

XPath

T. Pitner, L. Bártek, A. Rambousek. L Grollich
FI MU Brno 2020

XPath – základní principy

- XPath je syntaxe pro specifikování částí dokumentů v XML
 - Primitivních hodnot, uzlů a sekvencí uzlů a hodnot
- XPath neumožňuje specifikovat části textových uzlů
- Jméno je odvozeno od *path expression* poskytující hierarchické adresování uzlů v XML stromu.
- Syntax je podobná cestám v souborovém systému.
- XPath poskytuje *standardní knihovnu funkcí*, stejně jako uživatelsky definované funkce v procesorech pro XPath 2.0 resp. i XPath 1.x.
- XPath nepoužívá XML syntaxi.
 - Zápis by byl dlouhý.

XPath - specifikace

- XPath 1.0 (revize ze 7. 9. 2015)
- XML Path Language (XPath) 2.0 a
- XML Path Language (XPath) 3.0 – doporučení W3C (8.4. 2014)
- XPath 3.1 – doporučení W3C (21.3. 2017)
- Zpětná kompatibilita: téměř všechny XPath 1.0 výrazy poskytují stejný výsledek s XPath 3.0
 - Výjimky viz kapitola [Zpětná kompatibilita](#) ve specifikaci.

XPath v jiných technologiích

- XPath se využívá jako základ pro XSLT od verze 1.0 a
- V XQuery od XPath verze 2.0.

Základní zdroje pro studium

- [XPath tutorial](#) na W3Schools
- [Zvon XPath 1.0 Tutorial](#) to learn step by step (by Miloslav Nič)
- [PathEnq](#) - XPath 2.0 online evaluator - nice for try&see
- [Online XPath Tester - Evaluator](#) by ExtendsClass

Doména XPath: Pokročilá navigace v XML

```
<?xml version="1.0"?>
```

```
<a>
```

```
  <b/> <!-- this is the //b[count(./*)=0] -->
```

```
  <b> <!-- this is the //b[./c] -->
```

```
    <c/>
```

```
  </b>
```

```
  <b> <!-- this is the //b[3] -->
```

```
    <!-- and also //b[./c] -->
```

```
    <c/>
```

```
  </b>
```

```
</a>
```

- Select the 3rd node `_b_`: ``//b[3]``
- Select a node "b", which has a child node "c": ``//b[./c]``
- Select an empty (eg. no child elements) node b: ``//b[count(./*)=0]``

Doména XPath - Transformation (XSLT)

- Výběr zpracovávaného uzlu
`<xsl:apply-templates match="para"/>`
- Výběr hodnoty
`<xsl:value-of select="para/@id"/>`

Doména XPath: Selekční část XQuery

- Část (F)or:
 - for \$para in \$doc//para – vybere všechny části para
- Část (L)et:
 - Let \$mypara := \$doc//para[@author='myself']
- Část (W)here:
 - where \$para[@class='task']
- Část (O)rder:
 - order by \$para/@created

Doména XPath: Modelovací jazyky

- XML Schema
- Schematron

XPath – cesty a navigace

- Cesta popisuje (naviguje) k místu v XML dokumentu.
 - Její konstrukce je podobná cestám na souborovém systému.
- Cesty jsou:
 - Relativní – vztažené ke kontextovému uzlu.
 - Absolutní – vztažené ke kořenu dokumentu
 - Predikáty jsou vyhodnocovány relativně

Datové typy v XPath

- Od XPath 3.0 sjednoceno s datovými typy XMLSchmatu a XQuery
- Viz doporučení W3C [XQuery and XPath Data Model 3.0](#), 8.4. 2014

Osy

- *Osy* – množiny elementů dokumentu
 - Většinou vztažené ke kontextovému uzlu.
- *Kontext* je tvořen dokumentem a aktuálním kontextovým uzlem.

