

转 Pinyin4j的基本用法

2013年10月19日 14:44:11 阅读数：2056

1. 简单介绍

有时候,需要将汉字编程对应的拼音,以方便数据的处理。比如在Android手机应用的开发上,要查询联系人的姓名,通常都是用拼音进行查询的。比如要查询"曹孟德",就可以输入"cmd",即"曹孟德"三个汉字的拼音"caomengde"各字的首字母。但是怎样才能将"曹孟德"翻译成"caomengde"呢?很简单的办法就是建立一个大的对照表(比如用关联容器Map),比如<"曹","cao">,<"孟","meng">,<"德","de">...但这样的做法,需要维护好一个比较大的对照表,同时一个汉字可能有多个发音,也就是说Map这样的容器时不行的,因为其<key,value>必须是一一对应的。在C++中可以用STL里面的multimap来解决这个问题,但Java中没有类似multimap这样的东西,除非自己实现一个。

Pinyin4j就是为了解决类似这样的问题的。它是sourceforge.net上的一个开源项目,功能非常强大:

+ 支持同一汉字有多个发音

+ 还支持拼音的格式化输出,比如第几声之类的,

+ 同时支持简体中文、繁体中文转换为拼音...使用起来也非常简单。下面是其官方网址,其中提供了下载:

<http://pinyin4j.sourceforge.net/>

2. 基本用法:

通常情况下,只需要用到其中的PinyinHelper类中的静态方法toHanyuPinyinStringArray就可以了,比如:

```
[java] 1. String[] pinyinArray =PinyinHelper.toHanyuPinyinStringArray('单');
2. for(int i = 0; i < pinyinArray.length; ++i)
3. {
4.     System.out.println(pinyinArray[i]);
5. }
```

就会输出:

dan1

chan2

shan4

这三种发音,后面的数字代表第几声。可以看到静态方法toHanyuPinyinStringArray返回的数据类型是一个String数组,它用来接收一个汉字的多个发音,如果toHanyuPinyinStringArray中的参数不是汉字,那么它会返回null。

3. 格式支持

Pinyin4j支持拼音输出的格式化,比如,"黄"可以输出为"huang"、"huang2"、"huáng"等等,下面的代码就似是输出"huáng"的示例:

```
[java] 1. HanyuPinyinOutputFormat format= new HanyuPinyinOutputFormat();
2. format.setToneType(HanyuPinyinToneType.WITH_TONE_MARK);
3. format.setVCharType(HanyuPinyinVCharType.WITH_U_UNICODE);
4.
5. String[] pinyinArray = null;
6. try
7. {
8.     pinyinArray = PinyinHelper.toHanyuPinyinStringArray('黄', format);
9. }
10. catch(BadHanyuPinyinOutputFormatCombination e)
11. {
12.     e.printStackTrace();
13. }
14. for(int i = 0; i < pinyinArray.length; ++i)
15. {
16.     System.out.println(pinyinArray[i]);
17. }
```

此外,还支持大小写转换、ü等等。详细情况,可以查看Pinyin4j自带的文档。

4. 实际示例代码

如果我们要将一段文字中的汉字全部转换成不带音调的拼音输出,而这段文字中又可能包含阿拉伯数字、英文、标点符号等等。如果完全靠自己写代码进行转换,那是非常麻烦的,其中一个首先就要区别,这段文字中那些是汉字,那些是非汉字。有了Pinyin4j,这个问题就不再困难了,因为对于非汉字,Pinyin4j会自动输出null。参考下面的示例代码:

```

1. import net.sourceforge.pinyin4j.*;
2. import net.sourceforge.pinyin4j.format.HanyuPinyinOutputFormat;
3. import net.sourceforge.pinyin4j.format.HanyuPinyinToneType;
4. import net.sourceforge.pinyin4j.format.exception.BadHanyuPinyinOutputFormatCombination;
5.
6. public class Chinese2PY
7. {
8.     public static void main(String[] args)
9.     {
10.         Hanyu hanyu = new Hanyu();
11.         // 中英文混合的一段文字
12.         String str = "荆溪白石出, Hello 天寒红叶稀。Android 山路元无雨, What's up? 空翠湿人衣。";
13.         String strPinyin = hanyu.getStringPinyin(str);
14.         System.out.println(strPinyin);
15.     }
16. }
17.
18. class Hanyu
19. {
20.     private HanyuPinyinOutputFormat format = null;
21.     private String[] pinyin;
22.
23.     public Hanyu()
24.     {
25.         format = new HanyuPinyinOutputFormat();
26.         format.setToneType(HanyuPinyinToneType.WITHOUT_TONE);
27.
28.         pinyin = null;
29.     }
30.
31.     //转换单个字符
32.     public String getCharacterPinyin(char c)
33.     {
34.         try
35.         {
36.             pinyin = PinyinHelper.toHanyuPinyinStringArray(c, format);
37.         }
38.         catch (BadHanyuPinyinOutputFormatCombination e)
39.         {
40.             e.printStackTrace();
41.         }
42.
43.         // 如果c不是汉字, toHanyuPinyinStringArray会返回null
44.         if (pinyin == null) return null;
45.
46.         // 只取一个发音, 如果是多音字, 仅取第一个发音
47.         return pinyin[0];
48.     }
49.
50.     //转换一个字符串
51.     public String getStringPinyin(String str)
52.     {
53.         StringBuilder sb = new StringBuilder();
54.         String tempPinyin = null;
55.         for (int i = 0; i < str.length(); ++i)
56.         {
57.             tempPinyin = getCharacterPinyin(str.charAt(i));
58.             if (tempPinyin == null)
59.             {
60.                 // 如果str.charAt(i)非汉字, 则保持原样
61.                 sb.append(str.charAt(i));
62.             }
63.             else
64.             {
65.                 sb.append(tempPinyin);
66.             }
67.         }
68.         return sb.toString();
69.     }
70. }

```

输出结果：

jingxibaishichu, Hellotianhanhongyexi. Android shanluyuanwuyu, What'sup? kongcuishirenyi.

原文地址：<http://blog.csdn.net/pathuang68/article/details/6692882>

文章标签：[Pinyin4j](#) [用法](#) [java](#) [汉字](#)

个人分类：[J2EE](#)

此PDF由spygg生成,请尊重原作者版权!!!

我的邮箱:liushidc@163.com