

XML - eXtensible Markup Language

Présentation générale

Julien Barnier

Février 2005

Plan

- 1 Historique
- 2 Structurer l'information
- 3 Syntaxe de XML
 - Éléments du langage
 - Règles du langage
- 4 XML comme métalangage
 - Comparaison avec HTML
 - Définition de document
- 5 Intérêt de XML

Plan

- 1 Historique
- 2 Structurer l'information
- 3 Syntaxe de XML
 - Éléments du langage
 - Règles du langage
- 4 XML comme métalangage
 - Comparaison avec HTML
 - Définition de document
- 5 Intérêt de XML

Historique de XML

- Histoire liée au développement de l'Internet et de l'échange de fichiers.

Historique de XML

- Histoire liée au développement de l'Internet et de l'échange de fichiers.
- Réflexion entamée en 1996 par le W3C.

Historique de XML

- Histoire liée au développement de l'Internet et de l'échange de fichiers.
- Réflexion entamée en 1996 par le W3C.
- Spécification publiée en 1998.

Historique de XML

- Histoire liée au développement de l'Internet et de l'échange de fichiers.
- Réflexion entamée en 1996 par le W3C.
- Spécification publiée en 1998.
- *Objectif* : structurer l'information pour permettre son traitement par un ordinateur.

Plan

- 1 Historique
- 2 Structurer l'information**
- 3 Syntaxe de XML
 - Éléments du langage
 - Règles du langage
- 4 XML comme métalangage
 - Comparaison avec HTML
 - Définition de document
- 5 Intérêt de XML

Structurer l'information

Exemple de document :

Structurer l'information

Exemple de document :

```
Supermarché Lastar
3 rue de la Libération, Givors
le 01/02/2005 à 18h17

3KG CHOUX DE BRUXELLES      2,30
STEAK D'AUTRUCHE             8,78

TOTAL EUROS :                 11,08
TOTAL FRF :                   72,68
```

Structurer l'information

- Document facile à lire et à comprendre pour un lecteur humain francophone.

Structurer l'information

- Document facile à lire et à comprendre pour un lecteur humain francophone.
- Difficile pour un programme informatique de repérer et de donner sens aux éléments d'information importants.

Structurer l'information

C'est la ponctuation et la mise en forme du document qui permettent de donner du sens à l'information qu'il contient :

Structurer l'information

C'est la ponctuation et la mise en forme du document qui permettent de donner du sens à l'information qu'il contient :

```
Supermarché Lastar 3 rue de la Libération Givors  
le 01/02/2005 à 18h17 BIENVENUE ! 3KG CHOUX DE  
BRUXELLES 2,30 STEAK D'AUTRUCHE 8,78 TOTAL  
EUROS : 11,08 TOTAL FRANCS : 72,68
```

Structurer l'information

```
Supermarché      Lastar
                  3 rue de la Libération, Givors

le      01/02/2005      à      18h17

                  3KG CHOUX DE BRUXELLES
2,30

                  STEAK D'AUTRUCHE
8,78

TOTAL EUROS :      11,08
TOTAL FRF :      72,68
```

Structurer l'information

Supermarché <nom>Lastar</nom>

3 rue de la Libération, Givors

le 01/02/2005 à 18h17

3KG CHOUX DE BRUXELLES

2,30

STEAK D'AUTRUCHE

8,78

TOTAL EUROS : 11,08

TOTAL FRF : 72,68

Structurer l'information

Supermarché `<nom>Lastar</nom>`
`<adresse>3 rue de la Libération, Givors</adresse>`

le 01/02/2005 à 18h17

3KG CHOUX DE BRUXELLES

2,30

STEAK D'AUTRUCHE

8,78

TOTAL EUROS : 11,08

TOTAL FRF : 72,68

Structurer l'information

```

<supermarche>
Supermarché <nom>Lastar</nom>
<adresse>3 rue de la Libération, Givors</adresse>
</supermarche>

```

```

le          01/02/2005          à          18h17

```

```

          3KG CHOUX DE BRUXELLES

```

```

2,30

```

```

          STEAK D'AUTRUCHE

```

```

8,78

```

```

TOTAL EUROS :          11,08

```

```

TOTAL FRF :          72,68

```

Structurer l'information

```

<supermarche>
Supermarché <nom>Lastar</nom>
<adresse>3 rue de la Libération, Givors</adresse>
</supermarche>
<commande>
le <date>01/02/2005</date> à <heure>18h17</heure>
<articles>
<article><nom>3KG CHOUX DE BRUXELLES</nom>
<prix>2,30</prix></article>
<article><nom>STEAK D'AUTRUCHE</nom>
<prix>8,78</prix></article>
</articles>
TOTAL EUROS : <total devise="euro">11,08</total>
TOTAL FRF : <total devise="franc">72,68</total>
</commande>

```

Plan

- 1 Historique
- 2 Structurer l'information
- 3 Syntaxe de XML**
 - Éléments du langage
 - Règles du langage
- 4 XML comme métalangage
 - Comparaison avec HTML
 - Définition de document
- 5 Intérêt de XML

Éléments du langage : balises

L'élément fondamental dans un document XML est la balise.

