

Rechercher sur le WEB

1 Table des matières

1	Table des matières.....	2
2	Distinguer les outils.....	3
2.1	Annuaire, moteurs et métamoteurs	3
2.1.1	Les annuaires	4
2.1.2	Les moteurs.....	4
2.1.3	Les métamoteurs	5
2.1.3.1	Métamoteur(s) cartographique(s).....	5
2.1.3.2	Métamoteurs textuels.....	6
2.2	Annuaire et moteurs pour la Belgique	7
3	Choisir ses mot-clés	8
3.1	Trouver de "bons" mots-clés	8
3.1.1	Des mots suffisamment précis sur les moteurs	8
3.1.2	Des mots suffisamment larges sur les annuaires.....	8
3.1.3	La langue	8
3.1.4	Le type de mots.....	8
3.1.5	Le nombre de mots.....	8
3.1.6	L'ordre des mots	8
3.2	Les saisir correctement	8
3.2.1	Aide en ligne des outils	8
3.2.2	Fiches du site Abondance	9
3.3	Réajuster sa requête en fonction du « bruit » et du « silence »	9
3.3.1	Restreindre la question	9
3.3.2	Élargir la question.....	9
4	Google.....	10
4.1	Options linguistiques	10
4.1.1	Accents (non / oui).....	10
4.1.2	Majuscules (non).....	10
4.1.3	Ligature (oui).....	10
4.1.4	Correction orthographique (oui).....	10
4.1.5	Ordre des mots (oui)	11
4.1.6	Troncature / Lemmatisation (non).....	11
4.1.7	Analogie (oui).....	11
4.2	Préfixes de recherche.....	11
4.3	Préfixes de fonction	12
4.4	Préfixe de recherche linguistique.....	13
4.5	Tableau d'extensions	13

2 Distinguer les outils

2.1 Annuaire, moteurs et métamoteurs

Beaucoup d'internautes ont tendance à utiliser de la même manière

- les annuaires (ou répertoires) comme Yahoo!, Nomade...
- et les moteurs de recherche, comme Google, Altavista...

Ces deux types d'outils sont pourtant de nature différente.

Il est essentiel de bien comprendre leurs spécificités pour choisir le plus adapté à votre requête.

ANNUAIRES	MOTEURS
Indexation de sites - par des documentalistes	Indexation de mots - par des robots
<i>Recherche</i> - sur des sites - et des catégories	<i>Recherche</i> - en texte intégral - sur des pages web
<i>Avantages</i> - choix des informations - classement raisonné par catégories et sous-catégories	<i>Avantages</i> - plus d'exhaustivité, - mise à jour plus rapide
<i>Inconvénients</i> - moins d'exhaustivité, - mise à jour moins rapide	<i>Inconvénients</i> - capture de pages web - sans classement raisonné
<i>À retenir</i> L'exploration des catégories s'avère souvent plus fructueuse que celle des sites.	<i>À retenir</i> La recherche par mots clés donne de meilleurs résultats sur les moteurs

MOTEURS	MÉTAMOTEURS
Répétition de la requête sur chaque moteur	Une seule requête simultanée sur plusieurs moteurs et annuaires
<i>Avantage</i> une syntaxe spécifique (résultats plus précis)	<i>Avantage</i> gain de temps
<i>Inconvénient</i> temps de réponse plus long	<i>Inconvénient</i> pas de syntaxe commune (résultats moins précis)

L'Internet (et le WEB) évoluent rapidement. La simple existence des adresses ci-dessous est régulièrement vérifiée (dernière

vérification le 27/12/2008), sans garantie du jour. Si vous constatez un problème (erreur 4xx), contactez-moi (<http://pascal.cambier.eu>). Si c'est une erreur 5xx, patientez.

Des adresses peuvent être détournées de leur objectif initial. Si c'est le cas, contactez-moi (<http://pascal.cambier.eu>).

