

用java抓取百度Top500歌曲及源码下载计算机等级考试 PDF转换可能丢失图片或格式，建议阅读原文

[https://www.100test.com/kao\\_ti2020/638/2021\\_2022\\_\\_E7\\_94\\_A8java\\_E6\\_8A\\_93\\_c97\\_638825.htm](https://www.100test.com/kao_ti2020/638/2021_2022__E7_94_A8java_E6_8A_93_c97_638825.htm) 编辑特别推荐: 全国计算机等级

考试（等考）指定教材 全国计算机等级考试学习视频 全国计算机等级考试网上辅导招生 全国计算机等级考试时间及科目预告 百考试题教育全国计算机等级考试在线测试平台 全国计算机等级考试资料下载 全国计算机等级考试论坛 在我整理完

在线听歌(<http://www.5a520.cn>)的模块后，剩下下来的工作就是如何通过程序抓取最新的Baidu好听的歌曲，抓取的工作主要包括3个属性：歌名、歌曲在线播放地址和歌词内容（符合LRC歌词格式），目前完成歌曲和歌曲地址抓取，由于百度的歌曲地址很多通过js获取，所以歌曲地址获取我这里使用搜狗音乐搜索方便些，所有的源码如下：

```
package
com.common.utils. import java.io.BufferedReader. import
java.io.ByteArrayOutputStream. import java.io.IOException. import
java.io.InputStream. import java.io.InputStreamReader. import
java.io.OutputStreamWriter. import
java.io.UnsupportedEncodingException. import
java.net.HttpURLConnection. import
java.net.MalformedURLException. import java.net.URL. import
java.net.URLConnection. import java.net.URLDecoder. import
java.net.URLEncoder. import java.util.ArrayList. import
java.util.HashSet. import java.util.List. import java.util.Set. import
java.util.TreeSet. import java.util.regex.Matcher. import
java.util.regex.Pattern. import org.htmlparser.Node. import
```

```
org.htmlparser.NodeFilter. import org.htmlparser.Parser. import
org.htmlparser.filters.NodeClassFilter. import
org.htmlparser.filters.OrFilter. import
org.htmlparser.nodes.TextNode. import
org.htmlparser.tags.LinkTag. import org.htmlparser.util.NodeList.
import org.htmlparser.util.ParserException. import
com.common.doc.FileOperUtils. class Song{ private String name.
private String url. private String lrc. public Song(String name,String
url){ this.name = name. this.url = url. this.lrc = "". } public String
getName() { return name. } public void setName(String name) {
this.name = name. } public String getUrl() { return url. } public void
setUrl(String url) { this.url = url. } public String getLrc() { return lrc.
} public void setLrc(String lrc) { this.lrc = lrc. } } 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
```