

A Tutorial on XProc: An XML Pipeline Language

Rui Lopes

LaSIGE
University of Lisbon
rlopes@di.fc.ul.pt

Abstract. This tutorial will present the syntax and semantics of a new XML processing language specification, XProc, which is being standardised by the World Wide Web Consortium, for the description of sequences of operations to be performed on XML documents.

An XProc pipeline specification processes a set of inputs through an ordered set of operations, resulting on a set of outputs. Each operation performs an atomic transformation on inputs, feeding its results to the next operation. Such operations include well-known technologies, such as XSLT and XInclude, as well as special purpose micro-operations, e.g. insert and delete.

The purpose of this tutorial is the dissemination of this new XML technology, including step-by-step examples on how to create pipelines, and how to leverage the language's constructs and operations, easing the task of building simple and complex XML processing applications.

1 Synopsis

XML processing languages have come a long way. With the increasing use of such languages, including XSLT 2.0 [1], XQuery [2], building complex and full-fledged XML applications entirely in XML is starting to become a necessity. Moreover, background support through XML Schema [3] and other data validation technologies enforce the correctness of XML applications. However, a last piece has still been missing, the glue for all of these technologies. Typically, developers connected these self-contained technologies either through programming APIs, or by hacking custom shell scripts or makefiles. XProc [4] is W3C's answer to the standardisation of XML pipeline processing technologies. With XProc, XML developers can specify entire applications without leaving the XML syntax, thus leveraging existing knowledge on already existing standards and tools (e.g., XPath). This tutorial will provide an overview on how to build simple XML processing pipelines, and dive into some more advanced concepts, showing the expressive power of XProc.

1.1 Pipeline Concepts

The first part of the tutorial will present a short background on the different concepts inherent of XML pipelines, namely *pipelines*, *steps*, *inputs*, *outputs*, *options*, *parameters*, and *XPath contexts*.

1.2 Syntax Overview

After the brief introduction to the main concepts inherent of XProc, a syntax overview will be presented. Small examples will build upon each others, in order to leverage its core concepts, including the different namespaces of the language, scoping, and binding.

1.3 Steps

The core of this tutorial will be spent by presenting the different core steps of XProc, i.e., its main elements. These include *p:pipeline*, *p:for-each*, *p:viewport*, *p:choose*, *p:group*, and *p:try*.

1.4 Other Pipeline Elements

Complementing steps, other pipeline elements provide additional value to the language. Such examples include using inputs and parameters, debugging and documentation helpers, declaring steps, as well as building complex XML applications with pipelines and libraries.

1.5 Standard Step Library

The last part of this tutorial will provide a flavour of the standard step library of XProc. The concept of micro-operations will be introduced, as a practical way to define simple procedures to be applied onto XML documents. Well-known processors will be approached too, such as XSLT and XInclude.

References

1. Kay, M.: XSL Transformations (XSLT) Version 2.0. Technical report (2007) <http://www.w3.org/TR/xslt20>.
2. Boag, S., Chamberlin, D., Fernandez, M.F., Florescu, D., Robie, J., Simeon, J.: XQuery 1.0: An XML query language. Technical report (2007) <http://www.w3.org/TR/xquery>.
3. Thompson, H.S., Beech, D., Maloney, M., Mendelsohn, N.: Xml schema part 1: Structures second edition. Technical report (2004) <http://www.w3.org/TR/xmlschema-1/>.
4. Walsh, N., Milowski, A., Thompson, H.: XProc: An XML Pipeline Language. Technical report (2006) <http://www.w3.org/TR/xproc>.