

# TP recherche sur Internet

## A. Un petit préalable: qu'est-ce qu'Internet ?

### A.1 Les réseaux informatiques

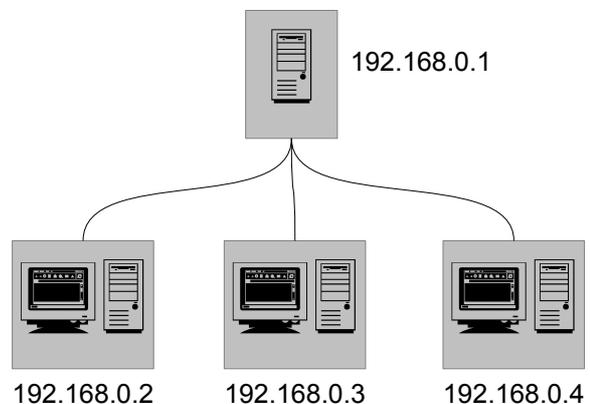
Un réseau informatique (*net* en anglais) est un ensemble d'ordinateurs reliés entre eux par des câbles, des fibres optiques, ou des liaisons radio. On peut le rapprocher du réseau téléphonique, sur lequel il est parfois basé.

Chaque ordinateur est identifié par une adresse unique; sur la plupart des réseaux, il s'agit d'une adresse IP<sup>1</sup>, constituée de 4 nombres (par exemple 192.168.0.1). On peut rapprocher cette adresse IP d'un numéro de téléphone.

- Essayez d'accéder au serveur web 173.194.35.102 . Les noms de domaine ont été inventés plus tard, par commodité. Quel est le nom de domaine associé à 173.194.35.102 ?

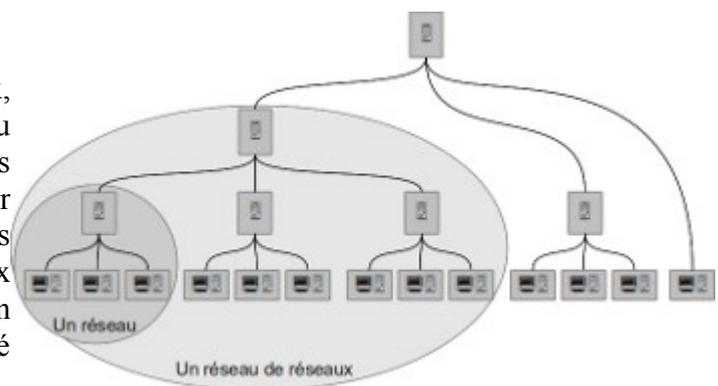
Exemple (ci-contre):

Un réseau de 4 ordinateurs. Ici, l'ordinateur dont l'IP est 192.168.0.1 sert de *routeur* pour les autres.



### A.2 Les réseaux de réseaux

Un *routeur* peut être relié à d'autres réseaux, formant ainsi un *réseau de réseaux*. Ce réseau de réseaux peut lui-même être relié à d'autres réseaux, et ainsi de suite à l'infini. Le premier grand réseau, qui liait quelques universités américaines, a été créé en 1969. Les réseaux ont continué à croître et à s'interconnecter. En 1982, le plus grand de ces réseaux a été baptisé *Internet*.



### A.3 Le web

À partir du langage HTML (créé en 1989) et du protocole HTTP (créé en 1990), Tim Berners Lee et Robert Cailliau, ingénieurs au CERN, vont donner naissance à un ensemble très populaire de pages liées entre elles par des *hyperliens*, qu'on appellera *World Wide Web* (et son contenu *pages Web* ou *sites Web*) à partir de 1991.

- La plus ancienne page du *Web* (13 novembre 1990) est conservée à l'adresse: <http://www.w3.org/History/19921103-hypertext/hypertext/WWW/Link.html>

1 Pour *Internet Protocol*.

## **A.4 Les navigateurs**

Des logiciels sont créés spécialement pour afficher ces pages: les *navigateurs*, par exemple Lynx (1992), Mosaic (1993), Netscape (1994) et Internet Explorer (1995), Opera (1996), Firefox (2004), Chrome (2008)...

## **A.5 Les moteurs de recherche**

Les moteurs de recherche apparaissent peu après le *Web* : W3Catalog (1993), Lycos (1994), AltaVista (1995), Yandex (1997), Google et MSN Search (1998), Baidu et Exaled (2000), Yahoo! Search (2004), DuckDuckGo (2008)...

Le web s'organise en deux espaces: le web de surface (20% du contenu), qui est indexé par les moteurs de recherche, et le web profond (80% du contenu) qui n'est pas accessible par les moteurs de recherche. L'objectif de ce TP est de perfectionner la recherche dans les 20% du web accessibles aux moteurs de recherche, et de faire quelques incursions ciblées dans les 80% inaccessibles de cette manière.

- Pourquoi une partie du web n'est pas accessible aux moteurs de recherche ?

## **B. Préparation du TP**

Lancez un navigateur *Web*.

### **Les favoris**

Les favoris (ou marque-pages, ou encore *bookmarks*) permettent de conserver l'adresse de sites intéressants, de leur donner un nom, et de les organiser.

## **C. Les ressources Internet**

### **C.1 Les formats de documents**

- Lisez la section 1.1 (page 30) du support de cours.
- Pouvez-vous citer des formats de:
  - texte:
  - image:
  - son:

### **C.2 Un thème de recherche**

Pour la suite, choisissez un thème de recherche qui vous intéresse. Par exemple, le sujet de votre dossier.

