

Internet et Recherche d'Information

INF0326 - Outils bureautiques, logiciels et Internet

Plan

- Partie 1 : Internet
 - Un bref historique
 - Fonctionnement
 - Configuration de client courriel
- Partie 2 : Recherche d'information
 - Problématique de la recherche d'information
 - Processus de la recherche d'information
 - Moteurs de recherche
 - Opérateurs de recherche
 - Métamoteurs
 - Annuaire
 - Évaluation de l'information trouvée sur le Web

Partie 1 : Internet

Un bref historique

Histoire de l'internet

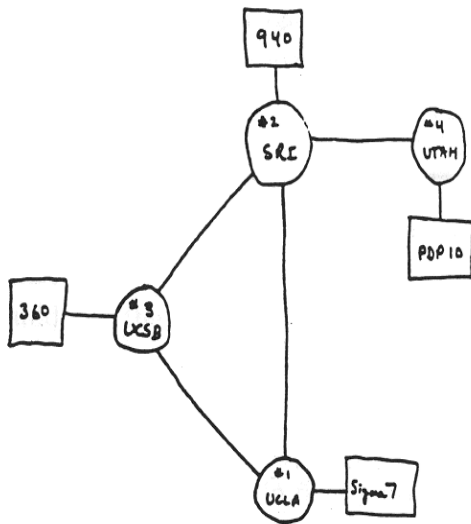
- Eisenhower annonce en 1955 que les États-Unis espèrent lancer un satellite en orbite
- Course contre la Russie
- Création de ARPA (Advanced Research Projects Agency dépendant du Department of Defense)
 - Mission: Appliquer la technologie « state-of-the-art » pour ne plus jamais être surpris par les avancées technologiques de l'ennemi.
- 1962: ARPA ouvre un programme de recherche en informatique avec à sa tête John Licklider.

Histoire de l'internet

- Dans ARPA, Leonard Klienrock développait déjà des idées pour l'envoi d'informations en les brisant en morceaux et en les reconstruisant à l'arrivée.
- 1965: Berkeley et MIT vont créer le premier réseau téléphonique (WAN)
- 1969: ARPAnet
 - Un réseau créé pour lier les principaux centres de recherche américains entre eux pour assurer leur communication en cas d'attaque militaire. Le réseau est décentralisé.
 - Créé par l'ARPA
 - Reliant 4 universités:
 - Le Stanford Institute
 - L'université de Californie à Los Angeles
 - L'université de Californie à Santa Barbara
 - L'université d'Utah
- 1972: 1^{ère} présentation d'Arpanet en conférence
 - Apparition du terme «internetting»

Histoire de l'internet (Suite)

- 1971: Ray Tomlinson envoie le premier mail entre différents hôtes sur ARPAnet
 - Contenu du message: "QWERTYUIOP"
 - Choix du caractère @ pour exprimer "at"
- 1972 : Amélioré par Lawrence G. Roberts avec la première application qui permet :
 - Affichage de liste
 - Archivage
 - Transfert
 - Réponse

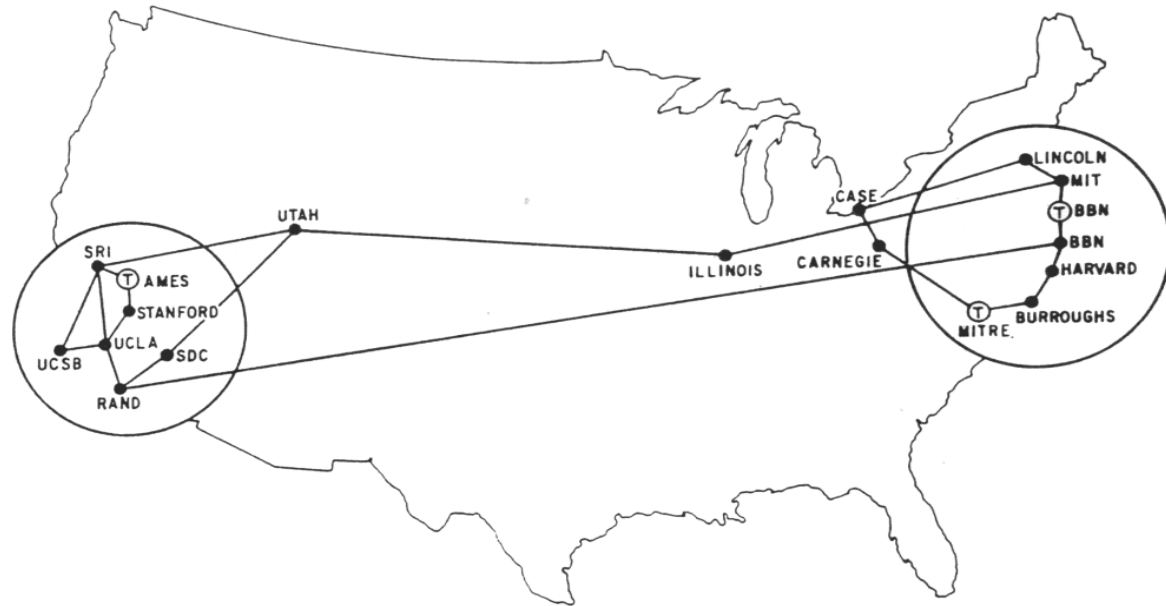


THE ARPA NETWORK

DEC 1969

4 NODES

FIGURE 6.2 Drawing of 4 Node Network
(Courtesy of Alex McKenzie)

