



# ADBS FORMATION

---

Les bases de la recherche d'information sur Internet

304-15

septembre 2015

Conception

STEPHANE COTTIN – SGG (Services du Premier ministre)

[stephane@cottin.nom.fr](mailto:stephane@cottin.nom.fr) / <http://cottin.tel>

twitter : @cottinstef / Scoop-it : <http://scoop.it/u/stephane-cottin>

# SOMMAIRE

<http://www.adbs.fr/les-bases-de-la-recherche-d-information-sur-internet-143070.htm?RH=ACCUEIL>

## 304-15 - LES BASES DE LA RECHERCHE D'INFORMATION SUR INTERNET

### PRE-REQUIS

Maîtriser l'environnement Windows et savoir utiliser un navigateur.

### OBJECTIFS

- Repérer les méthodologies de recherche d'information spécifiques au réseau Internet
- Identifier et organiser les ressources d'information essentielles
- Utiliser les différents outils de recherche de façon appropriée

### METHODE

Exposés théoriques avec démonstrations.

Travaux pratiques : recherche avec les divers outils.

### PROGRAMME

**Les principes de fonctionnement du web**

**Panorama des outils de recherche**

**Méthodologie**

**Optimiser sa recherche**

Contact et suivi : Stéphane Cottin : [stephane.cottin@gmail.com](mailto:stephane.cottin@gmail.com)

Compte twitter : [@cottinstef](https://twitter.com/cottinstef) / Compte ScoopIt : <http://www.scoop.it/u/stephane-cottin>



<b>1</b>	<b>LES PRINCIPES DE FONCTIONNEMENT DU WEB .....</b>	<b>5</b>
1.1	LES NOTIONS FONDAMENTALES : URL, NOM DE DOMAINE, ADRESSE IP, ? .....	5
	<i>Les spécificités de la recherche sur le web .....</i>	<i>5</i>
	<i>Le Vocabulaire du navigateur .....</i>	<i>6</i>
	<i>L'essentiel sur le courrier électronique .....</i>	<i>8</i>
	<i>Les raccourcis clavier .....</i>	<i>9</i>
1.2	WEB VISIBLE ET WEB INVISIBLE .....	16
	<i>Caractéristiques de l'information sur Internet .....</i>	<i>16</i>
1.3	- ORGANISATION DE L'INFORMATION ET OUTILS DE RECHERCHE .....	19
	<i>Diversité des besoins, des contenus et des outils .....</i>	<i>19</i>
1.4	- LES APPORTS DU WEB 2.0. ....	20
	<i>Apport des outils et pratiques du web 2.0 : en quoi sont-ils créateurs de valeur ? .....</i>	<i>20</i>
<b>2</b>	<b>PANORAMA DES OUTILS DE RECHERCHE .....</b>	<b>25</b>
2.1	- MOTEURS DE RECHERCHE, METAMOTEURS .....	25
	<i>Les moteurs .....</i>	<i>26</i>
	<i>Les métamoteurs .....</i>	<i>28</i>
2.2	- ANNUAIRES, ENCYCLOPÉDIES .....	30
	<i>Les répertoires de recherche (ou Annuaires) .....</i>	<i>30</i>
	<i>Les encyclopédies .....</i>	<i>31</i>
2.3	- LA RECHERCHE SUR LE WEB INVISIBLE .....	32
	<i>Informations de base sur les méta-données .....</i>	<i>32</i>
2.4	- ÉVOLUTION DES OUTILS DE RECHERCHE .....	35
<b>3</b>	<b>METHODOLOGIE .....</b>	<b>36</b>
3.1	- ANALYSE DE LA DEMANDE ET DELIMITATION DU SUJET .....	36
	<i>Comment s'organise la recherche d'information ? .....</i>	<i>36</i>
	<i>Planifier ressources et moyens .....</i>	<i>38</i>
3.2	- REPERAGE ET SELECTION DES SOURCES D'INFORMATION .....	40
	<i>Recourir à des experts / développer son expertise .....</i>	<i>40</i>
	<i>Astuces pour identifier rapidement des sources d'information, des experts, etc .....</i>	<i>41</i>
3.3	- CONSTRUCTION DE LA REQUETE .....	41
	<i>Choix des mots-clés .....</i>	<i>41</i>
3.4	- TECHNIQUES DE RECHERCHE AVANCEES .....	45
	<i>Opérateurs avancés de recherche .....</i>	<i>45</i>
	<i>Où trouver des archives du web ? .....</i>	<i>47</i>
3.5	- ÉVALUATION DES RESSOURCES INTERNET .....	48
	<i>Quelques questions clés à se poser .....</i>	<i>48</i>
	<i>Comment évaluer un site web ? .....</i>	<i>48</i>
3.6	QUELQUES OUTILS PRATIQUES .....	50
	<i>Comment gérer les problèmes fréquents avec les outils ? .....</i>	<i>50</i>
	<i>Peut-on circuler de façon anonyme sur le web ? .....</i>	<i>51</i>
	<i>Peut-on effectuer des traductions de textes sur le web ? .....</i>	<i>51</i>
<b>4</b>	<b>OPTIMISER SA RECHERCHE .....</b>	<b>51</b>
4.1	- ORGANISER ET GERER LES FAVORIS (OU MARQUE-PAGES) .....	51
4.2	- UTILISER LES FILS RSS .....	53
4.3	- NAVIGATION SOCIALE ET PARTAGE DE LIENS .....	54
	<i>Tags et folksonomie .....</i>	<i>54</i>
4.4	- APRES LA FORMATION : S'AUTOFORMER ET SUIVRE L'ACTUALITE DES OUTILS DE RECHERCHE .....	56
	<i>Veille sur l'actualité des outils de recherche d'information .....</i>	<i>56</i>
	<i>Veille sur les producteurs de sources d'informations .....</i>	<i>57</i>

<http://www.adbs.fr/net-recherche-2013-surveiller-le-web-et-trouver-l-information-utile-137911.htm?RH=OUVRAGEADBS>

[http://superieur.deboeck.com/titres/130394\\_3/net-recherche-2013.html](http://superieur.deboeck.com/titres/130394_3/net-recherche-2013.html)



## NET RECHERCHE 2013 : SURVEILLER LE WEB ET TROU- VER L'INFORMA- TION UTILE

Véronique MESGUICH et Armelle THOMAS

Collection : Ouvrages d'autres éditeurs diffusés  
par l'ADBS

2013, 264 page(s), ISBN 978-2-80418-228-1

**35,00 € TTC**

### PRESENTATION DE L'OUVRAGE

Nouvelle édition entièrement remaniée, complétée et mise à jour, incluant des méthodes de recherche dans le web social, l'open data, le web temps réel...

Qu'elle réponde à des besoins professionnels ou personnels, la recherche d'information représente une grande part de l'activité des internautes.

Pourtant, le fonctionnement des moteurs de recherche est encore mal connu des utilisateurs, souvent perdus face à des résultats nombreux et imprécis, à des contenus structurés ou non, validés ou non, gratuits ou payants...

Cet ouvrage offre un panorama des outils et méthodes existant à ce jour :

- des conseils, des « trucs et astuces » pour optimiser le processus de recherche ;
- des informations précises sur les évolutions actuelles ;
- des présentations d'outils et des listes d'adresses utiles.

# 1 LES PRINCIPES DE FONCTIONNEMENT DU WEB

## 1.1 LES NOTIONS FONDAMENTALES : URL, NOM DE DOMAINE, ADRESSE IP, ?

### LES SPECIFICITES DE LA RECHERCHE SUR LE WEB

Le "web" n'est pas une encyclopédie : il n'est pas conçu pour la recherche

On y trouve de tout et surtout du n'importe quoi.

On y trouve ce qu'on y apporte (forums, blogs...)

On y trouve ce qu'on n'y cherche pas (sérendipité)

Le web est "daté" : il ne commence qu'en 1992 (ce qu'il y a avant n'est pas natif du web)

Ne pas confondre :

- Web et Internet
  - o Le web est un produit (parmi d'autres : mail, newsgroup...) du réseau Internet
- Internet et protocoles hypertexte
  - o Le protocole IP (Internet Protocol) est basé sur l'identification du réseau et de la machine (avec son adresse IP : une suite de quatre nombre de 0 à 255, permettant donc d'adresser 4,2 milliards de machines = ce n'est pas suffisant)

URL = Uniform Resource Locator

http://	www.	Nom-de-domaine	.fr	/	Sous-repertoire	/fichier.html
Protocole (pourrait être ftp:// smtp:// nntp:// ..)	Sous domaine	Nom du domaine	TLD (top level domain)			Nom du fichier en ligne (.html = fichier web)
	Nom de domaine complet					

## LE VOCABULAIRE DU NAVIGATEUR

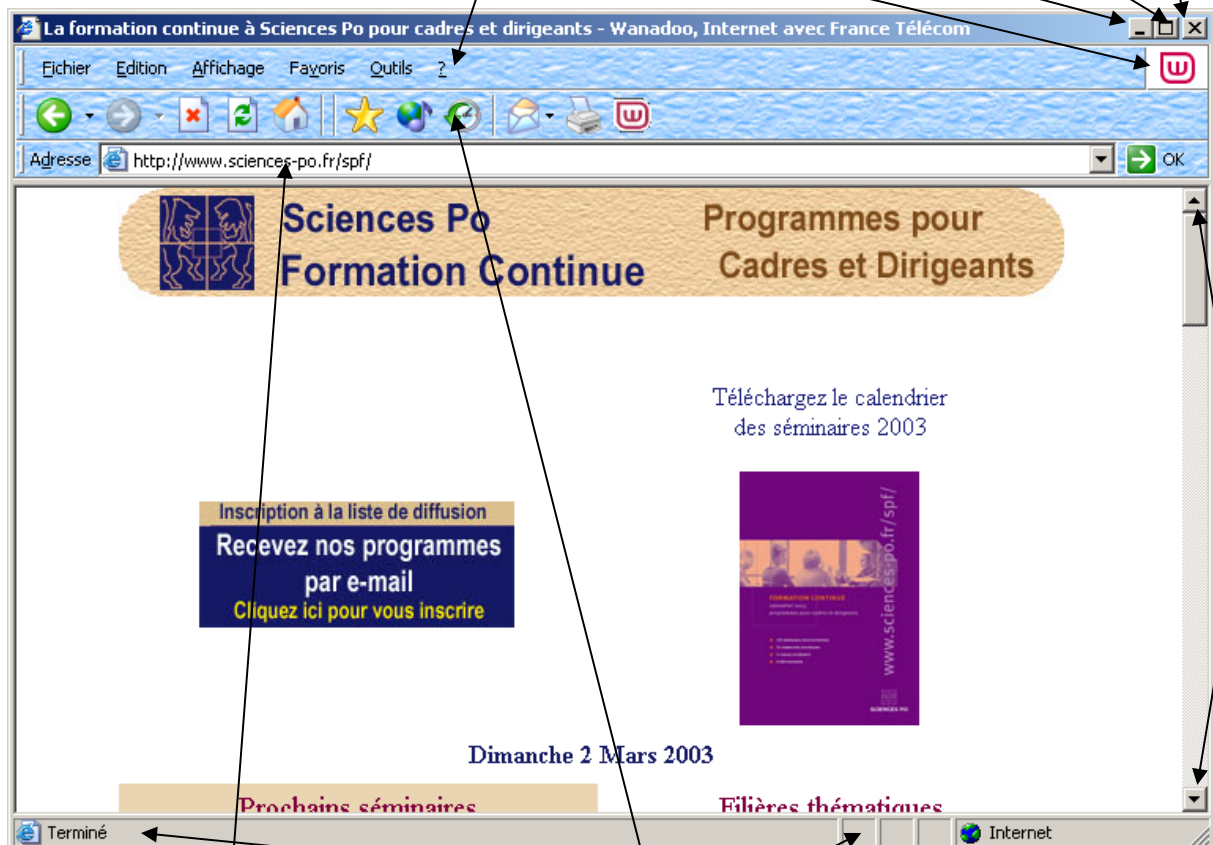
Ici une page de Microsoft Internet Explorer V.6

**Barre de titre** : contient le titre de la page visitée

**Boutons "système"** : réduire (à un bouton en bas) / agrandir ou réduire (à l'état de fenêtre) / fermer

Indicateur de fonctionnement du réseau (ici « Wanadoo »)

**Barre de menu**



Indicateur de sécurisation du site visité (ici non sécurisé)

**Barre de "Boutons"** dans l'ordre : Retour (d'une page) / Avancer (d'une page) / Arrêter (le chargement de la page en cours) / Recharger (en cas de blocage) / Retour Maison /

**Barre d'adresse** : c'est ici qu'il faut taper l'adresse du site web que vous voulez visiter ! Pas la peine de taper "http: // ", c'est sous-entendu  
Ensuite, taper sur la touche "ENTRÉE"

En cliquant ici, vous avez la liste des derniers sites visités (**historique** : voir aussi en tapant CTRL + H)

**Barre d'ascenseur** : monter et descendre la page en cours si elle dépasse de l'écran

(ici, l'utilitaire est Outlook Express, version 6)

Barre de titre (rappelle le titre de la rubrique sélectionnée)

Barre de menu

Barre de boutons

Zone de liste des messages reçus (les messages non lus sont en gras)

Dossiers  
 Outlook Express  
 Dossiers locaux  
 Boîte de réception  
 Boîte d'envoi  
 Éléments envoyés  
 Éléments supprimés  
 Brouillons  
 wanadoo.fr

À	Objet	Envoyé
Emmanuel Barthe	Lu : [juriconnexion] Tous les textes communautaires ...	21/09/20...
<b>stephane.cottin@conseil...</b>	<b>EUR-Lex Législation communautaire en vigueur - Doc...</b>	<b>21/09/20...</b>
Emmanuel Barthe	Re: [juriconnexion] Tous les textes communautaires ...	21/09/20...
stephane.cottin@conseil...	correction	21/09/20...
stephane.cottin@conseil...	Fw: [motrech] Envoi groupé n° 430	22/09/20...
stephane.cottin@conseil...	Fw: [Legifrance] J.O. NUMERO 221 DU 23 SEPTEMBR...	23/09/20...

