



# Technologies XML

LI – Semestre 4  
Département informatique  
UFR des Sciences et technologies  
Université de Thiès

1

1



## Plan du cours

1. Présentation du cours
2. Introduction
1. XML et ses applications

UFR SET – UT / Dep. Info

Pr Mouhamadou THIAM – mthiam@univ-thies.sn

2

2

2




## Plan du cours

1. Présentation du cours
2. Introduction
1. XML et ses applications

UFR SET – UT / Dep. Info      Pr Mouhamadou THIAM – mthiam@univ-thies.sn      3  
3

3



## Présentation du cours (1)

**Unité d'Enseignement**  
Titre : INFORMATIQUE  
Sigle : INF 243

**Élément constitutif**  
Titre : Technologies XML  
Sigle : INF 2431

**Autres éléments constitutifs de l'UE**  
Titre : Développement Web 2  
Sigle : INF 2432

UFR SET – UT / Dep. Info      Pr Mouhamadou THIAM – mthiam@univ-thies.sn      4  
4

4



## Présentation du cours (2)

### Horaires

- CM : 20H
- TD/TP : 20H
- TPE : 40H
- Coefficient de l'UE : 3
- Crédits de l'UE : 8

### Evaluation

- Contrôle des connaissances : 40%
- Examen écrit : 60%

UFR SET – UT / Dep. Info

Pr Mouhamadou THIAM – [mthiam@univ-thies.sn](mailto:mthiam@univ-thies.sn)

5  
5

5



## Présentation du cours (3)

### Responsable du cours

Pr Mouhamadou THIAM  
Maître de conférences en Informatique  
Intelligence Artificielle : Sémantique Web  
Email : [mthiam@univ-thies.sn](mailto:mthiam@univ-thies.sn)

### Travaux dirigés et pratiques

Dr. Moussa Déthié SARR  
Email : [mdsarr@univ-thies.sn](mailto:mdsarr@univ-thies.sn)

UFR SET – UT / Dep. Info

Pr Mouhamadou THIAM – [mthiam@univ-thies.sn](mailto:mthiam@univ-thies.sn)

6  
6

6

## Plan du cours



1. Présentation du cours
2. Introduction
  1. XML et ses applications

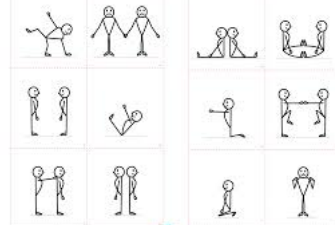
UFR SET – UT / Dep. Info

Pr Mouhamadou THIAM – mthiam@univ-thies.sn

7  
7

7

## XML, un Langage ...



UFR SET – UT / Dep. Info

Pr Mouhamadou THIAM – mthiam@univ-thies.sn

8  
8

8



## Langage ...

- Langage humain
- Langage animal
- Langage formel
- Langage naturel
- Langage de programmation
- Langage informatique
- Langage de script
- Langage SMS
- Langage des fleurs
- Langage interprété informatique

UFR SET – UT / Dep. Info

Pr Mouhamadou THIAM – mthiam@univ-thies.sn

9  
9

9



## Web...



UFR SET – UT / Dep. Info

Pr Mouhamadou THIAM – mthiam@univ-thies.sn

10  
10

10

## Web [wikipedia]...



- **World Wide Web (WWW)**, littéralement la « **toile (d'araignée) mondiale** », communément appelé le **Web**, et parfois la **Toile**, est un système **hypertexte** public fonctionnant sur **Internet**.
- Le Web permet de consulter, avec un navigateur, des pages accessibles sur des sites. L'image de la toile d'araignée vient des hyperliens qui lient les pages web entre elles

## Web...



- Applications d'Internet
- Distinct d'autres applications comme
  - courrier électronique,
  - messagerie instantanée,
  - partage de fichiers en pair à pair.
- Inventé par **Tim Berners-Lee** et **Robert Cailliau** au début des années 90

## La révolution du Web...



- Depuis vingt cinq ans, Internet révolutionne l'informatique « grand public »
- **HTML** est le langage du Web...
  - Même si on trouve aussi du .doc, .ps, .pdf... et des images (jpg, gif), du son, de la vidéo...

