
SOMMAIRE

SOURCES DE LA RECHERCHE	3
PRINCIPALES CARACTERISTIQUES DES SERVICES EN LIGNE ETUDIES	3
ALTAVISTA	3
ANTINEA	3
ATOMZ	4
BRAVENET	4
ECILA OFFSHORE	5
EXTENSE	5
FREEFIND	5
INDEXMYSITE	6
LE CGI	6
MYCOMPUTER.COM	7
OVH	7
PICOSEARCH	7
SEARCHBUTTON	8
VOILA	8
WHATUSEEK	9
X-RECHERCHE	9
SYNTHESE COMPARATIVE DES SERVICES EN LIGNE DE MOTEURS DE RECHERCHE INTERNES	11
SOLUTION PROPOSEE	11
BIBLIOGRAPHIE	11
ANNEXE 1 : TABLEAU COMPARATIF DES SERVICES EN LIGNE DE MOTEURS DE RECHERCHE INTERNES	12

INTRODUCTION

Le présent document correspond à une étude comparative des différentes solutions de moteurs de recherche internes aux sites Web, existant sur le marché.

Sources de la recherche

La recherche s'est faite directement au travers des principaux moteurs de recherche, au travers des outils de recherche proposant des services de moteurs internes, ou au travers de sites "spécialisés" tels que : abondance.com, "Le coin des Experts" (<http://perso.wanadoo.fr/coin.des.experts>) ou monmoteur.net.

Deux familles de solutions sont apparues :

- Les services de moteurs internes en ligne
- Les logiciels ou scripts à installer directement sur les serveurs Web

La première solution est la plus simple à tester et à mettre en œuvre. La deuxième solution a été étudiée avec les scripts proposés sur monmoteur.net. Mais ces derniers sont difficilement testables sans une étude approfondie nécessitant les installations de logiciels supplémentaires (Perl...). De plus les solutions trouvées semblent limitées et peu professionnelles.

Les principaux services en ligne testés sont les suivants : Altavista (www.altavista.fr), Antinea (www.antinea.org), Atomz (www.atomz.com), Bravenet (www.bravenet.com), Ecila Offshore (<http://offshore.ecila.com>), eXtense (www.extense.com), Freefind (www.freefind.com), IndexmySite (www.indexmysite4free.com), Le CGI (www.lecgi.net), MyComputer (www.mycomputer.com), Ovh (www.ovh.net), Pico Search (www.picosearch.com), SearchButton (www.searchbutton.com), Voila (www.voila.fr), whatUseek (www.whatuseek.com), X-recherche (www.x-recherche.com).

Principales caractéristiques des services en ligne étudiés

Altavista

Altavista ne semble pas proposer de solution réellement professionnelle, en dehors de la solution gratuite consistant à placer un champ de recherche 'altavista', en donnant comme consigne de taper 'host:' et le nom de domaine du site sur lequel s'effectue la recherche, avant de taper les mots clés.

Antinea

Il s'agit d'un moteur francophone hébergé par une université (Angers). Si la qualité du service est très intéressante et les améliorations à venir, prometteuses, on restera prudent quant à la pérennité du service (cf l'histoire de l'hébergeur "Mygale").

Points positifs :

- facilité de mise en œuvre
- gratuit
- francophone
- pas de bannière de publicité (juste un logo en bas de page)
- pas de limite du nombre de pages référencées
- possibilité de rajouter plusieurs sites et/ou plusieurs pages
- possibilité d'exclure des répertoires et des pages de l'indexation (robots.txt)
- réindexation du site automatique 2 fois par semaine (le mercredi soir et le dimanche soir)
- conditions booléennes basiques pour les recherches ("Tous les mots", "Chaque mot")
- gestion des caractères accentués
- affichage du nombre d'occurrences trouvées pour chaque mot demandé
- affichage des lignes où les mots ont été trouvés, ceux-ci étant mis en gras
- statistiques sur les recherches des visiteurs (sur le site Antinea)

Points négatifs :

- pas de possibilité de gérer les sites avec mot de passe pas de possibilité de recherche par secteurs (groupe de pages)
- pas de possibilité de recherche des expressions personnalisation (limitée) de la page de résultats

- pas d'indication sur le manière dont le tri est effectué (à part le nb de mots trouvés)

Atomz

Ce moteur assez riche possède un grand nombre d'options et permet la personnalisation complète de la page de résultats et cela sans publicité. Son seul défaut est la limitation à 500 pages dans sa version gratuite, car les sites qui ont le plus besoin d'un moteur peuvent dépasser largement ce chiffre.