### **C.3 Les outils de recherche sur Internet**

- Lisez la section 2 du support de cours.
- Faites une recherche à l'aide de quelques moteurs de recherche. Comparez les résultats obtenus avec <http://google.fr> et <http://exalead.fr> . Quelles différences constatez-vous ?

- Cherchez des sites parlant des *langues* sur le **moteur de recherche** <http://google.fr> . Maintenant, essayez sur l'**annuaire** <http://dmoz.org> . Puis sur <https://www.xmarks.com> . Puis sur l'**encyclopédie** <http://wikipédia.fr> . Enfin sur <http://scholar.google.fr> . En quoi les résultats divergent-ils ? Pourquoi ?
  - Google Search :
  
  - Annuaire Dmoz :
  
  - Annuaire Xmarks :
  
  - Wikipédia :
  
  - Google Scholar :
  
- N'oubliez pas de consulter aussi le catalogue en ligne des bibliothèques universitaires (voir page 37 du support de cours). Quelle est la différence avec Google Scholar ?

## **D. Méthodologie de recherche documentaire**

### **D.1 Que cherche-t-on ?**

- Lisez la partie 3.1 du support de cours.
- À l'aide des outils vus en C.3, essayez de trouver des ressources intéressantes pour votre thème de recherche (celui que vous avez défini en C.2).
- Sauvegardez quelques-unes de ces ressources en favoris.

### **D.2 Trouver précisément ce que l'on cherche**

- Lisez la partie 3.2 du support de cours.
- Utilisez le mode avancé et les outils de recherche de Google Search, et essayez d'affiner et d'enrichir votre recherche, de façon à obtenir une pleine page de résultats pertinents (plusieurs pages si c'est trop facile !). Sauvegardez la page de recherche obtenue en favori.

## **E. Évaluation des ressources**

### **E.1 Les noms de domaine**

Les noms de domaines (par exemple google.com) peuvent être utilisés à la place des adresses IP (par exemple 173.194.35.102). Ils sont beaucoup plus faciles à mémoriser.

Les noms de domaine sont constitués de niveaux. Par exemple, le nom de domaine fr.wikipedia.org est constitué de plusieurs parties, qui se lisent de droite à gauche :

- *org* : le nom de domaine de premier niveau (obligatoire),
- *wikipedia* : le nom de domaine de deuxième niveau (théoriquement facultatif, il y en a presque toujours un),
- *fr* : le sous-domaine (facultatif),
- il peut y avoir encore un sous-sous-domaine, un sous-sous-sous-domaine, *etc.* (facultatif).

Le nom de domaine de premier niveau a, par convention (personne ne vérifie), une signification particulière (voir p. 34 du support de cours).

- Trouvez un nom de domaine qui ne respecte pas cette convention.

## **E.2 Lecture d'une URL**

protocole://identifiant:mdp@adresse:port/répertoire/fichier?paramètres#ancrage

- Lisez le support de cours, partie 4.1 (page 33 et suivantes).
- Sans vous rendre sur le site, décomposez ces URL :

[http://www.lemonde.fr/voyage/video/2010/10/21/plongee-dans-le-metro-moscovite\\_1424766\\_3546.html](http://www.lemonde.fr/voyage/video/2010/10/21/plongee-dans-le-metro-moscovite_1424766_3546.html)

<mailto:Achille.Falaise@imag.fr>

<ftp://anonymous@ftp.rfc-editor.org/in-notes/rfc2396.txt>

## **E.3 Licence et référence**

- Imaginons que vous ayez besoin d'insérer, dans un dossier, une carte montrant la répartition des navigateurs *Web* par pays.
  - Cherchez une telle carte sur le *Web*, qui soit utilisable légalement. Comment la référencer ?
- Lisez le support, parties 4.2, 4.3 et 4.4 .
- Consultez le document suivant, et cherchez à l'identifier précisément (auteur, lieu, type d'info, date, organisation, objectif). Citez correctement ce document (voir les exemples page 36 du support de cours).

[http://www.tlfg.ulaval.ca/AXL/francophonie/HIST\\_FR\\_s1\\_Expansion-romaine.htm](http://www.tlfg.ulaval.ca/AXL/francophonie/HIST_FR_s1_Expansion-romaine.htm)

## **F. Le plagiat**

- Essayez de trouver dans un moteur de recherche spécialisé, par exemple <http://news.google.fr> , des cas de plagiat à l'université, et les conséquences pour les auteurs de ces plagiats.

## G. Lexicographie avec Google Trends

Google Trends est un outil permettant de visualiser les requêtes effectuées par les utilisateurs de Google Search. Il est intéressant en lexicographie, mais attention, ce n'est pas un outil fait pour cela ! Il ne faut pas perdre de vue qu'il se base sur les *recherches* des utilisateurs, pas sur le contenu des sites Web.

Allez sur le site de Google Trends: <http://www.google.fr/trends/explore>

- Dans quelles régions françaises utilise-t-on le terme *chocolatine* ? Comparez avec *pain au chocolat*.
- Pour *pain au chocolat*, que se passe-t-il en octobre 2012 ? Que pouvez-vous en déduire (attention, il y a un piège...) ?
- Comparez l'utilisation de *tchat* et de *clavardage* en France et au Canada.
- Étudiez les recherche du mot *grippe* dans le temps. Quand ce mot est-il le plus recherché ? Que se passe-t-il en 2009 ?