

Access to Memory (AtOM) - Bug #12886

Sitemap XML validation fails because of missing end tag

03/14/2019 11:29 AM - Corinne Rogers

Status:	Verified	Start date:	03/14/2019
Priority:	Medium	Due date:	
Assignee:		% Done:	0%
Category:	CLI tools	Estimated time:	0.00 hour
Target version:	Release 2.5.0	Tested version:	
Google Code Legacy ID:		Requires documentation:	No
Sponsored:	No		

Description

This was reported in the user forum here: <https://groups.google.com/forum/#topic/ica-atom-users/bYDBUwU1mkk>. When a sitemap is generated using the command line tool 'php symfony tools:sitemap' the resulting XML fails validation because the final tag </urlset> is missing. Adding the tag manually allows the XML to validate.

History

#1 - 04/10/2019 08:55 PM - Steve Breker

- Status changed from New to Code Review

Ready for CR:

<https://github.com/artefactual/atom/pull/877>

#2 - 04/10/2019 08:55 PM - Steve Breker

There was an issue with the sitemap CLI task where the last chunks of data were never written to the output files because the end() method registered with register_shutdown_function() was being called after both the sitemap index file and the last sitemap file were closed for writing - end() for the last files was being called after the "DONE" message was written.

I have removed the call to register_shutdown_function() in favour of calling end() directly from the sitemap task to complete writing the last sitemap.#.xml file, and added a call to flush() in the SitemapWriter class to flush the sitemap index file.

As per AtOM's documentation, a sitemap index file will always be created and the sitemap file will be broken into 50K line chunks. The sitemap index file can be ignored if there is only one sitemap chunk created.

A side effect of this fix is that static pages are now included in the sitemap file as intended, where before they were always in the last chunk and omitted due to the flush issue on the last file.

#3 - 04/10/2019 11:02 PM - Mike Cantelon

- *Status changed from Code Review to Feedback*
- *Assignee set to Steve Breker*

Looks good... nice detective work and fix!

#4 - 04/11/2019 09:55 AM - Steve Breker

- *Status changed from Feedback to QA/Review*
- *Assignee deleted (Steve Breker)*

Merged to qa/2.5.x. Ready for QA.

#5 - 04/12/2019 01:09 PM - Dan Gillean

- *Category set to CLI tools*
- *Status changed from QA/Review to Verified*
- *Target version set to Release 2.5.0*
- *Requires documentation set to No*

Nice work! Looks good. Interestingly, the extra sitemap file is also no longer empty! Contains pointer information, even when there's only 1 file.