

从Office2003的图片中抠出文字 PDF转换可能丢失图片或格式
， 建议阅读原文

https://www.100test.com/kao_ti2020/454/2021_2022__E4_BB_8EOffice2_c98_454353.htm Office在2003版中增加了Document Imaging工具，用它可以查看、管理、读取和识别图像文档和传真文本。其实，利用它的这个功能，我们还可以把网页或电子书中的文字给“抠”出来。打开电子书，尽量采用较大的字体，翻到想要获取的页面，用抓图软件SnagIt对相关的内容进行抓取，然后在“文件”菜单中选择“复制到剪贴板”命令(也可以用其他抓图软件，当然最简单的是Windows中自带的Print Screen键来抓取整个屏幕，然后在“画图”程序中对不要的部分进行裁剪并保存，然后复制)。在“开始”菜单的“microsoft Office工具”中打开Microsoft Office Document Imaging，在左侧窗口中单击鼠标右键，选择“粘贴页面”，把复制的图片粘贴到Document Imaging中，在“工具”中选择“使用OCR识别文本”，Document Imaging的OCR识别程序就会对图片进行识别，完成后选择“工具”中的“将文本发送到Word”，程序会自动打开Word文档，展现在你面前的就是从图片中“抠”出来的文字。提示：一般而言，识别的准确率可以达到95%以上，但对英文和数字的识别不是太好。100Test 下载频道开通，各类考试题目直接下载。详细请访问 www.100test.com