

基础入门 - JAVA字符集详解（三）PDF转换可能丢失图片或格式，建议阅读原文

https://www.100test.com/kao_ti2020/145/2021_2022__E5_9F_BA_E7_A1_80_E5_85_A5_E9_c104_145063.htm 5.3.

URLEncoder.encode() 这是Java本身提供对的URL编码函数，完成的工作和上述UTF-8选项有效时浏览器所做的工作相似。

值得说明的是，java已经不赞成不指定编码来使用该方法

（ deprecated ）。应该在使用的時候增加编码指定。当不指定编码的时候，该方法使用系统默认编码，这会导致软件运行结果得不确定。比如对于"中文"，当系统默认编码为"gb2312"时，结果是"N-e"，而默认编码为"UTF-8"，结果却是"涓"，后续程序将难以处理。另外，这儿说的系统默认编码是由运行tomcat时的环境变量LC_ALL和LANG等决定的，曾经出现过tomcat重启后就出现乱码的问题，最后才郁闷的发现是因为修改修改了这两个环境变量。建议统一指定为"UTF-8"编码，可能需要修改相应的程序。

5.4. 一个解决方案上面说起过，因为浏览器设置的不同，对于同一个链接，web server收到的是不同内容，而软件系统有无法知道这中间的区别，所以这一协议目前还存在缺陷。针对具体问题，不应该侥幸认为所有客户的IE设置都是UTF-8有效的，也不应该粗暴的建议用户修改IE设置，要知道，用户不可能去记住每一个web server的设置。所以，接下来的解决办法就只能是让自己的程序多一点智能：根据内容来分析编码是否UTF-8。比较幸运的是UTF-8编码相当有规律，所以可以通过分析传输过来的链接内容，来判断是否是正确的UTF-8字符，如果是，则以UTF-8处理之，如果不是，则使用客户默认编码（比

如"GBK") , 下面是一个判断是否UTF-8的例子 , 如果你了解相应规律 , 就容易理解。 public static boolean isValidUtf8(byte[] b,int aMaxCount){ int lLen=b.length,lCharCount=0. for(int i=0.i byte lByte=b[i].//to fast operation, now, ready for the following for(..) if(lByte>=0) continue.//>=0 is normal ascii if(lByte>(byte)0xfd) return false. int lCount=lByte>(byte)0xfc?5:lByte>(byte)0xf8?4:lByte>(byte)0xf0?3:lByte>(byte)0xe0?2:1. if(i lCount>lLen) return false. for(int j=0.j=(byte)0xc0) return false. } return true. } 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