De : Stéphane Cottin (Quincy) À : stephane.cottin@conseil-constitutionnel.fr  
 Objet : EUR-Lex Législation communautaire en vigueur - Document 301R1049

[http://europa.eu.int/eur-lex/fr/lif/dat/2001/fr\\_301R1049.html](http://europa.eu.int/eur-lex/fr/lif/dat/2001/fr_301R1049.html)

Contacts  
 benoit.tabaka@wanadoo.fr  
 Emmanuel Barthe  
 r.nerriere@cheuvreux-associes.fr  
 stephane.cottin@conseil-constitutionnel.fr

Thème du message à envoyer

Fichier Edition Affichage Insertion Format Outils Message ?

Couper Copier Coller Annuler Joindre Priorité Signer Crypter

Destinataires

À : prenom.nom@societe.fr  
 Cc : prenom.nom@societe2.com; prenom.nom@societe3.de  
 Cci : prenom@masociete.fr  
 Objet : Thème du message à envoyer  
 assurance.doc (22,0 Ko)

Barre de mise en forme

Arial 10

Bonjour,  
 Ceci est un message

Bonne journée  
 Ma signature...

Zone de liste des messages reçus (les messages non lus sont en gras)

Fenêtre de message

Bouton "Joindre" (permet "d'attacher" un fichier au message)

Pour envoyer le message  
 Cliquer ici



Le courrier électronique est l'outil le plus répandu d'abord dans les Intranets et les réseaux des entreprises, puis pour l'extérieur (Extranets et Internet), enfin pour le particulier avec Internet. Il permet d'acheminer des notes ou du courrier entre personnes éloignées. Le nom anglais 'e-mail' est passé directement en français, de même que le @ « at » caractéristique d'une adresse électronique. Mais l'académie souhaite que l'on dise respectivement 'mel' et 'arrobisque'. Les Canadiens disent joliment 'courriel'.

L'avantage du courrier électronique sur le téléphone ou sur le fax est qu'il permet de joindre un correspondant avec des informations écrites tout comme le fax, mais qui peuvent être recopiées dans un document directement sur un traitement de texte, un tableur ou une base de données. En outre vous pouvez lire votre courrier de n'importe où dans le monde.

Les logiciels de courrier électronique permettent d'envoyer des **documents attachés** à la note principale.

### Principes

Chaque connecté à Internet possède une ou plusieurs adresses de courrier Internet. On les appelle les adresses e-mail. A la différence du téléphone, le courrier électronique n'est pas anonyme dans la mesure où votre correspondant connaît votre adresse dès que vous lui adressez un message.

Ces adresses sont de la forme **nom@organisation.domaine**

[On ne peut **jamais** y voir de lettres accentuées, d'apostrophes, d'espaces ou de guillemets. Le **domaine** est un ensemble de deux, trois ou quatre lettres renvoyant à un ensemble géographique - deux lettres, comme *.fr* - ou thématique - trois ou quatre lettres, comme *.com*]

### Envoyer et recevoir des courriers

Un mail contient toujours (sous la forme de champs, ce n'est autre qu'un fichier informatique...) :

- l'adresse du destinataire, (champ A: ou TO:)
- le sujet du mail parfois appelé aussi objet du courrier (champ SUJET: ou OBJET:) (la nétiquette l'impose, pas la technique)

Et de façon optionnelle :

- La ou les adresses d'autres destinataires (champs CC - Carbon copy, ou BCC Blind Carbon copy, ou CCI copie invisible)
- les lignes correspondant au contenu du mail.
- Un ou plusieurs attachements qui correspondent à un ou plusieurs fichiers qui peuvent être de n'importe quelle forme (ASCII, Word, son etc).



## LES RACCOURCIS CLAVIER

Chaque logiciel a ses raccourcis spécifiques, qui ne marchent pas forcément partout. Certains sont génériques..

Les raccourcis génériques


F1	Aide en ligne du programme en cours (en général aide contextuelle)
ALT + F4 (rester appuyé sur ALT puis appuyer une fois sur F4)	Fermeture du programme en cours
CTRL + F4	Fermeture de la fenêtre, ou du fichier (pour word par exemple) en cours (mais pas fermeture du programme)
ALT + TAB	Passage d'un programme ouvert à un autre
⌘ + E	Ouvre le gestionnaire de fichier ou Explorateur
⌘ + M	Nettoie le bureau (réduit toutes les fenêtres ouvertes : ne les ferme pas)
⌘ + F	Ouvre l'outil de recherche de fichier sur le disque
⌘ + R	Ouvre la fenêtre d'exécution
⌘ + L	Verrouille immédiatement l'ordinateur
⌘ (tout seul)	Ouvre le menu démarrer
CTRL + C	Copier
CTRL + V	Coller
CTRL + X	Couper
CTRL + A	Tout sélectionner
CTRL + Z	Annuler dernière action

## Généralités sur le clavier



Qu'est ce qu'un "raccourci clavier" ?

Un raccourci clavier est une combinaison de touches du clavier, qui, pressées ensemble, offrent une fonctionnalité (ouvrir un fichier, faire une action comme imprimer, copier, couper, coller...)

Comment faire un raccourci clavier ?

Les raccourcis clavier sont en général la combinaison de une ou plusieurs touches normalement inactives lorsqu'elles sont pressées seules, avec une touche active. Aussi, la meilleure façon de faire un raccourci clavier consiste à presser D'ABORD la (ou les) touche(s) inactives (SHIFT (ou MAJ), CTRL, ALT, ALT GR, Touche Windows ) , DE RESTER APPUYÉ, PUIS DE PRESSER LA TOUCHE ACTIVE (en général une lettre, une action (suppr, inser...) ou une touche de fonction - F1 à F12 sur la ligne du haut du clavier)

### Les touches du clavier

esc		F1	F2	F3	F4		F5	F6	F7	F8		F9	F10	F11	F12		impr écran	Arrêt défil	Pause		verr num	maj lock	arrêt défil		
²	1	2	3	4	5	6	7	8	9	0	°	+	<--				inser	home	pgup		verr num	/	*	-	
tab		A	Z	E	R	T	Y	U	I	O	P	..	\$	entrée				suppr	fin	pgdn		7	8	9	+
maj lock		Q	S	D	F	G	H	J	K	L	M	%	μ											4	
maj	<	W	X	C	V	B	N	?	.	/	§	maj ou shift						^			1	2	3	entr	
ctrl		alt	barre d'espace							alt gr		bouton droit souris	ctrl		<	v	>		0		.				

les touches de fonction													
F1	F2	F3	F4		F5	F6	F7	F8		F9	F10	F11	F12
servent de touches complémentaires programmables par les logiciels (word, excel, etc...) et ont donc des fonctionnalités différentes selon les cas. Normalement, F1 sert toujours à l'AIDE													

les touches d'action (ou de déplacement)													
esc													
						<-- (ou backspace)				inser	home	Pgup (Page Up)	
tab						entrée				suppr	fin	Pgdn (Page Down)	
											^		
		barre d'espace								<	v	>	



la touche "esc" (ou échap., ou échappement) permet d'arrêter une action ou de l'annuler

la touche "tab" (tabulation) permet dans un traitement de texte d'avancer le point d'insertion jusqu'à la prochaine marque de tabulation (par défaut, 2,5 cm) ; dans un autre type de programme, elle permet de passer d'un champ à un autre ; combinée avec la touche "ALT", elle permet de basculer d'un programme ouvert à un autre

la touche backspace (<--) permet d'effacer le caractère à la GAUCHE du point d'insertion

la touche suppr (ou del) permet d'effacer le caractère à la DROITE du point d'insertion

la touche inser permet de basculer en mode traitement de texte entre le mode INSERTION et le mode REFRAPPE (voir la barre d'état tout en bas). Le mode normal est le mode INSERTION (tout caractère frappé s'insère au niveau du point d'insertion et chasse les caractères à sa droite). En mode REFRAPPE, tout caractère frappé EFFACE le caractère à sa droite. Combinée avec la touche "SHIFT" (ou MAJ) = COLLER

la touche HOME (ou flèche vers le nord ouest) : revient au début de la ligne. Combinée à la touche CTRL, revient au début du document ; combinée avec les touches SHIFT et CTRL : sélectionne le texte à partir du point d'insertion jusqu'au début.




la touche FIN : part à la fin de la ligne. Combinée à la touche CTRL, part à la fin du document ; combinée avec les touches SHIFT et CTRL : sélectionne le texte à partir du point d'insertion jusqu'à la fin.



les touches page up et page down : monte ou descend de quelques lignes (variable selon le logiciel : en général 5 à 15 lignes)



barre d'espace / touche entrée et curseur (flèches de déplacement). Notez que la touche entrée est doublée sur le pavé numérique et que, quand le pavé numérique est désactivé (touche verr num), le pavé numérique fait office de flèches de déplacement)

les touches du pavé alphabétique													
2	1	2	3	4	5	6	7	8	9	0	°	+	
		A	Z	E	R	T	Y	U	I	O	P	¨	\$
maj lock		Q	S	D	F	G	H	J	K	L	M	%	µ
maj	<	W	X	C	V	B	N	?	.	/	§	maj ou shift	

En mode normal, c'est le caractère du BAS de la touche qui s'imprime. Pour simplifier, les minuscules n'ont pas été imprimées sur la plupart des touches, seules les majuscules sont visibles en HAUT des touches, accessibles soit par la pression continue et préalable d'une quelconque des touches "MAJ" (ou "SHIFT") de part et d'autre du pavé, soit par l'utilisation du mode MAJ LOCK (ou CAPS LOCK) (ou symbole cadenas). Lorsque ce mode est actif, la lumière en haut du pavé numérique s'allume et c'est comme si une touche SHIFT était enfoncée. Ce mode s'éteint, selon les modèles de clavier soit par une pression sur MAJ LOCK, soit par une pression sur une des touches SHIFT

les touches de combinaison (d'altération)												
maj												maj ou shift
ctrl		alt	barre d'espace					alt gr		 bouton droit souris	ctrl	

Ces touches, utilisées seules, sont **inactives** ou ne modifient rien. Seules les deux touches Microsoft () et leur voisine "bouton droit de la souris" (ou "menu contextuel" , représenté par une icône symbolisant un menu et une flèche) font respectivement :

- pour  ouvrir le menu Démarrer (remarquez que le menu ne s'ouvre qu'au moment où vous lâchez la pression, preuve que ce bouton est surtout conçu pour combiner avec une autre touche, qu'il attend dans ce cas)
- pour  même action qu'un clic du bouton droit de la souris au niveau du point d'insertion (c'est-à-dire ouverture d'un menu contextuel, ne comprenant que les actions les plus courantes)

Les touches SHIFT (ou MAJ ou  $\hat{u}$ ), CTRL (lire "Control") ou ALT sont parfaitement inactives lorsqu'elles sont utilisées toutes seules. Vous pouvez les presser, les tenir appuyées, puis aller chercher l'autre touche pour le raccourci clavier

Notez que si les touches CTRL et SHIFT ( $\hat{u}$ ) sont identiques de part et d'autre de la barre d'espace, les deux touches ALT sont DIFFÉRENTES : celle de gauche, ALT (tout court), pour Alternate, est la touche Alt classique, celle de droite s'appelle ALT GR, pour Alternate Graphics, prononcez "Alt graphique", est utilisée pour les raccourcis d'appel aux caractères spéciaux (comme @ : ALT GR + à )

Situé à l'extrême droite du clavier, le pavé numérique double les caractères CHIFFRES qui se trouvent sur la premières lignes du pavé alphabétique, mais que l'on doit alors "shifter" (appuyer sur SHIFT ou MAJ). Le pavé numérique ne fonctionne en mode CHIFFRES que si la lumière "Verr Num" est allumée : appuyez sur la touche "Verr num" (ou Num Lock sur certains claviers) pour basculer l'état de la lumière [Note : si la lumière ne bouge pas alors que l'on appuie sur la touche Verr Num, c'est une indication caractéristique du blocage de votre poste = c'est une des actions que les informaticiens font pour détecter une panne]

Les quatre opérations arithmétiques sont présentes (attention, le signe "multiplié" est une étoile = \* ; le signe "divisé" est une barre oblique = /). Il se peut que la virgule (symbolisée à l'américaine par un point à côté du zéro) soit mal configurée selon les logiciels, et affiche un point au lieu d'une virgule ou vice-versa. Notez que le bouton "Entrée" a exactement les mêmes fonctionnalités que le bouton "Entrée" (ou "Retour") du pavé alphabétique.

le pavé numérique			
verr num	maj lock	arrêt défil	
verr num	/	*	-
7	8	9	+
4	5	6	
1	2	3	entr
0	.		

les touches pré-programmées "à voyant lumineux"

																impr écran	Arrêt défil	Pause	verr num	maj lock	arrêt défil	
																verr num						

Les touches "Pause" et "Impr Écran" n'ont que peu d'intérêt dans les fonctionnalités que vous utilisez  
 "Pause" en a beaucoup en programmation, car elle permet, combiner avec la touche CTRL, d'arrêter un processus ;  
 "Impr Ecran" fait une copie de l'écran que vous pouvez "coller" dans une application, ALT + "Impr Ecran" fait une copie de la fenêtre active (juste le logiciel que vous êtes en train d'utiliser, contrairement à "Impr Ecran" tout seul, qui copie tout le bureau)  
 Les trois autres touches commandent les trois petites lumières qui sont au-dessus du pavé numérique.  
 Maj Lock ou Caps Lock, ou symbole cadenas, verrouille le clavier en mode MAJUSCULES  
 Verr Num indique que le pavé numérique est en mode CHIFFRES (et non en mode flèches de déplacement)  
 Arrêt défil ou Scroll Lock, est une fonctionnalité qui n'a plus cours de nos jours (elle bloquait le défilement de l'écran dans le cas de listing trop long)



## 1.2 WEB VISIBLE ET WEB INVISIBLE

### CARACTERISTIQUES DE L'INFORMATION SUR INTERNET

- ✓ Grande hétérogénéité dans les contenus et dans les publics (grand public et professionnels)
- ✓ Contenus dynamiques et renouvellement continu
- ✓ Instabilité des localisations
- ✓ Fragmentation plus ou moins importante, selon les disciplines
- ✓ Multilinguisme et couverture géographique mondiale
- ✓ Information gratuite et payante (tendance à plus d'information, plus rapide, moins chère, avec une frange d'information à valeur ajoutée payante) : notion de freemium

### LA TAILLE DU WEB

Il est très difficile d'estimer la taille réelle du Web. Sa croissance se poursuit (encore ?) à un rythme excessivement rapide (on avait coutume de d'estimer à quelque 7 millions de pages supplémentaires par jour en 2002, certainement 10 fois plus en 2005, 100 à 1000 fois plus en 2014), mais de nombreuses pages ont une durée de vie très limitée. La plus grande difficulté liée au dénombrement provient de la part très majoritaires des pages dynamiques, et du "big data" (cf. le chapitre consacré au web invisible), et donc de la définition que l'on donne à une "page web".

