TextCritical.net - Bug #420

Perseus Importer Gets Nodes Out of Order

10/19/2012 07:17 AM - Luke Murphey

Status:	Closed	**Start date:**	
Priority:	Urgent	**Due date:**	
Assignee:		**% Done:**	100%
Category:		**Estimated time:**	0.00 hour
Target version:	0.1		

Description

The Perseus importer is getting some nodes out of order and places the text node just outside of the node where is should be. The speaker in the text below is Πυθιας but the name is after the speaker node.

```
<span id="verse_" class="verse_container ">
<!--?xml version="1.0" encoding="utf-8"?-->
<span class="verse">
<span class="milestone" data-ed="p" data-n="1" data-unit="card"></span>
<span class="sp" data-n="*puqia/s">
<span class="speaker"></span></span>Πυθιάς
<span class="l">πρῶτον μὲν εὐχῇ τῇδε πρεσβεύω θεῶν</span>
<span class="l">τὴν πρωτόμαντιν Γαῖαν: ἐκ δὲ τῆς Θέμιν,</span>
```

Related issues:

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

Associated revisions

Revision 83 - 10/20/2012 03:32 AM - Luke Murphey

Changed the way that the XML element appender works such that it:

```
1) Concatenates the contents of a the text nodes if the src and dst nodes are both text nodes
 2) Shims in the node to be copied at the parent of the a text node if the node to be copied is not a text nod
e but the dsetination is
```

This is part of that mitigation that closes #420

Revision 78 - 10/20/2012 03:32 AM - Luke Murphey

Changed the way that the XML element appender works such that it:

```
1) Concatenates the contents of a the text nodes if the src and dst nodes are both text nodes
 2) Shims in the node to be copied at the parent of the a text node if the node to be copied is not a text nod
e but the dsetination is
```

This is part of that mitigation that closes #420

Revision 84 - 10/20/2012 03:41 AM - Luke Murphey

Card milestones are now considered chunks.

Added the option (ignore_division_markers) that causes the Perseus importer to ignore division markers.

Removed unnecessary log message indicating that no verse content existed to save.

Fixed problem when the first text node was placed in the wrong location for the first verse. This closes #420.

Division markers are now included in the resulting original content.

Revision 79 - 10/20/2012 03:41 AM - Luke Murphey

Card milestones are now considered chunks.

Added the option (ignore_division_markers) that causes the Perseus importer to ignore division markers.

Removed unnecessary log message indicating that no verse content existed to save.

Fixed problem when the first text node was placed in the wrong location for the first verse. This closes #420.

Division markers are now included in the resulting original content.

History

#1 - 10/19/2012 10:42 PM - Luke Murphey

- Priority changed from Normal to Urgent

#2 - 10/19/2012 10:43 PM - Luke Murphey

It only seems to get the first node wrong. The nodes following are correct.

The problem is that we are not handling heirarchy very well. For example, the following causes an error:

```
<div1 type="choral">
<div2 n="1" type="strophe">
<sp><speaker>*xoro/s</speaker>
```

This causes the importer to fail because the div tags are dropped yet each div tag is followed by a text node (an endline). This means that importer essentially sees a text node with a text node as its child (which should not happen).

To address this, we could either:

1. Attach all XML nodes (don't filter anymore)
2. Ignore text nodes under other text nodes
3. Attach text nodes that appear to be in a hierarchy.

#3 - 10/20/2012 01:01 AM - Luke Murphey

- Status changed from New to Closed

- % Done changed from 0 to 100