



Rocky Enterprise Linux 9.2 Manual Pages on command 'HTML::Parse.3pm'

C:\>man HTML::Parse.3pm

HTML::Parse(3pm) User Contributed Perl Documentation HTML::Parse(3pm)

NAME

HTML::Parse - Deprecated, a wrapper around HTML::TreeBuilder

VERSION

This document describes version 5.07 of HTML::Parse, released August 31, 2017 as part of HTML-Tree.

SYNOPSIS

See the documentation for HTML::TreeBuilder

DESCRIPTION

Disclaimer: This module is provided only for backwards compatibility with earlier versions of this library. New code should not use this module, and should really use the HTML::Parser and HTML::TreeBuilder modules directly, instead.

The "HTML::Parse" module provides functions to parse HTML documents. There are two functions exported by this module:

parse_html(\$html) or parse_html(\$html, \$obj)

This function is really just a synonym for \$obj->parse(\$html) and \$obj is

assumed to be a subclass of "HTML::Parser". Refer to HTML::Parser for more documentation.

If \$obj is not specified, the \$obj will default to an internally created new "HTML::TreeBuilder" object configured with strict_comment() turned on. That class implements a parser that builds (and is) a HTML syntax tree with HTML::Element objects as nodes.

The return value from parse_html() is \$obj.

parse_htmlfile(\$file, [\$obj])

Same as parse_html(), but pulls the HTML to parse, from the named file.

Returns "undef" if the file could not be opened, or \$obj otherwise.

When a "HTML::TreeBuilder" object is created, the following variables control how parsing takes place:

\$HTML::Parse::IMPLICIT_TAGS

Setting this variable to true will instruct the parser to try to deduce implicit elements and implicit end tags. If this variable is false you get a parse tree that just reflects the text as it stands. Might be useful for quick & dirty parsing. Default is true.

Implicit elements have the implicit() attribute set.

\$HTML::Parse::IGNORE_UNKNOWN

This variable controls whether unknown tags should be represented as elements in the parse tree. Default is true.

\$HTML::Parse::IGNORE_TEXT

Do not represent the text content of elements. This saves space if all you want is to examine the structure of the document. Default is false.

\$HTML::Parse::WARN

Call warn() with an appropriate message for syntax errors. Default is false.

REMEMBER!

HTML::TreeBuilder objects should be explicitly destroyed when you're finished with them. See HTML::TreeBuilder.

SEE ALSO

HTML::Parser, HTML::TreeBuilder, HTML::Element

AUTHOR

Current maintainers:

? Christopher J. Madsen "<perl AT cjmweb.net>"

? Jeff Fearn "<jfearn AT cpan.org>"

Original HTML-Tree author:

? Gisle Aas

Former maintainers:

? Sean M. Burke

? Andy Lester

? Pete Krawczyk "<petek AT cpan.org>"

You can follow or contribute to HTML-Tree's development at

<<https://github.com/kentfredric/HTML-Tree>>.

COPYRIGHT AND LICENSE

Copyright 1995-1998 Gisle Aas, 1999-2004 Sean M. Burke, 2005 Andy Lester, 2006 Pete Krawczyk, 2010 Jeff Fearn, 2012 Christopher J. Madsen.

This library is free software; you can redistribute it and/or modify it under the same terms as Perl itself.

The programs in this library are distributed in the hope that they will be useful, but without any warranty; without even the implied warranty of merchantability or fitness for a particular purpose.

perl v5.28.1

2019-01-13

HTML::Parse(3pm)