



INTRODUCTION À XML

Elina Leblanc – Faculté des Lettres (UNIGE)



LE LANGAGE XML

Introduction théorique

Vous avez dit XML ?

- XML (W3C, 1998) : *eXtensible Markup Language*.
- Simplification d'un ancien langage : SGML (Standard Generalized Markup Language, ISO standard, 1986).
- Langage **libre** (non-propriétaire) et **indépendant** des outils qui l'utilisent. Peut être lu par n'importe quel outil ➔ **Stabilité** et **interopérabilité**.
- XML n'est pas tout à fait un langage, mais plutôt un métalangage. Il est à la base de nombreux autres standards (XML-TEI, XML-RDF, etc.).
- Séparation du contenu (structure, sens) de la forme (design).

Un langage polymorphe

- En raison de sa flexibilité, XML est partout :
 - Dans votre vie quotidienne : sur votre téléphone, votre ordinateur, votre liseuse, votre console de jeu, même les stations essences !
 - Dans le monde de l'édition : XML est un fichier source qui peut être transformé dans d'autres formats, qu'ils soient imprimés ou numériques (HTML, PDF, ePub, DocX).
- Quelques exemples d'utilisation multiple de XML :
 - *Aphrodisias in Late Antiquity*, Charlotte Roueché (King's College London)
 - <http://insaph.kcl.ac.uk/ala2004/index.html>
 - METOPES (MRSH de Caen)
 - http://www.unicaen.fr/recherche/mrsh/document_numerique/projets/metopes

De quoi ai-je besoin ?

- Pour faire du XML, vous avez besoin :
 - De connaître et de respecter la syntaxe XML (Slide suivante ;) ;
 - D'un éditeur XML.

Rien de plus, rien de moins ! (Pour le moment...)

- NB : Comment choisir un éditeur XML ? Vous pouvez faire du XML avec n'importe quel éditeur (*Notepad*, *Notepad++*, *SublimeText*, *Jedit*, *Oxygen XML editor*, etc.). Cependant, certains éditeurs ont été pensés spécifiquement pour XML : on les appelle des **éditeurs XML**. Ils proposent des outils qui facilite l'encodage en XML : on dit qu'ils sont « **XML aware** ».



XML EN PRATIQUE

La syntaxe XML

- Le XML repose sur une structure arborescente : une **racine**, plusieurs **nœuds**.

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

```
<library>
```

```
<book>
```

```
<author> J. R. R. Tolkien </author>
```

```
<title> The Lord of the rings : The fellowship of the ring </title>
```

```
</book>
```

```
<book>
```

```
<author> Lewis Carroll </author>
```

```
<title> Alice's Adventures in Wonderland </title>
```

```
</book>
```

```
</library>
```

Les éléments (I)

- Unité de base de la syntaxe XML.
- Un élément est composé de **2 balises** :
 - La balise ouvrante commence avec **<** et finit avec **>**.
 - La balise fermante commence avec **</** et finit avec **>**.

Ex. : **<sentence>**We love XML.**</sentence>**

- Attention ! Certains éléments sont composés d'une seule balise et sont appelés des **éléments vides**. On peut les voir comme la contraction d'une balise ouvrante et d'une balise fermante.

<emptyTag/>

Les éléments (II)

- **La règle des poupées russes** : Les éléments sont toujours imbriqués les uns dans les autres. Ils ne se chevauchent jamais.

<title><italic>Le seigneur des anneaux</title></italic>



<title><italic>Le seigneur des anneaux</italic></title>



Les attributs et les valeurs

`<sentence type="citation">Que la force soit avec toi !</sentence>`

- Un attribut décrit un élément. Il donne des informations supplémentaires sur la langue, la largeur, la hauteur, le type, *etc.*
- Les attributs fonctionnent par pair : `nom="valeur"`.
- Les attributs sont uniquement insérés dans la balise ouvrante.
- Un élément peut avoir plusieurs attributs.

Let's try : Encodons une bibliothèque !

- Dans *Oxygen*, ouvrir le projet `my-library.xpr` ;
- Dans le fichier `my-library.xml`, encoder les références bibliographiques contenues dans `my-library.txt`, en vous aidant du modèle proposer pour le roman *Shirley*.
- Durée : 30 min

Oxygen m'aide. Pourquoi ?

- Une seule réponse : le schéma !
- Un schéma est la grammaire et le vocabulaire d'un langage XML particulier. Il décrit :
 - Les éléments, les attributs et les valeurs qui **peuvent être utilisés** dans un fichier XML.
 - **La manière dont vous pouvez utiliser** ces éléments et attributs (obligatoires, optionnels, répétables).
- XML vous offre des règles d'écriture, mais les noms des éléments et des attributs sont définis dans le schéma (`.dtd`, dans notre cas).
- Exemple : « Library » est un langage qui utilise le métalangage XML, mais ses éléments et attributs sont définis dans `library.dtd`.

Trucs et astuces avec Oxygen

- Suggestions d'éléments, quand vous commencez à écrire une balise.
- Auto-complétion : *Oxygen* ferme automatiquement les éléments.
 - Pour désactiver cette fonctionnalité : Option > Preferences > Editor > Content Completion > Deselect "Close the inserted element".
- Le carré vert/rouge indique si votre document contient des erreurs ou non (c'est-à-dire s'il est **valide**, en respectant le schéma utilisé et la syntaxe XML).
- **CTRL + E** : Permet d'encoder une portion de texte.
- Mode auteur : pour contrôler votre encodage.

Oxygen : 7 reasons why

- *Oxygen* est actuellement considéré comme l'un des meilleurs éditeurs XML :
 - Rapide à installer ;
 - Facile à utiliser ;
 - Outil puissant, qui s'adresse aussi bien aux débutants qu'aux experts ;
 - « XML aware » : plusieurs fonctionnalités pour aider les utilisateurs à encoder ;
 - En accord avec les besoins des chercheurs en humanités ;
 - Recommandé par les instituts de recherche ;
 - L'un des moins chers du marché.