

Documents structurés Cours 4

Nassim ZELLAL

XPath - 1

Les expressions XPath

- Le langage XPath s'appuie sur des expressions désignant un ou plusieurs nœuds, en exprimant des chemins de localisation dans l'arbre d'un document XML.
- Une expression XPath représente donc un chemin de localisation pour naviguer dans l'arbre d'un document XML et sélectionner un ensemble de nœuds.
- Tester les expressions XPath avec « Exchanger XML Editor » :
 - https://www.softpedia.com/get/Internet/WEB-Design/HTML-Editors/Exchanger-XML-Editor.shtml

Le nœud racine (root node)

- Dans l'expression XPath « /livre/chapitre », la première barre oblique / (slash) indique que la recherche commence à partir du nœud racine (ou nœud de document) de l'arbre du document XML.
- Dans l'exemple ci-dessous, l'enfant direct du <u>nœud racine</u> / est le nœud d'élément racine (ou élément racine ou élément de document) <<u>livre></u>. Ce dernier est le seul enfant du <u>nœud racine</u>.
- Le <u>nœud racine</u> / contient tout le document XML et inclut donc le nœud d'élément racine <<u>livre</u>>.
- Le <u>nœud racine</u> représente donc tout le document.

</livre>

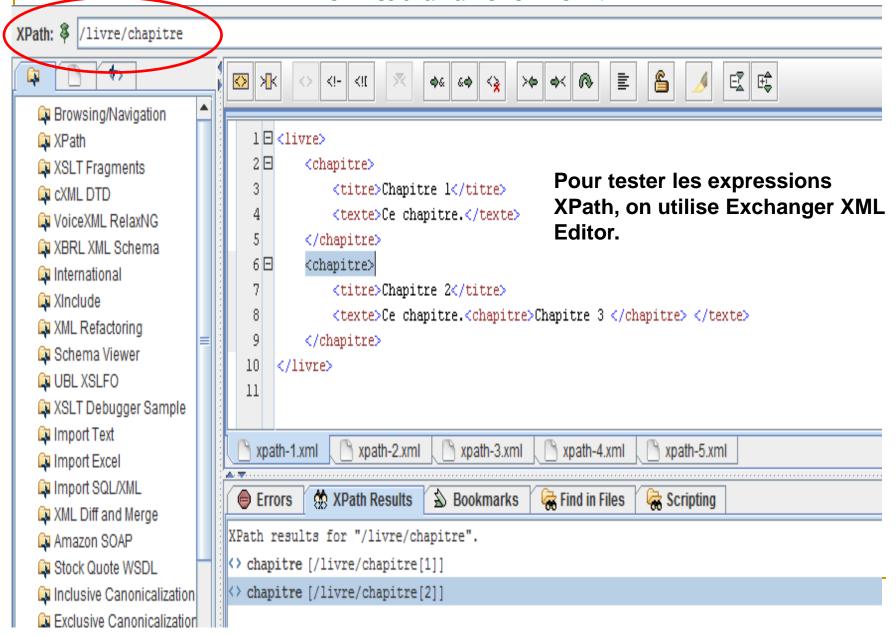
 L'évaluation de l'expression XPath composée d'un / retourne le document XML tout entier.

- Localiser les nœuds « chapitre » dans le document XML « xpath-1.xml » avec l'expression : /livre/chapitre
- livre>

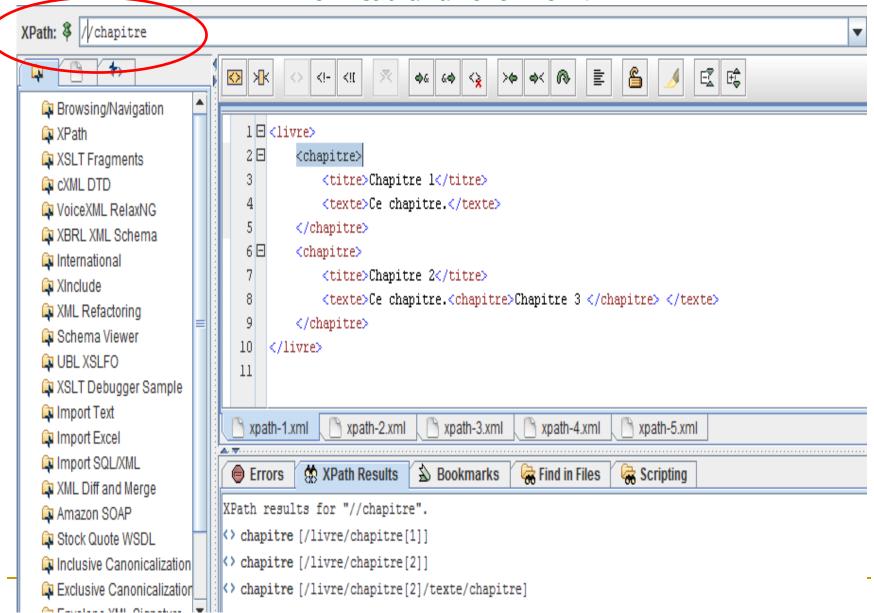
```
<chapitre>
```

```
<titre>Chapitre 1</titre>
```

