从 PDF 练习文档的内容来看，提取的文本并没有什么问题，主要是因为这个文档的内容比较简单。

如果文档结构比较复杂，比如说存在各种公式图表，使用了多国语言符号等情况，又或者使用了 pypdf2 不支持的 PDF 生成器，那么提取出来的内容就不一定正确了。这也没办法，如果遇到复杂的文档，就算使用专业的 PDF 编辑软件也不一定能正确提取。

还要注意的是，如果文档复杂，则提取出来的文本的顺序也不一定是正确的。