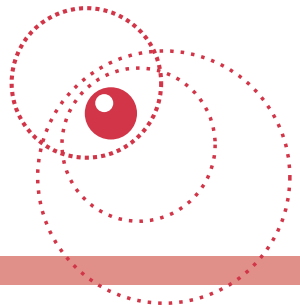


Ingénierie du Web (INF347)

Introduction aux technologies du Web

14 mai 2012

Licence de droits d'usage

Pierre Senellart

TELECOM ParisTech

Plan

Internet et le Web
- Généralités
- Le Web, un marché

Le langage HTML

CSS

Licence de droits d'usage

Plan

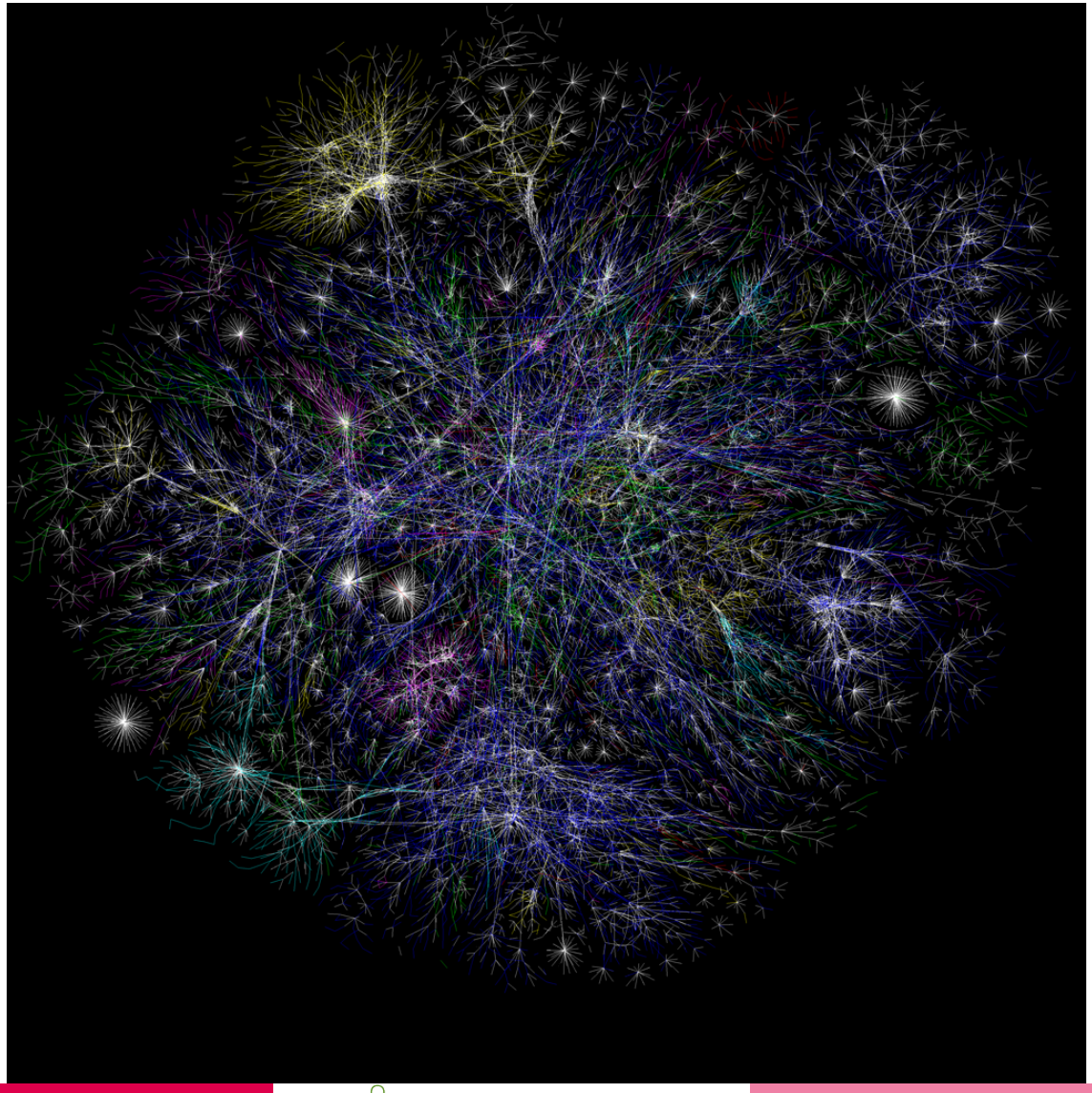
Internet et le Web
Généralités
Le Web, un marché

Le langage HTML

CSS

Licence de droits d'usage

Un réseau de réseaux : ordinateurs interconnectés



http://www.opte.org/

Licence de droits d'usage

La pile de protocoles Internet

Une pile de protocoles de communication, l'un au-dessus de l'autre

Application	HTTP, FTP, SMTP, DNS	
Transport	TCP, UDP, ICMP	(sessions, fiabilité. . .)
Réseau	IP (v4, v6)	(routage, adressage)
Lien	Ethernet, 802.11 (ARP)	(adressage local)
Physique	Ethernet, 802.11 (physique)	

Licence de droits d'usage

IP (Internet Protocol)

- Adressage de machines et routage sur Internet
- Deux versions du protocole IP utilisées sur Internet : IPv4 (très répandu) et IPv6 (support encore un peu expérimental)
- IPv4 : Adresses de 4 octets affectées à chaque ordinateur, p. ex., 137.194.2.24. Des gammes de telles adresses sont données aux diverses institutions (fournisseurs d'accès à Internet, universités, etc.), pour les affecter comme elles le souhaitent.
- Problème : seulement 2^{32} adresses possibles (un grand nombre d'entre elles ne peut pas être données à de nouveaux hôtes, pour diverses raisons). Cela signifie que beaucoup de machines connectées à Internet n'ont pas d'adresse IPv4 et une forme de traduction d'adresse (network address translation, NAT) est nécessaire.
- IPv6 : Adresses de 16 octets ; espace d'adressage beaucoup plus grand ! Les adresses ont la forme suivante : 2001:660:330f:2::18. Chacun des appareils connecté à l'Internet peut avoir sa propre adresse IP !

