NAME

weblint − pick fluff off web pages (HTML)

SYNOPSIS

weblint [**-d** *id*] [**-e** *id*] [**-f** *filename*] [**-i**] [**-l**] [**-s**] [**-stderr**] [**-t**] [**-todo**] [**-help**] [**-U**] [**-urlget** *command*] [**-v**] [**-version**] [**-warnings**] [**-x** *extension*] file1 .. fileN

DESCRIPTION

Weblint is a Perl script which picks fluff off HTML pages. Files to be checked are passed on the command-line:

```
% weblint foobar.html ./dodgy-files/ index.html
```

If any of the arguments are directories weblint will recurse in the directory, and check any HTML files found. If an argument is a URL, then weblint will get the file using a URL retrieval program, and then check the file:

```
% weblint http://www.foobar.com/
```

By default weblint will use lynx to retrieve URLs, but this can be over-ridden. A filename of '-' specifies that weblint should read from standard input:

```
% lynx -source http://www.foobar.com/ | weblint -
```

Warnings are generated a la lint:

```
home.html(9): unmatched </A> (no matching <A> seen).
```

Weblint includes the following features:

- by default checks for HTML 3.2 (Wilbur)
- 46 different checks and warnings
- Warnings can be enabled/disabled individually, as per your preference
- basic structure and syntax checks
- warnings for use of unknown elements and element attributes.
- context checks (where a tag must appear within a certain element).
- overlapped or illegally nested elements.
- do IMG elements have ALT text?
- flags obsolete elements.
- support for user and site configuration files
- stylistic checks
- checks for html which is not portable across all browsers
- flags markup embedded in comments, since this can confuse some browsers
- support for Netscape, and Microsoft HTML extensions

OPTIONS

-d *warning-identifier*
: Disable the warning associated with the identifier. Multiple identifiers can be specified, with a comma between identifiers.

-e *warning-identifier*
: Enable the warning associated with the identifier. Multiple identifiers can be specified, with a comma between identifiers.

-f *config-file*
: Specify a weblint configuration file which should be used in place of the user's default config file, or the site configuration file.

-help Show a short usage summary.

-i Ignore case of element tags.

-l When recursing in directories, ignore any files which are symlinks (also known as soft links). This will also cause files on the command-line to be ignored if they are symlinks, unless only one file is given.

-pedantic
: Turn on all warnings except the case-sensitive and bad-link warnings.

-s Generate 'short' warning messages, which do not include the filename.

-stderr Print warning messages to STDERR rather than STDOUT.

-t Enable terse warning mode, which is mainly useful for the weblint testsuite.

-U Same as -help.

-urlget *command*
: The command which should be used to retrieve HTML pages specified by URL.

-v Display the version number.

-version
: Display the version number.

-todo This prints out the URL for the online version of the weblint ToDo list. This includes known bugs, and requested/planned features.

-warnings
: List all supported warnings, with warning identifier, and whether the warning is enabled.

-x *extension*
: Include checks for the specified HTML extension; multiple extensions can be specified, separated with a comma. Currently the only extensions supported are *Netscape* and *Microsoft*. This can also be set in your weblint configuration file, described below.

HTML EXTENSIONS

Unless you specify otherwise, weblint assumes you are using HTML 3.2. Weblint supports the Netscape and Microsoft HTML extensions in addition. For example, weblint will complain that the BLINK element is not known, unless you enable the **Netscape** extension. The following extensions are currently supported:

Netscape
: The HTML extensions supported by the Netscape browser, version 4.

Microsoft
: The HTML extensions supported by Microsoft Internet Explorer, version 4.

To enable an extension, you can either use the **-x** command-line switch:

```
% weblint -x Netscape foobar.html
```

Or you can use the **extension** keyword in your *.weblintrc*:

```
# enable the Microsoft extensions
extension Microsoft
```

CONFIGURATION FILE

Weblint can be configured using a file *.weblintrc* in your home directory (or a file referenced by the WEBLINTRC environment variable). This file can be used to enable or disable specific warnings, set weblint variables, and include HTML extensions, as described above. Each warning has a short identifier string, used to refer to the warning in config files, and from the command-line. For example, if you want to enable the check for tags in upper-case, but disable the check for obsolete elements, then you would include the following lines in your .weblintrc:

```
# specify the command used to retrieve URLs (-urlget switch)
set url-get = lynx -source

# the style of warning message to generate (lint, short, or terse)
set message-style = lint

# enable warning for tags not in upper-case
enable upper-case

# disable the warning for obsolete tags
disable obsolete

# enable the Netscape HTML extensions
extension Netscape

# when recursing in a directory,
# ignore files which are symlinks (also known as soft links)
ignore symlinks
```

The keywords can be followed by any number of arguments, separated by spaces or tabs. Anything following a '#' is treated as a comment.

A sample configuration file is included in the weblint distribution (as of version 1.004), which mirrors the configuration built-in to weblint.

