

Institut Lallemand
Rue du Couvent, 2 – 1050 Ixelles
Année académique : 2016/2017

SECTION : BACHELIER
BIBLIOTHÉCAIRE-DOCUMENTALISTE
UNITÉ DE FORMATION 24/01 :
LABORATOIRE INFORMATIQUE DE
GESTION DOCUMENTAIRE

Présenté par Alexandre Thiry

Synthèse documentaire

Le XML

Extensible Markup Language

Le XML

*« Be strict in what you produce,
Be generous in what you accept »*

Table des matières

Introduction	4
1. Qu'est-ce que le XML ?	4
2. D'où vient le XML ?	4
3. À quoi sert le XML ?	5
4. Quels sont ses avantages ?	5
5. Quelles sont ses applications ?	6
6. Comment est structuré le XML ?	6
7. Les définitions	7
a. Le Document Type Definition	7
8. L'Affichage	7
9. Les éditeurs de XML	9

Introduction

À travers cette synthèse documentaire, nous allons tenter de définir le XML, d'expliquer son utilité, son fonctionnement et quelques-unes de ses applications.

1. Qu'est-ce que le XML ?

XML signifie « Extensible Markup Language » ou « Langage de balisage extensible » en français.

Le XML est un « langage informatique de balisage générique »¹. Générique renvoie au fait que les balises sont extensibles plutôt que figées (comme le sont celles du HTML), l'utilisateur est libre de générer de nouvelles balises.

2. D'où vient le XML ?

Le XML est apparu suite à la nécessité d'uniformiser la syntaxe de descriptions des données afin de mieux pouvoir en échanger sur le web ; il découle du SGML.

En 1969, C. Goldfarb, E. Mosher et R. Lorie inventent le GML. Ce langage sert à éditer des documents, les mettre en page et les partager au sein de systèmes de gestion éditoriaux².

En 1989 apparaît la norme ISO 8879 ou SGML (Standard Generalized Markup Language).

Cette norme définit un métalangage de balisage de documents, lisible par un être humain et traitable par une machine.

Ses objectifs sont la séparation du fond (contenu) de la forme (balises). Le support de traitements sur le contenu des documents sans prise en compte de la forme. Et l'intégration d'un format de stockage et d'échange normalisé.

Cependant le SGML est un format très lourd et complexe à mettre en œuvre. Sa syntaxe doit être suivie rigoureusement.

¹ <https://openclassrooms.com/courses/structurez-vos-donnees-avec-xml/qu-est-ce-que-le-xml#/id/r-1769953>

² REYNIER. « XML : documents et outils ». 2011, p11.

Un groupe d'informaticiens ayant de l'expérience dans le SGML et le web a alors décidé de se pencher sur le sujet. Le langage XML est donc né. Le XML 1.0 est devenu une recommandation du W3C (le "World Wide Web Consortium") le 10 février 1998.

Depuis, les spécifications du langage ont évolué, et la version 1.1 a été publiée le 4 février 2004. C'est pourtant la version 1.0 du langage XML qui est encore la plus utilisée³.

3. À quoi sert le XML ?

Le XML a été créé pour structurer l'information des documents. Il permet de décrire, structurer, stocker, transporter et échanger des données. Il permet de faciliter l'échange automatisé de contenus entre des systèmes d'informations variés grâce à ses balises de descriptions de contenu.

Il s'agit d'un méta-langage permettant d'écrire d'autres langages ; de nombreux langages sont aujourd'hui basés sur le XML.

Il n'est pas limité à un seul type d'application.

Il peut par exemple structurer les applications d'échanges de données entre machines MOM « Message Oriented Middleware ».

Ou bien être POP « Presentation Oriented Publishing », utilisé à des fins de présentation des données à un utilisateur humain⁴.

4. Quels sont ses avantages ?

Par rapport à la norme SGML ; il supporte nativement les différents codages internationaux. Il est plus facile d'utilisation et plus grand public que cette dernière puisque la définition du type de document (DTD) y est optionnelle, que les liens hypertextes se sont facilement, etc.

Il est également plus ouvert que l'HTML (langage spécialisé dérivé du SGML).

