

原版 PDF 格式电子书扫描制作指南 (第一稿)

1.引言

书海无边。

把想看的书都收入囊中，那是多少人的梦想。

我就是个怀揣着这样梦的人，不过我的钱包告诉我：这是不可能的！

科技创造了伟大的电子书，无私的扫友们通过自己的辛勤劳动，给我们制作了无数的电子书。电子书给我们的工作和学习带来了很大的方便。最初我也只是简单的收集，后来觉得自己总是索取而没有奉献，颇为过意不去，于是我就开始尝试自己动手制作电子书，搜集制作电子图书也成了我的一大爱好。很多网友也有扫书共享的爱好，不过由于每个人扫描制作的技术水平不同，不少网友费了很多时间和气力做出来的电子书却模糊不清。

经过不断的失败和探索，在原版 PDF 格式电子书制作上我也有了自己的一点小小的经验。这里把我在扫描制作电子书方面的一点经验写一下，希望对有各位扫友能有所帮助。

2.本文的讨论的范围

常见的电子书格式有：PDF、EXE、TXT、HTML、HLP、CHM、LIT、WDL、CEB、ABM、PDG、UMD、JAR 等，WORD 文档从一定意义上来讲也可以说是一种电子书。不同的电子书格式各有自己的优点和缺点，制作方法也不尽相同。目前，我个人还是最喜欢 PDF 格式的电子图书。

常见的 PDF 格式电子图书有两种：一种是由纸质出版物经过扫描制作而成的，还有一种是经过了文字识别或人工录入编辑而成的。前者图书内的文字编辑起来比较麻烦，但是最大程度的保留了原书的本色；后者更加清晰，而且便于文字编辑和复制引用，但是在二次输入或文字识别过程中出现错误几乎是不可避免的。

对专业书籍来说，精准应该是第一位的。因此，我更推崇原版扫描 PDF 格式电子图书。本文仅讨论原版扫描 PDF 格式电子图书的制作技巧。

3.硬件、软件及基础知识准备

3.1. 硬件

扫描设备一台，各种型号都行，不同型号扫描仪效率和效果会有差别。

普通平板扫描仪精度高，但是扫描速度太慢，它蜗牛一样的速度会使扫书变成一件极为痛苦的事情。如果有一台带扫描功能的快速复印机或一体机会给你的扫书工作带来不小的帮助。

如果图书可分解为散页，一台带有馈纸功能的扫描设备会极大的提高你的工作效率。

数码相机也可以，不过边角会有畸变。

我现在用的是单位的柯尼卡美能达 bizhub 421 黑白数码复合机。

3.2. 软件

编辑 PDF 格式电子图书，需要用到各种软件，推荐使用以下 6 款软件。后面的所有具体操作均以电脑已安装对应软件为前提。

1、ACDSee

我用的版本：ACDSee 2009

具有一定批处理功能，主要用来对原始图片进行剪裁、旋转和调整图像曝光度处理。

2、IrfanViewPortable

我用的版本：IrfanViewPortable 4.23

具有强大的批处理功能，主要用来对经过初步处理的图片进行批量调整色深，批量调整大小，批量重命名处理。

3、Adobe Acrobat

我用的版本：Adobe Acrobat 9.2

Adobe 官方 PDF 文档编辑工具，用来制作、编辑 PDF 文档。

这里需要说明一点，Adobe Acrobat 9.0 以后版本相对于之前版本文件存贮格式有了很大的改进，在相同清晰度条件下，9.0 版本文档的文件体积相对于 7.0 版本可缩小 1/3 至 1/2。强烈推荐使用老版本 Adobe Acrobat 的朋友进行升级。

4、FreePic2Pdf

这是一款国内高人制作的免费小软件，版本不太清楚，功能比较多，不过我用的最多的是它的书签摘挂功能，非常方便。

5、PDF Password Remover 和 Advanced PDF Password Recovery

目前我用的版本分别是：PDF Password Remover2.5 和 Advanced PDF Password Recovery 5.0。这两款软件都是用来清除 PDF 文档的密码的。

