

---

# Bureautique

## Introduction à Internet et à la recherche d'information



Tarek Melliti  
Laboratoire IBISC  
(Informatique Biologie Intégrative et Systèmes Complexes)  
tarek.melliti@ibisc.univ-evry.fr

---

Bureautique – Introduction à Internet et à la recherche d'information

### Plan du cours

---

- Séances
  - 1 : Introduction à l'informatique
  - 2 : Internet et la recherche d'information
  - 3 & 4 : Traitement de texte (Word)
  - 5 & 6 : HTML
  - 7 : Présentation (Powerpoint)
  - 8 & 9 : Tableur (Excel)
  - 10, 11, 12 : Soutenances
  - 13 : Bilan
- Fil rouge
  - Découverte de l'informatique
  - Elaboration d'un projet professionnel

---

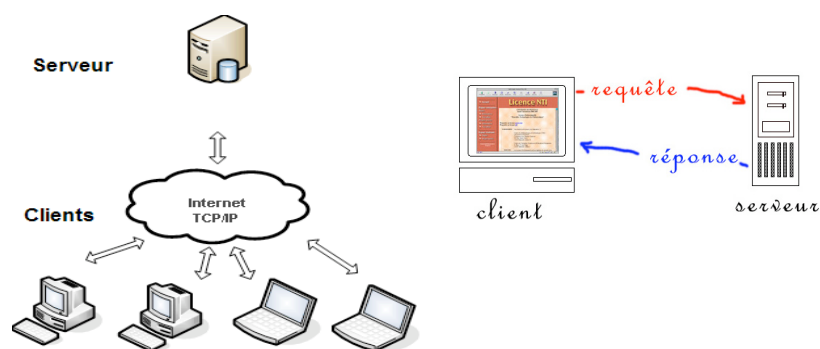
Bureautique – Introduction à Internet et à la recherche d'information

## Plan de la séance

- Introduction : Internet & le Web
  - ✓ Définition, Historique, survol
- Méthodes d'accès à l'information sur le Web
  - ✓ Thématique
  - ✓ Moteurs de recherche
  - ✓ Moteurs de recherche spécialisés
- Evaluation des méthodes d'accès
- Le Web invisible

Bureautique – Introduction à Internet et à la recherche d'information

## Internet : architecture Client/Serveur



Bureautique – Introduction à Internet et à la recherche d'information

## Un peu d'Histoire

---

- Les années 60 : **contexte de guerre froide**
  - ✓ ARPA (Advanced Research Project Agency) : l'idée est d'avoir des paquets passant de noeud en noeud ayant la même autorité sans système central
- Les années 70/80 : **Arpanet et le protocole TCP/IP**
  - ✓ Arpanet relie 4 instituts universitaires
  - ✓ TCP/IP est utilisé pour relier divers réseaux à Arpanet
- Les années 80/90 : **d'Internet au WWW**
  - ✓ **Tim Berners-Lee** (CERN) propose de créer un ensemble de documents reliés les uns aux autres par des liens hypertextes.
- A partir des années 90
  - ✓ le WWW devient le service le plus important sur Internet

---

Bureautique – Introduction à Internet et à la recherche d'information

## Les services d'Internet

---

- L'Internet est un ensemble de réseaux de toutes tailles composés d'ordinateurs hétérogènes interconnectés en utilisant un langage commun appelé **TCP/IP** (Transmission Control Protocol/Internet Protocol)
- Les protocoles de communication de la famille TCP/IP sont chacun spécialisés dans un service particulier de l'Internet (par exemple le **World Wide Web (HTTP)**, la messagerie électronique (SMTP) , l'IRC [le chat], le transfert de fichiers (FTP), les groupes de nouvelles (Newsgroups), telnet, ...)

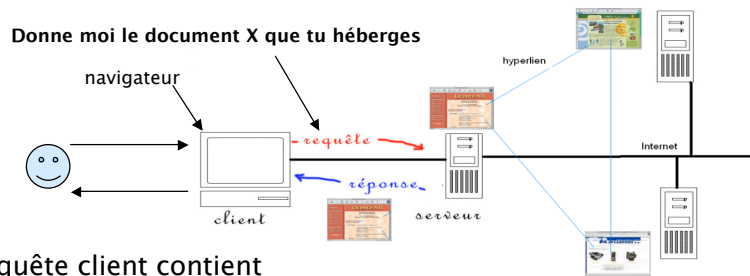
---

Bureautique – Introduction à Internet et à la recherche d'information

## Qu'est ce que le Web alors???

- Définition simple :