- <texte>Ce chapitre.</texte>
- </chapitre>
- <chapitre>
- <titre>Chapitre 2</titre>
- <texte>Ce chapitre.
- <chapitre>Chapitre 3<chapitre/>
- </texte>
- </chapitre>
- Voir xpath-1.xml



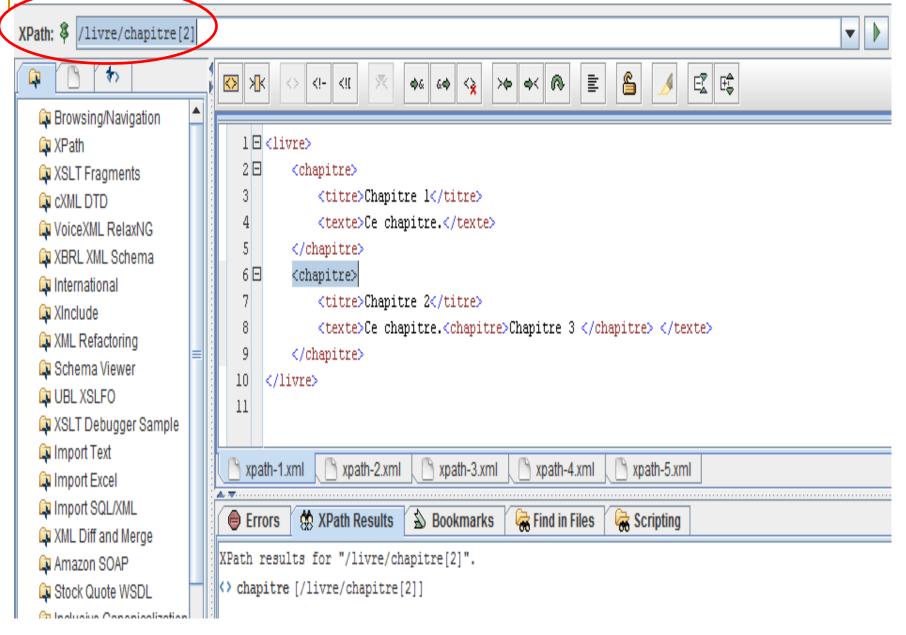
- Autre syntaxe : // parcourir <u>récursivement</u> tous les nœuds <u>descendants</u> à partir du nœud d'élément courant (ou contextuel).
- <u>Exemple</u>: //chapitre permet d'extraire (sélectionner) tous les nœuds d'élément « chapitre ».
- livre>
- <chapitre>
- <titre>Chapitre 1</titre>
- <texte>Ce chapitre.</texte>
- </chapitre>
- <chapitre>
- <titre>Chapitre 2</titre>
- <texte>Ce chapitre.<chapitre>Chapitre 3 </chapitre>
 - </texte>
- </chapitre>
- Voir xpath-1.xml



Pour orienter la recherche sur la <u>position</u> d'un nœud, on utilise les crochets []: /livre/chapitre[2]

