Metacat - Bug #2797

Character set (charset) problem when filling out form

03/12/2007 03:07 PM - Callie Bowdish

Status:	Resolved		**Start date:**	03/12/2007
Priority:	Normal		**Due date:**	
Assignee:	Saurabh Garg		**% Done:**	0%
Category:	metacat		**Estimated time:**	0.00 hour
Target version:	1.8.1		**Spent time:**	0.00 hour
Bugzilla-Id:	2797			

Description

The ESA data repository form has a problem whem people enter a number of different characters. It might look right when they type it in but it does not save and then display correctly. Decimal points, foreign letters such as Ã²,Ã¼, quotes, apostrophes, and greater than symbols can all cause problems. There is not an easy way to fix this for the ESA moderator. When people fill out forms they will cut and past from different character sets. I think some scripting in Perl was suggested as being used to help with this problem.

The current fix for this is to open the file in Morpho and use the Windows Character Map to put in the foreign language characters and others such as degrees that are displaying incorrectly. Quotes, and aposrophes can usually be fixed by just retyping them. The greater than and lesser than symbols are listed as another bug (http://bugzilla.ecoinformatics.org/show_bug.cgi?id=2517). I've been typing out the phrase "greater than" because I have not found any solution for problems with displaying them correctly.

History

#1 - 03/28/2007 02:55 PM - Callie Bowdish

knb-lter-luq.118
Here is a data set that shows the character set (charset) problem with text not displaying correctly. I am concerned that some important information will get lost. It will be hard in some cases to know what the text entry was. I do not know how this LTER entry was made. It may not of been through a form.

#2 - 12/17/2007 03:04 PM - Callie Bowdish

Looks like the forward slash needs to be added to the "special characters" that are problematic with viewing online.

http://knb.ecoinformatics.org/knb/metacat/nceas.347.1/nceas

This data package uses the forward slash and displays it as grassland Ã¢ï¿½â€ž meadow instead of grassland/meadow and RRX = ln (EÃ¢ï¿½â€ž C) instead of RRX = ln (E/C)

The ones below have given some trouble in the past.

There are 5 predefined entity references in XML:
<   <   less than
>   >   greater than
&   &   ampersand
'   '   apostrophe
"   "   quotation mark

I wonder why we can't wrap the contents entered in the forms text boxes around the CDATA element. I think this might help with some issues but others such as Windows proprietary character sets could still be a problem

http://www.w3schools.com/xml/xml_cdata.asp

#3 - 03/07/2008 09:17 AM - Callie Bowdish

Jing has fixed this and it has been tested on the Dev. test server using both Morpho and the Form.

Here is Jing's comment on a similar bug entry (2517)
------ Comment #2 from tao@nceas.ucsb.edu 2008-03-06 17:32 ------
This bug has been fixed (base on callie's testing). Here are somethings i did:
1)Use ServletOutputSteam to replace PrintWriter to send response of read and query actions. So Java wouldn't encode the special characters.
2)Fix a bug in normalize method in MetacatUtil class - it will encode special characters.
3)Fix a similar bug in normalize subroute in register-dataset.cgi.

#4 - 05/14/2008 11:36 AM - Jing Tao

Change the milestone to 1.8.1

#5 - 05/27/2008 11:04 AM - Callie Bowdish

Generally this is working much better. However, today I tried pasting special characters and quotes using from Word and I had some strange things happen when opening and saving. First I filled out the form and cut and pasted special characters in the the Sampling Description section of
http://dev.nceas.ucsb.edu/knb/metacat?action=read&qformat=nceas&docid=nrs.853.1

When I opened it in Morpho The special symbols and the quotes did not display correctly in Morpho. Then I added some special characters and quotes using Morpho (cut and pasted form Word) in the Abstract section.

http://dev.nceas.ucsb.edu/knb/metacat?action=read&qformat=nceas&docid=nrs.853.2

Online the Abstract section looked good but online the Sampling Description no longer displayed correctly.

I again opened the data package from the network on Morpho and the Abstract section no longer displayed correctly. I didn't try saving it to the network afterwards.

#6 - 05/27/2008 11:15 AM - Jing Tao

Callie. Yeah, it is known that the morpho has problem to handle special character. This is not metacat bug. It is morpho one. I will close the bug and open one in morpho.

#7 - 03/27/2013 02:21 PM - Redmine Admin

Original Bugzilla ID was 2797