

World wide web

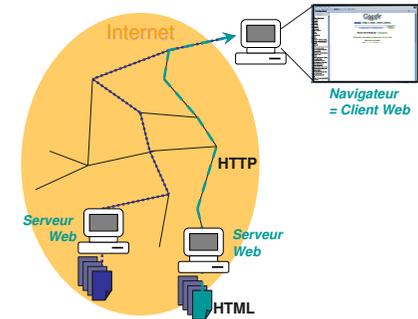
- ❑ PCI
- ❑ Permis de Conduire Informatique
- ❑ L1 1er semestre
- ❑ UFR d'Informatique



Objectifs du cours et des 3 TP

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

Le Web en un schéma



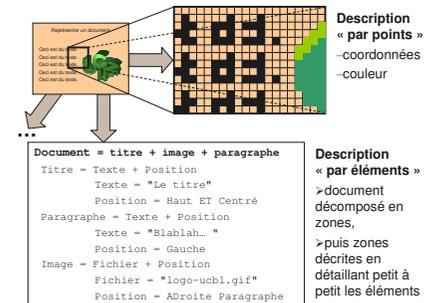
Différents médias

- ❑ Différentes formes pour l'utilisateur
 - texte
 - image (photo, dessin, graphique)
 - son (musique, voix)
 - vidéo (images animées + son)
- ❑ Différents formats de fichiers
 - texte simple : ASCII (.txt)
 - images : JPEG (.jpg), GIF (.gif)
 - sons : WAVE (.wav), MPEG3 (.mp3)
 - vidéo : MPEG (.mpg), AVI (.avi), Quicktime (.qt)...
- ❑ Différentes applications permettant de « lire » les fichiers (ouvrir les documents)
 - éditeur de texte
 - visualisateur d'images
 - lecteur de son, de vidéo

Document multimédia

- ❑ Combinaison de différents médias
 - textes + images + sons + vidéos
- ❑ 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 hypertexte, remplir une zone texte...
- ❑ Exemples d'utilisations
 - cours, jeux interactifs, logiciels éducatifs, encyclopédies interactives, bornes dans les lieux publics, pages Web, films DVD...

Plusieurs façons de décrire un document



Page Web - définitions

- ❑ Une page Web, c'est
 - un document multimédia
 - images, textes, possibilité d'interaction
 - décrit élément par élément
 - titre, morceaux de texte, images...
 - avec un langage de description
 - HTML
 - stocké dans un fichier
 - .html (ou .htm)
- ❑ Une page Web
 - est calculée et affichée par un navigateur
 - se trouve à un endroit sur Internet décrit par une adresse (URL)
 - permet d'accéder à d'autres pages en suivant des liens

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/>

Présentation d'une page HTML dans un navigateur

- ❑ Page Web
 - fichier html
- ❑ Navigateur (browser, butineur)
 - logiciel qui demande des fichiers HTML, les reçoit, les interprète et les présente à l'utilisateur
- ❑ Principaux navigateurs
 - Mozilla Firefox, Netscape Communicator, Opéra, Microsoft Internet Explorer.



1. Introduction 1. Généralités
2. Langage HTML 2. Balises
3. Navigation 3. Tableaux
4. Recherche d'information

Un 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

- ≈ 40 éléments de description dans le langage HTML
- (PCI : quelques éléments de base)

PCI 2005 – CM 03 World wide web 10

1. Introduction 1. Généralités
2. Langage HTML 2. Balises
3. Navigation 3. Tableaux
4. Recherche d'information

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>)
 - balises avec attributs (<TITLE AUTHOR="PCI">)
- Un élément
 - est en général délimité par deux balises
 - <P> blabla </P>
 - peut être décrit par une balise, avec des attributs
 -
 - peut en contenir un autre (récursivement)
 - <P> blabla truc blabla </P>
 - (= l'élément P contient l'élément B)

PCI 2005 – CM 03 World wide web 11

1. Introduction 1. Généralités
2. Langage HTML 2. Balises
3. Navigation 3. Tableaux
4. Recherche d'information

Exemple de description dans le langage à balise HTML

```

Balise ouvrante -> <html>
<head>
  <Title> Merveilles du monde </Title>
</head>
<body>
  <p> Dans l'image ci-après
  <img SRC = "zebre.jpg">
  nous présentons un zébre vivant
  dans la savane.
  </p>
  <p>
  <img SRC = "lion.gif">
  Le lion n'est pas vraiment son ami
  </p>
</body>
Balise fermante -> </html>
  