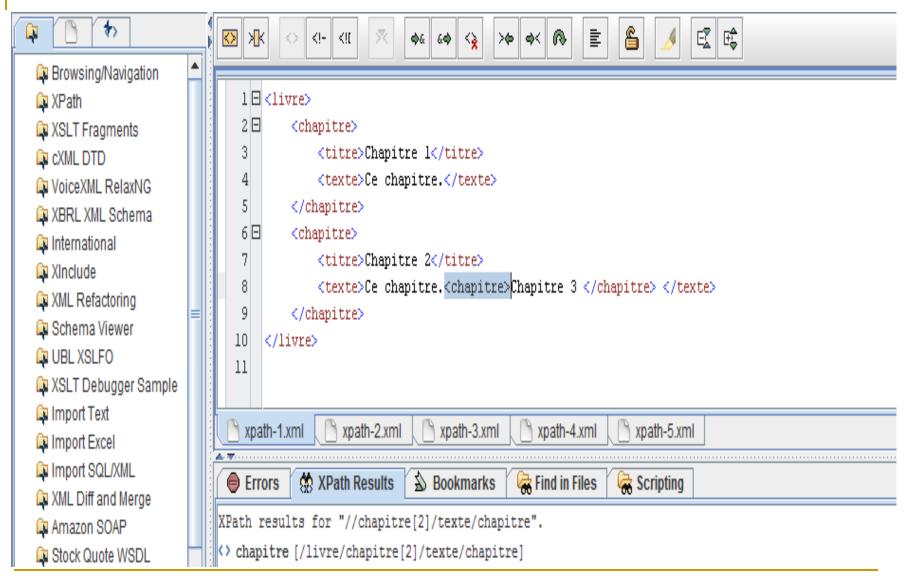
```
vre>
     <chapitre>
             <titre>Chapitre 1</titre>
             <texte>Ce chapitre.</texte>
     </chapitre>
     <chapitre>
             <titre>Chapitre 2</titre>
             <texte>Ce chapitre.<chapitre>Chapitre 3 </chapitre>
</texte>
     </chapitre>
</livre>
```

Voir xpath-1.xml



Question : comment retrouver Chapitre 3 ?

- Réponse :
- //chapitre[2]/texte/chapitre
- /livre/chapitre/texte/chapitre



- Pour localiser un nœud d'attribut, on emploie les crochets (Voir aussi les « Prédicats », slides 28, 29 et 30) :
- //personne[@prenom="Jack"]
- Pour tout localiser ==> //personne[@prenom]
- !
- < personne prenom ="Jack" nom= "Nolan" />
- < personne prenom="Jean" nom ="Palvadeau" />
- < personne prenom="Billy" nom = "Madison" />
- < personne prenom="John" nom ="Woo" date_de_naissance="10-121990"/>
- </liste>
- Voir xpath-2.xml

```
XPath results for "//personne[@prenom="Jack"]".
```

> personne [/liste/personne[1]]

```
XPath results for "//personne[@prenom]".

    personne [/liste/personne[1]]

    personne [/liste/personne[2]]

    personne [/liste/personne[3]]

    personne [/liste/personne[4]]
```

- /nœud_courant/@type (sélectionne l'attribut type du nœud nœud d'élément courant)
- Tester avec :
- /liste/personne/@nom

```
XPath results for "/liste/personne/@nom".
@ nom [/liste/personne[1]/@nom]
@ nom [/liste/personne[2]/@nom]
@ nom [/liste/personne[3]/@nom]
@ nom [/liste/personne[4]/@nom]
```

Question : quelle autre expression peut-on utiliser pour obtenir le même résultat ?

- Réponse :
- //personne/@nom

Wildcard (joker)

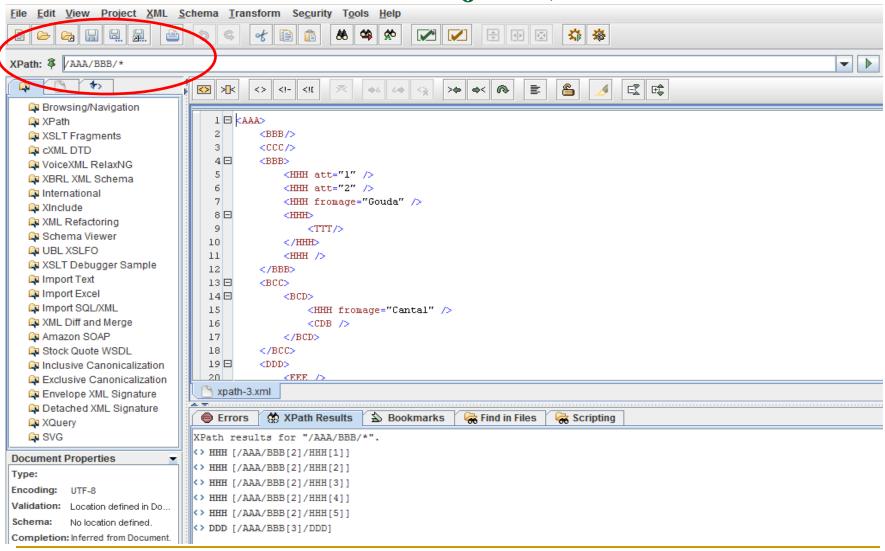
 Le caractère astérisque (Wildcard) appelé aussi joker * correspond à n'importe quel élément contenu dans le nœud d'élément courant.
 Cela correspond donc aux nœuds d'élément sans tenir compte de leur nom.