Cela dit, en toute logique, on doit dépasser actuellement les 1000 milliards de pages, sans compter les informations contenues dans les bases de données. Google a dépassé début 2005 les 8 milliards de pages indexées, date à laquelle il a cessé d'indiquer sur sa page d'accueil cette information, de toute façon certainement déjà sous-estimée à l'époque.

Cette question est très souvent posée et de nombreux débats en ressortent, qui permettent d'illustrer diverses évolutions récentes de la présentation des contenus sur internet.



## Application :

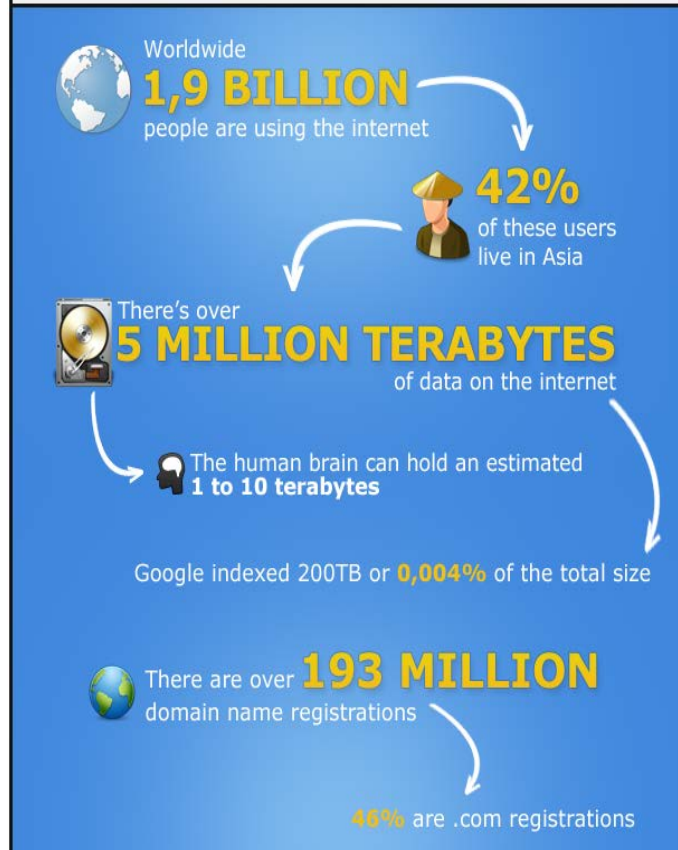
Pour avoir un aperçu du volume considérable de données indexées par les principaux moteurs de recherche, sans chercher à faire de comparatif, mais juste pour avoir des jalons mesurables, on peut régulièrement tenter de regarder les résultats de quelques requêtes tests sur les trois principaux moteurs (Google, Bing de Microsoft et Yahoo)

Nombre de résultats (depuis juin 2010)	Requête « Droit France » dans la version française du moteur	Requête « French Law » dans la partie anglophone du moteur	Nombre de pages indexées en gouv.fr
Google	<a href="http://www.google.fr/search?q=droit+france">http://www.google.fr/search?q=droit+france</a> 33 200 000 résultats 2010 149 000 000 résultats 2012 <b>158 000 000 en juin 2014</b> <b>195 000 000 en septembre 2015</b>	<a href="http://www.google.com/search?q=french+law">http://www.google.com/search?q=french+law</a> 104 000 000 résultats 2010 631,000,000 other results 2012 <b>410 000 000 en juin 2014</b> <b>620 000 000 sept. 2015</b>	42 200 000 (requête site:gouv.fr 2010) 47 400 000 en 2012 <b>26 300 000 en juin 2014</b> <b>73 000 000 en septembre 2015</b>
Bing	<a href="http://www.bing.com/search?q=droit+france">http://www.bing.com/search?q=droit+france</a> 70 000 000 résultats 2010 373 000 000 en 2012 122 000 000 en 2014 58 700 000 en 2015	<a href="http://www.bing.com/search?q=french+law">http://www.bing.com/search?q=french+law</a> 139 000 000 résultats en 2010 499 000 000 en 2012 109 000 000 en 2014 19 000 000 en 2015	3 230 000 (requête site:gouv.fr) en 2010 3 690 000 en 2012 2 950 000 en 2014 14 000 000 en 2015
Yahoo (id. Bing depuis 2012)	<a href="http://fr.search.yahoo.com/search?h?p=droit+france">http://fr.search.yahoo.com/search?h?p=droit+france</a> 465 000 000 résultats en 2010 379 000 000 en 2012 122 000 000 en 2014 58 300 000 en 2015	<a href="http://search.yahoo.com/search?p=french+law">http://search.yahoo.com/search?p=french+law</a> 281 000 000 résultats en 2010 568 000 000 en 2012 94 700 000 en 2014 8 640 000 en 2015	6 822 733 (requête domain:gouv.fr) en 2012 2 470 000 en 2012 3 020 000 en 2014 14 000 000 en 2015 Voir <a href="#">Rapport Riester</a> .

**La quantité de ces résultats ne permet pas une recherche professionnelle efficace : il y aura toujours trop de bruit documentaire. La disparité du nombre de réponses entre chaque moteur illustre aussi l'insécurité documentaire dans laquelle on se trouve face à ces masses de données non structurées.**

# THE AWESOME SIZE OF THE INTERNET

AN INFOGRAPHIC BY THEROXOR.COM



Analyse : Même en posant une recherche sur un moteur censé répondre en français : on a très vite des réponses en anglais, via des articles de blogs reprenant les infos ailleurs. Phénomène de la « documentarisation » = tout est information, document, réutilisation possible (techniquement et juridiquement)

La présence de blogs et des encyclopédies collaboratives est évidente. Il faut apprendre à les maîtriser, ainsi que les autres méthodes du web 2.0 et du « crowd-sourcing »

Dans l'ordre d'apparition :

- 1969 : des messages ? Difficilement exploitables, mais énormément de données. Quand elles sont structurées, cela peut donner de très belles choses (ex : projet Gutenberg)
- 1974 : des forums ? Idem. L'absence de structuration et de métadonnées en font des ressources peu exploitables, sauf quelques domaines très pointus : l'informatique, et quelques sciences dures, mais aussi la recherche et l'analyse de rumeurs.
- 1990 : des pages web ? La face émergée de l'iceberg, et pourtant déjà gigantesque
- Maintenant : du web structuré (sémantique) ! Arrivée peu à peu de « big data sets » = encore faut-il savoir les réutiliser.

### 1.3 - ORGANISATION DE L'INFORMATION ET OUTILS DE RECHERCHE

---

#### DIVERSITE DES BESOINS, DES CONTENUS ET DES OUTILS

---

##### LES BESOINS : LE CAHIER DES CHARGES DU PROTOCOLE IP ET DU WEB ?

Rappel = le protocole TCP/IP (support du Net, créé fin des années 60), n'est qu'un protocole de communication parmi de nombreux autres. Il n'est pas considéré comme le meilleur, il a longtemps (et est toujours) concurrencé par de nombreux autres protocoles.

Son utilité a été avérée dans son utilisation comme support de messagerie par envoi de paquets d'information (maillage en réseau).

Le TCP/IP est d'abord un outil d'envoi de messages (e-mail ou SMTP), puis très vite, les listes de discussion sont apparues, avec leurs limites.

L'usage des forums est apparu aussi très vite, contre l'avis des concepteurs, et c'est un autre protocole, l'USENET (ou NNTP) qui dès 1974 ouvrira les Usenet Newsgroups (ne pas confondre avec les google groups) qui en ont repris les archives de 1985 à nos jours.

Voir <http://boardreader.com> pour avoir une idée de l'activité des forums (ou bulletin board system (BBS)), puis <http://groups.google.com> pour la recherche très ancienne ou par auteur

<https://groups.google.com/forum/#!forum/fr.doc.divers>

Enfin, et seulement début 1990, arrivent les outils d'affichage graphique propres au protocole TCP/IP. Tim Berners Lee, et son équipe du CERN à Genève, crée le protocole HTTP (transfert hypertexte) et son langage de description de page (hypertexte markup language ou HTML), qui n'est qu'une version simplifiée du SGML des imprimeurs, uniquement conçue pour l'affichage sur écran : c'est donc une

méthode pour les **contenants**. Mais l'html semble suffire à structurer superficiellement les données (les **contenus**), ce qui permet aux premiers robots de venir indexer (crawler) le réseau des réseaux (ou Web), mais on est encore et toujours loin du web sémantique qui pourrait naître des relations qualifiées et de l'intelligence des liens entre les objets dans un cadre de relations de données (RDF ou Related Data Framework).

---

#### LES OUTILS : DE L'ADEQUATION AUX RESSOURCES ET AUX BESOINS ?

<http://outils.abondance.com/>

Il faut :

- Des outils pour **aller** sur les sites : les navigateurs (Internet Explorer, Mozilla Firefox (ex Netscape, ex Mosaic), Google Chrome, Safari, Opera, etc.)
- Des outils pour **concevoir** les sites : éditeurs html (Frontpage, Dreamweaver, Konqueror, Nvu, Blue Griffon...), Système de gestion collaborative de contenu web (CMS type EzPublish, Spip, Joomla, Typo3, Alfresco, Nuxeo, Drupal...)
- Des outils pour **analyser** ces sites et ces contenus : moteurs internes ou externes, lecteurs de fils rss, aspirateurs, agrégateurs, ...
- Des outils de recherche : les moteurs, les métamoteurs, les outils de curation (et, pour mémoire, les annuaires) :

#### 1.4 - LES APPORTS DU WEB 2.0.

---

#### APPORT DES OUTILS ET PRATIQUES DU WEB 2.0 : EN QUOI SONT-ILS CREATEURS DE VALEUR ?

---

#### ARCHIVES OUVERTES

Les ressources géantes de l'OAI-PMH

Le concept d'archives libres et de dépôts institutionnels

<http://repositories.webometrics.info/>

## Top Institutionals

World Rank ▲	Portal	Country	Size	Visibility	Files Rich	scholar
1	<a href="#">HAL Institut National de Recherche en Informatique et en Automatique Archive Ouverte</a>		21	11	2	4
2	<a href="#">University of California eScholarship Repository</a>		429	9	13	1
3	<a href="#">(1) Smithsonian/NASA Astrophysics Data System</a>		5	1	1393	183
4	<a href="#">NASA Technical Reports Server</a>		189	2	5	29
5	<a href="#">HAL Sciences de l'Homme et de la Société</a>		44	8	87	14
6	<a href="#">(1) Universitat Autònoma de Barcelona Dipòsit Digital de Documents</a>		105	28	9	2
7	<a href="#">Universidade de São Paulo Biblioteca Digital de Teses e Dissertações</a>		135	10	12	38
8	<a href="#">CERN Document Server</a>		4	22	1	48
9	<a href="#">Virginia Tech University Digital Library and Archives</a>		430	3	59	94
10	<a href="#">Queensland University of Technology Institutional Repository</a>		160	7	83	47
11	<a href="#">MIT Institutional Repository</a>		133	4	15	125

## Top Portals

World Rank ▲	Portal	Country	Size	Visibility	Files Rich	scholar
1	<a href="#">ResearchGate</a>		3	2	1	1
2	<a href="#">Academia.edu</a>		5	1	28	4
3	<a href="#">Scientific Electronic Library Online Brazil SciELO Brazil</a>		4	3	5	2
4	<a href="#">DIALNET</a>		8	4	3	3
5	<a href="#">HAL Hyper Article en Ligne</a>		7	9	2	5
6	<a href="#">DiVA Digitala Vetenskapliga Arkivet Academic Archive On-line</a>		11	10	6	6
7	<a href="#">Scientific Electronic Library Online Chile SciELO Chile</a>		13	13	19	7
8	<a href="#">Berkeley Electronic Press Bepress</a>		2	7	41	32
9	<a href="#">Scientific Electronic Library Online SciELO</a>		48	8	48	25
10	<a href="#">Érudit Consortium interuniversitaire</a>		35	17	7	15
11	<a href="#">Thèses en Ligne TEL</a>		25	18	12	16
12	<a href="#">(1) RedALyC Red de Revistas Científicas de América Latina y el Caribe, España y Portugal</a>		36	15	4	24
13	<a href="#">Revue.org</a>		15	5	67	53

Applications : les moteurs de recherche dédiés à ces archives ouvertes qui vont "moissonner" (le H de OAI-PMH = Harvest) les entrepôts.