前者操作简单，可以对付大多数加密的 PDF 文档，后者用起来稍麻烦一些，可以用来治疗各种疑难杂症。看了这么多解密利器你就会发现普通的加密根本就是徒劳的，而且毕竟加密有违共享精神，我拿来共享的 PDF 文档都是开放全部权限不做任何加密设置的。

3.3. 关键概念-色深

色深，也称之为色位深度，在某一分辨率下，每一个像素点可以有多少种色彩来描述，它的单位是“bit” (位)。典型的色深是 8-bit、16-bit、24-bit 和 32-bit。深度数值越高，可以获得更多的色彩，相同分辨率下文件的体积也会成倍增加。

色深是决定最终 PDF 格式图书清晰度和文件大小的核心。最终完成的 PDF 文档中彩色图片一般可以采用 24b 或 32b。灰度图片一般采用 8b，文字以纯黑白即 1b 色深为佳。通过扫描或 PDF 文档拆解得到的原始图片素材一般采用 8b 灰度图片。

4.原版PDF图书制作的基本步骤

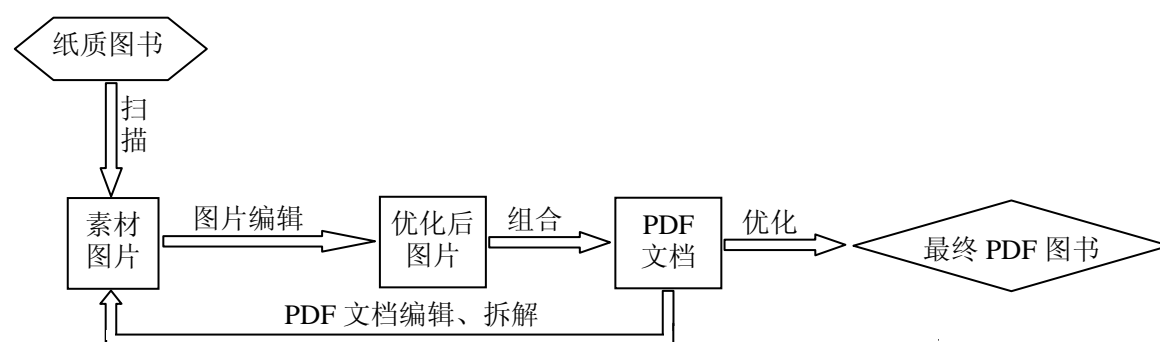


图 1 原版 PDF 图书制作流程图

通过扫描或者 PDF 文档拆解得到的图片是我们制作 PDF 图书的基本素材，这里我们称之为素材图片。

由于扫描参数设置或操作的原因，素材图片往往会存在曝光不足或过度、歪斜、黑边、噪点等问题，这就需要用图片处理软件对图片进行编辑。经过优化的图片可用 PDF 制作软件组合为 PDF 文档。

有些中间处理工序，如消除黑边，版心调正，双页拆分等在 PDF 编辑软件中操作更为方便。为使最终成果简洁美观，经过这些中间工序编辑的 PDF 文档一般需要重新

拆解为图片，以进入下一处理环节。有时甚至需要如此循环往复几个回合。

原版 PDF 图书制作的基本操作流程见图 1 原版 PDF 图书制作流程图。

4.1. 素材图片的获取

原版扫描 PDF 格式电子图书的制作基础是纸质图书的电子版影像，本质上就是一张张图片。扫描是这些图片的最常见的获取方式。另外，有时我们需要对别的朋友制作的 PDF 图书进行瘦身和清晰化处理，仅对 PDF 文件本身进行处理，往往难以达到预期的效果，这就需把已经存在的 PDF 图书拆解为图片，为进一步的处理作准备。

4.1.1. 扫描参数设置

不同型号的扫描仪有不一样的操作方法，说明书和网络上都有非常详细的介绍，这里就不再赘述。

下面主要说一下几个关键参数的设置：

- 1、分辨率，推荐 300dpi。太精细文件过大，太低不够清晰；
- 2、色深，彩页就用彩色（24b），非彩页建议用灰度（8b）或者纯黑白（1b）。

灰度或者黑白文字显示更加清晰，而且扫描得到的图片会小很多，除了彩色图片一般设为 8b 灰度即可。当然也可以在扫描时采用较高色深，在后期处理中通过软件进行色深调整。

