

TextCritical.net - Bug #434

Importer error when attempting to make verse

11/11/2012 08:03 PM - Luke Murphey

Status:	Closed	Start date:	
Priority:	Normal	Due date:	
Assignee:	Luke Murphey	% Done:	100%
Category:		Estimated time:	0.00 hour
Target version:	0.1		
Description			
The Perseus importer generates errors on some documents due to an encoding problem. See below for an example of the error:			
<pre>2012-11-09 01:21:42,533 [ERROR] reader.importer.PerseusBatchImporter: Exception generated when attempting to process file="plut.cat.ma_gk.xml"Traceback (most recent call last): File "C:\Users\Luke\Workspace\TextCritical.com\src\reader\importer\PerseusBatchImporter.py", line 269, in process_directory if self.__process_file__(os.path.join(root, f)): File "C:\Users\Luke\Workspace\TextCritical.com\src\reader\importer\PerseusBatchImporter.py", line 223, in __process_file__ return self.process_file(file_path, document_xml, title, author, language) File "C:\Users\Luke\Workspace\TextCritical.com\src\reader\importer\PerseusBatchImporter.py", line 396, in process_file perseus_importer.import_file(file_path) File "C:\Users\Luke\Workspace\TextCritical.com\src\reader\importer\Perseus.py", line 121, in import_file return self.import_xml_document(doc) File "C:\Program Files\Python27\lib\site-packages\django\db\transaction.py", line 209, in inner return func(*args, **kwargs) File "C:\Users\Luke\Workspace\TextCritical.com\src\reader\importer\Perseus.py", line 557, in import_xml_document verses_created = self.make-verses(divisions, current_state_set) File "C:\Users\Luke\Workspace\TextCritical.com\src\reader\importer\Perseus.py", line 575, in make-verses verses_created = verses_created + self.make-verses_for_division(division, state_set) File "C:\Users\Luke\Workspace\TextCritical.com\src\reader\importer\Perseus.py", line 591, in make-verses_for_division division_doc = parseString(division.original_content) File "C:\Program Files\Python27\lib\xml\dom\minidom.py", line 1924, in parseString return expatbuilder.parseString(string) File "C:\Program Files\Python27\lib\xml\dom\expatbuilder.py", line 940, in parseString return builder.parseString(string) File "C:\Program Files\Python27\lib\xml\dom\expatbuilder.py", line 223, in parseString parser.Parser.parse(string, True)</pre>			

History

#1 - 11/11/2012 08:05 PM - Luke Murphey

- Description updated

#2 - 11/11/2012 08:20 PM - Luke Murphey

The import began at 11/9/12 1:21:41.348 AM.

```
[INFO] reader.importer.PerseusBatchImporter: Importing a Perseus XML file, file_path="C:\Users\Luke\Workspace\TextCritical.com\var\resources\Perseus\Classics\Plutarch\opensource\plut.cat.ma_gk.xml"
```

It looks like it failed on verse 5 of division 12 of Marcus Cato.

```
2012-11-09 01:21:42,517 [DEBUG] reader.importer.Perseus: Making verse 5 in division 12 of Marcus Cato
```

This issue seems to affect 56 works at the current time. These can be identified by running the following search:

sourcetype="django" ERROR UnicodeEncodeError | table file

#3 - 11/12/2012 06:36 AM - Luke Murphey

- Status changed from New to In Progress

#4 - 11/14/2012 02:59 AM - Luke Murphey

- Status changed from In Progress to Closed

- % Done changed from 0 to 100