```

Élément p avec contenu textuel

Élément sans contenu (une seule balise)

Attribut de l'élément Image

Arborescence

PCI 2005 – CM 03 World wide web 12

1. Introduction 1. Généralités
2. Langage HTML 2. Balises
3. Navigation 3. Tableaux
4. Recherche d'information

Structuration de documents

```

<html>
<head>
  <title> Ceci est le titre </title>
</head>
<body bgcolor="#ffffff" <!-- ou "white" -->
  <p> Paragraphe A </p>
  <p> Paragraphe B </p>
  <h1> Titre de niveau 1 </h1>
  <p> Paragraphe C </p>
  <h2> Titre de niveau 2 </h2>
  <h3> Titre de niveau 3 </h3>
  <h4> Titre de niveau 4 </h4>
  <p> ... </p>
  <h6> Titre de niveau 6 </h6>
</body>
</html>
  
```

Élément titre

En-tête

balise de corps avec attribut couleur de fond (ici blanc)

commentaire

Paragraphe A

Paragraphe B

Titre de niveau 1

Paragraphe C

Titre de niveau 2

Titre de niveau 3

Titre de niveau 4

...

Titre de niveau 6

6 niveaux de titres

Corps

PCI 2005 – CM 03 World wide web 13

1. Introduction 1. Généralités
2. Langage HTML 2. Balises
3. Navigation 3. Tableaux
4. Recherche d'information

Mise en forme : gras, italiques, souligné et listes

```

<P>Ceci est en<b> gras (Bold)</b>
<P>Ceci en<i>italique et gras.</i><b></b>
Ceci est souligné (Underlined)
Voici une
  <ul>
  <li> ligne dans une liste à puces
  </li>
  <li> autre ligne
  </li>
  </ul>
Voici une<p>
  <ul>
  <li> ligne dans une liste à puces</li>
  <li> et une autre <u> ligne.</u></li>
  </ul>
  
```

Élément liste non numérotée (Unordered list)

Élément item de liste (List Item)

PCI 2005 – CM 03 World wide web 14

1. Introduction 1. Généralités
2. Langage HTML 2. Balises
3. Navigation 3. Tableaux
4. Recherche d'information

Différentes URL utilisées en HTML

- Forme principale (protocole HTTP)


```
http://www.orig.univ-lyon1.fr/livret/presentation.htm
```

Protocole Adresse machine Chemin fichier
- Forme pour désigner les fichiers locaux
 - Chemin relatif : fichier.html, dossier/toto.html
 - Sur un disque : file://C:/chemin/fichier.htm
- Forme pour le transfert de fichiers
 - ftp://www.orig.univ-lyon1.fr/livret/cycle1/DEUGsv.pdf
- Forme pour l'envoi de courrier électronique
 - mailto:jean.durand@univ-lyon1.fr

PCI 2005 – CM 03 World wide web 15

1. Introduction 1. Généralités
2. Langage HTML 2. Balises
3. Navigation 3. Tableaux
4. Recherche d'information

Insertion de liens

Syntaxe générale :

```
<A HREF="URL_du_lien"> texte du lien </A>
```

(A pour Ancre)

```
<P><A HREF="http://www.univ-lyon1.fr/">Ceci est un lien vers le serveur Web de Lyon 1.</A></P>
```

```
<P>Voici <A HREF="fichier.html">un lien</A> vers fichier.html situé dans le même dossier que le fichier en cours.</P>
```

```
<P>Voici un lien qui permet de <A HREF="ftp://www.orig.univ-lyon1.fr/livret/cycle1/DEUGsv.pdf">télécharger</A> un fichier par FTP.</P>
```

```
<P>Si je clique <A HREF="mailto:jean.durand@univ-lyon1.fr">ici</A>, j'envoie un courrier électronique à Jean Durand.</P>
```

