

Langages à balises

Objectifs du cours

- ▶ Comprendre les grands principes de la représentation de données et de documents numériques à l'aide d'un langage à balises
- ▶ Apprendre les bases d'HTML pour la génération de pages web
- ▶ Définir des langages basés sur XML à l'aide de DTD

Plan

Cette séance et la semaine prochaine :

- ▶ Introduction aux langages à balises
 - ▶ SMGL, XML, HTML.
- ▶ XML
- ▶ DTD

La prochaine séance

- ▶ HTML5
- ▶ CSS

Plan

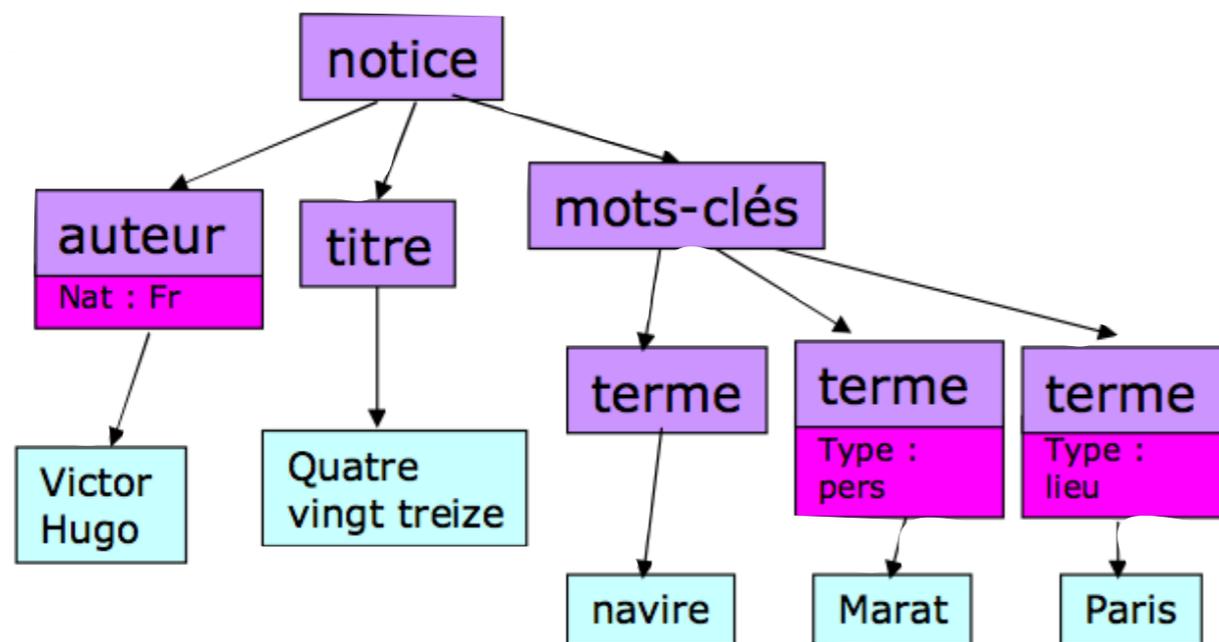
- ▶ Introduction aux langages à balises
 - ▶ **Langage à balises**
 - ▶ SGML
 - ▶ HTML
 - ▶ Aparté : le Web en 3 minutes
 - ▶ XML
- ▶ HTML
- ▶ CSS

Pourquoi des langages à balises?

