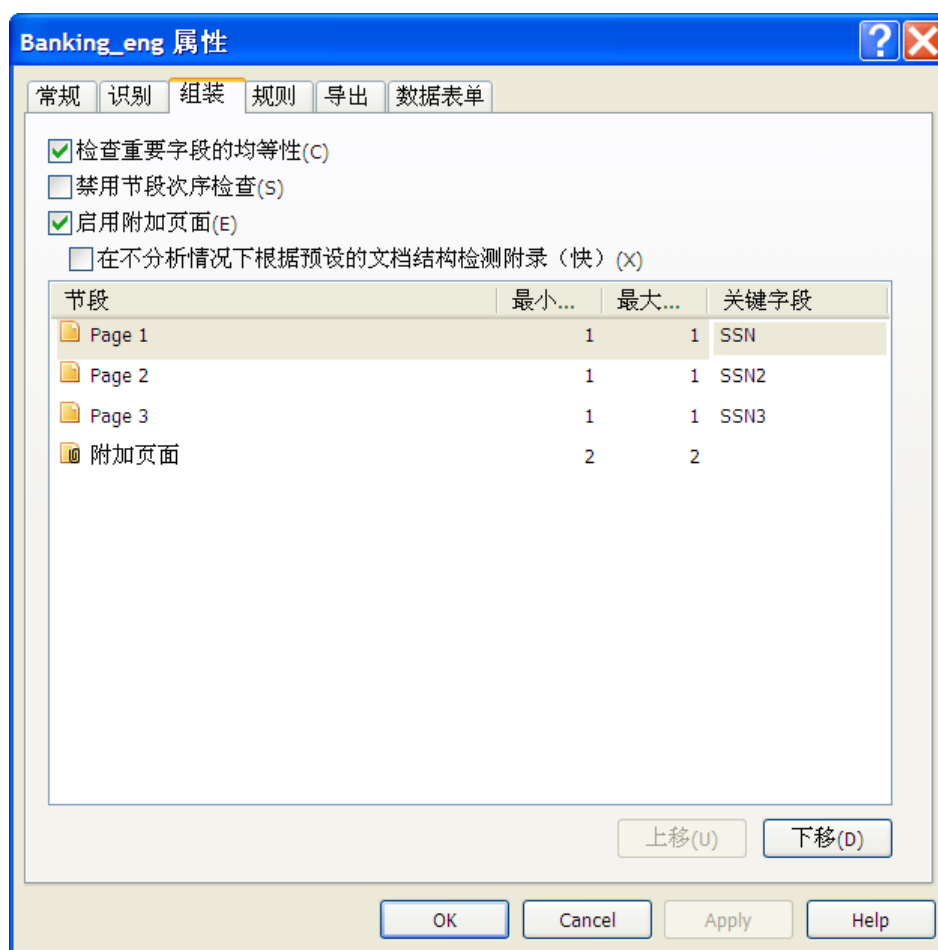


图 15. 文档定义属性，“组装”选项卡

3.1.5. 创建包含附录页面的文档定义

ABBYY FlexiCapture 9.0 允许为带附录的文档创建文档定义。

任何文档都可以附带附录页面。这些页面不要求进行独立字段搜索，因此不需要为附录页面匹配文档定义。但是，在文档组装期间同样会将附录页面考虑在内。例如，信贷申请表（结构化表单）可以附带以任意形式书写的推荐信。这类推荐信就可以作为附录页面处理。

要创建带有附录的文档定义？请执行以下操作：

- 在文档定义编辑器窗口右侧的**文档结构**窗口中，单击**页面**选项卡并选择**启用文档中的附录页面**。在节段名称右侧显示的方框中输入附录页面的数量（或范围）（图 14），或
- 在文档定义编辑器窗口中，单击**文档定义 > 文档定义属性...**，打开**组装**选项卡。选择**启用附录页面**。然后，指定最小和最大附录页面数（图 15）。

要保存附录页面图像，启用图像保存。在导出设置窗口中单击**编辑**，然后在打开的对话框中单击**保存文档图像**，可执行此操作。单击**图像格式**选项，将图像保存为图形格式或可搜索的 PDF。

3.1.6. 导出设置

要对文档处理期间获取的数据进行保存，需要为每个文档定义配置导出。导出类型有四种：导出至指定格式的文件、导出至 ODBC 兼容的数据库、导出至 MS Sharepoint 库及自定义导出（使用脚

本)。导出是在**导出设置**对话框（**文档定义 > 导出设置**）中配置的。要添加新的导出目标，单击**添加...**。要编辑现有导出目标，单击**编辑...**。

除了在文档定义中设置外，还可以在批处理文件和项目级别配置导出。在批处理文件级别配置的导出将用作默认的批处理文件。在操作员单击**文件 > 导出至**来启动导出时，将默认使用这些设置。