亮度、对比度、曝光度、色阶等参数的设置以尽量使得到的扫描图片逼真清晰为准则，如果对扫描设备不是很熟悉，使用默认也可以。强大的图片处理软件足以修正在这些细节上的小遗憾。

4.1.2. PDF图书拆解

所谓的 PDF 图书拆解，就是把 PDF 图书另存为图片，Adobe Acrobat 本身就有这个功能。可按以下步骤操作：

- (1) 用 Adobe Acrobat 软件打开需要进行拆解处理的 PDF 文档。
- (2) 选择“文件”|“另存为”，显示“另存为”对话框。
- (3) 单击“保存类型”，选择“PNG(*.png)”。
- (4) 单击“设置”按钮，弹出“另存为 PNG 设置”对话框，“文件设置”、“色彩管理”两项均用默认参数即可；“转换”|“色彩空间”，选择“灰度”；“转换”|“分辨率”，推荐选择“118.11 像素 / 厘米”。
- (5) “确定”，返回“另存为”对话框，指定文件名，设置保存路径。点击“保存”。

注：这里选择的“118.11 像素 / 厘米”相当于“300 像素 / 英寸”，即“300dpi”

需要说明一点，这里的文件“保存类型”选用 PNG 格式，是个人经验，选用其他格式也可。虽然常用 JPG 格式存储得到的图片本身相对较小，但是 JPG 格式的图片相对于 PNG 格式的图片在组成 PDF 的过程耗时和最终 PDF 文件大小上反而略为逊色。

4.2. 图片编辑

4.2.1. 用ACDSee批量调整图片曝光度

曝光度问题是素材图片最常见的问题，直接影响着最终成果的清晰度。不同设备和参数设置扫描得到的图片曝光度差异很大，在批量调整前可以先选取一张代表性图片试调，用试调得到的调整参数进行批量调整。

用 ACDSee 试调单张图片曝光度可按以下步骤操作：

- (1) 进入图片所在文件夹，右键单击要调整的图片，在弹出的菜单中选择“使用 ACDSee 相片管理器 2009 进行编辑”，进入图片编辑界面。
- (2) 单击“曝光”按钮，弹出“曝光”、“对比度”、“填充光线”三个参数，拉动滑块或直接输入参数进行调整。一般曝光不足的图片常用参数设置范围为“曝光度”：10~50，“对比度”：5~35，“填充光线”：-50~-85。具体参数可根据右侧的预览图像自行决定。
- (3) 试得最佳参数之后记录下来备用，然后放弃编辑，不保存，以保持所有图像一直便于后面统一调整。

用 ACDSee 批量调整图片曝光度可按一下步骤操作：

- (1) “开始” | “程序” | “ACD Systems” | “ACDSee 相片管理器 2009”，或通过其他途径启动“ACDSee 相片管理器”。
- (2) 在“ACDSee 相片管理器”选择素材图片所在文件夹。按下 Ctrl+A 选中文件夹内全部图片。
- (3) 单击“工具” | “调整图像曝光度”，弹出“批量调整曝光度”对话框。
- (4) 选择“曝光”选项，根据前面试调得到的最优参数设置“曝光”、“对比度”、“填充光线”三个参数。
- (5) 单击对话框中部的“选项”按钮，在弹出的“选项”对话框设置文件保存方式。单击“确定”按钮退出“选项”对话框。这里可根据个人习惯自行设定保存方式，一般建议在得到满意最终成果前保留中间成果，万一出了问题可以从出错

(6) 单击“过滤所有图像”按钮，批量调整选中图片曝光度。

有些图片经过手工批量调整之后可能效果仍不是很好，可以采用批量自动调整色阶和对比度的方法对经过手工调整的图片做进一步的处理，具体操作步骤与上述操作类似，只是“曝光”选项变为“自动色阶”选项。“自动色阶”功能还是非常强大的，强烈推荐使用。