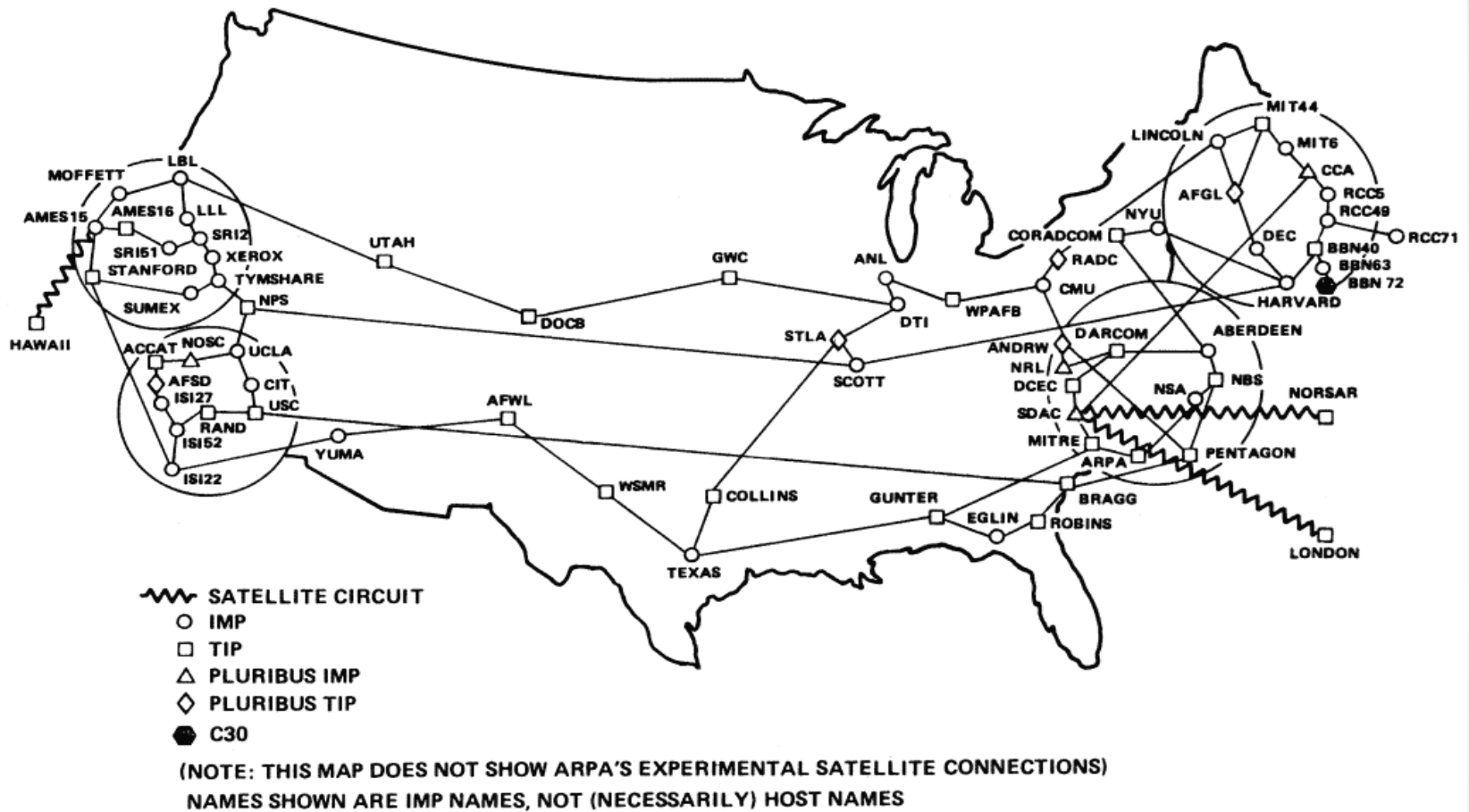


MAP 4 September 1971

ARPANET de 1969 à 1971

source: <http://www.let.leidenuniv.nl/history/ivh/chap2.htm>

ARPANET GEOGRAPHIC MAP, OCTOBER 1980



ARPANET en 1980

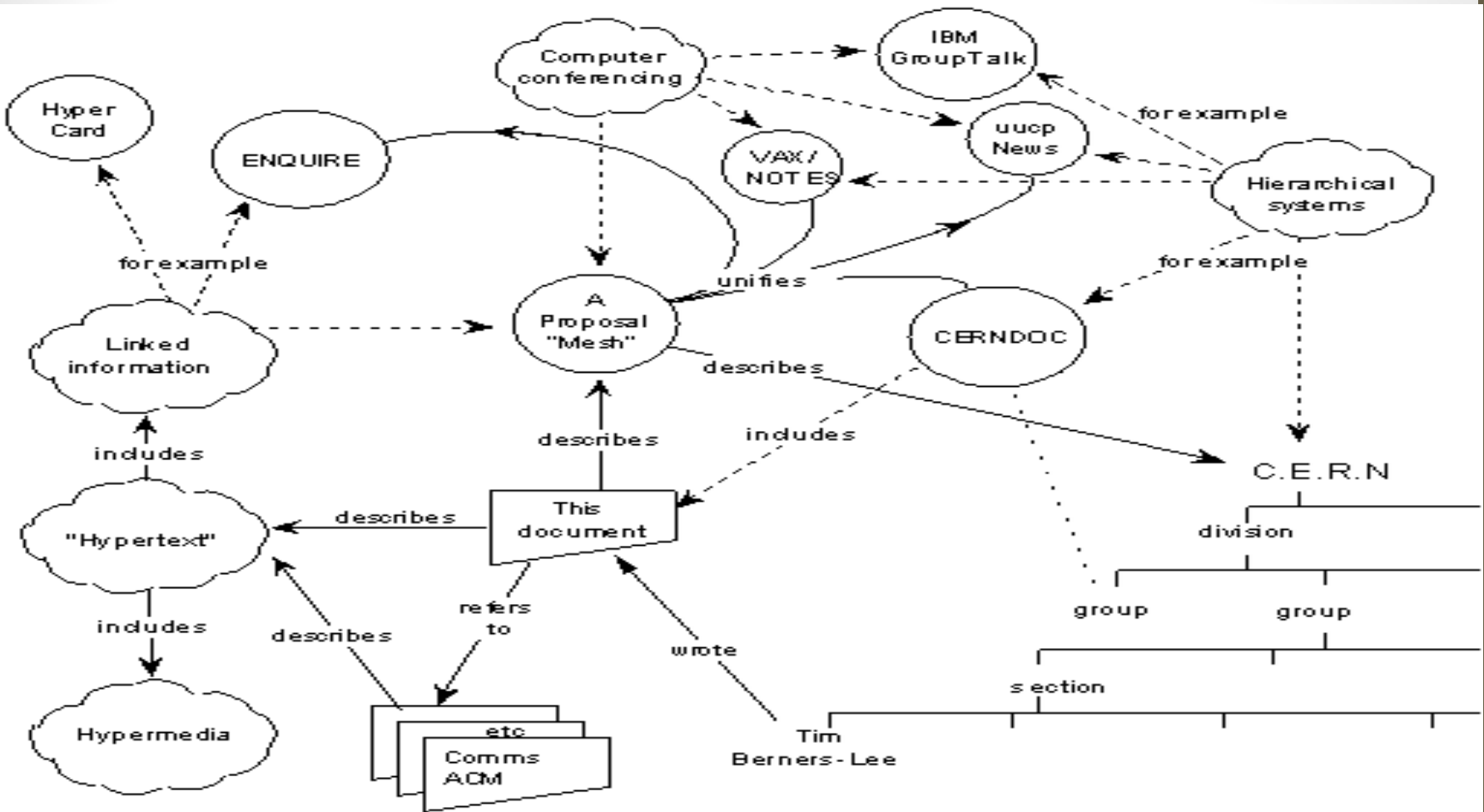
Images de : <http://www.let.leidenuniv.nl/history/ivh/chap2.htm>

Histoire de l'Internet (suite)

- 1983 : Séparation de la partie militaire de ARPANet (MILNet) pour assurer la confidentialité.
- 1986 : Adoption de la forme actuelle de l'Internet (protocole TCP/IP pour les envois de messages entre ordinateurs).
- 1989 : Invention du World Wide Web (WWW).
- 1990-1991 : Premières exploitations commerciales. Plus de machines de l'industrie que du monde académique.
- 1995-1997: Premiers utilisateurs de l'Internet à la maison au Québec.