Éléments du langage : balises

L'élément fondamental dans un document XML est la balise.

```
<balise>contenu</balise>
```

Éléments du langage : balises

L'élément fondamental dans un document XML est la balise.

```
<balise>contenu</balise>
```

Exemple :

```
<titre>Les pratiques culturelles des classes  
populaires</titre>
```

Éléments du langage : balises

L'élément fondamental dans un document XML est la balise.

```
<balise>contenu</balise>
```

Exemple :

```
<titre>Les pratiques culturelles des classes  
populaires</titre>
```

Forme courte :

```
<balise />
```

Éléments du langage : attributs

Les attributs servent à donner des informations supplémentaires sur une balise ou son contenu.

Éléments du langage : attributs

Les attributs servent à donner des informations supplémentaires sur une balise ou son contenu.

```
<balise attribut="valeur">contenu</balise>
```

Éléments du langage : attributs

Les attributs servent à donner des informations supplémentaires sur une balise ou son contenu.

```
<balise attribut="valeur">contenu</balise>
```

Exemple :

```
<total devise="euros">11,08</total>
```

```
<total devise="FRF">72,68</total>
```

Éléments du langage : entités

Les entités servent à encoder des caractères spéciaux :

Éléments du langage : entités

Les entités servent à encoder des caractères spéciaux :

`< = <`

`> = >`

`& = &`

`" = "`

`' = '`

Éléments du langage : autres éléments

Éléments du langage : autres éléments

Commentaires :

```
<!-- Texte de commentaire -->
```

Éléments du langage : autres éléments

Commentaires :

```
<!-- Texte de commentaire -->
```

Déclaration :

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

Éléments du langage : autres éléments

Commentaires :

```
<!-- Texte de commentaire -->
```

Déclaration :

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

Instructions de traitement :

```
<?xml-stylesheet type="text/xsl"  
href="article.xsl" ?>
```

Exemple de fichier

Exemple de fichier

```
<?xml version="1.0" encoding="UTF-8" ?>
<?xml-stylesheet type="text/xsl"
href="carnet.xsl" ?>
<!-- Début du carnet d'adresse -->
<carnet>
  <adresse>
    <nom>CRUSOE</nom>
    <prenom>Robinson</prenom>
    <telephone type="fixe">aucun</telephone>
    <telephone type="mobile">aucun</telephone>
  </adresse>
</carnet>
```

Règles du langage

Le document doit contenir un unique élément racine.

Règles du langage

Le document doit contenir un unique élément racine.

Contre-exemple :

```
<?xml version="1.0" encoding="UTF-8" ?>
<adresse>
  <nom>Robinson CRUSOE</nom>
  <telephone>Aucun</telephone>
</adresse>
<adresse>
  <nom>Vendredi LAVIESAUVAGE</nom>
  <telephone>Aucun</telephone>
</adresse>
```

Règles du langage

Les éléments ayant du contenu doivent avoir une balise ouvrante et une balise fermante.

Règles du langage

Les éléments ayant du contenu doivent avoir une balise ouvrante et une balise fermante.

Contre-exemple :

```
<?xml version="1.0" encoding="UTF-8" ?>
<adresse>
  <nom>Robinson CRUSOE
  <telephone>Aucun</telephone>
</adresse>
```

Règles du langage

Les attributs doivent avoir une valeur (éventuellement vide) et celle-ci doit-être entre guillemets simples ou doubles.

Règles du langage

Les attributs doivent avoir une valeur (éventuellement vide) et celle-ci doit-être entre guillemets simples ou doubles.

Exemples :

```
<total devise="euros">11,08</total>
```

```
<total devise='FRF'>72,68</total>
```

```
<total devise="">8215,30</total>
```

Règles du langage

Les attributs doivent avoir une valeur (éventuellement vide) et celle-ci doit-être entre guillemets simples ou doubles.

Exemples :

```
<total devise="euros">11,08</total>  
<total devise='FRF'>72,68</total>  
<total devise="">8215,30</total>
```

Contre-exemples :

```
<total devise=euros>11,08</total>  
<total devise=>72,68</total>
```

Règles du langage

Les balises ne doivent pas se recouvrir, *i.e.* toute balise B ouverte après une balise A doit être fermée avant cette balise.

