

WORLD WIDE WEB

Introduction / langage HTML

Lionel Médini

Licence professionnelle DIST



Objectifs de cette UE

- Découvrir ce qu'est le Web et comment ça marche
- Connaître (et pratiquer) les différentes technologies de base
- Découvrir quelques outils pour être rapidement opérationnels

Planning de l'UE

- Présentation du WWW
 - Historique
 - Fonctionnement
- Technologies côté client
 - HTML
 - CSS
- Programmation côté serveur
 - PHP
 - Lien avec une BD
- Frameworks côté serveur
 - CMS

Objectifs du cours et du TP

- Découvrir ce qu'est le Web et comment ça marche
- Connaître les différentes sources d'information
- Être capable
 - de créer une page Web simple
 - introduction au langage XHTML
 - notions de droit du Web
 - de personnaliser un navigateur
 - de rechercher des informations sur le Web
 - d'évaluer la fiabilité d'une source d'information

World Wide Web

- Principe original : accéder à des documents textuels
 - situés sur des machines accessibles par Internet
 - reliés entre eux par un mécanisme de lien « hypertexte »
 - Actuellement : servir des ressources
 - De différentes natures : texte, image, son, vidéo, contenu applicatif...
 - Hypermédia
 - Interactives
 - Permettant à l'utilisateur d'accéder à un service donné : rechercher de l'information, acheter un objet, accéder à ses mails, consulter ses comptes en banque...
- ➔ Nombreuses évolutions techniques

Rappel : Internet

- Un réseau de réseaux interconnectés (d'où le nom)
- Un ensemble de matériels, logiciels et protocoles (notamment IP)
- Un ensemble de **services**
 - **Application** qui utilise un **protocole** et un numéro de **port**
 - e-mail, transfert de fichiers, connexion à distance, WWW...
- Une somme « d'inventions » qui s'accumulent
 - Mécanismes réseau de base (TCP/IP)
 - Nommage et adressage des ressources (DNS, URL)
 - Outils et protocoles spécialisés
 - Langages d'échange d'informations standardisés (HTML, XML...)

Bref historique d'Internet

- **1959-1968 : Programme ARPA**
 - le ministère américain de la défense lance un réseau capable de supporter les conséquences d'un conflit nucléaire
 - **1969 : ARPANET, l'ancêtre d'Internet**
 - les universités américaines s'équipent de gros ordinateurs et se connectent au réseau ARPANET
 - **1970-1982 : Ouverture sur le monde**
 - premières connexions avec la Norvège et Londres
 - **1983 : Naissance d'Internet**
 - protocole TCP/IP : tous les réseaux s'interconnectent
 - les militaires quittent le navire
 - **1986 : Les autoroutes de l'Information**
 - la National Science Foundation déploie des super-ordinateurs pour augmenter le débit d'Internet
 - **1987-1992 : Les années d'expansion**
 - les fournisseurs d'accès apparaissent
 - les entreprises privées se connectent au réseau
 - **1993-2003 : L'explosion d'Internet**
 - ouverture au grand public
 - avènement du WEB et du courrier électronique
- } marché considérable

Genèse du Web : la notion d'hypertexte

- Principe
 - S'abstraire de l'aspect linéaire du document textuel
 - Mécanisme intellectuel permettant le cheminement d'une information à une autre → navigation, butinage, transclusion
- Historique
 - 1945 : invention de la notion d'hypertexte
 - Vannevar Bush, *As We may think*, *Atlantic Monthly*, 1945
 - 1965 : invention du terme d'hypertexte
 - Ted Nelson, projet Xanadu
 - Années 1960 : premier système hypertexte fonctionnel
 - NLS (oNLine System), Douglas Englebart
 - 1987-2004 : diffusion du logiciel HyperCard
 - Programme et environnement graphique de programmation, créé par Bill Atkinson pour Mac OS, livré avec les Mac
 - 1987 : première conférence HyperText
 - Sponsorisée par l'ACM

Naissance du Web (1989-1991)

- Mars 89 : projet de création d'un hypertexte documentaire distribué sur le réseau du CERN
 - Origine : Tim Berners-Lee, puis Robert Cailliau (1990)
 - Choix des technologies TCP/IP et ouverture de la première connexion du CERN avec Internet
 - Mise au point des **3 technologies de base du Web : URL, HTML et HTTP**
- Septembre 90 : 1^{er} site Web fonctionnel (mode texte)
 - 1^{er} serveur Web : nxoc01.cern.ch
 - 1^{er} navigateur Web : WorlWideWeb (rebaptisé plus tard Nexus), développé en Objective C
- Août 1991 : publication du projet WorldWideWeb dans un message sur UseNet
- Décembre 91 : démonstration publique à la conférence Hypertext'91 (San Antonio)



Le premier serveur Web,
un NeXT Cube
(source : Wikipédia)

Historique du Web

- 1993 : Mosaic : premier navigateur « grand public »
 - Marc Andreessen, NCSA : plateformes X, puis Mac et Windows
 - affichage d'images (GIF et XBM) dans les pages Web
 - prise en charge de formulaires interactifs
- 1994 : création du W3C
 - à l'initiative du CERN (Genève) et du MIT (Boston)
 - président : Tim Berners-Lee
 - but : standardisation et développement du Web
- 1994 : Apparition des navigateurs privés
 - M. Andreessen crée Netscape Communications Corp.
- 1995 : Microsoft lance la « guerre des navigateurs »
 - Apparition d'Internet Explorer pour Windows 95
- 1995 : Altavista : premier « gros » moteur de recherche
- 1996 : Navigateur Opera
- 1998 : Apparition de Google
- 2003 : Apple lance Safari
- 2004 : Première version de Mozilla Firefox
- 2004 : première conférence « Web 2.0 »
- 2008 : Google lance Google Chrome
- Actuellement : explosion du Web mobile
- En cours de développement : Web sémantique, de données, des objets...

