

Cleaning Up Word HTML in Adobe Dreamweaver

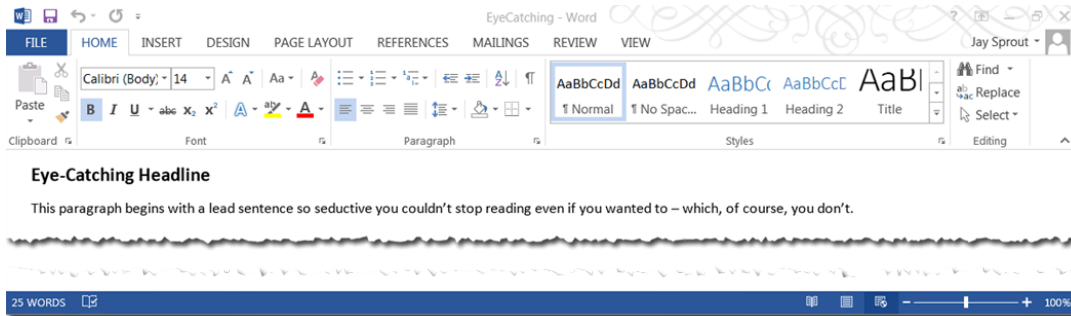
Sprout Means Grow



Why does Word HTML Need Cleaning?

I could say, "Word-generated HTML needs cleaning because it creates a lot of unnecessary code" but that really doesn't do justice to the horrific nightmare that is Word-generated HTML. It's better if I show you. If you can stomach it. You've been warned.

The following 25 words in Microsoft Word ...

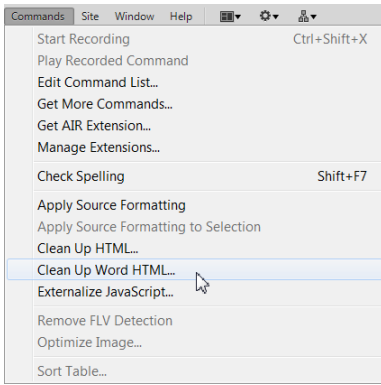


... become the following 758 lines of HTML code after using the diabolical **Save As Webpage** command.

```
DW File Edit View Insert Modify Format Commands Site Window Help Designer
EyeCatching.htm x C:\Users\jsprout\Desktop\from_Word\EyeCatching.htm
Source Code filelist.xml themedata.thmx colorschememapping.xml
Code Split Design Live Title:
726 mso-pagination:widow-orphan;
727 font-size:11.0pt;
728 font-family:"Calibri", "sans-serif";
729 mso-ascii-font-family:Calibri;
730 mso-ascii-theme-font:minor-latin;
731 mso-hansi-font-family:Calibri;
732 mso-hansi-theme-font:minor-latin;}
733 </style>
734 <![endif]--><!--[if gte mso 9]><xml>
735 <o:shapedefaults v:ext="edit" spidmax="1026"/>
736 </xml><![endif]--><!--[if gte mso 9]><xml>
737 <o:shapelayout v:ext="edit">
738 <o:idmap v:ext="edit" data="1"/>
739 </o:shapelayout></xml><![endif]-->
740 </head>
741
742 <body lang=EN-US style='tab-interval:.5in'>
743
744 <div class=WordSection1>
745
746 <p class=MsoNormal><b style='mso-bidi-font-weight:normal'><span
747 style='font-size:14.0pt;mso-bidi-font-size:11.0pt;line-height:107%'>Eye-Catching
748 Headline<o:p></o:p></span></b></p>
749
750 <p class=MsoNormal>This paragraph begins with a lead sentence so seductive you
751 couldn't stop reading even if you wanted to - which, of course, you don't. </p>
752
753 </div>
754
755 </body>
756
757 </html>
758
38K / 1 sec Western European
```

Delightful Trivia for Mac-Snobs

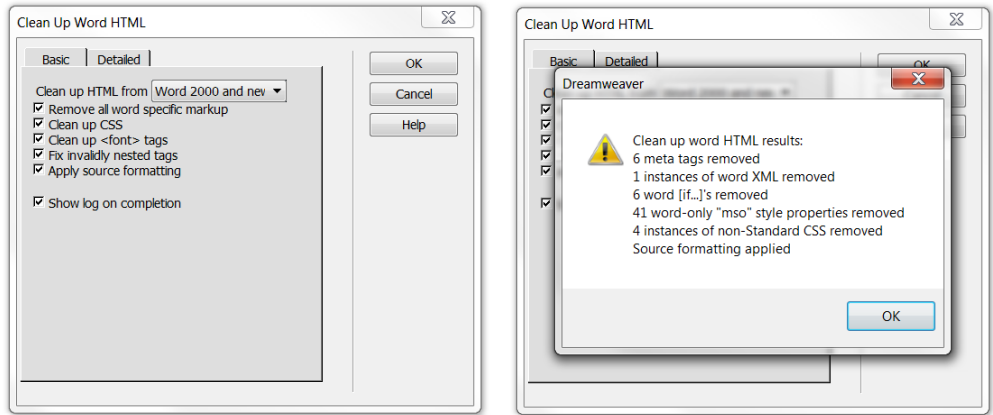
This only happens in the Windows version of MS Word.



Cleaning Up Word HTML

Word-generated HTML is so bad there's a specific command for fixing it in Dreamweaver! The process is simple.

1. Choose **Clean Up Word HTML** from the **Commands** menu.
2. Click **OK**.
3. Click **OK** again.



In my example, Dreamweaver ripped out 732 lines of code!

```
1 <html xmlns="http://www.w3.org/TR/REC-html40">
2 <head>
3 <meta http-equiv=Content-Type content="text/html; charset=windows-1252">
4 <style>
5 <!--
6 .MsoChpDefault {
7     font-family: "Calibri", "sans-serif";
8 }
9 .MsoPapDefault {
10     line-height: 107%;
11 }
12 -->
13 </style>
14 </head>
15
16 <body lang=EN-US>
17 <div class=WordSection1>
18 <p><b><span
19 style='font-size:14.0pt;line-height:107%'>Eye-Catching
20     Headline</span></b></p>
21 <p>This paragraph begins with a lead sentence so seductive you
22     couldn't stop reading even if you wanted to - which, of course, you don't. </p>
23 </div>
24 </body>
25 </html>
26
```

I'm even more strict and ripped out another 14 unnecessary lines.

```
1 <html xmlns="http://www.w3.org/TR/REC-html40">
2 <head>
3 <meta http-equiv=Content-Type content="text/html; charset=windows-1252">
4 </head>
5
6 <body lang=EN-US>
7
8 <h1>Eye-Catching Headline</h1>
9 <p>This paragraph begins with a lead sentence so seductive you couldn't stop reading even if
10 you wanted to - which, of course, you don't. </p>
11 </body>
12 </html>
```