

Saxon - Bug #4328

saxon:get-pseudo-attribute() returns empty string on no match

2019-10-01 22:04 - Tony Graham

Status:	Closed	Start date:	2019-10-01
Priority:	Normal	Due date:	
Assignee:	Michael Kay	% Done:	0%
Category:	Saxon extensions	Estimated time:	0:00 hour
Sprint/Milestone:		Spent time:	0:00 hour
Legacy ID:		Fixed in Maintenance Release:	9.9.1.6
Applies to branch:	9.9, trunk	Platforms:	
Fix Committed on Branch:	9.9, trunk		

Description

saxon:get-pseudo-attribute() is documented [1] as returning an empty sequence on no match. It instead returns an empty string. This complicates checking for pseudo-attributes since it is not possible to distinguish an absent pseudo-attribute from one with a zero-length value.

The result of running the attached stylesheet:

```
$ java -jar c:/SaxonEE9-9-1-5J/saxon9ee.jar -it -xsl:get-pseudo-attribute.xsl
version: EE 9.9.1.5
pseudo: 'value' (true)
bogus: '' (true)
```

[1] <http://www.saxonica.com/documentation/#!functions/saxon/get-pseudo-attribute>

History

#1 - 2019-10-03 00:13 - Michael Kay

- Category set to Saxon extensions
- Status changed from New to In Progress
- Assignee set to Michael Kay
- Priority changed from Low to Normal

Thanks for reporting it. I haven't looked at this code for about 20 years and it appears to be only very skimpily tested. This particular problem is easily fixed, but while I was about it I wrote some tests for other edge cases and found a few issues -- partly because the spec is very vague, e.g. what happens if there's more than one pseudo-attribute with the same name. The worst bug I found was that if the data contains

```
ham="eggs"
```

then a search for "m" will return "eggs".

#2 - 2019-10-03 09:45 - Michael Kay

We use the same code internally for parsing the xml-stylesheet processing instruction, so I think we should follow the spec for "pseudo-attributes" that appears in <https://www.w3.org/TR/2010/REC-xml-stylesheet-20101028/>

The simplest way to do this is to wrap the string value into an element `<e + value + />`, parse this as XML, and then look for the corresponding attribute. This means that bad syntax (including invalid character references, duplicated attribute names, etc) becomes an error. That's a breaking change so I will leave it till the next major release.

The xml-stylesheet spec allows colons in names (including at the start of a name). I don't know if we can handle that by configuring the XML parser to be non-namespace-aware. I'll do some experiments. If we can't then it's not really important.

#3 - 2019-10-03 10:14 - Tony Graham

That's all fine by me. I personally don't use duplicate pseudo-attributes (that's taking 'pseudo' a bit too far) or put colons in pseudo-attribute names.

I'm currently using:

```
<!-- saxon:get-pseudo-attribute('nocharge') with no 'nocharge' returns an empty string. -->
<let name="nocharge" value="saxon:get-pseudo-attribute('nocharge')[. ne '']" />
<assert test="empty($nocharge) or $nocharge = ('error', 'not-error')"
  >'nocharge' must be either 'error' or 'not-error'. Value is: '<value-of select="$nocharge"/>'.
</assert>
```

which will be forwards-compatible and which is better than when I had:

```
<assert test="$nocharge = ('', 'error', 'not-error')">...</assert>
```

#4 - 2019-10-03 10:31 - Michael Kay

- Status changed from In Progress to Resolved
- Applies to branch 9.9, trunk added
- Fix Committed on Branch 9.9, trunk added

Resolved. For 9.9 I have fixed the reported bug only; there is still some sloppy parsing and sloppy matching of pseudo-attribute names.

For 10.0 I have rewritten the code to parse the pseudo attributes using a non-namespace-aware XML parser, which as far as I can see ensures that the rules of the xml-stylesheet processing instruction are followed exactly. An error SXCH0005 is raised if the pseudo-attribute syntax is incorrect, e.g. if there are duplicate attributes, attributes containing <, or invalid character references.

#5 - 2019-12-06 14:47 - O'Neil Delpratt

- Status changed from Resolved to Closed
- Fixed in Maintenance Release 9.9.1.6 added

Patch committed to the Saxon 9.9.1.6 maintenance release.

Files

get-pseudo-attribute.xsl	1.05 KB	2019-10-01	Tony Graham
--------------------------	---------	------------	-------------