

Saxon - Bug #4841

HTML5 serialization should suppress escaping of a script element in the XHTML namespace

2020-11-26 00:18 - Michael Kay

Status:	Closed	Start date:	2020-11-25
Priority:	Normal	Due date:	
Assignee:	Michael Kay	% Done:	100%
Category:	Serialization	Estimated time:	0:00 hour
Sprint/Milestone:		Spent time:	0:00 hour
Legacy ID:		Fix Committed on Branch:	10, trunk
Applies to branch:	10, trunk	Fixed in Maintenance Release:	10.5

Description

The HTML serializer is suppressing escaping (e.g. of ampersand) in a script element only if the script element is in no namespace. With HTML5 serialization it should also suppress escaping when the script element is in the XHTML namespace.

Serialization 3.1 § 7.3 says "The HTML output method MUST NOT perform escaping for any text node descendant, nor for any attribute of an element node descendant, of a script or style element.", but this to be read in the context of §7.1, which says "the HTML output method will not output an element differently from the XML output method unless the element is to be serialized as an HTML element." The definition of "serialized as an HTML element" says this happens, in the case of HTML5, when the element is in no namespace or in the XHTML namespace.

History

#1 - 2020-11-26 01:24 - Michael Kay

In the QT3 tests, the method-html test set contains a whole string of tests that were laid out by Henry Zongaro in 2012 and never completed. I have been fleshing these out, following his schedule, to improve the test coverage a little.

The relevant test is Serialization-html-9. Changing the code to suppress escaping in the text node within script was very easy. But the spec also says that escaping should also be suppressed in attributes; that's something we've never done and is proving a bit more challenging.

Meanwhile quite a few of the other new tests are failing too.

#2 - 2020-11-26 17:22 - Michael Kay

- Status changed from New to In Progress

The test Serialization-html-5, created to test this issue, is now passing. Some of the other new tests are not. I'll deal with them under this tracker rather than raising separate issues.

The first one is Serialization-html-5, in which we serialize an "XML island" - specifically an element in a random non-HTML-related namespace. We are outputting the empty element as `<magic></magic>` whereas I think the spec requires `<magic/>`. The rule is: the HTML output method will not output an element differently from the XML output method unless the element is to be serialized as an HTML element.

Fixed this in the method `HTMLEmitter.emptyElementTagCloser()` - this is now conditional on `isHTMLElement()`

#3 - 2020-11-26 19:47 - Michael Kay

I needed to make one other fix, the handling of boolean attributes (`<option selected="SELECTED">` wasn't working properly if the element is in the XHTML namespace. The new test cases are now working.

#4 - 2020-11-26 19:57 - Michael Kay

- Status changed from In Progress to Resolved

- Applies to branch 10, trunk added

- Fix Committed on Branch 10, trunk added

#5 - 2021-04-14 17:49 - O'Neil Delpratt

- Status changed from Resolved to Closed

- % Done changed from 0 to 100

- Fixed in Maintenance Release 10.5 added

Bug fix applied to Saxon 10.5 maintenance release.