PCI 2005 – CM 03 World wide web 16

1. Introduction 1. Généralités
2. Langage HTML 2. Balises
3. Navigation 3. Tableaux
4. Recherche d'information

Insertion d'images

Syntaxe générale :

```
<IMG SRC="URL_de_l_image" ALT="texte alternatif">
```

```
<IMG SRC="vue-du-ciel.jpg" ALT="Lyon aérien"></P>
```

Élément image

```
<P>Voici une photo de Fourvière (URL distant) :
<IMG SRC="http://www.lyon-city.org/static/vdl/contenu/decouverte/au_fi_des_quartiers/sa_odeon.jpg"></P>
```

```
<P>Cliquez sur le logo <A HREF="http://www.univ-lyon1.fr/"><IMG SRC="logo-ucbl-3d.png"></A> pour visiter le site Web de Lyon 1.</P>
```

L'élément lien contient un élément image

PCI 2005 – CM 03 World wide web 17

1. Introduction 1. Généralités
2. Langage HTML 2. Balises
3. Navigation 3. Tableaux
4. Recherche d'information

Tableaux

```
<TABLE BORDER="1">
<TR>
  <TD> Haut gauche</TD>
  <TD> Haut droite</TD>
</TR>
<TR>
  <TD> Milieu gauche</TD>
  <TD> Milieu droite</TD>
</TR>
<TR>
  <TD> Bas gauche</TD>
  <TD> Logo UCBL
  <IMG SRC="logo-ucbl-3d.png">
</TD>
</TR>
</TABLE>
```

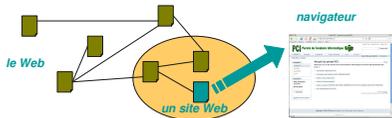
attribut « épaisseur de trait »

Élément cellule de tableau

Élément ligne de tableau

PCI 2005 – CM 03 World wide web 18

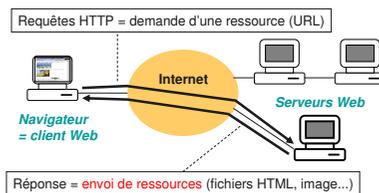
Technologie Web : une description plus formelle



- Ensemble de pages HTML formant un **réseau** (ou une **toile web** en anglais) sur lequel il est possible de naviguer
- WWW = World Wide Web
- Navigateur
 - outil permettant d'afficher des pages HTML et de passer d'une page à l'autre au moyen de liens hypertextes → navigation
- Site Web
 - regroupement de pages Web, accessible à partir d'une URL

Serveur Web

- Application dont le rôle est de fournir des ressources Web aux clients (navigateurs) qui les demandent
- Par extension, **machine** sur laquelle tournent ces services



Caractéristiques de la connexion

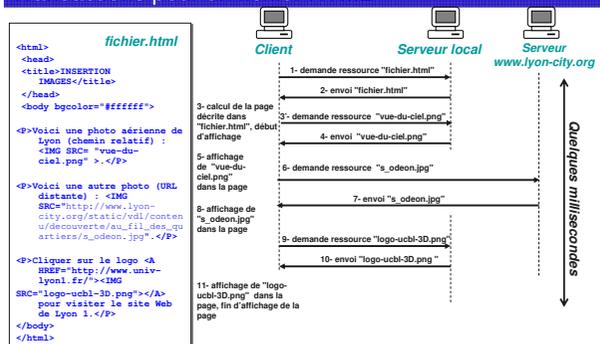
- Temps d'affichage d'une page Web lié
 - à la charge du serveur (réponse à n clients en même temps)
 - à l'encombrement du réseau
 - à la vitesse de la machine cliente
 - à la possibilité de présenter une page avant que toutes les données soient arrivées (ex. texte avant images)
- Deux possibilités pour fournir une page
 - à partir de fichiers existants sur le serveur
 - en générant la page HTML au moment de la requête

Pages fournies à partir de fichiers sur le serveur

- Cas le plus simple : le fichier est présent sur le disque du serveur, il suffit de le transférer vers votre machine
 - en général, le nom du fichier apparaît explicitement dans l'url : URL finissant en .html, .htm, .jpg...
 - erreurs : si on n'a pas le droit d'accéder (Error 403), si le fichier n'existe pas (not found : Error 404)
- Exemple de transaction

