

CSI 3540

Structures, techniques et normes du Web

Représentation des données du Web en XML

Objectifs :

- Maîtriser le langage XML 1.0
- Savoir utiliser les outils standard pour le traitement de XML

Lectures :

- Web Technologies (2007) § 7
Pages 402-418

Plan

1. Transformations

1. XPath

2. XSL

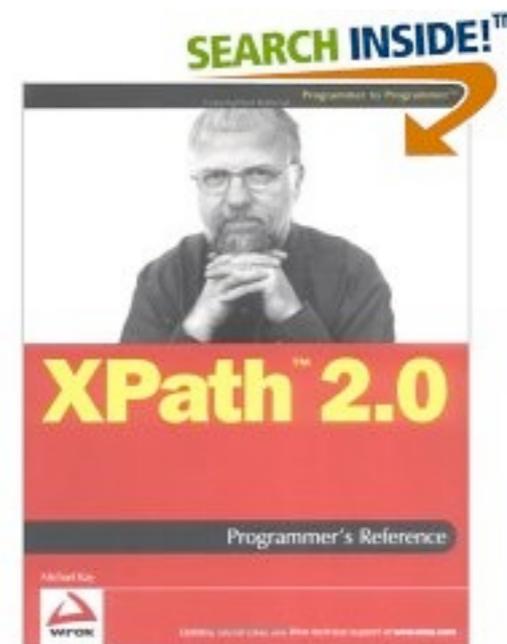
3. XSLT

XPath

- La présentation couvre **XPath 1.0**
 - qui est la version du **manuel du cours**
 - version supportée par **Java 1.5 et 1.6**
 - **<http://www.w3.org/TR/xpath>**
- **XPath 2.0** est une recommandation datée du 23 janvier 2007

XPath

- XPath 2.0 est un surensemble de XPath 1.0 qui offre des fonctionnalités accrues, notamment pour les types (XML Schema)
- Michael Kay (2004) XPath 2.0 Programmer's Reference. Wrox, 552 pages.



Un document XML

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

```
<catalog>
```

```
<album>
```

```
<name>OK Computer</name>
```

```
<track>
```

```
<name>Karma Police</name>
```

```
<artist>Radiohead</artist>
```

```
<genre>Alternative</genre>
```

```
</track>
```

```
<track>
```

```
<name>Exit Music (For A Film)</name>
```

```
<artist>Radiohead</artist>
```

```
<genre>Alternative</genre>
```

```
</track>
```

```
</album>
```

```
<album>
```

```
<name>Dark Side Of The Moon</name>
```

```
<track>
```

```
<name>Us And Them</name>
```

```
<artist>Pink Floyd</artist>
```

```
<genre>Progressive</genre>
```

```
</track>
```

```
<track>
```

```
<name>Brain Damage</name>
```

```
<artist>Pink Floyd</artist>
```

```
<genre>Progressive</genre>
```

```
</track>
```

```
</album>
```

```
...
```

```
...
```

```
<album>
```

```
<name>Requiem</name>
```

```
<track>
```

```
<name>Kyrie</name>
```

```
<artist>Wiener Philharmoniker - Herbert von Karajan</artist>
```

```
<composer>Wolfgang Amadeus Mozart</composer>
```

```
<genre>Classical</genre>
```

```
</track>
```

```
</album>
```

```
<album>
```

```
<name>African Guitar Summit</name>
```

```
<track>
```

```
<name>Obaa Y Ewa</name>
```

```
<artist>Pa Joe</artist>
```

```
<genre>World</genre>
```

```
</track>
```

```
<track>
```

```
<name>Mwajuma</name>
```

```
<artist>Adam Solomon</artist>
```

```
<genre>World</genre>
```

```
</track>
```

```
</album>
```

```
</catalog>
```

XPath

- **`/child::catalog/child::album/child::name`**
 - sélectionne 4 noeuds
- **`/descendant::name`**
 - sélectionne 11 noeuds
- **`//child::track/child::composer`**
 - sélectionne 1 noeud
- **`//child::track[child::artist='Radiohead']`**
 - sélectionne 2 noeuds
- **`//child::track[attribute::id="3"]/following-sibling::track`**
 - sélectionne 1 noeud
- **`//self::name/child::text()`**
 - sélectionne 11 noeuds

<oxygen/> - [//Users/turcotte/iDisk/teaching/csi3540/2008/Sandbox/Src/09_XML/07_music_lib_xpath/music_library.xml]

