Redmine - Feature #16962

Better handle html-only emails

2014-05-23 13:39 - Felix Schäfer

Status:	Closed		**Start date:**	
Priority:	Normal		**Due date:**	
Assignee:	Jean-Philippe Lang		**% Done:**	0%
Category:	Email receiving		**Estimated time:**	0.00 hour
Target version:	3.1.0			
Resolution:	Fixed			

Description	
Currently html-only emails are handled by just removing all html tags, this causes really munched together formatting, because html doesn't rely on newlines for formatting, and styles present in the head are also just passed over.	

Related issues:		
Related to Redmine - Feature #8462: Save received email as an attachment, not...	**New**	**2011-05-30**
Related to Redmine - Defect #12025: rake receive_pop3 is stripping formatting...	**Closed**	
Related to Redmine - Defect #18374: Incoming emails are truncating CRLF	**Closed**	
Related to Redmine - Defect #19737: HTML Sanitizer not working for Outlook mails	**Closed**	
Related to Redmine - Defect #19740: "Truncate emails after one of these lines...	**Closed**	
Related to Redmine - Defect #10201: MailHandler ignores line feeds when handl...	**Closed**	
Related to Redmine - Defect #13284: Should not try to strip HTML tags from te...	**Closed**	
Related to Redmine - Defect #6705: Incoming emails from Apple Mail not parsed...	**Closed**	**2010-10-20**
Related to Redmine - Patch #12937: mail_handler patch	**Closed**	
Related to Redmine - Defect #14668: Incoming emails has wrong line breaks	**Closed**	
Related to Redmine - Feature #13844: Option to enable HTML emails for issue c...	**Closed**	
Has duplicate Redmine - Feature #16034: Support for HTML in incoming mails	**Closed**	

Associated revisions

Revision 14313 - 2015-06-15 23:16 - Jean-Philippe Lang

Better handle html-only emails (#16962).

Revision 14542 - 2015-09-01 02:15 - Toshi MARUYAMA

Gemfile: remove loofah definition (#16962, #19419)

actionpack 4.2 depends on loofah.

History

#1 - 2014-05-23 13:49 - Felix Schäfer

- File 16962.patch added

The attached patch expands the html-only email used for testing by markup similar to what we have experienced at Planio. A growing number of users send html-only emails, especially Outlook users.

The test run with the current MailHandler results in:

```
p {font-size:12.0pt;}\r\n\r\n\r\nThis is a html-only email.With a titleand a paragraph.
```

There is content from the head and the actual email text is all squashed together.

The result with the patch applied is:

```
This is a html-only email.\r\nWith a title\r\n\r\nand a paragraph.
```

The loofah gem parses the html based on Nokogiri and tries to handle tags a little more intelligently than just removing them.

#2 - 2014-05-23 13:57 - Jan from Planio www.plan.io

- Tracker changed from Feature to Patch

- Status changed from New to Needs feedback

- Target version set to Candidate for next minor release

#3 - 2014-05-23 14:07 - Felix Schäfer

- File 16962.2.patch added

Here is an updated patch to only convert html parts to text and leave text parts alone.

#4 - 2014-05-23 14:16 - Toshi MARUYAMA

- Tracker changed from Patch to Feature

#5 - 2014-05-23 14:17 - Toshi MARUYAMA

- Has duplicate Feature #16034: Support for HTML in incoming mails added

#7 - 2014-06-13 10:49 - Joe Darkless

I think you are mixing things up. So to shed some light on this problem I'll be verbose. (if there are any misstatements feel free to correct them.) (Oh and I have redmine version 2.4.0)

There are 2 issues concerning HTML formatting from incomming mails.

The first one:

Parsing incoming mails and storing them into database

This process is made in app/models/mail_handler.rb. Around line 400 there is a method called plain_text_body, which (imho) takes care of formatting/stripping email body and is called by receive_imap method (indirectly)

This method really sux and doesn't work as it should at all. There can be 3 scenario email. Plaintext only email, HTML only email, plaintext with html both included in email.

If the plaintext email is sent, this method does its job more or less. Same with both formats included. It just grabs plaintext and its done. And It really had problem with HTML only email. And **thats** what you stated patch here is for... (though thew me an error that some method of loofa doesn't exist - but I think thats due to my version of redmine) Without the patch the HTML mail is just stored in DB as is, without any striping. Anyway it shoud do the parsing for new lines (or BRs) better.