World Wide Web

- 1980 : Tim Berners-Lee et Robert Caillau
 - Enquire, prédécesseur du World Wide Web
 - Pas adapté à un usage grand public
 - Problème pour le maintenir à jour (non indépendance des cartes entre elles)
- 1989 : World Wide Web (WWW), système de documents hypertextes interconnectés
 - HTTP (Hyper Text Transfer Protocol)
 - langage HTML (HyperText Markup Language)



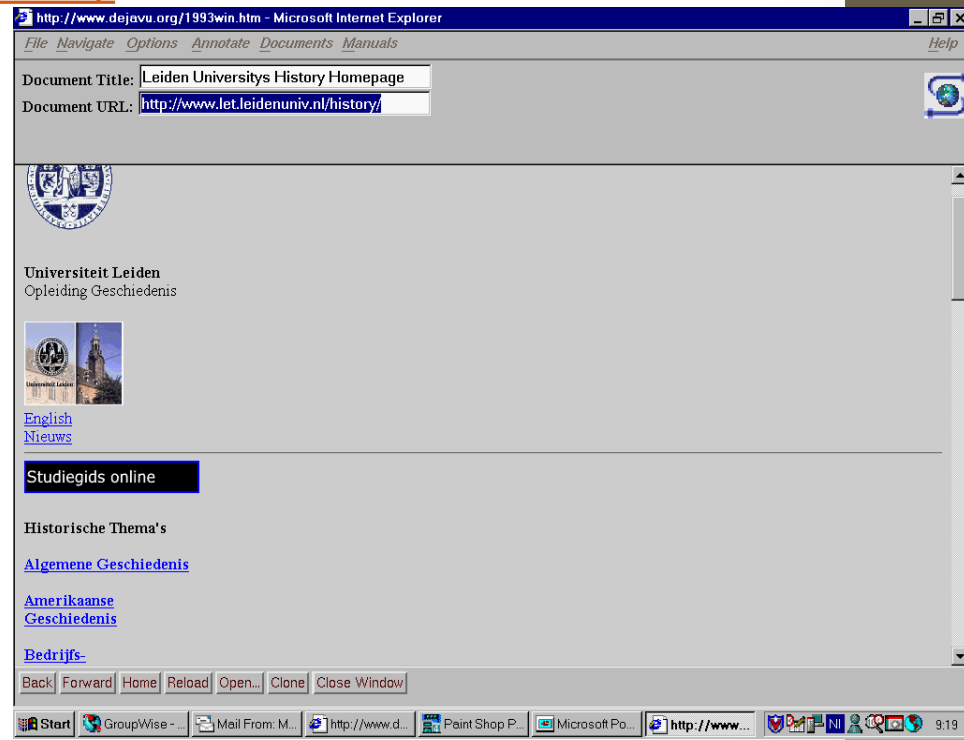
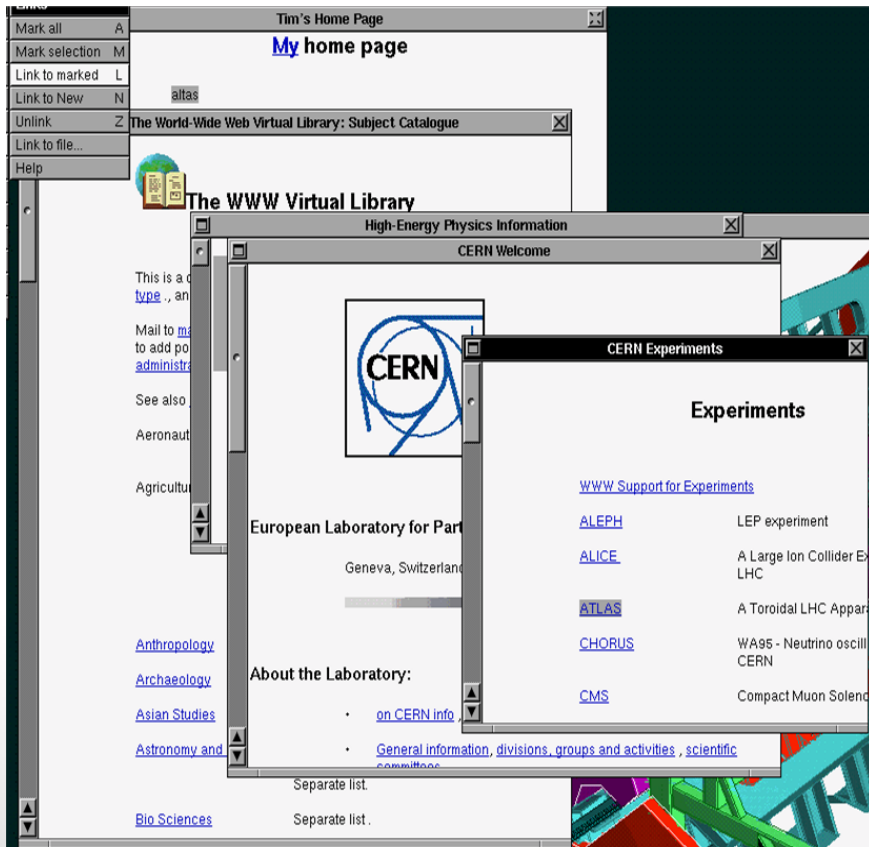
World Wide Web

Proposition de 1989

Image de : <http://www.let.leidenuniv.nl/history/ivh/chap2.htm>

Le Web à ses débuts

WWW : Toile (d'araignée) mondiale: système hypertexte public fonctionnant sur Internet et qui permet de consulter, avec un navigateur, des pages mises en ligne dans des sites. <http://web.archive.org/web/>



