

Export Wiki to ODT

2022/05/09 17:58 - Admin Redmine

ステータス:	New	開始日:	2022/05/09
優先度:	通常	期日:	
担当者:		進捗率:	90%
カテゴリ:	Wiki_1	予定工数:	0.00時間
対象バージョン:		作業時間:	0.00時間
Redmineorg_URL:	https://www.redmine.org/issues/22923	status_id:	1
category_id:	1	tracker_id:	2
version_id:	0	plus1:	0
issue_org_id:	22923	affected_version:	
author_id:	14446	closed_on:	
assigned_to_id:	0	affected_version_id:	
comments:	19		

説明

Attached you may find a patch, which adds Wiki to ODT export capabilities to Redmine.

The export works similar to the PDF export in that it mainly passes the generated HTML export to another library. In case of ODT, this is the @html2odt@ gem, which in turn is based on "xhtml2odt": <https://github.com/abompard/xhtml2odt>, which uses XSLT to transform HTML to OpenDocument compatible XML.

The Redmine integration mainly consists of code to handle image paths and a bit of clean up before passing the HTML to the library.

This change was implemented for a Planio customer, who wanted to use the Wiki to create simple templates, that users should then fill out using MS Word (and friends). We chose to export ODT (instead of DOCX e.g.) since the OpenDocument format is an open standard, the tool support was better and since it is supported by a wide range of word processing applications (MS Word 2010 and later, LibreOffice, OpenOffice, AbiWord, Pages).

Please note, that this feature benefits from the change proposed in #22898. Otherwise aligned images within a paragraph cause errors in the export, i.e. the following paragraph is missing from the ODT.

journals

I tested this patch with LibreOffice Mac and Word 2016 for Mac, works fine. I think it is very useful to write documents based on wiki contents.

@Jan-Philippe, could this be included in 3.3.0?

I tested the patch on a copy of the Redmine wiki:

- exporting the whole wiki doesn't respond/is too slow (had to kill ruby after 15 minutes), we should probably disable this option. In comparison, exporting every pages individually took about 2 minutes.
- some pages generate documents that seem invalid (at least for Word 2007 that complains about an unspecified error when opening the generated odt)
- some pages that contain URLs trigger this kind of error (seems to be caused by examples URLs that are not valid URLs):

```
ActionView::Template::Error (the scheme http does not accept registry part: proxy.domain.tld:port (or bad hostname?)):
  1: <%= raw wiki_page_to_odt(@page, @project) %>
```

```
c:/utils/ruby/lib/ruby/2.0.0/uri/generic.rb:1203:in `rescue in merge'
c:/utils/ruby/lib/ruby/2.0.0/uri/generic.rb:1200:in `merge'
html2odt (0.3.0) lib/html2odt/document.rb:270:in `block in fix_links'
nokogiri-1.6.7.2-x86 (mingw32) lib/nokogiri/xml/node_set.rb:187:in `block in each'
nokogiri-1.6.7.2-x86 (mingw32) lib/nokogiri/xml/node_set.rb:186:in `upto'
nokogiri-1.6.7.2-x86 (mingw32) lib/nokogiri/xml/node_set.rb:186:in `each'
html2odt (0.3.0) lib/html2odt/document.rb:269:in `fix_links'
html2odt (0.3.0) lib/html2odt/document.rb:209:in `prepare_html'
html2odt (0.3.0) lib/html2odt/document.rb:55:in `content_xml'
html2odt (0.3.0) lib/html2odt/document.rb:142:in `block (4 levels) in data'
rubyzip (1.2.0) lib/zip/entry.rb:495:in `get_input_stream'
html2odt (0.3.0) lib/html2odt/document.rb:139:in `block (3 levels) in data'
```

I attached a raw wiki page content. When formatting is set to Markdown, you should get the error, when formatting is set to textile, you should get the invalid odt (which is also attached).

I tested the patch on a copy of the Redmine wiki:

Thank you for testing it. The long response times seem quite odd. It works quite fast (also with larger wikis) for us. For comparison, you could try on a "new Planio account": <https://accounts.planio/signup/Bronze>, we already have this running in production.

We hadn't tested it on Windows though which it seems you have used for your tests, correct?

We will look into it and try to reproduce/fix these problems. Thanks again!

Jan from Planio www.planio wrote:

We hadn't tested it on Windows though which it seems you have used for your tests, correct?

Correct, ruby 2.0.0p481 (2014-05-08) [i386-mingw32]

Thanks a lot for testing this patch/feature.

I have just set up a Windows development machine to verify your feedback:

- exporting the whole wiki doesn't respond/is too slow (had to kill ruby after 15 minutes), we should probably disable this option. In comparison, exporting every pages individually took about 2 minutes.

