

HTMLTidy

HTMLTidy is a means to take badly formed HTML markup and generate well-formed XHTML.

There's a command-line utility, as well as a Java API.

This tool is *vital* if you want to 'screen scrape' data from HTML pages. Cocoon provides HTML Tidy as a [Generator](#).

See also

- [HTMLGenerator](#)
- [The Tidy project home page](#).