

Saxon - Bug #5107

Base URI of validated document is incorrect

2021-09-28 15:31 - Michael Kay

Status:	Resolved	Start date:	2021-09-28
Priority:	Normal	Due date:	
Assignee:	Michael Kay	% Done:	0%
Category:	Schema-Aware processing	Estimated time:	0:00 hour
Sprint/Milestone:		Spent time:	0:00 hour
Legacy ID:		Fixed in Maintenance Release:	
Applies to branch:	10, 11, trunk	Platforms:	.NET, Java
Fix Committed on Branch:	10, trunk		

Description

Unit test `TestValidator.testValidationWithEntities()` is failing, on both the 10 and 11 branches, both Java and C# (but failing in slightly different ways).

The `XdmDestination` used as the result of validation is supposed to copy the base URIs of the corresponding elements in the source document. That's achieved by having the underlying `TinyBuilder` have the property `useEventLocation=true`. The logic in `XdmDestination.setReceiver()` sets this property only if no `BaseUri` property is set on the `XdmDestination`. But if no `BaseUri` property is set on the `XdmDestination`, then the document node will have no base URI, rather than having its base URI be a copy of that on the source document. The base URIs of element nodes are OK on Java because the absent URI is delivered as "" (zero-length URI) whereas on C# it is delivered as null (see `XdmDestination.BaseUri` property); this is because `System.Uri` does not accept "" as a legal URI value. But I added a check for the base URI of the document node, and this fails on both platforms.

History

#1 - 2021-11-12 20:58 - Michael Kay

Revisiting this, debugging the unit test on SaxonJ 10.6.

First observation is that the whole validation pipeline is duplicated - we are actually doing the validation twice.

The first `DocumentValidator` is created by `SchemaValidatorImpl#692`; the second by `Sender.send()` at line 121.

I have fixed this somewhat pragmatically by registering the `DocumentValidator` as a component on the `PipelineConfiguration`, and changing `Sender.send()` so it doesn't create a new `DocumentValidator` if there is already one on the pipeline.

Next observation is that the (changing) `systemIDs` on element nodes are actually being passed through to the target document correctly; the test is failing because the `systemId` on the document node is wrong.

#2 - 2021-11-13 00:43 - Michael Kay

Having some trouble with this.

If we set a base URI (or system ID) on the builder used to construct the result tree (validated tree), then it is used for all nodes in that tree. If we don't, then element nodes are getting the correct base URI set, but the document node isn't -- because in the `Receiver` interface, the `startDocument` event doesn't have a `Location` argument.

The problem with the setting `useEventLocation=true` is that the `startDocument` event has no event location.

#3 - 2021-11-29 12:10 - Michael Kay

I have finally found a solution to this as follows.

- When creating a `LocationCopier`, supply a system ID.
- In `TinyBuilder.open()`, if there is a `LocationCopier` on the `PipelineConfiguration`, use its `systemId` as the `systemId` of the `TinyBuilder`.
- When creating a `LocationCopier` in `Sender.sendDocumentInfo()`, set the `systemId` on the `LocationCopier`.

This basically compensates for the fact that the `Receiver.startDocument()` call does not have a `Location` argument, and therefore has no direct way of transmitting the base URI of the document node; instead we pass it indirectly via the `LocationCopier`.

#4 - 2021-11-29 12:38 - Michael Kay

The solution doesn't work "as is" on the 11 branch, because the CopyInformee on the PipelineConfiguration is now a simple lambda expression.

#5 - 2021-11-29 17:34 - Michael Kay

- *Status changed from New to Resolved*
- *Applies to branch 10, 11, trunk added*
- *Fix Committed on Branch 10, trunk added*
- *Platforms .NET, Java added*

Adapted the solution slightly to work on the 11 branch. Marking as resolved.