



XML

Extensible Markup Language

Module SOA
A.U 2019-2020



Objectifs



- Savoir le rôle de XML.
- Apprendre la structure et les règles syntaxiques d'un document XML.
- Comprendre la notion des espaces de nom.



Plan



- Introduction
- Structure de données
- Présentation de XML
- Structure d'un document XML
- Espace de noms XML



Introduction 1/2




Les pages jaunes



Introduction 2/2




- On doit organiser d'une certaine manière les données ce qui permet un traitement automatique de ces dernières plus efficace et rapide.
 utilisation d'une structure de données.



Structure de données 1/3



- Structure de données:
 - Une organisation des informations.
 - est destinée à contenir des données, afin de leur donner une organisation permettant de simplifier leur traitement.
-  baisser de manière significative la complexité d'une application informatique et diminuer le taux d'erreurs.



Structure de données 2/3

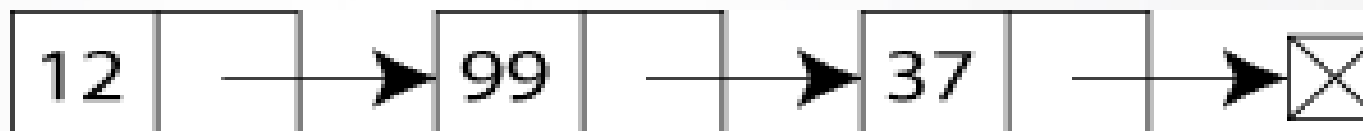


- Différentes structures de données existent:

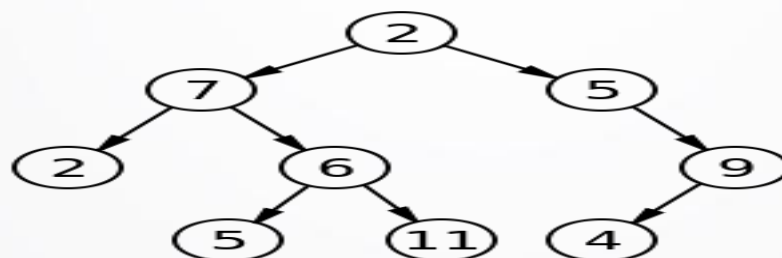
- Tableau:

Valeur	45	154	58	78	31	5	74
Index	0	1	2	3	4	5	6

- Liste chaînée:



- Arbre:





Structure de données 3/3



- Les documents structurés sont des documents qui contiennent de l'information à propos de leurs **structures logiques** et **physiques**:
 - **Structure physique** : apparence visuelle (texte sur deux colonnes, texte justifié ou non, etc.)
 - **Structure logique** : organisation du contenu intellectuel du document (chapitre, section, sous-section, etc.)

Langages de description de document structuré 1/3

- Les langages les plus couramment utilisés permettant d'encoder un document structuré à l'aide des balises sont:





Langages de description de document structuré 2/3



- SGML est un langage servant à préciser la structure d'un document quelconque. Il est compréhensible mais il était inadapté à l'écriture de documents pour internet. Il a donc été nécessaire d'en dériver le langage HTML.^[1]
- HTML est le standard du développement web mais il n'est ainsi pas possible de définir autre chose qu'une page Web. ^[1]
- Le XML est un dérivé du SGML. Il tente à être plus souple que HTML et plus simple que SGML.



Langages de description de document structuré 3/3

- Diffusion électronique de documents

⊖ Syntaxe complexe

SGML

- Présentation des documents sur le web

1986

⊖ Non flexible, figé

HTML

1996

- Structuration, échange des documents

+ Plus simple que SGML

+ Plus souple que HTML

XML



Présentation de XML



- eXtensible Markup Language.
- Langage de balises.
- Recommandation de W3C.
- XML décrit, structure, échange des données.
- Archiver des données.



