QGIS Application - Bug report #11287
composer' html items do not respect the character encoding of HTML pages loaded via URL

2014-09-28 04:27 AM - Mathieu Pellerin - nIRV

Status:	Closed		
Priority:	High		
Assignee:	Nyall Dawson		
Category:	Map Composer/Printing		
Affected QGIS version:	master	**Regression?:**	No
Operating System:		**Easy fix?:**	No
Pull Request or Patch supplied:		**Resolution:**	
Crashes QGIS or corrupts data:	No	**Copied to github as #:**	19583

Description

The composer' html items do not respect the character encoding declared within HTML pages when loaded via entering a URL.

Steps to reproduce:

1. Create a new composer, and add an html item onto it
2. In the html item's URL text box, enter "http://www.licadho-cambodia.org/?khmer"
3. Hit Refresh HTML
4. When the page is drawn, you'll notice a lot of ÁÁÁÁÁÁÁ characters, indicating that the character encoding used to render the page is set to something other than UTF-8 (the declared encoding of the loaded page)

Load the above-mentioned page in your browser to see how it should normally look with proper encoding. I'm attaching a screenshot which should also help.

Whatever underlying QT web tech the html item is using, if it can't detect a given page's character encoding, it surely allows for a default character encoding to be set, which should be UTF-8 to insure greatest compatibility with what's out there.

It's a pretty significant issue when it comes to making sure QGIS works for everyone around the globe.

Setting the priority to high as I am not sure if this regressed, or if it was always like that.

Associated revisions

Revision 18412257 - 2014-09-30 02:27 PM - Nyall Dawson

[composer] Correctly handle encoded HTML source (fix #11287)

History

#1 - 2014-09-28 05:38 AM - Mathieu Pellerin - nIRV

- File html_item-broken_arabic.png added

As to express the wide-ranging impact of this issue, I'm attaching a screenshot showing broken character encoding of an Arabic wikipedia page.

#2 - 2014-09-30 05:27 AM - Nyall Dawson

- Status changed from Open to Closed

Fixed in changeset commit:"184122571ac1d310de8c8906e512788ebbbb29a9".

Files

html_item-wrong_character_encoding.png	295 KB	2014-09-28	Mathieu Pellerin - nIRV
html_item-broken_arabic.png	232 KB	2014-09-28	Mathieu Pellerin - nIRV