Nombre de sites référencés

1990 : 1 (CERN)

1991 : premier site hors d'Europe (SLAC, Stanford)

1992 : 26

Juin 1993 : 130

Juin 1994 : 2738

Juin 1995 : 23 500

Janvier 1996 : ~ 100 000

Avril 1997 : > 1 million

Février 2000 : > 11 millions

Février 2007 : > 100 millions

Février 2009 : ~ 216 millions



Aspects techniques du Web

- Les 3 mécanismes de base du Web
 - URL
 - Le Web permet d'accéder à un ensemble de **ressources**
 - Le mécanisme de localisation peut faire appel au protocole DNS
 - HTTP
 - Protocole de niveau applicatif
 - Paradigme client-serveur
 - Protocole sans état (pas de « mémoire » des transactions précédentes)
 - HTML
 - Langage de description de « pages Web »
 - Texte, images et autres objets
 - Liens hypermédias entre les pages
 - Programmation déclarative

Aspects techniques du Web

- Les forces en présence
 - Côté client : l'utilisateur utilise un navigateur
 - Client HTTP
 - Interprète les pages Web et les affiche à l'utilisateur
 - Peut effectuer des traitements plus complexes (plugins, moteur de scripts...)
 - Côté serveur : le Webmaster gère un serveur Web
 - Attend les requêtes HTTP et y répond
 - En renvoyant des ressources dont il dispose
 - En interrogeant plusieurs modules (sécurité, scripting, redirection...)
 - En interrogeant d'autres outils pour les traitements complexes
 - On parle aussi de serveur Web pour désigner la machine qui héberge le programme serveur (abus de langage)

Le Web : historique rapide

- 1946 : invention de la notion d'hypertexte
- 1970 : développement de l'Internet
- 1989-91 : application de la notion d'hypertexte à Internet (CERN, Tim Berners-Lee)
- 1993 : diffusion universitaire (navigateur Xmosaic, 50 serveurs dans le monde)
- 1993 : création du W3C pour normaliser le Web
- 1994-95 : premiers navigateurs privés (Netscape puis IE)
- 1998 : 2,2 millions de sites
- 2000 : 20 millions de sites (2,5 milliards de pages)
- 2002 : 3 / 550 milliards de pages (web de surface / web profond)
- 2004 : première conférence « Web 2.0 »
- 2005 : Google prétend indexer 8 milliards de pages