可以使用导出目标向导添加新的导出目标。首先，必须指定常规设置（类型和接收要求），然后设置选定类型。

您可以为一个文档定义指定多个导出目标。导出目标可能重要，也可能不重要。如果在导出至重要目标时出错，将终止处理。如果在导出至非重要目标时出错，将继续处理，并导出此文档。因此，建议至少设置一个重要导出目标，否则导出错误将不被注意到。您可以在**导出设置**窗口中将某个导出目标设置为重要导出目标。

也可以导出页面图像 — 与数据一起导出或作为独立的导出过程导出。可指定每个导出目标所要编辑的敏感信息。要同时导出图像和数据，请在“导出目标向导”对话框的第一个页面上，选择“类型”下拉框中的**导出到图像文件**。

注：如果使用的是分布式版本，必须以 UNC 格式（\\server_name\share_name[path]）指定已上传项目中的文件路径，才能正确完成导出。如果指定的是相对路径，导出结果仅在使用应用程序界面 (API) 时可用。如果以 <disk>:[path] 格式指定绝对路径，将发生错误并终止导出。

如果指定的是相对导出路径，则可在上传项目到服务器时以 UNC 格式指定根路径。在这种情况下，相对路径将根据相对导出路径来形成。

您也可以这样配置目标，即只导出没有错误的文档或只导出有错误的文档，否则将导出所有文档。根据配置情况使用**导出设置**中的**无错误的文档**和**有错误的文档**。

可以自动执行导出，也可以根据操作员的命令执行导出。如果使用的是分布式版本，需要在项目设置中创建一个导出确认队列，以允许操作员启动导出。如果使用的是单机版本，请选择**导出设置**窗口中的**启用自动报告**。

3.1.6.1. 导出至文件

如果要配置导出至文件，选择**类型**字段中的**导出到数据文件**。指定常规导出设置，单击**下一步**。

在**路径**字段中输入存储导出文件的文件夹路径，或单击**浏览...**来使用文件夹树状图。

您可以将一个批处理文件的文档导出至一个文件，或者将每个文档导出至一个独立文件（选择为**每份文档创建独立的文件**）。也可以选择为每个批处理文件创建一个独立文件夹（选择为**每个批处理文件创建独立文件夹**）。

如果希望在导出期间用新文件覆盖同名的现有文件，选择**覆盖现有文件**。

接下来，指定所导出文件的文件命名选项。单击**文件命名选项...**并选择所需选项。

单击**下一步**继续操作。

选择文件类型（CSV、DBF、TXT、XLS 或 XML），并调整其它属性，例如更改文本编码。

单击**下一步**继续操作。

如果要一同导出图像和数据，请配置图像导出参数（如“导出图像”一节所述）。

输入导出目标的名称，单击**完成**。

3.1.6.2. 导出至数据库

要启用导出至数据库并进行配置，选择**类型**字段中的**导出至 ODBC 兼容数据库**。选择所需选项，单击**下一步**。

输入**连接字符串**或单击**设置连接...**，在**数据链接属性**对话框中配置必要参数。然后，从下拉列表中选择方案。

单击**测试连接**，可测试数据库连接。

然后，需要定义“文档描述”字段值所要导出到的数据库的表格和表格列。要执行此操作，单击**设置字段映射...**

在**字段映射**对话框的左侧，可以看到文档的各节段和字段。将右侧的表格和数据库字段与这些节段和字段进行匹配。如果该数据库已包含文档导出表格，请为每个节段和每个字段的表格列选择一个数据库表格。表格列格式必须与所导出字段的值一致。

如果数据库没有文档导出表格，则可在第一个导出时启用创建自动数据库表格。要执行此操作，请单击**自动创建表格**。程序将采用适合导出的结构来创建表格。表格列将与文档字段进行对应。

字段组、带有多处重复的字段、文档表格和节段将被导出至独立的数据库表格。主键和外键用于将主要表格与附属表格相链接。在主要表格中，主键添加至每个条目中。在附属表格中，每个条目均有一个外键，其记录了主要表格中相应条目主键的值。如有必要，将自动添加主键和外键。您只需指定要将键导出到的字段。

默认情况下，**显示链接的列**选项已选中。如果清除此选项，树状图右侧不会显示所配置导出的列。

单击**下一步**继续图像导出设置。

您可以将图像保存至数据库或作为文件保存。如果将图像保存为文件，请根据“导出图像”一节所述来配置图像导出参数。

注：如果选择将图像保存至数据库，将为这些图像创建一个 **BLOB** 类型表格。如果以后更改了图像导出并选择将图像保存至文件夹，由于图像文件的文本路径将被导出至 **BLOB** 类型字段，因此导出时将发生错误。

