

## Morpho - Bug #1536

### Special character problems

04/27/2004 08:02 AM - Dan Higgins

<b>Status:</b>	Resolved	<b>Start date:</b>	04/27/2004
<b>Priority:</b>	Normal	<b>Due date:</b>	
<b>Assignee:</b>	Saurabh Garg	<b>% Done:</b>	0%
<b>Category:</b>	morpho - general	<b>Estimated time:</b>	0.00 hour
<b>Target version:</b>	1.5.1	<b>Spent time:</b>	0.00 hour
<b>Bugzilla-Id:</b>	1536		

#### Description

Users may import 'special characters' such as the degree symbol or the micro symbol (Greek mu) into data packages, especially by copying from MS word or other sources which use custom character sets. These special characters can cause various problem when stored in XML documents because they are not unicode and because one cannot determine just what characted set they originated in.

When pasting in degree and mu symbols on Windows, the display looks fine until the document is saved locally, If Morpho is then exited and re-opened, the document with the special characters does NOT appear in the list of local documents. The problem is:

```
java.io.UTFDataFormatException: Invalid byte 1 of 1-byte UTF-8 sequence
```

caused by the coding of the deg symbol (and the mu symbol is missing from the XML text).

Previously a special 'normalize' method was used to handle the strings used in data packages. The problem with Morpho 1.5 is that the DOM routines used handle much of the character/entity translations automatically - e.g. '&' automatically becomes '&' and is converted back to '&' when displayed.

#### History

##### #1 - 04/28/2004 08:56 AM - Dan Higgins

This problem has been fixed within Morpho by adding a new 'getDOMTreeAsString' routine which include the 'normalize' command when a DOM is converted to a string for saving as an XML file.

##### #2 - 03/27/2013 02:17 PM - Redmine Admin

Original Bugzilla ID was 1536