



FOX **path**

An expression language for selecting files and folders

Hans-Jürgen Rennau, Traveltainment GmbH
Presented at Balisage 2016, August 3, 2016



XPath is an expression language for navigating the contents of XML documents. Using XPath, we can select contents in an elegant, concise and very intuitive way. What is more: XPath is not a loose collection of expressions and rules, but a complete and fully composable expression language. Therefore there is virtually no limit to the selectiveness we may achieve, constructing XPath expressions. XML documents and a file system have much in common – they both expose a tree-structured collection of names. So – should we not have something like „XPath for files and folders“ – FOXpath for short? This presentation reports my efforts to design and implement such an expression language.



Outline

- XPath – a model for file system navigation?
- FOXxpath 1.0 – an expression language for navigating the file system
- FOXxpath 3.0 – FOXxpath merged into XPath
- Beyond the file system ...

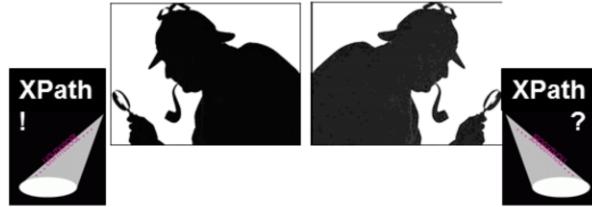
2016-08-03

FOXxpath - an expression language

2

First, we check the basic assumption that XPath is an adequate model for file system navigation. Then we explore FOXxpath, a new expression language for navigating the file system. We shall see that this first version of FOXxpath is a modified copy of XPath. But then we merge FOXxpath back into XPath, obtaining an extended version of XPath supporting file system navigation. And finally we shall see that FOXxpath is not restricted to physical file systems, but can navigate logical file systems as well, as for example used to organize the contents of a NOSQL database, version control repositories or the URI collection exposed by REST-ful web services.

XPath – a model for *file system investigation?*



tree



file system

2016-08-03

FOXPath - an expression language

3

XPath is a device for exploring and navigating XML documents. A closer look at XPath reveals core concepts not based on XML, but on the general notion of a tree-structured collection of names. XML is a particular instance of that notion, and the file system is another instance. We conclude that it should be possible to design an expression language which is similar to XPath, but deals with files and folders, rather than XML nodes.



Goal: examples

```
/xsdbase/niem-3.0//descendant~::*:iso*.xsd/parent~:::*
```

```
/xsdbase/niem-3.0//(*.xlsx, *.txt, *.xml)
```

```
/xsdbase/niem-3.0//*[is-dir()][not(*)]
```

```
sort(distinct-values(/xsdbase/niem-3.0//*.xml/xroot()))
```

```
/xsdbase/niem-3.0//*.xsd/file-name() except  
/xsdbase/niem-2.1//*.xsd/file-name()
```

Let us start with a few examples illustrating the desired expression language. They look like XPath, and they are as powerful as XPath expressions.

The AFFe principle



XPath:
within-tree direction

Axis

A F Fe

XPath:
name test | kind test

Filter

Filter expressions

XPath:
effective boolean value
of arbitrary expression

descendant::employees [address/city eq 'NewYork']
ancestor::element() [metaData]

2016-08-03

FOXPath - an expression language

5

If we want to use XPath beyond XML, we must understand the core principles of XPath in an abstract way, which is independent of XML. „Affe“ is the German word for monkey, and most monkeys love to traverse trees. It's natural to summarize the core principles of XPath under the term AFFe. It's an acronym referring to the building blocks of selection: axis + filter + filter expression. The axis collects everything visited when traversing the tree in a particular direction (children, descendants, parent, ancestors, and so forth). Collected items are submitted to a simple filter, which in the case of XPath is either a name test or a kind test. If this filtering is still too crude, further filtering of unlimited selectiveness can be achieved by adding filter expressions.

