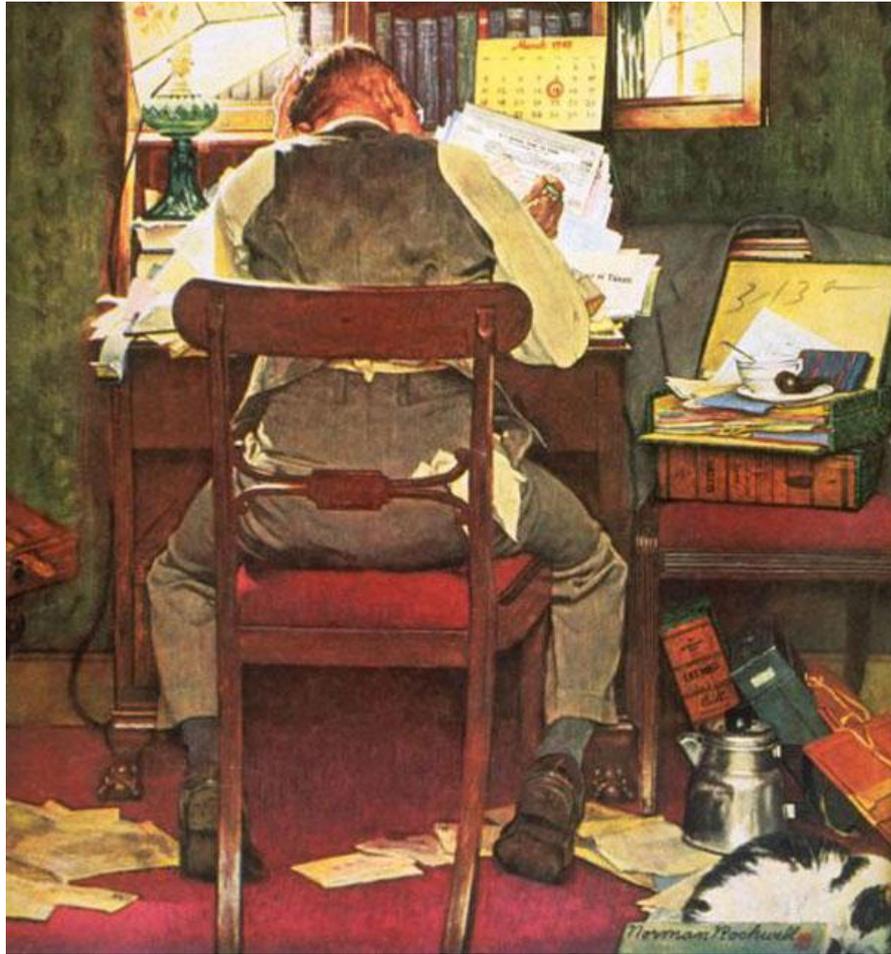


Recherche d'informations sur Internet (perfectionnement) méthodologie et outils disponibles