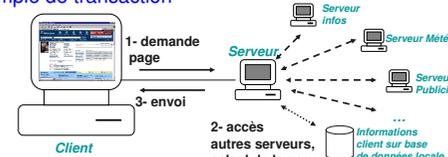


Transactions pour un fichier HTML

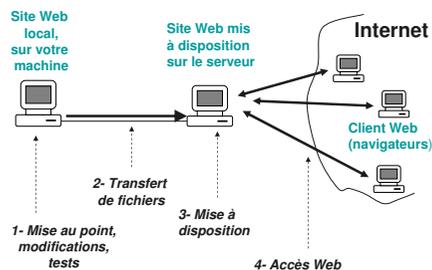


Pages générées à la demande

- Cas plus complexe : il faut créer une page à partir d'informations
 - présentes sur le serveur (base de données)
 - distantes du serveur (le serveur Web est le **client** d'autres serveurs)
- Nécessite un programme pour la création de la page (extension des fichiers : .php, .pl, .asp)
- Exemple de transaction



Conception, transfert et accès

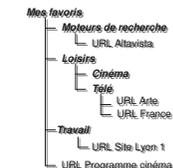


Plusieurs situations type

- Cas d'un particulier
 - site chez le Fournisseur d'Accès Internet
 - conception en local à la maison
 - transfert des fichiers chez le FAI par FTP
 - site sur la 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
- Cas de votre page à l'Université
 - site sur le serveur Web b201ntb.univ-lyon1.fr
 - un dossier de 2 Mo mis à votre disposition
 - transfert soit par FTP, soit par copie de fichiers vers un lecteur réseau (U :

Organisation d'URL par signets

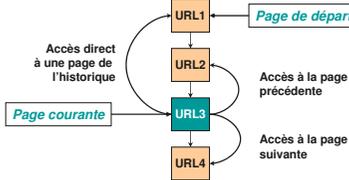
- Signets (marque-pages, favoris, bookmarks)
 - stocker des URL pour pouvoir retrouver des pages Web plus tard (possibilité d'ajouter des commentaires)
 - organiser ses signets
 - arborescence (hiérarchie) de « dossiers »
- Exemple



Organisation d'URL dans l'historique

Historique

- pendant une navigation, on consulte une suite de pages Web, le navigateur en garde trace comme liste d'URL
- possibilité d'aller en arrière, en avant, ou directement à une page



Impression de pages Web

Une page Web n'est en général pas prévue pour l'impression

- attention au **gaspillage** de papier
- plus un document est complexe, moins il a de chances de s'imprimer de façon satisfaisante

Quelques conseils

- attention aux couleurs de fond et de texte
- si seul le texte est intéressant
 - copier /coller vers un éditeur de texte, remettre en page, puis imprimer
- les sites les mieux faits proposent des pages prévues pour l'impression (ex. bouton « Pour imprimer »)

Interpréteur Javascript

Javascript

- programmes à l'intérieur de fichiers HTML (entre les balises `<SCRIPT>` et `</SCRIPT>`)

Le navigateur lit le programme et l'exécute

- pour gérer le navigateur
 - lancer des fenêtres
 - créer des info-bulles...
- pour gérer la navigation
 - vérifier qu'un formulaire est bien rempli avant de l'envoyer...
- documents plus interactifs et dynamiques qu'avec HTML seul

Il est possible de désactiver l'interpréteur Javascript du navigateur

Exécution d'applets JAVA

JAVA

- langage de programmation

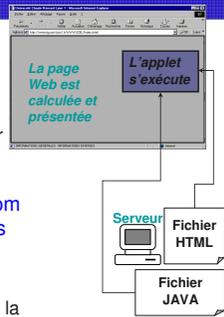
Applet

- programme JAVA qui peut fonctionner sur **tout** type de machine (systèmes d'exploitation différents)

Dans une page Web, on spécifie le nom de l'applet à exécuter entre les balises `<APPLET>` et `</APPLET>`

Le navigateur sait gérer une applet

- l'applet s'exécute dans une portion de la page présentée à l'utilisateur, comme n'importe quelle application



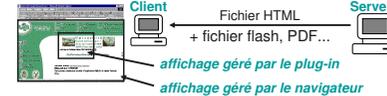
Plug-in

Principe

- étendre les possibilités du navigateur pour présenter des éléments HTML (décrits par la balise `OBJECT`) non gérés par le navigateur
- le navigateur donne toute ou une portion de la page Web à un programme « branché » (plug-in) qui prend en charge l'exécution ou l'affichage (cf. applet JAVA)
- S'il n'est pas disponible, on peut en général le télécharger et l'installer

Exemples

- Flash / Shockwave : documents multimédias
- Real Audio / Vidéo : sons ou vidéos
- Acrobat : documents PDF



Qu'est-ce que naviguer sur le Web ?

