



Séance 6

Xpath

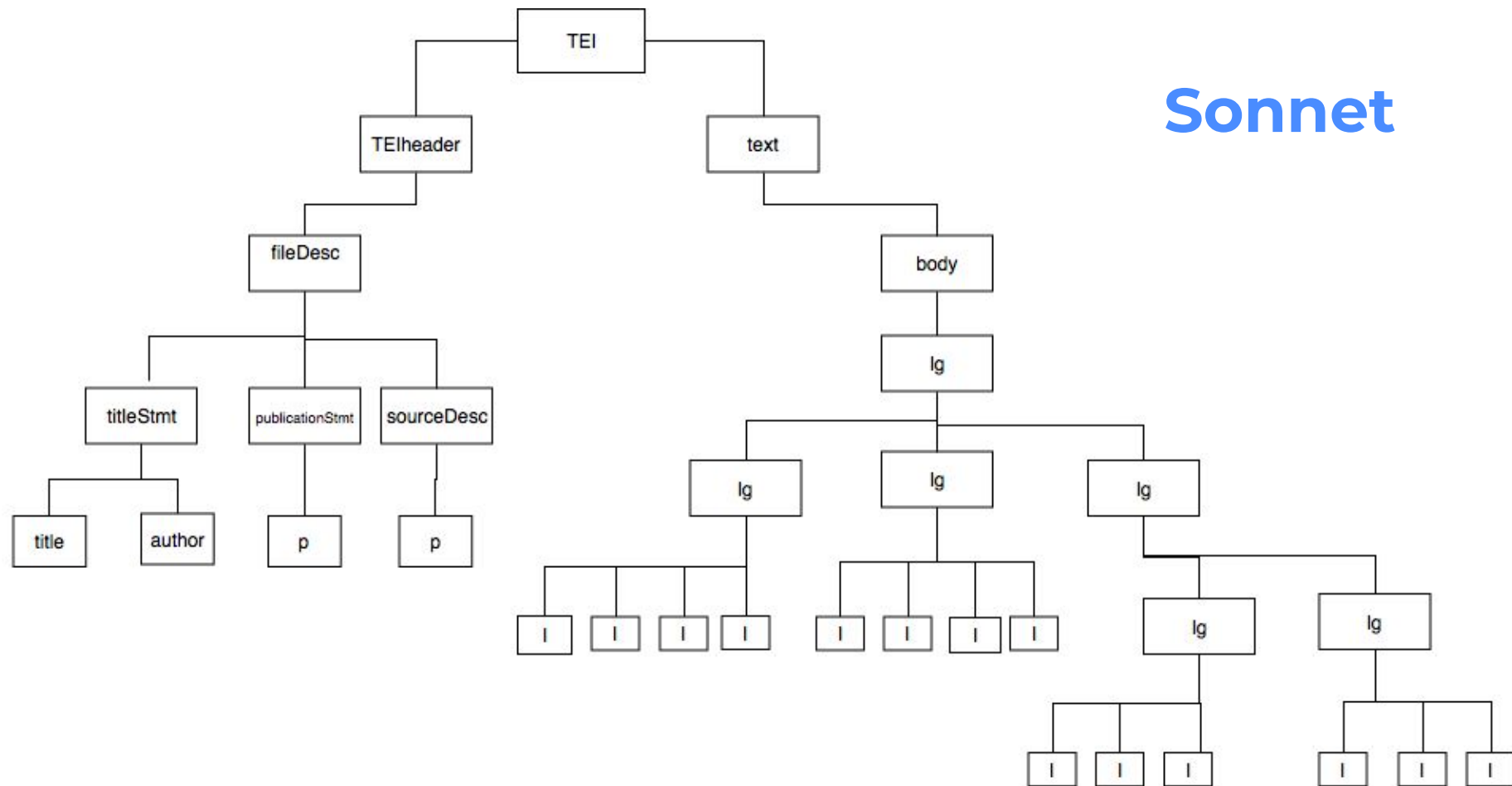
Localisation dans le document XML

A decorative graphic on the left side of the slide consisting of two overlapping squares. The bottom-left square is a dark blue, and the top-right square is a lighter blue, creating a cross-like shape.