Représenter des documents de manière structurée

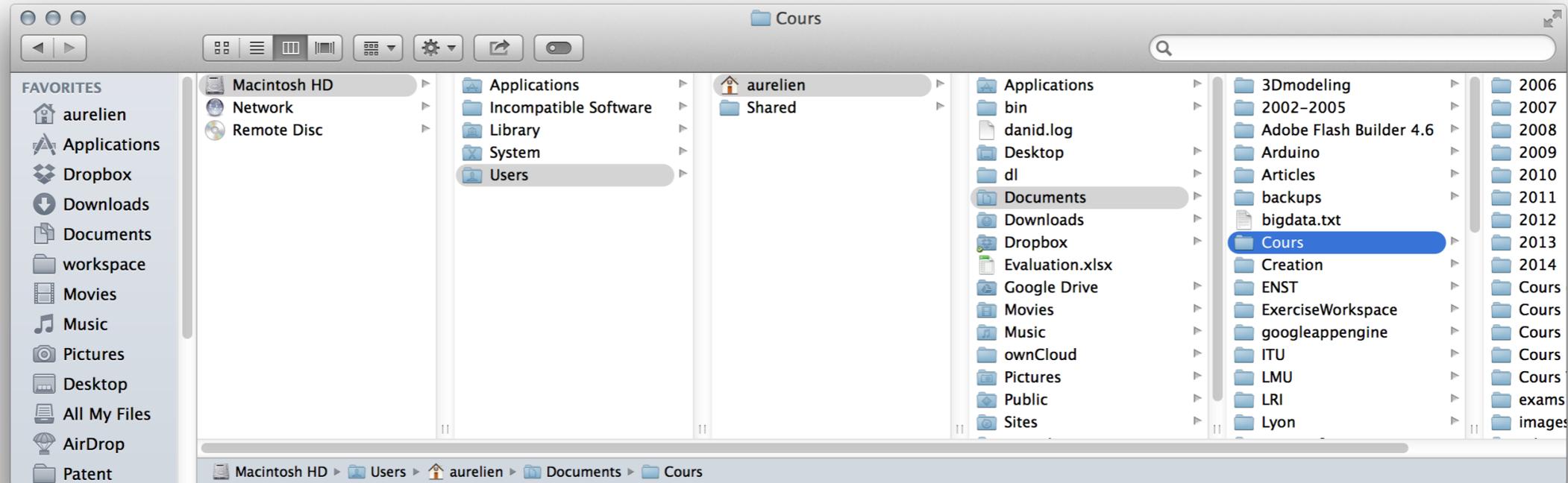
- ▶ Structure arborescente
- ▶ Stockée dans des fichiers *échangeables*

Exemple :



```
<?xml version="1.0"
encoding="ISO-8859-1"?>
<notice>
<auteur nat="fr">Victor Hugo
</auteur>
<titre>Quatre vingt treize</titre>
<mots-clés>
<terme>navire</terme>
<terme Type="pers">Marat</terme>
<terme Type="lieu">Paris</terme>
</mots-clés>
</notice>
```

Les arbres sont partout!



```
<!DOCTYPE html>
<html itemtype="http://schema.org/WebPage" itemscope="">
  <head></head>
  <body id="gsr" class="hp vasq" vlink="#61c" text="#222" link="#12c" bgcolor="#fff" alink="#dd4b39"
  onload="try{if(!google.j.b){document.f&&document.f.q.focus();documen...if(document.images)new Image().src='/images
  /nav_logol170.png'}">
    <div id="viewport" class="ctr-p"></div>
    <script src="/xjs/_/js/k=xjs.s.en_US.EeLgqkzqnSg.O/m=sy15,cdos,sy32,gf,vm..m=EAIaow/rt=j/d=0
    /sv=1/rs=AItrSTPtusUgvcCk9AJNpSwQcRE5i0aZ0Q" gapi_processed="true"></script>
    <table class="gstl_0 gssb_c" cellspacing="0" cellpadding="0" style="width: 571px; display: none; top: 340px;
    position: absolute; left: 269px;"></table>
    <style type="text/css"></style>
  </body>
</html>
```

