Morpho - Bug #5238

Use UTF-8 for file reading and writing

11/11/2010 01:28 PM - ben leinfelder

Status:	Resolved		**Start date:**	11/11/2010
Priority:	Normal		**Due date:**	
Assignee:	ben leinfelder		**% Done:**	0%
Category:	morpho - general		**Estimated time:**	0.00 hour
Target version:	1.8.1		**Spent time:**	0.00 hour
Bugzilla-Id:	5238			

Description

Rather than rely on the "default" character encoding used on individual platforms, Morpho should explicitly read and write text files using UTF-8 character encoding. When non-determinate encodings are used across different systems, special characters (accents, tildes, umlauts, Chinese, etc..) can become garbled and misinterpreted. Using the same encoding for all Morpho reading and writing will mitigate these encoding issues.
Note: this does not address character encoding issues that arise from copy/paste actions from other systems that use non-UTF-8 encoding (i.e. Word).

Related issues:

Related to Morpho - Bug #6045: Morpho not correctly handling Chinese characters	**In Progress**	**08/05/2013**

History

#1 - 11/11/2010 01:30 PM - ben leinfelder

created a tag of the 1.8.1 code before committing the update to use UTF-8 across the board.
https://code.ecoinformatics.org/code/morpho/tags/BEFORE_UTF-8/

#2 - 11/11/2010 02:26 PM - Jim Regetz

I'll put in a vote for including an explicit encoding declaration in the EML docs that Morpho creates:

<?xml version="1.0" encoding="utf-8" ?>

#3 - 11/15/2010 02:10 PM - ben leinfelder

This works:
Morpho now using UTF-8 for all reading and writing.
Additionally, special characters are not being escaped because we can encode them with UTF-8.

This does not:
Saving to Metacat (and the subsequent read) result in ????? for characters that should be, say, Chinese. This means (as I suspected) Metacat uses the default character encoding rather than explicitly using UTF-8.

#4 - 01/17/2013 10:50 AM - ben leinfelder

This should be closed. UTF-8 is used exclusively in Morpho - especially important now that we have so much internationalization support

#5 - 03/27/2013 02:29 PM - Redmine Admin

Original Bugzilla ID was 5238