Structure en arbre

Maîtriser la hiérarchie XML

Sonnet



Sélectionner des nœuds

Le langage Xpath fournit une syntaxe pour sélectionner un nœud ou un ensemble de nœud au sein d'un document XML.

A decorative graphic on the left side of the slide consisting of two overlapping squares. The bottom-left square is a dark blue, and the top-right square is a lighter blue, creating a cross-like shape.

Qu'est-ce que le Xpath ?

Les principes du langage

Principes Xpath

W3C

Xpath 2.0 est publié par le *World Wide Web Consortium* comme [standard XML](#)

Requêtage

Xpath est un **langage de requêtes** qui permet de parcourir un arbre XML

Intégré

Xpath a été conçu comme un **langage intégré**, et non pas un langage autonome

Syntaxe chemin

```
node/childNode
```

Une expression de chemin correspond à une séquence d'étapes séparées par l'opérateur `/`.

Sans indication particulière la relation se fait d'un élément parent vers un élément enfant.

Chaque étape devient le nœud courant (nœud de contexte) pour l'étape suivante.



Absolu / relatif

Un chemin de localisation **absolu** commence au nœud document : il commence par une barre oblique /

Les chemins de localisation **relatifs** commencent à n'importe quel nœud de l'arborescence ou par une double barre oblique //

Désignation d'une localisation

Une expression XPath est un chemin de localisation, constitué de pas de localisation (“étapes du chemin”) séparés par le caractère `/`. Les pas sont définis par :

Axe

Sens dans lequel l'arbre doit être parcouru
parcoursu (**parent**, descendant)

Test de nœud

Nom d'un nœud ou **type** de nœud situé dans la direction de l'axe dans l'arborescence

Prédicat

Indiqué entre [...] pour préciser une caractéristique du nœud (sorte de **filtre**)

Syntaxe pas de localisation

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

A decorative graphic on the left side of the slide consisting of two overlapping squares. The bottom-left square is a dark blue, and the top-right square is a lighter blue, creating a cross-like shape.

Syntaxe Xpath

Les différents éléments manipulés en Xpath



Les nœuds

Les documents XML sont traités comme des **arbres de nœuds**. L'ordre des nœuds est déterminé par la séquence du document et par l'imbrication des éléments XML. L'élément le plus élevé de l'arbre est appelé l'élément racine (**root**).

Types de nœuds courants

element	Désigne un élément (balises et contenu) Ne concerne pas les déclarations (<? . . . ?> et <! . . . >)
attribute	Désigne les attribut d'un élément XML
document	Désigne l'élément abstrait comprenant l'ensemble du document
text()	Désigne le contenu textuel d'un élément
comment()	Désigne le contenu d'un commentaire (pas son balisage <!-- . . . -->)

Un élément et son contenu

```
/teiHeader/titleStmt/title
```

L'élément `title` et son contenu

Le texte d'un élément

```
/teiHeader/titleStmt/title/text()
```

La chaîne de caractère contenu dans l'élément `title`

Exercice

Sélectionner les nœuds

Consigne

Trouver les chemins Xpath dans l'encodage TEI d'un sonnet :

Les vers de quatrain

Le paragraphe indiquant la provenance

Les tercets

Le titre du poème

Le texte des vers du poème

Abréviations pour les nœuds

- * Remplace n'importe quel nœud
- | Pour combiner deux ensembles de nœuds

N'importe quel contenu

```
/teiHeader/*
```

Tout ce qui est contenu dans le `teiHeader`

Concaténation de chemins

```
/teiHeader/titleStmt | /teiHeader/sourceDesc
```

Le titleStmt et le sourceDesc



Axes Xpath

Indique la **direction** dans laquelle se déplacer dans l'arbre XML, relativement au nœud courant ou depuis la racine. Quand l'axe n'est pas précisé, il s'agit implicitement de l'axe des enfants (`child::`).