XPath 2.0 `//child::track[artist='Radiohead']`

Project: newProject.xpr

music_library.xml

```

1 <?xml version="1.0" encoding="UTF-8"?>
2 <catalog>
3   <album>
4     <name>OK Computer</name>
5     <track id="1">
6       <name>Karma Police</name>
7       <artist>Radiohead</artist>
8       <genre>Alternative</genre>
9     </track>
10    <track id="2">
11      <name>Exit Music (For A Film)</name>
12      <artist>Radiohead</artist>
13      <genre>Alternative</genre>
14    </track>
15  </album>
16  <album>
17    <name>Dark Side Of The Moon</name>
18    <track id="3">
19      <name>Us And Them</name>

```

Attributes: track

Attribute	Value
id	1

Outline:

- catalog
 - album
 - name OK Computer
 - track "1"
 - track "2"
 - album
 - album

Info: Description - 2 items

Info	Description - 2 items	Resource	System I
-	/catalog[1]/album[1]/track[1] - id="1"	music_library.xml	file:/Use
-	/catalog[1]/album[1]/track[2] - id="2"	music_library.xml	file:/Use

XPath - music_library.xml

StatusBar: /Users/turcotte/iDisk/teaching/csi3540/2008/Sandbox/Src/09_XML/0... XPath - successful U+003C 5:5

widgets.xml

```
<?xml version='1.0' encoding='UTF-8'?>
<widgets>
  <widget>
    <manufacturer name="Acme"/>
    <dimensions width="1.0" height="2.0"/>
  </widget>
  <widget>
    <manufacturer name="K-Tel"/>
    <dimensions width="1.5" height="1.5"/>
  </widget>
</widgets>
```

javax.xml.xpath

- Depuis la version 1.5, Java supporte XPath (version 1.0)

```
import javax.xml.xpath.*;
import org.xml.sax.InputSource;
import org.w3c.dom.NodeList;
import org.w3c.dom.Document;
...
```

```
XPath xpath = XPathFactory.newInstance().newXPath();
```

```
String expression = "/child::widgets/child::widget";
```

```
QName returnType = XPathConstants.NODESET;
```

```
InputSource inputSource = new InputSource( "widgets.xml" );
```

```
NodeList nodes;
```

```
node = (NodeList) xpath.evaluate( expression, inputSource, returnType );
```

XPath

- Conçu afin d'être le langage **commun** pour l'**adressage** des parties (éléments, attributs, texte, etc.) d'un document XML
- L'adressage est défini par rapport à l'arbre d'analyse du document
- **Au contraire du DOM, les attributs sont des noeuds de l'arbre**
- **La racine est un noeud supplémentaire qui désigne l'élément racine du document**

XPath

- XPath **supporte les espaces de nommage**, ainsi, le nom d'un noeud est un couple composé d'un espace de noms et d'un nom local ; URI désignant l'espace de nommage peut avoir la valeur null

XPath : contexte

- Conçu afin d'être le langage **commun** aux applications devant manipuler des documents XML : **XSLT, bases de données XML, Schematron, xfind, ...**
- L'évaluation d'une expression se fait dans un certain contexte (**environnement d'évaluation**) qui, au départ, est fourni par l'application

XPath : 4 types d'objets

- **Ensemble de noeuds**
(liste non ordonnée de noeuds sans doublons)
- Booléen
- Nombre réel
- Chaîne de caractères

XPath : chemin de localisation

- Le **chemin de localisation** est le type d'expression prépondérant
- **Contexte** = noeud de contexte ou tout simplement **noeud contextuel**
- Le **chemin de localisation** sélectionne un ensemble de noeuds par rapport au **noeud contextuel** (relativement)
- L'application hôte fournit le noeud contextuel (de départ)

DocumentBuilder builder;

`builder = DocumentBuilderFactory.newInstance().newDocumentBuilder();`

Document document = `builder.parse(new File("widgets.xml"));`

XPath xpath = `XPathFactory.newInstance().newXPath();`

String expression = `"/child::widgets/child::widget";`

QName returnType = `XPathConstants.NODE;`

Node widgetNode = `(Node) xpath.evaluate(expression, document, returnType);`

`xpath = XPathFactory.newInstance().newXPath();`

`expression = "child::manufacturer";`

Node manufacturerNode;