FOXxpath 1.0 = modified copy of XPath 3.0



- Expression language
- Data model = XPath data model
- Grammar = XPath grammar

- *PathExpr*
+ *FoxxpathExpr*

- Semantics = XPath semantics

- *semantics (PathExpr)*
+ *semantics (FoxxpathExpr)*

2016-08-03

FOXxpath - an expression language

6

XPath is a masterly elaboration of the AFFe principle. I never cease to marvel at the expressiveness achieved by uncompromising rigour and consistency. Aspiring to a new „XPath for files and folders“, what would be more promising than to stick as closely to XPath as possible?



fox path expression - example

```
/xsdbase/  
niem-3.0/  
descendant~::*.xsd  
  [xpath('not(//xs:documentation)')]/  
  file-info(., 'n40. d')
```



```
conformanceTargets.xsd ... 2013-10-18T14:19:18Z  
de.xsd ..... 2013-10-18T14:19:20Z  
gml.xsd ..... 2013-10-18T14:19:20Z  
localTerminology.xsd .... 2013-10-18T14:19:22Z
```

2016-08-03

FOXpath - an expression language

7

So the fox path expression is the only difference between XPath and FOXpath! Before becoming formal, let us look at an illustrative example. The expression reports all Niem XSDs without documentation. Ignoring the tilde and the use of wildcards within names, this could be an XPath expression.



fox path expression - grammar

fox path expression :
one or more steps

step :
postfix expression | fox axis step

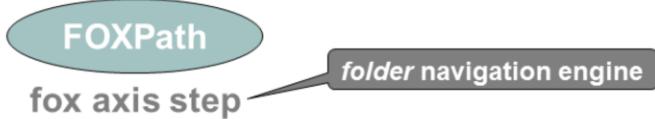
postfix expression : e.g. function call,...
concat(file-name(.), ' ', file-date(.))

fox axis step :
fox axis + fox name test [+ predicates]
descendant~:: .xsd [matches-xpath(. . .)]

That's because the grammar of the fox path expression is just a modified copy of the grammar of the path expression.



fox axis step –vs– axis step



fox axis + fox name test [+ predicates]

axis + name test [+ predicates]



So the core functionality of folder navigation, which is the fox axis step, is a modified copy of the axis step of XPath, which is the core functionality of node navigation..



fox axis step

- `child~::foo`
- `foo`
- `descendant~::zoo*bar`
- `descendant-or-self~::zoo?bar`
- `parent~::`foo bar``
- `ancestor~::`.git``
- `ancestor-or-self~::`2016``
- `following-sibling~::`201601??-*``
- `preceding-sibling~::foo`
- `self~::foo`

axis + name test
default axis: child

10

FOXpath supports most of the axes supported by XPath, excepting of course the attribute axis, as well as the preceding and following axis, which are not believed to be useful when navigating a file system. As in XPath, the child axis is the default axis.



fox name test

- Name pattern = characters and wildcards (*, ?)
- Canonical syntax (string in backquotes)
- Abbreviated syntax (without backquotes)
Escape certain characters by preceding ~
 - Escape always: ~[] \ / < > () += ! | , WHITESPACE
 - Escape if first character: . ` DIGIT

Canon.: `foo` `foo*` `foo(1)` `foo bar` `2016`

Abbrev.: foo foo* foo~(1~) foo~ bar ~2016

2016-08-03

FOXpath - an expression language

11

The fox name test filters by file or folder name. In contrast to the node name test of XPath, name patterns with inserted wildcards are supported. As file names can be almost arbitrary strings – not only NCNames – the syntax of a name test must make sure to avoid parsing ambiguity.



foxpath operator (/)

Example expression:

```
xsdbase / niem-3.0
```

E1 / E2

- evaluate E1
- atomize result
- **for every item P: evaluate E2, using P as context item**
- concat, remove duplicates, sort lexicographically

2016-08-03

FOXpath - an expression language

12

In XPath, the steps of a path expression are combined by the path operator, represented by a slash. In FOXpath, steps are combined by the foxpath operator, also a slash. The semantics are very similar to the path operator, but there is a key difference: what is pumped from left to right are URIs, not nodes.