## La révolution du Web...



- Des milliards de pages existent actuellement
  - public/privé,
  - statique/dynamique,
  - visible/caché
- Support naturel pour **l'information distribuée**
  - À destination d'êtres humains et
  - de plus en plus, pour des **applications**

## La révolution du Web...



- Exemples:
  - B2C (commerce électronique)
  - B2B (achats groupés)
  - Bibliothèques et fonds documentaires en ligne
- HTML n'est pas adapté pour ces applications
  - Ces applications (programmes) ont besoin de typage pour représenter la structure des données

## Prérequis ...



- Les bases de données (relationnelle, ...)
  - Conception de schémas : relationnel, UML, ...
  - Les langages de requêtes : SQL, OQL, ...
- Le langage HTML : tables, listes, formulaires, ...
- Le langage Java
  - Héritage
  - Utilisation d'API





## Problématique ...

- Besoins d'accès contenu du web de manière précise
- Recherche mot-clef utilisée : pas satisfaisant;
- Méthodes ne garantissent pas
  - Pertinence de la réponse,
  - Réponse : document lui-même
  - Nombre exorbitant de réponses
  - Majeure partie répondent pas à la question posée
- Le langage
  - HTML → **human**
  - XML → **machine readable**

UFR SET – UT / Dep. Info

Pr Mouhamadou THIAM – mthiam@univ-thies.sn

17  
17

17



## Résumé historique du web

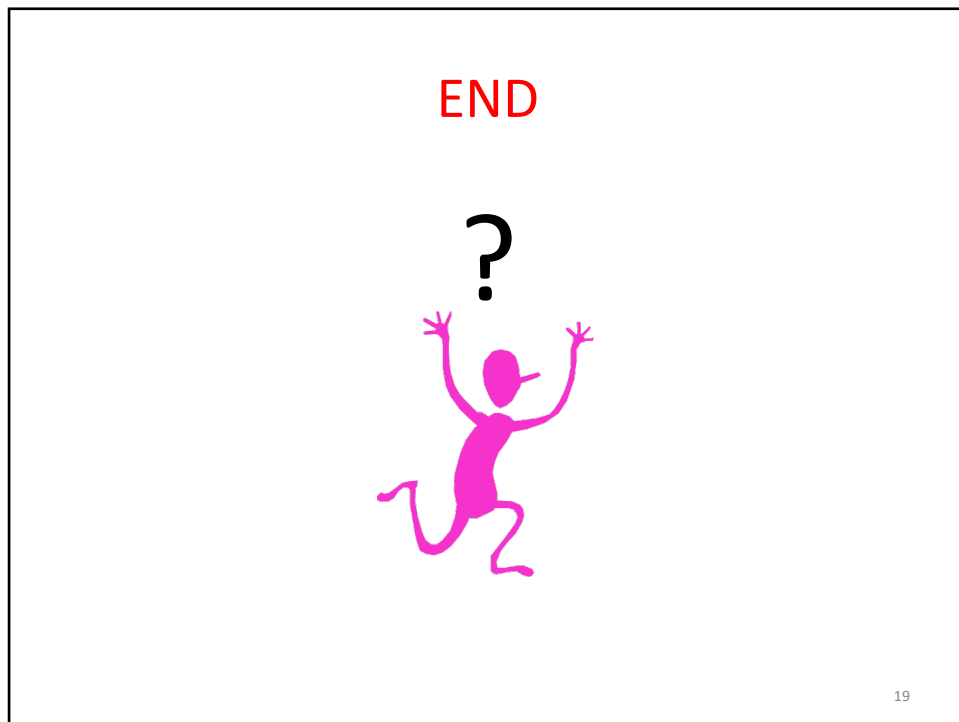
- **1989** : **Tim Berners-Lee** (CERN, Genève) commence le développement d'un **système hypertexte**.
- **1990** : Premières définitions pour **HTTP, HTML, URL**.
- **1992** : Premier **annuaire de sites Web. 26 sites**.
- **1994** : **Netscape Navigator 1.0** ; Fondation du **W3C**
- **1998** : Plus de **2 millions de sites** ; Création de **Google**.
- **2000** : **XHTML 1.0**.
- **2004** : **Firefox 1.0**.
- **2005** : Plus de 60 millions de sites !!!!