Exemple 1:

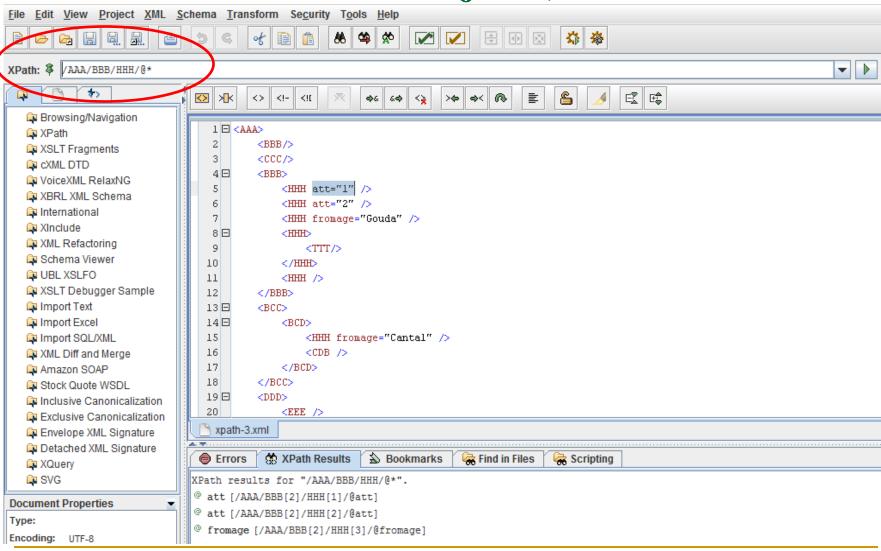
- /AAA/BBB/* (sélectionne tous les enfants directs du nœud d'élément courant /AAA/BBB).
- Exemple 2:
- /AAA/BBB/HHH/@* (sélectionne tous les attributs (ou nœuds d'attribut) du nœud courant /AAA/BBB/HHH).
- Exemple 3:
- /AAA/BBB//* (sélectionne tous les descendants du nœud d'élément courant /AAA/BBB).

Voir xpath-3.xml

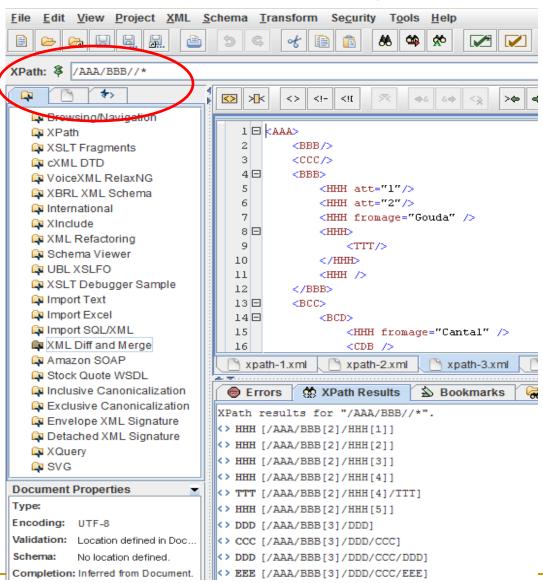




Wildcard (joker) *



Wildcard (joker)



Les nœuds attributs (opérateurs logiques)

