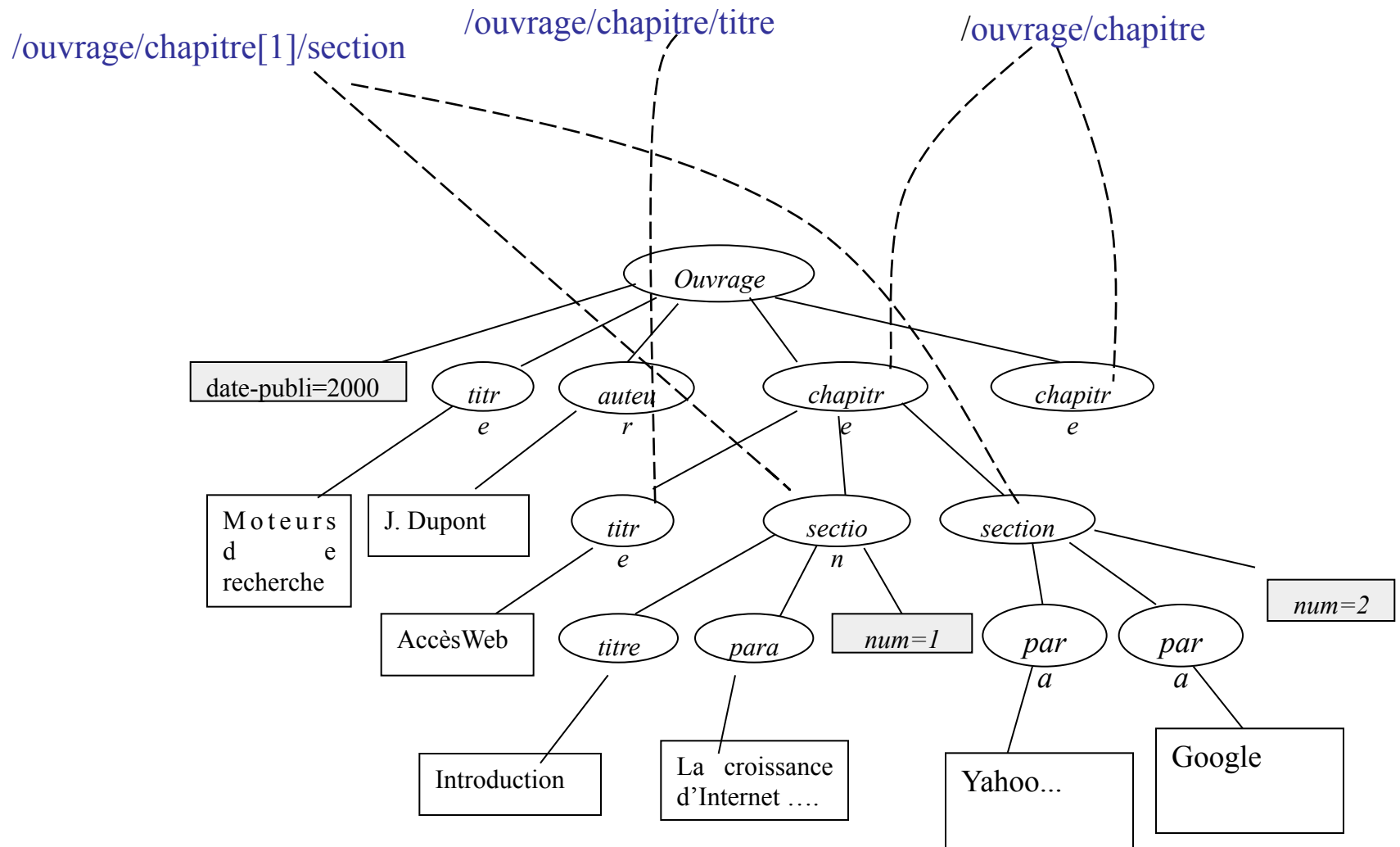

Chapitre 4 :

Xpath

Introduction

- Document XML = ensemble de balises organisées hiérarchiquement (structure d 'arbre)
- XPath
 - Langage d'expressions de chemins permettant de sélectionner des parties d'un document XML
 - Opère sur l'arbre d'un document
 - Objectif : référencer les noeuds (éléments, attributs, commentaires, ...) dans un document XML

