

得到了清晰的文字纸稿图片后，我们就需要使用 office 软件中的一个小工具“Microsoft Office Document Imaging”来进行文字识别了，一般安装好 office 后会默认安装在系统中的。

打开 Word，点击菜单“插入 → 图片 → 来自文件”，把纸稿图片插入到文档中，再点击菜单“文件 → 打印”，在弹出窗口的“打印机”下拉列表框中选择“Microsoft Office Document Imaging Writer”，点击“确定”按钮会弹出一个保存窗口，我们设置一下文件保存的路径和文件名，Word 会把文档保存为 MDI 格式的（如果使用的是 office2007，则 Word 会把它保存为 TIFF 图片格式）。如图 2：

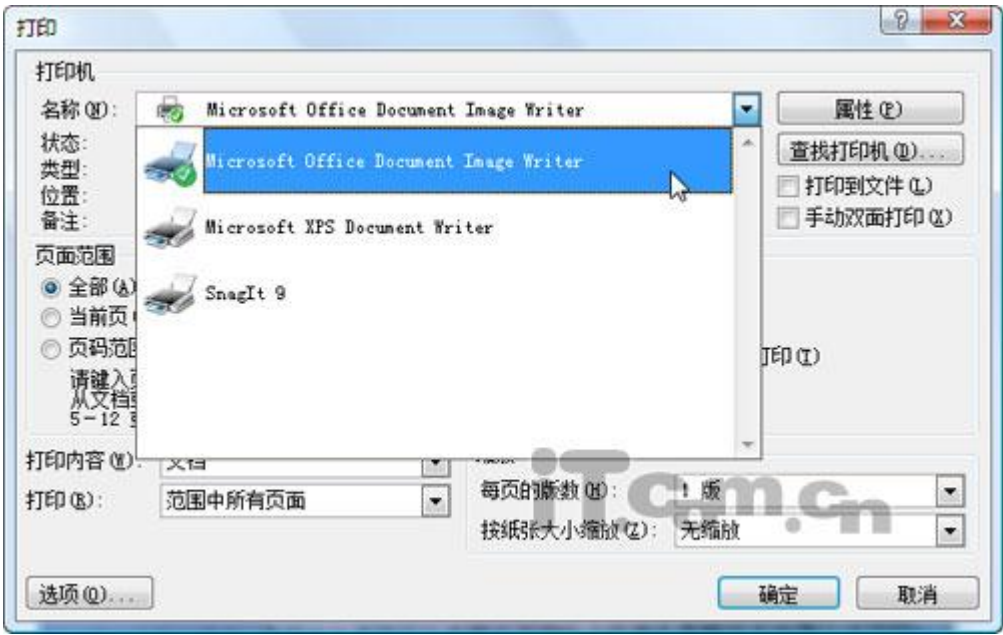


图 2

接下来双击打开前面保存的 MDI 文档，可以看见图片显示在左侧栏中，在右侧栏用鼠标把要进行识别的文字用红色线框起来，选好后点击菜单“工具 - 使用 OCR 识别文本”，这时软件就会对所选区域进行文字识别了（如图 3）。

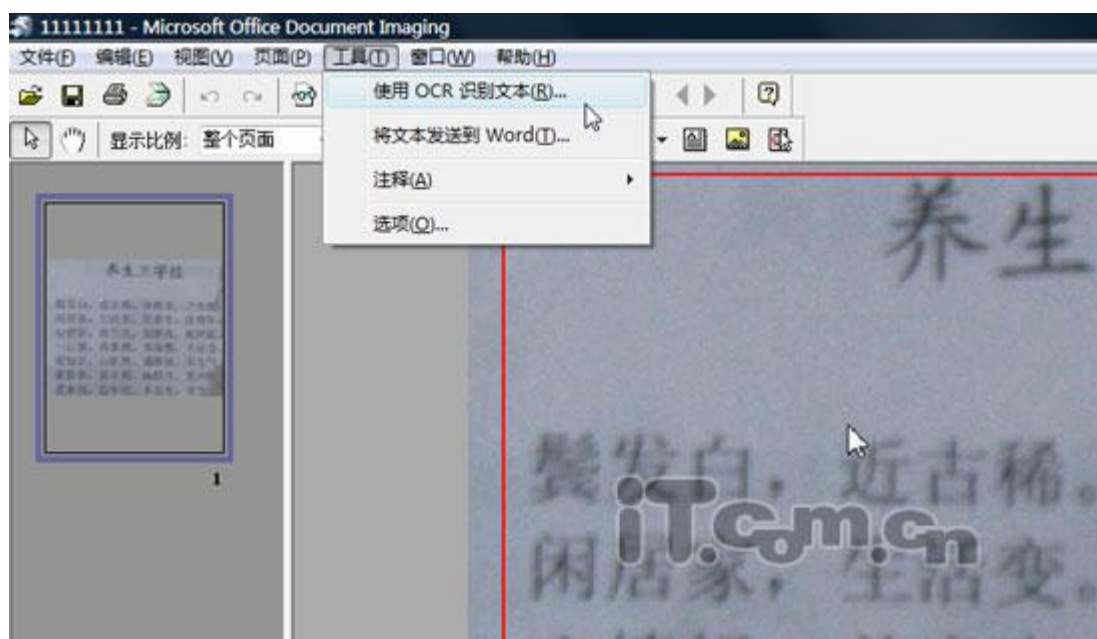


图 3

小提示：如果使用的是 office2007，则需要从程序菜单中点击“Microsoft Office → Microsoft Office 工具 → Microsoft Office Document Scanning”，在打开的“Document Scanning”软件中点击菜单“文件 → 打开”来打开前面保存的 TIFF 文件。

识别好后图片中的文字就可以进行选择了，把需要的内容全部选中，再点击菜单“工具”，将文本发送到 Word”，这时会弹出一个对话框，我们选择里面的“当前选择范围”，并把下面的“在输出时保持图片版式不变”选中，完成后点击“确定”按钮，这样就可以在新建的 Word 文档中看见识别出来的文字（如图 4）。



图 4

小提示：网上的一些精简版的 office 软件是没有 “Microsoft Office Document Scanning” 这个小工具的，需要安装完整版的 office 才有。