

La syntaxe du XML

Le XML impose des règles de syntaxe très strictes et très spécifiques par rapport au HTML. On retrouvera ces mêmes règles de syntaxe dans tous les langages dérivés du XML, dont le XHTML.

Passons ces règles en revue :

- Le XML est un langage de balises. Mais au contraire du HTML où les balises sont définies, vous devez ou pouvez inventer vos propres balises. Rappelez-vous, le XML est eXtensible. Il faut donc définir soi-même le nom des balises utilisées.

Il y a quelques règles pour la composition des noms (mais elles ne déroutent pas les habitués du JavaScript) :

- Les noms peuvent contenir des lettres, des chiffres ou d'autres caractères.
- Les noms ne peuvent débuter par un nombre ou un signe de ponctuation.
- Les noms ne peuvent commencer par les lettres xml (ou XML ou Xml...).
- Les noms ne peuvent contenir des espaces.
- La longueur des noms est libre mais on conseille de rester raisonnable.
- On évitera certains signes qui pourraient prêter à confusion, comme le tiret, le point virgule, le point, la virgule, le signe plus grand que (>) et le signe plus petit que (<).
- Les balises sont sensibles aux majuscules et minuscules (*case sensitive*). Ainsi, la balise <Message> est différente de la balise <message>. La balise d'ouverture et la balise de fermeture doivent donc être identiques. Ainsi par exemple <Message> ... </message> est incorrect et <message> ... </message> est correct.

Un consensus se dégage pour n'écrire les balises qu'en minuscules, limitant ainsi les erreurs possibles.

- Toute balise ouverte doit être impérativement fermée.

Les mauvaises pratiques du HTML, avec lesquelles on pouvait dans certains cas omettre la balise de fin, comme pour le paragraphe <p> ou l'élément de liste , sont révolues.

Ainsi en HTML, ce qui suit est affiché correctement :

```
<p>
<ul>
<li>Point 1
<li>Point 2
```

Le XML est beaucoup plus strict. On devrait avoir :

```
<p>
<ul>
<li>Point 1</li>
<li>Point 2</li>
</ul>
</p>
```

Les éventuelles balises uniques, appelées aussi balises vides, comme
, <meta> ou en HTML, doivent également comporter un signe de fermeture soit balise/. Ainsi une balise <meta/> serait correcte en XML.

- Les balises doivent être correctement imbriquées. Le XML attachant beaucoup d'importance à la structure des données, des balises mal imbriquées sont des fautes graves de sens.

Ainsi, l'écriture suivante est incorrecte car les balises ne sont pas bien imbriquées :
<parent><enfant>Marine</parent></enfant>

L'écriture correcte avec une bonne imbrication des éléments est :
<parent><enfant>Marine</enfant></parent>

- Tout document XML doit comporter une racine.

En fait, la première paire de balises d'un document XML sera considérée comme la balise de racine (*root*).

Par exemple :

```
<racine>
... suite du document XML ...
</racine>
```

En comparaison avec le HTML, l'élément racine est `<html> ... </html>`.

Tous les autres éléments sont imbriqués entre ces balises de racine.

Par exemple :

```
<racine>
<parents>
<enfants>
<petits_enfants> ... </petits_enfants>
</enfants>
</parents>
</racine>
```

- Les valeurs des attributs doivent toujours être mises entre guillemets. Le XML peut, comme le HTML, avoir des attributs comportant des valeurs. En XML, les valeurs des attributs doivent obligatoirement être reprises entre guillemets, à l'inverse du HTML où leur présence ou leur absence n'a plus beaucoup d'importance pour les navigateurs.

Ainsi, l'écriture suivante est incorrecte car il manque les guillemets.