Navigation proprement dite

- aller directement à une page, aller d'une page à une autre en suivant un lien
- remplir et envoyer des formulaires
 - attention à la communication de données personnelles
 - l'échange de données peut être sécurisé (ex. paiement électronique par Carte Bleue)

Activités liées à la navigation

- télécharger des fichiers
- arrêter des chargements trop longs (bouton Stop), recharger une page qui a été modifiée (bouton Reload)
- mener plusieurs navigations en parallèle (dans plusieurs fenêtres)
- définir et organiser ses signets
- envoyer des courriers électroniques en cliquant sur des liens

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
 - création du W3C pour normaliser le Web
- 1994-95
 - 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)
- 2005 : Google prétend indexer 8 milliards de pages

Le Web : une définition étendue

World Wide Web (WWW, Toile)

Définition

- ensemble des données disponibles sur les serveurs accessibles sur le réseau Internet
- pouvant être visualisées et/ou utilisées à partir d'un navigateur Web

Attention : Web ≠ Internet

- Internet = support de communication général
- Web = une partie des contenus circulant sur l'Internet

Droit du Web - Les états et le Web

Surveillance

- police, services secrets

Le point de vue de la justice (en France)

- la loi qui s'applique est celle du pays où un contenu est lu et/ou fourni
- responsabilité de la personne qui crée et diffuse un contenu
- les FAI ne sont responsables des contenus délictueux que s'ils contribuent à les créer, ou refusent de collaborer avec la justice
 - le FAI accède à toute demande de renseignements de la Justice

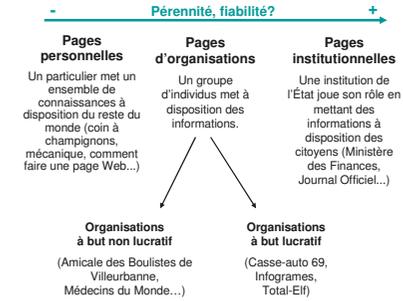
Droit du Web - Quelques règles de bon sens

- ❑ Toute activité informatique peut être enregistrée
 - l'anonymat n'existe pas
- ❑ Mettre à disposition une page Web, c'est la publier (≠ courrier électronique)
 - le droit s'applique
 - diffamation, respect de la vie privée
 - droits d'auteur (ex. images) : demander l'autorisation des ayants-droit avant d'utiliser

Rechercher des informations ?

- ❑ Nous passons notre temps à le faire
 - exemple : s'inscrire à la fac
 - trouver les formulaires, les bureaux...
- ❑ On fait plus ou moins confiance aux informations
 - suivant les enjeux
 - plus ou moins grand besoin de fiabilité
 - suivant la source de l'information (confiance)
 - document officiel
 - prospectus publicitaire
 - extrait de conversation entendue...

Sources d'information sur Internet



Différents types de documents sur le Web

- ❑ Pages Web personnelles
 - document libre
- ❑ FAQ
 - Foire Aux Questions, Frequently Asked Questions
 - ensemble de questions/réponses classiques sur des sujets très variés (science, vie pratique...)
 - lié à des communautés
- ❑ How-to
 - description de la manière de faire une certaine action (ex. comment installer un logiciel)
- ❑ Articles (cf. journaux, documents d'information)
 - forme ± standard et figée (auteur, titre, date...)
- ❑ Tableaux synoptiques, listes
 - réponse à une requête (offres d'emplois...)
- ❑ Cartes et plans

Confiance en l'information disponible sur le Web

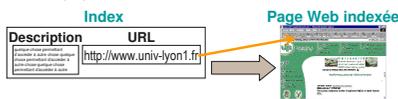
- ❑ Quelle information sur le Web ?
 - à peu près tout, sur tout...
 - ...et son contraire (fiabilité discutable)
- ❑ Il faut analyser l'information en fonction
 - de sa source
 - de sa forme (type de document)
 - de la confiance que vous lui accordez
- ❑ Questions
 - auteur identifié ?
 - auteur identifiable ?
 - institution ?
 - importance de la validité (puis-je utiliser l'information comme preuve) ?
 - date de rédaction ?
 - durée de validité de l'information ?