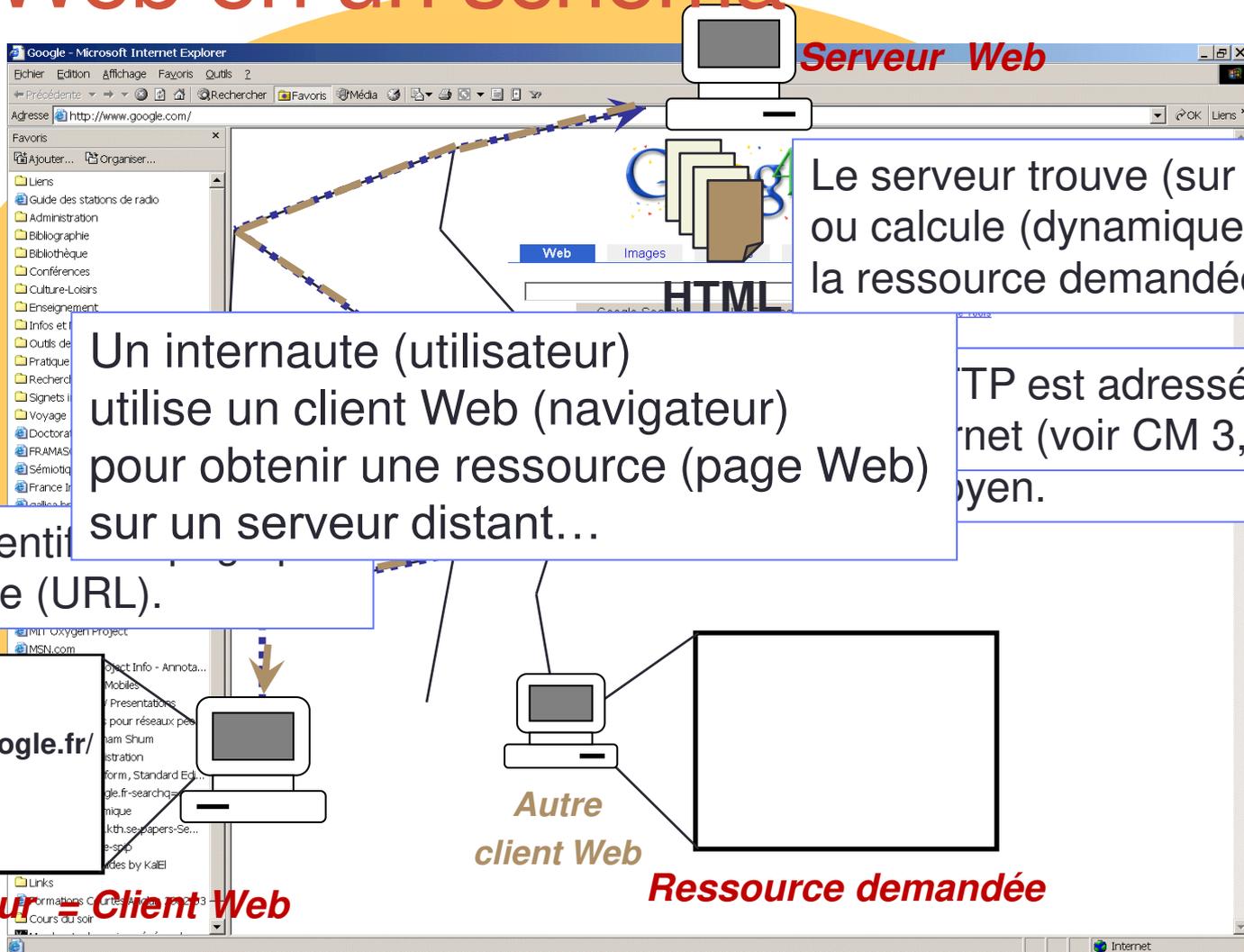
Le Web : première définition

- World Wide Web (WWW, Toile)
- Définition
 - ensemble de données disponibles sur les serveurs accessibles sur le réseau Internet
 - pouvant être visualisées et/ou utilisées avec un navigateur Web
- Attention : Web \neq Internet
 - Internet = support de communication général
 - Web = une partie des contenus circulant sur l'Internet

- 1. Introduction
- 2. Aspects techniques
- 3. Page Web
- 4. Conclusion

- 1. Objectifs du cours
- 2. Principes généraux

Le Web en un schéma



Serveur Web

Le serveur trouve (sur un disque) ou calcule (dynamiquement) la ressource demandée.

Un internaute (utilisateur) utilise un client Web (navigateur) pour obtenir une ressource (page Web) sur un serveur distant...

Le client identifie son adresse (URL).

TP est adressée au net (voir CM 3, routage).
oyen.

<http://www.google.fr/>

Autre client Web

Ressource demandée

Navigateur = Client Web

URL (Uniform Resource Locator)

- Permet d'identifier une **ressource** sur le réseau, c'est-à-dire :
 - une page Web
 - une image (seule ou utilisée dans une page Web)
 - un programme
 - un fichier à télécharger...
- Indique
 - un **protocole** (langage de communication entre deux programmes sur deux machines).
 - Exemple: FTP (File Transfert Protocol), HTTP (HyperText Transfert Protocol)...
 - une adresse et un chemin
 - forme générale : `protocole://adresse`
 - exemple : `http://www.univ-lyon1.fr/`

Différents types d'URL

- Forme principale (protocole *HTTP*)

`http://pci.univ-lyon1.fr/TP/sujets-TP.pdf`

Protocole *Adresse machine* *Chemin fichier*

- Forme pour désigner les fichiers *locaux*

- Chemin relatif : `fichier.html` ou `dossier/toto.html`
- Sur un disque : `file://C:/chemin/fichier.htm`

- Forme pour le *transfert* de fichiers

- `ftp://ftp.inria.fr/INRIA/tech-reports/RR-5645.pdf`

- Forme pour l'envoi de *courrier* électronique

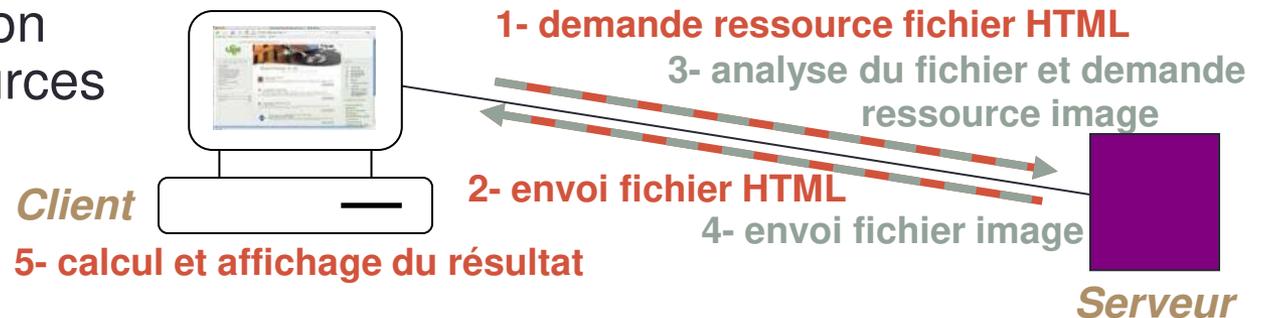
- `mailto:jean.durand@univ-lyon1.fr`

DEMO...

HTTP (HyperText Transfer Protocol)

- Rappels du CM 3
 - Protocole de type client-serveur
 - Protocole de niveau applicatif
- Mécanisme de requête / réponse HTTP
 - **Requête** du client : URL de la ressource
 - **Réponse** du serveur : code de statut / d'erreur + ressource (si « OK »)
 - tout s'est bien passé : « OK » (code de statut : 200)
 - pas le droit d'accéder : « forbidden » (code d'erreur : 403)
 - la ressource n'existe pas : « not found » : (code d'erreur : 404)