输入目标名称，单击**完成**。

3.1.6.3. 导出至 SharePoint

ABBYY FlexiCapture 9.0 允许将文档导出至 Microsoft™ SharePoint 库。每个文档都有匹配的列，这些列包含有这些文档字段的值。这些值可用于文档搜索和建立索引。

注：

1. 必须具备管理员权限方可配置导出至 SharePoint。但是，如果是执行导出，那么具备添加者 (Contributor) 权限就足够了。

2. 要用于数据存储的 SharePoint 列，必须为单行文本或多行文本类型。

要配置导出至 SharePoint，从**类型**字段选择**导出至 SharePoint**。选择所需的选项，单击**下一步**。

输入存储 SharePoint 库的**服务器 URL**。根据需要，单击**连接设置...** 配置验证参数（默认为采用 Windows 验证）和代理参数。

如果连接丢失，单击**连接**以重新连接服务器。

从下拉列表选择**选择一个 SharaPoint 库**。

在**内容类型**字段中选择文档**内容类型**（SharePoint 2007 或更高版本）。在这里，可以将值导出配置为导出至当前值类型对应的字段。单击**设置字段映射...**，在**字段映射**对话框中配置必要的设置。

单击**下一步**。

配置图像导出参数，再次单击**下一步**。

输入目标名称，单击**完成**。

3.1.6.4. 导出图像

要配置图像导出，从**类型**字段选择**导出至图像文件**。选择所需的选项，单击**下一步**。

在**路径**字段中输入存储导出图像的文件夹路径，或单击**浏览...** 来使用文件夹树状图。

您可以将一个批处理文件的文档导出到一个文件（选择为**每个批处理文件创建独立文件夹**）。如有必要，选择**覆盖现有文件**。

配置文件命名选项。单击**文件命名选项...**并在打开的对话框中选择所需标准。

如果需要编辑导出图像的机密信息，选择**编辑图像中的敏感性数据**。然后，单击**选择要编辑的字段...**，并选择要编辑的字段。

单击**下一步**。

在**类型**字段中选择图像存储格式。然后，配置所保存图像的质量和颜色类型。

如果要更改初始图像分辨率来减小文件尺寸，选择**将分辨率更改为**并输入一个分辨率值。

如果选择 PDF 格式并且选中了**创建可搜索 PDF** 选项，文档图像将以全文本模式搜索，并且识别的文本将以选定格式保存。在这种情况下，可以指定识别语言：可以选择保留文档定义中使用的语言，或从列表中选择一种或多种语言（单击**选择...**）。

单击**下一步**。

输入目标名称，单击**完成**。

3.1.6.5. 自定义导出（脚本）

通过采用脚本编写方法，可以创建自定义导出程序，这比采用标准方法更灵活。

要创建自定义导出程序，请选择**类型**字段中的**自定义导出（脚本）**。选择所需的选项，单击**下一步**。

选择脚本编写语言（JScript® 或 VBScript），在单击**编辑脚本...**时显示的脚本编辑器中输入脚本编写文本（有关脚本编写语言的详细说明和示例，请参阅帮助文件）。

3.1.7. 配置识别数据的显示

在识别数据后，操作员将在文档窗口中看到所识别的数据。这些数据在默认情况下以列表的形式显示，并且标题对应文档定义字段的名称。但是，也可以更改数据显示，以更加方便的方式排列数据或更改元素标题。数据显示可以在文档定义编辑器窗口的右下角进行编辑。

字段可以移动、重命名和更改。要添加文本，单击快捷菜单上的**插入标签框**。

在文档定义属性对话框（**文档定义 > 文档定义属性...**）的**数据表单**选项卡上，可以指定字段名称和已识别数据值的文本字体和字号。

3.1.8. 测试文档定义

在启动批量输入文档前，必须测试所创建的文档定义。

可以使用文档定义所创建的图像，直接从文档定义编辑器测试该文档定义。要启动测试，单击**测试 > 运行测试**。

测试涵盖添加至测试批处理文件的文档，其原因是这些文档可使用未发布的文档定义版本。这些批处理文件可以从本程序的主窗口（**视图 > 测试批处理文件**）进行访问。您也可以单击**批处理文件 > 切换至测试批处理文件...**，从文档定义编辑器访问测试批处理文件。