UFR SET – UT / Dep. Info

Pr Mouhamadou THIAM – mthiam@univ-thies.sn

18  
18

18



19

## Plan du cours

1. Présentation du cours
2. Introduction
1. XML et ses applications

UFR SET – UT / Dep. Info

Pr Mouhamadou THIAM – mthiam@univ-thies.sn

20  
20

20

## Lectures

- Obligatoires
  - Recommandations du W3C - [http://www.w3.org/TR/tr-technology-stds#w3c\\_all](http://www.w3.org/TR/tr-technology-stds#w3c_all)
    - XML, XML Schema
    - XPath, XQuery
    - XSLT
  - <http://www.w3.org/History/1989/proposal.html>
  - *The Semantic Web : A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities* by TIM BERNERS-LEE, JAMES HENDLER and ORA LASSILA
- Facultatifs
  - Recommandations HTML du W3C

21

21



## XML et ses applications : plan

1. Introduction à XML
2. Schémas de documents XML
  - a. DTD
  - b. XML Schema
3. Le modèle DOM
4. Interrogation : XPath
5. Transformation et présentation : XSLT
6. Programmation : XML et Java (API DOM, SAX, ...)

UFR SET – UT / Dep. Info

Pr Mouhamadou THIAM – mthiam@univ-thies.sn

22

22

22

# Introduction à l'eXtensible Markup Language –XML

23

23

## De SGML à XML en passant par HTML



- SGML[1969, ISO 8879 1986]
  - Inventé par *Charles Goldfarb*, juriste chez IBM.
  - Du balisage typographique au balisage logique.
- HTML[1989-1992, HTML5 depuis 2007]
  - Inventé par *Tim Berners-Lee* comme langage de description des pages du Web.
  - Concept d'ancrage : hypertexte.
- XML[1996, Rec. W3C 1998, Dan Cannolly]
  - Une synthèse de SGML et de HTML.

UFR SET – UT / Dep. Info

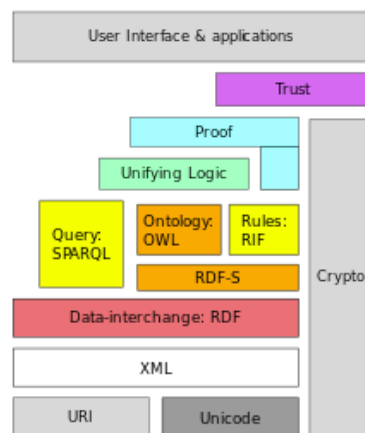
Pr Mouhamadou THIAM – mthiam@univ-thies.sn

24

24

24

## Modèle en couches de Tim Berners-Lee



UFR SET – UT / Dep. Info

Pr Mouhamadou THIAM – mthiam@univ-thies.sn

25  
25

25

## XML : Document semi-structuré



- Un Document Semi-Structuré peut se définir par la définition de ses principaux composants (Poulet et al, 1997): **contenu** et **structure**
  - **Contenu** : le contenu d'un document semi-structuré désigne le contenu textuel ou multimédia par exemple des paragraphes, des figures ou des images
  - **Structure**
    - **Logique** : elle définit l'organisation hiérarchique des données du document
    - **Physique** : elle définit la présentation du document sur le support

UFR SET – UT / Dep. Info

Pr Mouhamadou THIAM – mthiam@univ-thies.sn

26  
26

26



## XML : le modèle de données du Web

- HTML versus XML
  - Attention : HTML et XML ne sont pas concurrents !!!
    - **Présentation des données** pour HTML
    - **Représentation des données** pour XML : les données et leur visuel est indépendant
  - Pour s'en convaincre, la présentation d'un document XML nécessite sa transformation en HTML via XSLT
    - XSLT= eXtensible Stylesheet Language Transform

