

TP 2 – XML schéma et Xpath

Objectif : voir l'alternative XMLS-DTD et savoir former les expressions XPATH.

1. Composer les expressions XPath

1.1. Pour votre fichier XML sur les documents réglementaires (TP1), désignez un arbre correspondant.

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE annotationRegle SYSTEM "file:/D:/sept2007/data/annotationSchema.dtd">
<base>
  <document >
    <thematique typeThematique ="assurance construction"/>
    <thematique typeThematique ="amiante"/>
    <actualite typeActualite = "modifie"/>
    <titre>
      <sigle>NF P</sigle>
      <numero>31-313</numero>
      <date dateCreation="aout 1996" datePublication ="aout 1996"/>
      <title>
        Produits de couverture - Tuiles en béton à glissement à emboîtement longitudinal de classe montagne –
        Définition, caractéristiques, marquage
      </title>
      <relatifA>Tuiles en béton à glissement à emboîtement longitudinal de classe montagne</relatifA>
    </titre>
  </document>
  <document >
    <thematique typeThematique ="assurance construction"/>
    <actualite typeActualite = "modifie"/>
    <titre>
      <sigle>NF P</sigle>
      <numero>00-002</numero>
      <date dateCreation="Septembre 1994" datePublication ="Septembre 1994"/>
      <title>
        bâtiment et génie civil - vocabulaire - partie 2 : Termes relatifs aux marchés
      </title>
      <relatifA>Bâtiment, contrat, vocabulaire</relatifA>
    </titre>
  </document>
  <document >
    <thematique typeThematique ="acoustique"/>
    <actualite typeActualite = "modifie"/>
    <titre>
      <sigle>DTU P</sigle>
      <numero>52-221</numero>
      <date dateCreation="février 1969" datePublication ="février 1969"/>
      <title>
        prescriptions techniques relatif aux chaufferies aux gaz et aux hydrocarbures liquéfiés +
        Additif n° 1 de décembre 1971, additif n° 2 de septembre 1978 incorporés.
      </title>
      <relatifA>
        chaufferies aux gaz, chaufferies aux hydrocarbures liquéfiés
      </relatifA>
    </titre>
  </document>
</base>
```

1.2. Extraire les fragments de ce document (écrire les expressions XPath) pour sélectionner

- La racine du document
/
- Tous les éléments de type /A/B[@att1]
les nœuds /base/document[@att1] il n'y a pas d'attributs de « document »
- Tous les nœuds qui n'ont pas des attributs
//*[not(@*)]
- Toutes les sous-parties de l'élément « date »
nœuds /base/document/date/*
- Le premier nœud de type Text descendant d'un /A/B
/A/B/descendant::text()[1]

1.3. Interprétez ces expressions en termes de votre problématique

2. XPath : sélectionner par valeur

2.1. Extraire les fragments correspondant

- Les documents réglementaires dont la thématique n'est pas « acoustique »
/base/document[not(thematique[typeThematique='acoustique'])]
- Les documents avec leurs thématique et date de publication dans un Journal Officiel
/base/document[thematique][date/@datePubli]
- Le premier document de norme (possédant le sigle NF)
/base/document[titre/sigle=NF][1]
- Les nœuds « titre » en troisième position qui possède un élément fils « relatifA »
/base/document/titre[3][relatifA]
- Retournez le nombre des documents datés 2005
count(/base/document[date[@dateCreation='2005']])

3. XML schéma

3.1. Définissez un XML schéma pour votre XML fichier sur les documents réglementaires

