

pdf生成的解决方案（H2p）之H2P技术细节计算机等级考试
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_645281.htm H2P在很多人看来是个令人激动的方案，而且javaei网站提供h2p文件资源又是相当自然的一件事情，一想到h2p，我都很高兴，因为这个方案是我提出来的，尽管实现h2p的技术很成熟。在这篇文章里，说一下实现h2p所涉及到的技术吧，既有核心的j2se应用，又有开源框架的使用。（1）dtd的应用，为了规范使用者正确的编辑h2p文件，我定义了h2p文件的dtd，当然我还有一个奢望，希望能成为一种规范。（2）使用根据dtd校验xml（h2p文件），h2p-tool需要解析xml（h2p文件）提取url生成pdf并合并，不合法的xml文件肯定不能正确的生成pdf，所以必须对xml文件进行校验。（3）xml解析，编辑h2p，用JTree展示xml的树形结构，这一部分用的是dom解析（具体的是JDom），获取书签的目录结构并构建书签的数据结构，这一部分用的是SAX解析，并借助于Stack。保存xml文件用的是dom。（4）swing的应用，h2p-tool编辑功能用的是swing，展示并编辑树形结构用的是JTree（5）C#组件，根据url生成单个的pdf，用的是c#组件，并在vs.net开发平台上进行了简单的开发。Javaei提供的h2p文件通常包含几十条url，单就每条url生成pdf也需要一点时间，何况是几十条，所以运用多线程是必须的方式，研究了一下c#的多线程（c#的多线程方式也很有意思，感觉比java的简单），可是别人的组件在多线程下表现欠佳，无赖之下，只好还是采用单线程。本想用jni把c#的调用封装起来，但经过研究，据说比较麻烦，就放弃了，所以

采用简单的办法，用批处理来调用。（6）iText的应用，生成单个pdf页面以及合并pdf生成书签用的是iText。在合并的时候，也研究过另一个框架就是pdfbox。IText确实强悍，理论上，应用iText可以实现任意效果的输出，应该可以实现浏览器的效果，但比较麻烦。iText在生成书签这一块，做得相当漂亮，书签可以指向任意一个页面的任意部分，还可以设置书签对应页面的打开效果，对书签的开发也很便利，直接构造书签的树形数据结构我认为是最简单的，iText还支持用xml描述书签的结构。（7）Classloader的应用，h2p-tool的运行依赖于很多的jar包，通常我们开发的时候，都是把jar包放在指定的目录，应用服务器会去加载依赖的jar包。在h2p-tool中，需要自己处理jar包。因为这是提供给用户使用的工具，不应要求使用者去配置除了jdk环境变量以外的变量。所以解决jar包加载的问题通常的办法是在主类（带有main方法的类）所在jar包里的Manifest.mf文件里写上jar包的相对路径。这种办法不太灵活，所以在h2p-tool中，对类加载进行了改写，使其自动加载指定目录下的jar包。（8）jvm参数的应用，因为要合并的pdf很多，再合并的时候会占用大量内存，很容易out of memery，所以在批处理文件里，需要配置适当的jvm参数，主要是两个：-Xmx512m -Xms512m，这两个参数的意义这里不再赘述，网上有很多的。总结一下，实现h2p-tool的技术细节貌似很多，实则很少，除了JDOm和iText，再就是j2se核心的东西了，出列上面列到的，还包括流、字符集的处理等。

编辑特别推荐: 计算机等级考试二级JAVA考前密卷及参考答案 全国计算机等级考试二级JAVA全真模拟试题 全国计算机等级考试二级JAVA在线模拟考试 计算机等级考试二级Java初

级练习(精选25题) 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