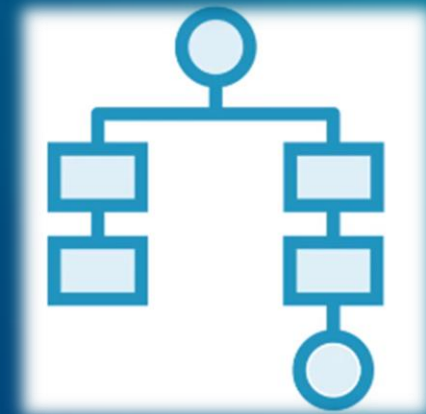


Université de BATNA 2  
Département d'informatique

# Cours : Les données semi-structurées

## L3 ISIL



# Chapitre 01 : Contexte et problématiques

Djennane, A  
Djennane.am@gmail.com

2023 -2024

# Contenu

**01**

**Contexte et  
problématique**

**02**

XML Core : syntaxe de base,  
espaces de noms.

**03**

La validation d'un document  
XML (DTD, XML Schema)

**04**

Xml et les base de données

**05**

La programmation XML :  
SAX, DOM, ...etc

# Contenu

**01**

Contexte et problématique

**02**

XML *Core* : syntaxe de base, espaces de noms.

**03**

La validation d'un document XML (DTD, XML Schema)

**04**

Xml et les base de données

**05**

La programmation XML : SAX, DOM, ...etc

# Contenu

**01**

Contexte et problématique

**02**

XML Core : syntaxe de base,  
espaces de noms.

**03**

**La validation d'un  
document XML  
(DTD, XML  
Schema)**

**04**

Xml et les base de données

**05**

La programmation XML :  
SAX, DOM, ...etc

# Contenu

**01**

Contexte et problématique

**02**

XML Core : syntaxe de base,  
espaces de noms.

**03**

La validation d'un document  
XML (DTD, XML Schema)

**04**

**Xml et les base de  
données**

**05**

La programmation XML :  
SAX, DOM, ...etc

# Contenu

**01**

Contexte et problématique

**02**

XML Core : syntaxe de base,  
espaces de noms.

**03**

La validation d'un document  
XML (DTD, XML Schema)

**04**

Xml et les base de données

**05**

**La programmation  
XML : SAX, DOM,  
...etc**

# Assessment

- CC (40%) = Test+ Project (3-5 students) + quiz\ test.
- Final exam (60%).

# Objectives de la matière

Cette matière va permettre aux étudiants :

- Faire la différence entre les données **structurées/semi-structurées** et **non structurées** (introduction pour le BIG DATA ...).
- Avoir des connaissances liées aux **formats d'échanges et de transportation de données entre les systèmes d'informations (XML, JSON, CSV)**.
- **Savoir utiliser les technologies XML (XPath, XQuery....)**.



# CHAPTER 1

On va voir dans ce chapitre :

- **Les données** : notion des documents, les bases de données, ..etc.
- **Problématiques** : *échange de données entre applications*
- **Catégories de données** : structurées, non-structurées, et semi- structurées,
- **Intro à XML.**

- **Les données d'une application** : les utilisateurs, les produits, historiques d'opérations => stockées généralement dans des BD.
- **Les documents informatiques** : **Hypertexte**, multimédias, hypermédia...
- **Les réseaux sociaux** : commentaires, opinions, ...
- **Les données géospatiales** : google Maps.

➤ **Plus de données** => **des bonnes décision**, meilleurs expérience utilisateur (*user experience*), compétitivité,.....



➤ Comment **stocker** ces données ?

📁 Comment **Échanger** ces données avec les différentes **applications** du SI ?

📁 Comment **analyser** ces données ?

➤ Différentes solutions :

- Les structures volatile (tableaux, matrices, ....)
- Les fichiers (textes, multimédias, ...)
- **Les bases de données**
- Les DataWerhouse (pour des besoins d'analytique)