2.1.1 Les annuaires

- explorer un sujet
- trouver des sites ressources dans un domaine
- trouver des sites similaires

Mondiaux	Francophones
http://www.about.com/ http://www.alex.com/ http://galaxy.einet.net/ http://www-english.lycos.com/ http://www.looksmart.com/ http://www.lii.org/ http://nbc.msnbc.com/nbc.asp http://dmoz.org/ http://www.webbrain.com/html/default_win.html http://dir.yahoo.com/	http://www.ahalia.com http://www.recherche.aol.fr/ http://www.boussole-fr.com/ http://www.francite.com/ http://www.indexa.fr/ http://www.meilleurduweb.com/ http://www.mirti.com/ http://www.ousurfer.com/ http://www.souany.com/ http://guide.voila.fr/

2.1.2 Les moteurs

- trouver une information précise sur une personne, un organisme, un produit...
- effectuer des recherches complexes
- repérer des sites récents ou non indexés par les annuaires

Mondiaux	Francophones
http://a9.com http://www.accoona.com/ http://www.antidot.net/ http://www.ask.com/ http://www.clush.com http://www.dipsie.com/ http://www.exalead.com/ http://www.eurekster.com/ http://www.gigablast.com/ http://www.google.com/ http://www.grokker.com http://www.infotiger.com/	moteur européen http://fr.abacho.com/ http://fr.altavista.com/ http://www.aol.fr/ http://www.fr.deepindex.com/ http://fr.dir.com/ http://www.exalead.fr/ http://www.google.fr/ http://www.laposte.net/portail/Rechercher/ http://www.mirago.fr/ http://www.misterbot.com/ http://search.msn.fr/ http://www.mozbot.fr/

Rechercher sur le WEB

http://www.kazzoom.com http://www.lexxe.com/ http://mooter.com/moot http://search.msn.com/ http://nokodo.com/ http://www.redzee.com/ http://www.teoma.com/ http://search.thunderstone.com/taxis/websearch/ http://www.ujiko.com/ http://www.wisenut.com http://search.yahoo.com	http://www.numika.com/ http://www.voila.fr/ http://www.perso-search.com/ http://www.seekport.fr/ http://fr.search.yahoo.com
---	---

2.1.3 Les métamoteurs

- débroussailler rapidement un sujet
- repérer les outils qui répondent le mieux à un sujet

Francophones	Mondiaux
<i>Présentation cartographique</i>	
http://www.kartoo.com/	
<i>Présentation textuelle</i>	<i>Présentation textuelle</i>
http://www.eo.st/ http://fr.ez2find.com http://www.ixquick.com/ http://www.seek.fr http://www.netoo.fr/	http://www.dogpile.com/index.gsp http://ez2find.com http://www.ixquick.com/ http://www.metacrawler.com/index.html http://www.oneseek.com/ http://www.qbsearch.com/ http://www.queryserver.com/web.htm http://www.searchwiz.com/ http://vivisimo.com/

2.1.3.1 Métamoteur(s) cartographique(s)

KartOO

<http://www.kartoo.com>

KartOO effectue les recherches à partir d'une quinzaine de moteurs : il analyse votre requête dès son lancement, interroge les moteurs activés les plus pertinents.

- Les résultats obtenus apparaissent dans une carte thématique représentant les sites sous forme de boules plus ou moins grosses selon le degré de pertinence trouvé par l'outil. Lorsque le pointeur survole une boule, la description du site s'affiche. En cliquant sur la boule, vous accédez au site.

- Au survol d'un thème, apparaissent deux boutons + et -. Le + ajoute le thème à la recherche. Le - élimine les sites concernant ce thème (la traduction des opérateurs booléens est automatique). Cette fonction est très pratique pour affiner sa requête : la restreindre ou l'élargir.

2.1.3.2 Métamoteurs textuels

EO

<http://www.eo.st/>

EO : un métamoteur prometteur

"Avez-vous essayé le métamoteur en ligne EO (pour "Eh Oh !") ? Il s'agit d'un métamoteur web très sobre, mais très efficace, interrogeant Google, AllTheWeb, Lycos, MSN, Teoma entre autres, proposant un look "Mac" très sympa, et affichant des résultats très pertinents. Une fonction "Preview" permet notamment de visualiser la page proposée dans une fenêtre insérée dans la page de résultats. Bref, un nouvel outil de recherche à essayer..."

Source : Abondance <http://actu.abondance.com/>

EO, le métamoteur Feng Shui du Web

"Dans la famille des outils de recherche, se côtoient des moteurs, des annuaires et des métamoteurs. Ces derniers permettent de lancer une requête simultanément sur plusieurs outils de recherche et dédoublonnent ensuite les résultats ; ils sont souvent utilisés lors de la phase initiale de débroussaillage d'un sujet."