`manufacturerNode = (Node) xpath.evaluate(expression, widgetNode, returnType);`

noeud contextuel



```
import java.io.*;
```

```
import javax.xml.xpath.*;
```

```
import javax.xml.parsers.DocumentBuilderFactory;
```

```
import javax.xml.parsers.DocumentBuilder;
```

```
import javax.xml.parsers.ParserConfigurationException;
```

```
import javax.xml.namespace.QName;
```

```
import org.xml.sax.InputSource;
```

```
import org.xml.sax.SAXException;
```

```
import org.w3c.dom.NodeList;
```

```
import org.w3c.dom.Node;
```

```
import org.w3c.dom.Document;
```

XPath : chemin de localisation

- **`/child::catalog/child::album[position()=1]`**
- Un **chemin de location** (location path) est une séquence d'**étapes de localisation**, séparées par le symbole «/»

XPath : chemin de localisation

étape de localisation

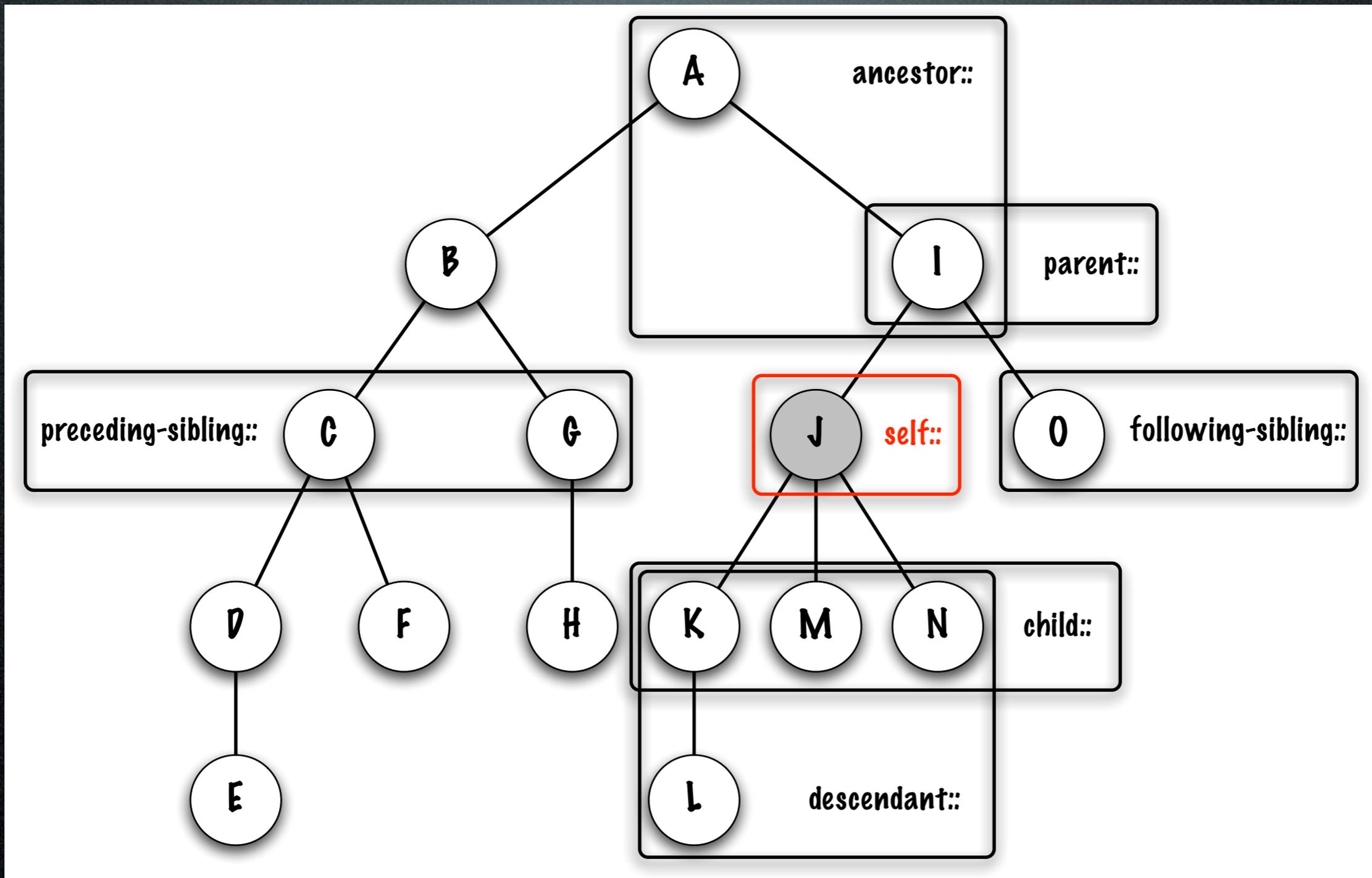
axe **noeud** **prédicat**
 de test

- **child::catalog/child::album [position()=1]**
- Chaque **étape** spécifie **1)** un **axe** (type de relation arborescente, sens de la recherche), **2)** un **noeud de test**, et optionnellement **3)** un **prédicat**

Relatif et absolu

- Un **chemin de localisation relatif** est une séquence d'une ou plusieurs étapes de localisation séparées par le symbole «/»
- Un **chemin de localisation absolu** débute par le symbole «/», qui sélectionne la racine du document, et est optionnellement suivi d'un chemin relatif

1. Axe



1. Axe

- l'**axe enfant** sélectionne les enfants directs du noeud contextuel
- l'**axe descendant** sélectionne tous les noeuds descendants sauf les attributs
- l'**axe parent** contient le parent du noeud contextuel, sauf s'il s'agit de la racine
- l'**axe ancestor** sélectionne tous les ancêtre, contient toujours la racine, sauf si le noeud contextuel est la racine

1. Axe

- l'**axe cible suivante**, sélectionne tous les noeuds du même niveau qui suivent le noeud contextuel
- l'**axe cible précédente**, même que ci-haut, mais pour les noeuds qui précèdent
- l'**axe attribut**, sélectionne les attributs du noeud contextuel (vide si l'élément n'a pas d'attributs)

1. Axe

- **l'axe réflexif (self::)** = noeud contextuel
- **descendant-or-self::** et **ancestor-or-self::**