```
<?xml version="1.0" encoding="UTF-8"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema">
  <xs:element name="base">
    <xs:complexType>
      <xs:sequence>
        <xs:element ref="document" />
      </xs:sequence>
    </xs:complexType>
  </xs:element>
  <xs:element name="document">
    <xs:complexType>
      <xs:sequence>
        <xs:element ref="thematique" minOccurs="0" maxOccurs="unbounded" />
        <xs:element ref="actualite" />
        <xs:element ref="titre" />
      </xs:sequence>
    </xs:complexType>
  </xs:element>
  <xs:element name="thematique">
    <xs:complexType mixed="true">
      <xs:attribute name="typeThematique" use="required">
        <xs:simpleType>
          <xs:restriction base="xs:NMTOKEN">
            <xs:enumeration value="accessibilité_handicapés" />
            <xs:enumeration value="acoustique" />
            <xs:enumeration value="aération" />
          </xs:restriction>
        </xs:simpleType>
      </xs:attribute>
    </xs:complexType>
  </xs:element>
</xs:schema>
```

```

        <xs:enumeration value="amiante" />
        <xs:enumeration value="ascenseurs" />
        <xs:enumeration value="assurance_construction" />
        <xs:enumeration value="sécurité_incendie" />
    </xs:restriction>
</xs:simpleType>
</xs:attribute>
</xs:complexType>
</xs:element>
<xs:element name="actualite">
    <xs:complexType mixed="true">
        <xs:attribute name="typeActualite" use="optional" default="nouveau">
            <xs:simpleType>
                <xs:restriction base="xs:NMTOKEN">
                    <xs:enumeration value="modifie" />
                    <xs:enumeration value="nouveau" />
                </xs:restriction>
            </xs:simpleType>
        </xs:attribute>
    </xs:complexType>
</xs:element>
<xs:element name="titre">
    <xs:complexType>
        <xs:sequence>
            <xs:element ref="sigle" minOccurs="0" />
            <xs:element ref="numero" />
            <xs:element ref="date" />
            <xs:element ref="title" />
            <xs:element ref="relatifA" />
        </xs:sequence>
    </xs:complexType>
</xs:element>
<xs:element name="sigle">
    <xs:complexType mixed="true" />
</xs:element>
<xs:element name="numero">
    <xs:complexType mixed="true" />
</xs:element>
<xs:element name="date">
    <xs:complexType mixed="true">
        <xs:attribute name="typeDate" use="required">
            <xs:simpleType>
                <xs:restriction base="xs:NMTOKEN">
                    <xs:enumeration value="dateCreation" />
                    <xs:enumeration value="datePubli" />
                </xs:restriction>
            </xs:simpleType>
        </xs:attribute>
    </xs:complexType>
</xs:element>
<xs:element name="title">
    <xs:complexType mixed="true" />
</xs:element>
<xs:element name="relatifA">
    <xs:complexType mixed="true" />
</xs:element>
</xs:schema>

```

3.2. Liez votre XML à ce schéma

```

<base xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:noNamespaceSchemaLocation="base.xsd">

```

4. Questions.

- 4.1. Pour quel objectif il est important de former les expressions XPath ? Comment je peux les utiliser ?
- 4.2. Est-ce que ces expressions sont équivalentes ? Pourquoi ?

- DOCUMENT[TITRE and position()=1]
- DOCUMENT[position()=1][TITRE]
- DOCUMENT[TITRE][position()=1]

1 et 2 sont équivalents

- 4.3. Quels sont les avantages du XML Schéma par rapport à la DTD ? Exemples
- 4.4. Comment déclarer un type complexe (des séquences d'éléments, des types de choix, des contraintes d'occurrences) ?
- 4.5. Comment lier un fichier XML à plusieurs schémas ? Pourquoi le faire (ou ne pas le faire)?

5. Enrichissement d'un schéma XML pour le problème « réglementaire »

Continuer le développement de votre exemple

- Concrétisez les types (utilisation de xsd:restriction, xsd:maxLength, xsd:pattern, ...)
- Définissez les types complexes (utilisation de xsd:sequence, xsd:choice, ...)
- Ecrivez 5 requêtes XPath *différentes* permettant de « récolter » la majorité de l'information de votre XML (voir ex.1, ex.2 pour composer les expressions de différents types)