Source : Archimag, juillet-août 2004

¹http://www.archimag.com/articles/176/176internet_internet.html

Ez2Find.com

En français : <http://fr.ez2find.com>

En anglais : <http://ez2find.com>

- Ce métamoteur recueille des résultats de AlltheWeb, Teoma, Google, Yahoo !, AltaVista, Wisenut, ODP, et MSN.
- Traitement des résultats
 - Ez2Find.com analyse les résultats, enlève les reproductions, inclut des liens aux catégories appropriées de l'annuaire (OPD) et regroupe les résultats.
 - Des étoiles sont attribuées aux sites trouvés dans les moteurs, un site avec 5 étoiles arrive en tête car il a été trouvé dans 5 moteurs de recherche différents.
 - La pertinence est calculée par un algorithme qui compte le nombre de fois que le site a été trouvé dans les moteurs de recherche et son positionnement dans chacun d'eux.

¹ Plus disponible le 10/10/2006

Rechercher sur le WEB

- Ez2Find.com regroupe automatiquement et cloisonne les résultats de la recherche dans les catégories choisies parmi les mots et les expressions contenues, afin d'affiner les résultats.
- Autres caractéristiques
 - Ez2Find peut trouver l'information également dans une multitude de bases de données différentes, soit par langues, soit par spécialités. Il recherche dans les bases de données en temps réel (web invisible²).
 - Une interface appropriés propose des informations relative au pays, comme des informations locales, un annuaire de téléphone, des radios, une recherche spécifique à la langue... Il est également possible de choisir un autre pays ou de paramétrer ses préférences.

Seek.fr

<http://www.seek.fr/>

SEEK.fr a été principalement conçu pour le marché francophone.

Le site est séparé en deux parties : un métamoteur de recherche, un annuaire "classique" (les deux sont pour le moment indépendants).

En développement actuellement : Service de News, Une Toolbar, Optimisation du temps de recherche, ..

2.2 Annuaires et moteurs pour la Belgique

	http://www.advalvas.be/fr/index.php
	http://be-fr.altavista.com/
	http://www.google.be/
	http://www.sharelook.be/
	http://www.webwatch.be/
	http://fr.dir.yahoo.com/Exploration_geographique/Pays/Belgique/

² Le Web invisible correspond à l'ensemble des documents (textes, vidéos, images...) qui n'est pas indexé par les outils de recherche (moteurs, annuaires...). Ces données peuvent être de nature non indexable (animations), non référencées (volontairement ou non), ou dynamiques.

3 Choisir ses mot-clés

3.1 Trouver de "bons" mots-clés

3.1.1 Des mots suffisamment précis sur les moteurs

- éviter les termes trop généraux qui risquent de générer trop de "bruit" dans les réponses,
- choisir de préférence des termes spécifiques, quitte à élargir par la suite votre requête.

3.1.2 Des mots suffisamment larges sur les annuaires

- pour trouver facilement les catégories et pour identifier les sites fédérateurs traitant du sujet,
- quitte à préciser ensuite avec des termes plus précis.

3.1.3 La langue

- éviter les mots d'une langue autre que celle du moteur.

3.1.4 Le type de mots

- choisir de préférence des noms,
- utiliser les autres mots (verbes, adjectifs, adverbes, pronoms) plutôt dans les expressions entre guillemets.

3.1.5 Le nombre de mots

- plus il y a de mots, plus la question se restreint,
- un seul mot peut suffire,
- essayer de ne pas dépasser 3 mots au départ.

3.1.6 L'ordre des mots

joue un rôle sur certains moteurs :

- quels mots-clés sont prioritaires ?
- commencer par les mots les plus importants

(voir aussi la syntaxe propre à chaque moteur).

3.2 Les saisir correctement

3.2.1 Aide en ligne des outils

Se référer à l'aide en ligne de l'outil utilisé.

Faut-il saisir la requête :

- en minuscules ou en majuscules ?
- avec des lettres accentuées ou non accentuées ?
- au singulier ou au pluriel ? (utiliser la troncature)

3.2.2 Fiches du site Abondance

Les fiches descriptives des principaux annuaires et moteurs mondiaux et francophones présentent des tableaux très bien faits sur la syntaxe de saisie :

- minuscules/majuscules, lettres accentuées,
- ordre des mots, opérateurs booléens, expressions, troncature,
- recherches spécifiques (linguistique, titre du document, domaine, nom du serveur, intitulé d'une url, applet Java, nom des images, adresses des liens, intitulé des liens hypertextes, texte visible de la page).

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

3.3 Réajuster sa requête en fonction du « bruit » et du « silence »

3.3.1 Restreindre la question

- affiner avec des termes plus pointus,
- avec un moins (-) devant le(s) mot(s) parasite(s),
- préciser éventuellement le pays et/ou la langue.

3.3.2 Élargir la question

- élargir avec des termes plus larges
- utiliser des synonymes (avec OU)
- voire des antonymes (avec OU).