- **preceding::, following::** sélectionne les noeuds qui précèdent ou suivent dans le document, mais pas les descendants
- **namespace::** sélectionne tous les noeuds du même espace de nommage que le noeud contextuel

2. test (de nom) de noeud

- Un **test de noeud** est vrai si le type du noeud testé est le même que celui du test
 - **child::artist**, l'axe **child** sélectionne tous les noeuds enfants du noeud contextuel, le test **artist** sera vrai pour les noeuds enfants dont le type est **artist**

2. test de noeud

- Le type spécifié doit correspondre à l'axe de cette étape de localisation:
 - Pour les axes **child::**, **parent::**, ... le type doit être **un nom d'élément**
 - Pour l'axe **attribute::**, la valeur spécifiée doit être **un nom d'attribut**

2. test de noeud

- «*» est vrai pour n'importe quel noeud
 - **child::*** sélectionne tous les noeuds enfants
 - **attribute::*** sélectionne tous les attributs du noeud contextuel

2. test (de type) de noeud

- **text()** sélectionne les noeuds «textuels»
- **node()** sélectionne tous les éléments
- **comment()** sélectionne les noeuds de type «commentaire»

3. prédicat

- Entre « [» et «] » et il peut y en avoir plusieurs
- **Filter** un ensemble de noeuds :
 - **[child::name]**
 - **[child::name="value"]**
 - **[attribute::name]**
 - **[attribute::name='value']**
 - **[position() <= 3], [position() = last()]**

Les prédicat de proximité dépendent du sens de l'axe (vers l'arrière ou vers l'avant)

Remarque

- Une étape de location peut contenir plusieurs prédicats, l'évaluation de chaque prédicat produit une liste filtrée pour le prochain prédicat (de gauche à droite)
- Les expressions **XPath** sont abondamment utilisées dans les documents **XSLT**; on doit alors utiliser des appels d'entités pour les symboles tels que < des prédicats

Syntaxe abrégée (compacte)

- Pour alléger la notation, les abréviations suivantes ont été définies
 - **child::** peut être omis, de sorte que **/child::catalog/child::album [position ()=1]** et **/catalog/album [position ()=1]** sont deux expressions équivalentes
 - de même **position ()=** peut être omis **/catalog/album [1]**

Syntaxe abrégée (compacte)

- «.» signifie **self::**
- «..» signifie **parent::**
- «//» signifie **/descendant-or-self::node()**

