

Contraindre le XML : présentation de la DTD

Séance 3

Master 2 - UVSQ - 2024-25

Pourquoi contraindre le XML ?

- **Assurer la Cohérence**
- **Éviter les Erreurs**
- **Faciliter le Traitement Automatisé**

Exemple :

<https://edition-testaments-de-poilus.huma-num.fr/>

Un document valide

Un document XML-TEI qui respecte un schéma de validation comme une DTD est dit "valide", ce n'est pas la même chose qu'un document dit "bien formé".

Un document bien formé :

- 1 balise ouvrante = 1 balise fermante
- respect de l'imbrication
- un attribut du même nom par balise
- un seul élément racine

Exemple :

```
<book>  
  <title>Les Misérables</title>  
  <author ref="https://fr.wikipedia.org/wiki/Victor_Hugo">Victor Hugo</author>  
</book>
```

Document Valide :

Un document XML est "valide" lorsqu'il est non seulement bien formé, mais qu'il respecte aussi les contraintes définies par une DTD ou un autre schéma de validation.

- Validation avec une DTD : La DTD définit quelles balises sont permises, comment elles peuvent être imbriquées, et quels attributs sont autorisés ou obligatoires. Si le document respecte toutes ces règles, il est considéré comme valide.

Le Document Type Definition (DTD)

Une DTD se présente sous la forme d'une **liste de déclaration** qui définissent la structure que devront suivre les documents qui s'y réfèrent.

La DTD qui sert de modèle à un document doit être déclarée au début de celui-ci, **avant l'ouverture de l'élément racine**.

Déclaration de la DTD

La déclaration de la DTD peut être faite au **début du document** XML dont elle contraint l'encodage, ou dans un **fichier externe** dont il est fait référence dans le document XML

Déclaration de la DTD Interne

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

<!DOCTYPE book [
  <!ELEMENT book (title, author)>
  <!ELEMENT title (#PCDATA)>
  <!ELEMENT author (#PCDATA)>
  <!ATTLIST author ref CDATA #IMPLIED>
]>

<book>
  <title>Les Misérables</title>
  <author ref="https://fr.wikipedia.org/wiki/Victor_Hugo">Victor Hugo</author>
</book>
```


Déclaration de la DTD externe

Fichier `book.dtd` :

```
<!ELEMENT book (title, author)>
<!ELEMENT title (#PCDATA)>
<!ELEMENT author (#PCDATA)>
<!ATTLIST author ref CDATA #IMPLIED>
```

Document XML faisant référence à la DTD externe :

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE book SYSTEM "book.dtd">

<book>
  <title>Les Misérables</title>
  <author ref="https://fr.wikipedia.org/wiki/Victor_Hugo">Victor Hugo</author>
</book>
```

Contenu d'une DTD

Une DTD contient des déclarations concernant :

- **Éléments**: Définit les éléments qui peuvent être utilisés, leur nombre, leur imbrication, leur ordre, etc.
- **Attribus**: Définit les attributs autorisés pour un élément précis, leur valeurs et le type de valeur autorisé
- **Entités**: Définit les entités qui pourront être utilisées dans le document XML

L'<élément> dans la DTD

1. Définir un Élément dans une DTD

Syntaxe Générale :

```
<!ELEMENT nom_de_l_element type_de_contenu>
```

Exemple :

```
<!ELEMENT title (#PCDATA)>
```

Donc par exemple dans le document XML-TEI :

```
<title>Les Misérables</title>
```

L'<élément> dans la DTD

2. Types de contenus des éléments

Expression	Signification
EMPTY	Contenu vide
ANY	Contenu quelconque
(#PCDATA)	Contenu textuel
(elt)	Un seul élément
(elt1, elt2, ..., eltn)	Séquence d'éléments pris dans cet ordre
(elt1 elt2 ... eltn)	Un des éléments au choix