Recherche d'information sur le Web

- ❑ Recherche d'information
 - navigation simple (plus ou moins au hasard)
 - en accédant directement à des URL connues
 - en naviguant de lien en lien
 - assistance à la recherche : outils
 - annuaires
 - moteurs de recherche
 - portails

Indexation

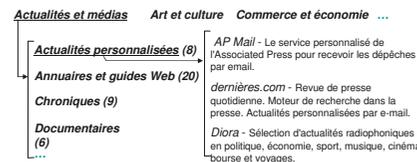
- ❑ Index
 - quelque chose permettant d'accéder à autre chose
 - marque-page, table des matières
 - ex. des bibliothèques : Auteur, Titre, ISBN, Mot-clés...
- ❑ Sur le Web, un index c'est
 - une description du site
 - ex. « Ce site est le site officiel de l'UCBL, il présente les diverses composantes de la faculté... »
 - un identifiant qui permet l'accès au site : son URL



- Problème
 - sur le Web, comment indexer (décrire) des documents pour aider un utilisateur à les trouver ?

Annuaires Web - Principe

- ❑ Principe
 - des documentalistes professionnels indexent une partie du Web
 - classification des sites
 - ❖ créer une hiérarchie (arborescence) de thèmes de plus en plus spécifiques
 - description de documents en quelques lignes



Annuaires Web - Exemples

Lycos Home > Actualité, Médias > Télévision > Émissions > Documentaires et émissions scientifiques

Sites web : sélectionnés et commentés par Claude...

- **E=Mc²** - Possibilité de visionner de nombreux reportages de cette émission, certains bénéficient d'une fiche technique un peu détaillée. Lien vers le site de E=Mc² spécial qui fournit de nombreuses informations et des vidéos sur un thème particulier.
- **Ushuaia** - Très beau site proposant de nombreuses photos et des reportages pour découvrir la nature aux quatre coins du monde des carnets de voyage, des musiques, des économiseurs d'écran...
- **Archimède** - La vulgarisation scientifique selon Arta. On retrouve sur le site des émissions, des liens pour approfondir les sujets, des forums de discussion, des carnets de la revue « Sciences actualités ».
- **Canal + Docs** - Les horaires de diffusion et résumés des documentaires de Canal +.

Internet: Actualités, Recherche, Jeux, Jeux de rôle, Jeux vidéo, Loisirs: Collections, Humour, Tourisme, Maison: Récupération, Jardinage, Informatique: Ordinateurs, Logiciel, Références: Bibliothèques, Langues...

1. Introduction 1. Considérations générales
2. Langage HTML 2. Moteurs de recherche
3. Navigation 3. Portails
4. Recherche d'information 4. Avenir

Indexation « en texte intégral »

- Principe
 - un texte est un ensemble de mots
 - ces mots peuvent décrire le texte
 - donc un texte contient ses propres index
- Application
 - chercher automatiquement dans un texte les mots-clés qui vont le décrire, puis les stocker comme index du texte
- Traitements statistiques
 - chercher les mots les plus pertinents (ceux qui n'apparaissent pas trop souvent)
 - éliminer les « mots vides » (est, au, le...)

PCI 2005 – CM 03 World wide web 46

1. Introduction 1. Considérations générales
2. Langage HTML 2. Moteurs de recherche
3. Navigation 3. Portails
4. Recherche d'information 4. Avenir

Définitions

- Principe
 - des « robots » (logiciels spécifiques) parcourent le Web, lisent le texte des pages, les indexent automatiquement et stockent les résultats
- Moteur de recherche
 - serveur Web permettant de rechercher dans les index mis en place
- Interrogation
 - requête : ensemble de mots fournis au moteur grâce à un formulaire pour l'interroger
- Réponse
 - ensemble de pages Web répondant à la requête, classées par pertinence décroissante (ex. nombre de mots-clés justes)
- Cycle de recherche
 - si trop de réponses, ou réponses non satisfaisantes, modifier la requête pour affiner les résultats