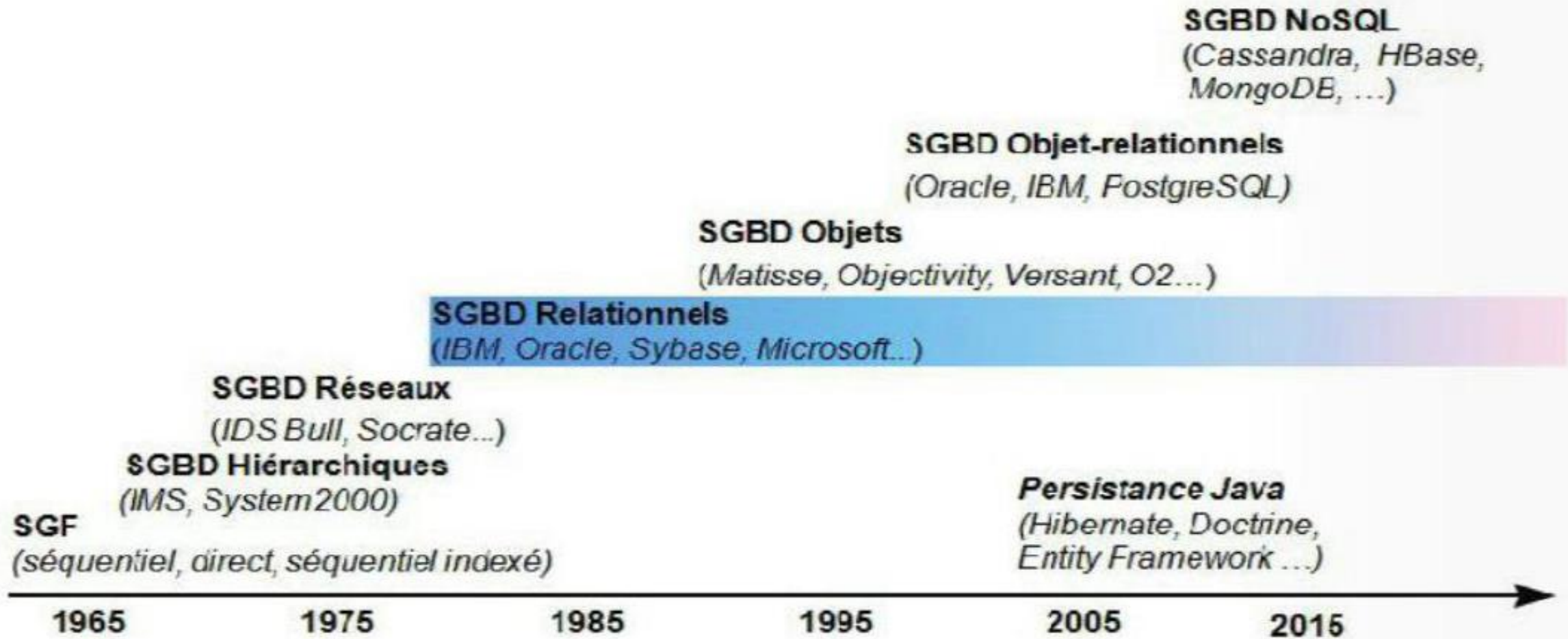


Figure 0-1. Historique des bases de données

- Chaque table est un fichier.
- Fichiers **binaires** **lisibles seulement** par le SGBD,
- SQL pour interroger la base.

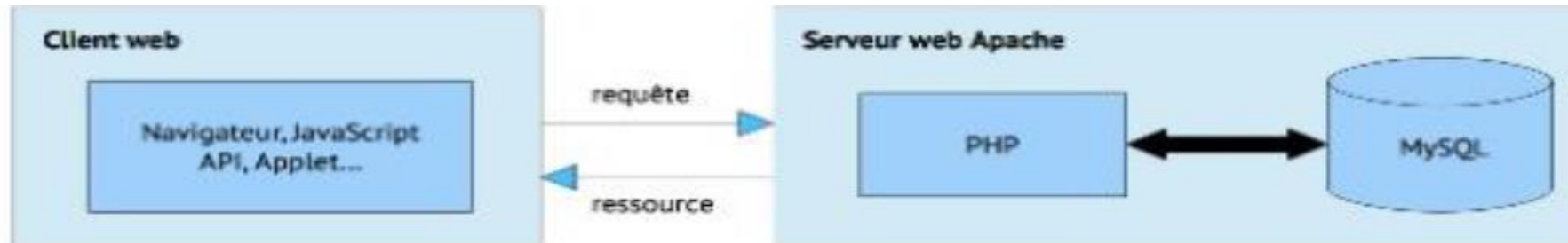


Figure : La stack LAMP.

### Stacks habituelles :

- LAMP : Linux, Apache, MySQL, PHP — la plus commune ;
- LEMP : Linux, Nginx, MySQL, PHP — commence à remplacer LAMP ;
- MEAN : MySQL, Express, AngularJS, Node.js — pour du JavaScript côté serveur.

➤ **Echange par Document HTML.**



# Problème : l'échange des données

## 01. Contexte et problématique



The image shows a screenshot of an Amazon India product page for MacBooks. The page displays three laptop models with their specifications and prices. The browser's developer tool is open, showing the HTML structure of the price information.

**Product Listings:**

- Apple MacBook Air MQD32HN/A 13.3-inch Laptop 2017 (Core i5/8GB/128GB/MacOS Sierra /Integrated Graphics)**  
Price: ₹55,990 (Original: ₹77,200) | You Save: ₹21,210 (27%)
- Apple Macbook Air MQD42HN/A 13-inch Laptop (Core i5/8GB/256GB/Mac OS/Integrated Graphics), Silver**  
Price: ₹72,990 (Original: ₹92,500) | You Save: ₹19,510 (21%)
- Apple MPXQ2HN/A Laptop (Core i5/8GB/128GB/Mac OS/Integrated Graphics), Space Grey**  
Price: ₹98,106 (Original: ₹1,04,900) | You Save: ₹6,794 (6%)

**Developer Tool HTML Snippets:**

```
<a class="a-link-normal a-text-normal" href="https://www.amazon.in/Apple-MacBook-Air-13-3-inch-Integrated/dp/B073Q...6VR/ref=sr_1_1?s=computers&ie=UTF8&qid=1526905827&sr=1-1&keywords=macbook">  
.. <span class="a-size-base a-color-price s-price a-text-bold"> == $0  
   <span class="currencyINR">&nbsp;&nbsp;&nbsp;</span>  
     "55,990"  
   </span>  
</a>
```

*Exemple 1 Une organisation publie des données financières stockées dans une BD relationnelle et des pages web sont créées après une requête SQL. Une autre organisation veut une analyse financière de ces données, mais elle a accès seulement aux pages HTML. Pour cela, elle ne peut qu'écrire du logiciel qui transforme du texte HTML en une structure de données adaptée à l'analyse.*

- Une petite modification du format d'un élément d'une page web peut casser ce logiciel !*
- Même si on a besoin seulement de la valeur moyenne d'une colonne d'une table, on peut avoir besoin de charger une base entière via plusieurs requêtes de pages HTML.*

- En plus les données habituelles (multimédias, clients,...), stockées au serveur ,on a besoins de stocker et manipuler les données **concernant le client** :

**Ex:** Les données de l'**historique de navigation** des utilisateurs (les cookies, les sessions, les champs de formulaire de connexion,...).

- **Utilisation** : enregistrer les activités des user (comme le contenu d'un panier d'achat d'une session précédente), afin de de **personnaliser** les **préférences** du site, **recommandations**, .....

- **Hypertexte** : Système de renvois permettant de passer directement d'une partie d'un document à une autre (**noeuds et de liens**), ou d'un document à d'autres document (lecture non linéaire grâce à la présence de liens entre les documents).
- **Hypermédia** : Extension de l'hypertexte à des données multimédias, permettant d'inclure des liens entre des éléments textuels, visuels et sonores.