4 Google



<http://www.google.fr/support?hl=fr>

Note: Le nombre maximum de mots dans une requête est de trente-deux. Les préfixes de recherche, chaque partie d'un mot séparé par un trait d'union et chaque mot entre guillemets sont comptés. Certains mots vides (articles, prépositions,...) et les signes de ponctuation sont ignorés.

4.1 Options linguistiques

4.1.1 Accents (non / oui)

Pour rechercher un mot accentué, mettez le signe + accolé au mot comme ceci : +fête, sans espace entre + et fête. Si vous recherchez un mot ayant des paronymes³ distincts par des accents, éliminez l'indésirable.

Ex.: cotes -côtes

4.1.2 Majuscules (non)

Google ne reconnaît pas les majuscules même avec le signe + utilisé pour le mot exact.

4.1.3 Ligature (oui)

La ligature est la réunion de deux lettres telles que oe en œ.

boeufs = bœufs

4.1.4 Correction orthographique (oui)

Le correcteur orthographique est établi sur des statistiques et non sur un dictionnaire. Seules les versions en anglais disposent d'un dictionnaire dans la barre d'état. Préférez un dictionnaire établi pour vérifier une orthographe.

Ex . pacal -> Essayez avec cette orthographe :pascal

³ subst. masc. et adj. (TLFI) Linguist. (Mot) qui présente avec un autre mot une certaine analogie phonétique, mais sans avoir le même sens

4.1.5 Ordre des mots (oui)

Google tient compte de l'ordre des mots pour classer les pages. Utilisez l'ordre naturel des mots. « Honnête homme » est différent de « homme honnête ».

4.1.6 Troncature / Lemmatisation (non)

Pour garantir des résultats aussi précis que possible, Google n'applique pas de « lemmatisation » (réduction des mots au masculin et/ou au singulier, à l'infinitif, etc.) et ne supporte pas les recherches à base de caractères joker/wildcard. Autrement dit, Google utilise les mots exactement tels que vous les entrez dans le champ de recherche. Exemple : Si vous entrez le terme « cheval » ou « cheva* », Google ne fait pas porter la recherche sur « chevaux », « chevaline » ou « chevalerie », mais uniquement sur le terme « cheval » ou « cheva* » (soit la chaîne de caractères « cheva » suivie d'un astérisque). Pour plus de sûreté, entrez les formes qui vous intéressent, par exemple : « cheval » et « chevalerie ».

4.1.7 Analogie (oui)

~bateau = bateaux, croisière

Le tilde [~] est un opérateur analogique et synonymique. Il est placé immédiatement avant le mot, sans espace. Il sert à élargir une recherche. Cet opérateur ne donne pas tout le lexique associé au mot. Préférez un dictionnaire établi comme le <http://elsap1.unicaen.fr/cgi-bin/cherches.cgi>

4.2 Préfixes de recherche

Les préfixes de recherche permettent de resserrer la recherche. Voici les plus usuels.

Préfixes	Exemples	Commentaires
site:	Herman de Croo <u>site:be</u>	Cette requête trouvera toutes les pages belges dans lesquelles se trouvent les mots Herman Croo.
	"Herman De Croo" <u>site:lachambre.be</u>	Cette requête trouvera toutes les pages du site lachambre.be dans lesquelles se trouve l'expression « Herman De Croo »
allintitle: intitle:	<u>allintitle:moteurs</u> électriques	Les deux mots sont dans le titre du fichier. Utilisez allintitle: au début de la requête pour inclure tous les mots dans le titre

Préfixes	Exemples	Commentaires
	"moteurs électriques" <u>intitle:moteurs</u>	La page web contient l'expression « moteurs électriques » et son titre moteurs. Utilisez intitle: pour inclure un mot de la requête dans le titre
allinurl: inurl:	<u>allinurl:moteur</u> recherche	L'URL de la page devra contenir les deux mots. Utilisez allinurl: au début de la requête pour inclure tous les mots dans l'adresse url.
	moteur <u>inurl:recherche</u>	La page devra contenir le mot moteur et son URL le mot recherche. Utilisez inurl: pour inclure un mot de la requête dans le l'URL.
filetype: ext:	"moteur électrique" <u>filetype:pdf</u>	La page devra contenir l'expression « moteur électrique » et elle devra être un document PDF Voir 4.5 ci-dessous

4.3 Préfixes de fonction

Préfixes	Exemples	Commentaires
cache:	<u>cache:www.lachambre.be</u>	La page en cache est celle sauvegardée dans le catalogue. Elle est active. Ce n'est pas une copie d'écran. Elle est différente de la page originale quand celle-ci a été modifiée après l'indexation. Elle nous montre la date d'indexation et la localisation des termes de recherche dans la page. Elle est utile lorsque la page originale n'est pas accessible, lorsqu'elle a été modifiée ou retirée du web ou pour accéder à une page devenue privée.
related:	<u>related:lachambre.be</u>	La page de référence ne doit avoir qu'un sujet de préférence court pour que ce préfixe puisse donner un bon résultat.

