

Saxon - Feature #4078

Fast XPath parsing for simple expressions

2018-12-20 19:22 - Michael Kay

Status: Closed	Start date: 2018-12-20
Priority: Normal	Due date:
Assignee: Michael Kay	% Done: 100%
Category: Performance	Estimated time: 0:00 hour
Sprint/Milestone:	Spent time: 0:00 hour
Legacy ID:	Fixed in Maintenance Release: 9.9.1.1
Applies to branch: 9.9	Platforms:
Fix Committed on Branch: 9.9	

Description

In XSLT, many path expressions are extremely simple. I collected some data and found that about 50% of path expressions in XSLT stylesheets belong to simple patterns such as @x or \$x or function() that can be readily detected after tokenization without going into the full recursive-descent parser, which has a lot of overhead for such simple expressions.

I've implemented a plug-in to the XPath parser for Saxon-EE that does this pre-scanning. The gains are not huge (about 4%), but they are worthwhile. The feature may make a bigger difference for applications that execute XPath expressions directly from Java, and fail to use caching to reduce XPath parsing costs.

History

#1 - 2018-12-20 19:26 - Michael Kay

- Status changed from New to Resolved

- Fix Committed on Branch 9.9 added

#2 - 2018-12-21 19:03 - Michael Kay

Fixed a couple of cases where this was leading to test failures (e.g. syntax errors not detected).

#3 - 2019-01-21 19:34 - O'Neil Delpratt

- % Done changed from 0 to 100

- Fixed in Maintenance Release 9.9.1.1 added

Bug fix applied to the Saxon 9.9.1.1 maintenance release.

#4 - 2019-01-21 19:38 - O'Neil Delpratt

- Status changed from Resolved to Closed