Extended semantics – examples



```
/xsdbase//*[@*.xsd]
```

```
/xsdbase/niem-3.0//*[@.xsd]/file-name()
```

```
except
```

```
/xsdbase/niem-2.1//*[@.xsd]/file-name()
```

2016-08-03

FOXpath - an expression language

13

As we want FOXpath expressions to be just as elegant and concise as XPath expressions, we are motivated to introduce a slight extension to the expression semantics of XPath. Consider these examples, which look elegant and intuitive, yet would produce type errors if we retained XPath semantics without any changes.

Extended semantics – definition



Definition Extended semantics:

- Evaluation as XPath yields a value => evaluation as FOXxpath yields the same value
- Evaluation as XPath yields a type error **yet** evaluation as FOXxpath yields a value

2016-08-03

FOXpath - an expression language

14

Extended semantics are defined in terms of a comparison between the outcomes of evaluating the expression as an XPath or as a FOXpath expression.

Extended semantics – rules



- Effective boolean value

Case: $\text{value}(E) = \text{sequence of } >1 \text{ atomic items}$

XPath: $\text{EBV}(E)$ raises type error

FOXpath: $\text{EBV}(E)$ = $\text{EBV}(E[1])$

- E_1 or E_2 contains atomic item =>

- E_1 union E_2 = distinct-values $((E_1, E_2))$
- E_1 except E_2 = distinct-values $(E_1[\text{not}(. = E_2)])$
- E_1 intersect E_2 = distinct-values $(E_1[. = E_2])$

2016-08-03

FOXpath - an expression language

15

FOXpath defines four semantic extensions.



Function library

- All XPath 3.0 functions
- Plus ...
 - `is-dir()`, `is-file()`
 - `file-name()`, `file-date()`, `file-size()`, `file-info()` ...
 - `grep()`
 - `xpath()`
 - `xatt`, `xelem`, `xroot()`
 - `rpad()`, `lpad()`

2016-08-03

FOXpath - an expression language

16

The power of XPath relies on a standard library of built-in functions. FOXpath supports all XPath functions, plus a few additional functions believed to be especially useful when navigating and reporting file system contents.



Using functions ...

```
count(/xsdbase/niem-3.0//*.xsd)
```

```
/xsdbase/niem-3.0//*.xsd  
[xatt('attributeFormDefault', 'qualified')]  
/file-info('n40. s10 d")“
```

```
/sort(distinct-values(  
/xsdbase/niem-3.0//*[is-file()]  
/replace(file-name(),':*\'.',"")  
, lower-case#1)
```

2016-08-03

FOXpath - an expression language

17

A few examples illustrate the use of XPath as well as FOXpath functions.



Bringing It All Back Home



2016-08-03

FOXpath - an expression language

18

If FOXpath is almost identical to XPath, it is natural to ask if we cannot merge FOXpath back into XPath, obtaining an extended version of XPath, rather than a modified copy..

FOXxpath 3.0 – foxpath merged into XPath



- Concept
 - Do not *replace* path expression by fox path expr
 - *Extend* path expression:
 - Step: postfix expr | axis step | fox axis step
 - Operator: path operator | fox path operator
- Syntax changes
 - fox path operator: \ (not /)
 - fox name test: constrain use of abbreviated syntax
(HelloWorld – an axis step or a fox axis step?)

Indeed – it's possible! This new version of FOXxpath is called FOXxpath 3.0, as it is an extended version of XPath 3.0.