Weblint also supports a site configuration file. If a user does not have a personal configuration file, then weblint will check for a local site configuration file. To provide such a file, create a directory such as /usr/local/weblint, and create a file global.weblintrc. You need to edit the weblint script and modify the $SITE_DIR variable, which you will find near the top of the file. For example:

```
$SITE_DIR = '/usr/local/weblint';
```

At some point in the future there will be configuration support for weblint, so you won't have to modify the script directly yourself.

If you have a site configuration file, then users can inherit the site defaults by adding the following line at the top of their .weblintrc file:

```
use global weblintrc
```

WARNINGS

All warnings generated by *weblint* are listed below, along with the associated identifier, and whether the warning is enabled or disabled by default.

tag <...> is not in upper case.
Identifier: upper-case
Default: disabled

tag <...> is not in lower case.
Identifier: lower-case
Default: disabled

***foo* attribute is required for <...>**
Identifier: required-attribute
Default: enabled

expected an attribute for <...>
Identifier: expected-attribute
Default: enabled

unknown element <...>
Identifier: unknown-element
Default: enabled

unknown attribute '...' for element <...>.
Identifier: unknown-attribute
Default: enabled

should not have whitespace between '<' and '...>'
Identifier: leading-whitespace
Default: enabled

bad form to use 'here' as an anchor!
Identifier: here-anchor
Default: enabled

no <TITLE> in HEAD element.
Identifier: require-head
Default: enabled

tag <...> should only appear once. I saw one on line XX!
Identifier: once-only
Default: enabled

<BODY> but no <HEAD>.
Identifier: body-no-head
Default: enabled

outer tags should be <HTML> .. </HTML>.
Identifier: html-outer
Default: enabled

<...> can only appear in the HEAD element.
Identifier: head-element
Default: enabled

<...> cannot appear in the HEAD element.
Identifier: non-head-element
Default: enabled

<...> is obsolete.
Identifier: obsolete
Default: enabled

unmatched </...> (no matching <...> seen).
Identifier: mis-match
Default: enabled

IMG does not have ALT text defined.
Identifier: img-alt
Default: enabled

<...> cannot be nested.
Identifier: nested-element
Default: enabled

Did not see <LINK REV=MADE HREF=mailto:...> in HEAD.
Identifier: mailto-link
Default: disabled

</...> on line XX seems to overlap <...>, opened on line YY.
Identifier: element-overlap
Default: enabled

no closing </...> seen for <...> on line XX.
Identifier: unclosed-element
Default: enabled

markup embedded in a comment can confuse some browsers.
Identifier: markup-in-comment
Default: enabled

odd number of quotes in element <...>.
Identifier: odd-quotes
Default: enabled

heading <H?> follows <H?> on line N.
Identifier: heading-order
Default: enabled

target for anchor
Identifier: bad-link
Default: disabled

unexpected < in <...> -- potentially unclosed element.
Identifier: unexpected-open
Default: enabled

illegal context for <...> - must appear in <...> element.
Identifier: required-context
Default: enabled

unclosed comment (comment should be: <!-- ... -->
Identifier: unclosed-comment
Default: enabled

element <...> is not a container -- </...> not legal.
Identifier: illegal-closing
Default: enabled

<...> is physical font markup -- use logical (such as XXX)
Identifier: physical-font
Default: disabled

attribute XYZ is repeated in element <...>
Identifier: repeated-attribute
Default: enabled

empty container element <...>
Identifier: empty-container
Default: enabled

use of ' for attribute value delimiter is not supported by all browsers (attribute XYZ of tag ABC)
Identifier: attribute-delimiter
Default: enabled

closing tag <...> should not have any attributes specified.
Identifier: closing-attribute
Default: enabled

directory DIR does not have an index file (index.html)
Identifier: directory-index
Default: enabled

<...> must immediately follow <...>
Identifier: must-follow
Default: enabled

setting WIDTH and HEIGHT attributes on IMG tag can improve rendering performance on some browsers
Identifier: img-size
Default: disabled

leading/trailing whitespace in content of container element ...
Identifier: container-whitespace
Default: disabled

first element was not DOCTYPE specification
Identifier: require-doctype
Default: disabled

'>' should be represented as '>'
Identifier: literal-metacharacter
Default: enabled

malformed heading - open tag is <H?>, but closing is </H?>
Identifier: heading-mismatch
Default: enabled

illegal context, <...>, for text; should be in XXX.
Identifier: bad-text-context
Default: enabled

illegal value for AAA attribute of XXX (...)
Identifier: attribute-format
Default: enabled

<...> is extended markup (use '-x <extension>' to allow this).
Identifier: extension-markup
Default: enabled

attribute '...' for <...> is extended markup (use '-x <extension>' to allow this).
Identifier: extension-attribute
Default: enabled

value for attribute XYZ (xyz-value) of element FOOBAR should be quoted (i.e. XYZ='xyz-value')
Identifier: quote-attribute-value
Default: enabled

you should use '>' in place of '>', even in a PRE element.
Identifier: meta-in-pre
Default: enabled

<A> should be inside <H?>, not <H?> inside <A>.
Identifier: heading-in-anchor
Default: enabled

The HTML spec. recommends the TITLE be no longer than 64 characters.
Identifier: title-length
Default: enabled

TESTSUITE

A simple regression testsuite is included with weblint, in the Perl script `test.pl`. You can run the testsuite with either of the following commands:

```
% make test
% ./test.pl
```

The results are printed to STDERR, with a more complete report generated in test.log.