³ <https://openclassrooms.com/courses/structurez-vos-donnees-avec-xml/qu-est-ce-que-le-xml#/id/r-1769982>

⁴ REYNIER. « XML : documents et outils ». 2011, p26.

5. Quelles sont ses applications ?

Il est à la base du XHTML⁵ qui était vu il fut un temps comme le successeur désigné de l'HTML (aujourd'hui c'est HTML5 qui est à la pointe). Le XHTML est une reformulation du HTML4 avec la syntaxe et les règles rigoureuses du XML.

WML ou Wireless Markup Language, lié au protocole WAP de transmission sans fil de données pour téléphones mobiles ;

MathML ou Mathematical Markup Language qui permet l'affichage des symboles mathématiques ;

SOAP : Protocole d'appel de procédure à distance

6. Comment est structuré le XML ?

Comme le HTML, le XML est basé sur le SGML ou Standard Generalized Markup Language qui est le langage normalisé de description à balises de référence. Tous deux servent à transporter des données sous forme textuelle. Le XML est une version simplifiée du SGML et plus adaptée au Web.

C'est un langage professionnel assez abstrait.

Le fait que le XML soit extensible signifie que l'utilisateur est libre de créer ses propres balises, elles ne sont pas figées. Il existe cependant des règles de syntaxe à respecter. On parle alors de document bien formé.

- Les balises sont encadrées par des chevrons <>exemple</>
- Les noms de balises ne peuvent pas commencer par un chiffre ou un signe de ponctuation ;
- Ils ne peuvent commencer par les lettres XML ;
- Ils ne peuvent contenir d'espace ;
- Les titres sont sensibles à la case (majuscules) ;
- Il n'existe pas de balises solitaires, une balise est toujours accompagnée d'une balise de fermeture. Ex : liste ;

⁵ Université d'Artois. « Cours XML + XSL », 2009, p. 1.

- Les balises doivent être correctement imbriquées, c'est-à-dire qu'une balise ouverte au sein d'une autre balise, doit être également fermée au sein de celle-ci.

Ex : `<grande_balise><petite_balise</grande_balise></petite_balise>`

- C'est un langage hiérarchisé, la première balise est englobante, elle est dite balise racine. Les balises comprises entre ses chevrons sont des balises subordonnées.
- Les attributs accompagnants les balises doivent impérativement être mis entre guillemets.

7. Les définitions

Un document qui respecte les définitions est dit document valide⁶.

a. Le Document Type Definition

Le Document Type Definition, ou DTD est un ensemble de règles et de propriétés que l'on impose au document XML. Ces règles définissent le nom et le contenu de chaque balise

Il y a des DTD interne, propre aux documents. Mais aussi des DTD externes, utilisés par différents développeurs pour garantir une norme commune de transfert de données et pouvoir automatiser le traitement par un logiciel⁷.

8. L'Affichage

Seul, le XML n'est que du texte brut, il peut être codé dans un simple bloc-note. Une fois le fichier enregistré et ouvert dans un navigateur (Firefox, Chrome, Opera...), les balises s'afficheront généralement dans une couleur différente des attributs et du contenu. Cela reste cependant du code difficilement visualisable. Il est toutefois possible, comme pour le HTML, de créer une feuille de style (CSS, XSL) pour mettre en forme le

⁶ <https://openclassrooms.com/courses/structurez-vos-donnees-avec-xml/introduction-aux-definitions-et-aux-dtd#/id/r-1770312>

⁷ Université d'Artois. « Cours XML + XSL », 2009, p. 6.

code XML. Il suffit pour cela de créer un élément de style lié à chaque balise.

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<racine>
  <enfant>
    <nom>Loïc</nom>
    <lien>garçon</lien>
    <date>07/11/83</date>
    <data>Le petit qui me dépasse d'une tête.</data>
  </enfant>
  <enfant>
    <nom>Marine</nom>
    <lien>fille</lien>
    <date>20/12/85</date>
    <data>La petite fille chérie à son papa.</data>
  </enfant>
</racine>
```

Figure 1. XML de départ⁸.

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
- <racine>
-   <enfant>
      <nom>Loïc</nom>
      <lien>garçon</lien>
      <date>07/11/83</date>
      <data>Le petit qui me dépasse d'une tête.</data>
    </enfant>
-   <enfant>
      <nom>Marine</nom>
      <lien>fille</lien>
      <date>20/12/85</date>
      <data>La petite fille chérie à son papa.</data>
    </enfant>
  </racine>
```

Figure 2. Version navigateur.

