Generative XPath

One XPath to rule them all

Oleg Parashchenko
Saint-Petersburg State University, Russia

olpa@ http://uucode.com/blog/
http://xmlhack.ru/
Outline

- Introduction
- Approach
- Architecture
- Correctness and performance
- Deploying
Use case: FrameMaker+SGML

Using XPath:

@attribute
// Return the value of an attribute
//
Sub GetAttributeValue Using vElement vAttributeName
     Local vValue; // Returns
     Local vIdx;
     Local vAttr;
     Local vAttrValList;
     If Not vElement.Attributes
         LeaveSub;
     EndIf
     Set vIdx = 1;
     Loop While (vIdx <= vElement.Attributes.Size)
         Get Member Number(vIdx) From(vElement.Attributes) NewVar(vAttr);
         If vAttr.AttrName = vAttributeName
             Set vAttrValList = vAttr.AttrValues;
             If vAttrValList
                 If 1 = vAttrValList.Size
                     Get Member Number(1) From(vAttrValList) NewVar(vValue);
                 EndIf
             EndIf
         LeaveLoop;
     EndIf
     Set vIdx = vIdx + 1;
     EndLoop
EndSub;

Using FrameScript
More use cases

- Compilers
- Text processors
- Any tree processing
XPath rule

Derived from Greenspun's Tenth Rule of Programming:

Any sufficiently complicated tree navigation library contains an ad hoc informally-specified bug-ridden slow implementation of half of XPath.
The need: portable XPath

One implementation for all trees and languages.

Generative programming is a software engineering paradigm based on modeling software families such that, given a particular requirements specification, a highly customized and optimized intermediate or end-product can be automatically manufactured on demand from elementary, reusable implementation components by means of configuration knowledge. — Krzysztof Czarnecki and Ulrich W. Eisenecker.
Outline

• Introduction
• **Approach**
• Architecture
• Correctness and performance
• Deploying
How?

Pseudocode (Virtual Machine):

- concise
- powerful
Code example

\[
\begin{align*}
  \text{(define (fac n)} \\
  &\text{(if (< n 2)} \\
  &\quad 1 \\
  &\quad (* n (fac (- n 1))))}) \\
\end{align*}
\]

\text{(fac 1)} ; Evaluates to 1

\text{(fac 6)} ; Evaluates to 720
Scheme R5RS
Outline

- Introduction
- Approach
- Architecture
- Correctness and performance
- Deploying
Two components

- Compiler
- Runtime
Runtime

Application layer

Customization layer

Virtual machine layer
Interfaces

Application layer to customization layer

- Load VM
- Execute XPath
- Data conversion
Interfaces

VM layer to customization layer

- Get an axis
- Compare document order
- Get a node property
XPath functions

- string
- namespace-uri
- local-name
- name
- lang
- id
string string(node) vs

The string function converts an object to a string as follows:

A node-set is converted to a string by returning the string-value of the node in the node-set that is first in document order. If the node-set is empty, an empty string is returned.

A number is converted to a string as follows

- NaN is converted to the string NaN
- positive zero is converted to the string 0
- negative zero is converted to the string 0
- positive infinity is converted to the string Infinity
- negative infinity is converted to the string -Infinity

If the number is an integer, the number is represented in decimal form as a Number with no decimal point and no leading zeros, preceded by a minus sign (−) if the number is negative

otherwise, the number is represented in decimal form as a Number including a decimal point with at least one digit before the decimal point and at least one digit after the decimal point, preceded by a minus sign (−) if the number is negative; there must be no leading zeros before the decimal point apart possibly from the one required digit immediately before the decimal point; beyond the one required digit after the decimal point there must be as many, but only as many, more digits as are needed to uniquely distinguish the number from all other IEEE 754 numeric values.

The boolean false value is converted to the string false.

The boolean true value is converted to the string true.

An object of a type other than the four basic types is converted to a string in a way that is dependent on that type.

If the argument is omitted, it defaults to a node-set with the context node as its only member.
Technical details

- See the paper

- See the example (C and Guile)
Compiler

Straightforward, but in generated code...

Morphisms instead of recursion

Usual algebra: \( x^2 - 10x + 21 = (x - 3)(x - 7) \)

There is also algebra of programming
Outline

- Introduction
- Approach
- Architecture
- Correctness and performance
- Deploying
Correctness is the must

Even for such clauses:

*If the argument is less than zero, but greater than or equal to -0.5, then negative zero is returned.*
Standard compliance

DocBook XSLT

xsltproc (XSieve) +

Generative XPath as the XPath engine

Works!
Performance

Today: it sucks :-(
unfair measurements: 30, 20, 2 times slower

In future: very, very fast
Outline

- Introduction
- Approach
- Architecture
- Correctness and performance
- Deploying
Finding a virtual machine

66 implementations listed on schemers.org

Recommended:

- C: Guile
- Java: SISC

From scratch: two weeks in free-time
Customization layer

Few hours (for me) or few days
In practice

XPath over S-expressions

(XLinq for LISP)
Wrap-up

- Universal XPath implementation
- Secret alien technology inside
- It works
Thank you!

Generative XPath

Oleg Parashchenko
Saint-Petersburg State University, Russia

olpa@ http://uucode.com/blog/
http://xmlhack.ru/