« Le Web est un réseau de documents reliés entre eux par des liens hypertextes. Chaque page est située physiquement sur un noeud (serveur) du réseau mondial Internet »



- Une requête client contient

- un nom de méthode
- l'adresse d'un document
- le protocole utilisé

GET http://www.ibisc.univ-evry.fr/~hutzler/L1info/index.html HTTP/1.0

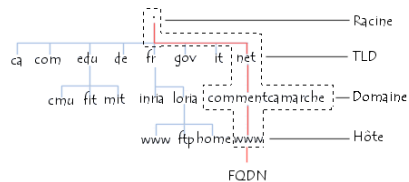
Bureautique – Introduction à Internet et à la recherche d'information

## Requête : Comprendre l'URL d'un site

- Une adresse URL : Uniform Resource Locator

protocole://chemin du hôte/chemin relatif de la ressource

- ✓ Protocole : http, ftp, gopher
- ✓ Ordinateur : Nom hôte (domaine et sous domaine)  
Ex. domaines : .com commercial, .fr en France, .org ONG



- Exemple :

<http://www.ibisc.univ-evry.fr/~hutzler/L1info/contact.html>

hôte ibisc sous domaine univ-evry sous www de France / Répertoire hutzler/L1info/  
document contact.html

Bureautique – Introduction à Internet et à la recherche d'information

## Réponse : un objet multimedia

Que trouve-t-on sur le Web ?

- De tout et n'importe quoi :
  - ✓ Les articles scientifiques
  - ✓ Les recherches en cours dans le Laboratoire Informatique Biologie Intégrative et Systèmes Complexes de l'Université d'Evry
  - ✓ Des dictionnaires et des encyclopédies
  - ✓ L'adresse e-mail de votre professeur
  - ✓ Les journaux de tous pays
  - ✓ La radio ou la TV en direct
  - ✓ Les résultats sportifs en direct
  - ✓ De la musique, des films,...
  - ✓ Des jeux en ligne
  - ✓ ....



Bureautique – Introduction à Internet et à la recherche d'information

## Documents et formats

- Sur l'Internet, on peut trouver tout type de document :
  - ✓ du texte,
  - ✓ des images simples, animées ou en 3D
  - ✓ du son
  - ✓ des vidéos
  - ✓ des programmes
  - ✓ ...
- Les navigateurs récents intègrent désormais de nombreux "plug-ins" permettant de visualiser certains documents.

Bureautique – Introduction à Internet et à la recherche d'information

## Pour écrire une page sur le Web

---

- Format HTML (Hypertext Markup Language)
  - Interprété par tous les navigateurs
- HTML permet en particulier d'insérer des hyperliens dans du texte, donc de créer de l'hypertexte, d'où le nom du langage.
- Balises HTML
  - Entête : Meta-données (Titre, Auteur, Description, etc.)
  - Corps : essentiellement des balises de formatage