Licence de droits d'usage

TCP (Transmission Control Protocol)

- L'un des deux protocoles de transport utilisés sur IP, avec UDP (User Datagram Protocol)
- Contrairement à UDP, fournit une transmission fiable des données (acknowledgments)
- Les données sont divisées en petits datagrammes qui sont envoyés sur le réseau, et éventuellement réordonnés à destination
- Comme UDP, chaque transmission TCP indique un numéro de port source et destination (entre 0 and 65535) pour la distinguer du reste du trafic
- Un client sélectionne usuellement un numéro de port aléatoire pour établir une connexion à un numéro de port fixe du serveur
- Le numéro de port du serveur identifie de manière conventionnelle un protocole d'application au-dessus de : 22 pour SSH, 25 pour SMTP, 80 pour HTTP, 110 pour POP3...

Licence de droits d'usage

DNS (Domain Name System)

- Les adresses IPv4 sont difficiles à mémoriser et un service donné (p. ex., un site Web) peut changer d'adresses IP (p. ex., nouveau fournisseur d'accès à Internet)
- Problème accentué pour les adresses IPv6 !
- DNS : un protocole basé sur UDP/IP pour associer des noms lisibles par les humains (p. ex., `www.google.com`, `weather.yahoo.com`) à des adresses IP
- Noms de domaines hiérarchiques : `com` est un domaine de niveau principal (top-level domain, TLD), `yahoo.com` un sous-domaine...
- Résolution hiérarchique de nom de domaine : serveurs racines avec des IP fixes savent qui est responsable des TLDs, les serveurs en charge d'un domaine savent qui est responsable d'un sous-domaine, etc.
- Rien de magique dans `www.google.com` : juste un sous-domaine de `google.com`.

Licence de droits d'usage

Généralités

- Qu'est-ce que le Web (ou World Wide Web, Toile, WWW, W3) ?
  - Système hypertexte public : système contenant des documents liés entre eux par des hyperliens permettant de passer automatiquement d'un document à l'autre.
- Différence entre le Web et Internet ?
  - Internet : réseau mondial d'ordinateurs permettant aux utilisateurs de communiquer (courrier électronique), de publier des informations (Web), de transférer des données (FTP), de travailler à distance (telnet et ssh)...
  - Web : un aspect d'Internet.

Licence de droits d'usage

Un bref historique du Web

1969 ARPANET (ancêtre d'Internet)

1974 TCP (Vinton G. Cerf & Robert E. Kahn, Turing award 2004)

1990 World Wide Web, HTTP, HTML (Tim Berners-Lee, Robert Cailliau)

1993 Mosaic (premier navigateur graphique à succès, ancêtre de Netscape)

1994 Yahoo ! (David Filo, Jerry Yang)

1994 Fondation du W3C

1995 Amazon.com, Ebay

1995 Internet Explorer

1995 AltaVista (Louis Monier, Michael Burrows)

1998 Google (Larry Page, Sergey Brin)

2001 Wikipedia (Jimmy Wales)

2004 Mozilla Firefox

2005 YouTube

Licence de droits d'usage

Architecture client-serveur

Le client (navigateur : Internet Explorer, Firefox, Safari...)

- demande au serveur des informations
- affiche des pages pour l'utilisateur

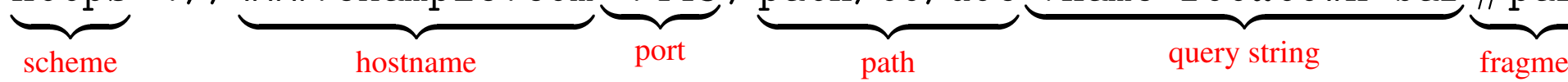
Le serveur (Apache, Microsoft IIS...)

- reçoit en permanence les requêtes des clients
- renvoie les documents correspondants

Le protocole de communication est HTTP (ou HTTPS).

Licence de droits d'usage

URL (Uniform Resource Locator)



scheme : manière dont on accède à la ressource ; généralement http or https

hostname : nom de domaine du serveur (cf. DNS)

port : port TCP ; défaut : 80 pour `http` et 443 pour `https`

path : chemin logique du document

query string : paramètres additionnels optionnels (documents dynamiques)

fragment : sous-partie optionnelle du document

URLs relatives à un contexte (p. ex., l'URL ci-dessus)

```
/titi   https ://www.example.com/titi
tata    https ://www.example.com/path/to/tata
```


Licence de droits d'usage

Le Web : un mélange de technologies

- Pour le contenu : HTML/XHTML, mais aussi PDF, documents Word, fichers texte, XML (RSS, SVG, MathML, etc.)...
- Pour présenter ce contenu : CSS, XSLT
- Pour animer ce contenu : JavaScript, AJAX, VBScript...
- Pour du contenu interactif riche : Flash, Java, Sliverlight, ActiveX, API `<canvas>` ...
- Contenu multimédia : images, sons, vidéos...
- Et côté serveur : n'importe quel langage de programmation (p. ex., PHP, JSP, servlets Java, ASP, ColdFusion, etc.) et technologie de base de données (p. ex., MySQL, Oracle) pour servir ce contenu

Devient complexe à gérer ! Un développeur Web de nos jours a besoin de maîtriser grand nombre de technologies Web ; similairement, un client Web a besoin de comprendre beaucoup de technologies différentes.

Licence de droits d'usage

Plan

Internet et le Web
Généralités
Le Web, un marché

Le langage HTML

CSS

Licence de droits d'usage

Clients Web

- Navigateurs graphiques (cf. transparent suivant)
- Navigateurs textuels : w3m, lynx, links (libres, Windows, Mac OS, Linux, Unix) ; leur usage n'est guère plus répandu, mais ils simulent assez bien ce que « voit » un robot ou un navigateur auditif
- Autres navigateurs : navigateurs auditifs, etc.
- Mais aussi : robots des moteurs de recherche, logiciels de traduction automatique. . .

Licence de droits d'usage

Clients Web

- Navigateurs graphiques (cf. transparent suivant)
- Navigateurs textuels : w3m, lynx, links (libres, Windows, Mac OS, Linux, Unix) ; leur usage n'est guère plus répandu, mais ils simulent assez bien ce que « voit » un robot ou un navigateur auditif
- Autres navigateurs : navigateurs auditifs, etc.
- Mais aussi : robots des moteurs de recherche, logiciels de traduction automatique. . .

Une très grande variété de clients ! En théorie, tout site Web doit être testé avec la plupart de ceux-ci (et au moins les plus répandus), dans leurs différentes versions.

