

Faculté des Mathématiques et Informatique,

Département informatique, 3<sup>ième</sup> Année LMD IIM

Matière : Données semi-structurées

## Partie 1 : Contexte et problématiques

- L'échange de données entre SI
- Les BD, avantages et limites
- Applications XML

Le but du TP est : d'apprendre à modéliser différentes informations en XML (syntaxe de base). Pour cela, le TP est structuré en plusieurs parties. Dans un premier temps, vous allez observer les limites d'utiliser HTML pour l'échange de données entre les SI (applications WEB, ...), ainsi que la différence entre XML et les BD, puis vous allez créer vos propres fichiers XML.

Exercice N° 01 HTML pour le transfert des données**1. Analyse du document HTML**

- Visualiser ce code HTML sur un Navigateur (un exemple des sites dites : *base documentaires, comme Wikipedia, EMS,...etc*).

- Ouvrir un éditeur WEB, ou le bloc note même,

- Copier \ Coller le code, enregistrer le code : test.html.

- Ouvrir le document web (html).

- Inspecter le code source.

- Toujours, dans le navigateur, cliquer sur inspecter,

- Quelle est la partie **Données**, et la partie **Présentation** dans ce code ?  
C'est quoi le rôle de CSS ?

- Les balises `<p>`, `<title>`, sert à structurer le texte

- Le CSS pour la mise en forme (y a pas de CSS dans ce code),

- Ce document est-il structuré ?

- Oui, il est structuré, le HTML structure ce document, MAIS : les balises, HTML,

**n'offre aucun sens, pour connaître le sémantique du contenu. Par exemple :** `<p><strong> Beurre de cacahuete </strong>`  
`</p>`, ne donne aucune indication, qu'il s'agit un produit alimentaire.

```
<!doctype html>
<html>
<head>
<meta charset="utf-8">
<title>Aliments pour sportifs</title>
</head>
<body bgcolor="#D1F7F7">
<h1 align="center">Les Aliments pour le <strong>RUNNING</strong></h1>
<p align="center">
  <header>
    <nav>
      <a href="#presentation">Présentation des aliments pour le sport</a>
      <a href="#produits">Beurre de cacahuète</a>
      <a href="produits2.html">Le riz</a>
      <a href="affiche.php">Les Livres</a>
      <a href="contacts.html">Contact</a>
    </nav>
  </header>
  <section id="presentation">

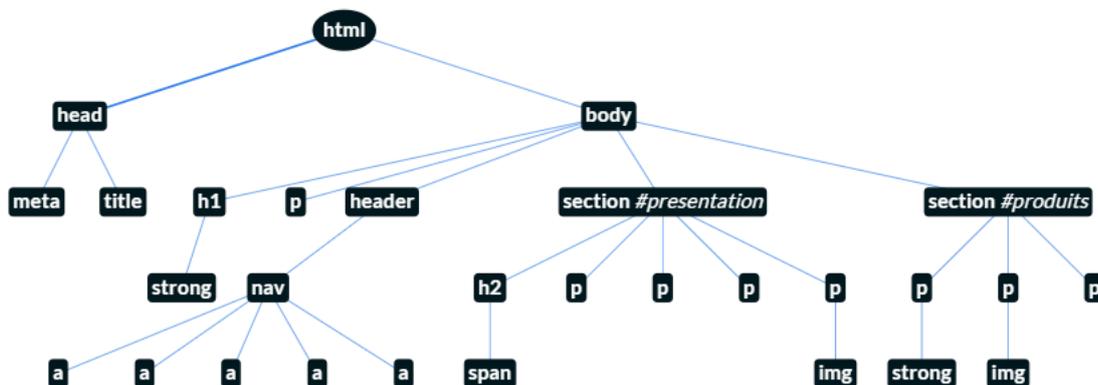
    <h2><span>Les produits alimentaires pour le sport de RUNNING</span></h2>
    <p>En course à pied, l'&rsquo;alimentation a un rôle fondamental pour vous permettre de bien vous entraîner, de récupérer, de progresser, et d'&rsquo;arriver en forme pour votre objectif, sans vous blesser.</p>

    <p>Que cela soit pour le 10 km, le semi-marathon, le marathon, ou la longue distance, vous devez connaître le programme nutrition conseillé chez le coureur à pied.</p>

    <p>Dans cet site nous allons vous guider à retrouver le régime alimentaire pour la course à pied, l'&rsquo;alimentation avant, pendant, et après la course, ainsi que l'&rsquo;hydratation et les compléments alimentaires.</p>

    <p>&nbsp;</p>
  </section>
  <section id="produits">
    <p><strong> Beurre de cacahuete </strong> </p>
    <p>&nbsp;</p>
    <p>Le beurre de cacahuète est l'un des meilleurs aliments pour les sportifs. Le beurre de cacahuète est pour les bodybuilders une source naturelle de calories et de protéines à un prix abordable qui t'apportera en plus de l'acide folique (vitamine B9), de la vitamine E, du magnésium et du resvératrol. Une petite quantité de zinc est également présente, rappelons que le zinc aide au bon fonctionnement de ton système immunitaire et participe à la production de testostérone.</p>
  </section>
</body>
</html>
```

- **DOM (Document Object Model)** : Un document HTML se compose d'un ensemble de nœuds. Ces nœuds sont structurés sous forme hiérarchique. Cette structure repose sur le modèle DOM. Pour pouvoir manipuler **dynamiquement** le contenu d'une page HTML, cette structure hiérarchique peut être utilisée.
- Visualiser la structure DOM, de ce document.