- Exemple de transaction avec plusieurs ressources



- Remarque : le protocole **HTTPS** (sécurisé) fonctionne de la même façon

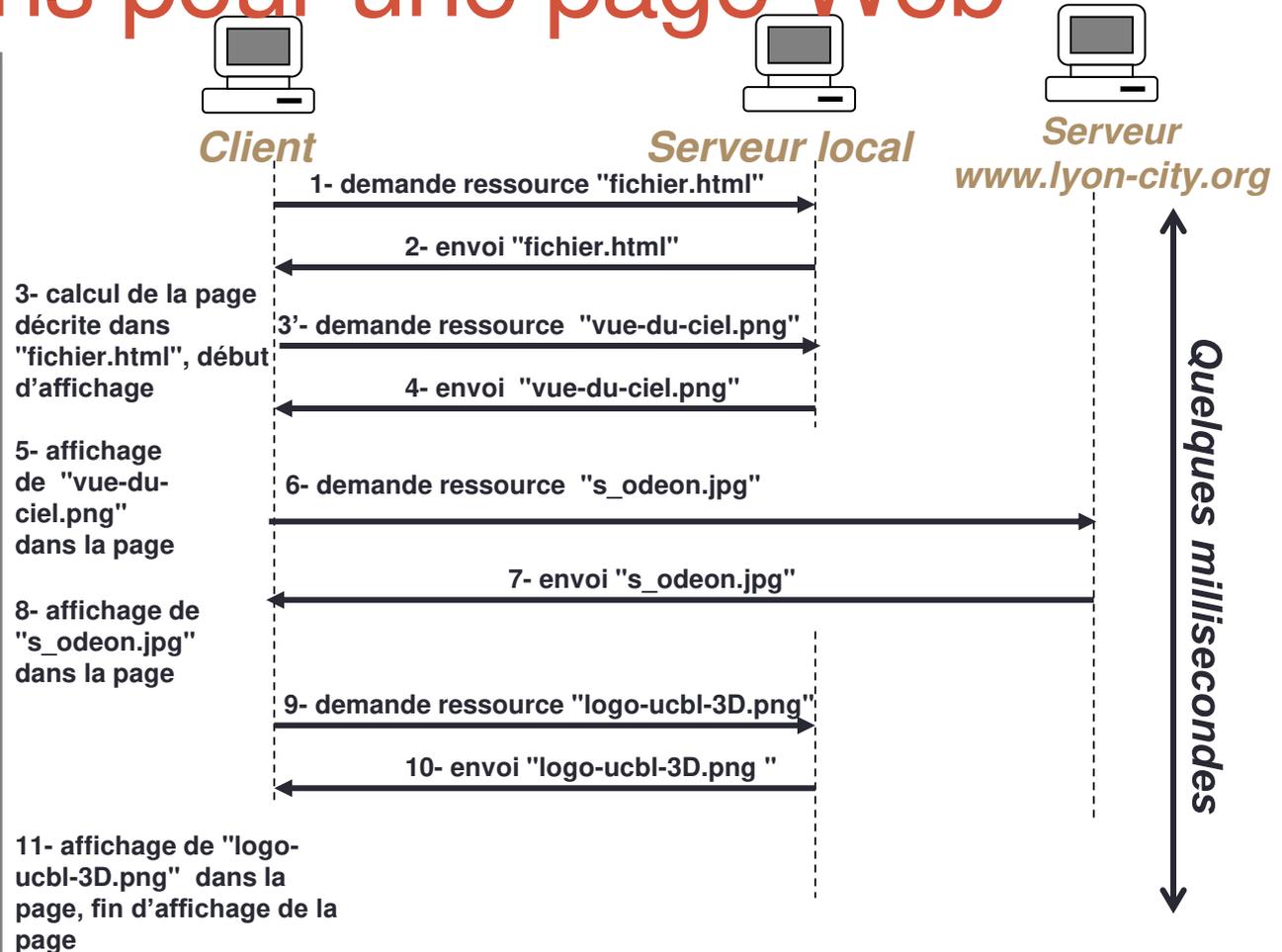
1. Introduction
2. Aspects techniques
3. Page Web
4. Conclusion

1. Ressource et URL
2. Protocole HTTP
3. Serveur Web
4. Client Web

Transactions pour une page Web

fichier.html

```
<html>
<head>
<title>INSERTION
  IMAGES</title>
</head>
<body bgcolor="#ffffff">
<p>Voici une photo aérienne de
  Lyon (chemin relatif) : <img
  src= "vue-du-ciel.png"
  alt="vue du ciel" />.</p>
<p>Voici une autre photo (URL
  distante) : .</p>
<p>Cliquer sur le logo <a
  href="http://www.univ-
  lyon1.fr/"></a> pour
  visiter le site Web de Lyon
  1.</p>
</body>
</html>
```



Vitesse d'une transaction HTTP

- Temps d'affichage d'une page Web lié
 - à la complexité de la page à afficher
 - nombre et volume des ressources composant la page
 - traitements éventuellement nécessaires pour générer dynamiquement ces ressources
 - au débit et à l'encombrement des réseaux parcourus
 - à la possibilité de présenter une page avant que toutes les données soient arrivées
(ex. texte avant images, chargement asynchrone...)
 - à la charge du serveur
 - réponse à de nombreux clients en même temps
 - cryptage ou non des données (HTTP / HTTPS)
 - autres applications tournant sur la même machine
 - à la vitesse de la machine cliente