⁸ Université d'Artois. « Cours XML + XSL », 2009, p. 10.



Figure 3. Après application d'une feuille de style.

9. Les éditeurs de XML

Afin de faciliter l'édition du code XML, il existe de nombreux logiciels dédiés.

On peut les classer en deux catégories⁹ :

1. Les éditeurs de texte, basés sur la ligne de code ;
2. Les éditeurs visuels, basés sur une interface graphique.

Les premiers affichent les balises de façon brute. Ils permettent d'afficher les éléments constituant le code et permettent des manipulations précises au sein de ce dernier. Le texte est cependant rarement complètement brute, et la plupart des éditeurs de texte proposent depuis les années 80's la coloration syntaxique. Cette dernière permet la mise en valeur des mots-clés d'un langage de programmation (ici les balises), ce qui aide à en repérer plus rapidement les éléments constitutants.

Une autre fonction courante est l'affichage du numéro de ligne de code.

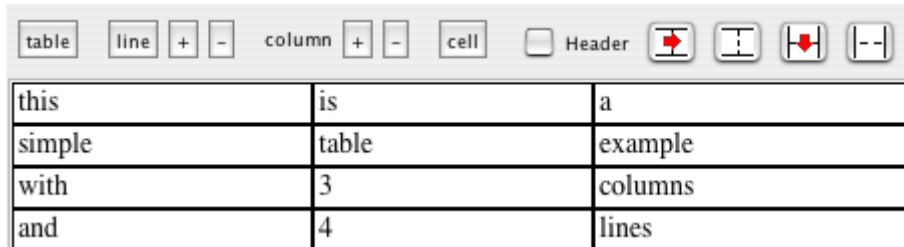
```
<TABLE>
<TR><TD>this</TD><TD>is</TD><TD>a</TD></TR>
<TR><TD>simple</TD><TD>table</TD><TD>example</TD></TR>
<TR><TD>with</TD><TD>3</TD><TD>columns</TD></TR>
<TR><TD>and</TD><TD>4</TD><TD>lines</TD></TR>
</TABLE>
```

Figure 4. Éditeur de texte

Les éditeurs visuels, eux, permettent de schématiser les données pour aider l'utilisateur à les encoder sans se plonger dans la syntaxe. Ils sont une aide non-

⁹ XML editor. (2016, September 10). In Wikipedia, The Free Encyclopedia.

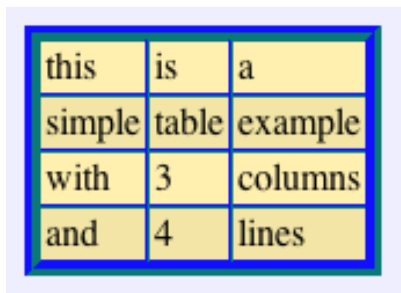
négligeable pour le débutant puisqu'ils évitent les erreurs de syntaxe. Cependant ils sont beaucoup plus rigides que les éditeurs de texte, alors que ces derniers laissent le champ libre à plus de créativité.



this	is	a
simple	table	example
with	3	columns
and	4	lines

Figure 5. Éditeur visuel

Parmi les éditeurs visuels certains permettent d'obtenir le rendu visuel final après traitement par une feuille de style. Se baser sur le rendu visuel peut cependant induire en erreur, puisque ce rendu est subjectif et dépend du système utilisé (l'affichage d'un navigateur à l'autre d'un même code n'est pas identique).



this	is	a
simple	table	example
with	3	columns
and	4	lines

Figure 6. Avec feuille de style

Conclusion

XML est un langage de description puissant qui s'adapte au besoin de l'utilisateur grâce à son côté extensible. Une fois les balises définies, elles pourront être réutilisées dans des applications XML.

Il permet l'échange de données entre des systèmes variés dans un langage lisible. Enfin XML est un langage non-propriétaire dont la norme et les spécificités ont été conçues par le W3C, un consortium international dédié au développement des normes et spécifications du Web.

Aujourd'hui de très nombreuses applications reposent sur le XML.