Points positifs :

- facilité de mise en oeuvre
- pas de bannière de publicité
- gratuit
- possibilité de rajouter plusieurs sites et/ou plusieurs pages
- possibilité d'exclure des répertoires et des pages de l'indexation (robots.txt)
- possibilité de gérer les sites avec mot de passe
- possibilité de déclarer des synonymes pour faciliter la recherche
- possibilité de recherche par secteurs (groupe de pages)
- référencement des fichiers TEXTE, PDF, FLASH et MP3
- réindexation automatique paramétrable et/ou manuelle
- conditions booléennes basiques pour les recherches ("Tous les mots", "Chaque mot")
- options de recherche étendues
- possibilité de recherche des expressions
- gestion des caractères accentués
- formulaire de recherche avancé très riche
- personnalisation du formulaire accessible à tous grâce aux explications détaillées sur chaque zone (en anglais)
- personnalisation complète de la page de résultats grâce à des "templates" (paramétrage en HTML)
- accès à des "templates" déjà existants
- contrôle du tri de présentation des résultats (très précis)
- affichage des lignes où les mots ont été trouvés, ceux-ci étant mis en gras
- affichage du texte optionnel (possibilité de ne laisser que les titres)
- statistiques sur les recherches effectuées lors des 2 derniers mois (envoi automatique d'un email et consultation sur leur site)

Points négatifs :

- anglophone (anglais nécessaire pour le paramétrage)
- limitation du nombre de pages référencées à 500 pages dans la version gratuite

Bravenet

Ce site qui est avant tout généraliste, offre énormément de services gratuits pour les webmasters et en contrepartie affiche systématiquement sur toutes ses pages au minimum un bandeau de publicité. Pour la partie moteur de recherche interne, on est loin de la qualité des services spécialisés.

Points positifs :

- Gratuit

Points négatifs :

- mise en œuvre pénible (saisie fastidieuse de toutes les URL, une par une !)
- anglophone
- bannière de publicité
- limitation à 60 pages de référencées
- pas de possibilité de rajouter plusieurs sites et/ou plusieurs pages
- pas de possibilité d'exclure des répertoires et des pages de l'indexation

- pas de possibilité de gérer les sites avec mot de passe
- pas de possibilité de recherche par secteurs (groupe de pages)
- réindexation manuelle page par page (pas de planification)
- pas de conditions booléennes pour les recherches (ET, OU, NOT)
- pas de possibilité de recherche des expressions
- pas de gestion des caractères accentués
- personnalisation (du formulaire et de la page de résultats) réduite au minimum
- pas d'affichage du nombre de résultats trouvés
- la liste est affichée sur une seule page quelque soit le nombre de liens trouvés

Ecila Offshore

Il s'agit du moteur de recherche interne proposé par Ecila et reposant globalement sur la même technologie que ce dernier. Les résultats des recherches sont pertinents. La mise en œuvre de la solution est simple et efficace (rapide). Cependant, cette offre (gratuite) est actuellement suspendue, pour des raisons inconnues.

eXtense

Ce moteur, développé par l'entreprise Echo interactive (E-stats, Voila), existe en deux versions : l'une gratuite et limitée, l'autre payante et professionnelle. La version gratuite est limitée à 50 pages.

Points positifs :

- facilité de mise en œuvre
- indexation très rapide
- nombreuses options pour les recherches comme les expressions booléennes
- gestion des caractères accentués
- possibilité de personnalisation de la page de résultats, uniquement dans la version payante
- possibilité de recherche uniquement dans des sous-ensembles du site, uniquement dans la version payante
- solution francophone
- indication sur l'ordre du tri des résultats

Points négatifs :

- limite du nombre de pages référencées (50) dans la version gratuite
- pas de possibilité de rajouter plusieurs sites et/ou plusieurs pages (toutes les pages du site doivent être sous le même nom de domaine)

Freefind

Ce service est de qualité par sa gestion simple et efficace, mais sa page de résultats est trop chargée (entête et bas de page). A noter que ce service offre deux options supplémentaires : une carte du site avec 3 présentations possibles et un système permettant d'être averti du changement d'une page par e-mail.