Représentation

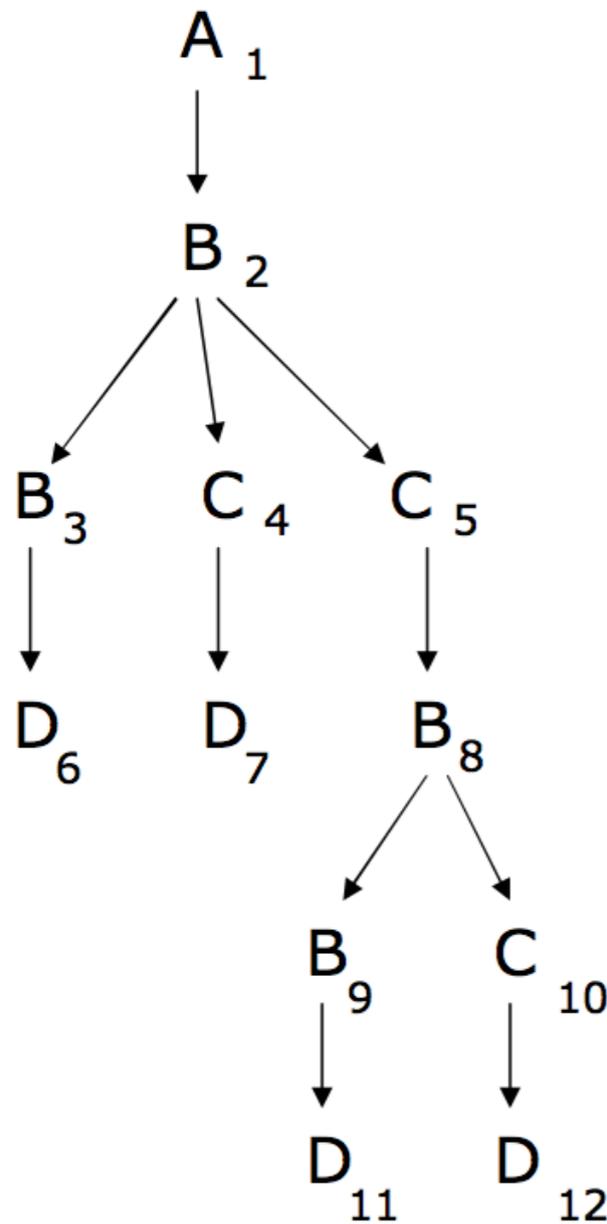
Fichier : suite linéaire d'octets

Comment représenter une structure d'arbre dans un fichier ?

Solution

- ▶ décrire l'arbre comme un ensemble d'éléments qui se contiennent les uns les autres.
- ▶ représenter les éléments entre deux balises
 - ▶ balises ouvrantes, on les notera par exemple <nom>
 - ▶ balises fermantes, on les notera par exemple </nom>