Serveur Web

□ Définition

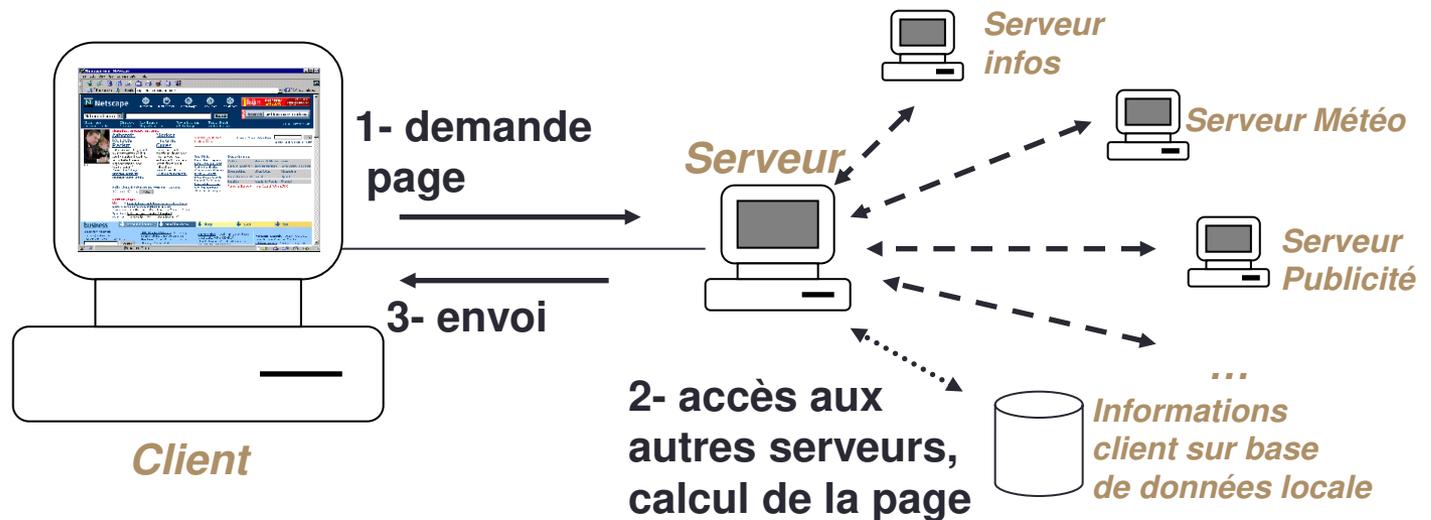
- **Application** dont le rôle est de fournir des ressources Web aux clients qui les demandent
 - par le protocole HTTP(S)
 - sur Internet ou sur un réseau local
- Par extension, **machine** sur laquelle s'exécute cette application

□ Deux façons de fournir des ressources

- statique : à partir de fichiers existants sur le serveur
- dynamique : en générant la ressource au moment de la requête

Pages générées dynamiquement

- Le serveur doit calculer la ressource avant de la renvoyer
 - en fonction des données
 - indiquées par le client (formulaires, cookies)
 - présentes sur le serveur (fichiers de données, pages statiques)
 - présentes sur d'autres serveurs (web ou bases de données) : le serveur considéré est le **client** d'autres serveurs
 - avec un langage de programmation côté serveur (PHP, ASP, Java, Python...)
- Le mécanisme de requête / réponse reste le même
- Exemple



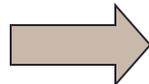
Navigateur Web (browser, butineur)

- Caractéristiques
 - logiciel client pour les protocoles HTTP et HTTPS
 - demande des fichiers (X)HTML à un serveur
 - interprète ces fichiers
 - demande éventuellement d'autres ressources
 - les présente à l'utilisateur
 - peut aussi réaliser quelques traitements (scripts)
- Principaux navigateurs
 - Mozilla Firefox, Chrome, Safari, Internet Explorer, Opera

Fichier HTML

```
<html>
<head>
<title>Université Claude Bernard Lyon 1</title>
<meta http-equiv="Content-Type" content="text/html; charset=iso-8859-1">
</head>
</html>
```

Description du document



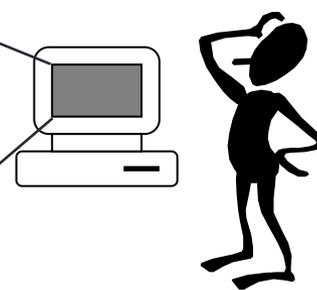
Calcul

Navigateur



Présentation à l'utilisateur

Utilisateur



Différents médias accessibles sur le Web

- Formes du point de vue de l'utilisateur
 - texte
 - image (photo, dessin, graphique)
 - son (musique, voix)
 - vidéo (images animées + son)
- Formats de fichiers correspondants
 - texte simple : ASCII (.txt)
 - images : JPEG (.jpg), GIF (.gif)
 - sons : WAVE (.wav), MPEG3 (.mp3)
 - vidéo : MPEG (.mpg), AVI (.avi), Quicktime (.qt)...
- Applications permettant de « lire » les fichiers (ouvrir les documents)
 - éditeur de texte
 - visualisateur d'images
 - lecteur de son, de vidéo

➔ Question : comment un navigateur sait-il quoi faire d'un type d'une ressource récupérée sur un serveur ?

Document multimédia

- Combinaison de différents médias
 - texte, image, son, vidéo...
- Deux caractéristiques principales
 - « **ça bouge** » : le document se déroule temporellement (vidéo, son, objets animés)
 - « **c'est interactif** » : on peut cliquer sur un bouton, suivre un lien hypermédia, remplir une zone texte...
- Exemples d'utilisations
 - cours, jeux interactifs, logiciels éducatifs, encyclopédies interactives, bornes dans les lieux publics, pages Web, films DVD...

