Anyone have a diff algorithm for rendered HTML? [closed]

Asked 16 years, 3 months ago   Modified 9 years, 11 months ago

Viewed 41k times

93

Improve this question

I'm interested in seeing a good diff algorithm, possibly in Javascript, for rendering a side-by-side diff of two HTML pages. The idea would be that the diff would show the differences of the *rendered* HTML.

To clarify, I want to be able to see the side-by-side diffs *as* rendered output. So if I delete a paragraph, the side by side view would know to space things correctly.

@Josh exactly. Though maybe it would show the deleted text in red or something. The idea is that if I use a WYSIWYG editor for my HTML content, I don't want to have to switch to HTML to do diffs. I want to do it with two WYSIWYG editors side by side maybe. Or at least display diffs side-by-side in an end-user friendly matter.

`javascript` `html` `diff`

Share

Improve this question

Follow

1   Is this really haacked of microsoft ? the original one ? :D
    – Ahmed Khalaf Oct 1, 2009 at 9:48

1   Sounds similar to this question: stackoverflow.com/questions/1061468/html-compare
    – rjmunro Nov 17, 2011 at 11:39

2   Old thread, but figured I'd give my 2 cents. I've been dealing with it a lot lately, and have found a bunch of libs: web.onassar.com/blog/2012/11/21/htmldiff-software-discoveries – onassar Feb 23, 2013 at 6:04

2   @Haacked, ever find a satisfactory solution? – DG. Mar 27, 2014 at 7:06

@Haacked Found a working solution ? I tried `prettydiff.com` but looks like it's broken. – coding_idiot
May 17, 2019 at 4:46 ✏

12 Answers

Sorted by: Highest score (default) ⇕

▲

18

▼

🔖

🕘

Over the weekend I posted a new project on codeplex that implements an HTML diff algorithm in C#. The original algorithm was written in Ruby. I understand you were looking for a JavaScript implementation, perhaps having one available in C# with source code could assist you to port the algorithm. Here is the link if you are interested: htmldiff.codeplex.com. You can read more about it here.

UPDATE: This library has been moved to GitHub.

Share Improve this answer

Follow

edited Apr 11, 2012 at 7:08

answered Nov 2, 2009 at 10:21

Rohland
1,417 ● 14 ● 20

4 The HTMLDiff algorithm has been ported in JavaScript (CoffeeScript), here. – Florian Parain Nov 15, 2012 at 8:59

Hi @pate, I believe these links aren't working anymore...
– Tiago Cardoso Nov 25, 2013 at 11:32

@TiagoCardoso deleted my dated comment.
– Petrus Theron Nov 25, 2013 at 12:43

@Rohland Awesome library! Really saved my day :)
– Sirar Salih Sep 22, 2015 at 8:31

3   @Florian Parain: Thanks for sharing the link to github.com/tnwinc/htmldiff.js it works great! Tip for anyone who wants the JS version, simply install via "npm install htmldiff" and then use htmldiff.js in the "src" directory.
– Elijah Lofgren Feb 9, 2016 at 20:26

17

There's another nice trick you can use to significantly improve the look of a rendered HTML diff. Although this doesn't fully solve the initial problem, it will make a significant difference in the appearance of your rendered HTML diffs.

Side-by-side rendered HTML will make it very difficult for your diff to line up vertically. Vertical alignment is crucial for comparing side-by-side diffs. In order to improve the vertical alignment of a side-by-side diff, you can insert invisible HTML elements in each version of the diff at "checkpoints" where the diff should be vertically aligned. Then you can use a bit of client-side JavaScript to add vertical spacing around checkpoint until the sides line up vertically.

Explained in a little more detail:

If you want to use this technique, run your diff algorithm and insert a bunch of `visibility:hidden` `<span>`s or tiny `<div>`s wherever your side-by-side versions should match up, according to the diff. Then run JavaScript that finds each checkpoint (and its side-by-side neighbor) and

adds vertical spacing to the checkpoint that is higher-up (shallower) on the page. Now your rendered HTML diff will be vertically aligned up to that checkpoint, and you can continue repairing vertical alignment down the rest of your side-by-side page.