Licence de droits d'usage

Navigateurs graphiques

Navigateur	Moteur	Part	Distribution
Internet Explorer	Trident	35%	distribué avec Windows
Chrome + Android	WebKit	30%	libre, Windows, Mac OS, Linux, Android
Firefox	Gecko	22%	libre, Windows, Mac OS, Linux, Unix
Safari	WebKit	9%	distribué avec Mac OS, iOS
Opera	Presto	4%	gratuit, Windows, Mac OS, Linux, mobile

Parts de marché : diverses sources, chiffres précis difficiles à obtenir. Internet Explorer est beaucoup plus répandu aux États-Unis qu'en Europe. Toujours dominant, mais en très nette baisse depuis plusieurs années.

Trident est (traditionnellement) le moteur de rendu supportant le moins bien les standards du Web. En évolution.

Licence de droits d'usage

Actualité des navigateurs graphiques

- Impressionnant succès de Google Chrome, devenu le deuxième navigateur le plus utilisé, seulement trois ans après sa sortie initiale.
- Internet Explorer 6, 7, 8, 9 tous encore utilisés (IE6 est le navigateur par défaut de Windows XP) ; les parts d'utilisations des anciennes versions de Firefox, Chrome, Safari sont relativement négligeables. Internet Explorer 10 sortira avec Windows 8.

Licence de droits d'usage

Serveurs Web

Server	Part	Distribution
Apache	65%	libre, Windows, Mac OS, Linux, Unix
Microsoft IIS	15%	avec certaines versions de Windows
nginx	10%	libre, Windows, Mac OS, Linux, Unix
lighthttpd	1%	libre, Windows, Mac OS, Linux, Unix

- Part de marché : selon des études par Opera et Netcraft, chiffres précis ayant peu de signification.
- Beaucoup de grandes entreprises de logiciel ont soit leur propre logiciel, soit leur version modifiée de Apache (p. ex., GFE/GWS pour Google).
- nginx et lighthttpd sont plus légers (moins riches en fonctionnalités, mais plus rapides en général) que Apache.
- Les versions de Microsoft IIS fournies avec les versions classiques de Windows sont très limitées.

Licence de droits d'usage

Moteurs de recherche

- Un grand nombre de moteurs de recherche différents, avec des parts de marché changeant beaucoup d'un pays à l'autre.
- Au niveau mondial :

  Google dominant (autour de 80%)
  Yahoo !/Bing (peut-être 10% du marché global)
- Dans certains pays, des moteurs de recherche locaux dominent (Baidu en Chine, Naver en Corée. . .)

Licence de droits d'usage

Yahoo ! et Bing

En juillet 2009, Microsoft et Yahoo ! annoncent un accord majeur :

- Yahoo ! arrête de développer son propre moteur de recherche (lancé en 2003, après les rachats de Inktomi et Altavista) et utilise Bing à la place ;
- Yahoo ! fournit les services de publicités de Bing.

Ne concerne pas Yahoo ! Japan, qui utilise au contraire Google comme moteur.

Licence de droits d'usage

Plan

Internet et le Web

Le langage HTML

- Présentation générale
- Le corps d'un document
- Les différentes variantes d'HTML

CSS

Licence de droits d'usage

Plan

Internet et le Web

Le langage HTML

Présentation générale

Le corps d'un document

Les différentes variantes d'HTML

CSS

Licence de droits d'usage

HyperText Markup Language

- normalisé par le W3C (World Wide Web Consortium) regroupant industriels (Microsoft, Google, Apple. . .) et académiques (INRIA, MIT. . .)
- format ouvert : lecture possible dans des conditions correctes sans contrainte matérielle ou logicielle
- un fichier texte avec des balises
- description de la structure et du contenu d'un document, accent sur l'accessibilité
- on ne décrit pas la présentation (ce sera le rôle de CSS)
- on ne décrit pas de comportement dynamique (ce sera le rôle de JavaScript et des langages côté serveur)

Licence de droits d'usage

- HTML est un langage qui alterne texte et balises (`<blabla>` ou `</blabla>`)
  - Les balises permettent de structurer chaque partie du document et servent par exemple au navigateur pour réaliser la mise en page du document.
- Les fichiers HTML
  - sont structurés en deux parties principales : l'en-tête (`<head> ... </head>`) et le corps (`<body> ... </body>`)
- En HTML, les blancs (espace, tabulations, retours à la ligne) sont en général équivalents et servent juste à délimiter mots, balises... Leur nombre n'a pas d'importance.

Licence de droits d'usage

Balises

- Leur syntaxe est (balises ouvrante et fermante)

```
<balise attributs>contenu</balise>
```

ou (balise sans contenu)

```
<balise attributs>
```

balise mot clé désignant un élément particulier

contenu peut contenir du texte ou d'autres balises

attributs représente les différents paramètres associés à l'élément, sous la forme d'une liste de `nom="valeur"` ou `nom='valeur'`, séparés par des espaces (les guillemets ne sont pas toujours indispensables, mais elles le deviennent dès que `valeur` contient des caractères exotiques)

Licence de droits d'usage

Balises

- Les noms des éléments et des attributs sont souvent écrits en minuscule, mais `<head>` et `<HeAd>` sont équivalents.
- Les balises sont ouvertes et refermées dans l'ordre (`<b><i></i></b>` et non `<b><i></b></i>`).
- Des règles strictes déterminent quelles balises peuvent être mises à l'intérieur de quelles balises.
- Sous certaines conditions, une balise peut être implicitement refermée, mais ces conditions sont assez complexes à décrire.
- `<!--zut.-->` dénote un commentaire, qui ne sera ni affiché ni interprété par le client Web (utile pour documenter une partie d'une page Web).

Licence de droits d'usage

Balises : exemples

Exemples

- `<title>coucou</title>` pour attribuer le titre *coucou* au document
- `<em>cuicui</em>` pour mettre en *emphase* le texte *cuicui* (cela sera rendu, le plus souvent, par une mise en italique).
- `<strong>cuicui</strong>` pour indiquer que le texte **cuicui** est important (cela sera rendu, le plus souvent, par une mise en gras).

Contre-exemple

```
<strong><em>bouh</strong></em>
```


Licence de droits d'usage

Structure d'un document

```
<!DOCTYPE html PUBLIC "-//W3C//DTD HTML 4.01//EN"
  "http://www.w3.org/TR/html4/strict.dtd">
<html lang="fr">
  <head>
    <!-- En-tête du document -->
  </head>
  <body>
    <!-- Corps du document -->
  </body>
</html>
```

- La déclaration `<!DOCTYPE ...>` précise la version d'HTML utilisée.
- La langue du document est précisée avec l'attribut `lang` de la balise principale `<html>`.