En 1994

<http://www.let.leidenuniv.nl/history/ivh/chap2.htm>

Partie 1 : Internet

Fonctionnement

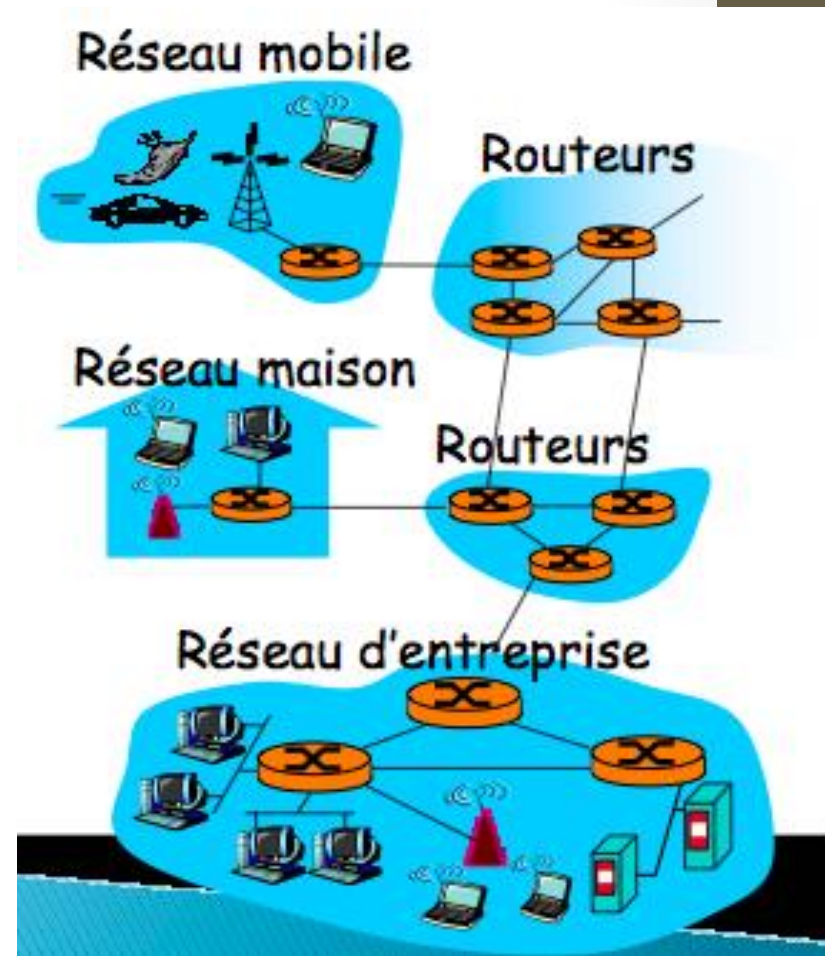
Qu'est ce que l'Internet ?

Internet est un système d'interconnexion de machines qui constitue un réseau informatique mondial, utilisant un ensemble standardisé de protocoles de transfert de données.

Terme d'origine américaine « *Internet* » est dérivé du concept d'*internetting*
en français : « interconnecter des réseaux »

Inter-connection de réseaux

- Millions d'équipements connectés
- Deux types d'ordinateurs
 - **Serveurs:** fournissant des données ou des services
 - **Clients:** cherchent à obtenir des données ou des services

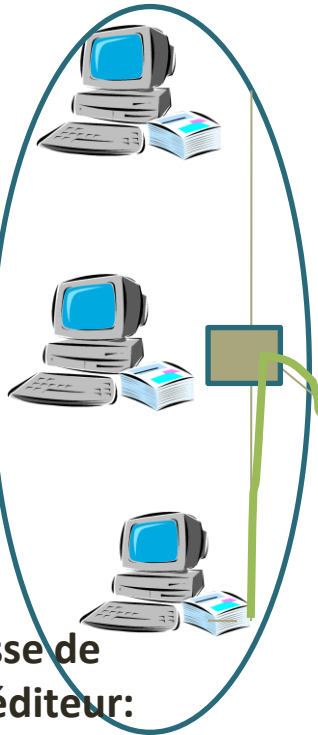


Adresses IP et Routage

- Chaque ordinateur a une adresse: l'**adresse IP**
- Les messages envoyés entre ordinateurs ont une adresse d'émission et de destination
- Des ordinateurs appelés **routeurs** s'occupent de diriger les messages par le chemin le plus court jusqu'au destinataire

Communication entre deux ordinateurs

Réseaux locaux
(ex.: PK-S1585)



Réseaux universitaires
(ex.: UQAM)

Réseaux provinciaux
ex.: RISQ

Réseaux nationaux
ex.: CA*NET

...

...

...

...

Adresse du destinataire:
123.123.123.2



Ordinateur individuel

Adresse de l'expéditeur:
1.22.332.555

Ordinateur individuel



 = Routeur

Serveurs et Services

Serveurs: ordinateurs généralement branchés en permanence sur Internet dont le but est d'offrir des services.

Exemple: les serveurs de uqam.ca envoient les pages Web du site Uqam à votre navigateur Web quand vous consultez le site Web

La plupart des services internet que vous utilisez sont hébergés sur des serveurs

DNS et noms de domaines

Se souvenir d'une adresse IP est difficile

"Je vais lire mes emails sur 74.125.226.248"

-- Personne

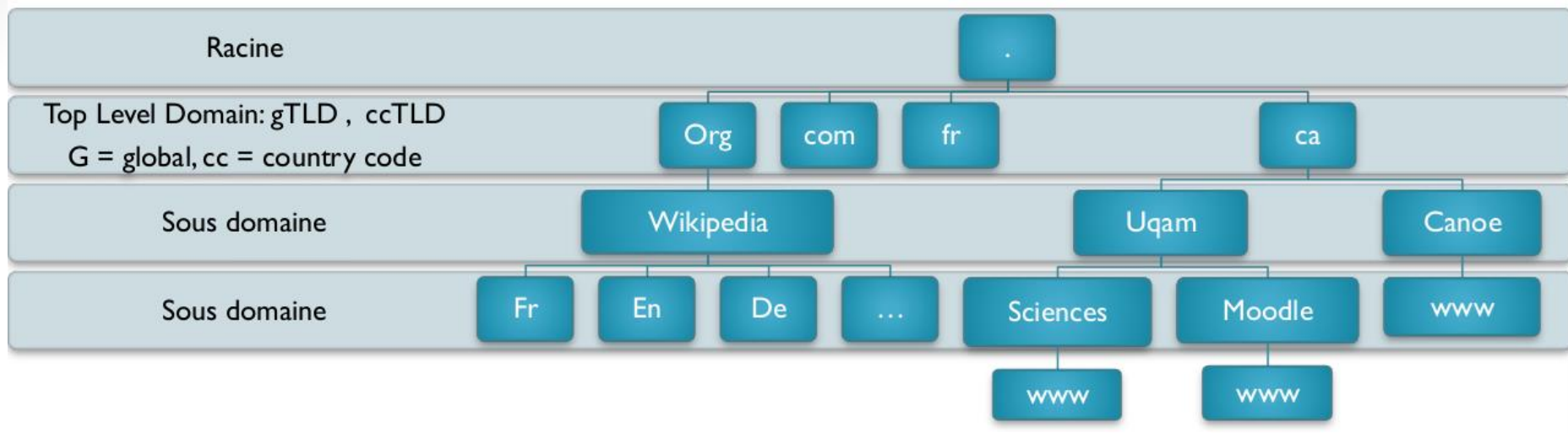
On utilise des noms de domaine pour accéder aux services

74.125.226.248 = google.ca

173.252.110.27 = facebook.com

DNS: Domain Name System

- Système de Noms de Domaines
- 1983 - Paul Mockapetris
- Traduit les noms en adresse IP



[Transmission de paquets sur internet:](https://www.youtube.com/watch?v=xluBmOufbls)

<https://www.youtube.com/watch?v=xluBmOufbls>

Espace de noms hiérarchisés

Exemple de nom de domaine

Exemple: www.sciences.uqam.ca

ca Domaine de premier niveau
(Top Level Domain)

uqam Sous domaine de « ca »

sciences Sous domaine de « uqam »

www Sous domaine de « sciences »

Nouvelles extensions de domaines

<https://iwantmyname.com/domains/new-gtld-domain-extensions>

Navigation Internet

Un **navigateur Web** (ou fureteur) est un logiciel conçu pour consulter le World Wide Web.