Share   Improve this answer

Follow

4

I ended up needing something similar awhile back. To get the HTML to line up side to side, you could use two iFrames, but you'd then have to tie their scrolling together via javascript as you scroll (if you allow scrolling).

To see the diff, however, you will more than likely want to use someone else's library. I used DaisyDiff, a Java library, for a similar project where my client was happy with seeing a single HTML rendering of the content with MS Word "track changes"-like markup.

HTH

Share   Improve this answer

Follow

daisydiff seems nice – May 17, 2019 at 4:58

Consider using the output of links or lynx to render a text-only version of the html, and then diff that.

4

Share   Improve this answer

Follow

answered Jan 12, 2009 at 22:38

Arafangion
11.9k ● 1 ● 43 ● 74

What about [DaisyDiff](#) ([Java](#) and [PHP](#) vesions available).

2

Following features are really nice:

- Works with badly formed HTML that can be found "in the wild".

- The diffing is more specialized in HTML than XML tree differs. Changing part of a text node will not cause the entire node to be changed.

- In addition to the default visual diff, HTML source can be diffed coherently.

- Provides easy to understand descriptions of the changes.

- The default GUI allows easy browsing of the modifications through keyboard shortcuts and links.

answered Oct 20, 2009 at 8:54

elhoim
6,511 ● 2 ● 26 ● 31

So, you expect

```
<font face="Arial">Hi Mom</font>
```

and

```
<span style="font-family:Arial;">Hi Mom</span>
```

to be considered the same?

The output depends very much on the User Agent. Like [Ionut Anghelcovici suggests](), make an image. Do one for every browser you care about.

edited May 23, 2017 at 11:54

Community Bot
1 ● 1

answered Aug 28, 2008 at 6:43

Josh
8,016 ● 5 ● 43 ● 63

If it is XHTML (which assumes a lot on my part) would the Xml Diff Patch Toolkit help? http://msdn.microsoft.com/en-us/library/aa302294.aspx

0

For smaller differences you might be able to do a normal text-diff, and then analyse the missing or inserted pieces to see how to resolve it, but for any larger differences you're going to have a very tough time doing this.

For instance, how would you detect, and show, that a left-aligned image (floating left of a paragraph of text) has suddenly become right-aligned?

Using a text differ will break on non-trivial documents. Depending on what you think is intuitive, XML differs will probably generate diffs that aren't very good for text with markup. AFAIK, DaisyDiff is the only library specialized in HTML. It works great for a subset of HTML.

If you were working with Java and XHTML, [XMLUnit](#) allows you to compare two XML documents via the [org.custommonkey.xmlunit.DetailedDiff](#) class:

> Compares and describes all the differences between two XML documents. The document comparison does not stop once the first unrecoverable difference is found, unlike the Diff class.

0

Share Improve this answer

Follow

answered Oct 5, 2008 at 23:52

[Ates Goral](#)

140k ● 27 ● 141 ● 191

Use the markup mode of Pretty Diff for HTML. It is written entirely in JavaScript.

[http://prettydiff.com/](#)

0

Share Improve this answer

Follow

edited Feb 8, 2011 at 22:34

[austincheney](#)

9 ● 3

answered Sep 10, 2009 at 19:07

austin cheney

---

1 Does it support rendered diffs? I can't find the option.
– [Joel Peltonen](#) Dec 10, 2012 at 13:57

2   Downvoted as I couldn't find the rendered diff.
    – Tiago Cardoso Nov 25, 2013 at 11:33

I believe a good way to do this is to render the HTML to an image and then use some diff tool that can compare images to spot the differences.

Share   Improve this answer

Follow

answered Aug 28, 2008 at 6:42

Ionut Anghelcovici

-7

Any misalignment in these two images will of course produce massive differences, where the actual difference is just miniscule, like a table being one pixel higher in one of the two pages. – Lasse V. Karlsen Sep 15, 2008 at 19:42