# Exemple : Encyclopédie Medico-Chirurgicale (EMC)

## 01. Contexte et problématique

**- Problématiques :**  
*Recherche et indexation*

**1**

www.banq.qc.ca/ressources\_en\_ligne/index.html

Les plus visités Débuter avec Firefox À la une

Catalogue Archives **Ressources en ligne** Collections Services Activités Espace Jeunes Services adaptés Mon dossier

Accueil > Ressources en ligne

### Ressources en ligne

#### Revue, journaux et bases de données

- [Journaux québécois, canadiens et étrangers](#)
- [Livres](#)
- [Musique et vidéo](#)
- [BREF 1800 sites Internet de référence](#)

Voir toutes les bases de données Revues, encyclopédies, dictionnaires...

**Livres numériques**

Découvrez nos collections >>>

**2**

BIBLIOTHÈQUE ET ARCHIVES NATIONALES DU QUÉBEC

## Ressources en ligne

Catalogue Archives **Ressources en ligne** Collections Services Activités Espace Jeunes Services adaptés Mon dossier

Accueil > Ressources en ligne > Revues, journaux et bases de données

### Revue, journaux et bases de données

Rechercher une revue ou un journal

Revue, journaux et bases de données constituent une mine d'information sur les événements de l'actualité à la musique, des sciences humaines et sociales aux nouvelles technologies, en passant par les langues, la littérature et bien d'autres domaines.

Forts du désir d'offrir une collection de ressources électroniques riche et variée, les bibliothécaires de BANQ ont créé un ensemble important de ressources à partir des produits actuellement offerts sur le marché. Les titres disponibles étant plus nombreux en anglais qu'en français, nous demeurons vigilants afin d'équilibrer davantage cette collection au rythme de la création de nouvelles ressources de langue française.

En saisissant votre numéro de client et votre mot de passe, vous pouvez accéder aux bases de données demandant une authentification à partir de votre domicile. Vous devez être abonné à BANQ pour obtenir votre numéro de client et votre mot de passe.

**De A à Z**

A	B	C	D
<b>E</b>	F	G	H
I	J	K	L
M	N	O	P
Q	R	S	T
U	V	W	X

**Britannica IMAGE QUEST**

**Image Quest**

Image Quest regroupe près de trois millions d'images libres de droits, qui peuvent donc être utilisées pour des présentations, des projets pédagogiques, des activités de communication, etc. [Lire la description complète...](#)

# Exemple : Les bases documentaires

## 01. Contexte et problématique

### - Problématiques : Recherche et indexation

The screenshot shows the homepage of 'Techniques de l'Ingénieur'. At the top left is the logo with the text 'TECHNIQUES DE L'INGÉNIEUR' and the tagline 'L'expertise technique et scientifique de référence'. To the right is a search bar with the placeholder 'Tapez votre recherche ici'. Below the search bar is a navigation menu with items: 'RESSOURCES DOCUMENTAIRES' (highlighted in red), 'DOC QUIZ', 'PARCOURS PRATIQUES', 'LE MAGAZINE D'ACTUALITÉ', and 'CERT.I.'. On the far right of the menu is the text 'Contenu gratuit'. Below the menu, a paragraph states: 'Plus de 10 000 articles de référence, fiches pratiques, articles interactifs, Parcours Pratiques d'actualité : en quelques clics, accédez à une solution validée par des experts'. The main content area is divided into two columns. The left column is titled 'PAR DOMAINE D'EXPERTISE' and lists ten categories: AUTOMATIQUE - ROBOTIQUE, BIOMÉDICAL - PHARMA, CONSTRUCTION ET TRAVAUX PUBLICS, ÉLECTRONIQUE - PHOTONIQUE, ÉNERGIES, ENVIRONNEMENT - SÉCURITÉ, and GÉNIE INDUSTRIEL. The right column is titled 'PAR CARTOGRAPHIE INTERACTIVE' and features a blue box with a network diagram icon, a red 'EXPLORER' button, and a short text: 'La cartographie interactive est une manière plus ludique d'accéder aux contenus du site Technique de l'Ingénieur. Classés en 8 grands groupes de mots-clés, eux même'.

TECHNIQUES DE L'INGÉNIEUR  
L'expertise technique et scientifique de référence

Tapez votre recherche ici

RESSOURCES DOCUMENTAIRES DOC QUIZ PARCOURS PRATIQUES LE MAGAZINE D'ACTUALITÉ CERT.I. Contenu gratuit

Plus de 10 000 articles de référence, fiches pratiques, articles interactifs, Parcours Pratiques d'actualité : en quelques clics, accédez à une solution validée par des experts

PAR DOMAINE D'EXPERTISE

AUTOMATIQUE - ROBOTIQUE

BIOMÉDICAL - PHARMA

CONSTRUCTION ET TRAVAUX PUBLICS

ÉLECTRONIQUE - PHOTONIQUE

ÉNERGIES

ENVIRONNEMENT - SÉCURITÉ

GÉNIE INDUSTRIEL

INNOVATION

MATÉRIAUX

MÉCANIQUE

MESURES - ANALYSES

PROCÉDÉS CHIMIE - BIO - AGRO

SCIENCES FONDAMENTALES

TECHNOLOGIES DE L'INFORMATION

PAR CARTOGRAPHIE INTERACTIVE

→ EXPLORER

La cartographie interactive est une manière plus ludique d'accéder aux contenus du site Technique de l'Ingénieur. Classés en 8 grands groupes de mots-clés, eux même

- Logiciel Word de Microsoft exemple.doc
- Lecture avec Bloc Notes de Microsoft (ou un autre logiciel) exemple.txt
- ... ?
  - Impossible d'échanger les documents
  - ...

Les programmes peuvent difficilement interpréter un contenu non structuré :

The book 'Fundations of Databases', written by Serge Abiteboul, Rick Hull and Victor Vianu, published in 1995 by Addison-Wesley



- **Traitement automatique** des emails, pour détecter les emails frauduleux, ????
- **Traitement automatiquement** des commentaires pour connaître les sentiments des usagers ( *sentiment analysis* ) ???



- On a besoins des techniques d'analyses avancées (IA), comme le traitement de langage naturel (NLP), pour le traitement de ces données.

