

Le langage d'interrogation XPath

Emeline DOREY

16 janvier 2014

emeline.dorey@gmail.com

Sommaire

Le langage
d'interrogation
XPath

Emeline DOREY

Présentation

Syntaxe générale

Evaluation

Les axes

Les filtres

Les prédicats

Les fonctions

Syntaxe abrégée

Exemple

- 1 Présentation
- 2 Syntaxe générale
 - Evaluation
 - Les axes
 - Les filtres
 - Les prédicats
 - Les fonctions
- 3 Syntaxe abrégée
- 4 Exemple

Présentation

Le langage
d'interrogation
XPath

Emeline DOREY

Présentation

Syntaxe générale

Evaluation

Les axes

Les filtres

Les prédicats

Les fonctions

Syntaxe abrégée

Exemple

XPath est :

- une spécification du W3C ;
- l'expression de chemins XML standards ;
- le prédicat pour spécifier les valeurs d'éléments ou d'attributs ;
- la brique de base pour d'autres technologies XML :
 - XLink : spécification des hyperliens ;
 - XPointer : pointer des éléments de documents avec des expressions XPath dans les URL
 - XSLT : langage de transformation ;
 - XQuery : langage de requêtes, ...

Présentation

Le langage
d'interrogation
XPath

Emeline DOREY

Présentation

Syntaxe générale

Evaluation

Les axes

Les filtres

Les prédicats

Les fonctions

Syntaxe abrégée

Exemple

XPath est un langage d'interrogation des documents XML. Il permet de sélectionner certaines parties d'un document XML : des sous-arbres, des noeuds, des attributs... Il permet de naviguer dans l'arbre du document, à partir d'un noeud origine, vers un ou plusieurs noeuds destinations, en sélectionnant des chemins dans l'arbre.

Syntaxe générale

Le langage
d'interrogation
XPath

Emeline DOREY

Présentation

Syntaxe générale

Evaluation

Les axes

Les filtres

Les prédicats

Les fonctions

Syntaxe abrégée

Exemple

L'endroit d'où l'on part se nomme le noeud courant. En première lecture on peut imaginer qu'il s'agit de la racine du document, mais n'importe quel noeud peut jouer ce rôle. Ensuite, il y a trois autres types d'éléments :

- un axe : la direction dans laquelle on se dirige à partir du noeud courant (vers le père, vers les fils, vers les frères de gauche...);
- un filtre : le type de noeuds qui nous intéresse dans l'axe choisi (des noeuds quelconques, des éléments quelconques ou un élément précis, des commentaires...);
- un prédicat (optionnel) : des conditions supplémentaires pour sélectionner des noeuds parmi ceux retenus par le filtre dans l'axe.

Syntaxe générale

Le langage
d'interrogation
XPath

Emeline DOREY

Présentation

Syntaxe générale

Evaluation

Les axes

Les filtres

Les prédicats

Les fonctions

Syntaxe abrégée

Exemple

- A eux trois, ils constituent une étape.
- Le principe de la syntaxe XPath est semblable à celui de l'adressage du système de fichiers, c'est l'enchaînement de plusieurs étapes, séparées par des "/" .

```
[/]etape1/etape2/.../etapeX
```

Syntaxe générale

Le langage
d'interrogation
XPath

Emeline DOREY

Présentation

Syntaxe générale

Evaluation

Les axes

Les filtres

Les prédicats

Les fonctions

Syntaxe abrégée

Exemple

- Si le chemin commence par un `"/`, il s'agit d'un chemin absolu, c'est-à-dire prenant son origine à la racine du document et non pas au noeud courant.
- Il est possible de faire une disjonction de requêtes XPath avec le signe `|` ; on obtient alors l'union des deux ensembles de noeuds correspondants.
- Chaque étape renvoie un ensemble de noeuds et pour chacun d'entre eux, on applique les étapes suivantes.

Syntaxe générale

Le langage
d'interrogation
XPath

Emeline DOREY

Présentation

Syntaxe générale

Evaluation

Les axes

Les filtres

Les prédicats

Les fonctions

Syntaxe abrégée

Exemple

Une étape a la forme :

```
axe::filtre [predicat1] [predicat2] ...
```

- L'axe définit le sens du parcours des noeuds ;
- Le filtre indique le type des noeuds qui seront retenus dans l'évaluation de l'expression ;
- Le (ou les) prédicat(s) exprime(nt) des propriétés que doivent satisfaire les noeuds retenus.

