TextCritical.net - Bug #416

Perseus Importer Doesn't Handle Sectioning Correctly with Multiple Levels of Divs

10/17/2012 06:32 PM - Luke Murphey

Status:	Closed		**Start date:**	10/17/2012
Priority:	Normal		**Due date:**	
Assignee:			**% Done:**	100%
Category:			**Estimated time:**	0.00 hour
Target version:	0.1			

Description

Eumenides by Aeschylus does not currently import correctly. It seems to be that the importer is not correctly detecting cases where a div does not contain any actual content.

Consider the following snippet:

```
<div1 type="choral">
<div2 type="strophe" n="1">
<sp><speaker>*xoro/s</speaker>
<l>i)alto\s e)k do/mwn e)/ban</l>
<l>xoa\s propompo\s o)cu/xeiri su\n ktu/pw|.</l>
<l>pre/pei parhi\s foini/ois a)mugmoi=s</l>
<l n="25">o)/nuxos a)/loki neoto/mw|:</l>
<l>di' ai)w=nos d' i)ugmoi=si  bo/sketai ke/ar.</l>
<l>linofqo/roi d' u(fasma/twn</l>
<l>laki/des e)/fladon u(p' a)/lgesin,</l>
<l>proste/rnw| stolmw=|</l>
<l n="30">pe/plwn a)gela/stois</l>
<l>cumforai=s peplhgme/nwn.</l>
<milestone ed="p" n="32" unit="card"/></sp></div2>
```

This correctly produces two divisions. The first contains the following content:

```
<div1 type="choral">
</div1>
```

The second contains the actual content:

```
<?xml version="1.0" ?><chapter><div2 n="1" type="strophe">
<sp><speaker>*xoro/s</speaker>
<l>i)alto\s e)k do/mwn e)/ban</l>
...
```

The problem is that a verse is created out of the first division since it contains content. However, this verse is empty since no actual verse content is included.

Related issues:

Related to TextCritical.net - Feature #403: Perseus Book Importer	**Closed**

History

#1 - 10/17/2012 07:26 PM - Luke Murphey

- Subject changed from Perseus Importer Doesn't handle sectioning correctly to Perseus Importer Doesn't Handle Sectioning Correctly with Multiple

Levels of Divs

#2 - 10/17/2012 09:55 PM - Luke Murphey

- Status changed from New to In Progress

#3 - 10/17/2012 09:59 PM - Luke Murphey

The problem is that we cannot determine if the verse is going to have content until we begin processing it. We may want to have two references to verses and only save the verse until we get actual content.

#4 - 10/17/2012 10:15 PM - Luke Murphey

We could delay creation of a verse instance until we get actual verse content (like a milestone or text data). The problem with this is that it is possible to have texts that have divisions but no milestones.

#5 - 10/18/2012 06:48 AM - Luke Murphey

I currently build up the XML content by attaching it to an XML document. I am testing changes such that where the XML document is appended to even if no verse has yet been created. The content will be included in the verse once it is created.

This way, I don't have to filter out nodes that could trigger creation of a verse until I figure out that a verse should be created.

I may need to create a verse if one does not exist when I see nodes that are clearly indicative of verse content to work with texts that do not have verse milestones.

#6 - 10/18/2012 07:16 AM - Luke Murphey

- Target version set to 0.1

#7 - 10/19/2012 06:50 AM - Luke Murphey

This work has the milestone indicators at the end of the prior division:

```
        <milestone ed="p" n="143" unit="card"/>
    </sp>
</div1>
<div1 type="choral">
    <div2 type="strophe" n="1">
        <sp><speaker>*xoro/s</speaker>
```


I would have expected the milestone to be at the beginning of the following chapter.

#8 - 10/20/2012 03:46 AM - Luke Murphey

- Status changed from In Progress to Closed

- % Done changed from 0 to 100