

TextCritical.net - Bug #1246

Some works have no spaces after commas or are missing dashes

03/07/2016 04:42 AM - Luke Murphey

Status:	New	Start date:	03/07/2016
Priority:	Normal	Due date:	
Assignee:	Luke Murphey	% Done:	0%
Category:		Estimated time:	0.00 hour
Target version:	Plus_3		

Description

Some works include mdash and ndash entities. These are currently not converted by the front-end and are therefore ignored. This causes the commas to be placed directly next the words without a space.

```
<span class="placename" data-key="tgn,7012913">Danube</span>
,which after all is but a small fraction of this
continent,&and, by the destruction of the Persian Empire,
they afterwards added to that the dominion of
<span class="placename" data-key="tgn,1000004">Asia</span>
```

Here is the original XML:

```
<placeName key="tgn,7012913">Danube</placeName>,&mdash;which after all is but a small fraction of
this
continent,&mdash;and, by the destruction of the Persian Empire,
they afterwards added to that the dominion of <placeName key="tgn,1000004">Asia</placeName>.
```

The actual parsed content in the database does not include the dashes. Thus, this is getting lost in the import process:

```
<placeName key="tgn,7012913">Danube</placeName>,which after all is but a small fraction of this
continent,&and, by the destruction of the Persian Empire,
they afterwards added to that the dominion of <placeName key="tgn,1000004">Asia</placeName>
```

Apparently, mdash and ndash are various lengths of dashes:

```
&ndash; is an en dash (-)
&mdash; is an em dash (—)
```

See <http://stackoverflow.com/questions/16550678/is-it-necessary-to-use-mdash-and-ndash-in-xhtml-or-html5>.

History

#1 - 03/07/2016 04:45 AM - Luke Murphey

A couple of options:

1. Add a filter that adds spaces after commas
2. Fix the importer to include the &mdash entities

I believe the importer is dropping the mdash entities because they are not valid entities; that may be the harder fix.

#2 - 05/10/2016 06:37 AM - Luke Murphey

- Subject changed from *Some works have no spaces after commas* to *Some works have no spaces after commas or are missing dashes*

#3 - 05/10/2016 06:37 AM - Luke Murphey

Some works are missing dashes. Compare:

- <http://www.perseus.tufts.edu/hopper/text?doc=Perseus%3atext%3a1999.04.0093>
- <http://textcritical.net/work/commentary-on-plato-protogoras-adam>

#4 - 11/23/2018 05:02 PM - Luke Murphey

- Target version changed from 3.0 to Plus_3