Axes courants

ancestor	Sélectionne tous les ancêtres du nœud courant
ancestor-or-self	Sélectionne tous les ancêtres du nœud courant ainsi que le nœud lui-même
attribute	Sélectionne tous les attributs du nœud courant
child	Sélectionne tous les enfants du nœud courant
descendant	Sélectionne tous les descendants du nœud courant
descendant-or-self	Sélectionne tous les descendants du nœud courant ainsi que le nœud lui-même
following	Sélectionne tous les nœuds du document après la balise fermante du nœud courant
following-sibling	Sélectionne tous les nœuds suivants qui sont au même niveau que le nœud courant
parent	Sélectionne le nœud parent du nœud courant
preceding-sibling	Sélectionne tous les nœuds précédents qui sont au même niveau que le nœud courant
self	Sélectionne le nœud courant

Syntaxe axe attribut

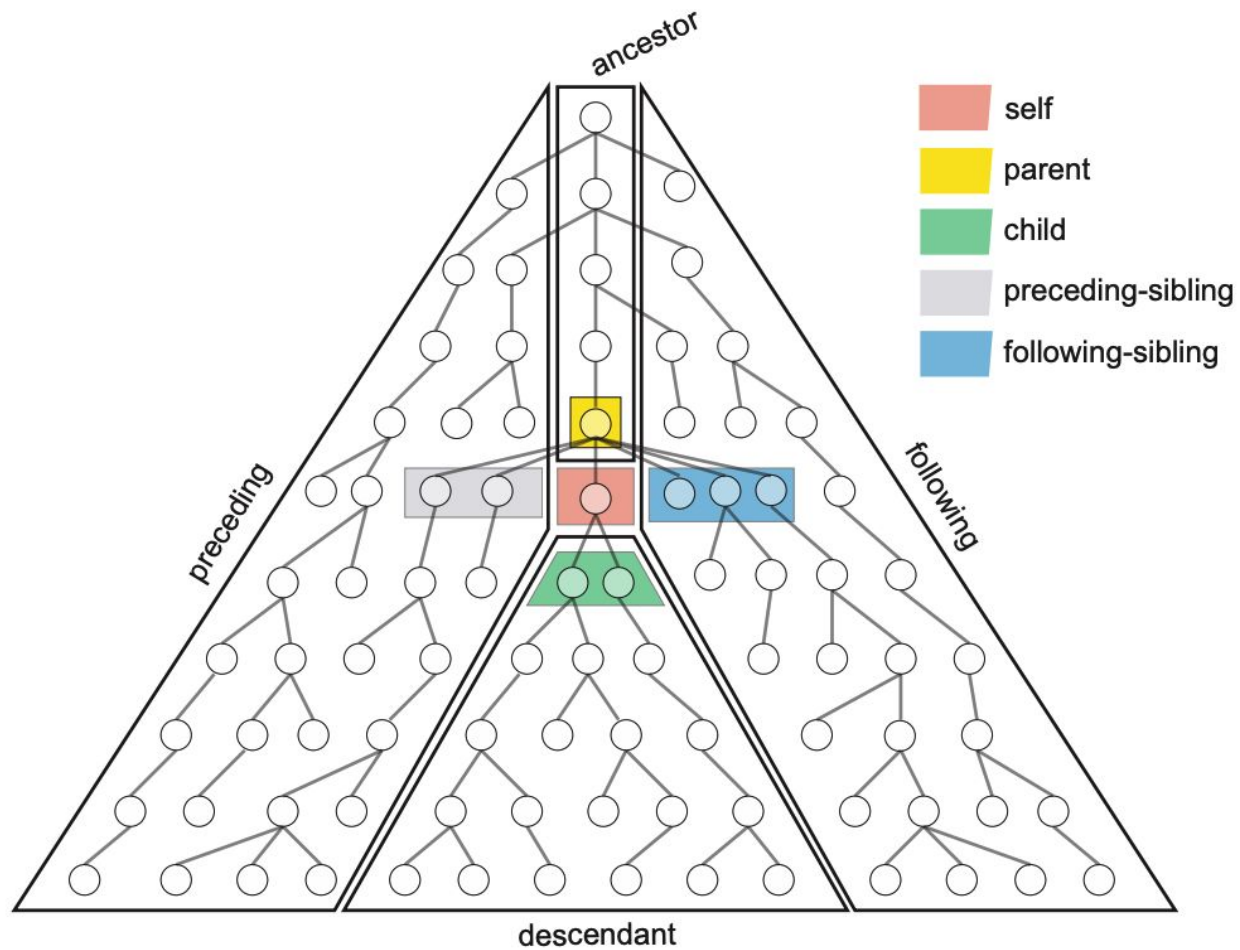
`p/attribute::n`

La valeur de l'attribut n de l'élément p

Syntaxe axe

```
//sibling::lg
```

Les éléments au même niveau que le lg courant



Les abréviations d'axe

vide	<code>child::</code>	Axe par défaut
@	<code>attribute::</code>	Sélectionne la valeur d'un attribut
//	<code>/descendant-or-self::node()/</code>	Sélectionne le nœud document et tous les descendants
..	<code>parent::node()</code>	Sélectionne le nœud parent du nœud courant
.	<code>self::node()</code>	Sélectionne le nœud courant

Sélection d'attribut

$p/@n$

La valeur de l'attribut n de l'élément p

N'importe où dans l'arborescence

```
//p
```

Tous les éléments p peu importe leur endroit dans l'arborescence

Parmi les descendants

`./p`

Tous les éléments p descendant du nœud courant

Noeud parent

../p

L'élément p parent de l'élément courant

Exercice

A decorative graphic on the right side of the slide, consisting of a light gray square positioned above a blue square, both of which are partially cut off by the right edge of the frame.

