HTML Sanitizer not working for Outlook mails.
There are some unwanted tags not stripped from the mail body.

Related issues:
Related to Redmine - Feature # 16962: Better handle html-only emails

History
#1 - 2015-04-29 15:54 - Toshi MARUYAMA
- Target version set to 3.0.3

#2 - 2015-04-29 15:55 - Toshi MARUYAMA
This is Redmine 3.0 regression according to #19537#note-9.

#4 - 2015-04-29 15:59 - Toshi MARUYAMA
- Related to Feature #16962: Better handle html-only emails added

#5 - 2015-04-29 16:56 - Deoren Moor
+1

I can confirm that Outlook emails are improperly handled:

```html
{{collapse(View details)}
@if !mso>-
<style>"[behavior:url(#default#VML);]
'o:"[behavior:url(#default#VML);]
\w:"[behavior:url(#default#VML);]
.shape [behavior:url(#default#VML);]
</style>-[
@end]

/* Font Definitions */
@font-face {font-family:Wingdings;
panose-1:5 0 0 0 0 0 0 0 0 0;}
@font-face {font-family:Wingdings;
panose-1:5 0 0 0 0 0 0 0 0 0;}
@font-face {font-family:Calibri;
panose-1:2 15 5 2 2 4 3 2 4 2;}
@font-face {font-family:Tahoma;
panose-1:2 11 6 4 3 5 4 2 4 2;}
@font-face {font-family:Verdana;
panose-1:2 11 6 4 3 5 4 2 4 2;}
```
The actual human-readable text is three lines. However a similar issue is also happening with iOS (v7.x) emails as I reported on #15716.

Another issue I noticed along with this one is that the **Truncate emails after one of these lines** setting does not appear to be working. I've opened #19740 for that.

Additional notes:

- I tested sending from Outlook and converting to Plain text when replying and the email was scraped properly. The truncate option also seemed to work properly.
- Same (good, expected) results when responding via Exchange web mail and first converting format to Plain text before replying.
- I tested using Exchange web mail and got similar results (see below).
- I tested replying using Outlook and leaving the email as-is and the email was rejected with **Validation failed: Documented cannot be blank**. Presumably this is because of the Truncate option not working properly because the custom **Documented** field is listed well below the Truncate phrase that we're using.

Reply from Exchange webmail, left in HTML format

```
-->
body {font-family:Verdana,sans-serif; font-size:0.8em; color:#484848}
h1, h2, h3 {font-family:"Trebuchet MS",Verdana,sans-serif; margin:0px}
h1 {font-size:1.2em}
h2, h3 {font-size:1.1em}
a, a:link, a:visited {color:#2A5685}
a:hover, a:active {color:#c61a1a}
fieldset.attachments {border-width:1px 0 0 0}
hr {width:100%; height:1px; background:#ccc; border:0}
span.footer {font-size:0.8em; font-style:italic}

P {margin-top:0;margin-bottom:0;}
```

Responding via web interface. I'm leaving the format as HTML and not manually pruning existing text in the reply.

---

**#6 - 2015-05-08 09:11 - Jean-Philippe Lang**
- Status changed from New to Needs feedback

**HTML Sanitizer not working for Outlook mails.**

Please attach such an email (with dummy text), so we can start working on this.

---

**#7 - 2015-05-08 16:16 - Rupesh J**
- File sample2.htm added

I saved the outlook msg file as HTML, let me know if this attachment is ok?
Thanks.

#8 - 2015-05-09 10:41 - Jean-Philippe Lang

A .eml file that contains the full email source would be better, thanks.

#9 - 2015-05-10 09:23 - Jean-Philippe Lang

- Target version changed from 3.0.3 to 3.0.4

#10 - 2015-05-11 18:03 - Deoren Moor

- File Test email.msg added

Jean-Philippe Lang wrote:

'A .eml file that contains the full email source would be better, thanks.'

From the attached msg file the following was scraped and inserted into the OP of a new issue on our Redmine v3.0.2 installation (raw text, grabbed by choosing to "edit" the OP so I could copy/paste here):

```
{{collapse(View details)

[if gte mso 9]><xml>
  <o:OfficeDocumentSettings>
    <o:AllowPNG/>
    <o:OfficeDocumentSettings/>
  </xml>
[if gte mso 9]><xml>
  <w:WordDocument>
    <w:Zoom>140</w:Zoom>
    <w:SpellingState>Clean</w:SpellingState>
    <w:TrackMoves/>
    <w:TrackFormatting/>
    <w:EnvelopeVis/>
    <w:PunctuationKerning/>
    <w:ValidateAgainstSchemas/>
    <w:SaveIfXMLInvalid>false</w:SaveIfXMLInvalid>
    <w:IgnoreMixedContent>false</w:IgnoreMixedContent>
    <w:AlwaysShowPlaceholderText>false</w:AlwaysShowPlaceholderText>
    <w:DoNotPromoteQF/>
    <w:LiDThemeOther>EN-US</w:LiDThemeOther>
    <w:LiDThemeAsian>X-NONE</w:LiDThemeAsian>
    <w:LiDThemeComplexScript>X-NONE</w:LiDThemeComplexScript>
    <w:Compatibility/>
    <w:BreakWrappedTables/>
    <w:SnapToGridInCell/>
    <w:WrapTextWithPunct/>
    <w:UseAsianBreakRules/>
    <w:DontGrowAutofit/>
    <w:SplitPgBreakAndParaMark/>
    <w:EnableOpenTypeKerning/>

2020-04-11
7/18
```
Simple, unadorned test email generated by Outlook 2010. It is in HTML format, but no special formatting has been chosen. I'm going to save this as a draft and then manually drop it into the Inbox for scraping by Redmine 3.0.2.

Thanks for your work on this!

#11 - 2015-05-19 16:43 - Deoren Moor

If I can provide any further information please let me know. Thanks.


For another datapoint, I've uploaded another .eml file (and included the original content below), pulled via IMAP right out of the mailbox: I've sanitized the company info, but it's otherwise unmodified. I believe you'll find this easier to work with than Deoren Moor's message, as that seems to have some outlook-specific content. This one is plane jane ASCII. (With HTML embedded, of course).

I had attempted to include some friendly expanding code blocks to show what arrived, but when I did so I got blocked as a spammer, so... I'm trying a bit less:

Note that the below "What I get in Redmine as source" is also what I get if I "edit" the comment.

{{collapse(What I get in Redmine as source [expand])}}
A mess.

--Geoff Maciolek

MYCOMPANYNAME, LLC
As you can see, it looks like it's not parsing the HTML comment inside HEAD as a comment, but then something else weird happens that's leading to the P {...} section showing up as well. It looks to me like this would be resolved by having everything inside <HEAD>...</HEAD> get ignored - maybe easier said than done?

#14 - 2015-06-11 20:16 - Deoren Moor

Is there any further information we can provide?
Thanks for the examples, these 2 are now fixed as part of #16962 and added as tests.

Files

<table>
<thead>
<tr>
<th>File name</th>
<th>Size</th>
<th>Date</th>
<th>Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>sample2.htm</td>
<td>129 KB</td>
<td>2015-05-08</td>
<td>Rupesh J</td>
</tr>
<tr>
<td>Test email.msg</td>
<td>51 KB</td>
<td>2015-05-11</td>
<td>Deoren Moor</td>
</tr>
</tbody>
</table>