Exemple



Éléments

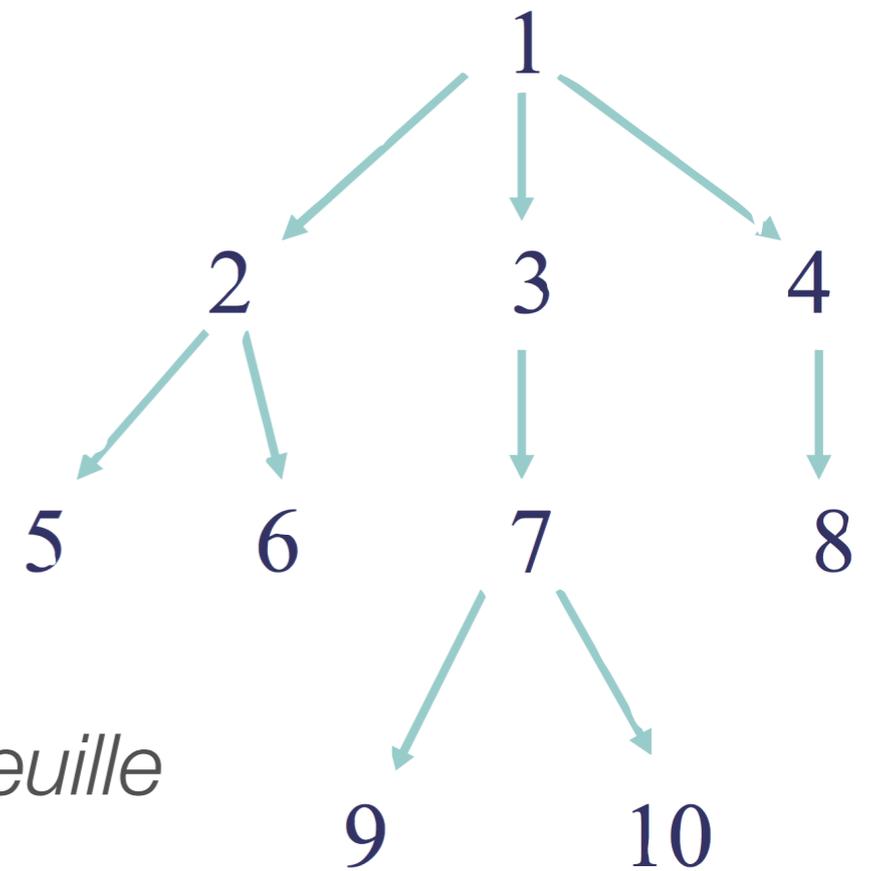
A₁ ⊂ B₂
B₂ ⊂ B₃ C₄ C₅
B₃ ⊂ D₆
C₄ ⊂ D₇
C₅ ⊂ B₈ C₁₀
B₈ ⊂ B₉
B₉ ⊂ D₁₁
C₁₀ ⊂ D₁₂

Éléments et balises

```
<A>  
  <B>  
    <B>  
      <D></D>  
    </B>  
    <C>  
      <D></D>  
    </C>  
    <C>  
      <B>  
        <B><D></D></B>  
        <C><D></D></C>  
      </B>  
    </C>  
  </B>  
</A>
```

Parler des arbres

- ▶ Arbre
- ▶ Noeud : nœuds fils et pères
- ▶ Racine : *nœud initial*
- ▶ Feuille : *nœud terminal*
- ▶ Chemin : *suite de nœuds dont chaque élément a pour père le précédent*
- ▶ Branche : *chemin se terminant sur une feuille*
- ▶ Ancêtres et descendants
- ▶ Taille d'un arbre : *nombre de nœuds*
- ▶ Profondeur d'un nœud



Langage à balises

Tous les langages ayant pour objectif de représenter de l'information en utilisant des balises sont définis par :

- ▶ vocabulaire
 - ▶ noms des éléments
- ▶ grammaire
 - ▶ mode d'organisation des éléments des éléments en contiennent d'autres
- ▶ + attributs des éléments
 - ▶ un peu plus de structure (voir cours XML)
- ▶ Une description
 - ▶ ensemble d'éléments organisés dans un fichier
 - ▶ contenus terminaux (texte)

Balisage de texte

Idée

- ▶ marquer des zones des textes pour les qualifier
- ▶ les balises ouvrantes et fermantes délimitent les éléments de description
- ▶ la structure logique est un arbre « ajouté » au texte

```
<p> Il est de tradition de présenter un langage de
programmation à l'aide d'un premier exemple
comme : <eg> CHAR*20 GRTG GRTG = 'BONJOUR TOUT LE
MONDE' PRINT *, GRTG END </eg>
</p>
<p> Dans cet exemple, on commence par déclarer la
variable <ident>GRTG</ident>, dans la ligne
<kw>CHAR*20 GRTG</kw>, qui identifie <ident>GRTG</
ident> comme formée de 20 octets de <kw>CHAR</kw>.
On affecte alors à cette variable la valeur finale
<kw>END</kw>.
</p>
```

p : servira à la mise en page

eg, kw,mentioned : seront
mis en évidence dans la
structure physique

kw,mentioned : utilisés pour
construire un index

Plan

- ▶ Introduction aux langages à balises
 - ▶ Langage à balises
 - ▶ **SGML**
 - ▶ HTML
 - ▶ Aparté : le Web en 3 minutes
 - ▶ XML
- ▶ HTML
- ▶ CSS