The second one

Displaying description content on issue page

This bug is relevant mainly to CKEditor. If you have disabled CKEditor for Project the "plain-text" mail created issues are displayed corectly. But if you turn the "WYSIWYG" editor on there are no LINE BREAKS and the text is rendered on one line. Referring to issue 12025 (http://www.redmine.org/issues/12025) there is probbably no processing of plain_text to HTML to display the content properly. And this bug is still not solved by stated patch.

#8 - 2014-06-13 11:08 - Felix Schäfer

Joe, you are right that this patch only treats your stated point 1, I don't think I stated otherwise either.

Your second point is irrelevant to core Redmine currently, as (at least last time I checked) Redmine ever only stores text with markup (textile or markdown) but at no time full html. If you use a rich text editor with your Redmine or if your Redmine stores plain HTML, it comes from a plugin and that would be the concern of the plugin author, not of core Redmine. This second point is thus not relevant here and I will not engage in further discussion about it in this ticket. If you think this is relevant to Redmine, please open another ticket about it.

#9 - 2014-06-13 11:26 - Felix Schäfer

Joe Darkless wrote:

Parsing incoming mails and storing them into database

This process is made in app/models/mail_handler.rb. Around line 400 there is a method called plain_text_body, which (imho) takes care of formatting/stripping email body and is called by receive_imap method (indirectly)

MailHandler#plain_text_body is exactly the method the proposed patch changes.

This method really sux

That's a little more harsh than what I'd say, but yes.

> and doesn't work as it should at all.

No, it works as it should: provide a plain-text version of an incoming email's body. You might not agree with the method to achieve that goal, but it does achieve it.

> There can be 3 scenario email. Plaintext only email, HTML only email, plaintext with html both included in email.

> If the plaintext email is sent, this method does its job more or less. Same with both formats included. It just grabs plaintext and its done. And It really had problem with HTML only email.

If the email has at least one text-only part, then all text parts are kept. If the email has no text part but html parts, then all html parts are kept. Any html part then is stripped of all tags before being returned source:app/models/mail_handler.rb#L431, but this tag-stripping is very basic and lacking.

> And **thats** what you stated patch here is for...

Yes. The patch changes the handling of said html parts and instead of just blindly removing tags and tries to make some more sense of the html markup (still at a very basic level though).

I don't know what in the title or description made you think otherwise.

> (though thew me an error that some method of loofa doesn't exist - but I think thats due to my version of redmine)

It shouldn't. Did you run bundle install after applying the patch and restarted your Redmine instance before trying it out? If this doesn't help, could you please provide me with the error you got?

> Without the patch the HTML mail is just stored in DB as is, without any striping.

No. Vanilla Redmine doesn't store plain HTML anywhere in the DB, if that is the case for you this is caused by a plugin.

> Anyway it shoud do the parsing for new lines (or BRs) better.

That is exactly what this patch addresses: currently the MailHandler will just drop for example <br/>, which means this

```
This is the first line.<br/>This is the second line.
```

will currently just become

```
This is the first line.This is the second line.
```

whereas with the patch it would become

```
This is the first line.
```
```
This is the second line.
```

I hope I was able to clear up any misunderstanding you had about this ticket or the attached patch.

#10 - 2014-06-13 12:24 - Joe Darkless

Hello,
thanks for the fast response. I admit that I was a bit harsh, because as someone who learnt ruby during last 3 days by just diggin in the code I was upset. So sorry for that, I didn't mean it.

The reason that I have included reference to the second problem (which is connected to external plugin) was because to fix problem you need to fix both of the issues, and I think that also other shoud know that there is a problem. And I definitely understad that this has nothing to do with redmine core. So as a quickfix for the second stated issue I'm going to redirect others here: https://github.com/a-ono/redmine_ckeditor/issues/89

Ok, and concerning the first issue, which is related to redmine core:
Yes I have run bundle install respectively (rvm 2.0.0 exec bundle install) And it finished sucessfuly with loofa gem installed.
Yes I have restarted httpd server (with passenger)
I have posted the error on pastebin http://pastebin.com/Cp3fCMhM (not sure about pasting politic so please feel free to post it here or include as a file)
I really wasn't able to troubleshoot this error with my 3day knowledge of ruby and didn't find relevant issue on google.