Licence de droits d'usage

En-tête

- L'en-tête du document est délimitée par les balises `<head> ... </head>` .
- L'en-tête contient des méta-informations concernant le document telles que son titre, son encodage, les fichiers annexes, etc. Les deux informations les plus importantes sont :
  - Le jeu de caractères de la page, à mettre tout au début de l'en-tête

```
<meta http-equiv="Content-Type"
      content="text/html; charset=utf-8">
```

  - Le titre de la page (la seule information obligatoire à préciser) ; celui-ci sera par exemple affiché dans la barre de titre du navigateur, il n'apparaît pas dans la page elle-même.

```
<title>Mon super site</title>
```


Licence de droits d'usage

Jeux de caractères

Unicode : répertoire de caractères, assignant à chaque caractère, de quelque langue que ce soit, un nombre entier.

A	$\rightarrow$	65	ε	$\rightarrow$	949
é	$\rightarrow$	233	$\aleph$	$\rightarrow$	1488

Jeu de caractères : moyen de représenter concrètement, par une suite de 0 ou de 1, un caractère Unicode.

Exemples (é)

iso-8859-1	`11101001`	Seulement pour certains caractères
utf-8	`11000011 10101001`	
utf-16	`11101001 00000000`	

utf-8 présente l'avantage de pouvoir représenter tous les caractères d'Unicode, de manière compatible avec l'ancien encodage ASCII.

Licence de droits d'usage

Plan

Internet et le Web

Le langage HTML

- Présentation générale
- Le corps d'un document
- Les différentes variantes d'HTML

CSS

Licence de droits d'usage

Corps d'un document

- Les balises `<body> ... </body>` délimitent le corps du document, structuré en sections, paragraphes, listes, etc.
- Il existe 6 balises permettant de représenter les titres de sections, par importance décroissante :
  - `<h1>Titre de la page</h1>`
  - `<h2>Titre de section principale</h2>`
  - `<h3>Titre de sous-section</h3>`
- Les balises `<p> ... </p>` permet de délimiter un paragraphe. Tous les paragraphes de texte doivent être balisés ainsi.
- La balise `<hr>` (horizontal rule) indique une séparation majeure dans le document (rendue par exemple graphiquement par une ligne horizontale).
- Directement à l'intérieur de `<body> ... </body>` n'apparaissent que des balises de bloc : `<p>`, `<h1>`, `<form>`, `<hr>`, `<ul>`, `<table>` ainsi que `<div>` qui dénote un bloc sans sémantique particulière.

Licence de droits d'usage

Listes

- HTML possède plusieurs balises permettant de présenter le texte sous forme de listes.
- On en distingue trois types :
  - les listes non numérotées,
  1. les listes numérotées,

  les listes de définitions (ou lexiques)
- Ces listes peuvent être emboîtées les unes à l'intérieur des autres.

Licence de droits d'usage

Listes (suite)

- Les listes classiques :
  - Les listes non numérotées délimitées par les balises `<ul> ... </ul>` (unordered list).
  - Les listes numérotées délimitées par les balises `<ol> ... </ol>` (ordered list).
  - Tous les éléments d'une liste numérotée ou non sont délimités par les balises `<li> ... </li>` (list item)
- Les lexiques sont délimités par les balises `<dl> ... </dl>` (definition list) et leurs entrées par les balises `<dt> ... </dt>` (term) et `<dd> ... </dd>` (definition).

Exemples

```
<ol> <li>un</li> <li>deux</li> </ol>

<dl> <dt>lapin</dt> <dd>rongeur à oreilles</dd> </dl>
```


Licence de droits d'usage

Tableaux

- Les tableaux sont délimités par les balises `<table>...</table>`.
- Les balises `<tr>...</tr>` (table row) délimitent les lignes.
- Les balises `<td>...</td>` (table data) délimitent les cellules.
- Attention ! On déclare les lignes à l'intérieur du tableau, les cellules à l'intérieur des lignes.

Exemple

```
<table>
  <tr> <td> l1, c1 </td> <td> l1, c2 </td> </tr>
  <tr> <td> l2, c1 </td> <td> l2, c2 </td> </tr>
</table>
```


Licence de droits d'usage

Ajouter de la structure à un tableau en :

- donnant une légende au tableau avec les balises `<caption>... </caption>` juste après la balise ouvrante `<table>`.
- remplaçant les `<td> ... </td>` qui contiennent des en-têtes (de ligne, de colonne) par des `<th> ... </th>` (table header).

Licence de droits d'usage

Images

- Pour insérer une image dans un document HTML, on utilise la balise `<img>` .
  - L'attribut `src` permet de préciser où se trouve l'image.
  - L'attribut `alt` permet de remplacer l'image par un texte quand elle n'est pas disponible. Il est obligatoire de l'utiliser, pour que tout agent (malvoyants, navigateur texte, incidents techniques, robots) ne pouvant voir votre image puisse avoir un texte alternatif.

```
<img src="http://www.lri.fr/images/lri.png" alt="LRI">
<img src="../images/montblanc.png" alt="Mont Blanc">
```

- Les formats d'images utilisables pour le Web sont :
  - Le JPEG (.jpg), un format adapté aux photos.
  - Le GIF (.gif) et le PNG (.png), des formats adaptés aux autres types d'image ; le PNG est à préférer dans tous les cas (transparence, profondeur de couleurs. . .) sauf besoin d'images animées (à utiliser avec parcimonie !).

Licence de droits d'usage

Liens

- Ce qui différencie une page Web (page HyperTexte) d'un banal document : ce sont les liens !
- Ils sont introduits par la balise `<a> ... </a>` (cette balise est une balise en ligne, il faut la placer à l'intérieur d'un bloc).
- En cliquant sur un lien, on peut se déplacer vers :
  - un autre serveur ou un fichier du même serveur
  - une autre partie du même document

Licence de droits d'usage

- Pour faire un lien, on utilise l'attribut `href` de la balise `<a>` dont le contenu formera le lien :

```
<a href="http://www.cnrs.fr/">
  <img src="images/cnrs.gif" alt="CNRS">
</a>

<a href="bio/indexbioinfo.html">Bioinformatique</a>
```

- Les ancres servent à atteindre un endroit précis dans le document.
  - On commence par définir les ancres, soit sur une balise existant déjà grâce à l'attribut `id`, soit avec un `<a id="...">` :

```
<h3 id="tutorials">Tutorials</h3>
<a id="tutorials">
```

  - Ensuite, on fait le lien avec cette ancre.

```
<a href="#tutorials">tutorials</a>
<a href="http://www.w3.org/#tutorials">tutorials</a>
```

  - On voit parfois l'ancienne syntaxe `<a name="...">`.

Et aussi...