UFR SET – UT / Dep. Info

Pr Mouhamadou THIAM – mthiam@univ-thies.sn

27  
27

27



## XML : le modèle de données du Web

- Exemple
  - Le fichier [HTML](#) contenant présentation et données
  - Le fichier [XML](#) ne contenant que les données
  - Le fichier [XSL](#) transformant le fichier XML en HTML

UFR SET – UT / Dep. Info

Pr Mouhamadou THIAM – mthiam@univ-thies.sn

28  
28

28

## Exemple de document (text / HTML)

Construire une application XML

Domaines d'application d'XML  
Vue d'ensemble

XML : battage ou réalité ?

Un document

En HTML

```

<!--extrait d'un document HTML-->
<HTML>
<HEAD>
  <TITLE> Construire une application XML </TITLE>
</HEAD>
<BODY>
  <center>
    <P>Domaines d'application d'XML <br>
    Vue d'ensemble</P>

    <P>XML : battage ou réalité ? </P></center>
  </BODY>
</HTML>

```

UFR SET – UT / Dep. Info
Pr Mouhamadou THIAM – mthiam@univ-thies.sn
29  
29

29

## Exemple de document (HTML / XML)

```

<!--extrait d'un document HTML-->
<HTML>
<HEAD>
  <TITLE> Construire une application XML </TITLE>
</HEAD>
<BODY>
  <center>
    <P>Domaines d'application d'XML <br>
    Vue d'ensemble</P>

    <P>XML : battage ou réalité ? </P></center>
  </BODY>
</HTML>

```

En HTML

En XML

Presentation  
Contenu  
Structure


```

<!--extrait d'un document XML-->
<?xml version="1.0" ?>
<livre>
  <couverture>
    <titre> Construire une application XML </titre>
    <auteur> Jean-Christophe Bernadac </auteur>
    <auteur> François Knab </auteur>
  </couverture>
  <chapitre n="1">
    <p> XML : battage ou réalité ? </p>
  </chapitre>
</livre>

```

UFR SET – UT / Dep. Info
Pr Mouhamadou THIAM – mthiam@univ-thies.sn
30  
30

30



## HTML versus XML

Document HTML

Présentation

Contenu

Structure

Document XML


Présentation

Contenu

Structure

UFR SET – UT / Dep. Info
Pr Mouhamadou THIAM – mthiam@univ-thies.sn
31

31



## Exemple de document

Demande d'information

Moussa Dethie Sarr (univ-thies.sn)

---

Demande d'information

---

Bonjour Moussa,

Peux tu m'expliquer ce qu'est le langage XML ?  
 Il me faudrait en effet ton avis éclairé sur le sujet.  
 J'attends ta réponse, à bientôt

-----

Pr Mouhamadou Thiam  
 Maître de conférences en informatique  
 Département Informatique  
 Vice-Directeur UFR SET  
 Université de Thiès  
 Email : [mthiam@univ-thies.sn](mailto:mthiam@univ-thies.sn)  
 Tél. Mob : (+221) 77.573.94.39

UFR SET – UT / Dep. Info
Pr Mouhamadou THIAM – mthiam@univ-thies.sn
32

32





## Le mail “demande d’infos” en XML

```
<?xml version='1.0' encoding='iso-8859-1'?>
<mail>
  <de>mthiam@univ-thies.sn</de>
  <a>mdsarr@univ-thies.sn</a>
  <objet classe="urgent">Demande d'information</objet>
  <date>11-01-2021</date>
  <corps>
    <paragraphe>Bonjour Moussa,</paragraphe>
    <paragraphe>Peux tu m'expliquer ce qu'est le langage XML ?
  </paragraphe>
    <paragraphe>Il me faudrait en effet ton avis éclairé sur le sujet.
  </paragraphe>
    <paragraphe>J'attends ta réponse, à bientôt</paragraphe>
  </corps>
</mail>
```

33



## En résumé

- Un **document** XML est découpé en **éléments** structurés hiérarchiquement.
- Un document a un élément racine appelé **élément du document**.
- Un élément est composé :
  - d'un **nom** qui spécifie son type,
  - d'**attributs**,
  - d'un **contenu** formé d'**éléments** ou de **textes**.