4.2.2. 用ACDSee手工校正歪斜图片

Adobe Acrobat 对 PDF 文档在进行文字识别的时候可以对轻度歪斜的页面进行自动纠正，但是歪斜角度稍大 Adobe Acrobat 就无能为力了。目前我还没有找到可以自动纠偏的软件，因此我只能对歪斜的图片进行逐页手工纠偏。

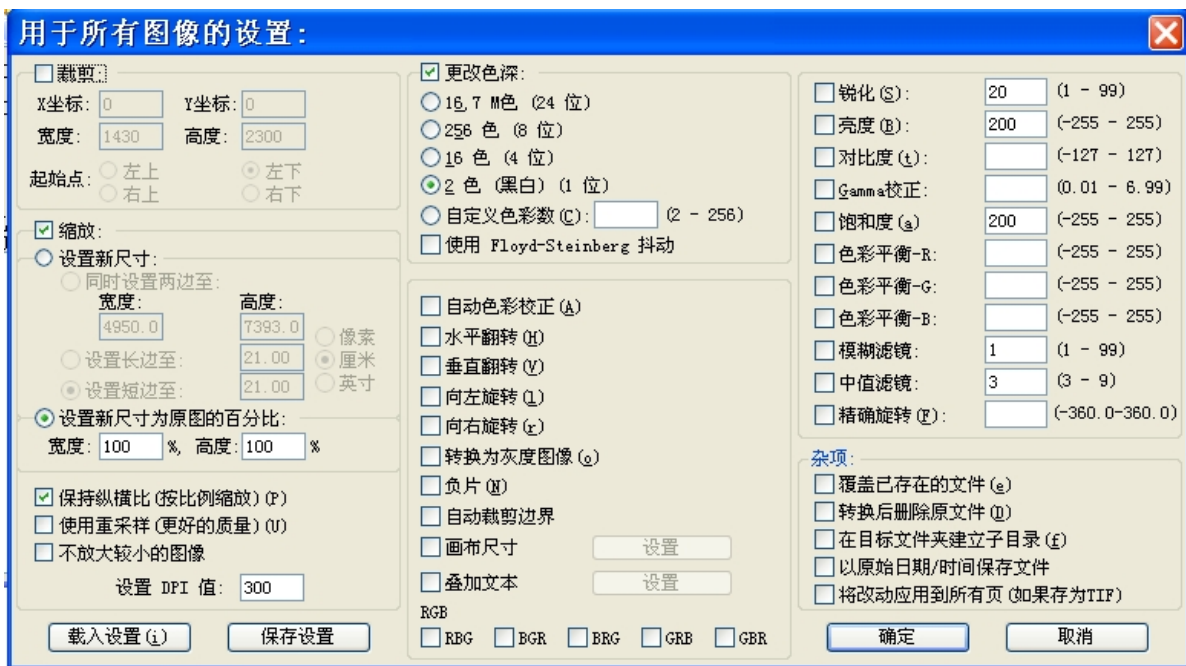
一般的图片编辑软件都可以进行该处理，用 ACDSee 操作起来也非常简单，具体操作步骤不再赘述。

4.2.3. 用IrfanView批量调整图片大小和色深

IrfanView 软件虽然程序体积不大，但是具有强大的批处理功能。这里主要用其调整图片大小和色深。

具体操作步骤如下：

- (1) 打开 IrfanView 软件，选择“文件”|“批量转换/重命名”，弹出“批量转换”对话框。
- (2) “操作”单选钮选择“批量转换-重命名”，“输出格式”下拉菜单选择“PNG”，勾选“使用高级选项”复选框。
- (3) 单击“高级”按钮，弹出“用于所有图像的设置”对话框，这里主要使用“缩放”和“更改色深”两项功能，其余功能有兴趣的朋友可自行摸索，不再赘述，常用设置如下：



(4) 通过 IrfanView 找到经过 ACDSee 编辑的素材图片存放文件夹, 添加需要进行编辑的图片, 设置好输出目录和命名规则。单击“开始”按钮进行批量编辑。