Règles du langage

Les balises ne doivent pas se recouvrir, *i.e.* toute balise B ouverte après une balise A doit être fermée avant cette balise.

Exemple :

```
<A><B></B></A>
```

Règles du langage

Les balises ne doivent pas se recouvrir, *i.e.* toute balise B ouverte après une balise A doit être fermée avant cette balise.

Exemple :

```
<A><B></B></A>
```

Contre-exemple :

```
<A><B></A></B>
```

Règles du langage

Les caractères de marquage (<, > et &) ne doivent pas être utilisés dans le contenu des balises mais être remplacés par l'entité correspondante.

Règles du langage

Les caractères de marquage (<, > et &) ne doivent pas être utilisés dans le contenu des balises mais être remplacés par l'entité correspondante.

Contre-exemples :

```
<titre>Ma vie & mon oeuvre</titre>  
<condition> x > y </condition>
```

Règles du langage

Les caractères de marquage (<, > et &) ne doivent pas être utilisés dans le contenu des balises mais être remplacés par l'entité correspondante.

Contre-exemples :

```
<titre>Ma vie & mon oeuvre</titre>  
<condition> x > y </condition>
```

À remplacer par :

```
<titre>Ma vie &amp; mon oeuvre</titre>  
<condition> x &gt; y </condition>
```

Fichier XML bien formé

Un fichier XML qui respecte les règles du langage est dit *bien formé*. Ceci signifie qu'il pourra être lu par tout analyseur XML (*XML parser*) conforme aux standards.

Plan

- 1 Historique
- 2 Structurer l'information
- 3 Syntaxe de XML
 - Éléments du langage
 - Règles du langage
- 4 XML comme métalangage**
 - Comparaison avec HTML
 - Définition de document
- 5 Intérêt de XML

Comparaison avec HTML

En apparence, un fichier XML semble assez proche d'un fichier (X)HTML :

Comparaison avec HTML

En apparence, un fichier XML semble assez proche d'un fichier (X)HTML :

```
<html>
  <head>
    <title>Article de revue</title>
  </head>
  <body>
    <h1>Titre de l'article</h1>
    <p>Voici du texte qui peut être mis <i>en
      italique</i> ou même <b>en gras</b>.</p>
  </body>
</html>
```

Comparaison avec HTML

- XML, contrairement à HTML, ne définit pas d'ensemble de balises.

Comparaison avec HTML

- XML, contrairement à HTML, ne définit pas d'ensemble de balises.
- XML ne définit que des règles de syntaxe, pas le vocabulaire.

Comparaison avec HTML

- XML, contrairement à HTML, ne définit pas d'ensemble de balises.
- XML ne définit que des règles de syntaxe, pas le vocabulaire.
- XML est un *métalangage*, il permet de décrire et définir d'autres langages, dont (X)HTML fait partie.

Définition de document

Définir un langage ou un type de document à l'aide d'XML revient à définir un vocabulaire et des règles supplémentaires :

Définition de document

Définir un langage ou un type de document à l'aide d'XML revient à définir un vocabulaire et des règles supplémentaires :

- Ensemble d'éléments et d'attributs utilisables.

Définition de document

Définir un langage ou un type de document à l'aide d'XML revient à définir un vocabulaire et des règles supplémentaires :

- Ensemble d'éléments et d'attributs utilisables.
- Ensemble de valeurs prédéfinies pour certains de ces éléments.

Définition de document

Définir un langage ou un type de document à l'aide d'XML revient à définir un vocabulaire et des règles supplémentaires :

- Ensemble d'éléments et d'attributs utilisables.
- Ensemble de valeurs prédéfinies pour certains de ces éléments.
- Séquence dans laquelle ces éléments peuvent apparaître.

Définition de document

Définir un langage ou un type de document à l'aide d'XML revient à définir un vocabulaire et des règles supplémentaires :

- Ensemble d'éléments et d'attributs utilisables.
- Ensemble de valeurs prédéfinies pour certains de ces éléments.
- Séquence dans laquelle ces éléments peuvent apparaître.
- Autres règles : éléments obligatoires, répétabilité...

Définition de document

La définition de ce vocabulaire et de ces règles supplémentaires s'effectue à l'aide d'un *langage de définition de document* :

Définition de document

La définition de ce vocabulaire et de ces règles supplémentaires s'effectue à l'aide d'un *langage de définition de document* :

- DTD (*Document Type Definition*)
- Schema XML
- Relax NG
- etc.

Définition de document

L'association de XML et d'un langage de définition de document permet à XML de décrire un très grand nombre d'objets :

Définition de document

L'association de XML et d'un langage de définition de document permet à XML de décrire un très grand nombre d'objets :

- Livres (Docbook, ...)
- Documents bureautiques (OpenDocument, ...)
- Données statistiques (DDI)
- Formules mathématiques (MathML)
- Images (SVG)
- Articles de revue
- Etc., etc., etc.

