



链滴

markdown 转文本，并保留换行

作者: [LyZane](#)

原文链接: <https://ld246.com/article/1567410788074>

来源网站: [链滴](#)

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



做了个监控系统，报警消息的原文是 markdown 格式的。

如果消息发送到企业微信，就推 markdown 格式；如果发送到邮箱，就转 html；如果发短信，就要文本。

在转文本后通过短信发送这里遇到了两个问题：

1. html 转文本后，丢掉了换行，导致短信内容没有断句。
2. 由于短信通道的问题，发出去的短信会丢掉所有不可见字符，包括换行、空格等。

markdown 转 html 再转 文本 再转短信的代码如下：

```
<!--用于 markdown 转 html-->
<dependency>
  <groupId>com.atlassian.commonmark</groupId>
  <artifactId>commonmark</artifactId>
  <version>0.13.0</version>
</dependency>
```

```
<!--用于 html 转 text-->
<dependency>
  <groupId>org.jsoup</groupId>
  <artifactId>jsoup</artifactId>
  <version>1.10.3</version>
</dependency>~~~~
```

```
import org.commonmark.parser.Parser;
import org.commonmark.renderer.html.HtmlRenderer;
import org.jsoup.Jsoup;
```

```

import org.jsoup.nodes.Document;
import org.jsoup.safety.Whitelist;

private static final Pattern MATCH_NOT_SYMBOL_PATTERN = Pattern.compile("[0-9a-z\\u4e0
-\\u9fa5]", Pattern.CASE_INSENSITIVE);

public static void main(String[] args) {
    String raw = "x-checkpoint" + System.lineSeparator() +
        System.lineSeparator() +
        "* 第三方服务不可用 22次" + System.lineSeparator() +
        System.lineSeparator() +
        "##### 失败原因: " + System.lineSeparator() +
        System.lineSeparator() +
        "1. 巴拉巴拉小魔仙! " + System.lineSeparator() +
        "2. 雨女无瓜";

    // markdown 转 html
    String html = HtmlRenderer
        .builder()
        .build()
        .render(
            Parser
                .builder()
                .build()
                .parse(raw)
        );

    System.out.println(html);
    System.out.println();

    /*
    <p>x-checkpoint</p>
    <ul>
    <li>第三方服务不可用 22次</li>
    </ul>
    <h5>失败原因: </h5>
    <ol>
    <li>巴拉巴拉小魔仙! </li>
    <li>雨女无瓜</li>
    </ol>
    */

    // html 转文本 (默认)
    String temp = Jsoup.parse(html).body().text();
    System.out.println(temp);
    System.out.println();
    /*
    x-checkpoint 第三方服务不可用 22次 失败原因: 巴拉巴拉小魔仙! 雨女无瓜
    */

    // html 转文本 (保留换行)

```

```

    String text = Jsoup.clean(html, "", Whitelist.none(), new Document.OutputSettings().prettyPrint(false));
    System.out.println(text);
    System.out.println();
    /*
x-checkpoint

第三方服务不可用 22次

失败原因：

巴拉巴拉小魔仙！
雨女无瓜
*/

    // 文本转短信
    String[] lines = text.split(System.lineSeparator());
    StringBuilder message = new StringBuilder();
    String separator = "; ";

    for (String line : lines) {
        line = line.trim();

        // 丢弃空行
        if ("".equals(line)) {
            continue;
        }

        // 检查行尾是否是符号
        String lastChar = line.substring(line.length() - 1);
        boolean lastCharIsSymbol = !MATCH_NOT_SYMBOL_PATTERN.matcher(lastChar).find();

        message.append(line);

        // 如果行尾不是符号，追加分隔符
        if (!lastCharIsSymbol) {
            message.append(separator);
        }
    }
    System.out.println(message);
    /*
x-checkpoint; 第三方服务不可用 22次; 失败原因：巴拉巴拉小魔仙！雨女无瓜;
*/
}

```

关键代码：

```

// 此处会在 html 转 text 时保留换行~~~~
String text = Jsoup.clean(html, "", Whitelist.none(), new Document.OutputSettings().prettyPrint(false));

```