Il existe de nombreux navigateurs web, pour toutes sortes de matériels et pour différents systèmes d'exploitation

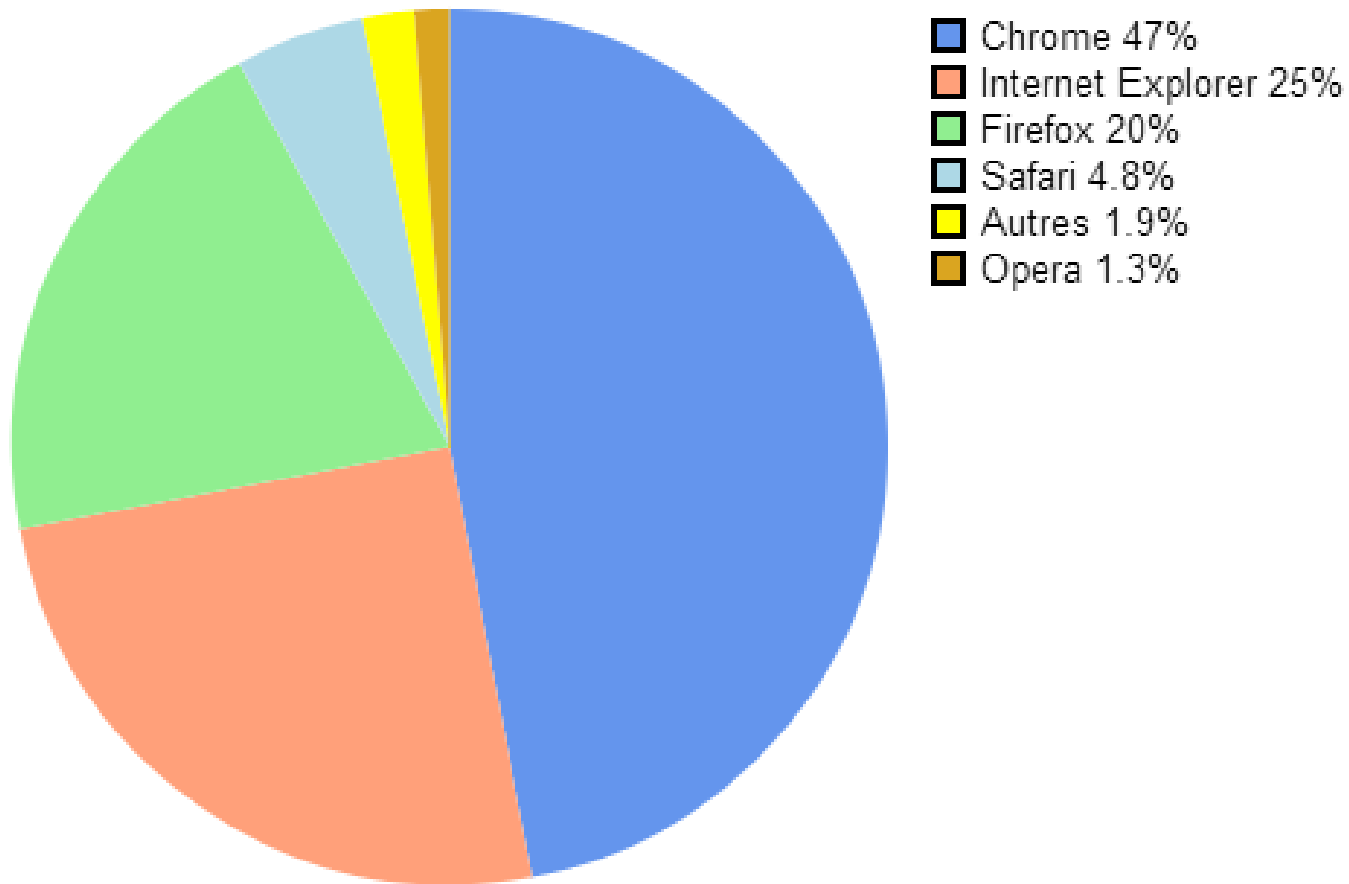


Les plus utilisés : Internet explorer, Firefox, Netscape, Opera, Maxthon, Safari et surtout le nouveau né de google : Google Chrome



Exemple d'utilisation...





Parts de marché des navigateurs

source: <http://statcounter.com/> (période 8/2013 à 12/2013)

Autres utilisations d'Internet

Internet n'est pas seulement le Web!

Il y a d'autres utilisations:

- transmission de courriels
- transfert et partage de fichiers
 - ex.: logiciel FTP, BitTorrent, Emule, etc.
- messagerie instantanée, vidéoconférence
 - ex.: MSN, Skype, GoogleTalk)
- ...

Éléments de sécurité

- Firewall (Pare-feu)

- un logiciel qui a pour fonction de faire respecter la politique de sécurité du poste de travail, celle-ci définissant quels sont les types de communication autorisés ou interdits.
- Firewall tiers : SmoothWall, CoreForce, iSafer, etc.