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01//EN"
"http://www.w3.org/TR/html4/strict.dtd">
<html lang="fr">
  <head>
    <title>Mon premier document HTML</title>
  </head>
  <body>
    <p>Bonjour tout le monde&nbsp;!</p>
  </body>
</html>
```

---

Bureautique – Introduction à Internet et à la recherche d'information

## Résumons la situation

---

- On a des machines reliées entre elles
  - ✓ Chacune est uniquement identifiable par un nom (adresse URL)
  - ✓ Chacune héberge des documents contenant de l'information destinée à être public
- Aujourd'hui des milliards de pages
- Problématique :
  - ✓ Comment trouver ce que je cherche dans cette masse grandissante d'information
- Solution :
  - ✓ Trouver un moyen d'organiser le contenu du Web
  - ✓ Automatiser le processus de localisation des sites

---

Bureautique – Introduction à Internet et à la recherche d'information

## Caractéristiques de l'information

---

- L'information sur Internet est
  - ✓ destinée aux humain
  - ✓ hétérogène (contenu et format)
  - ✓ instable et éphémère (liens morts, changement d'adresse...)
  - ✓ multilingue
  - ✓ renouvelée continuellement
  - ✓ gratuite ou payante
  - ✓ non structurée (pas un format uniforme)
- L'information n'y est pas toujours fiable (mais là n'est pas la question)

---

Bureautique – Introduction à Internet et à la recherche d'information

## Modes d'accès possible à l'information

---

1. La recherche par navigation en utilisant un répertoire thématique
  - ✓ Classification thématique (figée) manuelle, ressources limitées (intervention humaine irréaliste face à la masse d'information)
2. La recherche par mot-clef en utilisant un moteur de recherche
  - ✓ Indexation automatique, plus exhaustif mais sans classification (totalement automatique)
3. La recherche d'autres types d'information
  - ✓ Moteurs de recherche spécialisés (son, vidéo, etc.)

---

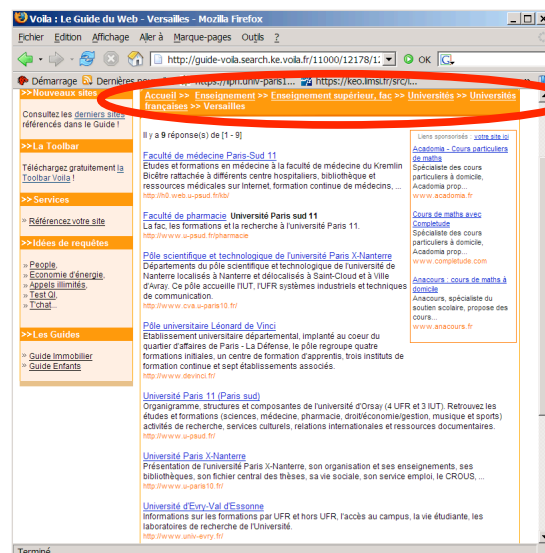
Bureautique – Introduction à Internet et à la recherche d'information

## La recherche par navigation

- Les répertoires sont des ressources classées par sujet dans une base de données hiérarchisée du plus général au plus spécifique
- Utilisation des taxonomies
- Ce type de classement permet à l'utilisateur de préciser progressivement sa recherche en passant successivement d'une sous-catégorie à une autre
- Classification **par des Humains**

Bureautique – Introduction à Internet et à la recherche d'information

## Exemple : voila.fr



Chemin de navigation

Bureautique – Introduction à Internet et à la recherche d'information



## Moteurs de recherche (1)

---

- Un moteur de recherche est divisé en quatre parties:
  - ✓ un programme appelé robot (ou bot) qui scanne les pages pour rentrer les données (les Web crawler)
  - ✓ une base de données pour stocker les informations (à de rares exceptions près, exclusivement des textes)
  - ✓ une structure qui organise ces pages (indexation)
  - ✓ une interface (une page Web) permettant de faire des recherches d'informations dans la base de données
- Il suffit de :
  - ✓ connaître l'adresse du moteur de recherche
  - ✓ savoir ce qu'on cherche !!!!
  - ✓ savoir comment formuler la question en fonction de...

