

Saxon - Bug #4419

fn:normalize-unicode fails when DTD-based validation of source data is turned on

2020-01-07 14:17 - Martin Lattner

Status:	Closed	Start date:	2020-01-07
Priority:	Normal	Due date:	
Assignee:	Michael Kay	% Done:	100%
Category:	Internals	Estimated time:	0:00 hour
Sprint/Milestone:		Spent time:	0:00 hour
Legacy ID:		Fixed in Maintenance Release:	9.9.1.7
Applies to branch:	9.9, trunk	Platforms:	
Fix Committed on Branch:	9.9, trunk		

Description

Hello!

I am running Saxon-EE 9.9.1.4 on Java from the command line.

When I turn on DTD-based validation of source data (-dtd:on), then my stylesheet using fn:normalize-unicode fails.

Here is XSL to demonstrate the problem:

```
<xsl:stylesheet xmlns:xsl="http://www.w3.org/1999/XSL/Transform" version="3.0">
  <xsl:output encoding="ASCII" omit-xml-declaration="yes"/>
  <xsl:template match="/">
    <xsl:value-of select="normalize-unicode('M&#252;ller', 'NFKD')"/>
  </xsl:template>
</xsl:stylesheet>
```

The source does not matter as long as it's XML that's valid against a DTD. For example:

```
<!DOCTYPE article SYSTEM "http://jats.nlm.nih.gov/archiving/1.2/JATS-archivearticle1.dtd">
<article><front><article-meta/></front></article>
```

When I run this with -dtd:off, I get the expected output:

```
Mu&#x308;ller
```

When I switch -dtd:don, then the transformation fails:

```
Recoverable error on line 2 column 13 of normalizationData.xml:
  SXXP0003: Error reported by XML parser: Document is invalid: no grammar found.
Recoverable error on line 2 column 13 of normalizationData.xml:
  SXXP0003: Error reported by XML parser: Document root element "UnicodeData", must match
  DOCTYPE root "null".
Error at char 18 in xsl:value-of/@select on line 7 column 68 of x.xsl:
  The XML parser reported two validation errors
net.sf.saxon.trans.XPathException: The XML parser reported two validation errors
```

It seems to me I should be able to use fn:normalize-unicode with -dtd:on.

History

#1 - 2020-01-07 15:48 - Michael Kay

Thanks for reporting it. It looks as if it's incorrectly applying DTD validation to one of the internal documents that holds the normalization tables.

#2 - 2020-01-17 17:38 - Michael Kay

- Category set to Internals

- Status changed from New to Resolved
- Assignee set to Michael Kay
- Priority changed from Low to Normal
- Applies to branch 9.9, trunk added
- Fix Committed on Branch 9.9, trunk added

I have fixed this on 9.9 and trunk.

I have reviewed all the places where we are parsing internal or system files where DTD validation is likely to be inappropriate, not just the issued files containing normalization and regex data, but also for example SEF files, configuration files, catalog files, and so on, and explicitly set DTD (and where necessary schema validation) off to override a request for DTD validation at the global configuration level.

#3 - 2020-03-05 11:09 - O'Neil Delpratt

- Status changed from Resolved to Closed
- % Done changed from 0 to 100
- Fixed in Maintenance Release 9.9.1.7 added

Patch applied in the 9.9.1.7 maintenance release.

Files

random_valid_xml.xml	140 Bytes	2020-01-07	Martin Latterner
xsl.xsl	281 Bytes	2020-01-07	Martin Latterner