很多朋友制作得到的 PDF 文档会出现页面大小相差悬殊的问题。这往往是由于用 ACDSee 等图片编辑软件对某些页的图片进行了编辑, 后面合并起来组成 PDF 文档的时候这些页面就会变的尺寸异常。这里用 IrfanView 进行一遍统一的“缩放”处理就可以很好的解决这个问题。

如果进行缩放处理的图片规格一致, 也可以用输入边长的方法进行大小控制。由于前面个别页面进行了旋转操作, 图片页面变的不规则, 故采用百分比控制。

4.3. 优化后图片组合为PDF文档

经过前面的一系列处理, 图片变的不仅清晰而且体积也大大缩小, 下面就该 Adobe Acrobat 上场了。

具体操作步骤如下:

- (1) 打开存放图片的文件夹, Ctrl+A 全选, 单击右键, 在弹出的菜单中选择“在 Acrobat 中合并支持的文件”, 弹出“合并文件”对话框。
- (2) 在“合并文件”对话框右上角, 单选“单个 PDF”选项, 配合 Ctrl 或 Shift 键选择要合并的文件, 把它们排列至预期的顺序。
- (3) 单击“合并文件”按钮。

图片较多时, Adobe Acrobat 的处理会耗时较久, 耐性等待至弹出对话框提示保存

合并得到的 PDF 文档，给它取个名字，找个窝，PDF 文档就算是初步完成了。

4.4. PDF文档的优化

PDF 文档的优化操作在 Adobe Acrobat 的帮助文档和网络上都有很多介绍，这里就不做全面叙述，只挑几条立竿见影的。

4.4.1. OCR文本识别

很多朋友瞧不起原版 PDF 文档的一个重要原因就是里面的文字不能直接复制，其实原版 PDF 文档里面的文字也可以进行复制。

打开要处理的 PDF 文件，“文档”|“OCR 文本识别”|“使用 OCR 识别文本”，弹出“识别文本”对话框，单击“编辑”按钮，在弹出“识别文本—设置”对话框“设置识别语言”和“PDF 输出样式”，建议“PDF 输出样式”设置为“可搜索的图像（精确）”。然后一系列“确定”，开始识别。

Adobe Acrobat 的“OCR 文本识别”模块运行速度比较慢，而且识别精度也不是很高，但是识别具有一定的图像纠偏功能，能自动对轻度歪斜的图片进行纠正。如果电脑不是特别慢，一般都建议进行该处理。

4.4.2. 减小文件大小

本文前面说过在相同清晰度条件下，9.0 版本文档的文件体积相对于 7.0 版本可缩小 1/3 至 1/2，这个优势要经过“减小文件大小”这一道工序才能真正的体现出来，很多旧版的 PDF 文档即使不经任何其他处理，仅经过这一道工序就可以把体积缩小 1/3 至 1/2。

具体操作步骤很简单：打开 PDF 文档，选择“文档”|“减小文件大小”，弹出“减小文件大小”对话框，选择兼容于“Acrobat 9.0 和更高版本”，然后“确定”，命名，保存。

经过这步处理，你一定会有惊喜的发现。需要注意的是打开经过这道工序处理的 PDF 文档一定要用对应版本的 PDF 阅读工具打开，否则可能出现错误。

4.4.3. 优化扫描的PDF

选择“文档”|“优化扫描的 PDF”，可以弹出“优化扫描的 PDF”对话框。

Adobe Acrobat 的这一模块在我最初接触 PDF 文档的时候经常用，该模块可以进行的处理内容也非常多，不过处理效果并不是特别理想。对经过前面一系列图片处理流程制作而成的 PDF 图书来说，基本上就不需要用不到这个模块了。有兴趣的朋友可以试一下。

5.几个特殊问题的处理

前面讲了制作一本 PDF 图书的基本流程，但是有一些细节问题仍没有解决，如：制作的 PDF 文档页面大小不一，翻页扫描的图书书脊位置存在黑边，双页扫描的图书如何拆分为单页，怎样使图书的页码和图书的目录对应，如何添加或批量转移书签，网络上下载的图书上面有各种烦人的广告。下面介绍这些特殊问题的处理方法。