Traduire les expressions

Consigne

Traduire les expressions suivantes

```
//figure
```

```
//*/*
```

```
//book/title
```

```
chapter//footnote
```

```
./footnote
```




Prédicats

Les prédicats sont utilisés pour **filtrer** les nœuds sélectionnés par l'axe et le test de nœud. Les prédicats sont écrits entre crochets. Si le prédicat est évalué à vrai, tous les nœuds correspondants sont sélectionnés.

Noeud qui a un enfant qui...

```
titleStmt[title="Sonnet d'automne"]
```

L'élément(s) `titleStmt` dont l'enfant `title` a pour contenu "Sonnet d'automne"

Noeud qui a un enfant qui...

```
titleStmt[title="Sonnet d'automne"]/author
```

L'élément(s) `author` contenu dans l'élément `titleStmt` dont l'enfant `title` a pour contenu "Sonnet d'automne"

Test d'existence

```
titleStmt[title]
```

L'élément(s) `titleStmt` qui contient un élément `title`

Élément qui a pour attribut...

```
lg[@type="quatrain"]
```

L'élément(s) lg qui a un attribut type ayant pour valeur quatrain

Prédicat numérique

```
lg[@type="quatrain"][2]
```

Le deuxième élément `lg` qui a un attribut `type` ayant pour valeur `quatrain`

Abréviation de `lg[position()=2]`

 Contrairement aux autres langages de programmation, l'énnumération XPath commence avec 1

Exercice

Filtrer les nœuds



Consigne

Formuler ces expressions en xpath

Le second p de la première div du body

Les p au sein du text qui contiennent des persName

Les placeName dont la valeur est Paris

La troisième mention du persName "Gringoire"

La deuxième div de la quatrième div



Opérateurs Xpath

Les tests formulés dans les prédicats sont définis à l'aide d'opérateurs. Il en existe de trois types : les opérateurs **arithmétiques**, les opérateurs **logiques** et les opérateurs de **comparaison**.