- //personne[@prenom="Jack"] or //personne[@prenom="Nolan"]
- Renvoie : true
- //personne[@prenom="Jack"] and //personne[@prenom="Nolan"]
- Renvoie : false
- <te>
- < personne prenom ="Jack" nom= "Nolan" />
- < personne prenom="Jean" nom ="Palvadeau" />
- < personne prenom="Billy" nom = "Madison" />
- < personne prenom="John" nom ="Woo" date_de_naissance="10-12-1990"/>
- </liste>

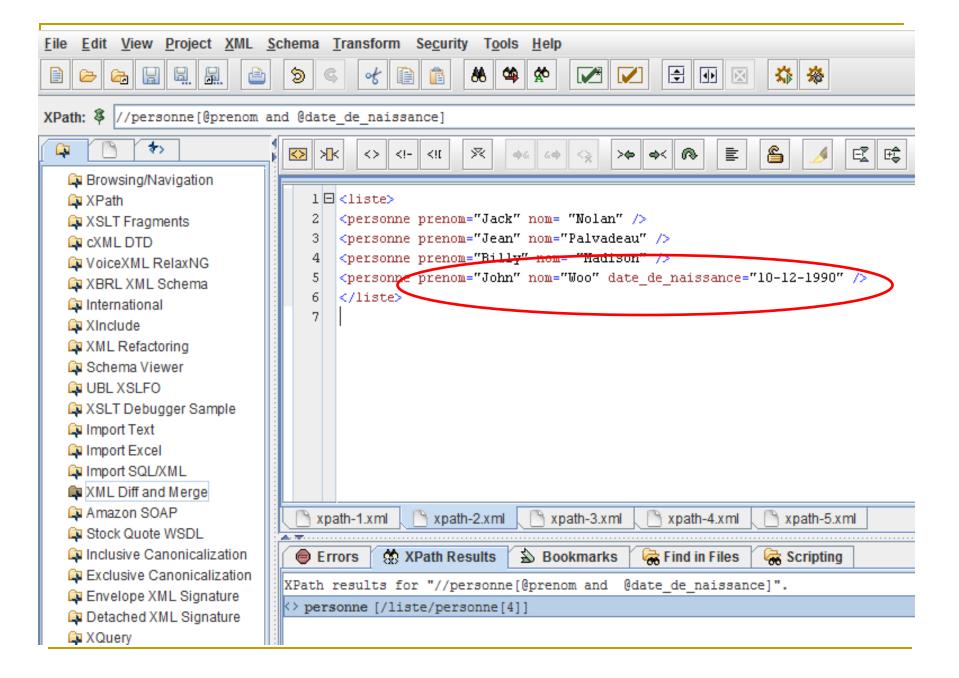
Voir xpath-2.xml

Le nœud d'attribut (opérateurs booléens)

```
XPath results for "//personne[@prenom="Jack"] or //personne[@prenom="Nolan"]
true
```

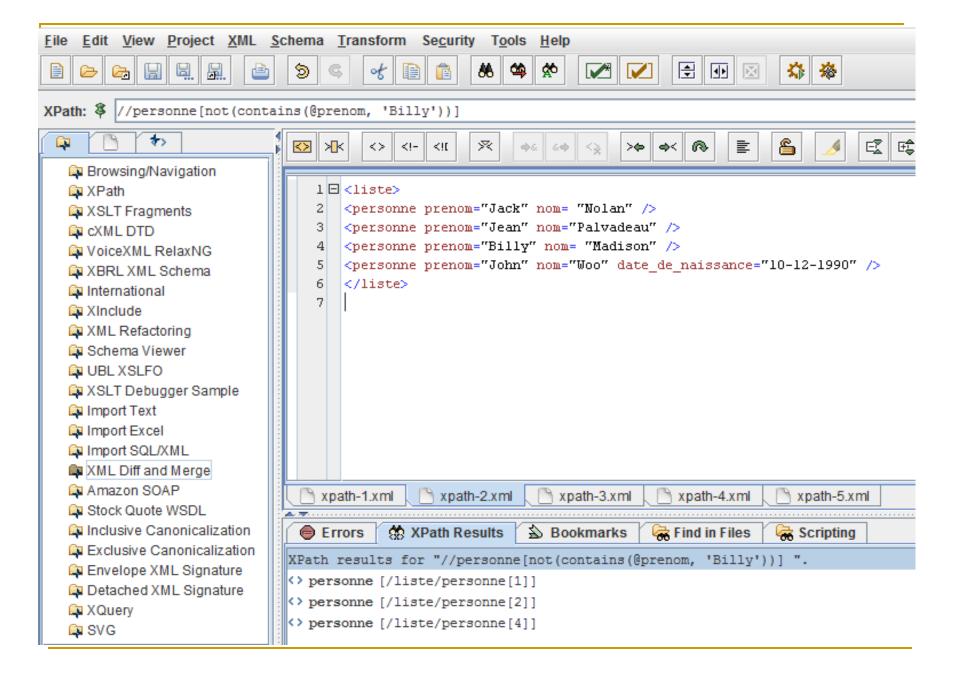
Les nœuds attributs (opérateurs logiques)