Syntaxe générale

Evaluation d'une expression

Le langage
d'interrogation
XPath

Emeline DOREY

Présentation

Syntaxe générale

Evaluation

Les axes

Les filtres

Les prédicats

Les fonctions

Syntaxe abrégée

Exemple

- L'évaluation d'une étape peut être une valeur, ou un ensemble de noeuds ;
- Les étapes sont évaluées les unes après les autres :
 - A partir du noeud contexte, on évalue l'étape 1, qui renvoie un ensemble de noeuds ;
 - Chaque noeud de cet ensemble est considéré comme noeud contexte pour l'évaluation de l'étape 2 ;
 - On procède de la même manière pour les autres étapes.

Syntaxe générale

Les axes

Le langage
d'interrogation
XPath

Emeline DOREY

Présentation

Syntaxe générale

Evaluation

Les axes

Les filtres

Les prédicats

Les fonctions

Syntaxe abrégée

Exemple

- **elf** : le noeud courant lui-même ;
- **child** : les enfants du noeud courant ;
- **descendant, descendant-or-self** : tous les descendants du noeud courant ;
- **parent** : le père du noeud courant ;
- **ancestor, ancestor-or-self** : les ancêtres du noeud courant ;
- **attribute** : les attributs du noeud courant ;
- **preceding, following** : les noeuds, précédants ou suivants du noeud courant, dans l'ordre de lecture du document ;
- **preceding-sibling, following-sibling** : les frères, précédant ou suivant, le noeud courant ;
- **namespace** : les espaces de noms.

Syntaxe générale

Les filtres

Le langage
d'interrogation
XPath

Emeline DOREY

Présentation

Syntaxe générale

Evaluation

Les axes

Les filtres

Les prédicats

Les fonctions

Syntaxe abrégée

Exemple

- **node()** : tous les noeuds ;
- **text()** : les noeuds textuels ;
- ***** : tous les éléments ;
- **nom** : les éléments portant ce nom ;
- **comment()** : les noeuds commentaires ;
- **processing-instruction('cible')** : les noeuds instructions, seulement les instructions cibles si cet argument est fourni.

Syntaxe générale

Les prédicats

Le langage
d'interrogation
XPath

Emeline DOREY

Présentation

Syntaxe générale

Evaluation

Les axes

Les filtres

Les prédicats

Les fonctions

Syntaxe abrégée

Exemple

Ils prennent la forme de tests que les noeuds sélectionnés devront vérifier. Ces tests peuvent impliquer des fonctions ou de nouveaux chemins XPath.

```
child::variete[position()=2]  
child::variete[@precision='opaque']
```

Syntaxe générale

Les prédicats

Le langage
d'interrogation
XPath

Emeline DOREY

Présentation

Syntaxe générale

Evaluation

Les axes

Les filtres

Les prédicats

Les fonctions

Syntaxe abrégée

Exemple

Une expression entre crochets permet de spécifier plus précisément un élément.

```
/nuages/nuage[1] selectionne le premier element  
'nuage' de 'nuages'.
```

Un nombre entre crochets donne la position d'un élément dans le jeu sélectionné.

```
/nuages/nuage[last()] selectionne le dernier  
element 'nuage' de 'nuages'.
```

Syntaxe générale

Les prédicats

Il est possible de combiner ces tests à l'aide des opérateurs logiques classiques (**and**, **or** et **not**) ou de les enchaîner.

```
child::variete[@precision='translucide' or
               @precision='opaque']
child::espece[contains(text(), 'flocus') and
               position()=1]
child::espece[contains(text(), 'flocus')][
               position()=1]
```

Les deux dernières requêtes ne sont pas équivalentes :

- la première renvoie le premier fils 'espece' si celui-ci contient le texte 'flocus',
- la seconde sélectionne tous les fils 'espece' qui contiennent le texte **flocus** et parmi ceux-ci renvoie le premier.