34



## En résumé

- Un **texte** est une chaîne de caractères.
- Un **attribut** a un **nom** et une **valeur** qui est une chaîne de caractères.
- Syntaxiquement, les éléments d'un document XML sont marqués dans le document lui-même par des paires de balises ouvrantes et fermantes.



## Structure d'un document XML

- Un document XML est composé :
  - d'un **prologue facultatif** (voir ci-après « Organisation d'un document XML »),
  - de **l'élément du document** qui est lui-même composé **d'éléments et de textes**

## Structure d'un document XML



- Dans le prologue et dans le contenu d'un élément, on peut insérer :
  - des **commentaires**,
  - des **instructions de traitement** destinés aux applications traitant le document.
- Un document peut être découpé en **entités** enregistrées dans un ou plusieurs fichiers.

## Caractères de nom



- Un **caractère de nom** est soit
  - une lettre,
  - un chiffre,
  - un point,
  - un tiret,
  - un espace souligné,
  - un deux-points.
- {[A..Z], [0..9], . , -, \_ , : }



## Nom et tokens de nom

- Un **nom** est une suite de un ou plusieurs caractères dont : le premier est soit
  - une lettre,
  - un espace souligné,
  - un deux-points {\_, :}
- Exemple de caractère de nom.  
xml:lan, **extrait\_de titre**, poeme-79
- Un **token de nom** est une suite de un ou plus caractères de nom.



## Un élément XML

- Un élément est composé :
  - d'une **balise de début** qui contient le nom de l'élément et éventuellement ses attributs **<elt attr="valAttr">**,
  - d'un **contenu**,
  - d'une **balise de fin** **</elt>**.



## Un élément XML

- Par exemple :  
`<objet classe="urgent">`  
 Demande d'information  
`</objet>`
  - balise de début : `<objet classe="urgent">`
  - nom : `objet`
  - attribut : `classe="urgent"`
  - contenu : `Demande d'information.`
  - balise de fin : `</objet>`



## Contenu d'un élément XML

- Vide :  
`<elt></elt>` ou `<elt/>`

Exemple :

`<renvoi cible="I15.1"></renvoi>`

ou

`<renvoi cible="I15.1"/>`

`<fruit type="bio"/> alternative <fruit> biologique </fruit>`

## Contenu d'un élément XML



- **Composé** d'éléments :

```
<intro>
  <para>Vallon situé à l'est du refuge ...</para>
  <para>Le vallon le plus utilisé pour la traversée ...</para>
</intro>
```

## Contenu d'un élément XML



- **Mixte** : mélange de textes et d'éléments

```
<nom>Col de Névache</nom>
<para>Du col de Névache <renvoi cible="I15.1"/>,
  suivre la ligne de crête qui mène à la pointe de Névache.
  <note type="prudence">Attention : corniches possibles. </note>
  <note type="matériel">Crampons utiles au printemps.</note>
  A la fin sortir par la gauche pour éviter les autres visiteurs
</para>
```



## Elément mixte

- Le contenu d'un élément mixte est constitué d'une chaîne de caractères dans laquelle peuvent être insérés des éléments. Cette insertion découpe ce contenu en deux types de constituants :
  - les plus longues suites d'au moins un caractère dans lesquelles ne sont pas insérés des éléments : nous les appellerons **textes**,
  - les éléments.



## Elément mixte

- Par exemple, le contenu :  
Du col de Névache <renvoi cible="I15.1"/>, suivre la ligne de crête qui mène à la pointe de Névache.
- comprend dans l'ordre :
  - le **texte** : Du col de Névache
  - l'**élément** : <renvoi cible="I15.1"/>
  - le **texte** : , suivre la ligne de crête qui mène à la pointe de Névache.