Mixing axis & fox axis steps

```
sort(distinct-values( fox axis steps
  \projects\xsd\niem*\/* .xsd\doc(.)  

  //xs:element/@name axis steps
path expression
) , lower-case#1)
```

2016-08-03

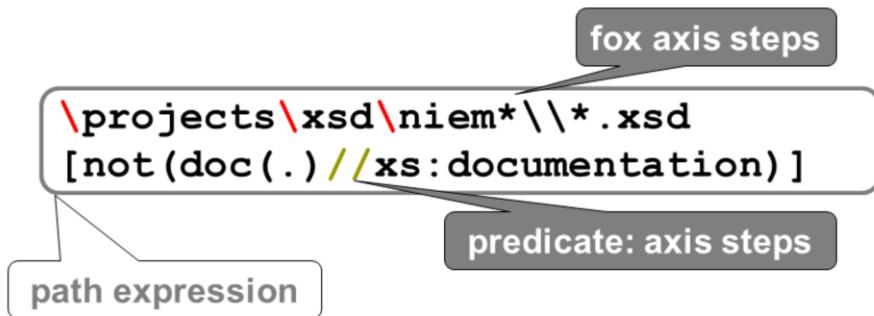
FOXpath - an expression language

20

With FOXpath 3.0, you can not only use fox axis steps, you can also mix axis and fox axis steps within a single path expression. This enables expressions accomplishing a two-phase navigation: initial steps select resources, following steps navigate into their contents.



Mixing axis & fox axis steps 2



2016-08-03

FOXpath - an expression language

21

Another example of mixing is the use of fox axis steps with predicates containing axis steps, which means resource selection based on resource contents.



FOXpath 3.0 - implementation

- Reference implementation:
written in XQuery 3.1 (5331 LOC)
- Dependency on XQuery processor
Currently: BaseX 8.5 or greater
- <https://github.com/hrennau/foxpath>

2016-08-03

FOXpath - an expression language

22

A reference implementation of FOXpath 3.0 is available.



```
for $d in /xsdbase/niem-3.0//*[is-dir()] return  
concat(  
  for $i in 1 to (count($d/ancestor~::*) - 2) return '.',  
  file-name($d))
```

```
. niem  
. . adapters  
. . . edxl-cap  
. . . . 3.0  
. . . . edxl-de  
. . . . 3.0  
. . . . edxl-have  
. . . . 3.0  
. . . . geospatial  
. . . . 3.0  
. . appinfores  
. . . 3.0  
. . codes  
. . . ansi_d20  
. . . . 3.0
```

The implementation in action – a complex expression producing a tree representation of the folder tree found under the niem-3.0 folder.



Moving beyond the file system

- File system – tree of resource URIs
- Other types of resource trees:
 - Resource URIs exposed by a REST-ful web service
 - URIs of documents stored in a NOSQL database
 - URIs of resources managed by version control
 - ...
- **Folder navigation** defined by FOXpath –
not restricted to the file system!

2016-08-03

FOXpath - an expression language

24

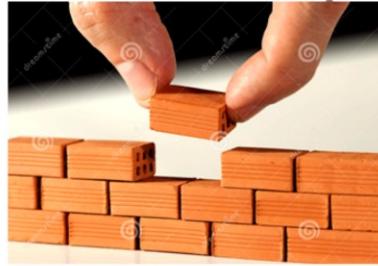
The folder navigation supported by FOXpath is not restricted to physical file systems – it can also deal with other kinds of resource trees, which may be regarded as logical file systems.