脚本编写规则可从脚本编辑器（单击**工具 > 检查脚本**）来直接检查。您可以利用这项功能来检查有不同字段值的脚本活动。

如果在测试期间发生错误，或者如果发现无效字段属性，请编辑文档定义来解决这些错误。纠正完所有错误时，即可发布文档定义，并继续处理批量文档输入。

3.1.9. 编辑和发布文档定义

在创建文档定义并对多个图像成功测试之后，可发布该文档定义，以便在识别工作批处理文件时访问此文档定义。

要发布文档定义，单击**文档定义**窗口（**项目 > 文档定义...**）中的**发布**。

要返回来编辑文档定义，请在**文档定义**对话框中选中该文档定义并单击**编辑...**。其他用户不可编辑该文档定义，并且将采用最新发布文档定义版本来进行识别。因此，在编辑文档定义时，不能按操作员的方式进行工作。完成文档定义编辑后进行发布，新版本即可对各个用户可用。如果不想发布已编辑的版本，可放弃所有更改，返回最新发布版本。要执行此操作，单击**文档定义**对话框中的**放弃更改**。

在发布新版文档定义时，将采用此版本来识别文档。对于先前已识别的独立文档，仅限此版本的更改不涉及更改的字段位置和识别属性（例如，可以更改导出设置）时，您可以更改所使用的文档定义版本且不需重新识别。要执行此操作，单击主菜单上的**工具 > 更新至最新版本**。

3.2. 创建 FlexiLayout

配置 FlexiLayout 处理是一项较为细致的工作。发票、订单和其它相似文档属于非结构化文档（不同副本上的字段的位置、大小和数目均有不同），因此不能创建有固定字段位置的常用布局。对于这类文档，将在预识别整页后搜索字段，并且使用关键字和相对字段位置来创建文档定义（这些定义称作 FlexiLayout）。ABBYY FlexiLayout Studio 9.0 这款程序设计特别，其目的就是为了创建 FlexiLayout。有关创建 FlexiLayout 的详细信息，请参阅 ABBYY FlexiLayout Studio 9.0 帮助文件。

您创建的 FlexiLayout 将导出至 .afl 文件，然后在 ABBYY FlexiCapture 9.0 中使用。FlexiLayout 将用作文档定义的一个节段。因此，文档定义可以同时包含结构化和非结构化的节段。FlexiLayout 可以由多页组成。

在以后创建文档定义时，可以将 FlexiLayout 附加到文档定义。要执行此操作，在创建文档定义的第二步（创建新文档定义向导），添加文档图像，然后选择**加载 FlexiLayout**，并输入带有 FlexiLayout 的 .afl 文件的路径。

您也可以使用“属性”对话框，从文档定义编辑器加载 FlexiLayout。要打开此对话框，单击节段快捷菜单上的**属性...**。单击**FlexiLayout**选项卡，然后单击**加载...**。选择带有 FlexiLayout 的 .afl 文件。

在加载 FlexiLayout 时，所有字段及其几何标记都会显示在图像上。

文档描述中的字段数必须与 FlexiLayout 中的字段数一致。建议不要添加或删除字段（在图像上不占区域的字段除外），只能通过更改 FlexiLayout 本身来更改所识别字段的设置。

现在，需要指定字段属性，其过程与结构化文档的一样。配置识别、校验、导出属性及规则。

3.2.1. 分类器

分类器是一个特殊的 ABBYY FlexiLayout Studio 项目，旨在初步识别独立的页面，并将这些页面标为特定类型（例如通过使用 FlexiLayout/FlexiLayout 变体）。分类器在 ABBYY FlexiCapture 中用于为批处理文件类型级别的文档进行分类（对于默认的批处理文件，将在项目级别添加分类器）。

分类器可以用于：

- 加快文档定义选择的过程。通过使用分类器，可以根据分类符号选择文档定义。您不必遍历所有的文档定义来选取所需的文档定义。在处理包含多个文档定义的大型项目时，速度提高尤其明显。
- 在分类器树状图中发布分类功能。例如，您可以在最顶层根据一个包含该公司名称的文档节段，来选择在该公司内部发送的文档。在下一级别，可根据文档的标题或特定页面上的分隔符数对文档进行分类。
- 管理第三方 FlexiLayout 应用程序。第三方开发人员创建的 FlexiLayout 可能与其文档不匹配。在这种情况下，您可使用分类器来补救，无需联系开发人员。
- 加速包含多语种文档定义的项目。如果使用了分类器，则仅执行两次文档预识别：当分类文档时及在应用选定的文档定义时。否则，必须为项目中每个语种的文档定义执行预识别。

有关分类器的详细信息，请参见 ABBYY FlexiLayout Studio 帮助文件。

可以在批处理文件类型级别或在项目级别添加分类器（后面一种情况仅适用于默认类型的批处理文件）。

要添加分类器，执行如下操作：

- 在主窗口中，单击**项目 > 项目属性...**，在项目级别添加分类器（或单击**项目 > 批处理文件类型...**，然后单击**编辑...**），在批处理文件类型级别添加分类器；
- 在“批处理文件类型属性”对话框中，单击**识别**选项卡，然后单击**加载...**；
- 选择包含分类器的 *.cfl 文件。