PCI 2005 – CM 03 World wide web 47

1. Introduction 1. Considérations générales
2. Langage HTML 2. Moteurs de recherche
3. Navigation 3. Portails
4. Recherche d'information 4. Avenir

Syntaxe des requêtes (1/2)

- OU
 - certains des termes **doivent** faire partie des index des documents trouvés
 - l'ordre des réponses dépend du nombre de termes trouvés
 - syntaxe
 - recherche OR information
- ET
 - les termes doivent **tous** faire partie des index des documents trouvés
 - syntaxes
 - +recherche +information
 - recherche AND information
 - utilisé par défaut par la majorité des moteurs
- NON
 - le terme **ne doit pas** faire partie des index des documents trouvés
 - syntaxes
 - internet
 - NOT internet

PCI 2005 – CM 03 World wide web 48

1. Introduction 1. Considérations générales
2. Langage HTML 2. Moteurs de recherche
3. Navigation 3. Portails
4. Recherche d'information 4. Avenir

Syntaxe des requêtes (2/2)

- Recherche avancée
 - pour préciser la requête
 - pour limiter les domaines de recherche (sites, pays, dates de modification...)
- Accents / majuscules
 - ete permet de retrouver été, Eté, ETE...
 - Ete permet de retrouver Eté mais pas été, ni eté
 - été permet de retrouver été mais pas ete
- Mots composés, morceaux de phrases
 - "guy marchand" ne permet pas de retrouver une page avec le texte : « Guy Béart, en visite chez le marchand de guitares, acheta des cordes ».

PCI 2005 – CM 03 World wide web 49

1. Introduction 1. Considérations générales
2. Langage HTML 2. Moteurs de recherche
3. Navigation 3. Portails
4. Recherche d'information 4. Avenir

Autres indexations

- Indexation utilisant les liens entre pages
 - plus il y a de liens vers une page, plus celle-ci est considérée comme importante (« vote »)
 - description d'une page en cherchant les mots-clés dans les pages qui pointent vers elle
- Utilisation du contexte
 - indexer des documents non textuels (vidéo, son, image) avec les mots-clé trouvés dans les pages textuelles alentour
- Indexation « par le contenu »
 - chercher des images qui « ressemblent » à d'autres images
 - taux de couleurs, textures, formes...
 - pas encore au point

PCI 2005 – CM 03 World wide web 50

1. Introduction 1. Considérations générales
2. Langage HTML 2. Moteurs de recherche
3. Navigation 3. Portails
4. Recherche d'information 4. Avenir

Qu'est-ce qu'un portail ?

- 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, volia)
 - thématique (santé, enfants)...
 - exemples de services : moteur de recherche, annuaires, météo, informations, jeux, publicité, accès courrier électronique, forums, discussion...
- Principe financier
 - fournir un nombre de services suffisants pour que le portail devienne l'accès principal au Web de certains utilisateurs
- Nécessite une appréciation critique pour ne pas rester bloqué dans la sous-partie du Web accessible depuis le portail

PCI 2005 – CM 03 World wide web 51

1. Introduction 1. Considérations générales
2. Langage HTML 2. Moteurs de recherche
3. Navigation 3. Portails
4. Recherche d'information 4. Avenir

Exemple : edu.univ-lyon1.fr

EDU LE PORTAIL ETUDIANTS DE LYON 1 Université Claude Bernard

PCI 2005 – CM 03 World wide web 52

1. Introduction 1. Considérations générales
2. Langage HTML 2. Moteurs de recherche
3. Navigation 3. Portails
4. Recherche d'information 4. Avenir

Quelle problématique ?

- Partage d'informations entre utilisateurs et communautés
 - exemple : partager ses favoris, ses annotations, accéder à ceux des autres
- Agents de recherche d'information
 - robots qui cherchent pour vous sur Internet
- Visualisation de l'information : cartes
 - représenter l'organisation des sources d'information de la meilleure façon possible

PCI 2005 – CM 03 World wide web 53

1. Introduction 1. Considérations générales
2. Langage HTML 2. Moteurs de recherche
3. Navigation 3. Portails
4. Recherche d'information 4. Avenir

Visualisation de l'information

PCI 2005 – CM 03 World wide web 54