Voir Google Scholar ([scholar.google.com](http://scholar.google.com)) ; l'initiative française : Recherche Isidore ([www.rechercheisidore.fr](http://www.rechercheisidore.fr))

Et surtout l'initiative allemande Base Search de Bielefeld : ([base-search.net](http://base-search.net))

0 Computer science, information & general works (136376)	<a href="#">View Records</a>	30 Social sciences, sociology & anthropology (133318)	<a href="#">View Records</a>	340 Law (44911)	<a href="#">View Records</a>
1 Philosophy & psychology (86395)	<a href="#">View Records</a>	31 Statistics (7793)	<a href="#">View Records</a>	342 Constitutional & administrative law (6)	<a href="#">View Records</a>
2 Religion (40296)	<a href="#">View Records</a>	32 Political science (64910)	<a href="#">View Records</a>	343 Military, tax, trade & industrial law (10)	<a href="#">View Records</a>
3 Social sciences (486547)	<a href="#">View Records</a>	33 Economics (109076)	<a href="#">View Records</a>	344 Labor, social, education & cultural law (12)	<a href="#">View Records</a>
4 Language (22642)	<a href="#">View Records</a>	34 Law (44967)	<a href="#">View Records</a>	345 Criminal law (2)	<a href="#">View Records</a>
5 Science (518257)	<a href="#">View Records</a>	35 Public administration & military science (21752)	<a href="#">View Records</a>	346 Private law (9)	<a href="#">View Records</a>
6 Technology (592250)	<a href="#">View Records</a>	36 Social problems & social services (3270)	<a href="#">View Records</a>	347 Civil procedure & courts (2)	<a href="#">View Records</a>
7 Arts & recreation (219131)	<a href="#">View Records</a>	37 Education (85292)	<a href="#">View Records</a>	348 Laws, regulations & cases (14)	<a href="#">View Records</a>

<http://www.base-search.net/about/en/index.php>

CREER UN BLOG : C'EST TRES (TROP ?) SIMPLE

Liste des blogs > Créer un blog



Titre

Adresse

.blogspot.com

Vous pourrez ajouter un domaine personnalisé ultérieurement.

Modèle

The image shows a grid of six different blog templates. Each template is represented by a small preview image and a text label below it. The templates are: 'Dynamic Views' (a dark theme with a large image), 'Simple' (a clean, light theme), 'Picture Window' (a theme with a large image of a dog), 'Awesome Inc.' (a dark theme with a blue abstract image), 'Watermark' (a light theme with a red and white image), and 'Ethereal' (a light theme with a blue and white image).

Vous pourrez découvrir d'autres modèles et personnaliser votre blog par la suite.

Créer un blog

Annuler

<http://www.blogger.com>

Voir aussi avec le plus célèbre des outils de création de blogs : [www.wordpress.com](http://www.wordpress.com)

Aide [Discussion](#) Lire [Voir le texte source](#) [Afficher l'historique](#)  

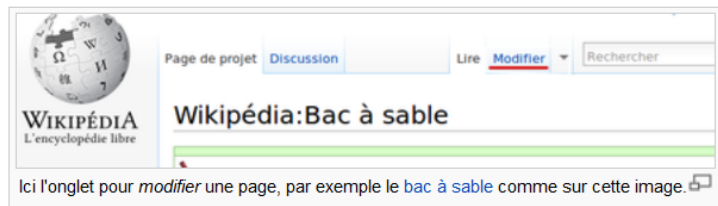
## Aide:Comment modifier une page

Le principe de base du fonctionnement de Wikipédia est que son contenu peut être modifié par les [internauts](#). Pratiquement toutes les pages de Wikipédia peuvent être directement corrigées ou enrichies par tous les visiteurs <sup>1</sup>, même sans [inscription préalable](#). Si vous voulez faire des tests, vous pouvez utiliser le [bac à sable](#)<sup>2</sup> ou votre [page personnelle](#).

[Raccourci \[+\]](#)  
[WP-MODIF](#)

### Sommaire [\[masquer\]](#)

- 1 [Préambule](#)
- 2 [Accès à la page de modification](#)
- 3 [Modifier le texte](#)
- 4 [Prévisualiser les modifications](#)
- 5 [Commenter les modifications dans la boîte de résumé](#)
- 6 [Publier les modifications \(ou les annuler\)](#)
- 7 [Voir aussi](#)
- 8 [Notes](#)



Application : survol des techniques wiki et notamment de la wikipedia. Apprentissage de la modification de pages et de l'analyse des historiques.



## 2 PANORAMA DES OUTILS DE RECHERCHE

### 2.1 - MOTEURS DE RECHERCHE, METAMOTEURS



Le principal annuaire mondial était [www.yahoo.fr](http://www.yahoo.fr) (YAHOO! Signifiant Yet Another hierarchically Organized Oracle), mais ce n'était pas le plus efficace (voir [www.dmoz.org](http://www.dmoz.org))

Depuis plusieurs années, la notion d'annuaire (notamment avec Yahoo et Voila) s'est complètement fondue dans celle de moteur, et que ces principaux annuaires (ou répertoires) sont en fait des moteurs.



Mais le principal moteur de recherche est [www.google.fr](http://www.google.fr) : il en existe plusieurs autres (le second est "Bing" de Microsoft, mais on trouve de très bonne chose aussi sur [www.exalead.fr](http://www.exalead.fr), [www.duckduckgo.com](http://www.duckduckgo.com),

Un MOTEUR est un outil permettant une recherche systématique d'une expression dans des milliards de pages web existantes et référencées par des robots.

Un ANNUAIRE est un outil permettant une recherche thématique dans les notices descriptives de milliers de sites web, réalisées par des documentalistes 'référenceurs'



Plus d'informations sur le site spécialisé [www.abondance.fr](http://www.abondance.fr)

Typologie des outils de recherche : moteurs généralistes et spécialisés, métamoteurs, annuaires généralistes et spécialisés, portails?

## LES MOTEURS

L'hégémonie de Google est telle qu'il n'est plus habituel de trouver des classements d'usage des moteurs. Cela a été à la mode à la fin de la première décennie 2000

Palmarès des moteurs de recherche (décembre 2011) en France

<http://www.journaldunet.com/ebusiness/le-net/top-5-moteurs-de-recherche/>

<http://www.les-infostrategies.com/actu/11121337/le-top-5-des-moteurs-de-recherche-en-france>

1. Google : "couverture active" (part d'audience) = 85 %
2. Google images 36,7 %
3. Voila 15,9 %
4. Bing 15,8 %
5. Ask 14,7 %
6. Yahoo

Le tri de pertinence des moteurs

### Principes

Les moteurs mettent au point des "tris de pertinence" pour classer de façon automatique leurs résultats de recherche, afin de présenter en début de liste ceux qui obtiennent le meilleur score pour une requête donnée. Les algorithmes de tri sont différents en fonction des outils et plus ou moins performants et complexes. Ils ne sont généralement pas connus de façon précise et varient dans le temps pour chaque moteur. Les principaux critères utilisés sont les suivants :

#### \* **Par rapport à la requête de l'internaute :**

- position des mots dans la requête : Ainsi, sur Google, l'ordre des mots de la question n'est pas neutre.
- correspondance d'expression : similarité entre l'expression de la requête et l'expression correspondante dans un document

#### \* **Par rapport aux pages de résultats**

- "densité" des mots-clés : nombre d'occurrences du (des) terme(s) demandé(s) / nombre de termes de la page en question, une fois éliminés les mots vides.
- présence dans le titre ou dans le premier tiers de la page

- mise en exergue du texte (gras, taille des caractères)
- présence dans les méta-données\* (ce critère tend à perdre de son importance). Des outils comme Google ou Fast n'utilisent pas du tout ce critère, et Voila ne leur donne plus beaucoup d'importance.
- présence dans l'adresse de la page
- proximité des mots-clés sur la page

**\* Par rapport à la base de données du moteur :**

- rareté des mots (déterminé par le nombre d'occurrences du mot dans l'index) : des mots rares dans une requête ont une pondération plus importante que des mots communs
- popularité des pages : indice de clic (basé sur l'audience) ou indice de popularité (basé sur le principe de citation).

**La popularité comme mesure de pertinence**

Depuis quelques années, on a assisté à la naissance, au développement, puis au franc succès de deux nouvelles mesures de pertinence appelées respectivement "indice de clic" et "indice de popularité". Ces mesures s'ajoutent le plus souvent à d'autres "ingrédients" pour classer les résultats des moteurs, mais ils constituent aussi le critère de tri primordial des nouveaux venus inventeurs de ces technologies. Ces nouveautés, issues du "filtrage collaboratif", sont symptomatiques d'un certain désarroi des acteurs et utilisateurs du réseau face aux multiples difficultés d'un recueil rapide d'informations pertinentes.

**\* L'indice de clic**

Il s'agit ici d'analyser le comportement des internautes posant la même question au moteur et de privilégier dans le classement les pages les plus "cliquées", et sur lesquelles le temps passé est le plus important. Il permet donc de classer les résultats des requêtes les plus populaires, en récupérant le jugement implicite de communautés d'utilisateurs. Fonctionne donc en "tâche de fond" sur un moteur existant, la base s'enrichissant ainsi.

Direct Hit, racheté par Ask Jeeves en 2001, puis devenu Teoma, était la référence dans ce domaine et fut utilisé par de nombreux moteurs comme Lycos et MSN (plus de 50 sites clients), mais aussi Ask Jeeves. Alta Vista et Inktomi ont développé leur propre système sur un principe similaire. Mais Ask Jeeves a ensuite décidé de centraliser ses efforts de développement sur le moteur Teoma.

**\* L'indice de popularité**

On s'intéresse ici aux "backlinks" ou "liens à l'arrivée", c'est à dire au nombre et à la qualité des liens pointant sur une page : on mesure ainsi sa popularité, et donc selon les concepteurs de ces technologies, sa pertinence. Les anglophones disent pour mieux expliquer le principe de l'indice de popularité : "It's not

what you know, it's who knows you". En d'autres termes, le plus important n'est pas ce que vous dites ou ce que vous savez, mais qui vous connaît.

Le principe, rendu célèbre par le moteur Google, n'est pas totalement nouveau. Ne mesure-t-on pas la crédibilité d'un auteur scientifique au nombre de citations qui sont faites sur ses articles ?

Google examine la structure des liens sur l'ensemble du web. Quand on fait une recherche, un URL avec un fort "page rank" a plus de chance d'être listée en premier. Chaque page de l'index de Google est notée : le "page rank" est une propriété de la page en elle-même, indépendante des requêtes effectuées : elle équivaut à la probabilité que l'internaute aboutisse à cette page sur Internet.

Le tri des résultats pour une requête intègre d'autres critères plus classiques, dont bien entendu la présence des termes de la requête dans les pages de résultat, ou identifiée comme pertinente via l'analyse du contexte des liens.

Le grand avantage du système est de donner une meilleure visibilité aux sites incontournables du domaine de recherche. L'inconvénient majeur est là encore, de pénaliser les nouveaux venus peu connus.

**Les sociétés spécialisées dans le référencement** cherchent bien entendu à connaître le plus précisément possibles les critères clés de chaque moteur. L'objectif est de faire apparaître en bonne position (ranking) les pages web de leurs clients sur les listes de résultats à une requête comportant certains mots-clés.

Ce travail de référencement se fait parfois au mépris de l'éthique et donne lieu à une activité de "spamdexing" ou "spamming". (Création ou modification d'un document avec l'intention de tromper un catalogue ou un système de classement électronique. Toute technique qui a pour objectif d'augmenter la position potentielle d'un site aux dépens de la qualité de la base de données du moteur de recherche. Définition issue du glossaire réalisé par les membres francophones de la liste de diffusion I-Search Digest hébergé par le fournisseur d'hébergement IDF [www.idf.net/mdr/glossaire.html](http://www.idf.net/mdr/glossaire.html)).

Voir les sites spécialisés dans le référencement comme Webrankinfo ([www.webrankinfo.com](http://www.webrankinfo.com)) ou Abondance.com, précité, qui donne de précieux conseils, notamment pour Google. Ces techniques sont aussi connues sous le nom de SEO pour Search Engine Optimisation.