Pour commencer

Principes

Moteurs de recherche

Sites internet

Bases de données bibliographiques

Autres bases de données textuelles

Images et multimédia

Web social

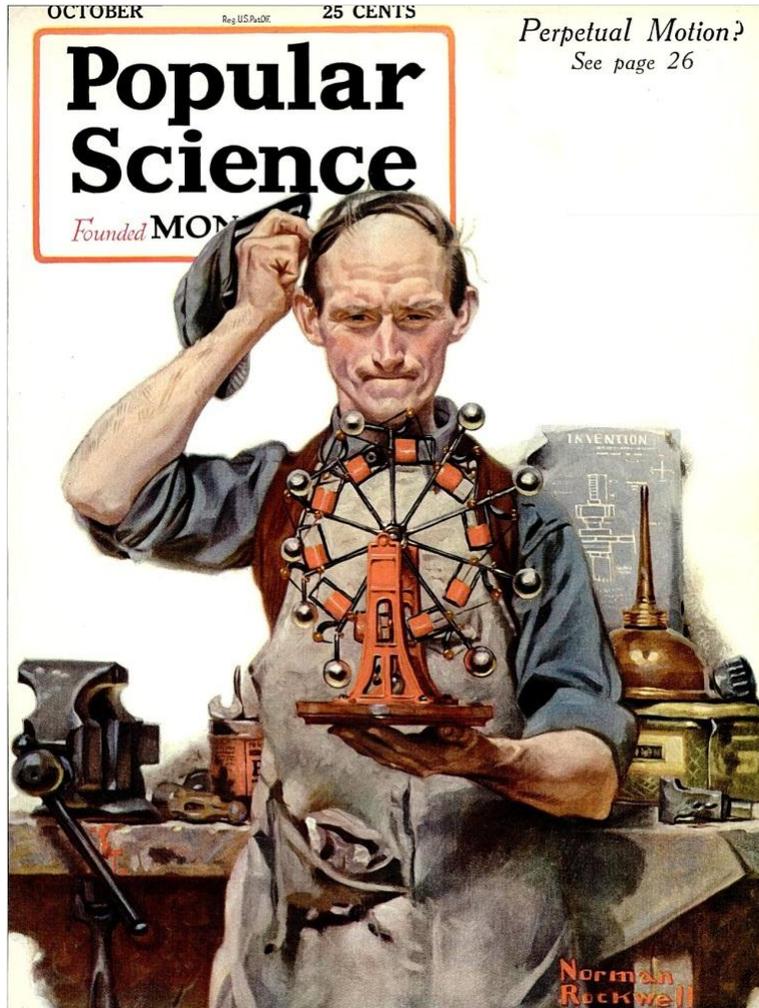
Temps réel

Quelques outils complémentaires

Veille automatisée

Exercices de synthèse

Bibliographie



Principles

Internet ? web ?

- internet
 - réseau de réseaux
 - fin des années 1960
 - applications et services divers : courrier électronique (mail), messagerie instantanée (IM), forums de discussion, transfert de fichiers (FTP), pair à pair (P2P), web (www)...
 - données
 - utilisateurs : 2,3 MM. dans le monde ([Internet World Stats](#))
 - langues : 56% du contenu en anglais, 4 % en français ([W3Techs](#))

Internet ? web ?

- Web
 - World Wide Web
 - ensemble de pages HTML (textes, images, liens...) avec une URL et accessibles avec le protocole HTTP
 - milieu des années 1990
 - web visible / web invisible
 - web invisible ou web profond : partie du web non indexée et qui ne peut être trouvée par les moteurs de recherche (pages protégées par un mot de passe, pages générées dynamiquement à la suite d'une requête...)
 - taille
 - 676,9 M. de sites ([Netcraft](#), 04/2012)
 - web indexé: au moins 8 milliards de pages ([Worldwidewebsize](#), 04/2012)
 - taille du web identifié (URL connues) ? web général ?

- définir le sujet

Questions préalables

Prendre du temps au départ pour en gagner par la suite...

Ne pas se lancer tout de suite dans la requête. Se poser des questions :

- Au minimum :

Qu'est-ce que je cherche exactement ? Le web est-il le moyen le plus rapide ?

- Pour aller plus loin :

Qui (profil du demandeur) ? Quoi (nature de l'information) ? Où (limites géographiques, lieux ressources) ? Quand ? Pourquoi (objectifs) ? Comment (outils, méthodologie, stratégie) ?

[Eduscol. Rechercher sur internet](#)

→ une stratégie de recherche

→ des critères d'évaluation

- définir le sujet
 - dictionnaires, encyclopédies et thésaurus :
abréviations, synonymes, hiérarchies, associations
 - RAMEAU : <http://catalogue.bnf.fr>
autorités > RAMEAU > choisir un critère : nom commun RAMEAU
mots-clés utilisés dans le catalogue de la BnF et dans le SUDOC
 - Grand dictionnaire terminologique :
http://www.granddictionnaire.com/btml/fra/r_motclef/index800_1.asp
dictionnaire québécois, propose également une traduction en anglais
 - TLF Trésor de la langue française : <http://atilf.atilf.fr/tlf.htm>
 - Dictionnaires Larousse : <http://www.larousse.fr/dictionnaires>
21 dictionnaires de langue (français et bilingues)
 - Acronymfinder : <http://www.acronymfinder.com/>