Page Web

- Une page Web, c'est
 - un document multimédia
 - images, textes, possibilité d'interaction, liens
 - décrit élément par élément
 - titre, morceaux de texte, images...
 - avec un langage de description
 - HTML (1992), XHTML (2002) ou HTML5 (?)
 - stocké dans un fichier
 - .html (ou .htm, .xhtml)
- Une page Web
 - est **calculée** et affichée par un navigateur
 - est localisée sur Internet à l'aide d'une adresse (URL)
 - permet d'accéder à d'autres pages en suivant des liens

Site Web

- Regroupement de pages Web
 - autour d'une thématique commune
 - reliées entre elles par des liens hypertextes
 - émanant d'une même entité (organisation, entreprise, particulier...)
 - accessibles à partir d'une page d'accueil commune
 - accessibles à partir d'une URL de base commune
 - en général, localisées sur un même serveur
- Organisation hiérarchique
 - structure sous forme d'arborescence de dossiers et de fichiers (comme un disque local)
 - permet de définir des chemins relatifs
 - **exceptions** (de plus en plus fréquentes) : pages générées dynamiquement
- Liens vers et à partir d'autres sites
 - permettent de naviguer d'un site à un autre
 - composent la « toile » mondiale (World Wide Web)

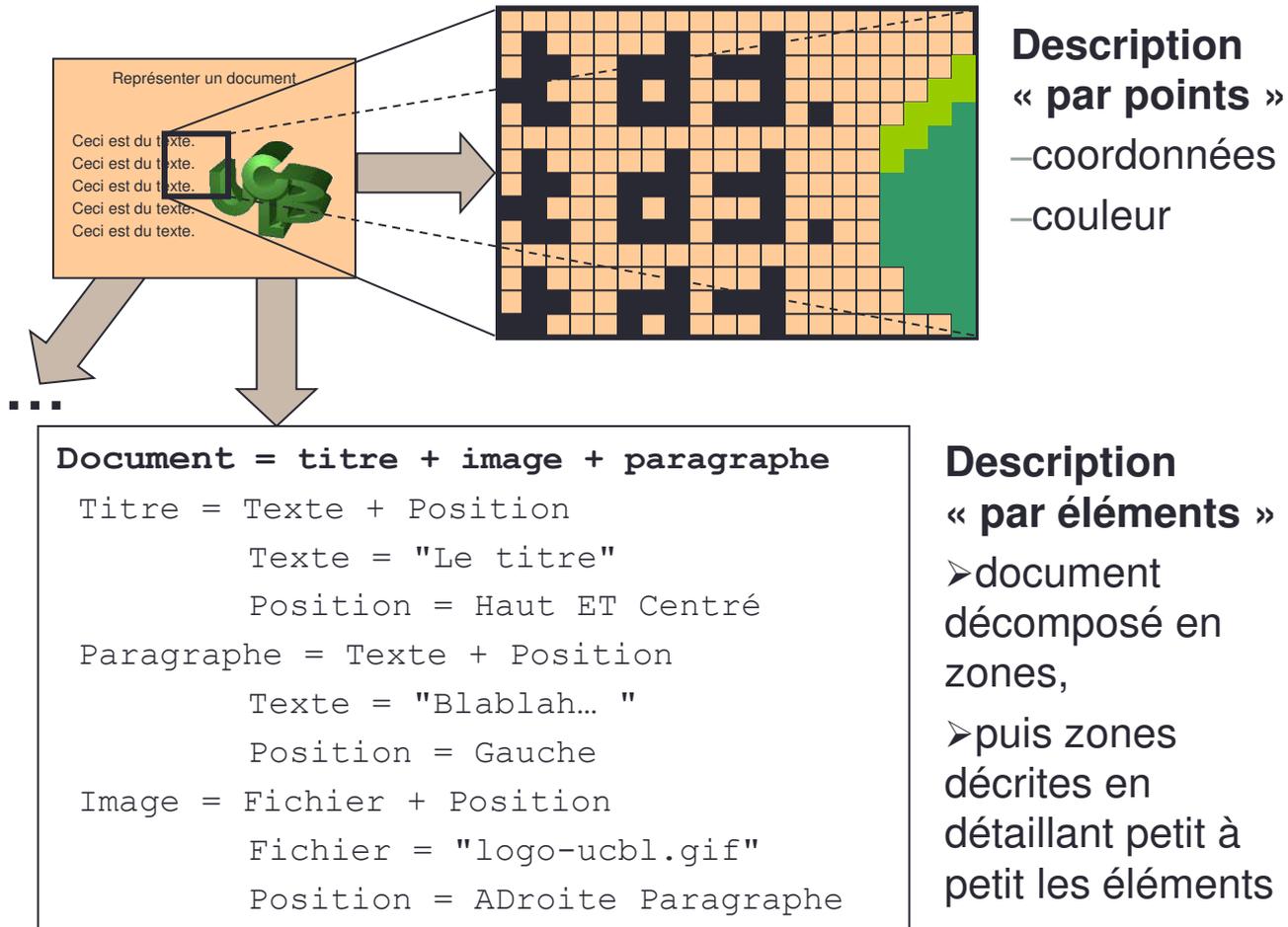
Portail Web

- Principe
 - proposer un ensemble de **services** pour une **communauté** sur un site Web
 - exemples de communauté
 - générale (Yahoo!)
 - liée à un fournisseur d'accès Internet (Infonie, Voilà)
 - thématique (santé, enfants)...
 - exemples de services : moteur de recherche, annuaires, météo, informations, jeux, publicité, accès courrier électronique, forums, discussion...
- Modèle économique
 - fournir un nombre de services suffisants pour que le portail devienne l'accès principal au Web de certains utilisateurs
 - proposer des publicités ciblées
- Nécessite une appréciation critique pour ne pas rester bloqué dans la sous-partie du Web accessible depuis le portail

Le « Web 2.0 »

- Principe du Web « classique »
 - ce sont les organisations qui détiennent des sites qui décident de l'information qui y figure
 - ➔ Communication de type « one-to-many » (= diffusion)
- Principe du « Web 2.0 »
 - donner le contrôle de l'information aux utilisateurs
 - faire émerger des « réseaux sociaux »
 - ➔ chacun peut déposer des contenus
 - pour donner son avis sur un sujet donné (blogs)
 - pour partager ses documents (images, vidéos...)
 - pour étiqueter (« tagger ») des contenus existants
- Aspects techniques
 - outils (protocoles, clients, serveurs) identiques à ceux du Web classique
 - nécessite plus de puissance de calcul (pages dynamiques)
 - nécessite plus d'espace de stockage (contenus envoyés par les utilisateurs)
- Exemples de sites « Web 2.0 »
 - Wikipédia, Del.icio.us, Technorati, Flickr, Picasa Web album, Dailymotion, YouTube, Kartoo...

