



XPath : l'essentiel

Date : 05/09/2017

Version : 0.1.0

Description du document

Titre : Xpath, l'essentiel

Version : 0.1.0

Auteur(s) : Blaise de Carné-Carnavalet

Historique du document

Date	Version	Description
12/08/2016	0.1.0	Création du document

Table des matières

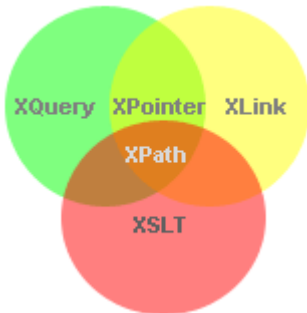
1. Introduction.....	4
1.1.XPath ?.....	4
1.2.Chemins.....	4
1.3.Terminologie.....	5
2. Syntaxe.....	6
2.1.Chemin de sélection.....	6
2.2.Axes.....	6
2.2.1. Liste des axes.....	6
2.2.2. Abréviations.....	7
2.3.Prédicats.....	8
2.4.Sélection à l'aveugle.....	9
2.5.Sélection multiple.....	9
3. Opérateurs.....	9
4. Exemples divers.....	10

1. Introduction

1.1. XPath ?

XPath est un élément majeur du standard XSLT (eXtensible Stylesheet Language Transformations).

Il est utilisé pour définir les règles d'extraction des éléments et attributs des documents XML envoyés sur la plateforme DATAtourisme.



- XPath signifie "XML Path Language"
- XPath utilise une syntaxe de "chemin" pour identifier et naviguer parmi les noeuds d'un document XML
- XPath contient plus de 200 fonctions intégrées
- XPath est une recommandation du W3C

1.2. Chemins

Le Xpath utilise une notation de "chemins" pour sélectionner les noeuds ou des ensembles de noeuds au sein de la hiérarchie d'un document XML.

Cette notation peut être comparée à la notation utilisée traditionnellement au sein du système de fichier d'un ordinateur.



/BOOKS/BOOK/AUTHOR/LASTNAME

1.3. Terminologie

```
<?xml version="1.0" encoding="UTF-8"?>

<bookstore>
  <book>
    <title lang="en">Harry Potter</title>
    <author>J K. Rowling</author>
    <year>2005</year>
    <price>29.99</price>
  </book>
</bookstore>
```

Noeuds

Dans Xpath, il existe plusieurs types de noeud : élément, attribut, text, namespace et commentaire.

Un document XML est traité comme un arbre de noeuds. L'élément le plus haut placé est appelé l'élément root.

```
<bookstore> (root element node)
```

```
<author>J K. Rowling</author> (element node)
```

```
lang="en" (attribute node)
```

Valeurs atomiques

Les noeuds peuvent contenir des valeurs texte, appelées valeurs atomiques. Dans l'exemple précédent il s'agira de "J K. Rowling" ou "en".

Lien entre les noeuds

- Parent
- Enfant
- Collateral
- Ancêtres
- Descendants

2. Syntaxe

```
<?xml version="1.0" encoding="UTF-8"?>

<bookstore>

<book>
  <title lang="en">Harry Potter</title>
  <price>29.99</price>
</book>

<book>
  <title lang="en">Learning XML</title>
  <price>39.95</price>
</book>

</bookstore>
```

2.1. Chemin de sélection

Un chemin de sélection peut être absolu ou relatif : un chemin absolu commence par un slash (/). Dans les deux cas le chemin consiste en une succession d'étapes séparées par un slash :

- Chemin absolu : /etape/etape/ etape
- Chemin relatif : etape/etape/ etape

Chaque étape est évaluée l'une après l'autre pour arriver au noeud(s) souhaité(s).

Une étape est composée de :

- **Un axe**, qui définit le lien hiérarchique à explorer.
- **Un test de noeud**, qui identifie les noeuds à sélectionner
- Zero ou plusieurs **prédicats** qui permet de filtrer l'ensemble de noeuds sélectionnés.

```
axisname::nodetest[predicate]
```

2.2. Axes

Un axe définit un ensemble de noeuds relatifs au noeud courant.