Utiliser les modèles de données des SGBD pose des problèmes :

- La structure est trop rigide
- les données peuvent ne pas être conformes au schéma  
( $\implies$  valeurs nulles, difficultés de traitement et ambiguïtés)
- l'évolution fréquente de la structure de données conduit à des évolutions de schéma pas toujours maîtrisées.
- les données du Web sont indexées par des moteurs de recherche dont les services sont limités, l'interrogation souvent imprécise.
- La majorité de données échangées ne sont pas **structurées**



**Nécessité d'un modèle général et souple, «sans schéma», avec un langage de requêtes associé : modèles semi-structurés**

- L'interrogation et l'échange de données entre systèmes d'information hétérogènes est souvent complexe
- XML contribue à résoudre ce problème
  - format d'échange normalisé indépendant de toute plateforme
- L'indexation et l'interrogation de grosses bases documentaires
  - informations structurelles en plus d'informations textuelles.

- Les données structurées sont des informations qui ont été formatées et transformées en un **modèle de données bien défini** => Les tables de bases de données constituées de lignes et des colonnes,
- Il constitue **20%** du total des données des entreprises.

# Les classes de données :

## 1. Données non structurées

### 01. Contexte et problématique

- Les données non structurées peuvent être tout ce qui **n'est pas dans un format spécifique**.
- **Exemple:** Word, PDF, texte, logs (fichier journal).
- Un exemple de données non structurées pourrait être les fichiers journaux qui ne sont pas faciles à séparer. Commentaires et publications sur les réseaux sociaux qui doivent être analysés.

Voici un exemple de données non structurées à partir d'un fichier journal.

```
38,P-R-38636-6-45,P-R-39105-1-11,P-R-38036-1-5,P-R-35697-1-13,P-R-35087-1-2  
Wed Sep 23 2020 05:21:01 GMT+0500
```



# Les classes de données :

## 1. Données semi-structurées

### 01. Contexte et problématique

Les données **semi-structurées** sont des informations qui **ne résident pas dans une base de données relationnelles**, mais qui **possèdent des propriétés organisationnelles facilitant leur analyse**. Avec certains processus, vous pouvez les stocker dans la base de données relationnelles

```
1.  <?xml version = "1.0" encoding="UTF-8" standalone="yes"?>
2.  <document>
3.    <employee>
4.      <name>Alex</name>
5.      <age>22</age>
6.    </employee>
7.    <employee>
8.      <name>Bob</name>
9.      <age>24</age>
10.   </employee>
11.   <employee>
12.     <name>Emily</name>
13.     <age>32</age>
14.   </employee>
15. </document>
```

Exemple : XML,  
JSON, Avro,  
Parquet, ORC,....

- Les données structurées nécessitent un schéma fixe qui est défini avant que les données puissent être chargées et interrogées dans un système de base de données relationnelle. Les données semi-structurées ne nécessitent pas de définition préalable d'un schéma et peuvent évoluer constamment, c'est-à-dire que de nouveaux attributs peuvent être ajoutés à tout moment.
  - Contrairement aux données structurées, qui représentent les données sous forme de table plate, les données semi-structurées peuvent contenir des hiérarchies d'informations imbriquées à  $n$  niveaux.
-



# Les classes de données : Comparaison

## 01. Contexte et problématique

Structurées	Semi-structurées	Non structurées
Il est basé sur les tables de base de données relationnelle	Il est basé sur XML/RDF	Il est basé sur des caractères et des données binaires
Il est dépendant du schéma et moins flexible	Il est plus flexible que les données structurées mais moins que les données non structurées	Très flexible et l'absence de schéma
Il est très difficile de mettre à l'échelle le schéma de base de données	La mise à l'échelle est plus simple que les données structurées	C'est très facile à mettre à l'échelle



XML : nouveau standard adopté par le World Wide Web Consortium (W3C) comme complément de HTML permettant un échange aisé de données de sur le web.

- Le but principal de XML n'est pas de décrire un format de texte, mais de structurer logiquement un contenu.
- Les balises ont le rle de classer des données selon une hiérarchie définie par l'auteur du document XML.

XML = Extensible Markup Language

C'est un langage permettant de représenter et structurer des informations à l'aide de balises que chacun peut définir et employer comme il le veut.

```
texte ... <BALISE> ... texte </BALISE> ...
```

### **Standard Generalized Markup Language**

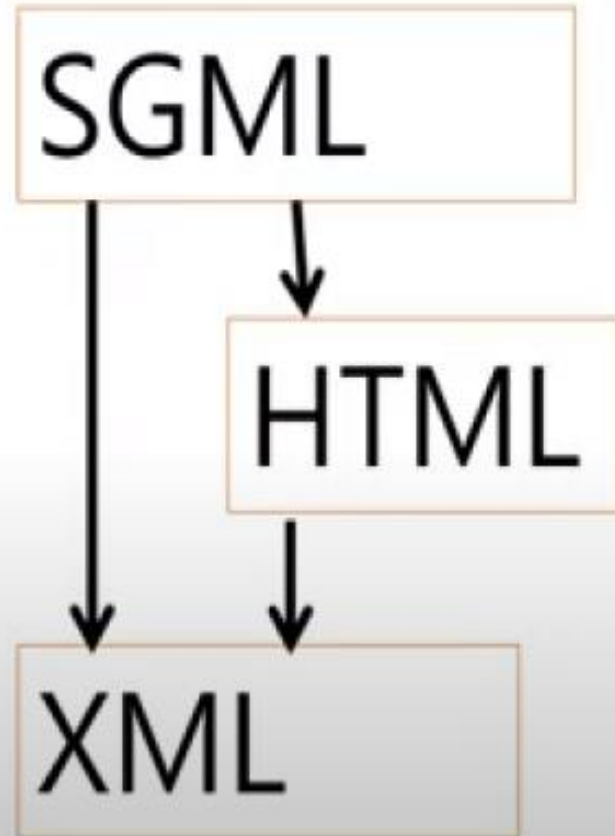
( Sépare les données la structure des données et la mise en forme )

### **Hyper Text Markup Language**

( Mélange les données et la mise en forme )

### **eXtensible Markup Language**

( Sépare les données la structure des données et la mise en forme )



- Avec XML, la mise en forme textuelle est effectuée dans une *feuille de style*, un document séparé qui associe des formes de présentations (texte en gras, en italique, centré, etc.) aux balises. Des feuilles différentes permettent des formattages différents du même document.
- Des outils permettent de convertir un document XML en HTML, afin de pouvoir afficher une page web.