SGML

- ▶ Objectif : représenter l'information contenue dans un document indépendamment
 - ▶ des systèmes utilisés pour la saisie et le traitement
 - ▶ de la forme physique qu'il sera amené à prendre (papier, CD-ROM, web...)
 - ▶ des langues et des alphabets, latins ou non
 - ▶ des applications
- ▶ Naissance chez IBM (années soixante)
 - ▶ GML
 - ▶ Gestion de la documentation technique
- ▶ Normalisation 1986 ISO-8879
 - ▶ une dizaine d'années de travail

SGML

Utilisation :

- ▶ Description des documents dans les grosses organisations
 - ▶ Complexité des langages
 - ▶ Lourdeur et cherté des outils (chaîne de traitement)
 - ▶ Journal Officiel, grosses entreprises/documentations, éditeurs...
- ▶ Echange des documents

SGML : principes

Métalangage

- ▶ permet de décrire des modèles (grammaires)

Notion de DTD (Document Type Definition)

- ▶ Permet de décrire un modèle, i.e. un type de document

Un document SGML

- ▶ Est une instance du type de document
- ▶ Doit être conforme à la DTD associée

SGML Exemple

Instance

```
<!DOCTYPE memo SYSTEM
  "memo.dtd">
<memo statut="conf">
<auteur>Serge Fleury</auteur>
<dest>
<nom>André Salem</nom>
<nom>Pollet Samvelian</nom>
</dest>
<sujet>Cours SLFE6</sujet>
<corps>
<par>Veuillez noter que le cours SLFE6
  sur les documents électronique aura
  bel et bien lieu au mois de mai
  2002</par>
<par>S'il y avait des changements de
  votre côté, veuillez m'en aviser
  dans les plus brefs délais.</par>
</corps>
</memo>
```

DTD (memo.dtd)

```
<!-- DTD utilisable pour baliser les
  memos en SGML -->
<!ELEMENT memo -- ((auteur & (date?)
  & sujet & dest & (cc?)), corps)>
<!ATTLIST memo statut (conf | pub)
  pub>
<!ELEMENT (dest | cc) -- (nom+)>
<!ELEMENT corps -- (par*)>
<!ELEMENT (auteur | date | sujet | nom
  | par) -- (#PCDATA)>
```

Métalangage

Langage avec lequel on peut définir d'autres langages

Pour les langages à balises

- ▶ langage exprimant la manière dont on peut décrire une famille de langages à balise
- ▶ comment exprimer les éléments ?
- ▶ comment organiser les éléments ?

Exemples de métalangages :

- ▶ SGML, permet de définir : TEI, HTML, ...
- ▶ XML, permet de définir : SVG, TEI, XHTML, ...

Plan

- ▶ Introduction aux langages à balises
 - ▶ Langage à balises
 - ▶ SGML
 - ▶ HTML
 - ▶ **Aparté : le Web en 3 minutes**
 - ▶ XML
- ▶ HTML en pratique
- ▶ CSS

Le Web

- ▶ 1991 – CERN – Tim Berners Lee : a "global information space."
- ▶ Basé sur
 - ▶ Principes de l'hypertexte
 - ▶ Client/serveur sur IP
- ▶ Principes
 - ▶ Des serveurs peuvent fournir des documents hypertextes
 - ▶ Les documents seront décrits en suivant une DTD SGML
 - ▶ HTML (HyperText Markup Language)
 - ▶ Les liens sont décrits avec leur cible (URL)
 - ▶ Un client (navigateur)
 - ▶ permet de lire (présenter) les documents HTML
 - ▶ charge un nouveau document quand on active un lien
 - ▶ Protocole d'échange : HTTP (HyperText Transfert Protocol)

