

根据批量URL生成有书签的PDF文档的解决方案计算机等级考试 PDF转换可能丢失图片或格式，建议阅读原文

https://www.100test.com/kao_ti2020/645/2021_2022__E6_A0_B9_E6_8D_AE_E6_89_B9_E9_c97_645285.htm

一、缘起 看到一篇好文章、一篇精彩的博文，我们最早期的做法有两种，添加到收藏夹，或者另存为，后来出现了新的方式，转贴到自己的博客里、或者收藏到一些收藏网站上（比如360doc）。近来又出现了生成PDF文档的方式，比如某些网站，你提交一个url，就为你生成对应的pdf，javaeye的电子书制作也很不错，甚至可以预测一下，浏览器的“另存为”窗口里有可能会有一个*.pdf选项。因为pdf实在是太好了，小巧而有表现力丰富。尽管如此，如果有多篇好的文章呢，比如一个非常优秀的连载（比如牛人写的教程、开发心得），我们能做的也就是添加到收藏夹或添加到自己博客的连接里头。想当年，没钱买书的时候，就在网上找一些教程来看，做得比较好的会有一个页面列出该连载的所有连接，这种情况我通常用迅雷以下载全部链接的方式（再去掉不相关的链接）批量下载下来，做的简单的只提供一个类似于上一篇下一篇的连接，这种情形到现在还是这样。（易用性啊，这个问题可大可小，往大的说，关系到应用模式和商业模式，往小的说也许就是个美观和方便。Javaei提供的快捷阅读采取左边树形菜单展示目录，右边展示内容的形式，这种形式在界面的设计上屡试不爽。）本文就是要阐述一种根据批量URL生成有书签的PDF文档的解决方案，这个批量URL就是一些好的文章的URL，根据这些URL生成一个合并的PDF文档，并且要有书签（也就是左边的树形菜单），而且必须要有书签。《java与

模式》这本书想必很多人都看过，老厚的一本书啊，我没钱买，看的是一个下载的pdf，这个pdf给我的印象太坏了，没有书签，要找个东西只能拉滚动条，尽管如此，我还是看完了，书写的不错，制作pdf的那个人功过相抵。

二、思路

目的是要根据批量URL生成有书签的PDF文档，实现这个目的分两步走：先解决根据一个URL生成一个PDF文档，再解决多个PDF合并并生成书签的问题。

（1）根据一个URL生成一个PDF文档

根据一个URL生成一个PDF文档看似很容易，因为我们有iText,pdfbox这些开源框架，实则不简单，因为要保证生成的pdf文档打开后的效果要和浏览器里的要一样，这无异于做一个浏览器，目前的浏览器尚且存在兼容的问题，所以自己写一个根据html生成pdf的想法很难实现。那么另一个思路就是利用一些网站来实现这个目的，经过尝试，有些网站是需要提供URL和email，人家把生成好的pdf发到你邮箱，这种形式无法通过代码来访问，也就无法批量处理；有些网站只需提交url，就把生成的pdf响应给客户端，这种形式可以通过程序来批量处理，但是生成的pdf与浏览器里的效果相差太远；还有一些网站根本就不支持中文。通过探索，终于找到一个网站提供的C#做的dll可以达到这个要求，利用这个dll，写个简单的C#程序就可以批量的生成pdf了，而且效果相当完美，美中不足的是生成的pdf里有人家的水印。

（2）多个PDF合并并生成书签

多个PDF合并并生成书签可以轻松的用iText搞定，合并是有一定的顺序的，并且书签是树形结构的，因此合并的顺序，书签的层次结构是需要事先确定好的。所以对批量的URL要进行一定的描述，那么很自然的选择xml。

三、实现

我现在越来越觉得，只要不是基础设施的

东西，技术上都很简单，关键是你有没有想法。这个实现先从xml描述开始。Xml描述分两步，先简单的描述一批URL（这里叫做href.h2p.xml），再来描述层次关系（这里叫做outline.h2p.xml）。h2p就是html to pdf的意思 先看href.h2p.xml 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