

## TextCritical.net - Bug #442

### Imported works with no descriptor on first verse fail

11/17/2012 04:29 PM - Luke Murphey

<b>Status:</b>	Closed	<b>Start date:</b>	11/17/2012
<b>Priority:</b>	Normal	<b>Due date:</b>	
<b>Assignee:</b>	Luke Murphey	<b>% Done:</b>	100%
<b>Category:</b>		<b>Estimated time:</b>	0.00 hour
<b>Target version:</b>	0.1		
<b>Description</b>			
For examples, see:			
<ul style="list-style-type: none"><li>Symposium By Lucian</li></ul>			

#### Associated revisions

##### Revision 199 - 11/27/2012 07:24 AM - Luke Murphey

Added option to ignore content before the first milestone. Reference #442.

##### Revision 199 - 11/27/2012 07:24 AM - Luke Murphey

Added option to ignore content before the first milestone. Reference #442.

##### Revision 192 - 11/27/2012 07:24 AM - Luke Murphey

Added option to ignore content before the first milestone. Reference #442.

#### History

##### #1 - 11/17/2012 06:42 PM - Luke Murphey

- Subject changed from *Imported works no descriptor on first verse fail* to *Imported works with no descriptor on first verse fail*

##### #2 - 11/18/2012 03:03 PM - Luke Murphey

These can be found by looking for events with the following in them: "Milestone observed that did not have an associated unit, the sequence number will be used instead"

##### #3 - 11/21/2012 02:00 AM - Luke Murphey

Many of the works by Lucian are affected by this.

##### #4 - 11/24/2012 06:13 AM - Luke Murphey

The problem is that the text within the div1 node is getting created in a division that is at the same level of the section node after it. The div1 should be one level above:

```
<text>
  <body>
    <pb id="v.5.p.476"/>
    <div1>
      <p>
a)pokhruxqei/s tis i)atrikh\n e)ce/maqen. mane/nta to\n pate/ra
kai\ u(po\ tw=n a)llwn i)atrw=n a)pegnwsme/non i)asa/menos farma/kou
do/sei a)nelh/fqh au)=qis e)s to\ ge/nos. meta\ tau=ta memhnui/an th\n
mhtruia\n i)a/sasqai keleuo/menos <gap/> a)pokhru/ttetai.</p>
      <p>
        <milestone unit="section" n="1"/> ou) kaina\ me\n tau=ta,
```

#### #5 - 11/24/2012 06:14 AM - Luke Murphey

- Assignee set to Luke Murphey
- % Done changed from 0 to 30

#### #6 - 11/24/2012 09:48 AM - Luke Murphey

Need to rethink how levels are assigned. Below is a list of the type of divisions that exist:

- text nodes
- divisions (div0, div1)
- milestone (chapters, sections)

We could assign the nodes a static range of levels. Something like:

- text = level 0
- div = level 1-10
- milestone = level 11

The only problem with this is that the document will have gaps in the levels.

We could store a value which indicates the maximum level assigned to a div or text node.

Finally, we could pass the closest division object down to the calls to process the downstream nodes when processing the milestones.

#### #7 - 11/24/2012 11:34 PM - Luke Murphey

Here is how the level is currently defined:

- text nodes are assigned a level 0
- division nodes are assigned the level that is parsed from the node tag name (e.g. div1 is level 1)
- milestones are assigned a level from the state info where the level is assigned based on the position in the state-info node (the first is level 1, the second level 2, etc.)

Assigning levels are important for: \* Determining if division is under or above another division \* Determining if divisions in a given work are under the same division

#### #8 - 11/24/2012 11:35 PM - Luke Murphey

Works by Epictetus may be suffering from this. It seems like the divisions are awkwardly assigned.

#### #9 - 11/24/2012 11:45 PM - Luke Murphey

From what I can tell, the divisions are assigned correctly. The refs declaration seem to comport with reality:

```
<refsDecl doctype="TEI.2">
  <state unit="text"/>
  <state delim="." unit="book"/>
  <state delim="." unit="chapter"/>
```

```
<state n="chunk" unit="section"/>
</refsDecl>
```

The only real problem is with verses that are not associated with milestones.

#### #10 - 11/25/2012 03:33 AM - Luke Murphey

The following works are affected:

- Abdicatus
- Anacharsis
- Bis accusatus sive tribunalia
- Catapulus
- Contemplantes
- De morte Peregrini
- De parasito sive artem esse parasiticam
- De saltatione
- Dearum iudicium
- Deorum concilium
- Dialogi Marini
- Eunuchus
- Fugitivi
- Gallus
- Icaromenippus
- Imagines
- Juppiter confuatus
- Juppiter trageodeus
- Lexiphanes
- Necyomantia
- Nigrinus
- Philopsuedes sive incredulus
- Piscator
- Pro imaginibus
- Prometheus
- Symposium
- Timon
- Toxaris vel amicitia
- Tyrannicida
- Vitarum auctio

Also, Dialogi Marini has a weird first verse with no content.

### #11 - 11/26/2012 05:41 AM - Luke Murphey

The question is whether I handle this when chopping up the divisions or when creating the verses. However, milestone node content is not included in the division since the functions that import the verse nodes take this to be a new verse.

### #12 - 11/28/2012 04:40 PM - Luke Murphey

- % Done changed from 30 to 70

After making this change, I am getting the following error:

```
2012-11-28 02:12:11,811 [ERROR] reader.importer.PerseusBatchImporter: Exception generated when attempting to process file="07_gk.xml"
Traceback (most recent call last):
  File "/Users/lmurphey/Documents/SP/Workspace/TextCritical.com/src/reader/importer/PerseusBatchImporter.py", line 327, in process_directory
    if self.__process_file__( os.path.join( root, f ) ):
  File "/Users/lmurphey/Documents/SP/Workspace/TextCritical.com/src/reader/importer/PerseusBatchImporter.py", line 278, in __process_file__
    return self.process_file(file_path, document_xml, title, author, language)
  File "/Users/lmurphey/Documents/SP/Workspace/TextCritical.com/src/reader/importer/PerseusBatchImporter.py", line 456, in process_file
    perseus_importer.import_file(file_path)
  File "/Users/lmurphey/Documents/SP/Workspace/TextCritical.com/src/reader/importer/Perseus.py", line 144, in import_file
    return self.import_xml_document(doc)
  File "/Library/Frameworks/Python.framework/Versions/2.7/lib/python2.7/site-packages/django/db/transaction.py", line 209, in inner
    return func(*args, **kwargs)
  File "/Users/lmurphey/Documents/SP/Workspace/TextCritical.com/src/reader/importer/Perseus.py", line 700, in import_xml_document
    raise Exception("No verses were discovered, title=%s" % (self.work.title) )
Exception: No verses were discovered, title=Nigrinus
```

### #13 - 11/29/2012 04:40 PM - Luke Murphey

- Status changed from New to Closed

- % Done changed from 70 to 100

When no division id is provided in the URL, get\_chapters\_list() returns all of the