Question

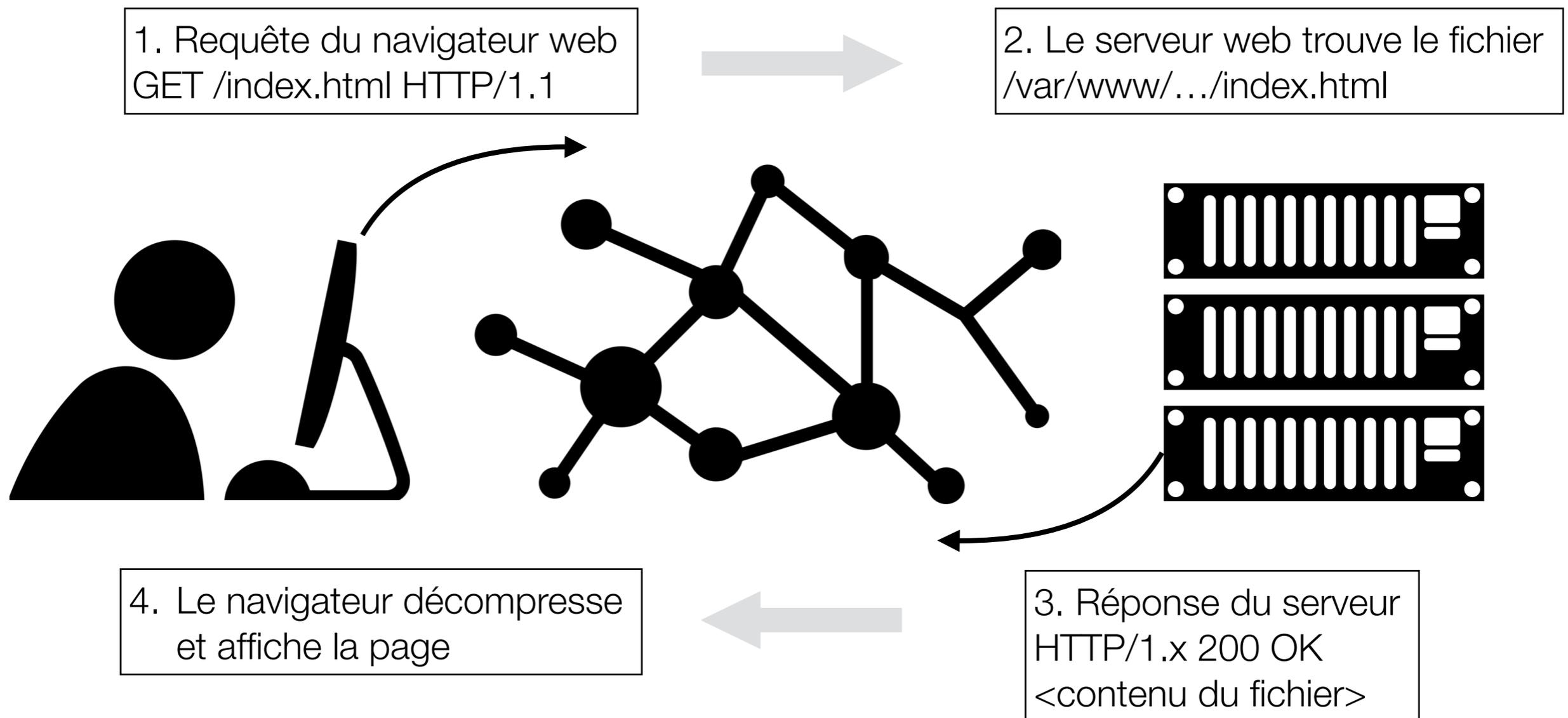
Différence entre Web et Internet?