Syntaxe abrégée (compacte)

- «@» signifie **attribute::**
- ainsi
 [attribute::color="red"] et
 [@color="red"]
sont deux prédicats équivalents

Les étapes de localisation

- **Lorsqu'un chemin est constitué de plusieurs étapes,**
 - l'évaluation se fait de gauche à droite
 - Chaque étape produit une liste de noeuds
 - Les noeuds de cette liste servent tour à tour de noeud contextuel pour l'étape suivante.

```

<?xml version="1.0" encoding="UTF-8"?>
<catalog>
  <album>
    <name>OK Computer</name>
    <track id="1">
      <name>Karma Police</name>
      <artist>Radiohead</artist>
      <genre>Alternative</genre>
    </track>
    <track id="2">
      <name>Exit Music (For A Film)</name>
      <artist>Radiohead</artist>
      <genre>Alternative</genre>
    </track>
  </album>
  <album>
    <name>Dark Side Of The Moon</name>
    <track id="3">
      <name>Us And Them</name>
      <artist>Pink Floyd</artist>
      <genre>Progressive</genre>
    </track>
    <track id="4">
      <name>Brain Damage</name>
      <artist>Pink Floyd</artist>
      <genre>Progressive</genre>
    </track>
  </album>
  ...

```

```

...
<album>
  <name>Requiem</name>
  <track id="5">
    <name>Kyrie</name>
    <artist>Wiener Philharmoniker - Herbert von Karajan</artist>
    <composer>Wolfgang Amadeus Mozart</composer>
    <genre>Classical</genre>
  </track>
</album>
<album>
  <name>African Guitar Summit</name>
  <track id="6">
    <name>Obaa Y Ewa</name>
    <artist>Pa Joe</artist>
    <genre>World</genre>
  </track>
  <track id="7">
    <name>Mwajuma</name>
    <artist>Adam Solomon</artist>
    <genre>World</genre>
  </track>
</album>
</catalog>

```

Exemples

- **/catalog/album/track/name**
- **/catalog/album [name="Requiem"]**
- **//name/..** : sélectionne les éléments de type "album" et "track"
- **//track[@id]** : retourne tous les éléments de type **track** ayant un attribut «id»

Exemples

- **//track[child::name]**
 - Sélectionne les éléments **track** ayant un descendant immédiat **name**
- **//track/name**
 - Sélectionne les éléments **name** qui sont des descendants immédiats de l'élément **track**

La valeur de position dépend de l'axe choisi

- **//track [position()=last()]** :
retourne les éléments **track** dont la valeur de l'attribut **id** est 2, 4, 5 et 7
- L'expression ci-haut équivaut à celle-ci :
/descendant-or-self::node()/track [position()=last()]
Étape 1 sélectionne tous les noeuds de l'arbre.
Étape 2, chaque noeud de la liste de l'étape 1 sert de noeud contextuel. Chaque nouveau noeud contextuel génère une nouvelle liste de noeuds que l'on filtre pour retenir que le noeud à la dernière position.

Examples

- `sum(//track[@duration])`
- ...

Examples

- **/child::body/child::strong**

- sélectionne s1

- **/child::body/descendant::strong**

- sélectionne s1, ..., s5

- **//para/strong**

- sélectionne s2, s3 et s5

- **//para[strong]**

- sélectionne p1, p3

```
<?xml version="1.0" encoding="UTF-8"?>
<body>
  <strong id="s1">Exit Music (For A Film)</strong>
  <para id="p1">
    Wake from <strong id="s2">your sleep</strong>,
    the drying of your tears,
    Today we escape, <strong id="s3">we escape</strong>.
  </para>
  <para id="p2">
    Sing us a <em><strong id="s4">song</strong></em>,
    a song to keep us warm,
    there's such a chill, such a chill.
  </para>
  <para id="p3">
    And now we are one
    in everlasting <strong id="s5">peace</strong>.
  </para>
</body>
```

Remarque

- Bien que les chemins de localisation ressemblent aux chemins dans un arbre de répertoires Unix, il y a des différences importantes :
 - Chaque étape produit un ensemble de noeuds qui sont possiblement situés à des niveaux différents de l'arbre du document source
 - **//foo/bar**

XPath côté-client

- **XPath** est-il supporté par les navigateurs?
- Bien sûr, puisque les feuilles de styles (**XSL**) sont supportées et qu'elles utilisent **XPath**!
- Est-ce **JavaScript** supporte **XPath**?