- //personne[@prenom and @date_de_naissance]
- <</p>
- < personne prenom ="Jack" nom= "Nolan" />
- < personne prenom="Jean" nom ="Palvadeau" />
- < personne prenom="Billy" nom = "Madison" />
- < personne prenom="John" nom="Woo" date_de_naissance="10-12-1990"/>
- </liste>
- Voir xpath-2.xml



Les nœuds attributs (opérateurs logiques)

- //personne[not(contains(@prenom, 'Billy'))]
- <</p>
- < personne prenom ="Jack" nom= "Nolan" />
- < personne prenom="Jean" nom ="Palvadeau" />
- < personne prenom="Billy" nom = "Madison" />
- < personne prenom="John" nom ="Woo" date_de_naissance="10-12-1990"/>
- </liste>
- Voir xpath-2.xml

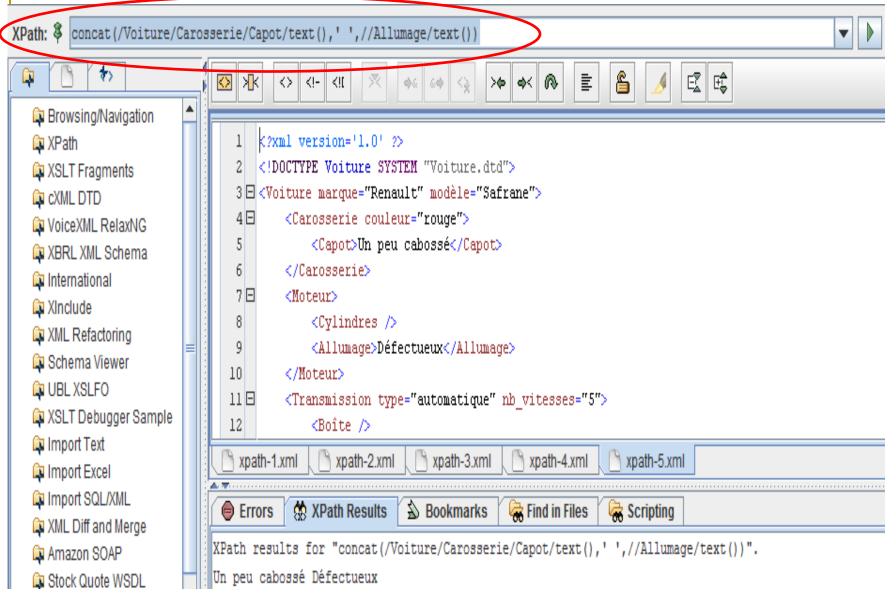


Les fonctions

```
node() → /livre/chapitre/texte/node()
position() → //nom[position()=3]
last() -> //nom[last()] == //nom[position()=last()]
text() → //Capot/text()
count() → count(//nom)
name() → //*[name()= 'BBB']
start-with() → //*[starts-with(name(),'B')]
string-length() >> string-length(//Capot)
contains() → contains(//Capot/text(), 'peu') renvoie un booléen
concat() → concat(/Voiture/Carosserie/Capot/text(),
',//Allumage/text())
```

comment() → //comment()