J'ai utilisée l'outil

<https://fritscher.ch/dom-css/>

- Vous pouvez utiliser l'inspecteur du navigateur, ou manuellement,

Resume : HTML, offre certaine structure, mais, utile que pour l'afficge du document a l'utilisateur, et non pas pour l'echange.

## 2. Echange de données

Un autre programme veut extraire **automatiquement** les informations : le titre du document, et les types de vitamines .

- Il a proposer le code javascript ci-dessous, basee sur DOM (tant qu'il ya une structure, on peut l'exploiter):

```

// Access the DOM to extract the title of the page
var pageTitle = document.getElementsByTagName("title")[0].innerText;

```

- **Compléter le code** pour extraire l'information : *les vitamines du beurre de cacahuète*.

Il est defficile d'analyser un texte non structure, comme :

Le beurre de cacahuète est pour les bodybuilders une source naturelle de calories et de protéines à un prix abordable qui t'apportera en plus de l'acide folique (vitamine B9), de la vitamine E, du magnésium et du resvératrol. Une petite quantité de zinc est également présente,

- D'autres techniques avancees, comme les : du plus simple (expresions regulieres , vers les techniques de traitement automatique des langues (NLP : Natural langage Processing),

- Discuter les limites de HTML pour le **transfert de données** entre les applications WEB, et quels sont les solutions alternatives.
- **HTML faite pour l'affichage des documents WEB,**
- **- y'a pas une separation entre, les donnees, et leurs structure, et la mise en forme,**
- **Malgré l'existence du DOM, les balises n'ont aucune indication de leurs contenu, → risque de conflit, et d'ambiguïté ...comment comprendre : livre ? livre, a book, or (UK devise : livre sterling ).**
- **On a besoins d'une langage qui modelise les donnees separement, de leurs structure et leurs mise en forme → XML,....**

## Exercice N° 02 *Les BD, avantages et limites*

Un développeur a créé la base de données **Biblio**, pour une application WEB de gestion d'une bibliothèque universitaire. Un extrait de la table *livre*, de la base de données relationnelle **Biblio**, contient les enregistrements suivants :

nom	titre	nom-ed	rue-ed	ville-ed	etat-ed	pays-ed
Abit	Data on the Web	Kauf	340, Pine St.	S. Fr.	Cal	USA
Bune	Data on the Web	Kauf	340, Pine St.	S. Fr.	Cal	USA
Suciu	Data on the Web	Kauf	340, Pine St.	S. Fr.	Cal	USA
:						

1. **Structure** : De quel type vous considérer la structure des données stockées dans la table ci-dessus (structurée, semi, ou non structurée) ? Justifiez.

- Structurée, il existe une schéma , connue à l'avance avant la saisie des données (schéma relationnelle) qui structure les données ( tous les enregistrement suivre cette schéma).

2. **Quels sont les avantages des BD ?**

- Les avantages offerts par les SGBD : Performance, langages de requêtes, ...etc.
- On va pas remplacer les BD, dans cette matière, mais on étudier les cas ou XML compléter les BD, ou utiliser toute seule comme BD ( dernier chapitre)
- L'**extensibilité** : Comment le développeur peut enrichir la BD avec les informations de : L'université/filière /section/ groupe ? comment vous évaluer la complexité de cette extension, avec une requêtes comme : Ex, : je veux afficher l'université, la filière, le groupe, et la section de l'étudiant Mohamed qui a emprunter le livre Java) ?

- 3.

- Il doit ajouter des tables, et établir les relations entre les tables , par exemple (juste un exemple, le but n'est pas faire de la modélisation BD):

Table Univ ==1 ou plusieurs → Table filière == un ou plusieurs → Tables section ==1 ..+= > Table groupe 1..+ → étudiants < == plusieurs a plusieurs == → Livres

- L'évaluation : Pour certains informations qui doivent être ajouter, le concepteur, a ajouter plusieurs tables, donc : l'extensibilité, n'est pas facile.
- La recherche (exécution de la requête SQL): est très coûteuse, → des jointures entre plusieurs tables (resultats uploader a la RAM).

#### 4. Échange de données :

- Si on veut réaliser une application WEB **hypermédia** pour la lecture des livres en ligne. Comment présenter les informations des livres au navigateur, pour permettre aux lecteurs de lire tout le contenu des livres ?
- L'application dans sa version actuelle, ne permet pas de ce genre d'applications, sauf si on peut stocker (numérisés; en images, Pdf) les livres, mais ça reste l'aspect hypermédia (la navigation); et la recherche par contenu (HTML ne le permet pas).

### Exercice N° 03 : Initiation a XML

- Modéliser les données de l'exercice précédent, utilisant XML.
- Comparer les deux approches.

### Exercice N° 04 supplémentaire : La description de données

#### **Travail de Recherche (travail renvoyé par email) :**

XML a été utilisé pour développer les vocabulaires suivants :

- eb-XML (*electronic business XML*) pour le commerce électronique ;
- CheeseML pour l'agro-alimentaire ;
- MathML (*Mathematical Markup Language*) pour les notations mathématiques ;
- AIML (*Astronomical Instrument Markup Language*) pour les coordonnées célestes des galaxies ;
- BSML (*Biosequence ML*) pour les séquences de codes génétiques ;
- CML (*Chemical Markup Language*) pour la structure moléculaire des composés chimiques.

Question : Citez un exemple pour chaque langage décrit précédemment.

- Quelle est la sémantique ajoutée par les balises sémantiques HTML 5 ? des exemples ?