- suggestions des moteurs de recherche
 - moteurs de recherche



Exalead

- outils spécifiques
 - KwMap <http://www.kwmap.com/>
carte de mots-clés du web
 - générateur de mots-clés Google :
http://adwords.google.fr/o/Targeting/Explorer?_c=1000000000&_u=1000000000&ideaRequestType=KEYWORD_IDEAS
système Adwords

Exemple de listes de mots-clés

Exemple 1 : Traitements biologiques aérobies des eaux de rejets industriels pollués

Entre parenthèses les équivalents anglais des mots-clés

Concepts retenus	Termes génériques (termes plus larges)	Termes spécifiques (termes plus précis)	Termes associés (termes de sens voisin)
Mots trouvés par simple analyse des termes du sujet	Mots trouvés avec un dictionnaire de terminologie scientifique ou technique	Mots trouvés à l'aide d'un dictionnaire de terminologie scientifique ou technique, index "sujets" ou "mots clés" d'un catalogue ou d'une base de données	Mots trouvés à l'aide de l'index "sujets" ou "mots clés" d'un catalogue ou d'une base de données
<ul style="list-style-type: none"> - traitement biologique (biological treatment) - aérobie* (aerobic) - rejet* industriel* (industrial waste) - eau* (water) - polluant* (pollutant) 	<ul style="list-style-type: none"> - procaryote* (prokaryote*) - microorganisme* (microorganism*) - épuration des eaux usées (wastewater treatment) - assainissement (sanitation) 	<ul style="list-style-type: none"> - dépollution biologique, biodépollution, bioremédiation (bioremediation) - effluent industriel (industrial effluent) - digestion aérobie (aerobic digestion) 	<ul style="list-style-type: none"> - épuration (purification) - biodégradation (biodegradation) - bactérie* aérobie (aerobic bacterium/ bacteria) - bactérie* (bacterium/ bacteria, eubacterium / eubacteria)

- connaître les outils

Différents outils

Il n'y a pas que Google !

- bien comprendre les différences entre moteurs, métamoteurs et annuaires,
- distinguer les outils généralistes et les outils spécialisés,
- repérer les sites portails utiles pour vos recherches les plus fréquentes,
- essayer et comparer des outils de types différents, et d'autres du même type,
- lire l'aide en ligne, l'imprimer, mieux encore, faire sa propre fiche,
- mieux vaut maîtriser 1 ou 2 outils que sous-utiliser un grand nombre.

Utilisation réfléchie

En fonction des réponses apportées, utiliser de façon réfléchie

- l(es) outil(s) adapté(s), avec au besoin des filtres pays/langues
- de bons mots clés avec une syntaxe correcte (en saisissant bien la requête)

Syntaxe de recherche

Se familiariser avec la syntaxe

- les principes d'interrogation des bases de données (opérateurs booléens et de proximité, signes + et -, troncature...),
- la syntaxe de saisie spécifique à chaque outil (recherche simple et recherche avancée).

- syntaxe de recherche
 - opérateurs booléens
AND, OR, AND NOT/BUT NOT/ NOT, plus rarement ET/OU/SAUF
parenthèses
 - syntaxe
ordre des mots
guillemets : ex. : « prénom nom » et « nom prénom »
abréviations : ex. : « George Bush » pour George W. Bush et George H. W. Bush
troncature
 - orthographe
casse
+mot et -mot pour indiquer ou soustraire un terme exact, ex. +De Wilde
accents et signes diacritiques
pluriels : ex. : « textiles techniques » et « textile technique »
 - proximité
- syntaxe de saisie
 - recherche simple et avancée
 - critères de filtres lors des résultats

→ variable selon les outils

→ consulter l'aide



Moteurs de recherche

Moteurs de recherche

Documents	pages web sites documents (texte, images...)
Recherche	RECHERCHE PAR MOTS-CLES spécifique à chaque outil présence généralement d'une recherche avancée existence possible de tris et de filtres
Périmètre	généralement web visible peu d'interrogations du web invisible existence de moteurs généralistes et spécialisés, mais indexation et présentation des résultats automatiques difficile de trouver des documents originaux et peu connus
Outils	moteurs de recherche généralistes métamoteurs moteurs de recherche spécialisés (scientifiques notamment)