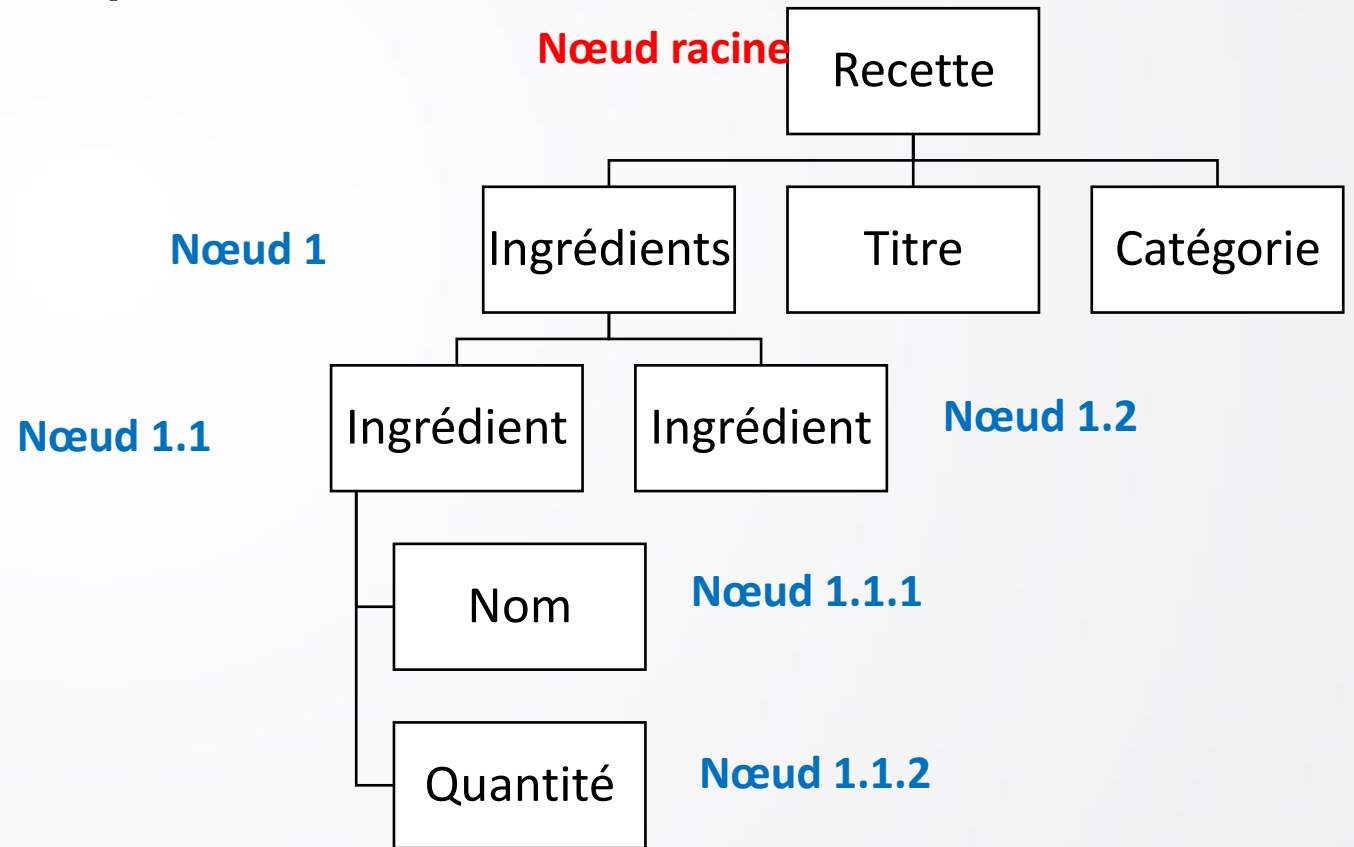
Pourquoi utiliser XML?



- Lisible : texte balisé avec marquage.
- Extensible : supporte les évolutions applicatives.
- Mise en forme avec des feuilles de style.
- Un méta langage permettant la définition de langages adaptés à des besoins variés.
- Supporté par les grands constructeurs: IBM, Microsoft .net, SUN, etc.

Arborescence XML

- Un document XML est composé de plusieurs nœuds.
- L'arborescence d'un document XML est la structure hiérarchique des nœuds.