3.3. 非结构化文档的规格

如果您要处理包含自由表单信息的非结构化文档，如合同、信函、订单、图表，ABBYY FlexiCapture 9.0 同样可以提供解决方案。包含文本或图像、由空页面分隔或带条形码的非结构化文档，将进行处理并导出至可搜索 PDF 文件或图像格式的文件。

处理此类文档通常要将文档转换为电子形式，并根据关键字段值来组织搜索。

通常是使用 ABBYY FlexiLayout Studio 中创建的 FlexiLayout 来搜索关键字段（例如合同号）。

如果不能自动搜索关键字段，可以由操作员输入关键字段值。要由操作员输入字段值，请创建包含一个（或多个，根据需要）字段的文档定义，并在该字段的识别属性中，选择**不要识别（将手动输入图像字段中的密钥）**。程序将提示操作员在启动校验时手动输入关键字段值。

要存储文档，您需要配置导出。关键字段值可以导出至文件或数据库，文档图像则可按所需的格式保存。要导出图像，请在**导出设置**对话框的**图像**选项卡上，单击**保存文档图像**并配置图像保存参数。您可将文档图像保存在图形文件或可搜索的 PDF 文件中。

在将页面组装到文档中时务必小心。如果是非结构化文档，可能较难定义一个页面源于哪个文档。建议用空页面或条形码页面来分隔文档，以便自动组装非结构化文档。然后，在向批处理文件添加图像时（从扫描仪、文件或在导入配置文件创建期间），选择**适用于由以下项分隔的图像**，然后根据用作分隔符的页面，从下拉列表中选择**空白页面**或**带有条形码的页面**。这些页面将被自动组装到文档中 — 在遇到分隔符页面前，这些页面将添加至当前文档。

4. 配置项目属性

项目属性可以在**项目属性**窗口中配置。单击**项目 > 项目属性...**，可打开此窗口。

“项目属性”窗口有六个选项卡。**常规**选项卡描述常规的项目属性 – 名称和描述等。其余选项卡与**批处理文件类型**属性窗口中的选项卡相同（请参阅“配置批处理文件类型”），用来描述默认类型批处理文件的属性。

常规选项卡允许配置如下项目属性：

- **名称；**
- **描述；**

- **导出根路径** — 文件夹路径，在项目上传到服务器后，相对文件导出路径就是源于该文件夹。此路径以 UNC 格式指定。

您还可以查看项目文件名及其在文件夹树状图中的位置（对于应用程序服务器上的项目，则是 GUID）。

其余选项卡与批处理文件类型属性窗口中的选项卡相同，它们是用于配置默认的批处理文件类型。

在配置项目属性时，可以指定文档定义所使用的顺序列表。这项操作是从**文档定义**窗口（项目 > **文档定义...**）执行的，与批处理文件类型相反。

FlexiLayout 按照其在此窗口中列出的相同顺序进行匹配（自上至下）。要更改文档顺序，请使用**向上**和**向下**按钮。

识别过程通常由以下过程组成：连续匹配添加的所有 **FlexiLayout**，对这些 **FlexiLayout** 的质量进行检查，选择一个最佳 **FlexiLayout**。要加快文档定义匹配，请选择**使用第一个可接受弹性布局**。在这种情况下，如果有一个 **FlexiLayout** 匹配，那么就不再考虑其余的 **FlexiLayout**。当您有很多个不同的 **FlexiLayout** 时，由于这样可以减少总识别时间，因此这个选项可能会很有用。

文档定义可以临时禁用。要禁用文档定义，请清除文档定义列表中的**已启用**选项。要启用文档定义，请选中此选项。

5. 配置批处理文件类型

ABBYY FlexiCapture 9.0 支持在批处理文件中执行文档处理：在处理前会创建一个批处理文件，文档图像会被添加至要该批处理文件中来一同处理。批处理文件既可以自动创建，也可以在图像导入期间手动创建（从扫描仪或图像文件）。批处理文件处理设置在批处理文件类型属性中进行配置。您可以为最常用的处理设置创建多种批处理文件类型。在这种情况下，操作员只需指定批处理文件类型。

批处理文件类型定义的文档处理参数如下：

- 图像处理方法；
- 文档定义的列表和匹配顺序，所使用的分类器；
- 批处理文件完整性检查；
- 手动导出的默认设置；
- 不可识别文档的导出设置；
- 批处理文件处理。