All tests should pass. If any tests fail, please email test.log to the address given in the AUTHOR section below.

ENVIRONMENT VARIABLES

WEBLINTRC

If this variable is defined, and references a file, then *weblint* will read the referenced file for the user's configuration, rather than $HOME/.weblintrc.

TMPDIR

The directory where weblint will create temporary working files. Defaults to /usr/tmp.

FILES

$HOME/.weblintrc

The user's configuration file. See the section 'CONFIGURATION FILE'.

SEE ALSO

perl(1)

VERSION

This man page describes *weblint 1.020*.

AVAILABILITY

`ftp://ftp.cre.canon.co.uk/pub/weblint/weblint.tar.gz`

`http://www.cre.canon.co.uk/~neilb/weblint/`

KNOWN BUGS

The list of known bugs can be found on the weblint home page:

http://www.cre.canon.co.uk/~neilb/weblint/todo/

Certain versions of Perl have bugs which are triggered by weblint. You shouldn't experience problems if you have 4.036, or 5.002.

AUTHOR

Neil Bowers, Canon Research Centre Europe
neilb@cre.canon.co.uk

CONTRIBUTIONS

Lots of people have contributed to weblint, in the form of suggestions, bug reports, fixes, and contributed code. Please email me if your name should appear in the roll call below.

Abigail <abigail@mars.ic.iaf.nl>; Anthony Thyssen <anthony@cit.gu.edu.au>; Axel Boldt <axel@uni-paderborn.de>; Barry Bakalor <barry@hal.com>; Bill Arnett <billa@netcom.com>; Bob Friesenhahn <bfriesen@simple.dallas.tx.us>; Mark Gates <mr-gates@uiuc.edu>; Bruce Speyer <bspeyer@texas-one.org>; Chris Siebenmann <cks@hawkwind.utcs.toronto.edu>; Clay Webster <clay@unipress.com>; Dana Jacobsen <dana@acm.org>; David Begley <david@bacall.nepean.uws.edu.au>; David J. MacKenzie <djm@va.pubnix.com>; Douglas Brick <dbrick@u.washington.edu>; Gil Citro; Eric de Mund <ead@ixian.com>; Richard Finegold <goldfndr@eskimo.com>; Joerg Heitkoetter <Joerg.Heitkoetter@germany.eu.net>; David Koblas <koblas@homepages.com>; John Labovitz <johnl@ora.com>; Eric Maryniak <E.Maryniak@rgd.nl>; John F. Whitehead <jfw@wral-tv.com> Juergen Schoenwaelder <schoenw@ibr.cs.tu-bs.de>; Frank Steinke <fsteinke@zeta.org.au>; Larry Virden <lvirden@cas.org>; Paul Black <black@lal.cs.byu.edu>; Doug Grinbergs <dougg@qualcomm.com>; Philip Hallstrom <philip@wolfe.net>; Craig Leres <leres@ee.lbl.gov>; Richard Lloyd

<R.K.Lloyd@csc.liv.ac.uk>; Charles F. Randall <crandall@dmacc.cc.ia.us>; Robert Schmunk <pcrxs@nasagiss.giss.nasa.gov>; Jeff Schave <schave@engr.wisc.edu>; Jon Thackray <jrmt@uk.gdscorp.com>; Jens Thordarson <thordurh@rhi.hi.is>; Ryan Waldron <rew@nuance.com>; Thomas Leavitt <leavitt@webcom.com>; Tom Neff <tneff@panix.com>; Victor Parada <vparada@inf.utfsm.cl>; Erick Branderhorst <branderhorst@fgg.eur.nl>; Bryan O'Sullivan <bos@serpentine.com>; Alan J. Flavell <FLAVELL@v2.ph.gla.ac.uk>; Raphael Manfredi <Raphael_Manfredi@grenoble.hp.com>; Keith Iosso <a-keithi@microsoft.com>; Chris Lambert <lambertc@sharelink.com>; Tristan Savatier <tristan@creative.net>; Phil Hooper <hooper@bcci.eng.sun.com>; Gerald Viers <grviers@csupomona.edu>; Dean Brissinger <brissing@bvsd.k12.co.us>; Dave Schmitt <dschmi1@gl.umbc.edu>; John Van Essen <vanes002@maroon.tc.umn.edu>; Brandon Bell <brandon@arcs.bcit.bc.ca>; Fumio Moriya and Toshiaki Nomura <dsfrsoft@oai6.yk.fujitsu.co.jp>; Vincent Lefevre <vlefevre@ens-lyon.fr>; Jason Mathews <mathews@nssdc.gsfc.nasa.gov>; Lars Balker Rasmussen <lbr@mjolner.dk>; Richard L. Hawes <rhawes@dmapub.dma.org>.