XPath - Exemples



Xpath : utilisation

- Le langage permet de désigner un ou plusieurs noeuds dans un document XML, à l'aide **d'expressions de chemin.**
- XPath est utilisé par
 - XSLT pour sélectionner des règles de transformation
 - XML Schéma pour créer des clés et références
 - XLink pour créer des liens entre documents/fragments XML
 - XQUERY pour interroger des collections de documents

XPath - Expression de chemins

- Une expression de chemins
 - spécifie une traversée de l'arbre du document (un chemin) :
 - depuis un nœud de départ
 -vers un ensemble de nœuds cibles
 - a pour résultat
 - un ensemble de noeuds
 - ou une valeur, numérique, booléenne ou alphanumérique
- Un chemin XPath est une suite **d'étapes** assemblées de gauche à droite et séparées par le caractère « / »
 - : [/]étape1 /étape2 /.../étapeN
- Deux variantes :
 - Absolu :
 - commence à la racine; /étape1/.../étapeN
 - Relatif :
 - commence à un nœud courant (contexte), étape1/.../étapeN,

Etapes Xpath

- Une étape = un Chemin élémentaire
 - [Axe::]Filtre[condition1] [condition2]..
- Axe de localisation
 - sens de parcours des noeuds (par défaut: child).
- Filtre
 - nom de nœud sélectionné (élément ou @attribut)
- Condition(prédicats)
 - que doivent satisfaire les noeuds retenus