- De nombreuses autres balises en ligne : `<abbr>`, `<cite>`, `<var>` ...
- Quelques autres balises blocs : `<blockquote>`, `<address>` ...
- Représentation des caractères spéciaux (ex., « é ») :
  - directement dans le codage du document (utf-8 de préférence, pour représenter tous les caractères possibles), à privilégier ;
  - par leur code en décimal (&#233 ;) ou en hexadécimal (&#xE9 ;) ;
  - par des entités caractères nommées (&eacute ;)

Licence de droits d'usage

Plan

Internet et le Web

Le langage HTML

- Présentation générale
- Le corps d'un document
- Les différentes variantes d'HTML

CSS

Licence de droits d'usage

Les différentes version d'HTML

- HTML 4.01 (1999) strict (vu plus haut) et transitionnel

```
<!DOCTYPE html PUBLIC "-//W3C//DTD HTML 4.01//EN"
"http://www.w3.org/TR/html4/strict.dtd">
```

```
<!DOCTYPE html PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN"
"http://www.w3.org/TR/html4/loose.dtd">
```

- XHTML 1.0 (2000) strict et transitionnel

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
```

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
```

- XHTML 1.1 et XHTML 2.0 : principalement un échec, inutilisable et inutilisé sur le Web actuel
- HTML 5 : le standard futur, en développement

Licence de droits d'usage

Soupe de balises

- On trouve encore beaucoup de documents HTML antérieurs à HTML 4.01
- En pratique : beaucoup de pages Web ne respectent aucun de ces standards (avec ou sans déclaration de type de document) $\Longrightarrow$ navigateurs ne respectent pas ces standards $\Longrightarrow$ soupe de balises !
- Pour de nouveaux développements : absolument indispensable de respecter les standards du Web (compatibilité avec les clients Web d'aujourd'hui et de demain), même si ce n'est pas suffisamment compris par les concepteurs de sites Web

Licence de droits d'usage

HTML contre XHTML

- XHTML : une application d'XML
- Les balises sans contenus `<hr>`, s'écrivent `<hr />` en XHTML.
- Certains éléments peuvent ne pas être refermées en HTML (`<ol> <li> un <li>deux </ol>`), mais la fermeture est obligatoire en XHTML.
- Les valeurs des attributs peuvent ne pas être entre guillemets (`<img src=toto.png alt=toto>`) en HTML, guillemets obligatoires en XHTML.
- Les noms des éléments et des attributs sont insensibles à la casse en HTML (`<HTMl laNg=fr>`), contrairement à XHTML ou tout doit être en minuscule.
- Attributs `xmlns` et `xml:lang` sur la balise `<html>` en XHTML.
- Et quelques autres petites subtilités. . .

Licence de droits d'usage

Avantages respectifs de HTML et XHTML

Avantages de HTML 4.01

- Meilleur support par les navigateurs (Internet Explorer 6/7 comprennent mal XHTML).
- Moins de contraintes. . .
- Beaucoup plus utilisé sur le Web.

Avantages de XHTML 1.0

- Plus de contraintes. . . (donc plus simple !).
- Syntaxe claire, sans ambiguïté.
- Familiarité avec XML, utile dans d'autres contextes.
- Facilité d'utilisation dans des contextes XML (p.ex. XSLT).

Licence de droits d'usage

Règles de compatibilité à respecter pour du XHTML

- Théoriquement, `<p />` et `<p></p>` sont synonymes en XML, donc en XHTML. En pratique, on utilisera la notation `<balise />` uniquement pour les balises n'ayant jamais de contenu (p.ex. `<br />`, `<hr />` ...).
- Théoriquement, un document XHTML peut commencer par une ligne `<?xml version="1.0"encoding="utf-8"?>`. En pratique, on l'omettra.
- cf. `http://www.w3.org/TR/xhtml1/#guidelines`
- On peut utiliser `http://qa-dev.w3.org/~bjoern/appendix-c/validator/` pour vérifier le respect de ces quelques règles.

Licence de droits d'usage

Strict contre Transitionnel

- À la fois en HTML 4.01 et en XHTML 1.0, on a le choix entre une déclaration de type de document Strict ou Transitional
- Différence : moins d'éléments autorisés en Strict
- Des éléments de présentation (`<u>` pour soulignement, `<center>`, `<font>`) en Transitional, héritage des anciennes versions d'HTML
- À éviter de nos jours : séparation fond/forme, présentation en CSS
- Pour de nouveaux projets : Strict
- Pour des modifications d'anciens sites avec beaucoup de balises présentationnelles : Transitional

Licence de droits d'usage

XHTML 2.0 contre HTML 5

- XHTML 2.0 : initiative du W3C, incompatible avec HTML 4.01/XHTML 1.0, changements majeurs
- HTML 5 : initiative des développeurs de navigateurs, compatibilité avec HTML 4.01/XHTML 1.0, changements incrémentaux mais nombreux
- XHTML 2.0 abandonné en juillet 2009
- Des fonctionnalités de HTML 5 font leur apparition dans les navigateurs récents (y compris Internet Explorer 9)
- HTML 5 laisse le choix entre les conventions HTML 4.01 ou XHTML
- Nouvelles fonctionnalités : dessin 2D (`<canvas>`), multimédia (`<audio>`, `<video>`), meilleurs éléments structurants (`<section>`, `<footer>`), etc.

Licence de droits d'usage

Plan

Internet et le Web

Le langage HTML

CSS

- CSS et HTML
- Sélecteurs de CSS
- Mise en forme
- Mise en page
- Fonctionnalités mal supportées

Licence de droits d'usage

Plan

Internet et le Web

Le langage HTML

CSS

CSS et HTML

Sélecteurs de CSS

Mise en forme

Mise en page

Fonctionnalités mal supportées

Licence de droits d'usage

Intérêt

- CSS : Cascading StyleSheets
- Recommandation du W3C
- Plusieurs niveaux : CSS1 (1996), CSS2 (1998), CSS2.1 (2005), CSS3 (en cours)
- Support par les navigateurs très inégal, jamais complet (en particulier, IE6 et IE7 ont plusieurs limitations importantes) $\Longrightarrow$ on se concentrera sur ce qui marche bien dans l'ensemble des navigateurs courants.

Principe

- HTML décrit la structure et le contenu
- CSS décrit :
  - la mise en forme du texte
  - la mise en page des boîtes les unes par rapport aux autres

Licence de droits d'usage

Styles en ligne

- Manière le plus simple d'utiliser les CSS.
- Rajouter un attribut `style` sur les balises HTML.
- On peut utiliser `<span>` si on a besoin d'une balise supplémentaire.
- Encombre le code HTML avec des indications de mise en forme : ce n'est pas ce qu'on veut !