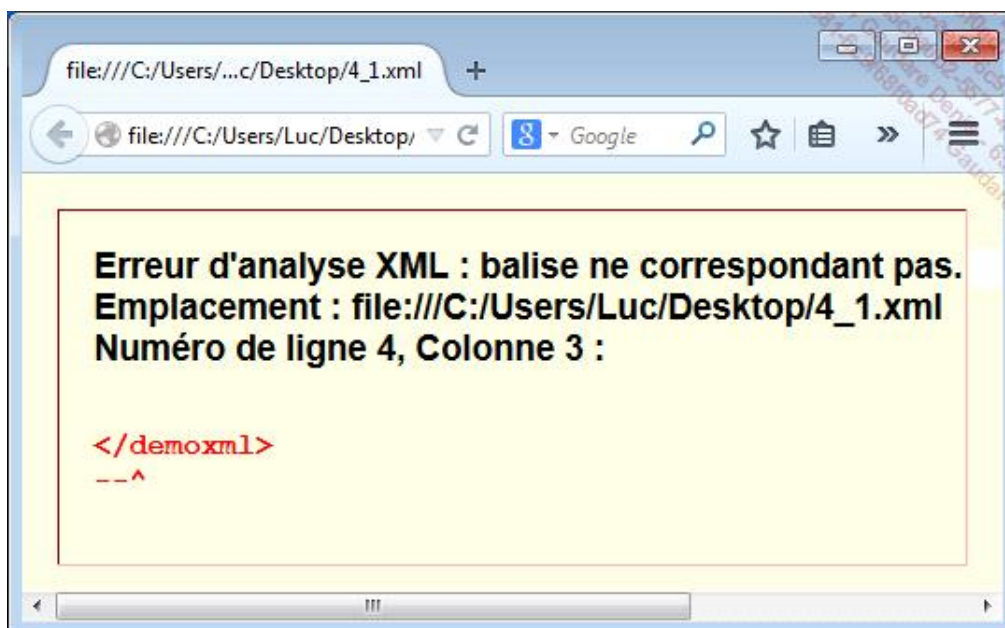
```
<enfant date_anniversaire=071185>
```

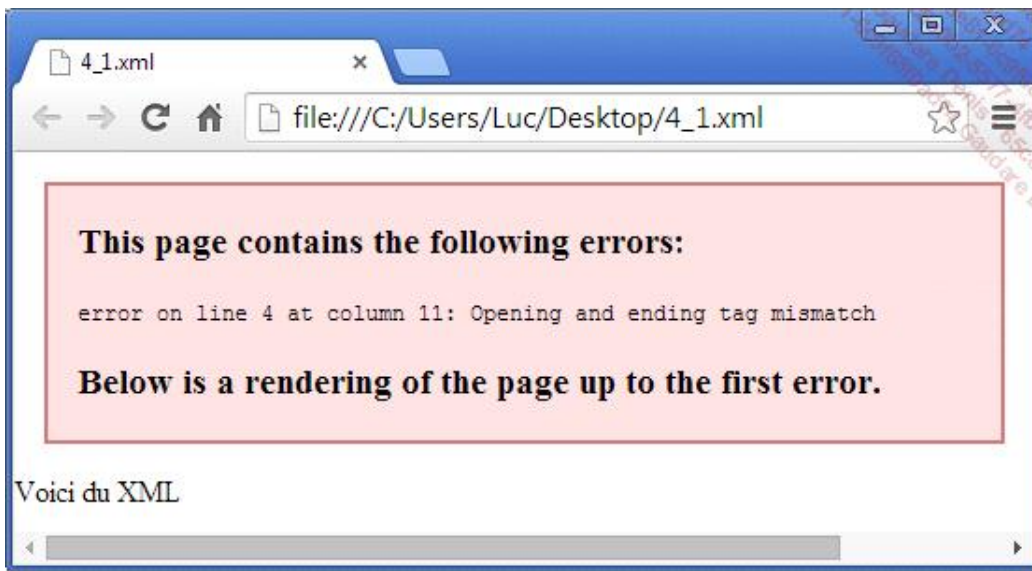
La bonne écriture est :

```
<enfant date_anniversaire="071185">
```

- Le XML est un langage strict. Votre document doit impérativement respecter la syntaxe du XML. On dit alors que le document est **bien formé** (*Well-formed*). Seuls les documents "bien formés" sont affichés correctement. À la moindre erreur de syntaxe, le document n'est pas affiché !

Voici ce qui est affiché dans Firefox et Google Chrome lorsqu'un document XML est incorrect.





➤ Les navigateurs sont aussi un outil de débogage rudimentaire du code XML.