- Virus informatiques

- un logiciel malveillant conçu pour se propager à d'autres ordinateurs en s'insérant dans des programmes légitimes appelés « hôtes ». Il peut perturber plus ou moins gravement le fonctionnement de l'ordinateur infecté.
- Anti-virus : AVG, ClamAV,

- Spywares (Logiciels espions)

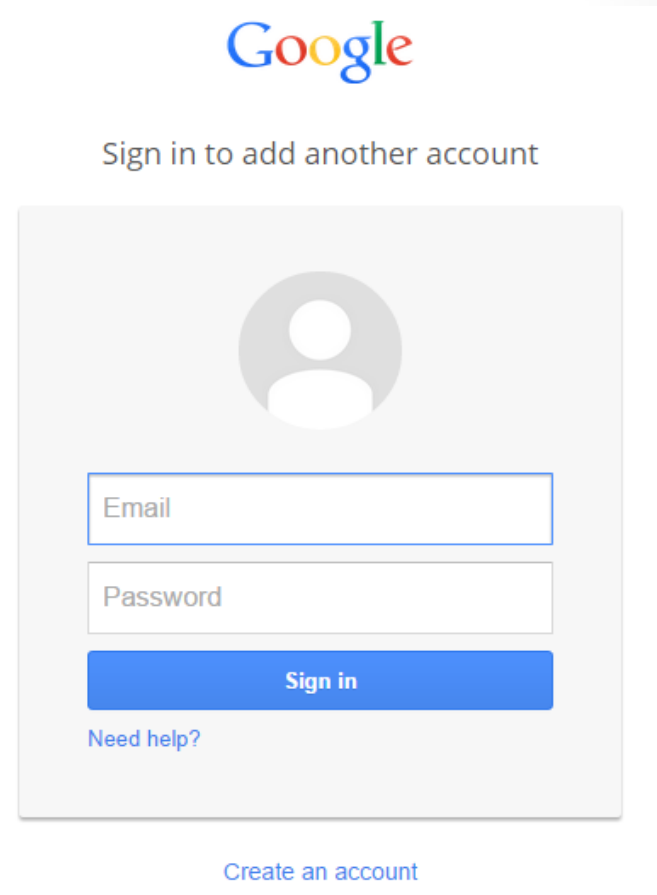
- un logiciel malveillant qui s'installe dans un ordinateur dans le but de collecter et transférer des informations sur l'ordinateur et/ou l'utilisateur.
- Protection: Ad-ware, MS AntiSpyware, HijackThis, etc.

Envoi/Réception de courriels

- Client local – logiciel
 - Outlook
 - Mozilla Thunderbird
 - DreamMail
 - Opera M2
- Client externe sur internet
 - Gmail
 - Yahoo
 - Hotmail
 - Foxmail
 - UQAM
 - ISP

Création d'un compte sur gmail

- <http://accounts.google.com/>



The image shows a screenshot of the Google sign-in interface. At the top, the Google logo is displayed in its multi-colored font. Below the logo, the text "Sign in to add another account" is centered. The main content area is a light gray box containing a circular profile picture placeholder. Below the placeholder are two input fields: "Email" and "Password". A blue "Sign in" button is positioned below the password field. At the bottom left of the gray box, there is a link that says "Need help?". Below the gray box, centered, is a link that says "Create an account".

Envoi/Réception de courriels

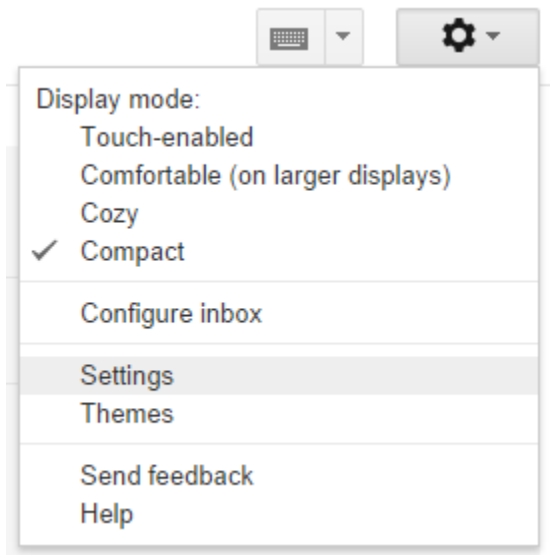
- POP
 - Permet de récupérer votre courrier sur une machine distante quand vous n'êtes pas connecté en permanence à Internet.
 - Bloque votre boîte aux lettres pendant que vous y accédez, ne permettant pas à une autre connexion d'accéder en même temps à votre courrier.
 - L'évolution du courrier électronique vers le multimédia et le manque de flexibilité de POP favorisent l'émergence d'un nouveau protocole : l'IMAP
- IMAP
 - Courrier et répertoires restent sur le serveur.
 - Synchronisation des messages
 - Le logiciel affiche une copie de votre boîte aux lettres, archives comprises. Comme l'original reste sur le serveur, vous pouvez avoir accès à votre courrier à partir de n'importe quel ordinateur.
 - Comme seul les en-têtes des messages sont téléchargés dans un premier temps, l'affichage de la liste des messages est souvent beaucoup plus rapide avec IMAP lorsqu'il y a un grand nombre de (nouveaux) messages sur le serveur.

Configuration d'un logiciel pour réception de courriels

- Configuration d'un client email pour l'UQÀM
 - Mozilla Thunderbird
 - Entrant
 - IMAP: imap.sitel.uqam.ca
 - POP3: pop.sitel.uqam.ca
 - Sortant: client internet (pour Bell: smtphm.sympatico.ca)
 - <http://www.sitel.uqam.ca/etudiants/Pages/AccesAuxServicesDeCourriel.aspx>
- Configuration d'un client email sur internet:
 - Gmail
- Ajout d'un Fil RSS (RSS Feed)
 - <http://www.actualites.uqam.ca/nouvelles.feed>

Ajoutez votre compte UQÀM à votre compte gmail

- Gmail



[General](#) [Labels](#) [Inbox](#) [Accounts and Import](#) [Filters](#) [Forwarding and POP/IMAP](#) [Chat](#) [Web Clips](#)

Partie 2 : Recherche d'informations

Avec Internet

Problématique de la recherche d'information (sur Internet)

Internet offre :

- une surabondance d'information
- de l'information non pertinente
- de l'information de qualité variable
- plusieurs sources
- plusieurs points de vue (pas toujours neutres)
- des moyens plus ou moins efficaces pour chercher l'information
- de l'information souvent non vérifiée

Le processus de la recherche d'information

Plusieurs étapes cruciales :

1. Définir ses besoins
2. Effectuer la recherche
3. Obtenir l'information
4. Évaluer l'information
5. Bien citer l'information trouvée pour éviter le plagiat

Définir ses besoins

Les besoins en information sont établis sur la base de plusieurs facteurs:

- l'utilisation de l'information envisagée
 - ex: TP, compte-rendu, thèse, etc.
- vos attentes et les attentes d'autres gens
 - ex: chargé de cours qui corrigera votre travail
- les délais de réalisation

Il importe donc de bien définir ses besoins et de se fixer des objectifs réalistes pour la recherche!

Effectuer la recherche

- Plusieurs types d'outils de recherche:
 - moteurs de recherche
 - méta-moteurs de recherche
 - annuaires
 - etc.
- Il y a des outils spécialisés et des outils généralistes
 - en fonction des domaines / thèmes
 - en fonction des supports / médias

Les moteurs de recherche

Moteur de recherche: logiciel permettant de retrouver des ressources (pages web, forums Usenet, images, vidéo, fichiers, etc.) associées à des mots clés.

