



Première partie

## **Le nouveau standard de stockage et d'échange de données : XML**

9



## Introduction à XML

- Document XML simple

```
<?xml version="1.0" ?>
```

```
<mail>
```

```
  <from>titi@truc.org</from>
```

```
  <to>toto@bidule.com</to>
```

```
  <date year="2000" month="10" day="06" />
```

```
  <subject>Bonnes nouvelles</subject>
```

```
  <body>
```

```
    Salut les amis !
```

```
  </body>
```

```
</mail>
```

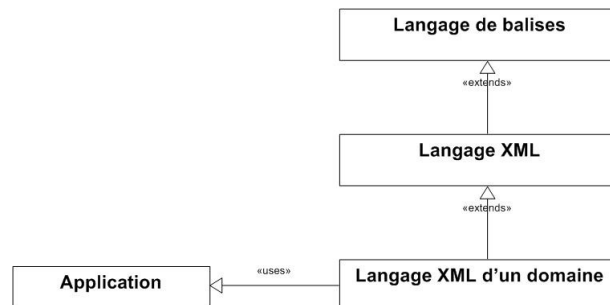
10



## Introduction à XML

### ■ Spécification XML

- XML est une spécification. Cette spécification définit un type générique de langage de balise que l'on appelle globalement « le langage XML »



11



## Introduction à XML

### ■ Spécification XML

En utilisant la métaphore de l'héritage de classes :

- Le « langage XML » est une classe de base abstraite
- Les divers langages XML utilisés pour créer des instances de documents sont des classes concrètes dérivées ;
- Les langages XML sont tous des langages de balises ;
- Les instances de documents sont tous des documents XML.

12



## Introduction à XML

---

- Balises XML
  - Dans l'exemple précédent `<mail>`, `<to>`, `</to>`, `<date>`, etc. sont des exemples de balises (*tags*).
  - Les balises sont des chaînes de caractères Unicode encadrées par les caractères « `<` » et « `>` ».
  - Il existe des balises de début et des balises de fin, comme `<from>` et `</from>`
  - Entre une balise de début et la balise de fin correspondante, on peut trouver du texte ou d'autres balises.

13



## Introduction à XML

---

- Balises XML
  - Pour schématiser, un fichier XML est composé de deux sortes de texte Unicode :
    - **Du texte de balisage.** Les balises sont simplement une technique d'écriture pour représenter la notion conceptuelle d'*élément*.
    - **Des données textuelles** séparées par des balises. Les données textuelles sont les contenus des éléments.
  - Lorsque des éléments sont imbriqués dans d'autres éléments, ils génèrent des structures arborescentes.

14



## Introduction à XML

---

- Balises XML
  - Les termes *year*, *month* et *day* sont les **attributs** de l'élément <date>.
  - Des valeurs placées entre "... " sont associées, le plus souvent explicitement, aux attributs.
  - La spécification XML ne définit pas les noms de balises et d'attributs.

15



## Introduction à XML

---

- Balises XML
  - Pour comprendre un document XML, il est nécessaire de disposer d'un répertoire des balises utilisées dans ce document, ainsi que d'une **grammaire** décrivant la structuration des balises entre elles.
  - Ces documents indispensables à la compréhension métier des fichiers XML sont les DTD (**Document Type Definition**) ou les **schémas XML**.
  - XML est souvent défini comme un « langage de balises sémantiques ».

16



## Utilisations du langage XML

- XML est largement diffusé pour décrire toutes sortes de structures arborescentes. Citons par exemple :
  - Les documents contenant des ensembles de données (bons de commande, factures, dossiers, manuels techniques, etc.) ;
  - Les documents décrivant des structures à installer de façon répétitive (descripteurs de déploiement des applications J2EE) ;
  - Les scripts de développement.

17



## Utilisations du langage XML

- Par ailleurs, XML est bien adapté pour
  - l'échange de données entre des logiciels, c'est-à-dire dans les domaines de l'intégration des applications (EAI – Enterprise Application Integration)
  - et dans le domaine de la communication inter-entreprises (B2B – Business to Business)
- Ces logiciels manipulent alors les documents commerciaux et techniques cités plus haut : bons de commande, etc.

18



## Utilisations du langage XML

- Les fichiers XML sont souvent des fichiers de grandes tailles contenant de très nombreuses informations reliées entre elles.
- Ils sont généralement générés automatiquement par des programmes appropriés.

19



## Principes du langage XML

- XML est compatible avec SGML, dont il est un sous-ensemble.
- Les documents XML sont destinés essentiellement à être traités par des programmes informatiques appelés ***processeurs XML***.
- Les processeurs XML doivent être faciles à écrire
- Les documents XML doivent être facilement lisibles par l'homme

20



## Principes du langage XML

- Un document XML doit obligatoirement être ***bien formé*** : obéir à la spécification XML
  - Chaque élément doit avoir une balise de début et la balise de fin correspondante
  - Les balises doivent être correctement imbriquées entre elles
  - Tout document XML doit avoir un et un seul ***élément racine*** : un élément qui contient tous les autres
  - Les valeurs associées aux attributs doivent être encadrées par les délimiteurs « " » ou « ' »

21



## Principes du langage XML

- Les **processeurs XML** sont des programmes permettant de vérifier que des documents XML sont bien formés.
- Ils traitent le contenu des éléments de manière déterministe. On parle aussi d'analyseurs syntaxiques XML (**XML parsers**)

22



## Principes du langage XML

---

- Le fonctionnement d'un processeur XML peut être décrit de la manière suivante :
  - À la rencontre d'une balise de début, entrer dans l'état de traitement de cette balise
  - À la rencontre de la balise de fin correspondante, sortir de cet état et revenir à l'état précédent
  - S'il n'y a pas d'état précédent, alors on vient de quitter l'état de traitement de l'élément racine : par conséquent, le document est bien formé.
  - Si une balise de fin ne correspond pas à l'état en cours, le document est mal formé.

23



### Deuxième partie

## **Documents XML : Structure Logique et Structure Physique**

24