Opérateurs courants

`+ - * div mod`

Pour effectuer des opérations arithmétiques

`= !=`

Pour tester l'égalité ou l'inégalité

`< >`

Strictement plus grand / plus petit que (`<lt;` et `>gt;` en XSLT)

`<= >=`

Inférieur / supérieur ou égal

`and or`

Pour accoler plusieurs expressions booléennes

Comparer des valeurs

```
1[@n > 2]
```

Les éléments 1 dont la valeur de l'attribut n est strictement supérieure à 2

Combiner des expressions

```
lg[@type != "tercet" and @n < 2]
```

Les éléments lg dont la valeur de l'attribut n est strictement inférieure à 2 et dont l'attribut type n'est pas "tercet"



Fonctions Xpath

Les fonctions Xpath, utilisées dans les prédicats, permettent de manipuler les données à tester. Elles peuvent s'appliquer à des données **textuelles**, **numériques** et **booléennes**. Elles peuvent également s'appliquer à des **ensembles de nœuds**.

Fonctions courantes

<code>count(node-set)</code>	Compte le nombre d'occurrences dans un ensemble de nœud
<code>sum(node-set)</code>	Somme des valeurs de type nombre dans un ensemble de nœud
<code>not(boolean)</code>	Renvoie la valeur inverse du booléen
<code>true() / false()</code>	Vrai / faux
<code>concat(str, str)</code>	Concaténation de plusieurs chaînes de caractère
<code>start-with(str1, str2)</code>	Renvoie vrai si <code>str1</code> commence par <code>str2</code>
<code>contains(str1, str2)</code>	Renvoie vrai si <code>str1</code> contient <code>str2</code>
<code>string-length(str)</code>	Renvoie la longueur d'une chaîne de caractère
<code>normalize-space(str)</code>	Retire les espaces en début et fin de chaîne, ainsi que les doubles espaces
<code>last()</code>	Dernier nœud du document correspondant au prédicat
<code>position()</code>	Nombre désignant la position d'un élément
<code>id(val)</code>	Renvoie vrai si l'élément a l'id donné en paramètre

Dénombrer le contenu

```
//lg[count(l) >= 4]
```

Les éléments `lg` qui contiennent 4 ou plus éléments `l`

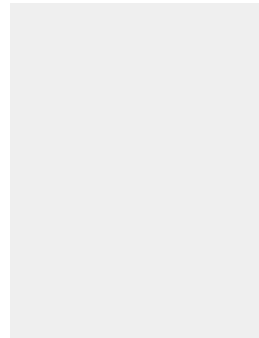
Filtrer la position

```
//lg[position() != 1]
```

Les éléments lg qui ne sont pas le premier dans l'axe du document

Exercice

Manipuler les données



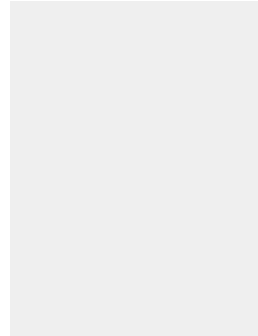
Consigne

Traduire les formules suivantes

```
//l[@n = count(lg)]  
//div[@n = sum(age)]  
//persName[text()="Trucmuche"]  
//persName[contains(., "Truc")]  
//listPerson[person[contains(., "Truc")]]
```

Exercice

Parcourir les encodages



Consigne

Donner les chemins les plus courts

Mon rêve familial

De la racine vers les vers d'un quatrain;

De la racine vers le titre;

Le Misanthrope

* De la racine vers les vers d'une scène;

* De la racine vers la déclaration d'un personnage;

* De la racine vers la dernière déclaration d'un nom dans la liste des personnages de la préface;