**A noter** : Le modèle du positionnement payant s'est aujourd'hui imposé, et tous les moteurs proposent des "résultats sponsorisés" (c'est aujourd'hui souvent l'indice de clic qui est favorisé) à ne pas confondre avec les résultats "normaux"

---

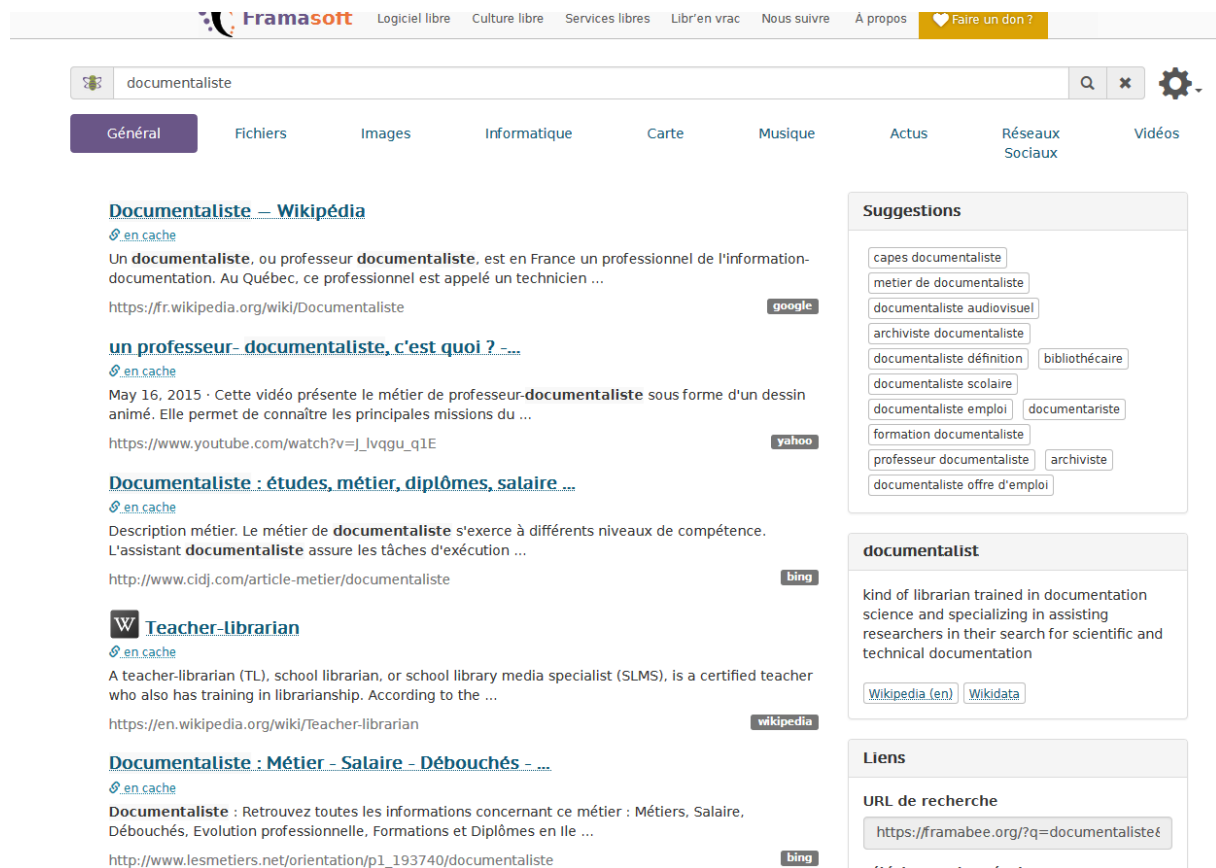
## LES METAMOTEURS

Un **métamoteur** (ou **méta-moteur**) ou un **méta-chercheur** est un moteur de recherche qui puise ses informations à travers plusieurs [moteurs de recherche](#) généralistes. De manière plus précise, le métamoteur envoie ses requêtes à plusieurs moteurs de recherche et retourne les résultats de chacun

d'eux. Le métamoteur permet aux utilisateurs de n'entrer le sujet de leur recherche qu'une seule fois tout en accédant aux réponses de plusieurs moteurs de recherche différents. (source wikipedia)

[https://fr.wikipedia.org/wiki/Liste\\_de\\_moteurs\\_de\\_recherche#M.C3.A9tamoteurs](https://fr.wikipedia.org/wiki/Liste_de_moteurs_de_recherche#M.C3.A9tamoteurs)

Exemple : Framabee.org



**documentaliste**

Général Fichiers Images Informatique Carte Musique Actus Réseaux Sociaux Vidéos

**Documentaliste – Wikipédia**  
en cache  
Un **documentaliste**, ou professeur **documentaliste**, est en France un professionnel de l'information-documentation. Au Québec, ce professionnel est appelé un technicien ...  
<https://fr.wikipedia.org/wiki/Documentaliste> **google**

**un professeur documentaliste, c'est quoi ? -...**  
en cache  
May 16, 2015 · Cette vidéo présente le métier de professeur **documentaliste** sous forme d'un dessin animé. Elle permet de connaître les principales missions du ...  
[https://www.youtube.com/watch?v=j\\_lvqgu\\_q1E](https://www.youtube.com/watch?v=j_lvqgu_q1E) **yahoo**

**Documentaliste : études, métier, diplômes, salaire ...**  
en cache  
Description métier. Le métier de **documentaliste** s'exerce à différents niveaux de compétence. L'assistant **documentaliste** assure les tâches d'exécution ...  
<http://www.cidj.com/article-metier/documentaliste> **bing**

**Teacher-librarian**  
en cache  
A teacher-librarian (TL), school librarian, or school library media specialist (SLMS), is a certified teacher who also has training in librarianship. According to the ...  
<https://en.wikipedia.org/wiki/Teacher-librarian> **wikipedia**

**Documentaliste : Métier - Salaire - Débouchés - ...**  
en cache  
**Documentaliste** : Retrouvez toutes les informations concernant ce métier : Métiers, Salaire, Débouchés, Evolution professionnelle, Formations et Diplômes en Ile ...  
[http://www.lesmetiers.net/orientation/p1\\_193740/documentaliste](http://www.lesmetiers.net/orientation/p1_193740/documentaliste) **bing**

**Suggestions**

- capes documentaliste
- metier de documentaliste
- documentaliste audiovisuel
- archiviste documentaliste
- documentaliste définition
- documentaliste scolaire
- documentaliste emploi
- formation documentaliste
- professeur documentaliste
- documentaliste offre d'emploi
- documentariste
- bibliothécaire
- archiviste

**documentalist**

kind of librarian trained in documentation science and specializing in assisting researchers in their search for scientific and technical documentation

[Wikipedia \(en\)](#) [Wikidata](#)

**Liens**

**URL de recherche**

<https://framabee.org/?q=documentaliste&...>

### LES REPERTOIRES DE RECHERCHE (OU ANNUAIRES)

#### Principe des répertoires de recherche

- \* "Collections" généralistes ou spécialisées de sites web classées par catégories organisées hiérarchiquement (le H de Yahoo signifie « hierarchically » Yet Another Hierarchically Organized Oracle )
- \* Filtrage et classement “ manuels ” : la sélection peut être plus ou moins rigoureuse, avec une évaluation et une description des sites éventuellement enrichies.
- \* Pas d’indexation en texte intégral des pages des sites.
- \* Outils de première approche : Donnent une vue d'ensemble d'un domaine à l'utilisateur, qui peut ensuite naviguer à l'intérieur des sites indiqués pour aller plus loin.
- \* Ne gèrent pas les requêtes complexes, mais permettent généralement de faire une recherche par mot-clé sur une catégorie seule.
- \* Problèmes de mise à jour et de “ désherbage ”.

#### Modes de recherche

- \* Recherche dans le plan de classement : Cette méthode est parfois complexe, aucune norme n'existant pour l'arborescence des répertoires. Les sites sont indiqués par ordre alphabétique.
- \* Recherche par mot clé : la recherche se fait sur les champs suivants : intitulés des catégories, titres des sites, résumé des sites, adresses URL des sites. Avec ce mode de recherche, les résultats bénéficient généralement d'un classement de pertinence opéré uniquement sur les fiches descriptives des sites. L'Open Directory ne recherche pas sur les catégories.

#### Utilisation

Les répertoires sont à réserver pour des recherches plutôt thématiques, ou sur des mots clés assez généralistes ; notons toutefois que les catégories deviennent au fil du temps de plus en plus "pointues" en fonction du sujet.

Si l'on utilise des mots clés trop précis, ou trop de mots clés, la plupart des répertoires passent le relais à des moteurs de recherche partenaires qui effectuent des recherches sur le texte intégral des pages web.

Les répertoires sont aussi utiles :

- \* pour se faire une idée du vocabulaire utilisé dans un domaine
- \* pour retrouver, à partir d'un site web donné, d'autres sites traitant du même sujet
- \* pour trouver des sites fédérateurs ou portails spécialisés
- \* pour obtenir rapidement tous les sites d'une organisation importante.

---

## LES ENCYCLOPÉDIES

Archétype : Wikipedia [https://fr.wikipedia.org/wiki/Wikipédia:A\\_propos](https://fr.wikipedia.org/wiki/Wikipédia:A_propos)

# Wikipédia:À propos

---

**Wikipédia** est une [encyclopédie](#) écrite par des volontaires sur internet, universelle, multilingue, à laquelle chacun peut collaborer, immédiatement et fonctionnant sur le principe du [wiki](#). Wikipédia a pour objectif d'offrir un [contenu libre](#), [objectif](#) et [vérifiable](#) que chacun peut modifier et améliorer, sans nécessité de s'enregistrer. Tous les articles de Wikipédia sont un *travail en progression* qui peut être modifié et amélioré par tout le monde.

**Pour plus d'informations, lire l'article [Wikipédia](#).**

---

Moteur de réponses (et non moteur de recherche)

En anglais seulement : [www.quora.com](http://www.quora.com) , [www.answers.com](http://www.answers.com)

En français <https://theconversation.com/fr/topics>

Voir aussi Projet Sindbad (BNF)

[http://www.bnf.fr/fr/collections\\_et\\_services/poser\\_une\\_question\\_a\\_bibliothecaire/s.sindbad\\_reponses\\_par\\_t\\_hemes.html?first\\_Art=non](http://www.bnf.fr/fr/collections_et_services/poser_une_question_a_bibliothecaire/s.sindbad_reponses_par_t_hemes.html?first_Art=non)

[http://bnf.libguides.com/chercher\\_trouver](http://bnf.libguides.com/chercher_trouver)

## Poser votre question

Le service SINDBAD (Service d'INformation Des Bibliothécaires A Distance) fournit gratuitement des références de documents sur tous les sujets et des informations factuelles (éléments biographiques, faits, chiffres, dates).



## 2.3 - LA RECHERCHE SUR LE WEB INVISIBLE

Il s'agit de l'ensemble des pages non localisables et/ou non indexables par les outils. Le web invisible correspond à plusieurs types de ressources :

- ✓ Pages dont les caractéristiques techniques rendent difficiles, sinon impossible l'indexation par les moteurs : frames, javascripts modifiant le contenu, technologies propriétaires.
- ✓ Pages qui n'ont fait l'objet ni d'un référencement direct, ni d'aucun lien d'une autre page.
- ✓ Pages nécessitant une identification de la part de l'internaute
- ✓ Pages dont le contenu indique aux moteurs qu'ils ne doivent pas l'indexer
- ✓ Page produite à partir de bases de données ou d'applications, et dont l'URL comporte des paramètres non exploitables par la plupart des moteurs
- ✓ Page produite à partir de données saisies par l'utilisateur via un formulaire html. Exemple : les résultats de l'interrogation d'une base de données avec des critères de recherche entrés par l'utilisateur.

*(définition mise au point par les formateurs internet ADBS)*

On ne connaît pas du tout la taille du web invisible

**Une certitude** : le web invisible croît plus rapidement que le web visible, du fait de la multiplication des bases de données à interface web, et de l'explosion du web dynamique.

**A noter** : les fichiers pdf ou flash, autrefois partie intégrante du web invisible, sont aujourd'hui indexés par plusieurs moteurs, Google en tête.

---

### INFORMATIONS DE BASE SUR LES META-DONNEES

Il s'agit au départ de balises du langage html qui permettent de donner des informations (description, mots-clés) sur le contenu d'une page web.

Elles se trouvent dans l'en-tête HTML de la page Web, (le "HEAD") et fournissent des informations qui ne sont pas visibles par les navigateurs. Les méta-tags les plus courants (et les plus utiles pour les moteurs de recherche) sont KEYWORDS (mots-clés) et DESCRIPTION.

Pour visualiser les méta-tags : Affichage Source (Explorer) / CTRL U (Firefox)



Le méta-tag KEYWORD permet à l'auteur de souligner l'importance de certains mots et phrases utilisés ou non dans sa page. Certains moteurs de recherche tiendront compte de cette information - d'autres l'ignoreront. Certains moteurs donneront en plus un « coup de pouce » dans le classement pour certains documents au cas où le mot clé de requête se trouve dans les méta-tags, mais ils peuvent pénaliser une page où un terme est répété plusieurs fois dans la balise meta keyword..

Le méta-tag DESCRIPTION permet à l'auteur de contrôler le texte affiché quand la page paraît au niveau des résultats d'une recherche. Certains moteurs de recherche peuvent ignorer cette information. Contrairement à KEYWORDS, DESCRIPTION est en langage naturel.

Pour pallier la "faiblesse" des balises méta classiques, certains groupements travaillent à mieux décrire les documents sur Internet. On pourra utilement se référer au "Dublin Core", métadonnée de 15 éléments destinée à la description générale des documents, qui est d'ores et déjà utilisée via les balises méta par certains organismes, y compris en intranet et normalisé dans la norme ISO 15836 [http://fr.wikipedia.org/wiki/ISO\\_15836](http://fr.wikipedia.org/wiki/ISO_15836) . Le Dublin Core, considéré comme un bon candidat pour une norme internationale, est le fruit du travail depuis 1995 d'une cinquantaine de chercheurs et professionnels issus du monde de la documentation et des bibliothèques, de l'informatique, de la codification des informations. L'ensemble fut initié par l'OCLC (Online Computer Library Center) en accord avec le NCSA (National Center for supercomputing applications). Le Dublin Core doit son nom à la première réunion de travail en juin 95 à Dublin Ohio dans les locaux de l'OCLC.

[http://fr.wikipedia.org/wiki/Dublin\\_Core](http://fr.wikipedia.org/wiki/Dublin_Core)

Élément	Élément (anglais)	Commentaire
1. Titre (métadonnée)	Title	Titre principal du document
2. Créateur (métadonnée)	Creator	Nom de la personne, de l'organisation ou du service à l'origine de la rédaction du document
3. Sujet (métadonnée) ou mots clés	Subject	Mots-clefs, phrases de résumé, ou codes de classement
4. Description (métadonnée)	Description	Résumé, table des matières, ou texte libre. Raffinements : table des matières, résumé
5. Éditeur	Publisher	Nom de la personne, de l'organisation ou du service à l'origine de la publication du document
6. Contributeur	Contributor	Nom d'une personne, d'une organisation ou d'un service qui contribue ou a contribué à l'élaboration du document. Chaque contributeur fait l'objet d'un élément Contributor séparé
7. Date (métadonnée)	Date	Date d'un évènement dans le cycle de vie du document
8. Type de ressource	Type	Genre du contenu
9. Format	Format	Type MIME, ou format physique du document

10. Identifiant de la ressource	Identifier	Identificateur non ambigu : il est recommandé d'utiliser un système de référencement précis, afin que l'identifiant soit unique au sein du site, par exemple les URI ou les numéros ISBN. Raffinement : Is Available At
11. Source	Source	Ressource dont dérive le document : le document peut découler en totalité ou en partie de la ressource en question. Il est recommandé d'utiliser une dénomination formelle des ressources, par exemple leur URI
12. Langue (métadonnée)	Language	
13. Relation (métadonnée)	Relation	Lien avec d'autres ressources. De nombreux raffinements permettent d'établir des liens précis, par exemple de version, de chapitres, de standard, etc.
14. Couverture (métadonnée)	Coverage	Couverture spatiale (point géographique, pays, régions, noms de lieux) ou temporelle
15. Droits (métadonnée)	Rights	Droits de propriété intellectuelle, Copyright, droits de propriété divers

Un 16<sup>e</sup> élément apparaît parfois, l'Audience, mais il ne figure pas dans la liste de la norme ISO 15836.