- La forme sérialisée est une représentation textuelle linéaire d'un arbre satisfaisant une syntaxe.
- Il existe un modèle orienté-objet permettant de représenter des données XML sous forme d'arbre : Document Object Model (W3C). Les applications "travaillent" avec ce genre de représentations sous forme d'arbre.




### Processing XML

- Application
- XML Processor
- XML Document





# XML Processor

- Read XML
-  Build structure
  - logical
  - physical
- Well formed
- Encoding

# XML Processor

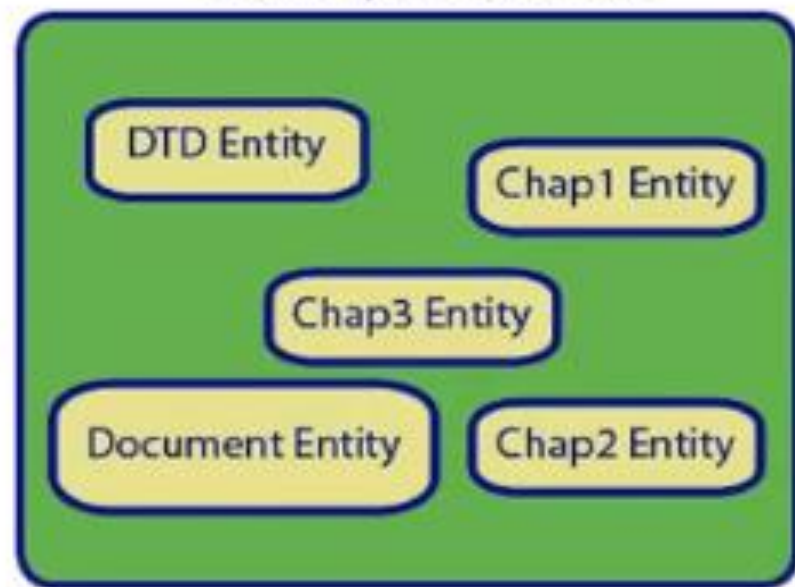


Read XML

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE book [
  <!ELEMENT book (chapter+)>
  <!ELEMENT chapter (title, p+)>
  <!ELEMENT p (#PCDATA)>
  <!ELEMENT title (#PCDATA)>
  <!ENTITY chap1 SYSTEM "chapter1.xml">
  <!ENTITY chap2 SYSTEM "chapter2.xml">
  <!ENTITY chap3 SYSTEM "chapter3.xml">
]>
<book>
  <chapter>&chap1;</chapter>
  <chapter>&chap2;</chapter>
  <chapter>&chap3;</chapter>
</book>
```



Book Physical Structure



Le format XML est au cœur de nombreux processus actuels :

- format d'enregistrement de nombreuses applications,
- échange de données entre serveurs et clients,
- outils et langages de programmation,
- bases de données XML natives.

XML est la norme d'enregistrement de nombreux logiciels parmi lesquels on peut citer :

- Bureautique
  - LibreOffice : format **OpenDocument**
  - Publication de livres et documentations : **DocBook**
- Graphismes
  - Dessin vectoriel avec Inkscape : format **SVG**
  - Équations mathématiques : format **MathML**
- Programmation
  - Interfaces graphiques : Android, **XUL** (mais le futur **WebExtensions** est tourné vers JS et JSON)



Les

➤ S

➤ N

➤ S\

➤ Xf

➤ ...