2.2.1. Liste des axes

Axe	Resultat
ancestor	Sélectionne les ancêtres (parents, grands-parent, etc.) du noeud courant
ancestor-or-self	Sélectionne les ancêtres (parents, grands-parent, etc.) du noeud courant et le noeud courant
attribute	Sélectionne tous les attributs du noeud courant
child	Sélectionne tous les enfants du noeud courant
descendant	Sélectionne tous les descendants (enfants, petits-enfants) du noeud courant
descendant-or-self	Sélectionne tous les descendants (enfants, petits-enfants) du noeud courant et le noeud courant
following	Sélectionne tous dans le document après la fermeture de la balise du noeud courant
following-sibling	Sélectionne tous les collatéraux après le noeud courant
namespace	Sélectionne tous les noeuds de namespaces après le noeud courant
parent	Sélectionne tous les parents du noeud courant
preceding	Sélectionne tous les noeuds apparaissant avant le noeud courant dans le document, à l'exception des ancêtres, attributs et namespaces.
preceding-sibling	Sélectionne tous les collatéraux avant le noeud courant
self	Sélectionne le noeud courant

Exemples

Example	Resultat
child::book	Sélectionne tous les éléments book qui sont enfants du noeud courant
attribute::lang	Sélectionne l'attribut lang du noeud courant
child::*	Sélectionne tous les éléments enfants du noeud courant
attribute::*	Sélectionne tous les attributs du noeud courant
child::text()	Sélectionne toutes les valeurs atomiques enfants du noeud courant
child::node()	Sélectionne tous les noeuds enfants du noeud courant
descendant::book	Sélectionne tous les éléments book descendants du noeud courant
ancestor::book	Sélectionne tous les éléments book ancêtres du noeud courant
ancestor-or-self::book	Sélectionne tous les éléments book ancêtres du noeud courant et le noeud courant
child::* / child::price	Sélectionne tous les éléments price petit-enfant du noeud courant

2.2.2. Abréviations

Il existe des abréviations pour certains axes :

Abbréviation	Axe	Description
(aucun axe)	child::	Sélectionne un enfant direct
@	Attribute::	Sélectionne un attribut
//	descendant-or-self::	Sélectionne les noeuds descendants dans toute la hiérarchie

.	self::node()	Sélectionne le noeud courant
..	parent::node()	Sélectionne le parent du noeud courant

Exemples

Path Expression	Resultat
bookstore	Sélectionne tous les éléments avec le nom "bookstore"
/bookstore	Sélectionne l'élément root bookstore Note: Si le chemin commence par un slash (/) cela représente toujours le chemin absolu vers un élément.
bookstore/book	Sélectionne tous les éléments book qui sont enfants de l'élément bookstore
//book	Sélectionne tous les éléments book du document
bookstore//book	Sélectionne tous les éléments book qui sont descendants directs ou indirects de l'élément bookstore
//@lang	Sélectionne tous les attributs lang du document

2.3. Prédicats

Les prédicats servent à filtrer les noeuds sélectionnés. Ils sont toujours entourés de brackets ([])

Path Expression	Resultat
/bookstore/book[1]	Sélectionne le premier élément book enfant de l'élément bookstore.
/bookstore/book[last()]	Sélectionne le dernier élément book enfant de l'élément bookstore
/bookstore/book[last()-1]	Sélectionne l'avant dernier élément book enfant de l'élément bookstore
/bookstore/book[position()<3]	Sélectionne les 2 premiers éléments book enfants de l'élément bookstore
//title[@lang]	Sélectionne tous les éléments title qui ont un attribut lang
//title[@lang='en']	Sélectionne tous les éléments title qui ont un attribut lang ayant pour valeur "en"
/bookstore/book[price>35.00]	Sélectionne tous les éléments book de l'élément bookstore et ayant un attribut price supérieur 35.00
/bookstore/book[price>35.00]/title	Sélectionne tous les éléments title d'un élément book enfant de l'élément bookstore et ayant un attribut price supérieur 35.00

2.4. Sélection à l'aveugle

Wildcard	Description
*	Sélectionne n'importe quel élément
@*	Sélectionne n'importe quel attribut
node()	Sélectionne n'importe quel noeud, quelque soit son type

Exemples

Path Expression	Result
/bookstore/*	Sélectionne tous les éléments enfants de l'élément bookstore
//*	Sélectionne tous les éléments dans le document
//title[@*]	Sélectionne tous les éléments title qui ont au moins un attribut

2.5. Sélection multiple

En utilisant l'opérateur | il est possible de sélectionner plusieurs éléments

Path Expression	Resultat
//book/title //book/price	Sélectionne tous les éléments title ET price de tous les éléments book
//title //price	Sélectionne tous les éléments title ET price dans le document
/bookstore/book/title //price	Sélectionne tous les éléments title des éléments book de l'élément bookstore ET tous les éléments price du document

3. Opérateurs

Operator	Description	Exemple
	Assemble 2 résultats de sélection	//book //cd
+	Addition	6 + 4
-	Soustraction	6 - 4
*	Multiplication	6 * 4
div	Division	8 div 4
=	Égal	price=9.80
!=	Différent	price!=9.80
<	Inférieur à	price<9.80
<=	Inférieur ou égal à	price<=9.80
>	Supérieur à	price>9.80
>=	Supérieur ou égal à	price>=9.80
or	ou	price=9.80 or price=9.70
and	et	price>9.00 and price<9.90
mod	Modulo (reste de la division euclidienne)	5 mod 2