**Application : trouver des pages web exposant des métadonnées en Dublin Core**

## 2.4 - EVOLUTION DES OUTILS DE RECHERCHE

Extrait du sommaire de Net Recherche 2013, Véronique Mesguich & Armelle Thomas, De Boeck, Adbs.

### 7. Évolution des moteurs de recherche : dix tendances actuelles . . . . .

- 7.1 Simplifier la syntaxe et aider l'utilisateur
- 7.2 Permettre une recherche « universelle »
- 7.3 Personnaliser son moteur de recherche
- 7.4 Exploiter les technologies de clustering
- 7.5 Permettre une visualisation cartographique des résultats
- 7.6 Rechercher en langage naturel.
- 7.7 Rechercher dans les fichiers multimédias
- 7.8 Rechercher depuis les « smartphones »
- 7.9 Traduire de manière toujours plus efficace
- 7.10 Permettre une recherche « temps réel »

#### COMMENT S'ORGANISE LA RECHERCHE D'INFORMATION ?

##### LES DIX REGLES D'OR DE LA RECHERCHE D'INFORMATION SUR INTERNET

1. **"Affiner"** savoir poser les bonnes questions : sa question (type de recherche, sujet précis et objectif, étude des concepts, recherches préliminaires éventuelles), choisir ses stratégies de recherche. (OA "lorsqu'on a une recherche à faire sur le web, la première chose à faire, c'est de ne pas aller sur le web")
2. **Maîtriser** les outils de navigation et de recherche : gestion des signets, récupération des données, répertoires, moteurs et méta-moteurs. Pour les moteurs, utiliser au moins deux moteurs ayant des approches différentes et complémentaires.
3. **Trouver** de bons points de repère : annuaires et "bons sites" (associations professionnelles, experts, usuels du domaine) dans un domaine :
  - Retrouver les équivalents de ses sources habituelles (d'où l'importance d'avoir une idée, même approximative, de l'offre documentaire dans le domaine recherché).
  - Compléter avec les sources originales
  - Trouver les répertoires et "méta-pages" spécialisées.

Une adresse fiable qui renvoie directement au sujet d'une recherche constitue un bon point de départ parce que :

L'administrateur d'un bon site spécialisé est généralement averti de l'existence et la création des autres sites de la spécialité : Il sélectionne les meilleures références et parfois les commente ; Il passe du temps sur le réseau dans son domaine de compétence ; Il met en jeu son expertise.
4. **Toujours analyser** l'information : recouper l'information, faire preuve d'esprit critique, évaluer rapidement
5. **Utiliser** en cours de recherche son carnet d'adresses pour **garder trace** des sites ou pages intéressants mais momentanément hors sujet, et "noter" rapidement les ressources enregistrées.

6. **Savoir se limiter** dans le temps : ne pas se rendre esclave d'une recherche d'exhaustivité à tout prix, ne pas s'obstiner en vain. Internet contribue souvent à répondre à la question "où trouver" (chercher l'info qui conduira à l'info).

7. **Choisir** les bons mots-clés

8. **Rester clair** sur ses objectifs, sa stratégie et ses critères de choix établis auparavant face à "l'hyper-choix". Rester vigilant sur la trajectoire parcourue et celle qui reste à parcourir. "on ne doit pas rechercher l'info de la même manière suivant que l'on est novice ou expert sur un sujet.

Le novice recherche les sites web les plus riches et les plus visités. Il n'a pas de temps à perdre et veut éviter le bruit. Il obtient des résultats rapides, après la phase d'acclimatation au problème.

L'expert n'est pas intéressé par les sites classiques. Il recherche au contraire le bruit afin de trouver le "signal faible" qui lui donnera l'avantage. Il est prêt à y consacrer beaucoup de temps. (il fait beaucoup d'efforts pour des résultats marginaux)

9. **Conjuguer harmonieusement** recherche dans les outils classiques, web invisible, presse et actualité et navigation hypertexte : la recherche d'information sur Internet est un processus itératif qui oblige à passer par différents modes d'accès à l'information.

10. **Etre "agile"** : développer une lecture rapide, lancer plusieurs recherches à la fois, savoir rebondir d'une information à l'autre, d'un outil à l'autre, d'un article à une institution. Se souvenir qu'il n'existe pas de méthode infaillible et que chercher l'information sur Internet, c'est avant tout un état d'esprit. Ainsi, si je cherche le premier producteur de statistiques en Irlande, je peux commencer, sans trop de risques d'erreurs, par faire l'hypothèse que l'INSEE propose des liens vers ses homologues européens.

---

#### PROBLEMATIQUE : FAUT-IL COMMENCER UNE RECHERCHE SUR INTERNET ?

Internet est-il complémentaire à d'autres supports ou se suffit-il à lui-même ? . On trouvera rarement matière à une étude complète d'un sujet via Internet (test : essayez avec un sujet que vous connaissez bien = vous serez toujours très déçu).

En revanche, bien (et rationnellement) utilisé, le Web sera souvent plus rapide et moins cher que d'autres supports pour des recherches de type "questions-réponses".

Enfin, Internet et ses différents services (mail, newsgroups, mailing lists) se prêtent bien à la pratique de la veille, de par leur caractère mouvant, décloisonné, international.

---

### SUR QUELS CRITERES ?

Les questions de base à se poser, à poser à son "client"

- Temps à consacrer à la recherche (recherches de long terme ou "recherche - blitz")
- Coûts / abonnement : quel budget, quelle préparation ?
- Confidentialité (ressources externes / internes) : dans quel environnement (ouvert ou fermé) vais-je pouvoir poser la question ? Notion d'anonymat / d'hétéronymat...
- Ressources documentaires déjà épuisées par le client lui-même (impasses, fausses routes...)

---

### QUAND UTILISER QUELS OUTILS ?

La réponse à cette question ne peut pas être définitive. Rappelons que la recherche d'information sur Internet n'est pas une science, et tout dépend aussi de son expérience de la recherche et du Web, et de sa façon de travailler.

Disons en simplifiant beaucoup...

#### En fonction du type de recherches

- \* Recherches larges ou première approche : annuaires généralistes, encyclopédies collaboratives en ligne, forum de questions/réponses
- \* Recherche d'information ponctuelle (tous secteurs) : moteurs généralistes
- \* Recherche sur des données de nature bien définie (statistiques, pays, presse, teurs...) : annuaires et outils spécialisés sur ce type de recherche
- \* Recherches récurrentes sur un sujet: identification de sites via pages de liens ou annuaires spécialisés, puis recherche par navigation / méta-moteur off-line
- \* Recherches précises sur noms ou chaînes de caractères (sans booléens) : méta-moteurs.

**En fonction de sa connaissance du sujet :**

	Faible connaissance du sujet	Bonne connaissance du sujet
"Question-réponse"	.Recherche sur les moteurs ou méta-moteurs  .Remonter à un concept plus généraliste et utiliser les annuaires	."Sites de référence" (Sites spécialisés sur le sujet, repérés au préalable)
"Tout savoir sur"	.Annuaires pour identifier les bons sites et les bons mots clés  .Recherche sur " sites de référence"  .Recherche sur moteurs	. "Sites de référence" complétés par recherches sur moteurs ou méta-moteurs  .Personne référence.

### RECOURIR A DES EXPERTS / DEVELOPPER SON EXPERTISE

Usage des listes de discussion

Exemple : <https://listes.adbs.fr/sympa/info/adbs-info>

Accès aux archives de la liste : <https://listes.adbs.fr/sympa/arc/adbs-info>

Application : savoir chercher dans des archives de listes, (s'abonner à une liste, participer à une liste, etc.)

Repérer des "moteurs" de liste ou utiliser la recherche avancée de Google (ou d'un autre moteur) en imposant la racine de l'adresse électronique contenant les archives

Exemple : usage de la requête avancée

veille google insite:<https://listes.adbs.fr/sympa/arc/adbs-info>

**Web**    Actualités    Vidéos    Images    Shopping    Plus ▾    Outils de recherche

Environ 773 résultats (0,43 secondes)

**adbs-info - INFO : Veille de C. Griset le 27/01/11 - arc**

<https://listes.adbs.fr/sympa/arc/adbs-info/2011-01/msg00119.html>

28 janv. 2011 - From: **Liste ADBS-INFO** <moderateur@adbs.fr> ...

<http://www.lemonde.fr/technologies/article/2011/01/27/google-limite-des-recherches-liees-...>

**adbs-info - INFO : Veille de Christine Griset du 28/7/8 - arc**

<https://listes.adbs.fr/sympa/arc/adbs-info/2008-07/msg00124.html>

31 juil. 2008 - From: **Liste ADBS-INFO** <moderateur@adbs.fr> ... **Veille** Technologies

... <<http://www.neteco.com/153370-anciens-google-lancent-moteur-cuil.html?xtor=EPR-1>> ... avec notamment : LePost, **site** collaboratif crée par Le Monde.

..... et aux archives <<<https://listes.adbs.fr/sympa/arc/adbs-info>><https://listes.adbs.fr/sympa/arc/adbs-info>>



## ASTUCES POUR IDENTIFIER RAPIDEMENT DES SOURCES D'INFORMATION, DES EXPERTS, ETC.

(Voir chapitre : Evaluer une information sur Internet)

- Usage des commandes google : related: , link: (popularité, similarité, recherche de « retours-liens »,...)
- Présence dans les forums, mentions dans les sites sociaux (bon ou « bad-buzz »)
- Twitter ou autres ressources Web 2.0
- Newsgroups ou liste de discussion...
- Nom de domaine (connaissance et respect des normes d'URL, occupation des espaces de noms, taille et graphie de l'URL, choix de l'extension du nom de domaine...)
- Présence de meta-tag dans le code source des pages (savoir lire le source : CTRL + U)
- Bons usages en matière d'ergonomie en général (couleurs, taille d'écran, langue, orthographe...)
- Respect de la législation et réglementation française et européenne en matière de diffusion d'information sur internet : mentions légales, CNIL, DAVDSI, LCEN,...
- Connaissance et liens vers les sites « pivots » (incontournables dans la matière : sites officiels, gouvernementaux,...)
- Plus généralement : E-réputation : présence sur le web, sur Wikipedia, ...

### 3.3 - CONSTRUCTION DE LA REQUETE

#### CHOIX DES MOTS-CLES

(extrait Net Recherche)

##### 1. Comment choisir ses mots-clés ?

**Quand ?** La sélection des mots-clés s'effectue après le choix d'une stratégie de recherche. En effet, le choix sera fondamentalement différent si l'on cherche un portail thématique, ou une source susceptible de fournir l'information ou l'information précise immédiatement. Pour simplifier, disons que

dans le premier cas, les mots-clés seront "le plus large possible", dans le second cas, ils seront "le plus précis possible".

**Un ou plusieurs ?** On procédera par étape pour affiner éventuellement sa recherche à l'aide de plusieurs mots-clés. Si le nombre de résultats est faible avec un seul mot-clé précis (exemple : 100 résultats sur un moteur), inutile de préciser davantage. Donc, utiliser d'abord un seul mot clé (ou expression) quand la terminologie ou l'association terminologique est très spécifique. Sinon, travailler du plus général au plus spécifique (mais choisir les synonymes appropriés pour le terme générique : par exemple si je m'intéresse à un film qui s'appellerait "Demain, dès l'aube", on pourrait écrire 'film OR cinéma "demain, dès l'aube"').

**Pour ou contre le SAUF ?** On peut aussi isoler les mots-clés à exclure absolument car générateurs de bruit (opérateur SAUF ou signe -). Attention toutefois à ne pas aller trop vite, de peur de passer à côté de documents pertinents : Ainsi, si je cherche des informations sur les énergies alternatives autres que solaires, je peux être tenté d'"envoyer" au moteur une équation du type +"énergies alternatives" –solaires. Mais je n'aurai pas alors les ressources qui abordent successivement **toutes** les énergies alternatives. C'est pourquoi il est parfois plus judicieux de repérer une notion discriminante de son sujet de recherche plutôt que de d'utiliser sans réflexion le SAUF.

**Majuscules, minuscules, accents ?** De façon générale, les moteurs sont insensibles à la casse des caractères et retourneront le même nombre de résultats pour python, PYTHON, ou Python. La situation est plus contrastée en ce qui concerne les accents : si Google traite de manière identique les mots accentués ou non (l'expérience montre que c'est loin d'être toujours évident), Yahoo par exemple procède différemment : ils ne retournent pour un mot-clé accentué que les mots contenant l'accent, mais pour une requête non accentuée, ils retournent les mots avec ou sans accent. Bref, il convient de faire attention à l'utilisation des accents avant d'utiliser un moteur.