Syntaxe générale

Les fonctions

Le langage
d'interrogation
XPath

Emeline DOREY

Présentation

Syntaxe générale

Evaluation

Les axes

Les filtres

Les prédicats

Les fonctions

Syntaxe abrégée

Exemple

- Positions relatives et information locale :
 - **position()** : position dans le contexte ;
 - **name()** : retourne le nom de l'élément ;
 - **count()** : compte le nombre d'éléments sélectionnés ;
 - **last()** : indicateur de dernière position.
- Fonctions booléennes : **and**, **or**, **not** ;
- Opérateurs : **mod**, **>**, **<=**, ...
- Fonctions de chaîne : **contains**, **substringbefore**, **stringlength**, ...
- Fonctions d'environnement : **normalize-space**. Elle supprime les espaces de début et de fin et remplace les séquences d'espaces blancs par un seul espace.

Syntaxe générale

Les fonctions

Le langage
d'interrogation
XPath

Emeline DOREY

Présentation

Syntaxe générale

Evaluation

Les axes

Les filtres

Les prédicats

Les fonctions

Syntaxe abrégée

Exemple

Exemples :

```
//*[name()='visibilite'] : elements 'visibilite'  
,  
//*[count(espece)='4'] : elements ayant 4  
enfants 'espece'  
//*[count(*)='2'] : elements ayant deux enfants  
//*[stringlength(name()='3')] : retourne tous  
les elements dont le nom a trois caracteres.
```


Syntaxe abrégée

Il existe une notation abrégée permettant de naviguer dans l'arborescence d'un document :

- **element** : sélectionne tous les éléments **element** fils du noeud courant (child : :element).
- ***** : sélectionne tous les éléments fils du noeud courant.
- **/** : représente l'élément racine.
- **//** : représente n'importe quel descendant de l'élément racine, donc tous les éléments (descendant-or-self : :node()).
- **.** : représente l'élément courant (self : :node()).
- **..** : permet de remonter d'un niveau dans l'arborescence du document par rapport à l'élément courant (parent : :node()).

Syntaxe abrégée

- **/element** : sélectionne tous les éléments element sous l'élément racine ().
- **./element** : sélectionne tous les éléments element sous l'élément courant (following : :element).
- **../element** : sélectionne tous les éléments element sous l'élément parent du noeud courant (preceding : :element).
- **//element** : sélectionne tous les éléments element descendants du noeud courant à quelque niveau de profondeur que ce soit.
- **@attribut** : sélectionne tous les attributs attribut du noeud courant (attribute : :attribut).
- **|** : correspond à un ou.

Syntaxe abrégée

Le langage
d'interrogation
XPath

Emeline DOREY

Présentation

Syntaxe générale

Evaluation

Les axes

Les filtres

Les prédicats

Les fonctions

Syntaxe abrégée

Exemple

Exemples d'expressions équivalentes :

```
parent::constitution  
../constitution
```

```
//composition  
/descendant-or-self::node()/child::composition
```

```
espece[2]  
child::espece[position() = 2]
```

Exemple

Le langage
d'interrogation
XPath

Emeline DOREY

Présentation

Syntaxe générale

Evaluation

Les axes

Les filtres

Les prédicats

Les fonctions

Syntaxe abrégée

Exemple

Soit une arborescence du type :

```
<bibliotheque>
  <livre annee="2007">
    <titre>XPath par le pratique</titre>
    <auteur>Thomas Frenz</auteur>
    <editeur>
      <nom>Paris Editions</nom>
      <ville>Paris</ville>
    </editeur>
    <prix>15 euros</prix>
  </livre>
</bibliotheque>
```

Exemple

- **//editeur/parent** : :* renvoie le parent du noeud editeur, c'est à dire livre.
- **/livre/editeur/descendant** : :* renvoie les descendants de editeur, c'est à dire nom et ville.
- **//editeur/following-sibling** : :* renvoie le noeud prix.
- **//editeur/preceding-sibling** : :* renvoie les noeuds auteur, titre.
- **//auteur/following** : :* renvoie les noeuds editeur, nom, ville, prix.
- **//auteur/preceding** : :* renvoie le noeud titre.