批处理文件类型还定义了批处理文件注册参数的列表，例如，对某个给定批处理文件进行特征化的文本信息。

批处理文件类型在项目中创建。要查看批处理文件类型的列表，单击主菜单上的**项目 > 批处理文件类型...**。

要创建新批处理文件类型，单击**新建...**。然后，在“创建新批处理文件类型”窗口中配置所需的参数。

要编辑批处理文件类型，选中该类型并单击**编辑...**。将打开**批处理文件类型属性**对话框。

批处理文件类型设置分为六组，每组都有一个独立的选项卡。在批处理文件类型创建期间，每组对应一个阶段。

5.1. 常规

常规属性如下所示：

- **名称：**
- **描述：**
- **注册参数。**注册参数属于文本属性，可用于识别给定的文本。扫描日期、扫描操作员的名称或其他信息都可以作为这类属性使用。在创建选定类型的批处理文件时，将指定注册参数。注册参数可以是必填参数，也可以是选填参数。

5.2. 图像预处理

可以为批处理文件类型配置如下的文档创建参数和图像处理参数：

- **创建新文档。**在为每个图像文件，或为由空白页面或（特定类型或特定值的）条形码页面分隔的图像文件匹配文档定义时，将自动创建一个新文档。如果分隔符页面未进行处理，可选中**删除分隔符页面**来删除这些页面。
- **图像处理。**可以指定的图像处理参数如下：
 - **图像去斑；**
 - **将彩色和灰色图像转换为黑白图像；**
 - **图像旋转 90°（逆时针）顺时针或 180°（倒置旋转）；**
 - **使用黑色分隔符纠正歪斜；**
 - **使用黑方块形纠正歪斜；**

当操作员在不适用导入配置文件的情况下向这类批处理文件添加文档时，将默认为使用这些设置。

5.3. 识别

可以指定的识别参数如下：

- **文档定义。**您可以使用项目中的所有文档定义，或选择当前批处理文件类型所要使用的文档定义。如果您在使用文档定义列表，则可以指定批处理文件类型的匹配顺序：文档定义将按照它们出现在该列表中的相同顺序（自上至下）来进行匹配。要更改文档定义顺序，请使用**向上**和**向下**按钮。
- **FlexiLayout。**通常，所加载的全部文档定义会连续进行匹配。然后，将检查匹配的质量，质量最佳的文档会选中。要加快文档定义匹配，请选择**使用第一个可接受弹性布局**。在这种情况下，如果有一个 FlexiLayout 匹配，那么就不再考虑其余的 FlexiLayout。
- **分类树。**可以为批处理文件类型或项目加载分类器以选择 FlexiLayout。要选择分类器，单击**加载...**，在打开的对话框中选择一个分类器文件。要断开该分类器，单击**断开**。
- **识别期间删除空白页** – 如果要删除空白页面，请选中此选项。

5.4. 完整性检查

如果要检查当前类型的批处理文件的完整性，请选中**导出之前运行批处理文件完整检查**。要编辑批处理文件完整性检查的脚本，单击**编辑检查脚本...**。请注意，执行此操作将导致 **workflow**选项卡（在分布式版本中）上的**导出整个批处理文件**选项被选中或被锁定，因而将无法更改。

如果未输入任何脚本文本，不论该选项的状态为何，均不会执行完整性检查。

5.5. 导出

对文档定义指定了导出参数。对于批处理文件类型，您可以手动指定默认导出参数。当前导出设置将显示在相应的部分中。要更改导出设置，单击**更改默认导出设置...**，并在打开的对话框中配置导出参数（请参阅“导出设置”）。

如果需要，可以在完成导出时自动触发脚本。要启用脚本触发，选择**导出后运行脚本**。要编辑脚本，单击**编辑...**（有关脚本编写语言的详细说明和示例，请参阅帮助文件）。

要导出不可识别文档的图像（没有匹配的文档定义），请选择**导出不可识别的文档**。要更改图像导出参数，单击**设置...**并指定新参数（请参阅“导出图像”）。

5.6. workflow

5.6.1. 单机

对单机版可配置的工作流参数如下：

- **批处理文件自动导出**（如果只有寥寥数个不确定识别的字符）。对于要自动导出的批处理文件，请选择**当出现错误或警告且不确定符号数量等于或小于以下项时自动导出批处理文件**，并为要自动执行的导出操作输入允许的不确定识别字符最大百分比。
- **导出后删除文档**。要在自动导出后删除文档，单击**导出之后删除已处理的文档**。

5.6.2. 分布式

对分布式版本可配置的工作流参数如下：

- **workflow模式**：无人照看或通过校验。在无人照看模式下，自动执行所有流程：导入文档、识别文档、导出文档，以及在发生错误时将文档添加至“异常”队列。在**通过校验**模式下，不确定识别的字符和错误将进行手动校验和纠正。也可以根据需要手动启动导出。
- **校验整个批处理文件**。使用来自同一个批处理文件的特定文档（最多 10 个）来组成校验任务。如果选择了此选项，校验任务将包含整个批处理文件。
- **当不确定字符百分比等于或小于以下项时跳过校验和数据校验** — 如果选择此选项，正确识别的文档不会包含在校验任务中。如果校验任务不包含任何要校验的字符，或者如果不确定字符的百分比小于指定值，将跳过这些文档。
- **导出整个批处理文件**。如果选择了此选项，则只导出整个批处理文件。不会导出单独的文档。
- **导出之后删除已处理的文档** — 选择此选项可自动删除导出文档。