This is very surprising. On my newly set up development virtual machine (on 5 year old Macbook) running Windows 7 and Ruby 2.0 (using Ruby Installer), the example wiki page you provided was generated within 0.4 seconds. This also matches my experience on Mac OS and Linux.

- some pages generate documents that seem invalid (at least for Word 2007 that complains about an unspecified error when opening the generated odt)

Unfortunately this is true. @xhtml2odt@ claims to handle only valid XHTML and we have already encountered cases where even valid HTML was not handled in the expected way. #22898 was one such case, the given example document is another. Attached you may find an updated one. All I did, was added new lines around the @pre@ tags, so that they are not entangled with the previous paragraph. But since we cannot expect the user to write perfectly formatted content, I will have a look at how I can fix the problem at hand in any case. I just wanted to let you know about the underlying reason.

- some pages that contain URLs trigger this kind of error (seems to be caused by examples URLs that are not valid URLs)

Thank you for this bug report. Indeed, handling of invalid URIs was missing in @html2odt@. The latest release (0.3.1) fixes that. Running @bundle update html2odt@ should add it to your installation.

I will have a look, at how I can handle the markup generated on your example content.

Could you try to narrow down, why generating the ODT is so slow on your machine? Should we gather more feedback by other developers? What do you think?

We have just released html2odt v0.3.3, which addresses all problems, that came up during your tests.

The only problem remaining would be the speed issues, you saw. But I am afraid, that I cannot isolate and reproduce those without further input.

Gregor Schmidt wrote:

This is very surprising. On my newly set up development virtual machine (on 5 year old Macbook) running Windows 7 and Ruby 2.0 (using Ruby Installer), the example wiki page you provided was generated within 0.4 seconds. This also matches my experience on Mac OS and Linux.

Yes, I had good response times when exporting single wiki pages, as I mentioned before. The problem was when exporting the whole wiki in one ODT file (by using the export link on the wiki page index). This was not a Windows issue as I get the same behaviour when testing under linux (I let it run for more than one hour before killing webrick).

This problem no longer occurs with the same wiki content and html2odt 0.3.3 so I guess it was a html2odt or nokogiri issue (html2odt 0.3.3 uses a different nokogiri version). Now I get the whole wiki export in a minute (with both windows and linux) but the resulting ODT seems to be invalid. You'll find it attached.

This feature seems to be usefull for a very few users and I prefer not to add it to the core. But I'd be happy to refactor a few things in order to make it easier for plugins to add new export formats, without having to patch views and controllers.

Thanks again for taking the time to review and test the changes and for giving such detailed feedback. This is very much appreciated.

Jean-Philippe Lang wrote:

This problem no longer occurs with the same wiki content and html2odt 0.3.3 so I guess it was a html2odt or nokogiri issue (html2odt 0.3.3 uses a different nokogiri version). Now I get the whole wiki export in a minute (with both windows and linux) but the resulting ODT seems to be invalid. You'll find it attached.

Thanks for the feedback. I am not aware of speed related improvements in html2odt - at least I was not aiming for them. But I am glad, that it's working better now.

Thanks for providing the ODT. I'll have a look and see if I can isolate the root cause of the error.

This feature seems to be usefull for a very few users and I prefer not to add it to the core.

I am sorry to hear that, but I can totally understand your decision.

But I'd be happy to refactor a few things in order to make it easier for plugins to add new export formats, without having to patch views and controllers.

That would be great. We would be happy to create a plugin with the same features. If Redmine had some kind of export registry, we could easily hook into that.

We are looking forward to those changes. Let me know if you would like to us to help with those refactorings.

Thanks again for providing the erroneous ODT. Now that I see, that it is more than 800 pages long, I can imagine, why it's taking a minute to generate. I assume, the PDF export would not be a lot faster.

Concerning the ODT error, I was able to identify the root cause. It is related to the HTML generated by the collapse macro. The error was located in the export of "this wiki page":/projects/redmine/wiki/InstallRedmineOnDebianStableApacheMysqlPassenger.

Attached you may find an updated patch[1], which further cleans up the HTML before handing it over to @html2odt@. I am leaving

this here for future reference.

fn1. replacing the first one

Previous versions of the patch contained a path traversal vulnerability, which allowed attackers to access image files outside of `@Rails.public_path@`. Attached you may find an updated patch, including the fix.

Wiki pages are converted without images. Runing on Debian stable, ruby 2.1.5, redmine 3.1.1

I install patch in my Redmine installation. In case of images inside wiki page, in the ODT file the images are missing.

related_issues

duplicates,New,16324,Wiki export as docx file.

blocks,Closed,22898,!>image.png! generates invalid HTML

履歴

#1 - 2022/05/10 17:07 - Admin Redmine

- カテゴリ を Wiki_1 にセット