---

Bureautique – Introduction à Internet et à la recherche d'information

## Moteurs de recherche (2)

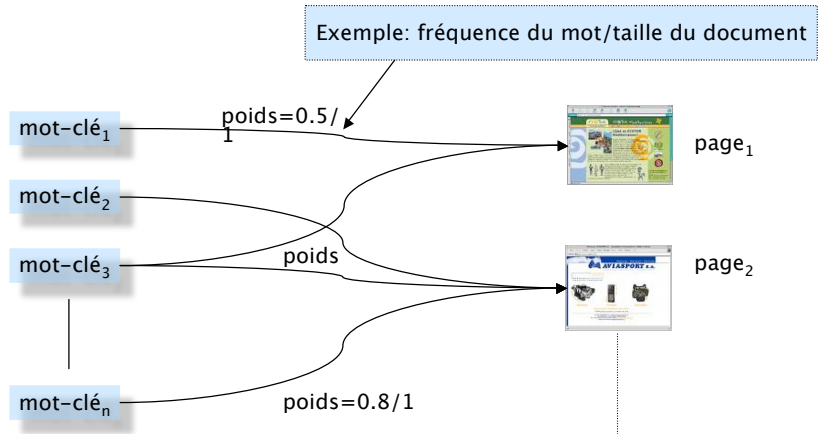
---

- La recherche s'effectue sur les mots contenus dans les ressources elles-mêmes
- Cette recherche affiche les sites correspondant aux termes demandés par le navigateur suivant un certain ordre qui dépend de divers paramètres (importance du site, nom, nombre de fois que le mot est repris dans la page, ...) et dans certains cas les sites qui payent (liens sponsorisés)
- Ce sont les outils à utiliser quand on cherche une information précise qui peut être située sur une page particulière d'un site

---

Bureautique – Introduction à Internet et à la recherche d'information

## Indexation des pages : une vision simple



Bureautique – Introduction à Internet et à la recherche d'information

## Un mot-clé

- Un mot ou composition de mots qui nous renseigne sur le sujet ou le thème du texte
- Quelques indicateurs utilisés :
  - Implicitement:
    - Les mots des titres sont significatifs
    - Les mots en gras sont significatifs
    - L'occurrence des mots
  - Explicitement : les méta-données
- Problème des mots :
  - Polysémie
  - Synonymie

Bureautique – Introduction à Internet et à la recherche d'information

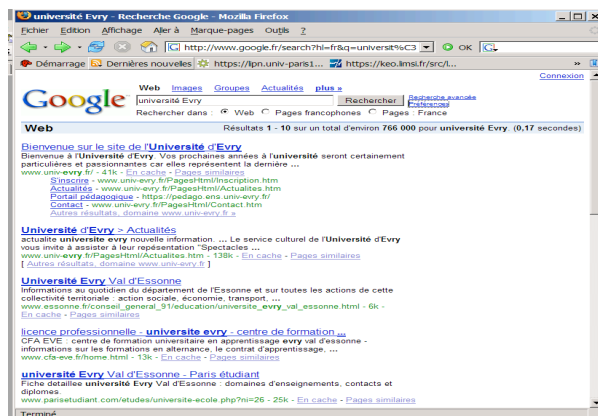
## Recherche par mots-clés

- Pour mieux chercher il faut exprimer les mots-clés en levant le maximum d'ambiguïté.
- Fonctionnalités :
  - Possibilité de recherche avec des opérateurs booléens (AND, OR, NOT)
  - Spécifier le type du document
  - Position des mots dans la page (titre, corp,...)
  - Analyse morphologique, fondée sur la seule reconnaissance de la forme des mots.
  - Possibilité de recherche par groupes de mots
    - Exemple : "traitement de l'information"
  - Possibilité de recherche par troncation
    - Exemple : Etudiant\* --> Etudiant, Etudiante, Etudiants, Etudiantes

Bureautique – Introduction à Internet et à la recherche d'information

## Exemples de moteur de recherche

- Google
- Lycos
- Altavista
- Yahoo



- Remarque : plusieurs portails permettent la recherche par mots-clés ainsi que la navigation thématique, exemple : Yahoo, Voila, etc.

Bureautique – Introduction à Internet et à la recherche d'information

## Exemple sur les fonctionnalités : Google

- Les opérateurs
  - ✓ Je veux les universités de Versailles autre que Evry :  
*université AND versailles -evry*
  - ✓ Je cherche melliti ou hutzler  
*melliti OR hutzler*
- Emplacement dans la page
  - ✓ les pages où "informatique" apparaît dans le nom du site  
*site: informatique*
  - ✓ les pages où l'informatique apparaît dans le titre  
*allintitle: informatique*
- Type de la ressource
  - ✓ Je cherche les documents de type pdf sur linux  
*Linux filetype:pdf*
  - ✓ Je cherche les définitions sur linux  
*define:Linux*

Bureautique – Introduction à Internet et à la recherche d'information

## Classement des réponses : Google

- PageRank
  - Indice de popularité d'une page Web en fonction
    - des liens externes (**popularité de liens**) pointant vers elle
    - des liens qu'elle fait vers elle-même (liens internes)
  - Pb : bombardement Google
