

Java HTML Parser

jsoup

output tidy HTML

manipulate the HTML elements, attributes, and Text

find and extract data

parse HTML from a URL, file, or string 🤝

parses HTML to the same DOM as modern browsers do

no JavaScript support ⚠️

implements the WHATWG HTML5 specification

working with real-world HTML

convenient API

liberal MIT license

✔️ Open source