2. Types de contenus des éléments

1. **EMPTY** - Contenu Vide

Exemple DTD:

```
<!ELEMENT linebreak EMPTY>
```

Exemple XML:

```
<linebreak/>
```

2. Types de contenus des éléments

2. **ANY** - Contenu quelconque

Exemple DTD:

```
<!ELEMENT note ANY>
```

Exemple XML:

```
<note>Cette note contient simplement du texte.</note>  
<note>  
  <paragraph>Cette note contient un paragraphe structuré.</paragraph>  
</note>  
<note/>
```


2. Types de contenus des éléments

3. #PCDATA - Contenu textuel

Exemple DTD:

```
<!ELEMENT title (#PCDATA)>
```

Exemple XML:

```
<title>Introduction to XML</title>
```

2. Types de contenus des éléments

4. (elt) - Un seul élément

Exemple DTD:

```
<!ELEMENT book (title)>
```

Exemple XML:

```
<book>  
  <title>Learning XML</title>  
</book>
```


2. Types de contenus des éléments

5. (elt1, elt2, ..., eltn) - Séquence d'éléments

Exemple DTD:

```
<!ELEMENT book (title, author)>
```

Exemple XML:

```
<book>  
  <title>XML Basics</title>  
  <author>John Doe</author>  
</book>
```

2. Types de contenus des éléments

6. (elt1 | elt2 | ... | eltn) - Un des éléments au choix

Exemple DTD:

```
<!ELEMENT content (paragraph | image)>
```

Exemple XML:

```
<content>  
  <paragraph>This is a paragraph.</paragraph>  
</content>
```

Ou

```
<content>  
  <image src="image.jpg"/>  
</content>
```


Exercice : lier un document XML à une DTD

L'<élément> dans la DTD

3. Les opérateurs d'occurrence

Opérateur	Signification
?	0 ou 1 occurrence
*	0 ou plusieurs occurrences
+	1 ou plusieurs occurrences
(rien)	1 seule occurrence

3. Les opérateurs d'occurrence

1. ? - 0 ou 1 Occurrence

DTD:

```
<!ELEMENT person (name, nickname?)>
```

XML Valide:

```
<person>  
  <name>John Doe</name>  
  <nickname>Johnny</nickname>  
</person>
```

XML Valide (sans `nickname`):

```
<person>  
  <name>John Doe</name>  
</person>
```

3. Les opérateurs d'occurrence

2. * - 0 ou Plusieurs Occurrences

DTD:

```
<!ELEMENT shoppingList (item*)>
```

XML Valide (avec plusieurs `item`):

```
<shoppingList>  
  <item>Bread</item>  
  <item>Milk</item>  
</shoppingList>
```

XML Valide (sans `item`):

```
<shoppingList/>
```


3. Les opérateurs d'occurrence

3. + - 1 ou Plusieurs occurrences

DTD:

```
<!ELEMENT paragraph (sentence+)>
```

XML Valide:

```
<paragraph>  
  <sentence>This is the first sentence.</sentence>  
  <sentence>This is the second sentence.</sentence>  
</paragraph>
```

XML Invalide (pas de `sentence`):

```
<paragraph/>
```

3. Les opérateurs d'occurrence

4. (rien) - 1 seule occurrence

DTD:

```
<!ELEMENT book (title, author)>
```

XML Valide:

```
<book>  
  <title>Les Misérables</title>  
  <author>Victor Hugo</author>  
</book>
```

XML Invalide (manque l'élément **author**):

```
<book>  
  <title>Les Misérables</title>  
</book>
```


