

pdf生成的解决方案（H2p）之H2P展望计算机等级考试 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_pdf_E7_94_9F_E6_88_90_E7_c97_645282.htm 《根据批量URL生成有书签的pdf文档的解决方案》一文作为一个完整的解决方案，主要有四个要点（1）详细描述了解决方案的思路，（2）提出了一种描述URL的xml文件，称之为h2p文件（3）根据h2p文件生成pdf文档的工具称为h2p-tool，同时h2p-tool还能对h2p文件进行编辑，从而制作自己的pdf电子书。（4）h2p文件是该解决方案的资源文件，收集url是件费时费力的事情，所以该解决方案还提供了下载h2p文件的链接 H2p是我自己定义的一种xml格式，我私下里是希望这种格式能够被大家接受，所以我定义了dtd，并且开发了一个编辑h2p文件的工具h2p-tool，而且在javaei网站上提供了大量的h2p文件作为资源下载。有了h2p，制作自己的pdf电子书将变得很容易。如果能让h2p为大家所接受，h2p必须不断的发展。因为现在h2p应用还是很简单很原始。主要表现在，生成的pdf是原html的完全再现，这是他的优点也是他的弱点。说是优点，是因为pdf的呈现效果和浏览器里风格一样。说是缺点，是因为既然生成了pdf，我们希望pdf里主要突出文章的主题内容，而不希望有原来网页里的那么多的图片和广告，而且还希望原来是分页的内容自动合并到一个pdf里。所以，就此缺点，我提出两个可能的展望。（1）制定一种规范，该规范要求网站为每篇博客提供一个为h2p专用的链接（姑且叫h2p链接吧），根据这个链接打开的博客文章是没有广告的，那么h2p文件的连接采用这种h2p链接，生成的pdf就没有上述的缺点了。这个想法可

能与网站的盈利相悖，因为如果提供h2p链接，很多人将会之点击h2p链接，这样，就避开了网站投放的广告。（2）第二种可能的展望就是，各网站提供一个webservice的编程接口，通过该编程接口，用程序访问每篇博客文章，这样得到的文章就是没有广告的干净的文章。这两种展望的目的主要是为了生成干净的pdf文档，显然，要实现这两个展望是很困难的，除非大家都接受h2p这个思路，并且对这种应用很迫切，才会有人来大力推动。想想rss的普及，那也是一波三折，最终还是因为其便捷的使用模式而被大家所接受。私下里是希望大家能接受并喜欢h2p的。编辑特别推荐: JAVA技术学习法 Java学习如何走出第一步 最值得学习的五种JAVA技术 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