Exemples

```
Ce mot en <em style="color: red;">emphase</em> est aussi en rouge.
```

```
<span style="text-decoration: underline">Cette expression de plusieurs mots</span> est soulignée.
```


Licence de droits d'usage

Styles en en-tête

- Intégration des propriétés de style à l'en-tête de la page avec la balise `<style>`
- Utilisation des sélecteurs pour définir à quels élément les propriétés s'appliquent
- Inconvénients : mélange HTML et CSS dans le même document, impossible de réutiliser les propriétés CSS dans plusieurs documents

Exemple

```
<!DOCTYPE ... >
<html>
  <head> ...
    <style type="text/css">
      em { color: red; }
    </style>
  </head>
  <body> ... </body>
</html>
```

Feuille de styles liée

- Mettre la feuille de style CSS dans un fichier à part (en général, on utilise l'extension *.css*).
- Permet d'utiliser la même feuille de style pour plusieurs pages Web.
- Rajouter une balise `<link>` dont l'attribut `rel` est positionné à `"stylesheet"` dans l'en-tête du document.
- Possibilité d'ajouter `media="screen"` ou `media="print"`, etc., pour choisir différentes feuilles de style suivant le mode d'affichage.

Exemple

```
<!DOCTYPE ... >
<html>
  <head> ...
    <link rel="stylesheet" href="feuille.css" type="text/css">
  </head>
  <body> ... </body>
</html>
```


Licence de droits d'usage

Classes

- On veut parfois rajouter encore plus de structure et de sémantique à un document HTML.
- On utilise l'attribut `class` sur n'importe quelle balise (ou, à défaut, sur une balise `<span>`).
- Après, on peut utiliser CSS pour appliquer une mise en forme commune à tout ce qui fait partie d'une `class` particulière.

Exemple (Mettre en bleu italique les noms de personnes)

- HTML

```
<p>Je voudrais remercier en particulier
<span class="personne">Madame Machin</span>
et <span class="personne">Monsieur Bidule</span>.</p>
```

- CSS

```
.personne { color: blue; font-style: italic; }
```


Licence de droits d'usage

Syntaxe de CSS

- Ensemble de règles de la forme :

```
sélecteur {
  propriété: valeur;
}
```

- `sélecteur` : indique à quelles parties du documents la règle s'aplique
- `propriété` : propriété spécifique de mise en forme à modifier
- `valeur` : son sens dépend de la propriété.
- Les feuilles de style peuvent être validées avec un validateur approprié, cf. `http://jigsaw.w3.org/css-validator/`
- Commentaires entre `/*` et `*/`.

Licence de droits d'usage

Cascade

- Plusieurs feuilles de style peuvent s'appliquer simultanément :
  - Plusieurs balises `<link rel="stylesheet">`
  - Directive `@import` d'une feuille de style

    ```
    @import url(feuille_annexe.css);
    ```
  - Feuille de style de l'utilisateur (Mozilla, Opera. . .)
- Au sein même d'une feuille de style, plusieurs règles peuvent être en conflit.
- Cascade : mécanismes gérant ces conflits.

Cascade

- Si !important est précisé après la valeur, la règle sera prioritaire.
- Sinon, plus une règle est spécifique, plus elle est prioritaire.
- Sinon, la dernière règle s'applique.

Licence de droits d'usage

Plan

Internet et le Web

Le langage HTML

CSS

- CSS et HTML
- Sélecteurs de CSS
- Mise en forme
- Mise en page
- Fonctionnalités mal supportées

Licence de droits d'usage

Sélecteurs simples, multiples, universel

Sélecteur simple : nom d'une balise.

Sélecteur multiple : plusieurs sélecteurs séparés par des virgules.

Sélecteur universel : '`*`', sélectionne tout.

Exemples

- `ul { color: blue; }` met l'ensemble du contenu des listes non ordonnées en bleu.
- `h1,h2,h3,h4,h5,h6 { color: red; }` met l'ensemble des titres de rubrique en rouge.
- `* { color: black; }` met tout en noir. Dans ce cas précis, on préférera une règle `body { color: black; }`.

Licence de droits d'usage

Sélecteurs de classes

Sélecteur de classe : nom d'une classe, préfixée d'un '.', tel qu'il apparaît dans un attribut `class` d'une balise HTML.

Exemples

- `.personne { font-weight: bold; }` met l'ensemble des balises de classe `personne` en gras.
- `p.comment { font-style: italic; }` met l'ensemble des balises `<p>` de classe `comment` en italique.

Licence de droits d'usage

Sélecteurs d'identifiants

Identifiant : défini par l'attribut `id` d'une balise HTML. Similaire aux classes, mais il ne peut y avoir qu'une seule balise ayant un `id` donné dans tout le document HTML.

Sélecteur d'identifiant : nom d'une classe, préfixée d'un '#', tel qu'il apparaît dans un attribut `id` d'une balise HTML.

Exemples

- `#introduction { font-size: 120%; }` met la balise d'identifiant `introduction` en plus gros.
- `p#introduction { font-size: 120%; }` met la balise `<p>` d'identifiant `introduction` en plus gros.

Licence de droits d'usage

Sélecteurs contextuels

Sélecteur contextuel : 2 sélecteurs ou plus séparés par des espaces. *A* *B* sélectionne les *B* seulement s'ils sont contenus dans des *A*.

Exemples

- `h1 em { color: blue; }` met les mots en emphase à l'intérieur d'un `h1` en bleu.
- `ul ol, ol ul, ul ul, ol ol { font-size: 80%; }` diminue la taille du texte des listes imbriquées.

Licence de droits d'usage

Pseudo-classes

- Pseudo-classes : permet de faire une sélection d'un élément uniquement dans certains contextes.
- `a:link` : les `<a>` qui sont des liens.
- `a:visited` : les `<a>` qui sont des liens vers des pages qui ont déjà été visitées.
- `a:hover` : les `<a>` qui sont des liens qu'on est en train de désigner (i.e., la souris est en train de passer dessus).
- `a:active` : les `<a>` qui sont des liens qu'on est en train d'activer (i.e., de cliquer).

Remarque

Attention ! Toujours définir ces pseudo-classes dans ce sens, à cause des règles de cascade. (Las Vegas (Forest) Has Animals).