Points positifs :

- facilité de mise en œuvre
- pas de limite du nombre de pages référencées (mais une limite globale selon la taille de l'ensemble des pages)
- possibilité de rajouter plusieurs sites et/ou plusieurs pages
- possibilité d'exclure des répertoires et des pages de l'indexation (tags et robots.txt)
- options avancées pour les recherches (+, -, ? et *)
- réindexation automatique paramétrable et/ou manuelle
- gestion des caractères accentués
- statistiques des requêtes effectuées par les visiteurs

- service supplémentaire de carte du site
- service supplémentaire d'information des modifications par e-mail

Points négatifs :

- anglophone (pas trop gênant)
- bannières de publicité
- pas de possibilité de gérer les sites avec mot de passe
- pas de possibilité de recherche par secteurs (groupe de pages)
- pas de conditions booléennes pour les recherches (ET, OU), le moteur choisit par défaut "ET" et ne passe au "OU" que si il n'a pas trouvé de résultat avec le "ET"
- pas de possibilité de recherche des expressions
- pas d'indication sur l'ordre de tri des résultats
- page de résultats trop chargée
- personnalisation (limitée) de la page de résultats

Indexmysite

Ce moteur de recherche est très intéressant par ses possibilités avancées de requêtes booléennes, mais ne supporte pas encore l'accentuation française (amélioration de la gestion des pages de codes prévue dans les mois à venir).

Points positifs :

- facilité de mise en oeuvre
- possibilité de rajouter plusieurs sites et/ou plusieurs pages
- possibilité d'exclure des répertoires et des pages de l'indexation (robots.txt et exclusion de liens)
- conditions booléennes basiques pour les recherches ("Tous les mots", "Chaque mot") ...
- mais aussi (ET, OU, NOT) dans les options avancées
- statistiques des requêtes effectuées par les visiteurs (globales uniquement)
- service supplémentaire de carte du site
- personnalisation très simple de la page de résultats (mais présentation globale figée)

Points négatifs :

- anglophone (pas trop gênant)
- bannière de publicité
- pas de possibilité de gérer les sites avec mot de passe
- pas de possibilité de recherche par secteurs (groupe de pages)
- réindexation manuelle (pas de planification)
- pas de possibilité de recherche des expressions
- gestion des caractères accentués annoncée, mais pas effective au vu des tests

Le CGI

Ce service se présente différemment des autres par son mode d'utilisation qui implique une saisie manuelle de tous les renseignements dans un fichier texte. En fait, ce service ne convient pas pour un moteur de recherche interne (surtout avec 200 liens maximum), mais doit être plutôt dédié à des besoins plus ciblés.

Points positifs :

- francophone
- possibilité de gérer les pages de son choix (même externes au site)
- gestion des caractères accentués
- conditions booléennes basiques pour les recherches ("Tous les mots", "Chaque mot")
- contrôle total des données référencées car la déclaration des titres, des liens et des commentaires est manuelle (mais voir défauts)
- possibilité d'utiliser une "roulette" (renvoie une URL au hasard parmi la liste)

Points négatifs :

- mise en œuvre particulière (déclaration manuelle des URL, des titres et des descriptions via un fichier texte)
- limitation à 200 liens
- pas de possibilité de gérer les sites avec mot de passe
- pas de possibilité de recherche par secteurs (groupe de pages)
- pas de notion de réindexation (il faut renvoyer le fichier texte mis à jour)
- pas de possibilité de recherche des expressions
- personnalisation de la page résultats limitée (logo, police, couleurs et target)
- pas de possibilité de ne pas faire apparaître les mots clés ou le texte servant à la recherche ou de réduire leur taille d'affichage (nb de caractères par exemple), ce qui éviterait d'avoir le contenu intégral des pages trouvées qui s'affiche dans celle de résultats
- pas de statistiques sur les recherches des visiteurs

MyComputer.com

Ce service, ne fonctionnant pas au moment de l'étude, n'a pas pu être testé.

Ovh

Comme pour le site Bravenet, les raisons qui peuvent motiver l'utilisation de ce moteur de recherche sont très faibles. Essentiellement, la saisie manuelle et l'impossibilité de réindexation du site en un seul coup (manuel ou automatique) le destinent à des sites dont les pages existantes ne bougent pas et dont le contenu importe peu pour la recherche car tout est basé sur des mots clés.