## Section CDATA

- Lorsqu'un texte contient des caractères qui jouent un rôle de **délimiteur** dans la syntaxe XML, il est nécessaire de pouvoir **inhiber** ce rôle. Ceci peut être fait en insérant le texte contenant ces délimiteurs dans une section **CDATA** sous la forme suivante :

**<![CDATA[*texte contenant des délimiteurs*]]>**



## Section CDATA

- Le texte inséré peut contenir n'importe quels caractères excepté la chaîne **]]**. Une section CDATA ne peut donc pas en contenir une autre.
- Par exemple, la phrase :  
« L'expression **<ALT>2794</ALT>** est un élément XML. »
- peut être représentée par l'élément suivant :  
**<phrase>**L'expression  
    **<![CDATA[<ALT>2794</ALT>]]>** est un élément XML.  
**</phrase>**





## Attributs

- Un **attribut** est un couple **nom-valeur** où :
  - le nom est un nom XML,
  - la valeur est une suite de caractères.
- Par exemple :  
`type="prudence"`
- Si une valeur d'attribut est placée entre guillemets, elle peut contenir des apostrophes et si elle est placée entre apostrophes, elle peut contenir des guillemets.
- Par exemple :  
`select="itinéraire[cotation=*****]"`  
`select='itinéraire[cotation="*****"]'`



## Exemple d'itinéraire

```
<itinéraire id="I15.1">
  <nom>Col de Névache</nom><alt>2794</alt><cotation>**</cotation>
  <num>1</num>
  <para>
    S'élever au-dessus du refuge des Drayères en direction est. Suivre la
    rive droite du torrent de Brune puis s'engager sur le flanc droit du ravin
    des Muandes que l'on quitte vers 2500 m pour rejoindre le col situé au
    nord. Descente possible sur Valmeinier.
    <note type="prudence">Départ assez raide.</note>
  </para>
</itinéraire>
```



## Exemple d'itinéraire

```
<itinéraire id="I15.2">
  <nom>Pointe de Névache</nom><alt>2892</alt><cotation>***</cotation>
  <num>2</num>
  <para>
    Du col de Névache <renvoi cible="I15.1"/>, suivre la ligne de crête qui
    mène à la pointe de Névache.
    <note type="prudence">Attention : corniches possibles.</note>
    <note type="materiel">Crampons utiles au printemps.</note>
  </para>
</itinéraire>
```

UFR SET – UT / Dep. Info

Pr Mouhamadou THIAM – mthiam@univ-thies.sn

51  
51

51



## Liens internes

- Tout élément peut avoir un attribut ayant pour valeur un token de nom qui l'identifie dans le document :
  - les tokens de nom I15.1 et I15.2 identifient les itinéraires n° 1 et n° 2 du Vallon des Muandes dans le document « Itinéraires skieurs ».
- L'identificateur d'un élément permet d'y faire référence depuis d'autres éléments :
  - <itinéraire id="I15.1"> ....</itinéraire>
  - l'élément <renvoi cible="I15.1"> renvoie à l'itinéraire n° 1 de ce même vallon).

UFR SET – UT / Dep. Info

Pr Mouhamadou THIAM – mthiam@univ-thies.sn

52  
52

52



## Commentaires

- Un commentaire est une phrase ayant la forme suivante :  
`<!--texte du commentaire-->`
- Un commentaire peut contenir n'importe quel caractère **excepté --**.  
Un commentaire ne peut donc pas inclure un autre commentaire.
- Un commentaire peut être inclus dans le contenu d'un élément mais pas à l'intérieur d'une balise.
- Exemple :  
`<!-- Les itinéraires sont classes par vallon -->`



## Résumé

- **langage de balises** (simple à implanter, mais extensible) : chacun définit ses propres balises
- **séparation** de la **présentation** et du **contenu**, conçu pour
  - **décrire données** en se concentrant sur leur **structure**
  - **assurer l'interopérabilité**
- Utilisation d'un **DTD** ou un **XML Schéma** : modèle de données
- XML et HTML complémentaires
  - **XML** ne remplace pas **HTML**
  - **XSL / XSLT** transforme **XML** en **HTML**