- fonctionnalités générales possibles
 - combinaison des différents opérateurs
 - filtres de recherche (B. Foenix-Riou, p. 58-81)
 - aux titres des pages
 - ! aux accents
 - à un nom de domaine ou un site
 - + fonctionnalités plus avancées que moteur de recherche sur le site
 - ! seulement sur les pages indexées par le moteur
 - à l'URL des pages
 - ! pas d'accent
 - à un type de fichier
 - à une date (Google, Exalead)
 - ! date de publication (1^e indexation des pages)
 - « backlinks » d'une page ou d'un site
 - suggestions de recherches associées
 - formulaires de recherche avancée
 - ne permettent pas des combinaisons très poussées
 - préférences de recherche (langue, résultats, sécurité)

Google

- révolutionnaire en 1998
mode de recherche (AND par défaut)
simplicité de l'interface (*one box*)
- données
chiffres inconnus (45-50 MM. de pages ?, [WorldWideWebSize](#))
ne concerne qu'une partie du web
- classement des résultats
algorithme tenu secret mêlant plusieurs éléments :
 - fréquence et place des mots
 - ordre des mots dans la requête
 - PageRank* (notoriété d'une page grâce aux liens)
 - pages connues et non les plus pertinentes dans les premiers résultats
 - prise en compte des nouveaux sites pas encore populaires ?

fonctionnalités de recherche	opérateurs booléens	<p>AND : tous les mots (généralement implicite)</p> <p>OR : au moins un des mots</p> <p>-mot : mot exclu</p>
	<p><u>critères de recherche avancés</u></p> 	<p>«mot» : mot exact, sans synonyme ni variante orthographique : <i>Mickael «Jacson» pour exclure Jackson</i></p> <p>« » : expression exacte : «<i>réchauffement climatique</i>»</p> <p>intitle : pour un mot ou une expression entre guillemets, allintitle : pour plusieurs mots dans le titre de la page</p> <p>inurl : pour un mot ou une expression entre guillemets, allinurl : pour plusieurs mots</p> <p>site : site ou domaine concerné sur le modèle : <i>site:google.fr</i></p> <p>filetype : type de fichier : <i>filetype:pdf</i></p> <p>chiffre..chiffre (+ unité de mesure au besoin) : fourchette de chiffres : <i>Michel Platini 1970..1980</i></p>
autres fonctionnalités	<p>troncature : implicite et aléatoire</p> <p>proximité : AROUND(n) mais aléatoires (B. Foenix-Riou)</p>	
résultats	filtres	<p>variables selon les moteurs de recherche (web, images..) : pays, langue, date</p> <p>+ pages en langue étrangère traduites</p> <p>+ mot à mot : chaîne exacte de caractères</p>
	rech. associées	des résultats différents selon que l'on regarde en bas de la page de résultats ET dans la colonne de gauche
autres fonctionnalités	cache, sites similaires	<p>en face de chaque réponse, en passant la souris sur le résultat</p> 
	vocabulaire	<p>define : définition</p> <p>~word : synonymes (anglais)</p>
	<u>préférences</u>	<p>filtre SafeSearch</p> <p>résultats par page</p> <p>blocage des résultats indésirables</p> <p>historique web</p>
	indexation d'un site par Google	info : sur le modèle info:google.fr (page en cache, pages similaires, liens entrants...)

Moteurs de recherche

- moteurs multilingues

- Bing : <http://www.bing.com/>

Microsoft

- moins de critères de recherche avancée que Google

- Yahoo! : <http://search.yahoo.com/>

technologie : Bing

- + suggestions orthographiques liées à des pages internet

- Ask : <http://www.ask.com/?o=312>

ancien moteur de recherche AskJeeves

- + indication de *related searches* et de *related questions* (en langage naturel)

- Blekko : <http://blekko.com/>

moteur de recherche participatif

souhaite présenter des résultats fiables à partir de sites de qualité et sans spam

déclare respecter les droits de recherche (algorithmes ouverts...)

- + *slashes* (fonctionnalité disponible sur inscription permettant de sélectionner des sites jugés pertinents et de limiter les résultats à ces sites)