

NAME

akfscraper – convert HTML/XHTML into plain text

SYNOPSIS

akfscraper [*options*] [*HTML files*]

DESCRIPTION

The program converts HTML or XHTML data into plain text (Markdown).

When no files are provided, it reads the standard input. Unless the option *-o* is used, it writes to the standard output. That means, it can be used as a filter.

The output is always encoded as UTF-8. For input it supports the encodings UTF-8, ISO-8859-1 or codepage 1252.

Links are added at the end as references.

OPTIONS

- h
- help
- Hilfe shows a short help
- V
- version
- Version shows the version
- u the input should be interpreted as UTF-8, if no meta tag or XML signature says otherwise
- c *charset*
- charset *charset*
- Zeichensatz *charset*
if the *charset* is set to UTF-8 or UTF8, the input should be interpreted as UTF-8, if no meta tag or XML signature says otherwise. (This is a longer variant of -u for compatibility)
- l
- Links
- links add external links as refereces
- L *number*
like -l, but ignores *number* links at the beginning
- n no automatic line breaks (long lines)
- e
- extended
- erweitert
output with terminal control codes (SGR codes)
- o "*file*"
- output="*file*"
- Ausgabe="*file*"
- Datei="*file*"
writes output to *file* instead of the standard output
- f
- force
- forciere
overwrite existing files

EXAMPLES

```
akfscraper -l -o article.text page1.html page2.html page3.html
akfscraper -e -l article.html | less -r
```

```
curl -sL "https://gnu.org" | akfscraper -e -l | less -r
```

Mailcap:

```
text/html; akfscraper -l -c %{charset}; copiousoutput
```

AUTHORS

Copyright © 2025 Andreas K. Förster

This program is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

This program is distributed in the hope that it will be useful, but **WITHOUT ANY WARRANTY**; without even the implied warranty of **MERCHANTABILITY** or **FITNESS FOR A PARTICULAR PURPOSE**. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License along with this program. If not, see <http://www.gnu.org/licenses/>.

SEE ALSO

more(1) **less(1)** **akfweb-dl(1)** **curl(1)**

<https://akfoerster.de/p/akfnetz/>