Folder navigation

Basic building block of folder navigation:

fox axis step



2016-08-03

FOXpath - an expression language

25

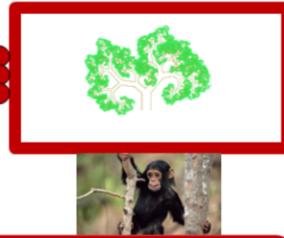
It is important to recognize the basic building block of folder navigation – which is the fox axis step. The question arises whether the fox axis step can deal with logical file systems as well.



fox axis step: implementation

tree interface

fox axis step



`fox:child-uri-collection($uri, $name) as xs:anyURI*`
`fox:descendant-uri-collection($uri, $name) as xs:anyURI*`
`fox:root-uri($uri) as xs:anyURI?`

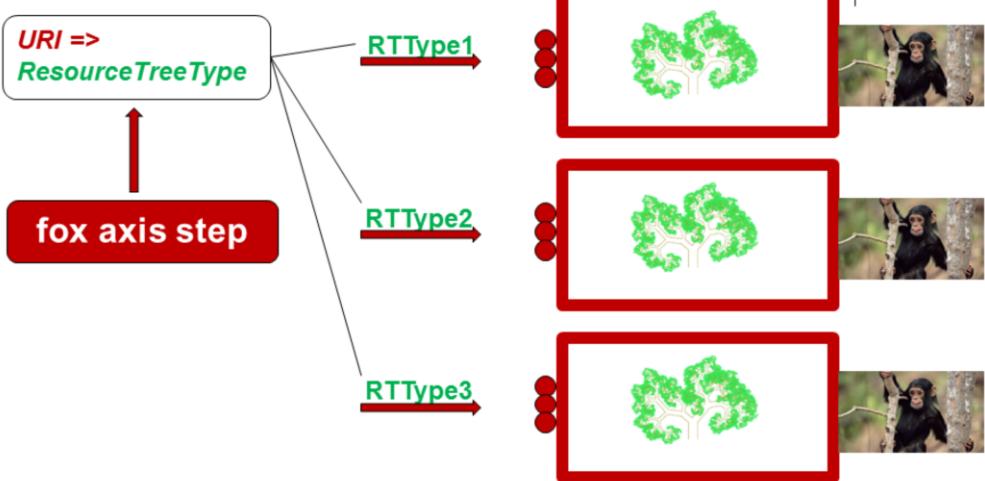
2016-08-03

FOXpath - an expression language

26

As you remember, the fox axis step combines a fox axis with a fox name test. A complete implementation can be built upon navigation primitives, three functions mapping an input URI to output URIs – root, child and descendant URIs. The functions constitute a sufficient interface to the tree, as far as navigation is concerned. This interface completely hides the nature of the resource tree – physical file system, SVN repository, etc. In principle, FOXpath can navigate any type of resource tree, for which implementations of these three navigation primitives are available.

Parallel support for multiple resource tree types



2016-08-03

FOXpath - an expression language

27

What is more: FOXpath can support in parallel any number of resource tree types. This presupposes that for a given URI, the type of containing resource tree can be determined. If this is the case, for a given context URI, the appropriate instances of the navigation primitives can be selected and the fox axis step can be executed.



Formal integration

Problem: set of navigation-supported resource tree types is implementation-defined!

Solution: New component of the **dynamic context**:

*[Definition: **Available navigation primitives.** This is a mapping of strings to function items representing the three navigation primitives.]*

2016-08-03

FOXpath - an expression language

28

Support for various resource tree types will certainly be implementation dependent. How to integrate this variability cleanly into the highly standardized XPath language?

FOXpath - an afterthought



Nested forest:

outer forest = resource trees

inner forest = node trees

Seamless navigation of the nested forest

Emerges:

*a continuous space of information
(the **info space**)*

2016-08-03

FOXpath - an expression language

29

An afterthought: FOXpath complements tree navigation of resource contents with tree navigation of resource collections. FOXpath encourages us to perceive a nested forest of information: an outer forest consisting of resource trees, an inner forest consisting of the node trees corresponding to leaves of the resource trees. FOXpath enables us to navigate the complete structure in a seamless way, and therefore we may experience this nested forest as a single, continuous space – the info space.



The end! (The end?)



2016-08-03

FOXPath - an expression language

30

There is a choice to be made: either XPath remains an ingenious tool for navigating XML documents, or it will be extended to become the engine of the info space.