XPath et JavaScript

```
function init() {
```

```
    var nextElem;
```

```
    var nodes = document.evaluate( '//strong', document, null, XPathResult.ANY_TYPE, null );
```

```
    var elem = nodes.iterateNext();
```

```
    while ( elem ) {
```

```
        elem.style.color = "red";
```

```
        var parent = document.evaluate( '../..', elem, null, XPathResult.FIRST_ORDERED_NODE_TYPE, null );  
        parent.singleNodeValue.style.color = "orange";
```

```
        elem = nodes.iterateNext();
```

```
    }
```

```
    return;
```

```
}
```

```

<head>
<title>La passante ; Emile NELLIGAN (1879-1941)</title>
<meta http-equiv="Content-Type" content="text/html; charset=latin1" />
<script type="text/javascript" src="index.js">
</script>
</head>
<body style="font-size:large" onload="init();">
<h1>Émile NELLIGAN (1879-1941)</h1>
<h2>La passante</h2>
<pre>
(1) Hier, j'ai vu passer, comme une ombre qu'on plaint,
En un grand parc obscur, une femme voilée :
Funèbre et singulière, elle s'en est allée,
Recélant sa fierté sous son masque opalin.
</pre>
<pre>
(2) Et rien que d'un regard, par ce soir cristallin,
J'eus deviné bientôt <i>sa douleur <strong>refoulée</strong>
Puis elle disparut en quelque noire allée
Propice au deuil profond dont son coeur était plein.
</pre>
<ul>
<li>Grand</li>
<li><strong>Jacques</strong> 1964</li>
</ul>
<pre>
(3) Ma jeunesse est pareille à la pauvre passante :
Beaucoup la croiseront ici-bas dans la sente
Où la vie à la tombe âprement nous conduit;
</pre>
<pre>
(4) Tous la verront passer, feuille sèche à la brise
Qui tourbillonne, tombe et se fane en la nuit ;
Mais nul ne l'aimera, nul ne l'aura comprise.
</pre>
</body>
</html>

```

La passante ; Emile NELLIGAN (1879-1941)

index.html Google

Émile NELLIGAN (1879-1941)

La passante

(1) Hier, j'ai vu passer, comme une ombre qu'on plaint,
En un grand parc obscur, une femme voilée :
Funèbre et singulière, elle s'en est allée,
Recélant sa fierté sous son masque opalin.

(2) Et rien que d'un regard, par ce soir cristallin,
J'eus deviné bientôt *sa douleur **refoulée*** ;
Puis elle disparut en quelque noire allée
Propice au deuil profond dont son coeur était plein.

- Grand
- Jacques 1964

(3) Ma jeunesse est pareille à la pauvre passante :
Beaucoup la croiseront ici-bas dans la sente
Où la vie à la tombe âprement nous conduit;

(4) Tous la verront passer, feuille sèche à la brise
Qui tourbillonne, tombe et se fane en la nuit ;
Mais nul ne l'aimera, nul ne l'aura comprise.

Ressources

- Langage de balisage extensible (XML) 1.0 [<http://pages.videotron.com/fyergeau/w3c/xml10/REC-xml-19980210.fr.html>] 2007
- Les espaces de nommage dans XML 1.1 [<http://www.yoyodesign.org/doc/w3c/xml-names11>] 2007

Ressources (suite)

- Java API for XML Processing (JAXP) Specification 1.3 [<http://java.sun.com/xml/downloads/jaxp.html>] 2007
- SAX [<http://www.saxproject.org>] 2007
- <oxygen/> XML Editor & XSLT Debugger [<http://www.oxygenxml.com>] 2007

Ressources (suite)

- The GNU JAXP Project [<http://www.gnu.org/software/classpathx/jaxp/jaxp.html>] 2007
- Langage XML Path (XPath) Version 1.0 [<http://xmlfr.org/w3c/TR/xpath>] 2007
- Transformations XSL (XSLT) [<http://xmlfr.org/w3c/TR/xslt/>] 2007
- XML: Looking at the Forest Instead of the Trees par Guy Lapalme [<http://www.iro.umontreal.ca/~lapalme/ForestInsteadOfTheTrees/>] 2007

Ressources (suite)

- Des exemples chez Mozilla :
https://developer.mozilla.org/en/Introduction_to_using_XPath_in_JavaScript