#11 - 2014-06-13 12:39 - Felix Schäfer

Joe Darkless wrote:

> thanks for the fast response. I admit that I was a bit harsh, because as someone who learnt ruby during last 3 days by just diggin in the code I was upset. So sorry for that, I didn't mean it.

The Redmine code has a lot of baggage and really shows its age in some parts, so it's probably not the best codebase to get started with Ruby and Rails (sadly). I'd say kudos for keeping at it for 3 days :-)

> Ok, and concerning the first issue, which is related to redmine core:
> Yes I have run bundle install respectively (rvm 2.0.0 exec bundle install) And it finished sucessfuly with loofa gem installed.
> Yes I have restarted httpd server (with passenger)
> I have posted the error on pastebin http://pastebin.com/Cp3fCMhM (not sure about pasting politic so please feel free to post it here or include as a file)
> I really wasn't able to troubleshoot this error with my 3day knowledge of ruby and didn't find relevant issue on google.

It seems I had developed the patch with Loofah 1.x and that since Loofah 2.0 has been released, which requires a little more legwork. Could you try adding

```
require "loofah/helpers"
```

at the top of the patched mail_handler.rb? I don't have the time to test it right now, but that should do it.

#12 - 2014-06-13 13:20 - Joe Darkless

Felix Schäfer wrote:

> It seems I had developed the patch with Loofah 1.x and that since Loofah 2.0 has been released, which requires a little more legwork. Could you try adding
>
> [...]
>
> at the top of the patched mail_handler.rb? I don't have the time to test it right now, but that should do it.

Yes that fixed the problem. Thank you very much.

#13 - 2014-06-16 08:59 - Joe Darkless

Excuse me for inconvenience but I don't know how to link issue 'Related to' another record.

This Issue is somehow related to Feature http://www.redmine.org/issues/8462 and in my opinion it is good to know.

#14 - 2014-07-13 03:08 - Toshi MARUYAMA

- Related to Feature #8462: Save received email as an attachment, not as a comment added

#15 - 2014-07-25 04:34 - Yehuda Katz

Patch 2 works for me (although I had to apply it manually, it did not patch cleanly). I also had to add the require "loofah/helpers" line.

#16 - 2014-08-10 02:11 - Toshi MARUYAMA

- Related to Defect #12025: rake receive_pop3 is stripping formatting and new lines added

#17 - 2014-11-29 11:35 - Felix Schäfer

Just a quick note that Rails starting with 4.2.0 will use Loofah internally http://edgeguides.rubyonrails.org/4_2_release_notes.html#html-sanitizer So this will only be an additional dependency until then.

#18 - 2014-12-18 04:31 - Toshi MARUYAMA

- Related to Defect #18374: Incoming emails are truncating CRLF added

#19 - 2014-12-23 03:53 - Krishna Gollamudi

Felix Schäfer wrote:

> Just a quick note that Rails starting with 4.2.0 will use Loofah internally http://edgeguides.rubyonrails.org/4_2_release_notes.html#html-sanitizer
> So this will only be an additional dependency until then.

After applying this patch and reverting the changes, I am getting Gem not found errorrs. Can you please help me? Here below is the complete log.
.0.21/helper-scripts/passenger-spawn-server   102

#20 - 2014-12-23 07:53 - Felix Schäfer

Make backups before applying patches.

Other than that, I think that running bundle install in your Redmine directory and restarting apache should do the trick.

#21 - 2014-12-23 08:42 - Krishna Gollamudi

Felix Schäfer wrote:

> Make backups before applying patches.
>
> Other than that, I think that running bundle install in your Redmine directory and restarting apache should do the trick.

I did, but still having the issues. The Gem is installed but somehow it is not being detected. Here is my bundle env. Do you find anything missing?

#22 - 2014-12-23 08:46 - Felix Schäfer

Check if you can run the built-in rails console or server for Redmine. If it works, the problem is with Passenger or your Passenger configuration.

#23 - 2014-12-23 08:47 - Felix Schäfer

(and this is the wrong place to get help with your installation, as it has nothing to do with the original ticket. For further help I suggest heading to the forums.)