5.1. 统一页面大小

由于前面对图片进行了旋转、剪裁等处理，组合得到的 PDF 图书往往页面尺寸大小不一。统一页面大小，不仅看起来舒服，而且也是后面用 Adobe Acrobat 进行黑边消除和双页拆分的基础。

具体操作步骤如下：

- (1) 打开需要处理的 PDF 文档，选择“文档”|“裁剪页面”，弹出“裁剪页面”对话框。
- (2) 选“更改页面”大小选项为“自定义”，设置需要设置的页面尺寸，“页面范围”选项根据实际需要设定。然后确定即可。

注意：这里自定义的页面尺寸一定要大于所有现有页面的尺寸，否则软件无法完成本处理。为了照顾最大页面，自定义的页面往往偏大，在“裁剪页面边距”选项中对页面四边进行裁剪即可得到最佳尺寸。

5.2. 消除页面黑边

消除页面黑边有多种途径可以解决，以前我用 ACDSee 对图片逐页进行剪裁，但是操作过于繁琐，而且效果不理想。利用 Adobe Acrobat 进行消除页面黑边处理操作更加方便，而且可以顺便把页面版心对中。

消除页面黑边操作与统一页面大小一样，都是利用“页面剪裁”功能。二者基本操作相同，只是在设置“裁剪页面边距”时仅保留希望保留的图书的内容部分，黑边自然就被切掉了。

由于扫描时操作的误差，各页面版心位置高低左右不一，而且往往没有规律。进行统一的剪裁时难免会伤及“内脏”。好在有“TouchUP 对象工具”这一利器可以解决这个问题，只是需要逐页进行繁复的手工操作。选择“工具”|“高级编辑”|“TouchUP 对象工具”即可对图片格式的页面内容进行剪切、复制、粘贴、移动、缩放等编辑。缩

放时把鼠标移至图片的任意一个角点，出现一个双箭头符号，即可对页面图片内容进行拉伸缩放。

有时个别黑边在页面内的延伸较深，统一的周边剪裁无法触及，有时需要消除页面中间的公章或其他个人信息，除了用图片编辑软件对图片进行编辑之外，在 Adobe Acrobat 也可以通过打补丁进行该处理，而且操作相对于图片编辑更加方便。

打补丁的操作方法为：选择“工具”|“注释和标记”，添加一个图形把需要消除的黑影或公章及个人信息盖住，然后选中输入的图形，单击右键，在属性中把图形的线条颜色和填充颜色均设置为白色即可。这里还有一个小技巧，设好一个图形的属性之后，选中该图形，单击右键，选择“设置当前属性为默认值”，后面再输入的图形都默认成白色的补丁了。

为使通过剪裁丢弃的页面黑边彻底从文档内消失，使打上的补丁融入图书中而不被揭穿，需要对经过消除页面黑边处理的 PDF 文档进行拆解，还原为图片，然后通过前面的一些列工序重新组合。

5.3. 双页拆分为单页

双页总是不如单页看起来方便。好多扫描仪可以自动在书本扫描过程中对书本进行拆分，具体设置可以参照对应设备的说明书。如果两页扫在一起了，就需要用软件进行双页拆分。双页拆分为单页可以利用 Adobe Acrobat 的“页面剪裁”，配合 IrfanView 的“批量重命名”实现。

具体操作步骤如下：

- (1) 把需要进行双页拆分为单页处理的 PDF 文档复制为 A、B 两份，A 文档通过剪裁保留左边，B 文档通过剪裁保留右边。
- (2) 把剪裁后的 A、B 文档分别拆解为图片，分别保存至两个独立的文件夹 A、B 内。
- (3) 打开 IrfanView 软件，选择“文件”|“批量转换/重命名”，弹出“批量转换”对话框。
- (4) “操作”单选钮选择“批量重命名”，在“批量重命名”设置选项内定义命名规则。A 文件夹的图片，起始编号设为 1，增量为 2；B 文件夹的图片，起始编号设为 2，增量为 2。
- (5) 把经过重命名的图片统一放在同一个文件夹内，这样页码顺序就顺下来了。然