Les fonctions

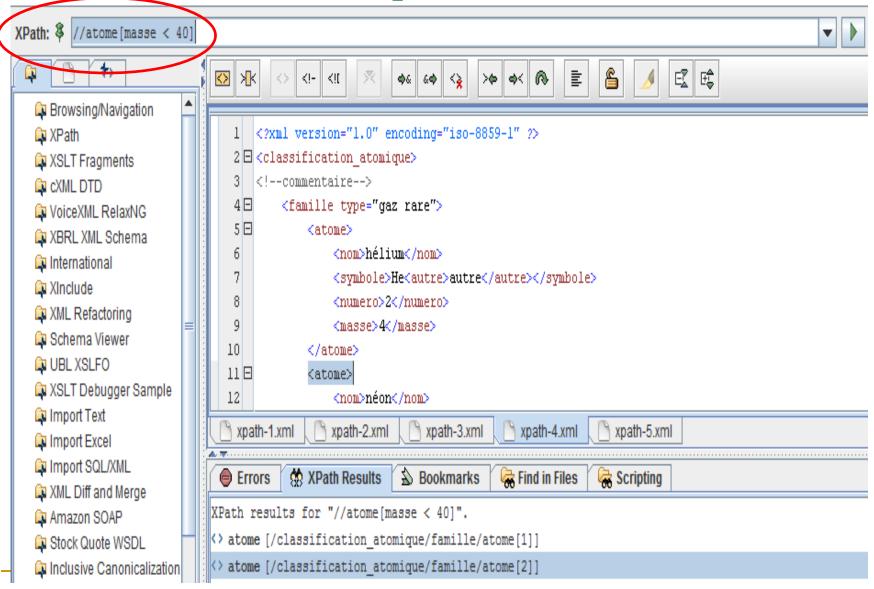


Les prédicats

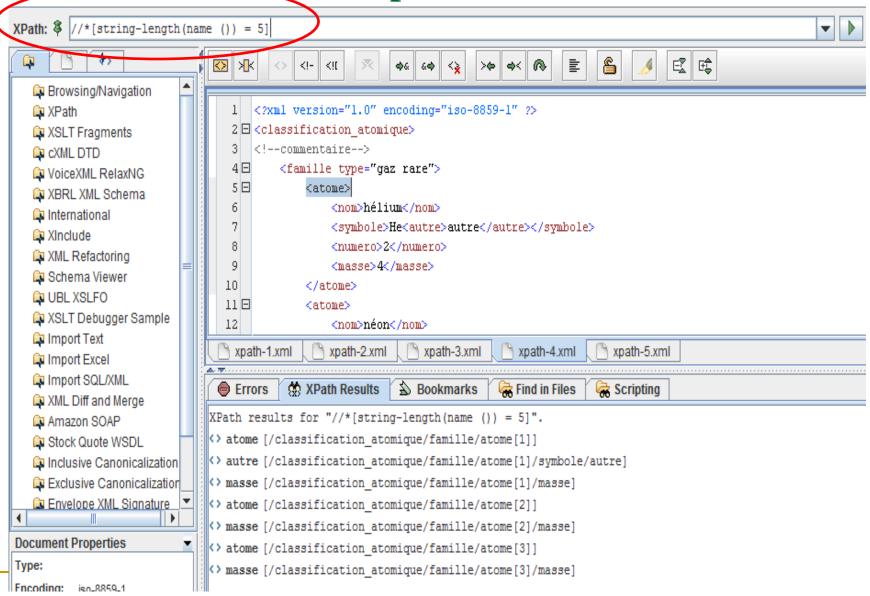
- L'ensemble des nœuds obtenus par un chemin XPath peut être filtré à l'aide d'un prédicat entre crochets []. Cela permet de raffiner notre sélection.
- Exemple :
- //atome[masse < 40]
- //*[string-length(name ()) = 5] (donne les éléments qui ont un nom dont le nombre de caractères est exactement 5)

Voir xpath-4.xml

Les prédicats



Les prédicats



Exercice

- 1-Extraire tous les nœuds de texte du document "xpath-1.xml".
 - Proposer deux solutions.
- 2-Extraire tous les nœuds d'élément "HHH" du document "xpath-3.xml".
- 3-Extraire tous les nœuds d'attribut du document "xpath-3.xml".
- 4-Extraire tous les nœuds d'élément "HHH", qui possèdent un nœud d'attribut du document "xpath-3.xml".
- 5-Extraire le nœud d'élément contenant un nœud d'attribut avec la valeur « Cantal » du document "xpath-3.xml".
- 6-Extraire les nœuds d'élément "atome" ayant une masse strictement inférieure à "20" du document "xpath-4.xml".
- 7-Extraire les nœuds d'élément vides du document "xpath-5.xml".