jsoup.org

Load a Document from a File: jsoup Java HTML parser

1-2 minutes

Problem

You have a file on disk that contains HTML, that you'd like to load and parse, and then maybe manipulate or extract data from.

Solution

Use the static `Jsoup.parse(File in, String charsetName, String baseUri)` method:

```
File input = new File("/tmp/input.html");
Document doc = Jsoup.parse(input, "UTF-8",
"http://example.com/");
```

Description

The `parse(File in, String charsetName, String baseUri)` method loads and parses a HTML file. If an error occurs whilst loading the file, it will throw an `IOException`,

which you should handle appropriately.

The `baseUri` parameter is used by the parser to resolve relative URLs in the document before a `<base href>` element is found. If that's not a concern for you, you can pass an empty string instead.

There is a sister method [parse(File in, String charsetName)](#) which uses the file's location as the `baseUri`. This is useful if you are working on a filesystem-local site and the relative links it points to are also on the filesystem.