Exercice : exemples avec des opérateurs d'occurrence

Exercice 1 : Rédiger des déclarations d'éléments pour de la poésie

```
<?xml version="1.0" encoding="UTF-8"?>
<poeme>
  <titre>Mon rêve familial</titre>
  <auteur>Paul Verlaine</auteur>
  <strophe>
    <vers>Je fais souvent ce rêve étrange et pénétrant</vers>
    <vers>D'une femme inconnue, et que j'aime, et qui m'aime</vers>
    <vers>Et qui n'est, chaque fois, ni tout à fait la même</vers>
    <vers>Ni tout à fait une autre, et m'aime et me comprend.</vers>
  </strophe>
  <strophe>
    <vers>Car elle me comprend, et mon coeur, transparent</vers>
    <vers>Pour elle seule, hélas ! cesse d'être un problème</vers>
    <vers>Pour elle seule, et les moiteurs de mon front blême,</vers>
    <vers>Elle seule les sait rafraîchir, en pleurant.</vers>
  </strophe>
  <strophe>
    <tercet>
      <vers>Est-elle brune, blonde ou rousse ? – Je l'ignore.</vers>
      <vers>Son nom ? Je me souviens qu'il est doux et sonore</vers>
      <vers>Comme ceux des aimés que la Vie exila.</vers>
    </tercet>
    <tercet>
      <vers>Son regard est pareil au regard des statues,</vers>
      <vers>Et, pour sa voix, lointaine, et calme, et grave, elle a </vers>
      <vers>L'inflexion des voix chères qui se sont tues.</vers>
    </tercet>
  </strophe>
</poeme>
```


Exercice 2 : Rédiger des déclarations d'éléments pour du théâtre

```
<?xml version="1.0" encoding="UTF-8"?>
<piece>
  <titre>Le Malade Imaginaire</titre>
  <auteur>Molière</auteur>

  <acte>
    <titre>Acte I</titre>
    <scene>
      <titre>Scène 1</titre>
      <replique>
        <personnage>Argan</personnage>
        <texte>Hé bien, Monsieur Purgon, que me conseillez-vous de faire ?</texte>
      </replique>
      <didascalie>(Il se tourne vers son médecin avec anxiété.)</didascalie>
      <replique>
        <personnage>Monsieur Purgon</personnage>
        <texte>Je vous conseille de continuer votre régime strictement.</texte>
      </replique>
    </scene>
    <scene>
      <titre>Scène 2</titre>
      <replique>
        <personnage>Toinette</personnage>
        <texte>Ah, Monsieur Argan, vous ne vous portez que trop bien !</texte>
      </replique>
      <didascalie>(Toinette, en aparté, fait signe qu'elle plaisante.)</didascalie>
    </scene>
  </acte>

  <acte>
    <titre>Acte II</titre>
    <scene>
      <titre>Scène 1</titre>
      <replique>
        <personnage>Argan</personnage>
        <texte>Mon coeur, vous êtes cruelle !</texte>
      </replique>
      <didascalie>(Il pose la main sur sa poitrine, feignant un malaise.)</didascalie>
      <replique>
        <personnage>Angélique</personnage>
        <texte>Père, je n'ai que de l'amour pour vous.</texte>
      </replique>
    </scene>
  </acte>
</piece>
```

L'attribut dans la DTD

1. Définir un Attribut dans une DTD

Syntaxe Générale :

```
<!ATTLIST nom_de_l_element nom_de_l_attribut type_de_l_attribut  
valeur_par_défaut_ou_contrainte>
```

Exemple :

```
<!ELEMENT title (#PCDATA)>  
<!ATTLIST title lang CDATA #IMPLIED>
```

Exemple de Document XML-TEI :

```
<title lang="fr">Les Misérables</title>
```


L'attribut dans la DTD

2. Type de contenu des attributs

Expression	Signification
CDATA	chaîne de caractères ne comprenant pas de balises
(val1 val2 ...)	liste de valeurs à utiliser
ENTITY / ENTITIES	entité déclarée dans la DTD (ou liste séparée par des espaces)
ID	pour identifier l'élément
IDREF / IDREFS	ID d'un autre élément (ou liste séparée par des espaces)

2. Types de contenus des attributs

1. **CDATA** - Chaîne de caractères ne comprenant pas de balises

DTD:

```
<!ELEMENT note (#PCDATA)>  
<!ATTLIST note content CDATA #REQUIRED>
```

XML Valide:

```
<note content="This is a simple note.">Take note of this information.</note>
```


2. Types de contenus des attributs

2. (*val1* | *val2* | ...) - Liste de valeurs à utiliser

DTD:

```
<!ELEMENT book (#PCDATA)>  
<!ATTLIST book genre (fiction | non-fiction | poetry) "fiction">
```