- **独立队列**。您可以为要组成的操作员任务添加独立队列。

6. 配置图像导入

在分布式版本中，图像将导入至特殊应用程序 — 扫描工作站。但是，校验操作员和高级校验操作员也可以将图像添加至批处理文件。也可以从热文件夹自动导入图像。

导入配置文件允许指定导入参数和图像处理参数。在分布式版本中，导入配置文件一般用于处理热文件夹，其原因是校验操作员添加图像是为了纠正组装错误，而不是为了处理流化文档。在单机版本中，操作员也可以按相应按钮，采用导入配置文件来快速导入图像。

要添加导入配置文件，单击本程序主菜单上的项目 > **图像导入配置文件...**。在打开的对话框中，可以创建新的导入配置文件，及编辑、删除和复制现有的导入配置文件（图 16）。

要创建新的导入配置文件，请单击**新建...**来启动“导入配置文件向导”

首先选择图像来源。您可以选择从扫描仪或热文件夹（将在该文件夹搜索新图像）获取图像。如果要从热文件夹添加图像，确保您具有打开和写入此文件夹的权限。如果选择将扫描仪作为图像来源，确保扫描仪正确连接。

使用扫描仪创建导入配置文件

1. 选择**扫描仪**作为图像来源，单击**下一步**。
2. 配置扫描和图像处理选项。从下拉列表选择与扫描仪通信的方法：选择 **FlexiCapture 扫描设置**，以使用 ABBYY FlexiCapture 的对话框；或选择**系统扫描设置**，以使用扫描仪驱动程序对话框。单击**下一步**。
3. 配置文档组装和图像预处理设置。例如，您可以选择为图像去斑（如果图像质量较差或者图像可能含有垃圾），将图像转换为黑白图像，旋转图像，以及禁用歪斜消除。当完成时，单击**下一步**。
4. 现在，您可以更改默认导入配置文件名称及添加注释。

使用热文件夹创建导入配置文件

1. 选择将**热文件夹**作为图像来源，并输入热文件夹路径（可能情况下，请采用 UNC 格式）。然后，单击**下一步**。
2. 配置图像加载选项。您可以选择**热文件夹检查时间间隔**，并输入图像检查的间隔时间。在**批处理文件设置**中，从中选择一个选项，以定义要添加新图像的一个或多个批处理文件。在**要添加的文件数**下拉列表中，选择添加数目，以定义要添加至指定或新建批处理文件的文件数。单击**下一步**。
3. 采用与使用扫描仪创建导入配置文件的相同方法，来配置文档组装和图像预处理设置，完成时单击**下一步**。
4. “导入配置文件向导”将提示您配置热文件夹清除选项：可以删除或保留/移动成功导入的图像和发生导入错误的图像。完成时单击**下一步**。
5. 现在，您可以更改默认导入配置文件名称及添加注释。



图 16. “图像导入配置文件”对话框

热文件夹自动检查的启动方法如下：

- 在单机版中 – 选择**图像导入配置文件**窗口中的**自动检查热文件夹**。
- 在分布式版本中 – 对于单个项目，使用**热文件夹**部分中的**处理服务器监控**。

7. 上传项目至应用程序服务器

在单机版系统中，操作员可以在配置项目就开始处理项目。在分布式版本中，必须将所创建的项目上传到应用程序服务器。要执行此操作，单击**文件 > 将项目上传至服务器...**，并在打开的对话框中输入 `http://<servername>` 格式的应用程序服务器地址。只有应用程序服务器管理员才能上传项目到服务器上。

当项目已上传到服务器时，必须使用**管理和监控控制台**来为操作员分配角色。如果项目中使用了热文件夹，请使用**处理服务器监控**来进行启用。有关系统管理的详细信息，请参阅《**管理员指南**》。