- Exemple : /child::ouvrage/child::chapitre

Etape 1

Etape 2

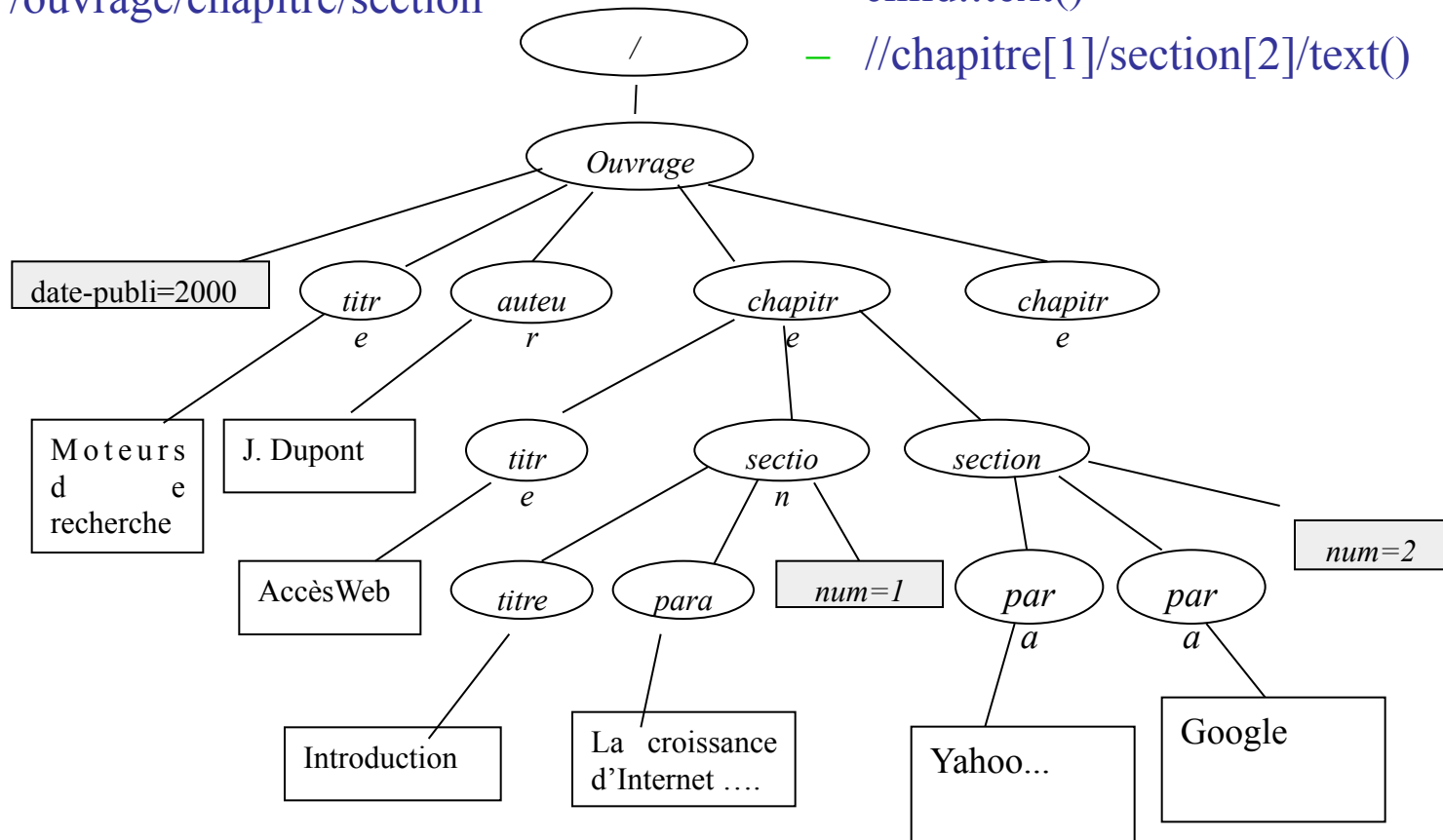
XPath - Exemples

- Sections d'un chapitre

- `/child::ouvrage/child::chapitre/child::section`
- `/ouvrage/chapitre/section`

- Texte du chapitre 1 section 2

- `/descendant::chapitre[position() = 1]/child::section[position() = 2]/child::text()`
- `//chapitre[1]/section[2]/text()`



Axe de localisation

- Un axe sélectionne, dans l'arbre du document et à partir du nœud contexte, l'ensemble des noeuds qui peuvent être atteints en suivant une certaine direction :
 - child : sélectionne les enfants du noeud contexte ;
 - descendant : sélectionne les descendants du noeud contexte ;
 - parent : sélectionne le père du noeud contexte ;
 - ancestor : sélectionne les ancêtres du noeud contexte ;
 - following-sibling : sélectionne les frères suivants du nœud contexte (si le noeud contexte n'est pas un élément, cet axe est vide) ;
 - preceding-sibling : sélectionne les frères précédents du nœud contexte (si le noeud contexte n'est pas un élément, cet axe est vide) ;