    - Créer des sites et référencer un site en utilisant un texte qui n'a rien à voir
    - Ex : tapez "Iznogoud" ou "miserable failure" dans Google
- TrustRank
  - degré de confiance attribué à une page web ou un site web pour le classement des sites
  - note comprise entre
    - 0 (qui équivaut à du spam)
    - 1 (page de confiance)

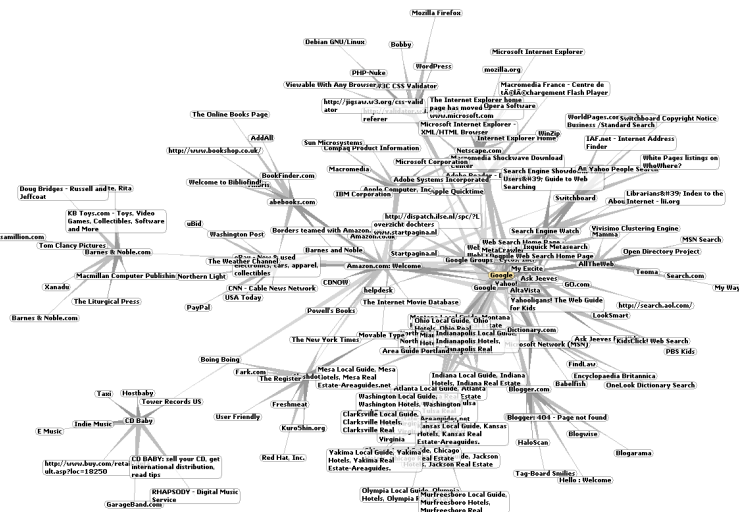
Bureautique – Introduction à Internet et à la recherche d'information

## Google FAQ (Frequently Asked Questions)

- Pourquoi le nbre de réponses varie pour la même requête ?
  - Plusieurs serveurs dispersés dans le monde
  - Pas synchronisés en temps-réel
  - Seul le site de Californie est constamment à jour
- Faut-il se fier au nbre de résultats renvoyés ?
  - Au-delà de 1000, les résultats seraient multipliés par 10 (ou +)

Bureautique – Introduction à Internet et à la recherche d'information

## Le Web autour de Google

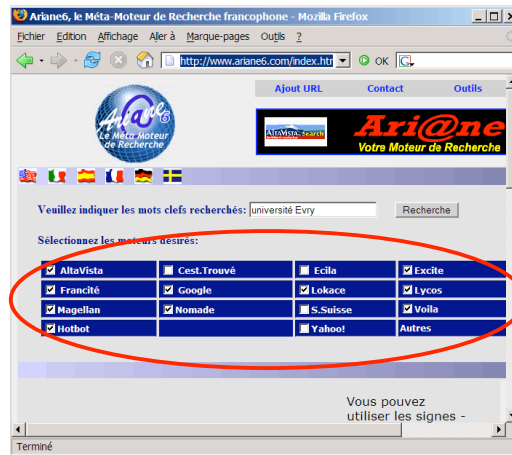


Bureautique – Introduction à Internet et à la recherche d'information

## Méta-moteur de recherche

- C'est un moteur de recherche qui lance la recherche simultanément sur plusieurs moteurs
- ! pas nécessairement plus utile qu'un seul moteur de recherche

trop d'information  
tue l'information



Bureautique – Introduction à Internet et à la recherche d'information

## Moteurs de recherche spécialisés

- Recherche dans le domaine pédagogique
  - Educasource (Ministère de l'éducation nationale, France) : <http://educasources.education.fr/>
- Recherche d'images et de sons
  - Image Search, d'autres moteurs de recherche « généraux » proposent ces fonctionnalités (eg. Google)
- Recherches de logiciels
  - shareware.com
- ...

Bureautique – Introduction à Internet et à la recherche d'information

## Critères d'évaluation des méthodes d'accès

---

- Pertinence des documents retournés
  - ✓ bruit : documents retrouvés non pertinents
  - ✓ silence : documents pertinents non retrouvés
- Temps de réponse
- Interface
  - ✓ conviviale ?
  - ✓ Intuitive ?
  - ✓ Aide en ligne ?
  - ✓ Possibilité d'affiner la requête

---

Bureautique – Introduction à Internet et à la recherche d'information

## Le Web invisible

---

- Généralement les fichiers qui ne sont pas au format HTML
- Les pages Web demandant un login ou restreintes à un domaine particulier
- Les pages dont le contenu est dynamique (créées en réponse à la demande d'un utilisateur par exemple)
- Les sites intranet
- Les pages qui ne sont pas référencées (aucun lien ne pointe vers cette page)
- Les ressources ne provenant pas du Web (site FTP, gopher...)
- ...

---

Bureautique – Introduction à Internet et à la recherche d'information

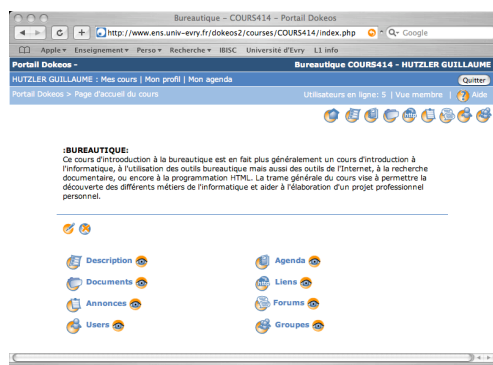
## Annexe

- La signification des messages d'erreurs
  - 400 Bad Request
  - 401 Unauthorized
  - 402 Payment required
  - 403 Forbidden
  - 404 Not Found
  - 500 Internal Server Error

Bureautique – Introduction à Internet et à la recherche d'information

## Dokeos

- Cours de bureautique
  - <http://www.ens.univ-evry.fr/dokeos2/courses/COURS414/>
- L1 Mention informatique
  - <http://www.ens.univ-evry.fr/dokeos2/courses/COURS415/>



Bureautique – Introduction à Internet et à la recherche d'information