5.4. 编排页码

这里所说的页码不是指页脚中的页码，而是指“页面导览”窗口中显示的 PDF 文档电子页码。PDF 文档默认的页码编排方式一般是从头至尾顺序而来，由于前面多了封面和目录等内容，“页面导览”窗口中的电子页码就和页脚中的纸质页码不一致，不能直接通过“页面导览”窗口进行页面定位。通过编排页码，使 PDF 文档的电子页码和页脚中的纸质页码相匹配，不仅阅读起来方便，而且可以通过这一操作检查制作的 PDF 文档是否有缺页、重页问题。

具体操作步骤如下：

- (1) 用 Adobe Acrobat 打开要进行编辑的 PDF 文档，选择“高级”|“文档处理”|“编排页码”，弹出编排页码对话框。
- (2) 在“页面”选项内，设置需要进行编排页码的页面范围，在“编码”选项内选“开始新节”，通过添加不同的前缀，或者采用不同的样式，将封面、前言、目录、正文等内容按个人习惯分别设置为不同的编码体系。
- (3) 对 PDF 文档进行抽检，检查页脚中的纸质页码与重新编排的电子页码是否对应，从而查得是否存在缺页、重页问题，并可迅速对问题部位精确定位。
- (4) 对缺页、重页问题，可以通过“文档”|“插入页面”、“提取页面”、“替换页面”、“删除页面”等操作进行处理。

另外，打开 Adobe Acrobat 左侧的“导览面板”最上面的“页面”选项板，点击双齿轮图标“选项”按钮，也可以进行编排页码操作，并可配合 Ctrl 和 Shift 键对需要编排的页面进行选择，操作更加直观、方便。

5.5. 编辑书签

添加了书签的图书阅读起来更加方便，在阅读过程中顺手加上书签也有利于对全书框架的把握。

常规的书签编辑只需打开 Adobe Acrobat 左侧的“导览面板”的“书签”选项板，可对书签进行编辑操作，在需要添加书签的页面单击右键也可以添加书签。对已经添加了书签的图书经过拆解重组等处理工序之后原来的书签就不见了，这时可以用 FreePic2Pdf 软件对书签进行摘挂操作，把书签从原来的文档上摘下来，挂到处理之后的新文档上。

具体操作步骤:

打开 FreePic2Pdf 软件, 单击“更改 PDF”按钮, 先摘后挂, 根据提示操作就可以了。

5.6. 去掉各种水印和广告图块

从网络上搜集来的 PDF 文档往往会被添加了各种水印或者广告图块, 并且很多都设置了密码, 限制了编辑、打印权限。

首先就是解密, 用前面提到的 PDF Password Remover 和 Advanced PDF Password Recovery 软件目前基本没有解决不了的。

对付一般的水印或背景及页眉页脚, 用 Adobe Acrobat 打开需要编辑的 PDF 文档, 应用“文档”|“页眉和页脚”、“背景”、“水印”等操作即可删除。

图片格式的图块, 目前我只会用 Adobe Acrobat 的“工具”|“高级编辑”|“TouchUP 对象工具”逐个手工删除。

另外还有一些和图书内容嵌合在同一图层上的颜色较浅的水印, 不能通过前面的消除水印操作删除, 那是因为这些水印已经变成图片上的内容了, 不再具有独立的水印属性。这时可以将图书拆解为图片, 通过图片的调整曝光度、对比度、色深等操作把这些浅色的水印给消掉。

还有一些位于页面顶部、底部或侧边的广告条可以采用前面消除黑边的办法裁掉。

6. 结语

本文主要是个人经验的总结, 里面有很多地方语言表述不是很清楚, 措辞也不太恰当, 有些操作甚至可能走了弯路, 把简单的问题搞复杂了。欢迎各位朋友批评指正, 后面附有我的联系方式。希望本文能为您提供一些新思路, 做出更完美的 PDF 文档。

最后, 向为了知识的传承而辛勤劳动的扫友们致以崇高的敬意!

njzy0532

2010 年 1 月 3 日

本文作者联系方式:

筑龙网 ID: njzy0532

QQ: 68607214

邮箱: 68607214@qq.com