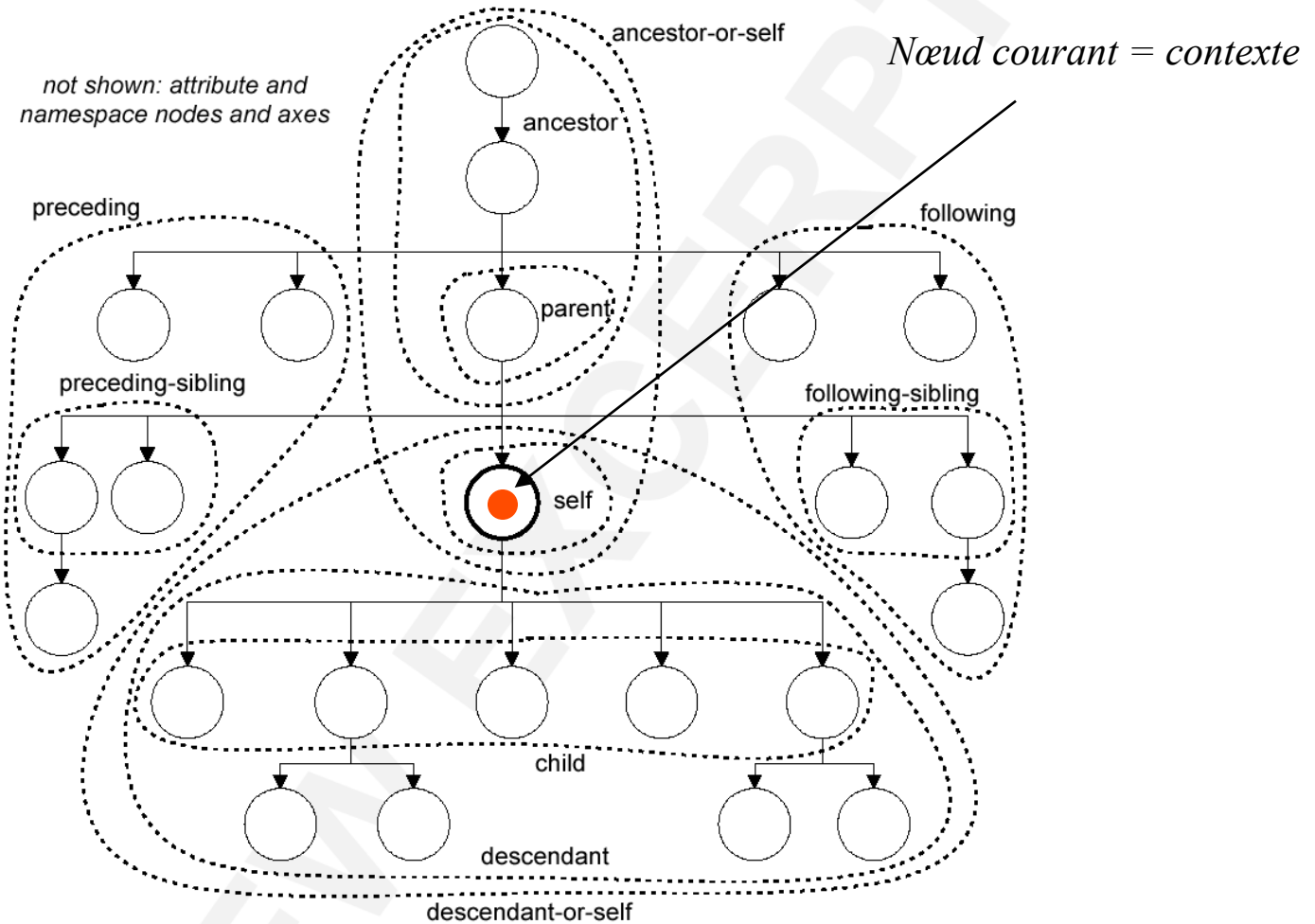
Axe de localisation

- following : sélectionne les noeuds qui sont après le nœud contexte dans l'ordre du document excepté ses descendants ainsi que les noeuds attributs et espace de noms ;
- preceding : sélectionne les noeuds qui sont avant le nœud contexte dans l'ordre du document excepté ses ancêtres ainsi que les noeuds attributs et espace de noms ;
- attribute : sélectionne les noeuds attributs, fils du nœud contexte (si le nœud contexte n'est pas un élément, cet axe est vide) ;
- namespace : sélectionne les noeuds attributs ou espaces de noms, fils du nœud contexte (si le nœud contexte n'est pas un élément, cet axe est vide) ;

Axe de localisation

- self : sélectionne uniquement le noeud contexte ;
- descendant-or-self : sélectionne le noeud contexte et les noeuds descendants du noeud contexte ;
- ancestor-or-self : sélectionne le noeud contexte et les noeuds ancêtres du noeud contexte.

Récapitulatif : Axe de localisation



Filtre

- Un **test de noeud** sélectionne parmi les noeuds de l'axe, ceux qui sont d'un certain type.
- Un test de noeud a l'une des formes suivantes :
 - n où n est un nom : sélectionne les noeuds de l'axe ayant le même type que le type principal de l'axe et dont le nom étendu est égal au nom étendu de n ;
 - $*$: sélectionne les noeuds de l'axe ayant le même type que le type principal de l'axe ;
 - `node()` : sélectionne tout noeud de l'axe ;
 - `text()` : sélectionne tout noeud de l'axe de type texte ;
 - `comment()` : sélectionne tout noeud de l'axe de type commentaire ;
 - `processing-instruction(n)` : sélectionne tout noeud de l'axe représentant une instruction de traitement de nom n .