Plusieurs façons de décrire un document



Document comme composition d'éléments

- Un **élément**
 - est une partie significative d'un document
 - TITRE, AUTEUR, PARAGRAPHE, IMAGE, LIEN, TABLEAU...
 - peut être composé d'autres éléments
 - un DOCUMENT est composé d'un TITRE et d'une suite de PARAGRAPHE
 - un TABLEAU est composé de LIGNES
 - une LIGNE est composé de CASEs...
 - peut avoir des attributs
 - un attribut "nom du fichier" pour un élément IMAGE
 - un attribut "URL de destination" pour un élément LIEN
- Pour les pages Web
 - \approx 40 éléments de description dans les langages (X)HTML
 - (PCI : quelques éléments de base, syntaxe XHTML)

HTML (HyperText Markup Language)

- HTML permet de décrire des pages Web dans des fichiers textuels (ASCII) en utilisant des balises
- Balises
 - mots-clés simples pour délimiter des descriptions d'éléments (`title`, `img`, `p`...)
 - balises ouvrantes (`<title>`), fermantes (`</title>`), vides (`<hr />`)
 - balises avec **attributs** (``)
- Un élément
 - est soit délimité par deux balises...
 - entre lesquelles se situe son contenu : `<p>Un paragraphe...</p>`
 - dans de cas, il peut contenir d'autres éléments (et ainsi de suite)
`<p>Un paragraphe avec un mot en gras (bold).</p>`
(= l'élément **p** contient l'élément **b**)
 - ...soit entièrement décrit par une balise vide : `
`
 - avec éventuellement des attributs : ``

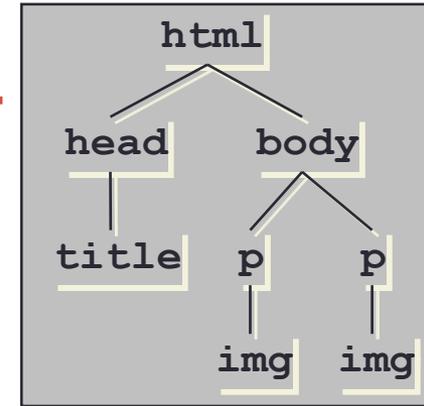
Exemple de structuration en HTML

Déclaration de type

de document → `<!doctype html>`

Balise ouvrante

```
<html>
  <head>
    <title>Merveilles du monde</title>
  </head>
```



Arborescence

Élément avec contenu

```
<body>
  <p>Dans l'image ci-après
    
    nous présentons un zèbre vivant
    dans la savane.
  </p>
```

Attributs de l'élément Image

Balise d'élément vide
(une seule balise)

```
<p>
  
  Le lion n'est pas vraiment son ami
</p>
```

Balise fermante

```
</p>
</body>
</html>
```

Les différents éléments HTML

- Catégories d'éléments

- Structuration d'un document
 - Titres, paragraphes, tableaux, menus, liens...
- Types de contenus
 - Textes, images, scripts, audio, vidéo...
- Métadonnées
 - Auteur, encodage...

- Liste des éléments

- http://liris.cnrs.fr/~lmedini/enseignement/PANIST1/CM/PANIST1_CM4-5.pdf

- Pour tester

- <http://w3schools.com/html/>

DEMO...

