

# jsoup

---

Jsoup é uma biblioteca Java para trabalhar com HTML. Ela permite extrair e manipular dados de HTML, usando métodos DOM, seletores CSS e métodos semelhantes ao jQuery. Jsoup é popularmente utilizada para web scraping, parsing de HTML e manipulação de dados.

## Maven

```
<dependency>
  <groupId>org.jsoup</groupId>
  <artifactId>jsoup</artifactId>
  <version>1.17.2</version>
</dependency>
```

## exemplos

### Obter o html de uma página e colher tags

```
import org.jsoup.Jsoup;
import org.jsoup.nodes.Document;
import org.jsoup.nodes.Element;
import org.jsoup.select.Elements;

public class JsoupExample {
    public static void main(String[] args) {
        try {
            // obter o documento HTML
            Document doc = Jsoup.connect("https://www.amazon.com").get();

            // Selecionar todos os parágrafos e imprimir o texto
            Elements paragraphs = doc.select("p");
            for (Element paragraph : paragraphs) {
                System.out.println(paragraph.text());
            }
        } catch (Exception e) {
            e.printStackTrace();
        }
    }
}
```

### Analiar o HTML de uma String

```
import org.jsoup.Jsoup;
import org.jsoup.nodes.Document;
```

```

public class JsoupFromStringExample {
    public static void main(String[] args) {
        String html = "<html><head><title>Exemplo</title></head>"
            + "<body><p>Parágrafo 1</p><p>Parágrafo 2</p></body></html>";

        Document doc = Jsoup.parse(html);
        System.out.println("Título: " + doc.title());
        System.out.println("Corpo: " + doc.body().text());
    }
}

```

## Modificar um documento HTML

```

import org.jsoup.Jsoup;
import org.jsoup.nodes.Document;
import org.jsoup.nodes.Element;

public class JsoupModifyExample {
    public static void main(String[] args) {
        String html = "<html><head><title>Exemplo</title></head>"
            + "<body><p>Parágrafo 1</p><p>Parágrafo 2</p></body></html>";

        Document doc = Jsoup.parse(html);

        // Adicionar um novo parágrafo
        Element body = doc.body();
        body.appendChild("p").text("Parágrafo 3");

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

```