XML Valide:

```
<book genre="poetry">The Road Not Taken</book>
```

2. Types de contenus des attributs

3. *Entity* / *Entities* -Entité déclarée dans la DTD (ou liste séparée par des espaces)

DTD:

```
<!ENTITY author "Victor Hugo">  
<!ELEMENT book (#PCDATA)>  
<!ATTLIST book author ENTITY #IMPLIED>
```

XML Valide:

```
<book author="author">Les Misérables</book>
```


2. Types de contenus des attributs

4. ID - Pour identifier l'élément

DTD:

```
<!ELEMENT chapter (#PCDATA)>  
<!ATTLIST chapter id ID #REQUIRED>
```

XML Valide:

```
<chapter id="chap1">Introduction</chapter>
```

2. Types de contenus des attributs

5. IDREF / IDREFS - ID d'un autre élément (ou liste séparée par des espaces)

DTD:

```
<!ELEMENT book (chapter)>
<!ELEMENT chapter (#PCDATA)>
<!ATTLIST book id ID #REQUIRED>
<!ATTLIST chapter bookRef IDREF #REQUIRED>
```

XML Valide:

```
<book id="book1">
  <chapter bookRef="book1">Les Misérables</chapter>
</book>
```


L'attribut dans la DTD

3. Type d'attributs

Opérateur	Signification
#REQUIRED	valeur requise dans l'élément
#IMPLIED	valeur facultative
#FIXED "valeur"	valeur fixe pour l'attribut
"valeur"	valeur par défaut de l'attribut (on peut la remplacer)

3. Types d'attributs

1. **#REQUIRED** - Valeur requise dans l'élément

DTD:

```
<!ELEMENT book (#PCDATA)>  
<!ATTLIST book author CDATA #REQUIRED>
```

XML Valide:

```
<book author="Victor Hugo">Les misérables</book>
```


3. Types d'attributs

2. #IMPLIED - Valeur facultative

DTD:

```
<!ELEMENT book (#PCDATA)>  
<!ATTLIST book author CDATA #IMPLIED>
```

XML Valide:

```
<book author="Victor Hugo">Les misérables</book>
```

Ou sans l'attribut, toujours valide :

```
<book>Les misérables</book>
```

3. Types d'attributs

3. #FIXED « valeur » - Valeur fixe pour l'attribut

DTD:

```
<!ELEMENT book (#PCDATA)>  
<!ATTLIST book lang CDATA #FIXED "fr">
```

XML Valide:

```
<book lang="fr">Les Misérables</book>
```


3. Types d'attributs

4. « valeur » - Valeur fixe pour l'attribut (remplaçable)

DTD:

```
<!ELEMENT book (#PCDATA)>  
<!ATTLIST book lang CDATA "fr">
```

XML Valide avec valeur par défaut :

```
<book>Les Misérables</book>
```

Ou en remplaçant la valeur par défaut :

```
<book lang="en">Les Misérables</book>
```

Exercice 1 : Rédiger des déclarations d'attributs pour de la poésie

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE poeme SYSTEM "Mon_reve_familier.dtd">