Cours sur architecture de l'Internet et du Web :

<http://courses.ischool.berkeley.edu/i153/su11/internet>

Requête HTTP simple

Creative Commons – Attribution (CC BY 3.0)
User designed by Luis Prado
Internet designed by Fernando Vasconcelos
Servers designed by Jaime Carrion from the Noun Project



HTML

Live demo : <https://fr.wikipedia.org/>,
plus dans la deuxième partie du cours

URL

Un élément constituant du Web!

Uniform Resource Locator

- ▶ permet d'identifier une ressource sur le réseau

Une ressource peut être

- ▶ une page Web
- ▶ une image (seule ou utilisée dans une page Web)
- ▶ un programme
- ▶ un fichier à télécharger...

URL

Exemple : <http://tabard.fr/courses/2015/cci/web.pdf>

Une URL indique

- ▶ un protocole (langage de communication entre deux programmes sur deux machines)
 - ▶ FTP (File Transfert Protocol),
 - ▶ HTTP (HyperText Transfert Protocol)...
- ▶ l'adresse d'un serveur
- ▶ un chemin dans l'arborescence des fichiers

Forme générale : [protocole://adresse/chemin](#)

Plan

- ▶ Introduction aux langages à balises
 - ▶ Langage à balises
 - ▶ SGML
 - ▶ HTML
 - ▶ **XML**
- ▶ HTML
- ▶ CSS

XML, motivation

Objectif : représenter et échanger des données et des documents sur le web

- ▶ SGML

- ▶ un peu vieux
- ▶ trop complexe

- ▶ HTML

- ▶ trop basique : document = en-tête + corps
- ▶ mélange logique et présentation : balise `b` = bold (mise en gras)
- ▶ bonne approche :
 - ▶ `<important>Attention !</important>`
 - ▶ présenter la chaîne de caractères importante avec une mise en forme particulière (italique, rouge, gras, etc.)

Objectifs de XML

- ▶ XML doit être facilement utilisable sur le Web
- ▶ XML doit supporter une grande variété d'applications
- ▶ XML doit être compatible avec SGML
- ▶ Il doit être facile d'écrire des programmes qui traitent des documents XML
- ▶ Le nombre d'options doit être réduit au minimum, idéalement à zéro
- ▶ Les documents XML doivent être lisibles et raisonnablement clairs
- ▶ La conception de XML doit être menée rapidement
- ▶ La description de XML doit être formelle et concise
- ▶ Les documents XML doivent être faciles à créer
- ▶ La concision du balisage XML est d'une importance minime

XML est un métalangage

Permet de décrire des types de documents

- ▶ DTD, Schémas

Permet de définir des instances

- ▶ Documents XML
- ▶ Répondant à un type de document : Classique
- ▶ Simplement bien construits : Nouveau

Les instances peuvent décrire

- ▶ des documents (texte balisé) : Classique, cf. SGML
- ▶ des données structurées quelconques : Nouveau !

Principe général XML

DTDs, Schéma

- ▶ Comment décrire les données et les documents ?