Règle à respecter en (X)HTML

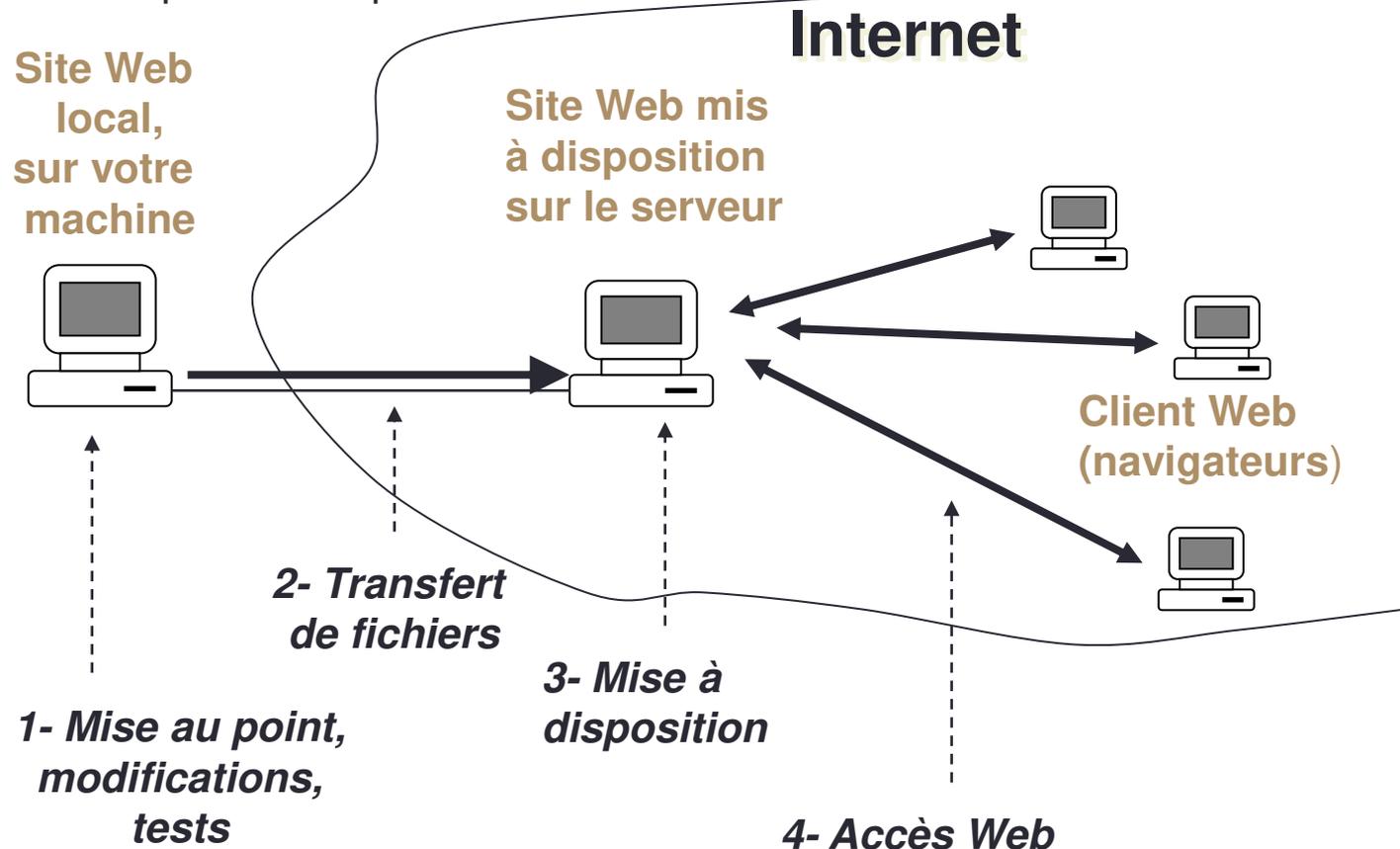
- Deux types d'éléments
 - Blocs
 - Peuvent être positionnés sur l'écran
 - Peuvent contenir d'autres éléments de bloc ou des élément en-ligne
 - L'élément body ne peut contenir que des balises de bloc
 - Exemples : h1-h6, p, div...
 - En-ligne
 - Suivent le flot documentaire
 - Ne peuvent contenir que des élément en-ligne
 - Exemples : a, span, img, object, input...
- ➔ On ne peut pas directement positionner une image sur l'écran ; il faut la mettre dans un div.

Plusieurs situations type

- Cas d'un particulier
 - site hébergé par un Fournisseur d'Accès Internet
 - conception en local à la maison
 - transfert des fichiers chez le FAI par FTP
 - site sur une machine à la maison
 - conception en local sur une partie du disque
 - copie de fichiers sur le site publié
- Cas d'une entreprise
 - site externalisé (conception, mise à disposition)
 - site conçu en interne et mis à disposition sur un serveur de l'entreprise

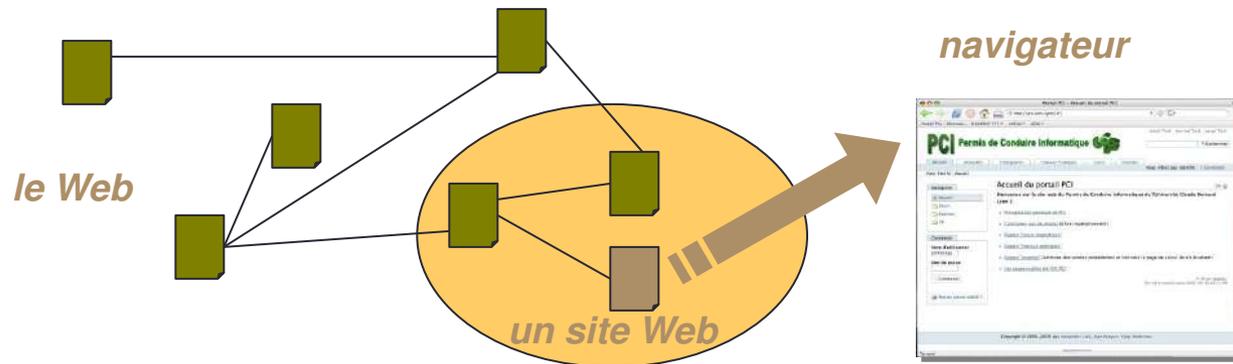
Conception, transfert et accès

- Dans tous les cas
 - Cycle de conception et de publication d'un site Web



1. Introduction
2. Aspects techniques
3. Page Web
4. Conclusion

Conclusion : une description plus formelle



- World Wide Web :
 - Ensemble de ressources interreliées formant un **réseau** (ou une toile) sur lequel il est possible de naviguer
- Navigateur
 - Outil permettant d'accéder à ces ressources et d'exploiter les liens hypermédias pour la navigation

Remerciements

- Sources de ces transparents
 - PCI (Pratiques et Compétences Informatiques, UCBL, 2006)
 - MIF13 (M1 informatique, UCBL, 2013)
- Images
 - deviantART : <http://e-designer.deviantart.com/>