```
<mrow>
  <msup>
    <mfenced>
      <mrow>
        <mi>a</mi>
        <mo>+</mo>
        <mi>b</mi>
      </mrow>
    </mfenced>
    <mn>2</mn>
  </msup>
</mrow>
```

Listing 2 – Extrait d'un document MathML



FIGURE 2: Document *interprété* par Firefox

# Application de XML : le langage SMIL pour le multimedia

## 01. Contexte et problématique

### Un exemple de document multimédia



Synchronisation

Affichage

Interactivite



### Un document multimédia :

- Dimension spatiale
- Dimension contenu
- Dimension temporelle
- Dimension hypermédia

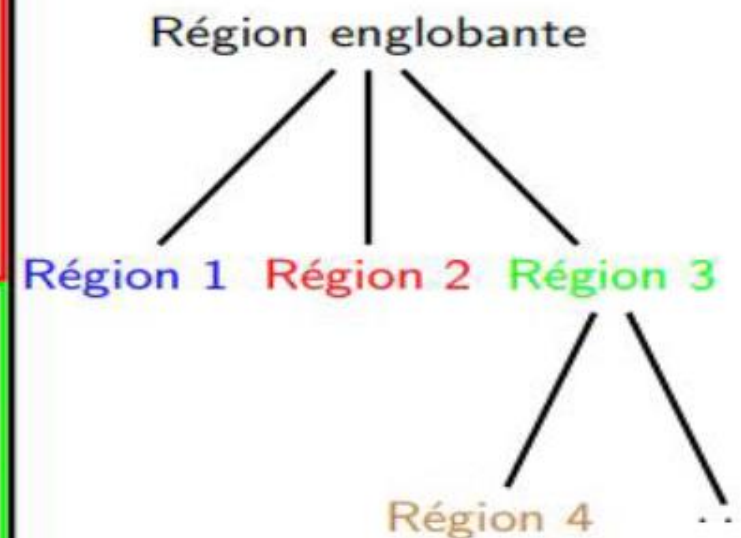
### Autres fonctionnalités

- Les animations

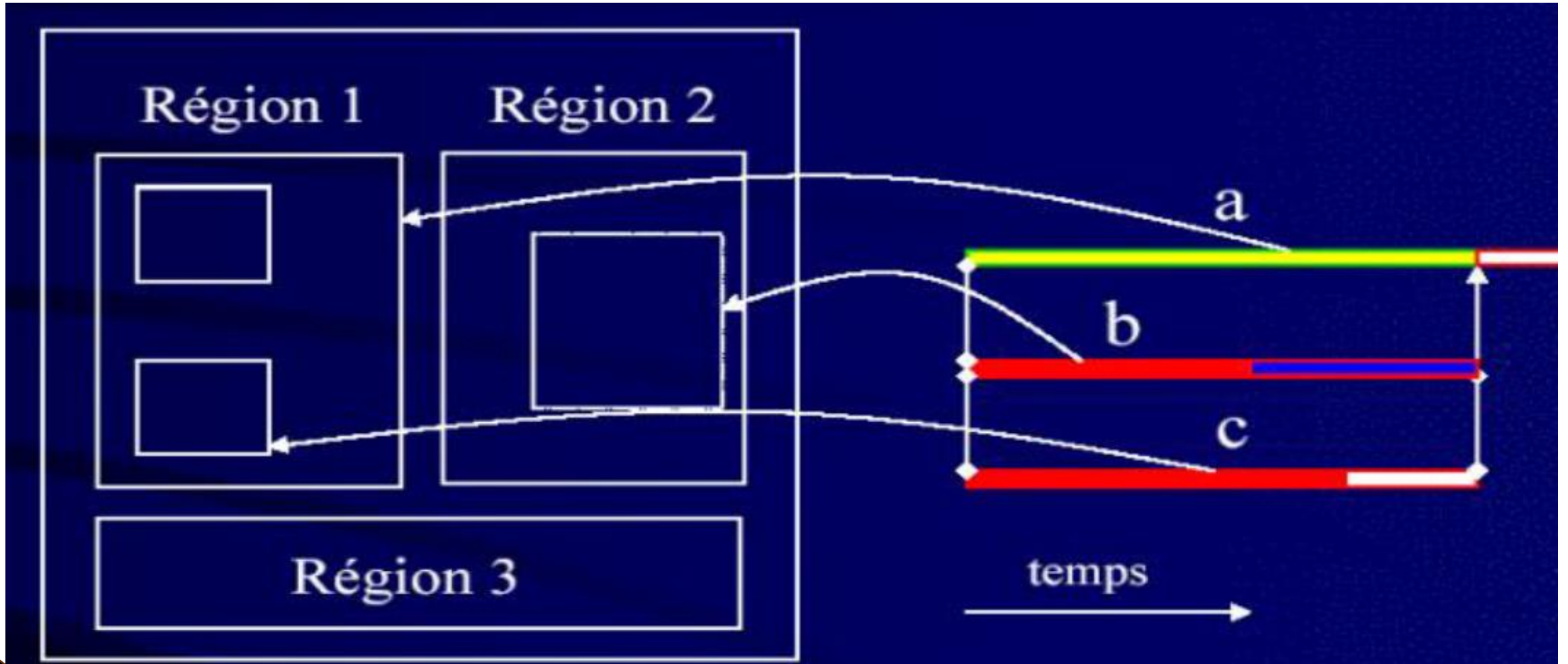
### L'aspect spatial ;

#### Spécification des zones d'affichage

The image shows a woman in a kitchen setting on the left. On the right, there is a recipe card for 'Creme anglaise'. The recipe card is divided into two sections: the top section contains the title 'Creme anglaise' and a photograph of the finished cream; the bottom section is titled 'Ingredients' and lists the following items: 25cl de creme liquide, 25cl de lait, 100g de sucre, 6 oeufs, and une gousse de vanille. A coordinate system with x and y axes is positioned in the top left corner of the overall image.



### L'aspect temporelle :



**SMIL est un standard W3C.**

**Exemple : Spécification des objets multimédias :**

- Audio :

```
<audio id="..." src="..." begin="..." dur="..." end="..." />
```

- Vidéo :

```
<video id="..." src="..." begin="..." dur="..." />
```

- Texte :