Quelques exemples

- **child::para** : sélectionne les éléments *para* enfants du noeud contextuel
- **child::*** sélectionne tous les éléments enfant du noeud contextuel
- **child::text()** : tous les noeuds textuels du noeud contextuel
- **child::node()** : tous les enfants du noeud contextuel, quel que soit leur type (élément ou autre)
- **attribute::name** : l'attribut *name* du noeud contextuel
- **attribute::*** : tous les attributs du noeud contextuel
- **descendant::para** : tous les descendants *para* du noeud contextuel
- **ancestor::para** : tous les ancêtres *para* du noeud contextuel
- **ancestor-or-self::section** : tous les ancêtres *section* du noeud contextuel et le noeud contextuel lui-même si c'est un *section*
- **descendant-or-self::para** : tous les descendants *para* du noeud contextuel et le noeud contextuel lui-même si c'est un *para*
- **self::para** : le noeud contextuel si c'est un élément *para*, et rien dans le cas contraire
- **child::chapitre/descendant::para** : les descendants *para* des éléments *chapitre* enfants du noeud contextuel
- **child::* / child::para** : tous les petits enfants *para* du noeud contextuel

Syntaxe abrégée

- Afin de faciliter la lecture des chemins de localisation, les abréviations suivantes sont autorisées :
 - **child** peut être omis (axe par défaut)
 - Exemple: `child::section/child::para` \equiv `section/para`
 - **attribute** peut être remplacé par **@**
 - Exemple: `child::para[attribute::type="warning"]` \equiv `para[@type="warning"]`
 - **//** \equiv **/descendant-or-self::node()/**
 - Exemple: `//para` \equiv `/descendant-or-self::node()/child::para`
 - `//para[1]` \neq `/descendant::para[1]`
 - **.** \equiv **self::node()**
 - **..** \equiv **parent::node()**

Condition (1)

- Condition :
 - expression booléenne constituée d'un ou plusieurs tests, composés avec les connecteurs logiques habituels *and* et *or*
- Test :
 - toute expression XPath, dont le résultat est convertie en booléen;
 - une comparaison, un appel de fonction.

Quelques exemples (1)

- **child::para[position()=1]** : le premier enfant *para* du noeud contextuel
- **child::para[position()=last()]** : le dernier enfant *para* du noeud contextuel
- **child::para[position()=last()-1]** : l'avant dernier *para* enfant du noeud contextuel
- **child::para[position()>1]** sélectionne tous les enfants *para* du noeud contextuel autres que le premier
- **following-sibling::chapitre[position()=1]** : le prochain *chapitre* cible du noeud contextuel
- **preceding-sibling::chapitre[position()=1]** : le précédent *chapitre* cible du noeud contextuel
- **/descendant::figure[position()=42]** : le 42ième élément *figure* du document
- **/child::doc/child::chapitre[position()=5]/child::section[position()=2]** : la 2ième *section* du 5ième élément *chapitre* de l'élément *doc* du document
- **child::para[attribute::type='warning']** : tous les enfants *para* du noeud contextuel qui ont un attribut *type* dont la valeur est *warning*

Quelques exemples (2)

- **child::para[attribute::type='warning'][position()=5]** : le 5ième enfant *para* du noeud contextuel qui a un attribut *type* dont la valeur est *warning*
- **child::para[position()=5][attribute::type="warning"]** : le 5ième enfant *para* du noeud contextuel si celui-là a un attribut *type* de valeur *warning*
- **child::chapitre[child::title='Introduction']** : l'enfant *chapitre* du noeud contextuel qui a un ou plus enfant *title* avec un contenu textuel égal à *Introduction*
- **child::chapitre[child::title]** : l'enfant *chapitre* du noeud contextuel qui a un ou plus enfant *title*
- **child::*[self::chapitre or self::appendix]** : tous les enfants *chapitre* et *appendix* du noeud contextuel
- **child::*[self::chapitre or self::appendix][position()=last()]** : le dernier enfant *chapitre* ou *appendix* du noeud contextuel
- **/A/B/descendant::text()[position()=1]** : Le premier noeud de type **Text** descendant d'un /A/B.

Condition (2)

- Il existe quatre formes possibles de conditions:
 - *axe::filtre[numéro]*
 - *axe::filtre[expression_XPATH]*
 - *axe::filtre[expression logique]*
 - Conditions composées

axe::filtre[numéro]

- Sélectionne les noeuds en fonction de leur position.
 - Exemple
 - `/ouvrage/chapitre/section[2]`
 - `//section[position()=last()]`
 - équivalent à
 - `//section[last()]`

axe::filtre[expression_XPATH]

- Sélectionne les noeuds pour lesquels la sous expression renvoie un ensemble de noeuds non vide.
 - Exemples
 - Chapitre qui comporte un texte
 - */ouvrage/chapitre[text()]*
 - Section comportant un attribut num
 - *//chapitre/section[@num]*

axe::filtre[expression logique]

