



链滴

Java的SaxReader解析XML遇到UTF-8有BOM出现异常的解决

作者: [Angonger](#)

原文链接: <https://ld246.com/article/1464597743226>

来源网站: [链滴](#)

许可协议: [署名-相同方式共享 4.0 国际 \(CC BY-SA 4.0\)](#)

今天解析XML遇到有人用编辑器编辑的xml中是UTF-8有BOM的编码格式，运用SaxReader解析时出了异常，异常原因是有BOM的解析前几位是EF BB BF，试了好几种方式不行，找了个折中的办法：内容重新写一个文件出来.....，代码来自网上，结尾给链接

```
import java.io.*;

public class UTF8ToAnsiUtils {

// FEFF because this is the Unicode char represented by the UTF-8 byte order mark (EF BB BF).
public static final String UTF8_BOM = "\uFEFF";

public static void main(String args[]) {
    try {
        if (args.length != 2) {
            System.out
                .println("Usage : java UTF8ToAnsiUtils utf8file ansifile");
            System.exit(1);
        }

        boolean firstLine = true;
        FileInputStream fis = new FileInputStream(args[0]);
        BufferedReader r = new BufferedReader(new InputStreamReader(fis,
            "UTF8"));
        FileOutputStream fos = new FileOutputStream(args[1]);
        Writer w = new BufferedWriter(new OutputStreamWriter(fos, "Cp1252"));
        for (String s = ""; (s = r.readLine()) != null;) {
            if (firstLine) {
                s = UTF8ToAnsiUtils.removeUTF8BOM(s);
                firstLine = false;
            }
            w.write(s + System.getProperty("line.separator"));
            w.flush();
        }

        w.close();
        r.close();
        System.exit(0);
    }

    catch (Exception e) {
        e.printStackTrace();
        System.exit(1);
    }
}

private static String removeUTF8BOM(String s) {
    if (s.startsWith(UTF8_BOM)) {
        s = s.substring(1);
    }
    return s;
}
}
```

Handle UTF8 file with BOM