Structure d'un document XML 1/8



- Un document XML comporte :
 - une prologue.
 - l'arbre des éléments.
 - éventuellement des commentaires



Structure d'un document XML 2/8



•La prologue

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

La prologue XML:

- est une instruction de traitement destinée à l'application chargée du traitement du document XML
- est facultative, mais fortement conseillée
- décrit:
 - la version du langage XML → version="1.0"
 - le codage des caractères (par défaut UTF-8) → encoding="UTF-8"
 - La dépendance à des document extérieurs → standalone="yes"



Structure d'un document XML 3/8



- Les nœuds XML

Il existe trois types de nœuds XML:

1 Les éléments

2 Les attributs

3 Les entités



Structure d'un document XML 4/8



1

Les éléments

- Un élément s'ouvre et se ferme par une balise
- Le nom de l'élément est repris dans la balise ouvrante et dans la balise fermante.

<categorie>Dessert**</categorie>**



Balise ouvrante



Balise fermante



Structure d'un document XML 5/8



2

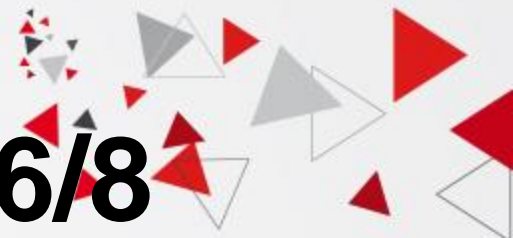
Les attributs

- L'attribut se trouve dans la balise ouvrante d'un élément
- L'attribut n'est pas repris dans la balise fermante
- Un élément peut contenir plusieurs attributs
- Un même attribut ne peut pas être présent qu'une seule fois dans un élément
- L'ordre des attributs n'a pas d'importance au sein d'un élément
- La valeur de l'attribut est indiquée entre guillemets

```
<quantite unite ="g" >100</quantite>
```



Structure d'un document XML 6/8



3

Les entités

- Certains caractères ont un sens particulier en XML (caractères spéciaux)

Exemple: >, &, "

- Les entités ont été prédéfinies afin de pouvoir utiliser les caractères réservés
- Une entité est une chaîne de caractère commençant par & et se terminant par ;&entite;
- Une entité est remplacée par la chaîne de caractère qu'elle représente.



Structure d'un document XML 7/8



Les entités prédéfinies

Caractère	Entité
&	&
<	<
>	>
"	"
'	'

Exemple:

```
<message>salaire &lt; 1000</message>
```

Structure d'un document XML 8/8

- Les commentaires

```
<!-- This is a comment -->
```

Les commentaires sont ignorés lors de l'interprétation du document XML.

Les règles syntaxiques

- Un document XML a un **seul** élément **racine**.
- Un élément peut:

- Être vide

```
<vide/>
```

- Contenir une chaîne de caractères

```
<categorie>Dessert</categorie>
```

- Contenir des éléments fils (qui doivent être correctement imbriqués)