**Troncatures ?** La troncature permet de remplacer plusieurs caractères sur la fin des mots, mais cette possibilité devient fort rare sur le web : les trois grands moteurs Google, Yahoo et MSN ne proposent pas cette option, et notamment, un mot-clé indiqué au singulier sera traité comme tel. Il est donc important de prévoir au moins l'alternative du mot au pluriel, sous peine d'occulter de nombreux résultats pertinents. En revanche, le challenger Exalead supporte la troncature avec le caractère \*

Attention : sur Google, l'astérisque \* n'a pas valeur de troncature, mais de "joker" (mot dans une expression entre guillemets). Ainsi "droit \* public" va trouver "droit international public", "droit pénal public" "droit du sport public" (avec les mots vides).

**Ordre des mots ?** Il peut avoir de l'importance selon les moteurs, non pas bien entendu pour le nombre de réponses, mais pour le classement des résultats : c'est par exemple vrai pour Google ou pour Voilà.

**Et les synonymes ?** Il est important d'explorer la terminologie du domaine de recherche, pour repérer les synonymes (très rares sont les moteurs travaillant sur les concepts). De façon générale, les premiers documents intéressants récupérés permettent de valider, compléter ou revoir ses mots-clés.

#### **Astuces pour identifier des synonymes et/ou mots associés**

- \* Utiliser un dictionnaire de synonymes tel celui du laboratoire de linguistique du CNRS pour les termes en français pour le français et l'anglais <http://dico.isc.cnrs.fr/> (ou <http://www.crisco.unicaen.fr> )
- \* Utiliser un thésaurus de son domaine (en ligne gratuit, ou acheté comme par exemple celui de la base INSPEC (<http://www.theiet.org/resources/library/index.cfm> ) ou la liste de Sylvie Dalbin sur <http://www.dmoz.org/World/Francais/Références/Thésaurus/>)
- \* Utiliser un moteur de recherche travaillant à partir de dictionnaires, encyclopédies, thesaurus, tel pour les termes en anglais FreeDictionary <http://www.thefreedictionary.com/> ou le canadien <http://www.granddictionnaire.com> . Voir aussi [www.thesaurus.com](http://www.thesaurus.com).
- \* Faire une recherche sur une base de données bibliographique du domaine dans lequel se situe le sujet, utilisant une indexation manuelle (dewey, autre plus spécialisée avec un thésaurus par exemple). Repérer alors comment sont indexés quelques documents pertinents, quelle est la terminologie retenue.
- \* Utiliser l'option define: sur Google (aujourd'hui disponible en français). On a aussi l'option Google Suggest

Ce site fait partie de **EuroVoc** Thésaurus multilingue de l'Union européenne

Europa &gt; Page d'accueil d'EuroVoc &gt; Télécharger la version thématique

Langue du contenu :

(en) English

Recherche simple

Recherche avancée

Navigation

Naviguer par domaine

Télécharger

Par domaine

Index permuté

Liste multilingue

Index alphabétique

SKOS/XML

Vos propositions

Contribuer

Nouveaux concepts  
approuvés

Registered User

**Télécharger la version thématique**Télécharger la présentation thématique au format PDF (dans un fichier Zip).  
Sélectionnez les domaines et langues qui vous intéressent.

- |  |   |
|--|---|
| <input type="checkbox"/> 04 VIE POLITIQUE                        | <input type="checkbox"/> BG Български   |
| <input type="checkbox"/> 08 RELATIONS INTERNATIONALES            | <input type="checkbox"/> ES Español     |
| <input type="checkbox"/> 10 COMMUNAUTÉS EUROPÉENNES              | <input type="checkbox"/> CS Čeština     |
| <input type="checkbox"/> 12 DROIT                                | <input type="checkbox"/> DA Dansk       |
| <input type="checkbox"/> 16 VIE ÉCONOMIQUE                       | <input type="checkbox"/> DE Deutsch     |
| <input type="checkbox"/> 20 ÉCHANGES ÉCONOMIQUES ET COMMERCIAUX  | <input type="checkbox"/> ET Eesti       |
| <input type="checkbox"/> 24 FINANCES                             | <input type="checkbox"/> EL Ελληνικά    |
| <input type="checkbox"/> 28 QUESTIONS SOCIALES                   | <input type="checkbox"/> EN English     |
| <input type="checkbox"/> 32 ÉDUCATION ET COMMUNICATION           | <input type="checkbox"/> FR Français    |
| <input type="checkbox"/> 36 SCIENCES                             | <input type="checkbox"/> IT Italiano    |
| <input type="checkbox"/> 40 ENTREPRISE ET CONCURRENCE            | <input type="checkbox"/> LV Latviešu    |
| <input type="checkbox"/> 44 EMPLOI ET TRAVAIL                    | <input type="checkbox"/> LT Lietuvių    |
| <input type="checkbox"/> 48 TRANSPORTS                           | <input type="checkbox"/> HU Magyar      |
| <input type="checkbox"/> 52 ENVIRONNEMENT                        | <input type="checkbox"/> MT Malti       |
| <input type="checkbox"/> 56 AGRICULTURE, SYLVICULTURE ET PÊCHE   | <input type="checkbox"/> NL Nederlands  |
| <input type="checkbox"/> 60 AGRO-ALIMENTAIRE                     | <input type="checkbox"/> PL Polski      |
| <input type="checkbox"/> 64 PRODUCTION, TECHNOLOGIE ET RECHERCHE | <input type="checkbox"/> PT Português   |
| <input type="checkbox"/> 66 ÉNERGIE                              | <input type="checkbox"/> RO Română      |
| <input type="checkbox"/> 68 INDUSTRIE                            | <input type="checkbox"/> SK Slovenčina  |
| <input type="checkbox"/> 72 GÉOGRAPHIE                           | <input type="checkbox"/> SL Slovenščina |
| <input type="checkbox"/> 76 ORGANISATIONS INTERNATIONALES        | <input type="checkbox"/> FI Suomi       |
|  | <input type="checkbox"/> SV Svenska     |
|  | <input type="checkbox"/> HR Hrvatski    |
|  | <input type="checkbox"/> SR Српски      |

<http://eurovoc.europa.eu/drupal/?q=fr>

## 3.4 - TECHNIQUES DE RECHERCHE AVANCEES

### OPERATEURS AVANCES DE RECHERCHE

Quels sont les opérateurs de recherche indispensables ? . .

1 Les guillemets « »... pour manipuler les expressions . .

2 Le +... pour imposer un mot ; le – ... pour l'exclure .

3 intitle:... pour chercher dans le titre . .

4 site:... pour cibler un domaine . .

5 filetype:... pour trouver les documents directement dans le bon format . . . . .



Recherche avancée

[Conseils de recherche](#) | [À propos](#)

Utilisez le formulaire ci-dessous pour lancer une recherche avancée ; les résultats apparaîtront ici.

#### Rechercher les pages contenant...

tous les mots suivants :

cette expression exacte :  [astuce](#)

au moins un des mots suivants :  OR  OR  [astuce](#)

#### Mais ne pas afficher les pages contenant...

l'un des mots suivants :  [astuce](#)

#### Besoin d'outils supplémentaires ?

Résultats par page :  Cette option ne s'applique pas à la [recherche instantanée Google](#).

Langue :

Type de fichier :

Recherche sur un site ou un domaine :

(exemple : youtube.com, .gouv.fr)

#### [Date, droits d'utilisation, région et autres](#)

Date : (ancienneté de la page)

Droits d'utilisation :

Vos mots clés s'affichent :

Pays/territoire :

[SafeSearch](#) :  Désactivé  Activé

Recherche avancée

#### Outils de recherche de pages spécifiques :

Pages similaires à la page suivante :

Pages avec un lien vers la page suivante :



Que cherchez-vous ? Fermer

- Phrases exactes ex : "être ou ne pas être"
- Termes exacts ex : +le parrain
- Termes optionnels ex : vache OPT folle
- Exclure des termes ex : développement -durable
- Recherche par préfixe ex : Stéphan\*
- Recherche par proximité ex : développement NEXT durable
- Logique ex : (développement AND durable) OR pollution
- Recherche phonétique ex : soundslike:ornitorinc
- Orthographe approchée ex : spellslike:exalead
- Choisissez une langue  ex : harry potter language:fr

Où cherchez-vous ?

- Recherche sur un site précis ex : star wars site:flickr.com
- Recherche dans le titre ex : intitle:"site officiel"
- Recherche dans l'URL ex : inurl:musique
- Recherche par liens ex : link:www.exalead.com

Quelle période vous intéresse ?

- Recherche avant une date ex : pollution before:01/01/2000
- Recherche après une date ex : pollution after:01/01/2000

Ajouter un ra

118 218

**Labs**

Nouveaux produits et technologies Exalead



[Bing Help Home >](#)

[Imprimer](#) | [Courrier é](#)

## Aide de Bing

### Découvrir la Bing

[Optimiser la recherche](#)  
[Search tips and techniques](#)  
[Bing FAQ](#)

### Fonctionnalités utiles de la Bing

[Obtenir des suggestions de recherche](#)  
[Afficher l'historique de vos recherches](#)  
[Trouver des réponses instantanées](#)  
[Découvrir la page d'accueil enrichie](#)  
[Décidez avec l'aide de vos amis](#)  
[Traduire les résultats de la recherche](#)

### Trouver ce dont vous avez besoin

[Retrouver les photos de vos artistes préférés](#)  
[Rechercher et regarder des vidéos](#)  
[Rechercher des produits en ligne](#)  
[Voir les avis des utilisateurs d'un coup d'œil.](#)  
[Obtenir les dernières nouvelles](#)  
[Trouver les prévisions météorologiques](#)  
[Rechercher des sites Web cinématographiques, de produits et de sociétés](#)  
[Find flights with low airfares](#)  
[Get alerted when airfares change](#)  
[Find flight deals](#)  
[Find hotel deals](#)  
[Compare travel dates and destinations](#)  
[Find out what's new in Bing](#)

### Limitier vos résultats

[Utiliser la recherche avancée](#)  
[Options de recherche avancée](#)  
[Recherche avancée par mots clés](#)

### Résoudre un problème

[La recherche n'a pas renvoyé les résultats appropriés](#)  
[Peu de résultats, voire aucun](#)  
[Trouver des informations sur des messages d'erreur sur le Web](#)  
[Why some results have been removed](#)

### Référence

[Notations mathématiques à utiliser avec les Réponses mathématiques](#)  
[Codes de pays, de région et de langue](#)

### En savoir plus sur la sécurité et la confidentialité

[À propos de la déclaration de confidentialité et de la stratégie de sécurité de Microsoft](#)  
[Méthodes utilisées par Bing pour proposer des résultats de recherche](#)  
[Influence des annonces sur les résultats de recherche Bing](#)  
[Bloquer des sites Web explicites](#)  
[Signaler un problème concernant un résultat](#)  
[Soumettre un problème de marque](#)

---

OU TROUVER DES ARCHIVES DU WEB ?



Rien n'est exhaustif dans le monde du web, mais le service proposé par l'association The Internet Archive (qui reçoit des donations et soutiens de différents acteurs, dont Alexa) est très impressionnant : on peut ainsi visualiser un site tel qu'il était à différentes dates depuis 1996, et même suivre des liens sur ces archives.

**The way back machine** : [www.archive.org](http://www.archive.org)

Depuis la fin 2003, un service en beta permettait d'aller beaucoup plus loin, en permettant une recherche plein texte, par date, sur plus de 11 milliards de pages archivées. Différentes fonctionnalités étaient accessibles à partir des résultats des sites répondant le mieux à la recherche : graphique permettant de voir la fréquence d'apparition du mot-clé sur la période, thèmes traités par le site, concepts proches, etc. : <http://recall.archive.org>

### 3.5 - EVALUATION DES RESSOURCES INTERNET

#### QUELQUES QUESTIONS CLES A SE POSER

Vérifier ses sources

Web bibliographie :

Information Quality Resources on the Internet, [Marcus P. Zillman](#), Published on December 2, 2011 sur LLRX <http://www.llrx.com/features/informationqualityresources.htm> (Legal Librarians Resources eXchange)

#### COMMENT EVALUER UN SITE WEB ?

L'évaluation de l'information sur Internet devient un enjeu important pour les professionnels. Il s'agit d'un acte d'expertise pour estimer la qualité des différentes ressources disponibles : le portail, le site web, la page web, l'article sur la page, la base de donnée accessible depuis la page, mais aussi le forum, la liste de discussion, le message posté sur une liste ou un forum, etc.

#### Les critères d'évaluation

Différentes catégories de critères sont à prendre en compte, sachant qu'il convient de croiser une évaluation de la source avec une évaluation du contenu :

- \* **Crédibilité** : Organisation émettrice, type d'émetteur, auteurs des documents, source de financement ou sponsoring, webmaster, cibles et objectifs du site, type d'accès, etc.
- \* **Fraîcheur** : Date de création et de mise à jour
- \* **Exhaustivité et l'exactitude** : Type de document, citations des sources, bibliographie, contextualisation de l'information, qualité de la langue, etc.
- \* **Adéquation** : pertinence et utilité par rapport à la recherche ou à la veille menées.
- \* **Ergonomie** : arborescence, navigation, orientation, frames, etc.
- \* **Design** : présentation visuelle, conception graphique.