Documents XML

- ▶ Les données et les documents eux-mêmes, dans des fichiers

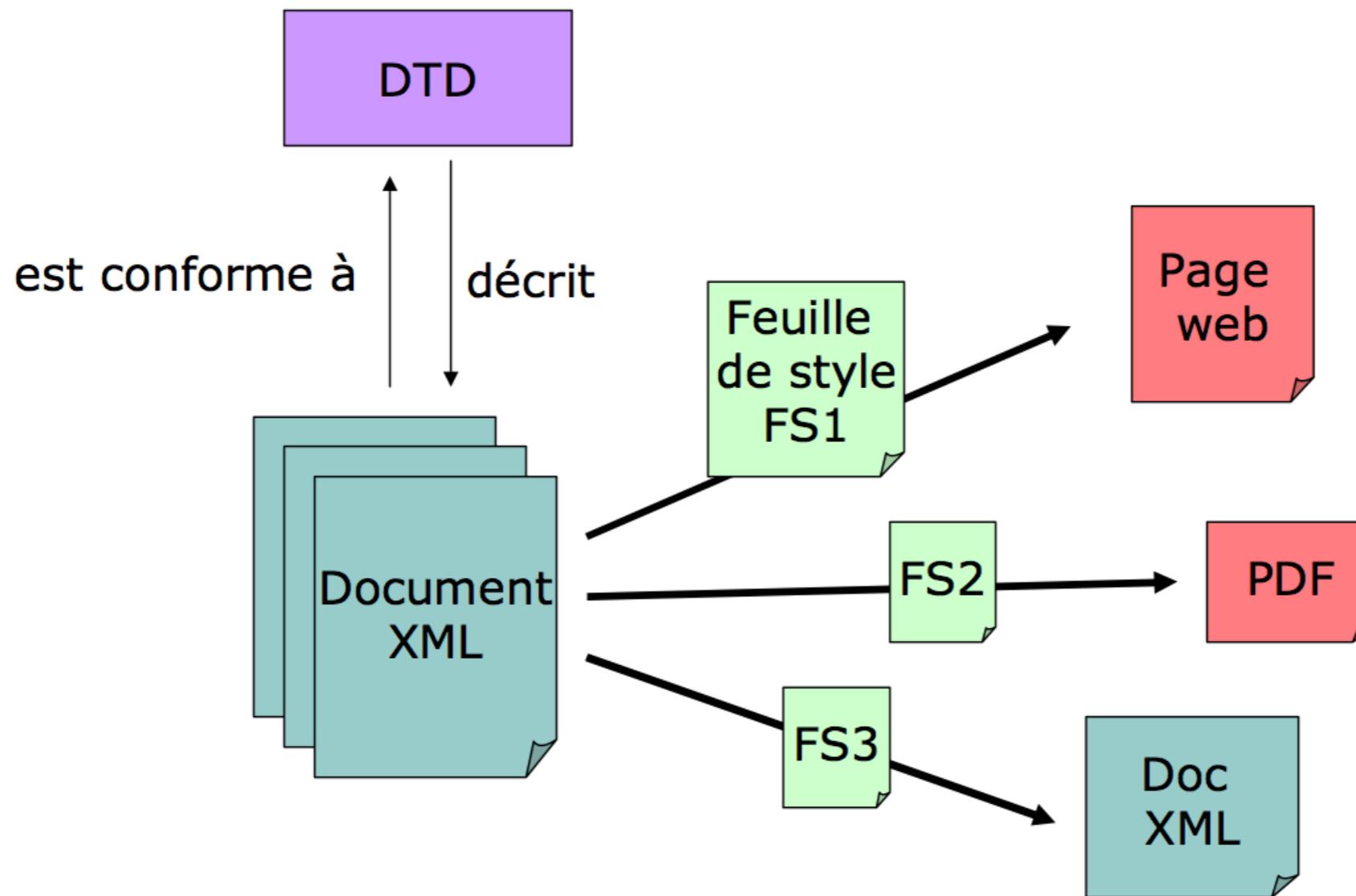
Feuilles de style

- ▶ Manière de présenter les données et les documents

Remarque

- ▶ On ne sait plus trop bien où sont les données, et les documents !

Schéma récapitulatif



Intégration de XML dans les SI

Stockage de données

- ▶ Simples fichiers (ex. configuration)
- ▶ Bases de données semi-structurées (requêtes, etc.)
- ▶ Bases de données documentaires
 - ▶ Documents XML, Documents XHTML (web)

Echange de données

- ▶ D'une base de données vers une autre (format d'échange)
- ▶ Serveur vers un navigateur : données + feuille de style

Remarque :

- ▶ Circulation de flux XML sur un réseau :
 - ▶ Utilisation de l'arbre entier (le document)
 - ▶ Utilisation à la volée de fragments pour les très gros documents