Notion de validité

- Un document XML est dit *valide* s'il est bien formé et conforme à une définition de type de document qui lui est associée.

Notion de validité

- Un document XML est dit *valide* s'il est bien formé et conforme à une définition de type de document qui lui est associée.
- Un document XML *bien formé* est manipulable par un analyseur XML, indépendamment du type de données qu'il contient.

Notion de validité

- Un document XML est dit *valide* s'il est bien formé et conforme à une définition de type de document qui lui est associée.
- Un document XML *bien formé* est manipulable par un analyseur XML, indépendamment du type de données qu'il contient.
- Un document XML *valide* est manipulable dans un second temps par un logiciel dédié au type de données qu'il contient.

Plan

- 1 Historique
- 2 Structurer l'information
- 3 Syntaxe de XML
 - Éléments du langage
 - Règles du langage
- 4 XML comme métalangage
 - Comparaison avec HTML
 - Définition de document
- 5 Intérêt de XML

Lisibilité

- Les fichiers XML sont des fichiers textes.

Lisibilité

- Les fichiers XML sont des fichiers textes.
- Leur format est public.

Lisibilité

- Les fichiers XML sont des fichiers textes.
- Leur format est public.
- L'encodage des caractères est spécifié dans le fichier lui-même.

Lisibilité

- Les fichiers XML sont des fichiers textes.
- Leur format est public.
- L'encodage des caractères est spécifié dans le fichier lui-même.
- Lisibles par les ordinateurs et les êtres humains.

Lisibilité

- Les fichiers XML sont des fichiers textes.
- Leur format est public.
- L'encodage des caractères est spécifié dans le fichier lui-même.
- Lisibles par les ordinateurs et les êtres humains.

⇒ *Garantie d'une conservation à long terme*

Universalité

- Possibilité de décrire une grande variété de données.

Universalité

- Possibilité de décrire une grande variété de données.
- Possibilité de formaliser et de partager ces descriptions *via* des langages de description de document.

Universalité

- Possibilité de décrire une grande variété de données.
- Possibilité de formaliser et de partager ces descriptions *via* des langages de description de document.
- Format indépendant d'un système d'exploitation ou d'un logiciel particulier.

Universalité

- Possibilité de décrire une grande variété de données.
- Possibilité de formaliser et de partager ces descriptions *via* des langages de description de document.
- Format indépendant d'un système d'exploitation ou d'un logiciel particulier.

⇒ *Facilité d'échange en particulier via l'Internet*

Dissociation forme/fond

Le balisage XML est un balisage *sémantique* (contrairement au HTML).

Dissociation forme/fond

Le balisage XML est un balisage *sémantique* (contrairement au HTML).

Il existe un langage particulier dédié à la mise en forme des informations contenues dans un fichier XML : le langage XSL (*eXtensible Stylesheet Language*).

Dissociation forme/fond

Le balisage XML est un balisage *sémantique* (contrairement au HTML).

Il existe un langage particulier dédié à la mise en forme des informations contenues dans un fichier XML : le langage XSL (*eXtensible Stylesheet Language*).

Celui-ci permet la transformation du XML en HTML, PDF, ou tout autre format.

Dissociation forme/fond

Le balisage XML est un balisage *sémantique* (contrairement au HTML).

Il existe un langage particulier dédié à la mise en forme des informations contenues dans un fichier XML : le langage XSL (*eXtensible Stylesheet Language*).

Celui-ci permet la transformation du XML en HTML, PDF, ou tout autre format.

⇒ *Facilité de mise à jour et de publication de l'information*

Conclusion

Conclusion

- XML est un langage simple et universel permettant de décrire une grande variété de données.

Conclusion

- XML est un langage simple et universel permettant de décrire une grande variété de données.
- Ses spécifications ouvertes et l'utilisation de langages de définition de types de document facilitent l'échange d'informations et leur conservation à long terme.

Conclusion

- XML est un langage simple et universel permettant de décrire une grande variété de données.
- Ses spécifications ouvertes et l'utilisation de langages de définition de types de document facilitent l'échange d'informations et leur conservation à long terme.
- La dissociation opérée entre le contenu et sa mise en forme facilite la publication des informations et leur mise à jour.

Liens utiles

- Spécifications officielles du langage :
<http://www.w3.org/XML/>
- Ressources en anglais :
<http://www.xml.com/>
- Ressources en français :
<http://www.xmlfr.org/>
- Tutoriaux en anglais :
<http://www.w3schools.com/xml/>
- Articles sur XML de Wikipedia :
<http://fr.wikipedia.org/wiki/XML>
<http://en.wikipedia.org/wiki/XML>