- Les conditions peuvent également s'écrire (les relations possibles sont =, !=, <, <=, >, >=) ; connecteurs logiques:

-

valeur1 relation valeur2

condition1 and condition2

condition1 or condition2

not(condition)

true()

false()

boolean(objet)

- Chapitre comportant une section ayant un attribut num = 1
 - *chapitre[section/@num = '1']*
- *//chapitre/section[@num != '1' and text()]*
- *//chapitre/section[@num>1 and titre/text()='Introduction']*
- *//chapitre[following::section[@num=1]]*

XPATH: Fonctions & opérations (1)

- Les expressions logiques peuvent également utiliser les fonctions suivantes:
 - Types de base:
 - booléen, chaîne de caractères, nombre réel, ensemble de noeuds.
 - Opérations sur les nombres:
 - +, -, div, mod
 - last() :
 - vraie ssi le noeud est le dernier.
 - position() :
 - renvoie la position du noeud courant. produit[(position() mod 2) = 0]
 - id(*nom*) :
 - renvoie le noeud identifié par l'étiquette *nom*.

XPATH: Fonctions & opérations (2)

- Autres fonctions:
 - $\text{local-name}(noeuds)$ $\text{namespace-uri}(noeuds)$
 $\text{name}(noeuds)$
 - $\text{string}(objet)$
 - $\text{concat}(chaîne1, \dots, chaîneN)$ $\text{string-length}(chaîne)$
 $\text{normalize-space}(chaîne)$ $\text{translate}(ch1, ch2, ch3)$
 - $\text{substring-before}(ch1, ch2)$: renvoie la chaîne *res* définie par $ch1 = res + ch2 + reste$
 $\text{substring-after}(ch1, ch2)$ renvoie la chaîne *res* définie par $ch1 = reste + ch2 + res$
 $\text{substring}(chaîne, début)$
 $\text{substring}(chaîne, début, len)$

XPATH: Fonctions & opérations (3)

- Autres fonctions:
 - `starts-with(ch1,ch2)` ssi *ch1* débute par *ch2*
 - `contains(ch1,ch2)` ssi *ch1* contient *ch2*
 - `number(object)` `sum(noeuds)` `count(noeuds)`
 - `floor(nombre)` arrondi par le bas
 - `ceiling(nombre)` arrondi par le haut
 - `round(nombre)` arrondi par le plus proche

Récapitulatif fonctions de base (1)

- **Sur les noeuds**

- *nombre* **last()**
- *nombre* **position()**
- *nombre* **count(*nœuds**)**
- *nœuds** **id(*objet*)**
 - `id("foo")/child::para[position()=5]`

- **Sur les chaînes de caractères**

- *string* **string(*object*?)**
- *string* **concat(*string*, *string*, *string**)**
- *string* **starts-with(*string*, *string*)**
- *boolean* **contains(*string*, *string*)**
- *string* **substring-before(*string*, *string*)**
- *string* **substring-after(*string*, *string*)**
- *string* **substring(*string*, *number*, *number*?)**
- *number* **string-length(*string*?)**

Récapitulatif fonctions de base (2)

- **Sur les booléens**
 - *boolean* **boolean**(object)
 - *boolean* **not**(boolean)
 - *boolean* **true**()
 - *boolean* **false**()
- **Sur les numériques**
 - *number* **number**(object?)
 - *number* **sum**(noeuds*)
 - *number* **floor**(number)
 - *number* **ceiling**(number)
 - *number* **round**(number)

Les conditions composées

- *axe::filtre[condition1] [condition2]...*
sélectionne les noeuds identifiés par *filtre* ssi les conditions sont respectées. Attention, ces deux expressions sont différentes:
 - *chapitre[2][para]*
sélectionne les noeuds chapitre en deuxième position qui possède un élément fils para.
 - *chapitre[text()][2]*
sélectionne le deuxième noeud chapitre qui possède un élément fils texte.

Fin

- Exercices
 - sélectionner
 - Les titres des sections
 - les chapitres ayant des sections
 - les sections avec attributs
 - le contenu des titres des sections
 - les sections ayant le titre « introduction ».
 - Les titres comportant le terme « introduction »