

Jsoup <https://www.cnblogs.com/zhangyinhua/p/8037599.html>

概述：jsoup 是一款Java 的HTML解析器，可直接解析某个URL地址、HTML文本内容。它提供了一套非常省力的API，可通过DOM，CSS以及类似于jQuery的操作方法来取出和操作数据。

主要功能：

- 1) 从一个URL，文件或字符串中解析HTML
- 2) 使用DOM或CSS选择器来查找、取出数据
- 3) 可操作HTML元素、属性、文本

注意：jsoup是基于MIT协议发布的，可放心使用于商业项目。

解析一个HTML

```
Document doc = Jsoup.parse(html);
```

从一个URL加载一个Document

```
Document doc = Jsoup.connect("http://example.com/").get();
```

从元素抽取属性，本文和HTML

方法：

要取得一个属性的值，可以使用Node.attr(String key) 方法

对于一个元素中的文本，可以使用Element.text() 方法

对于要取得元素或属性中的HTML内容，可以使用Element.html(), 或 Node.outerHtml() 方法

设置属性值

可以使用属性设置方法 Element.attr(String key, String value), 和 Elements.attr(String key, String value)