<poeme type="sonnet">
  <titre>Mon rêve familial</titre>
  <auteur>Paul Verlaine</auteur>
  <strophe type="quatrain" n="1">
    <vers>Je fais souvent ce rêve étrange et pénétrant</vers>
    <vers>D'une femme inconnue, et que j'aime, et qui m'aime</vers>
    <vers>Et qui n'est, chaque fois, ni tout à fait la même</vers>
    <vers>Ni tout à fait une autre, et m'aime et me comprend.</vers>
  </strophe>
  <strophe type="quatrain" n="2">
    <vers>Car elle me comprend, et mon coeur, transparent</vers>
    <vers>Pour elle seule, hélas ! cesse d'être un problème</vers>
    <vers>Pour elle seule, et les moiteurs de mon front blême,</vers>
    <vers>Elle seule les sait rafraîchir, en pleurant.</vers>
  </strophe>
  <strophe type="sizain">
    <tercet n="1">
      <vers>Est-elle brune, blonde ou rousse ? – Je l'ignore.</vers>
      <vers>Son nom ? Je me souviens qu'il est doux et sonore</vers>
      <vers>Comme ceux des aimés que la Vie exila.</vers>
    </tercet>
    <tercet n="2">
      <vers>Son regard est pareil au regard des statues,</vers>
      <vers>Et, pour sa voix, lointaine, et calme, et grave, elle a </vers>
      <vers>L'inflexion des voix chères qui se sont tues.</vers>
    </tercet>
  </strophe>
</poeme>
```


Exercice 2 : Rédiger des déclarations d'attributs pour du théâtre

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE piece SYSTEM "theatre.dtd">

<piece>
  <titre>Le Malade Imaginaire</titre>
  <auteur>Molière</auteur>

  <acte numero="1">
    <titre>Acte I</titre>
    <scene numero="1">
      <titre>Scène 1</titre>
      <replique personnageID="p1">
        <personnage id="p1">Argan</personnage>
        <texte>Hé bien, Monsieur Purgon, que me conseillez-vous de faire ?</texte>
      </replique>
      <didascalie>(Il se tourne vers son médecin avec anxiété.)</didascalie>
      <replique personnageID="p2">
        <personnage id="p2">Monsieur Purgon</personnage>
        <texte>Je vous conseille de continuer votre régime strictement.</texte>
      </replique>
    </scene>
    <scene numero="2">
      <titre>Scène 2</titre>
      <replique personnageID="p3">
        <personnage id="p3">Toinette</personnage>
        <texte>Ah, Monsieur Argan, vous ne vous portez que trop bien !</texte>
      </replique>
      <didascalie>(Toinette, en aparté, fait signe qu'elle plaisante.)</didascalie>
    </scene>
  </acte>

  <acte numero="2">
    <titre>Acte II</titre>
    <scene numero="1">
      <titre>Scène 1</titre>
      <replique personnageID="p1">
        <personnage id="p1">Argan</personnage>
        <texte>Mon coeur, vous êtes cruelle !</texte>
      </replique>
      <didascalie>(Il pose la main sur sa poitrine, feignant un malaise.)</didascalie>
      <replique personnageID="p4">
        <personnage id="p4">Angélique</personnage>
        <texte>Père, je n'ai que de l'amour pour vous.</texte>
      </replique>
    </scene>
  </acte>
</piece>
```

L'entité dans la DTD

1. Définir un Entité dans une DTD

Syntaxe Générale :

```
<!ENTITY nom_de_l_entité "définition_de_l_entité">
```

Exemple :

```
<!ENTITY author "Victor Hugo">  
<!ELEMENT book (#PCDATA)>  
<!ATTLIST book author ENTITY #REQUIRED>
```

Exemple de Document XML-TEI :

```
<!DOCTYPE book [  
  <!ENTITY author "Victor Hugo">  
<book author="author">Les Misérables</book>
```


L'entité dans la DTD

2. Utiliser des entités prédéfinies

Entité	Caractère
<code>&lt;</code>	<
<code>&gt;</code>	>
<code>&apos;</code>	'
<code>&quot;</code>	"
<code>&amp;</code>	&

Exemple avec l'esperluette :

```
<text>Il est important de comprendre les signes < et > dans un document XML.  
Utilisez &amp; pour représenter le caractère &.</text>
```

**Exercice : proposez une DTD pour le fichier
exemple-complexe.xml**

L'entité dans la DTD

3. Définir ses entité

DTD :

```
<!ENTITY uvsq "Université de Versailles Saint-Quentin-en-Yvelines">  
<!ELEMENT p (#PCDATA)>
```

Exemple de Document XML

Document XML :

```
<!DOCTYPE university [  
  <!ENTITY uvsq "Université de Versailles Saint-Quentin-en-Yvelines">  
<p>Bienvenue à &uvsq; !</p>
```

Le résultat affiché sera :

```
<p>Bienvenue à Université de Versailles Saint-Quentin-en-Yvelines !</p>
```