## Astuces pour l'évaluation des pages en cours de navigation

- \* *Chercher des informations sur l'éditeur sur le site.* En cas de difficulté, chercher le copyright en bas de page. On peut aussi repérer sur le plan du site la page Contact qui va fournir un email. Voir alors la seconde partie de l'adresse mail (après le @) qui peut renvoyer à un domaine particulier que l'on cherchera alors sur le web.
- \* *Chercher des informations sur la société indiquée.* On utilisera alors des bases de données d'informations sur les sociétés (R5CS, organismes de régulation boursiers).
- \* *Pour rechercher le propriétaire d'un nom de domaine* (noms des responsables techniques et administratifs). Attention, les informations sont loin d'être toujours mises à jour, donc il y a des risques d'erreur, et parfois besoin de recoupements.
  - o Pour les noms de domaine se terminant par un ".fr" on utilisera le moteur proposé par l'AFNIC, centre d'information et de gestion des noms de domaine pour la France (et pour l'île de la Réunion .re) : [www.afnic.fr](http://www.afnic.fr)
  - o Pour les noms de domaine "gTLD" (generic Top Level domains), c'est à dire les .com, .net, .org, et plus récemment les .biz et les .info, c'est plus difficile car les bases de données ne sont plus unifiées (auparavant, la base Whois gérée par l'Inter-nic). On utilisera donc un méta-moteur comme Betterwhois, qui permet d'interroger les bases des "régistrants" (prestataires assurant la gestion administrative et technique du nom de domaine) les plus importants : [www.betterwhois.com](http://www.betterwhois.com).
  - o Pour les autres noms de domaine par pays, on peut passer par un service générique <http://www.generic-nic.net/dyn/whois>, ou bien chercher préalablement l'organisme national pays par pays sur Yahoo : [http://dir.yahoo.com/computers\\_and\\_internet/internet/domain\\_name\\_registration/top\\_level\\_domains\\_\\_tlds\\_/registry\\_operators/International\\_Country\\_Codes/](http://dir.yahoo.com/computers_and_internet/internet/domain_name_registration/top_level_domains__tlds_/registry_operators/International_Country_Codes/)
- \* *Utiliser également le "URL info" de Fagan Finder* : <http://www.faganfinder.com/urlinfo>
- \* *Ne pas oublier non plus de faire des recherches sur le web* en prenant le nom du site comme mot-clé, et avec la fonction link : (recherche par popularité : qui a un lien sur cette page).
- \* *On peut aussi utiliser l'interface de recherche développé par un journaliste Jean-Marc Manack pour se simplifier la vie dans la validation des informations* : Plus de 200 outils classés par rubriques (moteurs de recherche, administratif – URL, dictionnaires, référence, actualités, blogs, etc.) sont disponibles à partir d'un seul formulaire, les résultats apparaissant dans la partie gauche de la page. La différence avec un méta-moteur classique, est que l'on peut mettre soit un mot-clé, soit une url. (utilisable aussi en mode "sidebar" dans le navigateur : <http://manhack.net>

## 3.6 QUELQUES OUTILS PRATIQUES

### COMMENT GERER LES PROBLEMES FREQUENTS AVEC LES OUTILS ?

\* **Erreurs 404, liens non valables** : remonter dans la hiérarchie du site. Si l'adresse de l'host est bonne, revenir à cette adresse et "tatonner" à l'intérieur du site pour retrouver la page cherchée et sa nouvelle URL. On peut aussi utiliser le lien "cached" sur Google ou les archives de Alexa.

\* **Signification des principaux messages d'erreurs** :

Erreur	Message	Signification
400	Bad Request	Erreur dans l'adresse
401	Access Denied	La consultation nécessite un nom d'utilisateur et un mot de passe
403	Forbidden	L'accès est réservé et vous n'avez pas les privilèges correspondants
404	Not found	La page correspondant à cette URL n'a pas été trouvée sur le serveur
500	Internal	Problème de serveur. Contacter l'administrateur du site
503	Read time out	Le temps alloué à la connexion est écoulé

\* **Réponses hors sujet** : reformuler sa question, rajouter des mots clés...

\* **La page proposée ne contient pas votre terme de recherche** .

Il peut y avoir plusieurs explications, mais la plus vraisemblable est que ce mot se trouvait dans la page lorsque celle-ci a été sauvegardée par le robot du moteur. Puis elle a été modifiée et le mot a disparu de la page. Mais par contre il est resté dans l'index de la base de données. Il se peut aussi que votre terme apparaisse dans un formulaire déroulant, ou enfin en méta-données.

Une solution pour être certain d'obtenir des résultats contenant les mots-clés de votre question consiste à utiliser un méta-moteur "off-line" avec la fonction "raffiner" ou "filtrer".

- \* **Non élimination des doublons** : les moteurs utilisent maintenant à peu près tous les techniques de clustering pour la présentation des résultats (une réponse = un site et non une réponse = une page) ou le proposent en option. Mais cela n'empêche pas toujours les doublons.
- \* **Problème d'accès à de l'information très récente** : attention, un moteur peut mettre plusieurs jours ou mêmes semaines avant d'indexer un nouveau site... Voir du côté des serveurs d'actualité, par exemple.

---

## PEUT-ON CIRCULER DE FAÇON ANONYME SUR LE WEB ?

On le sait, la navigation sur le web laisse des traces (voir notamment à ce sujet le site de la CNIL [www.cnil.fr](http://www.cnil.fr)). Il existe néanmoins des services permettant de masquer les adresse IP d'origine et d'empêcher les cookies et autres techniques de marquage de fonctionner, c'est à dire de garantir une meilleure confidentialité de surf sur internet

Anonymiser <http://www.anonymizer.com/> (payant)

Voir le TOR Project Anonymity online <https://www.torproject.org/>

Voir enfin le portail Stay Invisible qui propose définitions, actualités, tests, un forum de discussion sur le sujet ainsi qu'une liste d'outils : <http://www.stayinvisible.com>

---

## PEUT-ON EFFECTUER DES TRADUCTIONS DE TEXTES SUR LE WEB ?

Des outils gratuits sont disponibles en ligne pour traduire des textes, voire des pages web. Les résultats sont certes souvent discutables, mais pour une première approche, ces technologies peuvent être d'une aide réelle à la recherche.

**Sur Voila** (technologie Systran) <http://tr.voila.fr>

**Sur Google** (technologie Systran) <http://translate.google.com/>

**Sur Alta Vista** (technologie Systran) <http://babelfish.altavista.com/> /  
<http://fr.babelfish.yahoo.com/>

**Sur Reverso** (technologie Reverso) <http://www.reverso.net>

## 4 OPTIMISER SA RECHERCHE

### 4.1 - ORGANISER ET GERER LES FAVORIS (OU MARQUE-PAGES)

Encyclopédie de signets : <http://signets.bnf.fr/>



[http://wiki-urfist.unice.fr/wiki\\_urfist/index.php/G%C3%A9rer\\_ses\\_favoris\\_en\\_ligne](http://wiki-urfist.unice.fr/wiki_urfist/index.php/G%C3%A9rer_ses_favoris_en_ligne)



article discussion modifier historique

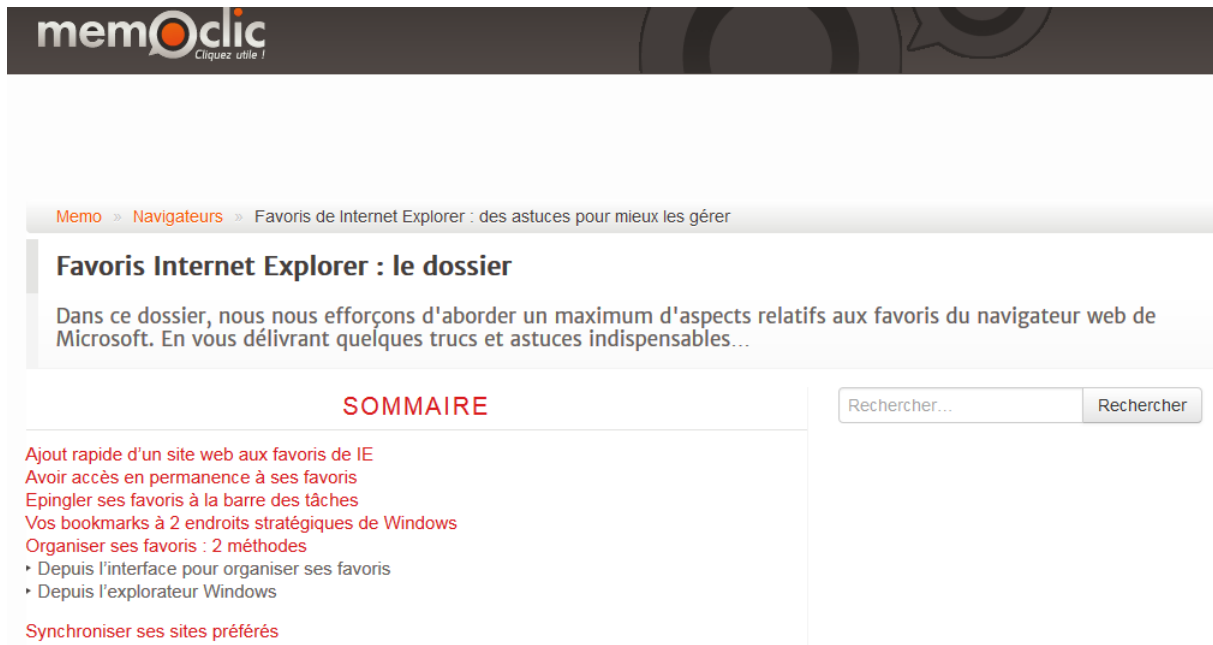
## Gérer ses favoris en ligne

pearltree

Sommaire [masquer]

- 1 Description
  - 1.1 fiches
  - 1.2 objectifs
  - 1.3 contenu
- 2 Le stage

<http://www.memoclic.com/814-navigateurs/18588-internet-explorer-favoris.html>



memoclic Cliquez utile !

Memo > Navigateurs > Favoris de Internet Explorer : des astuces pour mieux les gérer

## Favoris Internet Explorer : le dossier

Dans ce dossier, nous nous efforçons d'aborder un maximum d'aspects relatifs aux favoris du navigateur web de Microsoft. En vous délivrant quelques trucs et astuces indispensables...

**SOMMAIRE**

- Ajout rapide d'un site web aux favoris de IE
- Avoir accès en permanence à ses favoris
- Epingler ses favoris à la barre des tâches
- Vos bookmarks à 2 endroits stratégiques de Windows
- Organiser ses favoris : 2 méthodes
  - Depuis l'interface pour organiser ses favoris
  - Depuis l'explorateur Windows
- Synchroniser ses sites préférés

Rechercher... Rechercher

## 4.2 - UTILISER LES FILS RSS

Repérer les fils

Logo "feedburner" (extension de google dédiée au RSS)

Les lecteurs de fils RSS : veille assurée sur <http://www.scoop.it/t/rss> <https://twitter.com/rsscircus>

<http://blogs.crdp-limousin.fr/stage-veille/2011/05/17/didacticiels-diigo/>

Didacticiels DIIGO, Didier Pouzaud, sur le BiblioLab de la BNF

<http://bibliolab.fr/cms/content/comment-fonctionne-diigo>

Le Bibliolab est une plateforme de formation, d'expérimentation et d'information autour des TIC et du numérique en bibliothèque.

### *The Evolution of Diigo*



To learn more about Diigo V5.0, please check out this overview video: <http://www.vimeo.com/12687333>  
and click to check out ["What is New in Diigo V5.0" >>](#)

<http://cursus.edu/dossiers-articles/articles/9750/diigo-mode-emploi-pour-debutants/>

<http://search.twitter.com> (accessible sans compte twitter)

<https://twitter.com/search-advanced>

Usage d'un compte twitter

Lecture de Time Line

Reprise de mini messages

Exemple de LiveTweet : IFLA 2014 à Lyon. Le hashtag était #wlic2014

Application STORIFY

Exemple avec l'IFLA = <https://storify.com/WLIC2014Lyon/let-s-wlic2014#publicize>



#### 4.4 - APRES LA FORMATION : S'AUTOFORMER ET SUIVRE L'ACTUALITE DES OUTILS DE RECHERCHE

##### VEILLE SUR L'ACTUALITE DES OUTILS DE RECHERCHE D'INFORMATION

<http://www.abondance.com/> L'actualité des moteurs et du référencement... Olivier Andrieu

<http://www.les-infostrateges.com/> publie en permanence des actualités, des articles et des dossiers de fond qui rendent compte de l'expertise des auteurs en stratégies informationnelles (spécialisés en droit de l'information et en "e-reputation")

<http://saprستي-docinsa.insa-lyon.fr/guides-similaires> (Guide Sapristi, Insa de Lyon)

Les guides francophones

- L'URFIST de Lyon propose des documents très intéressants dont :
  - Des informations variées sur Internet
  - Un guide consacré aux sociétés et au marché des bases de données sur le web : SIN-BAD
  - Des documents pédagogiques relatifs à l'actualité du web
- L'URFIST de Rennes propose de nombreuses pages dont des supports de formation indispensables à la recherche sur Internet. : <http://www.sites.univ-rennes2.fr/urfist/ressources>
- L'URFIST de Paris propose CERISE : Conseil aux Etudiants pour une Recherche d'Information Spécialisée et Efficace. <http://giry.enc.sorbonne.fr/cerise/>
- ABCdoc est un guide méthodologique de recherche et de traitement de l'information scientifique et technique produit par la SUP (Structure Universitaire de Pédagogie) de l'Université Paul Sabatier (Toulouse 3)
- Form@doct est un guide destiné aux doctorants complément indispensable à Sapristi ! mais pouvant être utilisé par des d'étudiants d'autres niveaux. Il est produit depuis 2010 par l'UEB (Université Européenne de Bretagne)
- Méthodoc est un guide de méthodologie documentaire et de sciences de l'information. Il est destiné aux étudiants du supérieur. Il est co-produit par le SCD de l'Université Rennes 2 et par l'URFIST de Rennes.
- On peut aussi rajouter à cette liste non-exhaustive les pages signets de la BNF, de la BPI et du CERIMES <http://www.signets-universites.fr/>





GFII <http://www.gfii.fr/fr/>

Internet Actu et FING <http://www.internetactu.net/>



Le Journal du Net : e-Business, Informatique, Economie et ... [www.journaldunet.com](http://www.journaldunet.com)

Analyses, tendances, interviews : tous les jours, le Journal du Net vous propose le meilleur de l'actualité Internet et e-business.

Affordance.info : Notes, liens et réflexions en rapport avec les sciences de l'information et la gestion des connaissances, par Olivier Ertzscheid.