- Ex.: Google, Bing, Yahoo...
- Construisent des bases de données de pages Web et autres documents (ex.: PDF) en utilisant des **robots** qui parcourent l'Internet périodiquement
- **Indexent** les documents en fonction de leur contenu (mots-clés, titres, métadonnées, etc.)
- Ils offrent aux internautes des moyens de chercher dans ces indexes

Les moteurs de recherche (suite)

La façon la plus populaire de chercher sur Internet dans la vie de tous les jours.

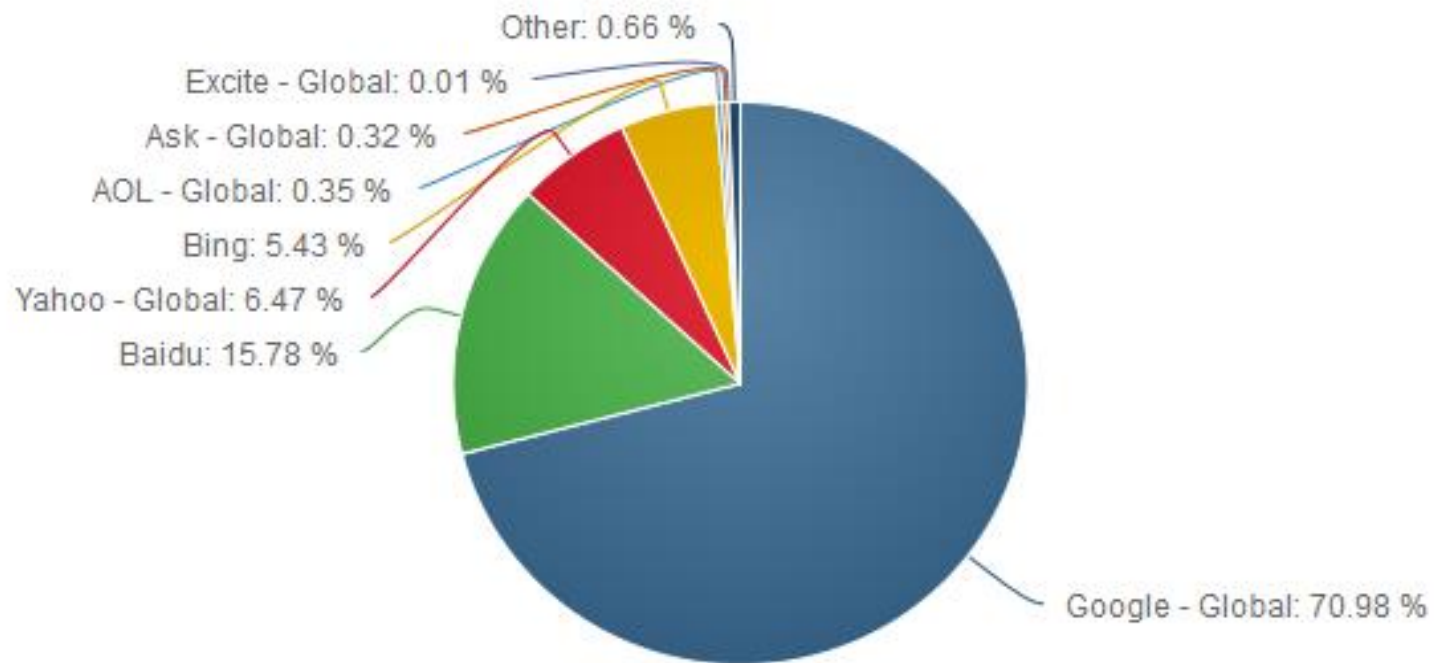
Avantages :

- accéder rapidement à des informations précises
- une énorme quantité d'information est indexée
 - dans certains cas, des milliards de pages
- résultats généralement classés par pertinence

Les moteurs de recherche (suite)

Limites :

- contrôle de l'information faible
 - liens périmés, qualité douteuse, etc.
- beaucoup (trop?) d'information
- les résultats peuvent contenir des synonymes
- résultats parfois décevants parce que l'indexation des pages Web est automatique
 - certains sites trompent les moteurs
- les documents ne sont pas toujours indexés
 - pages dynamiques, accès restreint, demandes d'exclusion...



Statistiques sur les moteurs de recherche

source: <http://www.netmarketshare.com/> (août 2013)

Les opérateurs de recherche

Avec plusieurs moteurs, il est possible d'utiliser des **opérateurs** pour restreindre la recherche.

- en fonction du type de document
- de la source
- ou pour des requêtes plus complexes
 - combiner plusieurs mots
 - utiliser des alternatives
 - éviter certains résultats

Opérateur	Sémantique	Exemple
-	exclure un mot des résultats	chien -chats
" "	trouver une expression exacte	"spectacles extérieurs"
*	compléter une requête	"un * vaut mieux que deux *"
OR ou	soit un mot-clé, soit l'autre	Montréal OR Toronto
AND ou +	un mot-clé et l'autre	Montréal AND Toronto
filetype:	chercher un type de document comme pdf, ppt et doc	filetype:pdf histoire internet
site:	limiter la recherche aux documents d'un site Web	site:cyberpresse.ca canadiens hockey
define:	obtenir la définition d'un mot	define: chiffrier

Principaux opérateurs de recherche sur Google

source: <https://support.google.com/websearch/answer/136861?hl=fr>

Autres fonctions pratiques de Google

Faire des conversions d'unités:

Ex.: 1 cm = ? inches

Faire des opérations mathématiques :

Opérateur	Sémantique	Exemple
+	additionner	300 +57
-	soustraire	100 - 4
/	diviser	25 / 5
*	multiplier	5 * 5
^	exposant	2^5
% of	pourcentage de	45 % of 39

Voir : [Une documentation de la calculatrice Google](#)

D'autres services de recherche de Google

- **Google Images:** pour les images
- **Google Scholar:** documents du milieu universitaire et scientifique
- **Google News:** actualités
- **Google Books:** livres (certains du domaine public, d'autres à accès restreint)
- **Google Translate:** Traduction de page ou de texte en plusieurs langues
- **Youtube:** vidéos

Quelques moteurs de recherche généralistes intéressants

- Microsoft [Bing](#)
- [Yahoo Search](#)
- [Wolfram Alpha](#): opérateurs de recherche complexes
- [Flickr](#): recherche de photos
- [Exalead](#): opérateurs de recherche complexes
- ...

Astuces pour chercher avec les moteurs de recherche

- La plupart des moteurs acceptent les caractères accentués
- Les majuscules ne sont pas toujours prises en compte
- La plupart des moteurs étendent les requêtes avec des synonymes (sauf si “expression”)
- Si trop de résultats, préciser la requête (opérateurs, ajouter des mots, etc.).
- Inversement, s’il n’y a pas assez de résultats, rendre la requête moins précise (enlever des mots, enlever les guillemets, etc.).
- L’ordre des mots-clés a parfois une importance (ex.: Google)
- Chercher les pages par langue ou par pays.