Licence de droits d'usage

Plan

Internet et le Web

Le langage HTML

CSS

- CSS et HTML
- Sélecteurs de CSS
- Mise en forme
- Mise en page
- Fonctionnalités mal supportées

Licence de droits d'usage

Propriétés de longueur

Les propriétés admettent différentes unités de mesure se répartissant suivant trois types de valeurs :

- sous forme de pourcentage (par rapport à la valeur courante)
- les valeurs relatives :
  - `em` la valeur de la propriété `font-size` de la police utilisée (hauteur du bloc dans lequel s'inscrit naturellement un caractère)
  - `ex` la hauteur du caractère `x` dans la police utilisée
- les valeurs relatives à l'écran (à éviter pour avoir une conception de la page indépendante de la résolution !)
  - `px` le nombre de pixels
- les valeurs absolues (N'ont aucun sens pour une page Web destinée à s'afficher sur un écran d'ordinateur !) : mm, cm, in, pt, pc

Licence de droits d'usage

Polices

- Il existe différents moyens pour contrôler les polices de caractères.
- Les principales propriétés (`font-family`, `font-size`, `font-weight` et `line-height`) sont responsables de l'aspect du texte.

Exemple

```
p {
  font-family: "Times New Roman";
  font-size: 130%;
  font-weight: bold;
  line-height: 150%;
}
```


Licence de droits d'usage

Alignement et indentation

`text-align` définit la justification et l'alignement d'un texte : `left`, `right`, `center` et `justify` (aligne le texte sur les marges gauche et droite).

`vertical-align` définit l'alignement vertical de la ligne de base d'un élément en fonction de la propriété `line-height` : `super` (texte en exposant), `sub` (texte en indice), `baseline` (normal), etc.

`text-indent` permet de décaler le début de la première ligne de texte d'une valeur fixe ou proportionnelle à la valeur du paragraphe.

Licence de droits d'usage

Couleurs

- La couleur doit être utilisée avec discernement.
- Son rôle est de différencier les éléments d'un texte en jouant sur la notion de contraste.
- Il existe différentes manières de spécifier les valeurs d'une couleur :
  - En indiquant le nom d'une couleur prédéfinie parmi `aqua`, `black`, `blue`, `fuchsia`, `gray`, `green`, `lime`, `maroon`, `navy`, `olive`, `purple`, `red`, `silver`, `teal`, `white` et `yellow`
  - En utilisant le code RGB pour (Red, Green, Blue) avec des valeurs de 0 à 255
  - En utilisant la valeur hexadécimale

Licence de droits d'usage

- La propriété *couleur* s'applique généralement au texte (`color`) ou à l'arrière-plan (`background-color`).
- La propriété `background-image` définit l'image d'arrière-plan d'un ou plusieurs éléments ou encore de tous les éléments. Il est fortement recommandé de spécifier aussi une couleur normale d'arrière-plan, qui sera utilisé en remplacement.

Exemple

```
background-image: url("images/monImage.jpg");
background-color: white;
```

Licence de droits d'usage

Plan

Internet et le Web

Le langage HTML

CSS

- CSS et HTML
- Sélecteurs de CSS
- Mise en forme
- Mise en page
- Fonctionnalités mal supportées

Licence de droits d'usage

Blocs et éléments en ligne

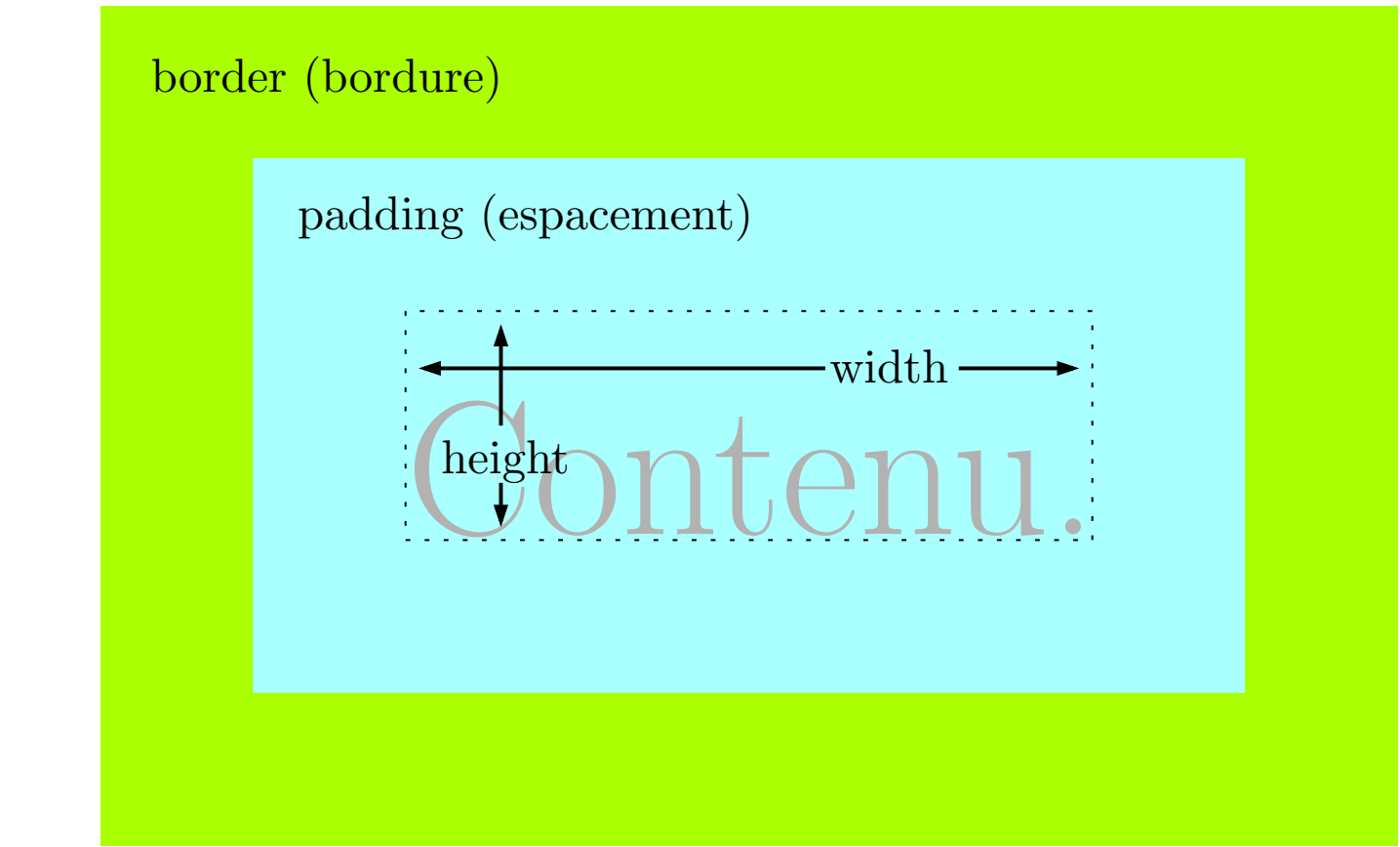
- Il existe deux sortes d'élements HTML :
  - Les blocs : `<p>` , `<h1>` , `<ul>` . . .