#24 - 2014-12-27 07:47 - Krishna Gollamudi

Felix:

I am able to identify the root cause and fix it. Has to do with the ruby and gems path but my emails are still loosing the HTML tags while loading into Redmine. Does your above patch works for Redmine 2.5.2 ?

#25 - 2014-12-27 12:38 - Jean-Philippe Lang

- Target version changed from Candidate for next minor release to Candidate for next major release

With the patch applied, new lines are inserted as expected but this doesn't solve 2. of #13209: new lines in incoming html should be removed, eg.

```
<p>This is
a paragraph</p>
```

should be converted to a single line: "This is a paragraph".

#26 - 2014-12-27 12:42 - Jean-Philippe Lang

A simple fix could be gsub(/\s+/, ' ') the html before processing it with Loofah.

#27 - 2014-12-27 15:42 - Krishna Gollamudi

Jean-Philippe Lang wrote:

> With the patch applied, new lines are inserted as expected but this doesn't solve 2. of #13209: new lines in incoming html should be removed, eg.
>
> [...]
>
> should be converted to a single line: "This is a paragraph".

Jean, Are you able to retain tables coming from the incoming emails?

#28 - 2014-12-27 15:59 - Felix Schäfer

Krishna Gollamudi wrote:

> I am able to identify the root cause and fix it. Has to do with the ruby and gems path but my emails are still loosing the HTML tags while loading into Redmine. Does your above patch works for Redmine 2.5.2 ?

That's the point: before the patch, if the incoming email has no text-part, the html-part is taken, but with all HTML tags removed. With the patch, in the case of no text-part, the html-part is taken, still with all HTML tags removed, but with certain tags replaced with newlines to keep paragraphs and alike.

#29 - 2014-12-27 16:23 - Felix Schäfer

Jean-Philippe Lang wrote:

> With the patch applied, new lines are inserted as expected but this doesn't solve 2. of #13209: new lines in incoming html should be removed, eg.
>
> [...]
>
> should be converted to a single line: "This is a paragraph".

I had not considered this case. I'll see if there is an option in loofah so that a wholesale replacing of all whitespaces in the whole document is not needed.

#30 - 2014-12-27 16:25 - Felix Schäfer

Krishna Gollamudi wrote:

> Jean, Are you able to retain tables coming from the incoming emails?

If you mean html tables: no, this method transforms the html-part of a text-part-less email to a somewhat equivalent text, but does not know about textile or markdown markup that might be used to render that text.

#31 - 2014-12-27 18:51 - Krishna Gollamudi

- File Snip20141227_1.png added

Felix Schäfer wrote:

> Krishna Gollamudi wrote:
>
> > Jean, Are you able to retain tables coming from the incoming emails?
>
> If you mean html tables: no, this method transforms the html-part of a text-part-less email to a somewhat equivalent text, but does not know about textile or markdown markup that might be used to render that text.

Felix, I mean if the incoming email has data like the attached, then the issue is getting created just with the text in separate line.

#32 - 2014-12-29 11:06 - Felix Schäfer

Krishna Gollamudi wrote:

> Felix, I mean if the incoming email has data like the attached, then the issue is getting created just with the text in separate line.

I guess that would be the case, but I'd have to test that. Could you either attach to this issue or send me (my email address should be on my user page) if you'd rather not have it accessible here the source of an email (in .eml format) with such a table so that I can test it?

#33 - 2014-12-29 17:39 - Krishna Gollamudi

- File Test Email.msg added

Felix Schäfer wrote:

> Krishna Gollamudi wrote:
>
> > Felix, I mean if the incoming email has data like the attached, then the issue is getting created just with the text in separate line.
>
> I guess that would be the case, but I'd have to test that. Could you either attach to this issue or send me (my email address should be on my user page) if you'd rather not have it accessible here the source of an email (in .eml format) with such a table so that I can test it?

No problem Felix. Here is the email in .msg format straight from my outlook account.

#34 - 2015-01-05 23:49 - Krishna Gollamudi

Got any update Felix??

#35 - 2015-01-09 14:31 - Felix Schäfer

Krishna, I had first to transform your .msg file to .eml format, I'm not sure how good the converters are, but I was able to see the same content as is visible in your screenshot.

