

NAME

HTML::Tree - build and scan parse-trees of HTML

VERSION

This document describes version 5.03 of HTML::Tree, released September 22, 2012 as part of HTML-Tree.

SYNOPSIS

```
use HTML::TreeBuilder;
my $tree = HTML::TreeBuilder->new();
$tree->parse_file($filename);

# Then do something with the tree, using HTML::Element
# methods -- for example:

$tree->dump

# Finally:

$tree->delete;
```

DESCRIPTION

HTML-Tree is a suite of Perl modules for making parse trees out of HTML source. It consists of mainly two modules, whose documentation you should refer to: [HTML::TreeBuilder](#) and [HTML::Element](#).

[HTML::TreeBuilder](#) is the module that builds the parse trees. (It uses [HTML::Parser](#) to do the work of breaking the HTML up into tokens.)

The tree that TreeBuilder builds for you is made up of objects of the class [HTML::Element](#).

If you find that you do not properly understand the documentation for [HTML::TreeBuilder](#) and [HTML::Element](#), it may be because you are unfamiliar with tree-shaped data structures, or with object-oriented modules in general. Sean Burke has written some articles for *The Perl Journal* (www.tpj.com) that seek to provide that background. The full text of those articles is contained in this distribution, as:

[HTML::Tree::AboutObjects](#)

“User’s View of Object-Oriented Modules” from TPJ17.

[HTML::Tree::AboutTrees](#)

“Trees” from TPJ18

[HTML::Tree::Scanning](#)

“Scanning HTML” from TPJ19

Readers already familiar with object-oriented modules and tree-shaped data structures should read just the last article. Readers without that background should read the first, then the second, and then the third.

METHODS

All these methods simply redirect to the corresponding method in [HTML::TreeBuilder](#). It’s more efficient to use [HTML::TreeBuilder](#) directly, and skip loading [HTML::Tree](#) at all.

new

Redirects to “new” in [HTML::TreeBuilder](#).

new_from_file

Redirects to “new_from_file” in [HTML::TreeBuilder](#).

new_from_content

Redirects to “new_from_content” in [HTML::TreeBuilder](#).

new_from_url

Redirects to “new_from_url” in HTML::TreeBuilder.

SUPPORT

You can find documentation for this module with the [perldoc\(1\)](#) command.

```
perldoc HTML::Tree
```

You can also look for information at:

- AnnoCPAN: Annotated CPAN documentation
<<http://annocpan.org/dist/HTML-Tree>>
- CPAN Ratings
<<http://cpanratings.perl.org/d/HTML-Tree>>
- RT: CPAN’s request tracker
<<http://rt.cpan.org/NoAuth/Bugs.html?Dist=HTML-Tree>>
- Search CPAN
<<http://search.cpan.org/dist/HTML-Tree>>
- Stack Overflow
<<http://stackoverflow.com/questions/tagged/html-tree>>

If you have a question about how to use HTML-Tree, Stack Overflow is the place to ask it. Make sure you tag it both `perl` and `html-tree`.

SEE ALSO

[HTML::TreeBuilder](#), [HTML::Element](#), [HTML::Tagset](#), [HTML::Parser](#), [HTML::DOMbo](#)

The book *Perl & LWP* by Sean M. Burke published by O’Reilly and Associates, 2002. ISBN: 0-596-00178-9

It has several chapters to do with HTML processing in general, and HTML-Tree specifically. There’s more info at:

<http://www.oreilly.com/catalog/perllwp/>

<http://www.amazon.com/exec/obidos/ASIN/0596001789>

SOURCE REPOSITORY

HTML-Tree is now maintained using Git. The main public repository is <<http://github.com/madsen/HTML-Tree>>.

The best way to send a patch is to make a pull request there.

ACKNOWLEDGEMENTS

Thanks to Gisle Aas, Sean Burke and Andy Lester for their original work.

Thanks to Chicago Perl Mongers (<http://chicago.pm.org>) for their patches submitted to [HTML::Tree](#) as part of the Phalanx project (<http://qa.perl.org/phalanx>).

Thanks to the following people for additional patches and documentation: Terrence Brannon, Gordon Lack, Chris Madsen and Ricardo Signes.

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 <http://github.com/madsen/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. (Except the articles contained in [HTML::Tree::AboutObjects](#), [HTML::Tree::AboutTrees](#), and [HTML::Tree::Scanning](#), which are all copyright 2000 The Perl Journal.)

Except for those three TPJ articles, the whole HTML-Tree distribution, of which this file is a part, is free software; you can redistribute it and/or modify it under the same terms as Perl itself.

Those three TPJ articles may be distributed 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.