```
<text id="..." src="..." region="..." />
```

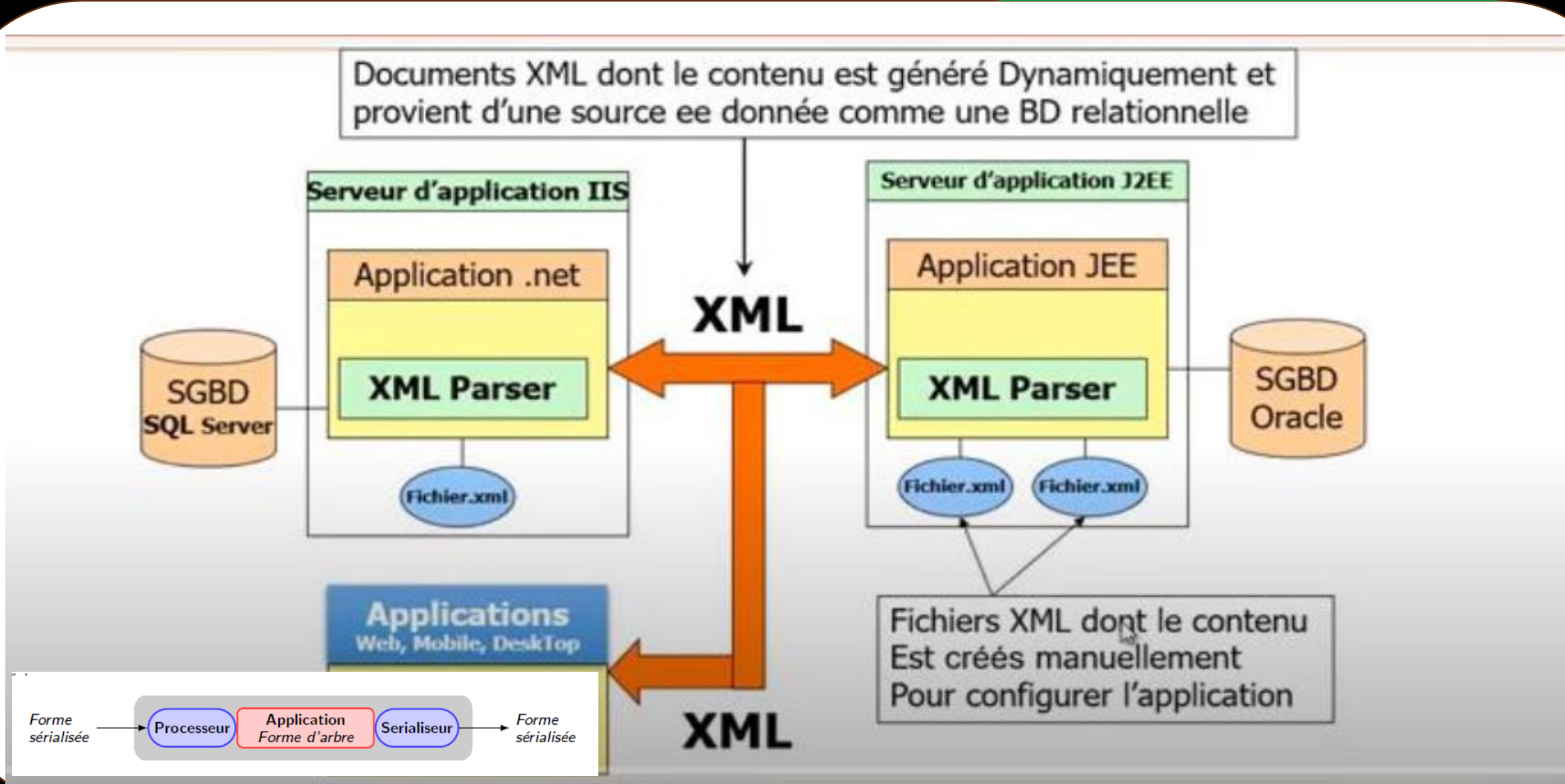
- Image :

```

```

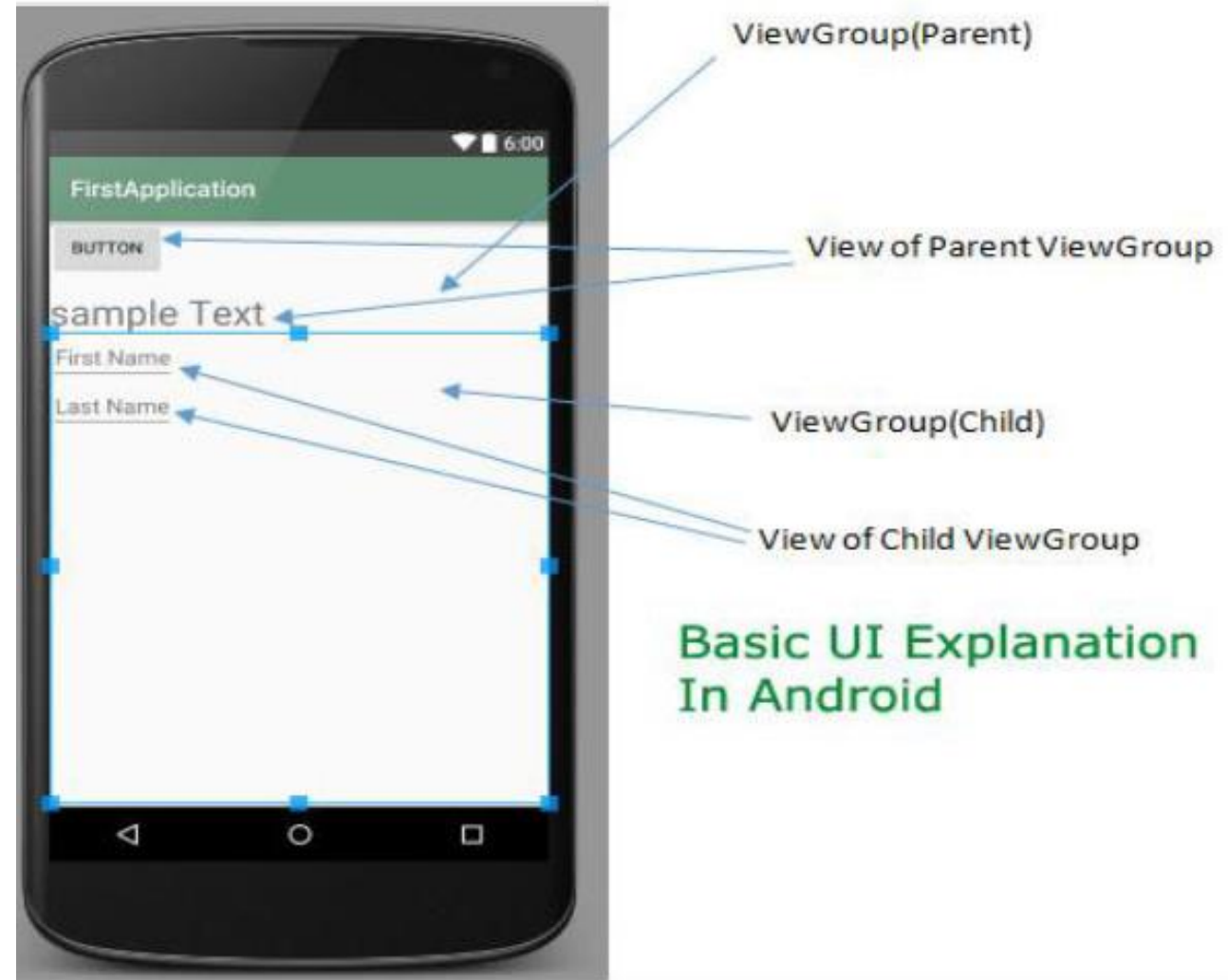
# Application de XML : échange de données entre les SI

## 01. Contexte et problématique





### Programmation ANDROID





# Programmation ANDROID