Les métamoteurs

Des moteurs de recherche utilisant les résultats d'autres moteurs.

Deux types :

- fusionnent les résultats
 - Ex.: [MetaCrawler](#), [DogPile](#), [Ixquick](#), [Copernic](#)
- présentent les résultats séparément
 - Ex.: [Bing vs Google](#)

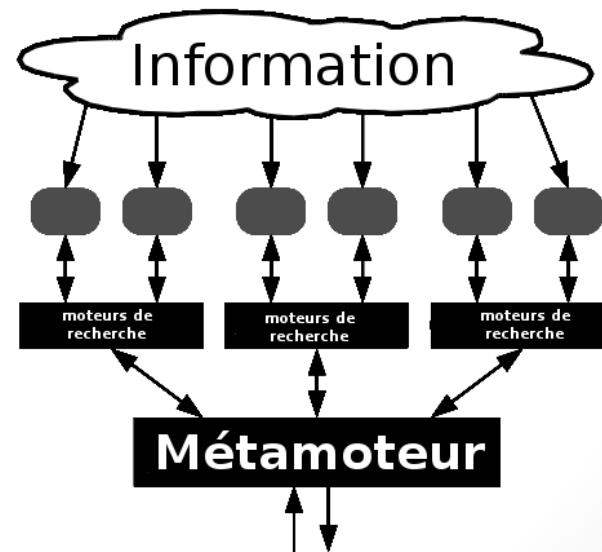
Les métamoteurs

Avantages :

- Gagner du temps en cherchant plusieurs moteurs à la fois
- Une recherche plus exhaustive

Limites :

- Impossible de profiter des particularités de chaque moteur
- Redondance dans les résultats



Les annuaires

Annuaire Web: site proposant une liste classée de sites Web

- arborescence de catégories par précision
- général > spécifique

Différentes caractéristiques :

- spécialisés vs généralistes
- inclusion d'un site par des professionnels, inclusion payante, etc.
- peuvent être accompagnés de description, d'une évaluation, etc.

Liste d'annuaires: <http://www.bibliotheques.uqam.ca/InfoSphere/sciences/module5/repertoires.html>

Les annuaires

Avantages:

- La classification est utile
- Des sites de qualité (en général)

Limites :

- Peuvent contenir des sites inexistant
- Pas toujours à jour
- Ne couvrent pas tout
- Recherche moins efficace
- Catégories imparfaites

Quelques exemples d'autres sites Web

Des exemples:

- [411.ca](#) : numéros de téléphones, adresses
- [Google Maps](#), [Yahoo Maps](#): cartes, trajets en voiture, à pied, transport en commun, etc.
- [Grand dictionnaire terminologique](#): recherche de termes techniques.
- [Infoprix.ca](#): comparateur de prix de pièces informatiques
- [IMDB](#): base de données sur les films
- [HoaxBuster](#): pour vérifier la véracité de rumeurs ou faits surprenants qui circulent sur le Net.
- Cyberpresse: nouvelles et contenus francophones télévisuels
- dictionnaires en ligne
- recherche de billets d'avion, d'hôtels, etc.
- ...

Wikipedia

Une encyclopédie collaborative et libre

- peut être éditée par tous
- contenu libre (pas de droits d'auteurs)
- certains articles sont de très bonne qualité
- parfois: erreurs, omissions, partialité, vandalisme, contenu périmé
- Souvent fiable pour des faits généraux ou sujets populaires
- Peut être plus fiable que des encyclopédies traditionnelles



Comment évaluer l'information trouvée sur le Web?

Deux questions principales :

- le contenu est pertinent?
- le contenu est fiable?

source:

<http://www.bibliotheques.uqam.ca/InfoSphere/sciences/module7/evaluer1a.html#qua>

Évaluer la qualité d'une source d'information

Quelle est la provenance?

- [l'adresse du site Web](#)
- codes de pays: .ca, .fr, ...
- sites généraux: .com, .org, .net, ...
- sites gouvernementaux: .gov.qc.ca, .edu. .mil, ...
- sites reconnus
 - gouvernements, entreprises, organismes internationaux, associations reconnues,...
- blogues, forums et listes de discussion

Est-ce que le site est connu? recommandé?

Évaluer la réputation d'un auteur

- A-t-il d'autres publications?
- Est-ce que l'auteur est cité par d'autres sources qui abordent le même sujet?
 - ex.: index de citations en sciences
- Est-ce qu'il a été recommandé?
- Les coordonnées de l'auteur sont-elles indiquées (incluant son courriel)?
- Quelle est son affiliation, formation et fonction?
- Fournit-il des références bibliographiques?

Évaluer la qualité du contenu

Vérifier si le contenu est objectif:

- est informatif?
- est publicitaire? inclut de la propagande?
- l'auteur peut avoir des motivations cachées ou conflits d'intérêts?
 - ex. compagnie de tabac vs cancer, information boursière
- l'information est biaisée
 - point de vue, information manquante, contradictoire face à d'autres sources, émotive...

Évaluer la qualité du contenu

Vérifier si le contenu est exact :

- bibliographie exhaustive?
- sources citées à jour?
- sources fiables?
- plusieurs points de vue?
- fautes dans le texte?

Vérifier si le contenu est actuel:

- date de création? date de mise à jour?

Utiliser les informations trouvées sans plagier

- Plagiaire: « personne qui s'approprie et copie les œuvres ou les idées d'autrui »
 - source: dictionnaire Antidote
- Une présentation multimédia résumant ce qu'est le plagiat en contexte académique:
<http://www.bibliotheques.uqam.ca/InfoSphere/sciences/module7/citer1.html>

Quiz sur le plagiat

- J'inclus dans mon texte une traduction que j'ai faite d'un paragraphe d'un article de périodique en anglais.
- J'explique dans mes propres mots l'idée émise par un auteur dans son texte sans indiquer qui est l'auteur.
- J'ajoute dans mon travail une image trouvée sur le Web sans dire de quel site elle provient.
- Je remets à mon professeur du cours X le même travail que j'ai remis pour mon cours Y.
- J'affirme dans mon travail que Christophe Colomb a découvert l'Amérique sans citer la source.

Quelques ressources utiles pour l'accès aux services Internet UQAM

Pour accéder hors campus aux banques de données à accès restreint des bibliothèques de l'UQAM (ex.: PubMed).

<http://www.bibliotheques.uqam.ca/acceder-aux-ressources>

Pour utiliser l'Internet sans fil (Wifi) à l'UQAM:

<http://www.sansfil.uqam.ca/>

FIN