

# **Présentation du langage XML**

- **Histoire et origine**
- **Caractéristiques**
- **Composants et fonctionnement**
- **Comparaisons et avantages**
- **Outils et utilisations**

*Liticia Hamrene*

*Adrien Dusserre*

*Jules Boissel*

*William Xu*

# Histoire et origine

Le langage XML, c'est à dire **Extensible Markup Language** est un métalangage de balisage dit extensible inventé en 1996 par le **World Wide Web Consortium** dans le but de créer un langage de structures de données simple et efficace.

En effet, il existe alors le **SGML ( Standard Generalized Markup Language )** qui est considéré comme pas assez simple d'utilisation et se doit donc d'être remplacé par ce nouveau langage que va être le XML.

Il finit d'être développé en 1998 et devient une norme W3C le 10 février 1998.

Il a pour caractéristique, comme son prédécesseur et en tant que métalangage, de pouvoir servir de structure de données lisible pour de nombreux langages de programmation et applications.

Il est aussi normé de façon à ce que tout document XML puisse être traité avec les codages de caractères UTF-8 et 16.

Il a donc été largement adopté dans les années 2000 puis a par la suite été concurrencé par des formats plus efficaces encore tels que le **YAML ( Yet Another Markup Language )** et le **JSON ( JavaScript Object Notation )**.

# Caractéristiques

Contrairement à d'autres métalangages, le XML ne demande pas de connaissances techniques en programmation pour être utilisé.

Etant naturellement structuré, il est facile à lire et à comprendre, ce qui contribue grandement à rendre ce langage très accessible.

En effet, pas de besoin d'un **IDE ( *Integrated Development Environment* )** pour l'éditer, un simple logiciel de traitement de texte comme le bloc-note peut faire l'affaire.

Il a également comme caractéristique d'être universel, c'est-à-dire qu'il ne rencontre pas de problème de compatibilité avec d'autres langages ou applications souhaitant l'utiliser.

Enfin, il est dit extensible, puisqu'il est toujours possible d'ajouter un nombre illimité de nouvelles balises suivant les besoins.

# Composants et fonctionnement

La syntaxe du langage XML repose sur une chaîne de caractères structurée en deux niveaux :

- Un pour le lecteur humain qui va avoir pour objectif de simplifier la lecture et donc l'utilisation du langage.
- Et un autre pour la machine qui aura donc comme but d'être efficace et fiable en terme de traitement de l'information.

Un document XML peut être vu sous la forme d'un arbre, dont le tronc sert de support à différents types d'éléments appelés "nœuds", qui peuvent être des textes, des attributs, des commentaires ou encore des éléments en tout genre.

Voyons maintenant différents composants du langage XML :

- Le Nœud Document : Il est la racine du document XML et se doit d'être unique de la même façon que le *main* dans un programme en C.
- Les Nœuds Élément : Ils sont qualifiés au sein d'un espace de noms et peuvent contenir des attributs en tout genre. Ces éléments peuvent être imbriqués mais pas entrecroisés. De plus, ces nœuds peuvent être Répétables, Ordonnés, Hiérarchisés, Mélangés et Qualifiés selon les besoins des développeurs.
- Les Balises : Les balises sont un nom délimité par deux chevrons qui ont pour rôle de contenir les éléments d'un fichier XML en permettant leur délimitation au sein de ce dernier.
- Les Nœuds Texte : Ils permettent de contenir du texte au sein d'un document XML. Ils servent également à définir l'encodage de ce dernier ainsi que les potentiels espacements.
- Les Attributs : Ils permettent de stocker des variables et ont les mêmes contraintes que les éléments bien qu'ils ne s'utilisent pas exactement de la même façon étant donné qu'ils sont systématiquement portés par un élément.

# Comparaisons et avantages

Le langage XML à de nombreux avantages par rapport à ses concurrents les plus communs que sont le **RTF ( Rich Text Format )**, le **TeX** ou encore le **Wiki**.

Ses principaux avantages sont les suivants :

- La Lisibilité : Il est facile pour un humain de lire un fichier XML car le code est structuré et facile à comprendre. En principe, il est même possible de dire qu'aucune connaissance spécifique n'est nécessaire pour comprendre les données comprises à l'intérieur d'un document XML.
- La Disponibilité : Ce langage est libre et un fichier XML peut être créer à partir d'un simple logiciel de traitement de texte, un simple bloc-notes suffit.
- L'Interopérabilité : Quel que soit le système d'exploitation ou la technologie utilisée, il n'y a pas de problème pour lire ce langage.
- L'Extensibilité : De nouvelles balises peuvent être ajoutée de façon illimitée.
- Plusieurs parseurs XML différent doivent en principe ( S'ils sont bien codés ) produire le même résultat.
- Tous les navigateurs internet récents intègrent un parseur XML permettant la lecture des documents de ce langage.

# Outils et utilisations

Afin de procéder à une édition correcte d'un fichier XML, il s'agit de suivre canoniquement les étapes suivantes :

- La Production : Un document XML peut être produit manuellement par un humain ou résulter d'un export d'une source de donnée depuis un autre format.
- L'Analyse : Un **parser** lit le fichier XML pour construire un objet informatique, il vérifie également que le document est bien conforme aux différentes normes du format XML.
- L'Inclusion : Il est parfois nécessaire d'inclure certains fichiers qui deviendront aux mêmes des nœuds.
- La Validation : Le document est ensuite vérifié par un algorithme qui confirme que la structure du fichier généré est bien conforme à celle du document XML traité.
- La Transformation : Le document peut désormais être transformé en un autre format à destination d'un navigateur, d'un traitement de texte ou encore d'un tableur.