Préfixes	Exemples	Commentaires
link:	<u>link</u> :www.lachambre.be	Utilisé pour connaître l'environnement d'une page liée. Qui le lie et dans quel contexte la page est liée. Le préfixe link: permet de trouver aussi des listes de liens. Il doit être utilisé seul sinon il est désactivé. Google trouvera toutes les pages qui citent l'adresse qu'elle soit ou non hyperliée
info: id:	<u>info</u> :www.lachambre.be	Ce préfixe est pratique pour accéder rapidement à de l'information complémentaire sans avoir à se souvenir des autres préfixes.

4.4 Préfixe de recherche linguistique

Préfixes	Exemples	Commentaires
define:	<u>define</u> : lemmatisation	Ce préfixe de recherche propose des Expressions contenant le terme de la requête. Ces expressions ne sont pas des suggestions de recherche mais des suggestions de définitions. Assurez-vous qu'elle provient d'une source fiable avant d'utiliser la définition.

4.5 Tableau d'extensions

Classement par extensions		Classement par applications	
Ext	Applications	Applications	Ext
ans	Text	Active Server Pages (dynamique)	asp
asp	Active Server Pages (dynamique)	Adobe Portable Document Format	pdf
cfml	ColdFusion Markup Language (dynamique)	Adobe PostScript	ps
cgi	Common Gateway Interface	AutoCAD	dwg
css	Cascading Style Sheet (dynamique)	Cascading Style Sheet (dynamique)	css
doc	Microsoft Word	ColdFusion Markup Language (dynamique)	cfml
dwg	AutoCAD	Common Gateway Interface	cgi
hdml	Handheld Device Markup Language	Extensible markup language	xml

Rechercher sur le WEB

Classement par extensions		Classement par applications	
Ext	Applications	Applications	Ext
jsp	Java Script Pages	Handheld Device Markup Language	html
lwp	Lotus 1-2-3	Java Script Pages	jsp
mw	MacWrite	Lotus 1-2-3	wku
pdf	Adobe Portable Document Format	Lotus 1-2-3	lwp
php	PHP Hypertext Processor (dynamique)	Lotus 1-2-3	wks
php3	PHP Hypertext Processor (dynamique)	Lotus 1-2-3	wki
ppt	Microsoft PowerPoint	Lotus 1-2-3	wk5
ps	Adobe PostScript	Lotus 1-2-3	wk3
rtf	Rich Text Format	Lotus 1-2-3	wk4
shtml	Server-side include Hypertext Markup Language file (dynamique)	Lotus 1-2-3	wk2
ssi	Server Side Includes (dynamique)	Lotus 1-2-3	wk1
swf	Macromedia Shockwave Flash (dynamique)	Macromedia Shockwave Flash (dynamique)	swf
txt	Text	MacWrite	mw
wdb	Microsoft Works	Microsoft Excel	xls
wk1	Lotus 1-2-3	Microsoft PowerPoint	ppt
wk2	Lotus 1-2-3	Microsoft Word	doc
wk3	Lotus 1-2-3	Microsoft Works	wps
wk4	Lotus 1-2-3	Microsoft Works	wdb
wk5	Lotus 1-2-3	Microsoft Works	wks
wki	Lotus 1-2-3	Microsoft Write	wri
wks	Lotus 1-2-3	PHP Hypertext Processor (dynamique)	php
wks	Microsoft Works	PHP Hypertext Processor (dynamique)	php3
wku	Lotus 1-2-3	Rich Text Format	rtf
wpd	WordPerfect	Server Side Includes (dynamique)	ssi
wps	Microsoft Works	Server-side include Hypertext Markup Language file (dynamique)	shtml
wri	Microsoft Write	Text	txt
xls	Microsoft Excel	Text	ans
xml	Extensible markup language	WordPerfect	wpd
gif	Graphic Interchange Format	Graphic Interchange Format	gif
jpg ;	Joint Photographic Group	Joint Photographic Group	jpg ;

Rechercher sur le WEB

Classement par extensions		Classement par applications	
Ext	Applications	Applications	Ext
jpeg			jpeg