Seznam os

- child – přímí potomci kontextového uzlu (CN).
- descendant – následníci CN s výjimkou atributů.
- parent – rodičovský element CN.
- ancestor – předci CN, rodiče, prarodiče, ..., až ke kořenovému uzlu.
- following-sibling – následující sourozenci
 - Je prázdná pro jmenné prostory a atributy.
- preceding-sibling – dtto, ale předcházející sourozenci.

Seznam os

- following
 - Obsahuje všechny následující uzly CN s výjimkou atributů, dceřiných uzlů a uzlů jmenného prostoru (NS).
- preceding
 - Dtto s výjimkou atributů, předků a NS.
- attribute
 - Obsahuje atributy (pouze pro elementy).
- namespace
 - Obsahuje všechny uzly NS daného CN.
- self
 - Pouze CN
- descendant-or-self
 - Obsahuje sjednocení os self a descendant.
- ancestor-or-self
 - Obsahuje sjednocení os self a ancestor.

Predikáty

- Příklad:
 - /article/para[3] – výběr 3. odstavce (element para) v článku (element article)
- Nejjednodušší predikát – přibližná specifikace pozice (viz preceding).
- Pozor u reverzních os (preeding, ancestor, ...)
 - Pozice se počítá od kontextového uzlu.
- Pozici 3 lze nahradit pomocí position()=3

Výrazy

- Používají se v *predikátech* pro výpočty.
 - Mohou používat funkce z knihovny Xpath.
- Výrazy mohou pracovat s:
 - Textovými řetězci
 - Čísly (v pohyblivé čárce)
 - Logickými hodnotami (boolean)
 - Uzly
 - Sekvencemi

Zkrácený zápis

- para – vybere všechny dceřiné elementy CN s názvem para
- * – vybere všechny dceřiné elementy CN
- text() – vybere všechny textové potomky CN
- @name – atribut *name* CN
- @* – všechny atributy CN
- para[1] – první přímý potomek para
- para[last()] - poslední přímý potomek para
- */para – všichni vnuci para CN

Zkrácený zápis

- `/doc/chapter[5]/section[2]` - druhá *section* páté *chapter* elementu *doc*
- `chapter//para` - všichni potomci elementu *chapter* se jménem *para*
- `//para` – všechny elementy *para* v dokumentu
- `//olist/item` – všechny elementy *item* s rodičem *olist*
- `./para` – všichni potomci CN se jménem *para*
- `..` – rodičovský uzel CN
- `../@lang` – atribut *lang* rodiče CN
- Používá se `article/para` místo `child::article/child::para`.
- U atributu se používá `para[@type="warning"]` místo `child::para[attribute::type="warning"]`
- Další často používaný zkrácený zápis je `//` místo `/descendant-or-self::node()/`
- A samozřejmě `.` a `..`
- Pro lepší čitelnost je vhodné použít nezkrácený zápis.

XPath 2.0

- Jiný úhel pohledu:
 - Vše je sekvence – odstraněn problém s pořadím prvků v množině.
- Zavádí:
 - podmíněné výrazy a cykly,
 - Uživatelsky definované funkce
 - dynamicky vyhodnocované XPath výrazů.
 - Obecný existenční kvantifikátor
 - exist student
 - all [student/@id](#)
 - Více viz www.saxonica.com

XPath 2.0 - ukázky

- Řetězcové funkce
- Numerické funkce
- Funkce nad sekvencemi
- Boolovské funkce

Další zdroje k XPath

- [Programming in XPath 3.0](#) (D. Novatchev)
- [XPath functions](#) (Mozilla)

On-line nástroje pro testování XPath

- XPath výrazy můžete otestovat on-line bez nutnosti instalovat na počítač např. na:
 - <http://codebeautify.org/Xpath-Tester>
 - [PathEng](#) - XPath 2.0 on-line vyhodnocovač.
 - [XPath on-line tester](#) - umožňuje také vyhodnocovat XPath proti nahranému dokumentu.