8. 键盘快捷键

8.1. 主窗口

新建项目...	Ctrl+Shift+N
打开项目...	Ctrl+Shift+O
保存	Ctrl+S
新建批处理文件...	Ctrl+N

加载图像...	Ctrl+O	
扫描图像...	Ctrl+K	
导入图像	Ctrl+I	
导出...	Ctrl+U	仅限单机版
数据导出至文件...	Alt+Shift+S	
导出至数据库...	Alt+Shift+D	
撤消	Ctrl+Z	
恢复	Ctrl+Y	
剪切	Ctrl+X	
复制	Ctrl+C	
粘贴	Ctrl+V	
删除	Del	
全选	Ctrl+A	
查找...	Ctrl+F	
查找下一个	F3	
转至下一个文档	Ctrl+D	
转至上一个文档	Ctrl+Shift+D	
图像去斑	Ctrl+Alt+K	
反转图像	Ctrl+Alt+V	
顺时针旋转图像 90°	Ctrl+W	
逆时针旋转图像 90°	Ctrl+Shift+W	
测试批处理文件	Ctrl+Shift+B	
批处理文件	Ctrl+B	
编辑...	Alt+1	
缩略图	Alt+2	
全屏	F11	
刷新	F5	
属性...	Alt+Enter	
文档编辑器：图像模式	Ctrl+1	
文档编辑器：数据模式	Ctrl+2	
放大	Ctrl +Num+	

缩小	Ctrl+Num-
缩略图：增大	Ctrl+Shift+Num+
缩略图：缩小	Ctrl+Shift+Num-
分析	Ctrl+E
匹配文档定义...	Alt+Shift+E
识别	Ctrl+R
运行校验	F7
下一个要校验的项目	F4
上一个要校验的项目	Shift+F4
下一个组装错误	F9
上一个组装错误	Shift+F9
下一个不确定字符	F8
上一个不确定字符	Shift+F8
下一个规则错误	F6
上一个规则错误	Shift+F6
文档定义...	Ctrl+T
批处理文件类型...	Ctrl+Shift+T
图像导入配置文件...	Ctrl+Shift+I
更新至最新版本	Alt+Shift+U
重新分析	Ctrl+Alt+E
重新识别	Ctrl+Alt+R
重新核对规则	Ctrl+F6
检查批处理文件完整性	Ctrl+Shift+E
帮助主题	F1
退出	Alt+F4

8.2. 文档定义编辑器窗口

添加页面...	Ctrl+P
添加文档节段...	Ctrl+Alt+S
文档定义属性	Ctrl+T

保存	Ctrl+S
撤消	Ctrl+Z
恢复	Ctrl+Y
剪切	Ctrl+X
复制	Ctrl+C
粘贴	Ctrl+V
删除	Del
删除区域	Shift+Del
全选	Ctrl+A
按类型选择	Ctrl+Shift+A
组	Ctrl+G
取消组	Ctrl+Shift+G
从图像复制文本	Ctrl+Alt+C
创建字段：文本	Alt+Shift+T
创建字段：复选标记	Alt+Shift+C
创建字段：复选标记组	Alt+Shift+M
创建字段：条形码	Alt+Shift+B
创建字段：图片	Alt+Shift+P
创建字段：表格	Alt+Shift+L
创建字段：组	Alt+Shift+G
图像去斑	Ctrl+Alt+K
反转图像	Ctrl+Alt+V
顺时针旋转图像 90°	Ctrl+W
逆时针旋转图像 90°	Ctrl+Shift+W
数据表单	Alt+F2
属性	Alt+Enter
静态元素	Ctrl+1
字段区域	Ctrl+2
显示文档结构	Alt+F1
字段	Alt+1
文档结构	Alt+2

放大	Ctrl+Num+
缩小	Ctrl+Num-
运行测试	Ctrl+R
自动检测对象	Ctrl+D
检查文档布局	F9
匹配节段	Ctrl+E
帮助主题	F1
关闭	Alt+F4

8.3. 组校验窗口

全部确认	Enter
全部推迟	Ctrl+Enter
切换	Space
下一页	Page Down
上一页	Page Up
撤消	Ctrl+Z
恢复	Ctrl+Y
全选	Ctrl+A
显示字符图像	F2
全屏	F11
显示字段图像	Ctrl+I
显示字段图像：在顶部	Alt+1
显示字段图像：在底部	Alt+2
图像比例：放大	Ctrl+Num+
图像比例：缩小	Ctrl+Num-
帮助主题	F1
退出	Alt+F4

8.4. 字段校验窗口

确认字段	Enter
推迟字段	Ctrl+Enter
下一个不确定字符	F4
上一个不确定字符	Shift+F4
下一字段	Page Down
上一字段	Page Up
撤消	Ctrl+Z
恢复	Ctrl+Y
剪切	Ctrl+X
复制	Ctrl+C
粘贴	Ctrl+V
删除	Del
全部删除	Alt+Del
全选	Ctrl+A
插入分行符	Shift+Enter
合并字符	Ctrl+M
类似字段	Alt+F3
显示字符图像	F2
全屏	F11
识别的文本	Alt+F1
字符图像切割	Alt+F2
显示字段图像	Ctrl+I
显示字段图像：在顶部	Alt+1
显示字段图像：在底部	Alt+2
图像比例：放大	Ctrl+Num+
图像比例：缩小	Ctrl+Num-
帮助主题	F1
退出	Alt+F4