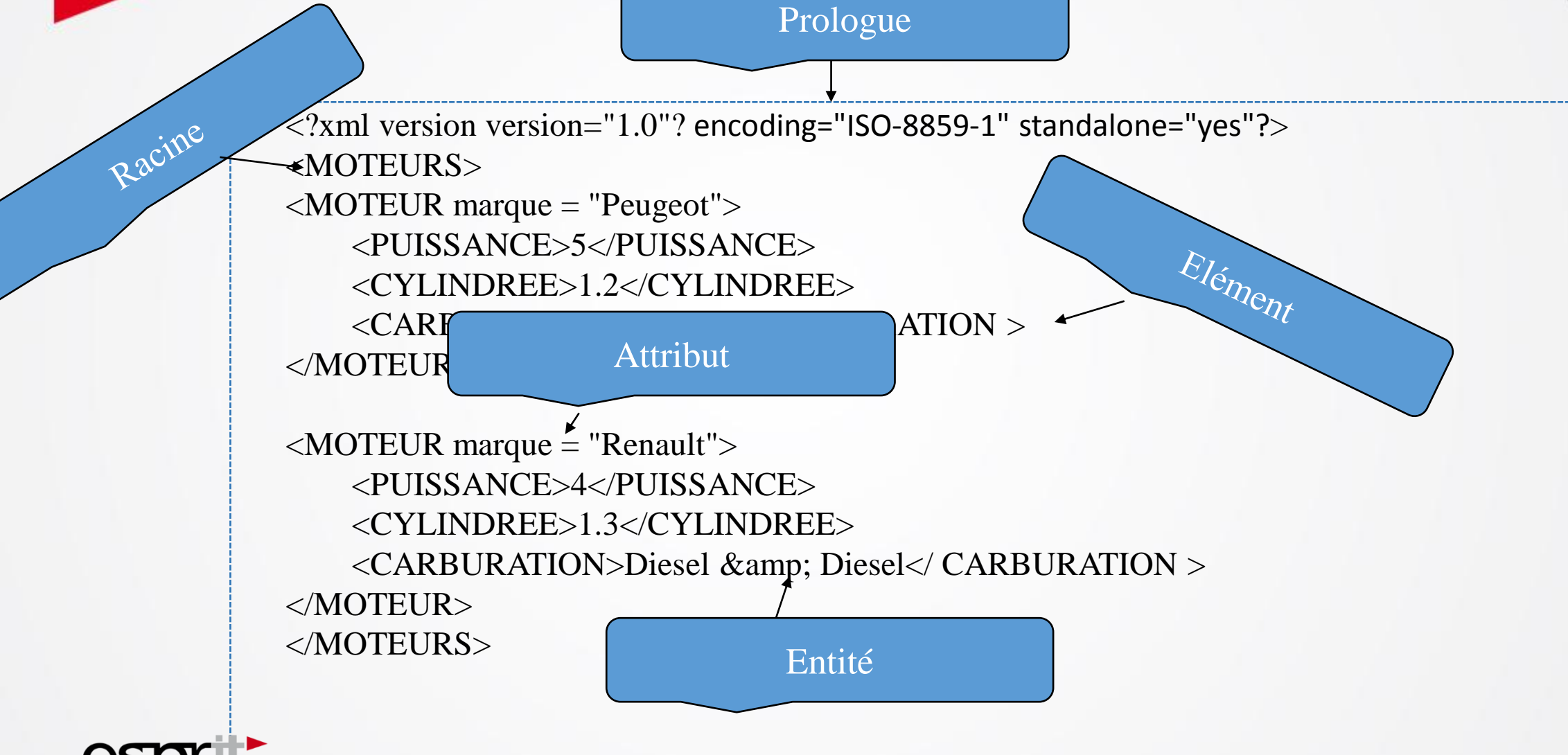
```
<ingredient>  
  <nom>beurre</nom>  
  <quantite>100</quantite>  
</ingredient>
```

- XML est sensible à la casse

<Categorie>incorrect**</categorie>**

XML

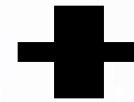
Exemple de document XML



Document XML bien formé

XML

```
<produit quantite= "80">  
  <id> 123</id>  
</produit>
```



DTD/XSD



Document XML **bien formé**

Respect de la syntaxe XML



Document XML **valide**

Respect une grammaire

- Un document XML est valide si et seulement s'il est bien formé



Grammaire



1 DTD

- Une DTD (Document Type Definition) est une grammaire qui permet de définir une structure type de document XML.

2 XSD

- XML Schema est un langage de description de format de document XML permettant de définir la structure et le type de contenu d'un document XML.[\[2\]](#)
- Cette définition permet notamment de vérifier la validité de ce document.



En résumé



- XML est un langage de structuration de données
- Un document XML est structuré à l'aide d'éléments et d'attributs
- Un document XML doit respecter les règles syntaxiques pour qu'il soit bien formé



Références



- [1] <http://www.gchagnon.fr/cours/xml/base.html>
- [2] http://fr.wikipedia.org/wiki/XML_Schem