```
<?xml version="1.0" encoding="utf-8"?>
<LinearLayout xmlns:android="http://schemas.android.com/apk/res/android"
    android:orientation="vertical" android:layout_width="match_parent"
    android:layout_height="match_parent">

    <Button
        android:id="@+id/buton1"
        android:layout_width="wrap_content"
        android:layout_height="wrap_content"
        android:text="Button"/>

    <TextView
        android:id="@+id/textView1"
        android:layout_width="wrap_content"
        android:layout_height="wrap_content"
        android:text="sample Text"
        android:layout_marginTop="15dp"
        android:textSize="30dp"/>

    <RelativeLayout
        android:layout_width="match_parent"
        android:layout_height="match_parent">

        <EditText
            android:id="@+id/editTextName"
            android:layout_width="wrap_content"
```

- Bases de données **XML native** : les données sont au format XML et les requêtes sont dans un langage (XQuery) permettant de réaliser l'équivalent de SQL.

# La Galaxy XML

## 01. Contexte et problématique

### Au centre

XML

### Premier Cercle

parcours/lecture : parseur(DOM, SAX...)

### deuxième Cercle

La grammaire/validation : DTD, XML Schema, Relax NG...

### Troisième Cercle

La sélection : Xpath

La transformation et requête : XSLT, XQUERY, DOM

### Quatrième Cercle

les liens : Xlink, Xinclude, XPointer...

### Cinquième Cercle

Producteur/porteur de XML : Xforms, XSQL, SOAP

### Sixième Cercle

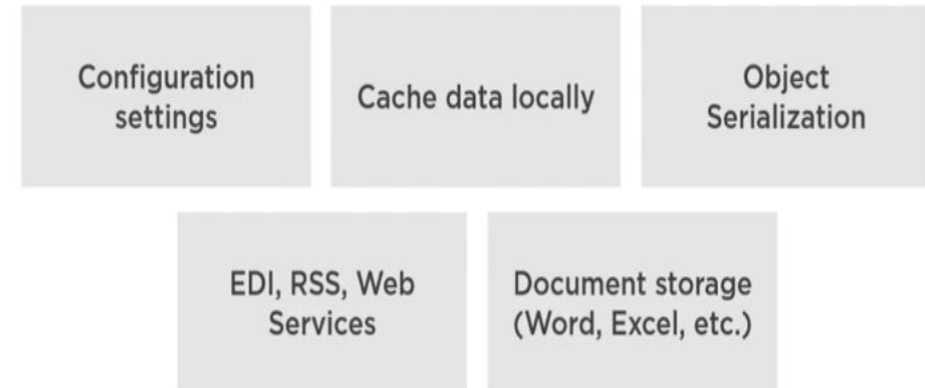
Les langages descriptifs de format XML : Open document, SVG, RSS, MathML, XBRL...

### Septième Cercle

Métalangage au service des métadonnées : RDF



Programming languages : JAVA, python, Javascript, PHP , .....



### **La sélection :**

XPath est un langage avec une syntaxe non XML, permettant d'adresser les différents nœuds ou groupes de nœuds particuliers d'un document XML.

### **La transformation :**

XSLT : langage de forme XML. Basé sur un système de templates, il permet de transformer un flux XML soit en un autre flux XML, soit HTML ou simplement texte

### **La requête :**

Xquery : langage qui n'utilise pas la forme XML. Il permet à partir d'un ou plusieurs documents XML de reconstruire un nouveau XML

***Time for coding,...***

Un fichier XML représente des informations structurées :



```
<?xml version="1.0" encoding="utf-8"?>
<!-- itinéraire fictif -->
<itineraire>
  <etape distance="0km">départ</etape>
  <etape distance="13km">tourner à droite</etape>
  <etape distance="22km">arrivée</etape>
</itineraire>
```

Cet exemple modélise un itinéraire composé d'étapes.

XML permet de choisir la représentation des données sans aucune contrainte. On choisit les balises et les attributs comme on le souhaite.



### Arborescence d'éléments

Un document XML est composé de plusieurs parties :

- Entête de document précisant la version et l'encodage,
- Des règles optionnelles permettant de vérifier si le document est valide
- Un arbre d'*éléments* basé sur un élément appelé *racine*
  - Un *élément* possède un *nom*, des *attributs* et un *contenu*
  - Le contenu d'un élément peut être :
    - rien : élément vide noté `<nom/>` ou `<nom attributs.../>`
    - du texte
    - d'autres éléments (les éléments enfants).
  - Un élément non vide est délimité par une *balise ouvrante* et une *balise fermante*.
    - une balise ouvrante est notée `<nom attributs...>`
    - une balise fermante est notée `</nom>`

### Choses interdites

Les règles d'imbrication XML interdisent différentes configurations qui sont plus ou moins tolérées en HTML :

- plusieurs racines dans le document,
- des éléments non terminés (NB: XML est sensible à la casse),
- des éléments qui se chevauchent.

```
<element1>  
  <element2>  
  </Element2>  
  <element3>  
  </element1>  
</element3>
```

En XML, cela crée des erreurs « *document mal formé* ».

# Structure d'un document XML : les attributs

## 01. Contexte et problématique

Un élément peut être porteur d'un ou plusieurs attributs.

xml

[Sélectionnez](#)

```
<Element att1="test1" att2="test2">...</Element>
```

Ici Element est porteur de deux attributs att1 et att2.

Les noms des attributs suivent les mêmes règles d'écriture que ceux des éléments.

Les attributs ne peuvent contenir que du texte.

Un élément ne peut posséder deux attributs de même nom, ainsi le code suivant est faux

xml

[Sélectionnez](#)

```
<Element att1="test1" att1="test2">...</Element>
```

# *Research*

- Les bases de données en biologie
- Exemple de données de capteurs : mesurant des données comme CO2, glycémie, GPS...etc.
- Les concepts suivants : Linked data, WEB mining
- Langage de programmation : AJAX

Thank you