The result from the .eml file is:

```
\n\n \n\nTeam project: \n\nTe\n\nArea: \n\n\\\\Te\\\\OH 30\n\nIteration: \n\n\\\\Te\\\\OH 30\\\\Iteration 1\n\
nAssigned to: \n\nKP\n\nState: \n\nNew\n\nReason: \n\nNew\n\nChanged by: \n\nGo\n\nChanged date: \n\n12/19/201
4 3:21:02 PM\n\nChanged fields \n\nField \n\nNew Value \n\nHistory\n\nPer triage QA confirmed this is happenin
g when user trying to see history.\n\n \n\nField \n\nNew Value \n\nOld Value \n\nTriage\n\nNeed More Info – De
v\n\nNeed More Info – QA\n\nAssigned To\n\nKP\n\nGo\n\n \n\n
```

which basically means that every table cell gets thrown on its own line. The first few lines would look like:

START

Team project:

Te

Area:

\\Te\\OH 30

Iteration:

\\Te\\OH 30\\Iteration 1

Assigned to:

KP
END

With the current method that would be like:

START
 Team project: TeArea: \\Te\\OH 30Iteration: \\Te\\OH 30\\Iteration 1Assigned to: KP
END

Please note that the converted .eml had a text part, so the patch proposed here would not change how this email is imported to Redmine, as the text part has precedence over the html part in any case.

#36 - 2015-01-09 14:40 - Felix Schäfer

Jean-Philippe Lang wrote:

> A simple fix could be gsub(/\s+/, ' ') the html before processing it with Loofah.

I though about it a bit more and that would also mangle whitespace in places where it is significant in html, for example in < pre> blocks. For whitespace that wouldn't be a problem as it would be just text in textile, where any number of whitespace will just get collapse in the rendered html afterwards, but < pre>foo\nbar\nbaz would then be changed to < pre>foo bar baz and converted to \nfoo bar baz\n instead of \nfoo\nbar\nbaz\n.

I have looked for a nicer solution with loofah or nokogiri but couldn't find one yet. I have asked on the loofah tracker for advice on this matter.

#37 - 2015-04-29 15:59 - Toshi MARUYAMA

- Related to Defect #19737: HTML Sanitizer not working for Outlook mails added

#38 - 2015-04-29 17:10 - Toshi MARUYAMA

- Related to Defect #19740: "Truncate emails after one of these lines" setting is not working added

#39 - 2015-06-15 23:19 - Jean-Philippe Lang

- Status changed from Needs feedback to Closed

- Assignee set to Jean-Philippe Lang

- Target version changed from Candidate for next major release to 3.1.0

- Resolution set to Fixed

Loofah is now used for processing HTML-only emails (r14313). HTML processing was moved from the MailHandler to the text formatter so we can convert it to wiki syntax.

#40 - 2015-06-16 01:54 - Deoren Moor

Jean-Philippe Lang wrote:

> Loofah is now used for processing HTML-only emails (r14313). HTML processing was moved from the MailHandler to the text formatter so we can convert it to wiki syntax.

Should the changes have also solved processing issues with Apple Mail emails (#15716) from Mac OS X and iOS?

#41 - 2019-02-09 08:12 - Go MAEDA

- Related to Defect #10201: MailHandler ignores line feeds when handling html-only emails added

#42 - 2019-02-09 10:08 - Go MAEDA

- Related to Defect #13284: Should not try to strip HTML tags from text/plain MIME part added

#43 - 2019-02-10 05:15 - Go MAEDA

- Related to Defect #6705: Incoming emails from Apple Mail not parsed correctly added

#44 - 2019-02-10 05:41 - Go MAEDA

- Related to Patch #12937: mail_handler patch added

#45 - 2019-02-10 06:19 - Go MAEDA

- Related to Defect #14668: Incoming emails has wrong line breaks added

#46 - 2019-02-19 03:10 - Go MAEDA

- Related to Feature #13844: Option to enable HTML emails for issue creation. added

Files

16962.patch	2.21 KB	2014-05-23	Felix Schäfer
16962.2.patch	2.59 KB	2014-05-23	Felix Schäfer
Snip20141227_1.png	28.6 KB	2014-12-27	Krishna Gollamudi
Test Email.msg	21 KB	2014-12-29	Krishna Gollamudi