

XML et Données Semi Structurées

Introduction

Introduction

L'utilisation d'internet a considérablement augmenté ces dernières années. Cette croissance est à associer à la naissance de deux standards:

- **HyperText Markup Language (HTML):** langage associant les données et la manière de les présenter
- **HyperText Transfert Protocol (HTTP):** protocole de transfert d'informations sur internet

Introduction

- Contrairement au protocole HTTP, le langage HTML a beaucoup évolué depuis son apparition.
- A l'origine, dans un document HTML, il n'y avait pas de prédominance entre les données et la présentation de ces données.
- Très rapidement, les normes de HTML ont laissé la présentation prendre de plus en plus de place dans un document HTML.

Introduction

- Il est donc apparu la nécessité de définir un nouveau langage qui serait centré sur la description des données: XML

eXtensible Markup Language

Introduction

❑ Qu'est-ce que XML (Langage à balise extensible) ?

C'est une spécification du stockage des informations et de la structure des informations stockées. Il ne définit aucune balise qui lui soit propre:

- *les utilisateurs peuvent créer celles dont ils ont besoin,*
- *les balises doivent toutes se conformer à la spécifications XML*

Introduction

❑ **Qu'est-ce que XML (Langage à balise extensible) ?**

C'est également un format standard pour les communications entre applications. Il est utilisé dans des projets de publication sur le web ainsi que les bases de données.

Introduction

□ Historique

- **XML** dérive du langage développé dans les années 80, le **SGML** (Standard Generalized Markup Language).
- L'idée principale de ce langage (**SGML**) était de séparer le contenu d'un document de sa forme.
- Ce langage (**SGML**) était si complexe que sa manipulation s'est trouvée restreinte aux spécialistes

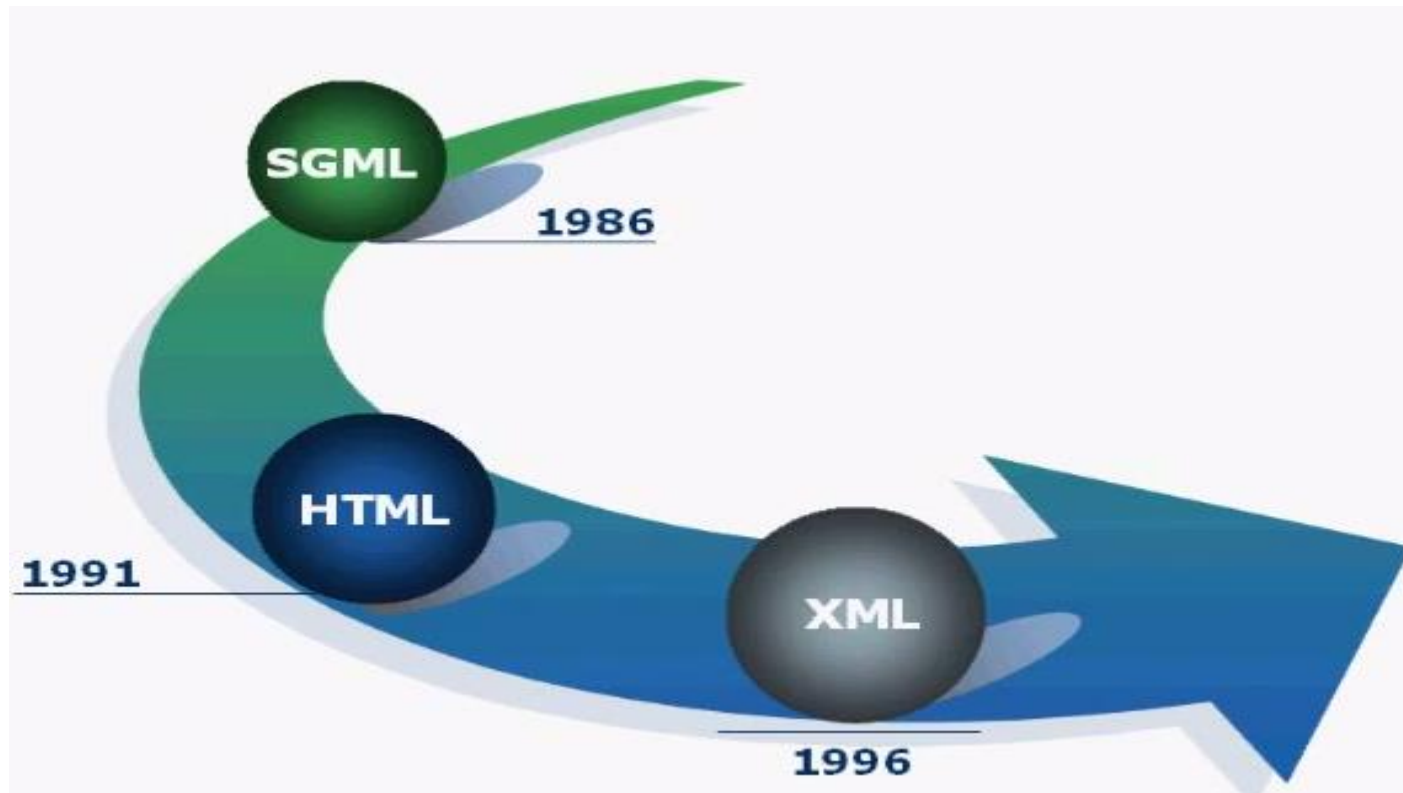
Introduction

❑ Historique

- En **1991** internet popularise **HTML**, mais ce dernier; malgré de nombreuses adaptations, ne pouvait pas être étendu à l'infini.
- C'est alors en **1996** que fut créé le **XML** par le **World Wide Web Consortium (W3C)**.

Introduction

❑ Historique



Introduction

❑ XML vs. HTML

- Ils sont deux langages à balises.
- Quelques points communs existent entre eux, mais il y a des différences importantes:
 - Le nom des balises
 - **XML**: les balises sont à définir en fonction des besoins
 - **HTML**: les balises sont fixées en vue de l’affichage

Introduction

❑ XML vs. HTML

▪ La syntaxe des balises

- **XML**: plutôt stricte
- **HTML**: assez tolérant, dépend du navigateur

▪ Utilisation du langage

- **XML**: tous types de besoins
- **HTML**: limité à la mise en forme de pages web

Introduction

❑ Intérêts

- Structurer l'information en fonction de la sémantique des données,
- Rendre explicite le contenu des documents:
 - Lisible par tous,
 - Modifiable simplement avec un éditeur de texte.

Introduction

❑ A quoi ressemble un document XML ?

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<employees>
  <employee>
    <employee_id>120</employee_id>
    <last_name>Weiss</last_name>
    <salary>8000</salary>
  </employee>
  <employee>
    <employee_id>121</employee_id>
    <last_name>Fripp</last_name>
    <salary>8200</salary>
  </employee>
</employees>
```