    `<div>` est un bloc générique.
  - Les éléments en ligne, qui doivent être placés à l'intérieurs de blocs : `<a>` , `<img>` , `<em>` . . .

    `<span>` est un élément en ligne générique.
- On s'intéresse ici principalement aux blocs et à la manière dont les placer les uns par rapport aux autres.

Licence de droits d'usage

Boîte CSS



Licence de droits d'usage

Flottement et positionnement

- Le flottement et le positionnement sont des outils permettant une mise en page complexe : mise en forme en colonnes, chevauchement d'éléments, etc.
- Le positionnement consiste à préciser où doivent apparaître des éléments de manière relative, par rapport à un autre élément ou encore par rapport à la fenêtre du navigateur.
- Le flottement n'est pas vraiment un positionnement : les éléments dits flottants sont pris dans le flux et les autres éléments les contournent.

Licence de droits d'usage

Flottement

- Le flottement est défini par la propriété `float` qui peut prendre les valeurs :
  - `left` : l'élément sera contourné par la droite.
  - `right` : l'élément sera contourné par la gauche
  - `none` : valeur par défaut, sert principalement pour écraser un style existant.
- Tout bloc (image, texte, etc.) peut être défini comme un élément flottant.
- Il existe une dizaine de règles précises qui gouvernent le comportement des éléments flottants.

Licence de droits d'usage

Positionnement statique et relatif

- Le positionnement `static` est le comportement par défaut :
  - l'élément est une boîte rectangulaire faisant partie du flux i.e. l'ordre des déclarations contenues dans le code source est respecté.
- Le positionnement `relative` permet de décaler un élément d'une certaine distance.

Licence de droits d'usage

Positionnement absolu et fixe

- Le positionnement `absolute` permet de retirer totalement un élément du flux et de le positionner par rapport à son conteneur (plus précisément, par rapport au conteneur le plus proche ayant un positionnement relatif).
- Le positionnement `fixed` permet de retirer totalement un élément du flux et de le positionner par rapport à la fenêtre d'affichage. Attention : Ne fonctionne pas avec Internet Explorer 6.

Remarque

Pour un positionnement absolu par rapport à la page (cas le plus fréquent), on utilisera `body { position: relative; }`.

Licence de droits d'usage

Décalage

- Une fois le type de positionnement choisi, on peut spécifier les propriétés de décalage : `top`, `right`, `bottom` et `left`.
- Le décalage se décrit à partir du bord le plus proche du bloc conteneur.
- Leur valeur peut être exprimée via une longueur, un pourcentage (par rapport à la taille du bloc conteneur) ou être choisie automatiquement `auto`.

Licence de droits d'usage

Visibilité

- La visibilité d'un élément peut être complètement contrôlée en utilisant la propriété `visibility` :
  - `visible` permet de rendre visible l'élément
  - `hidden` permet de rendre invisible un élément ; la mise en page continue à le prendre en compte

Licence de droits d'usage

Plan

Internet et le Web

Le langage HTML

CSS

- CSS et HTML
- Sélecteurs de CSS
- Mise en forme
- Mise en page
- Fonctionnalités mal supportées

Licence de droits d'usage

En CSS2.1

- `p>em` : `<em>` directement à l'intérieur d'un `<p>` (pas IE6)
- `li+li` sélectionne les `<li>` qui suivent juste un autre (pas IE6)
- `*:lang(fr)` sélectionne les éléments dont la langue est définie comme le français (avec un attribut `lang=`, pas IE6 et IE7)
- `em:after {content: "!"}` ajoute un « ! » après tous les `<em>` (`*:before` existe aussi, pas IE6 et IE7)
- Valeur `inherit` indique que la valeur est héritée des éléments de la hiérarchie (pas IE)
- `max-width` / `min-height` /etc. permettent de donner des hauteurs et largeurs minimales et maximales aux blocs (pas IE6)
- `display: table-cell;` (et `table`, `table-row`...) permet de mettre en page n'importe quels éléments comme des tables (pas IE6 et IE7)
- Compteurs automatiques (pas IE6 et IE7), etc.

Licence de droits d'usage

En CSS3

- Plus de noms de couleurs (marche presque dans IE !)
- Différentes feuilles de style suivant la résolution de l'écran, etc. (seulement Opera)
- Sélecteurs suivant la valeur d'un attribut (ex. `a[href$=".pdf"]`, pas IE6)
- Sélectionne les n^{es}, $2n^{es}$, etc., éléments à l'intérieur d'un autre élément (quasiment pas supporté)
- Polices de caractères téléchargées automatiquement `@font-face` (tous navigateurs, mais formats de polices différents requis par les différents navigateurs). cf `http://craigmod.com/journal/font-face/`
- Et énormément d'autres choses, toujours en développement

Licence de droits d'usage

Licence de droits d'usage

Contexte public } avec modifications

Par le téléchargement ou la consultation de ce document, l'utilisateur accepte la licence d'utilisation qui y est attachée, telle que détaillée dans les dispositions suivantes, et s'engage à la respecter intégralement.

La licence confère à l'utilisateur un droit d'usage sur le document consulté ou téléchargé, totalement ou en partie, dans les conditions définies ci-après et à l'exclusion expresse de toute utilisation commerciale.
Le droit d'usage défini par la licence autorise un usage à destination de tout public qui comprend :
– le droit de reproduire tout ou partie du document sur support informatique ou papier,
– le droit de diffuser tout ou partie du document au public sur support papier ou informatique, y compris par la mise à la disposition du public sur un réseau numérique,
– le droit de modifier la forme ou la présentation du document,
– le droit d'intégrer tout ou partie du document dans un document composite et de le diffuser dans ce nouveau document, à condition que :
    – L'auteur soit informé.

Les mentions relatives à la source du document et/ou à son auteur doivent être conservées dans leur intégralité.

Le droit d'usage défini par la licence est personnel et non exclusif.
Tout autre usage que ceux prévus par la licence est soumis à autorisation préalable et expresse de l'auteur : sitepedago@telecom-paristech.fr