Points positifs :

- francophone
- pas de limite du nombre de pages référencées
- possibilité de gérer les pages de son choix
- gestion des caractères accentués

Points négatifs :

- mise en oeuvre fastidieuse (saisie manuelle de toutes les URL de toutes les pages !)
- bannière de publicité
- pas de possibilité de gérer les sites avec mot de passe
- pas de possibilité de recherche par secteurs (groupe de pages)
- obligation de mettre des balises META sur toutes les pages à référencer
- recherche limitée à ces mots clés et aux titres des pages
- réindexation manuelle page par page ! (pas de planification)
- pas de conditions booléennes pour les recherches (ET, OU, NOT)
- pas de possibilité de recherche des expressions
- personnalisation (limitée) de la page de résultats
- pas de statistiques sur les recherches des visiteurs

PicoSearch

Ce moteur est l'un des meilleurs avec "Atomz". Il est sans publicité, hautement paramétrable et possède quelques petites options intéressantes. Il est limité à 1500 pages.

Points positifs :

- pas de bannière de publicité

- possibilité de rajouter plusieurs sites et/ou plusieurs pages
- possibilité d'exclure des répertoires et des pages de l'indexation (robots.txt)
- possibilité de gérer les sites avec mot de passe
- possibilité de recherche par secteurs (groupe de pages)
- indexation des fichiers TEXTE, MP3, Shockwave, Flash et Midi
- conditions booléennes basiques pour les recherches ("Tous les mots", "Chaque mot")
- possibilité de recherche des expressions
- gestion des caractères accentués
- personnalisation très avancée de la page de résultats
- affichage des lignes où les mots ont été trouvés, ceux-ci étant mis en gras
- choix de la langue d'affichage des résultats
- statistiques des 20 recherches les plus demandées par les visiteurs

Points négatifs :

- anglophone (anglais nécessaire pour tout personnaliser correctement)
- réindexation mensuelle ou manuelle (pas de planification)

Searchbutton

Ce service pose problème à cause de sa mauvaise gestion des pages de codes (caractères accentués), ce qui est dommage car sa gestion des recherches booléennes est intéressante. Bref, un service limité pour les sites francophones.

Points positifs :

- facilité de mise en œuvre
- conditions booléennes pour les recherches (ET, OU, NOT), mais uniquement avec le formulaire avancé
- possibilité de recherche des expressions
- formulaire de recherche avancé très riche
- contrôle du tri de présentation des résultats (score ou date)
- affichage des lignes où les mots ont été trouvés
- affichage du texte optionnel (possibilité de ne laisser que les titres)
- statistiques précises sur les recherches des visiteurs

Points négatifs :

- anglophone
- bannière de publicité
- pas de possibilité de rajouter plusieurs sites et/ou plusieurs pages
- possibilité d'exclure des répertoires et des pages de l'indexation
- pas de possibilité de gérer les sites avec mot de passe
- pas de possibilité de recherche par secteurs (groupe de pages)
- réindexation mensuelle ou manuelle (pas de planification)
- pas de gestion des caractères accentués (sauf si recherche avec codes HTML (´ pour é), ce qui n'est pas très pratique !)
- page de résultats un peu trop chargée
- personnalisation (limitée) de la page de résultats

Voila

Ce service est un peu spécial car ce n'est pas un vrai moteur de recherche interne, mais plutôt une utilisation du moteur de recherche global ciblée à un site. Les résultats sont donc fortement dépendants de la qualité du référencement du site sur le moteur de recherche.

Points positifs :

- facilité de mise en œuvre après l'inscription (voir défauts pour ce point)

- francophone
- possibilité d'exclure des répertoires et des pages de l'indexation (robots.txt)
- gestion des caractères accentués
- le moteur Voila étant un des plus utilisés en France, l'optimisation de son site pour améliorer la qualité des résultats dans le moteur de recherche interne bénéficie aussi à son référencement global et réciproquement

Points négatifs :

- bannières de publicité et les liens vers tous les services Voila
- pas de possibilité de rajouter plusieurs sites et/ou plusieurs pages
- pas de possibilité de gérer les sites avec mot de passe
- pas de possibilité de recherche par secteurs (groupe de pages)
- réindexation pas gérable du tout
- pas de conditions booléennes pour les recherches (ET, OU, NOT)
- pas de possibilité de recherche des expressions
- aucune personnalisation possible
- page de résultats chargée avec tout le graphisme de Voila
- nécessite une bonne gestion des balises META et des frames pour obtenir un bon référencement sur le moteur de recherche Voila
- pas de statistiques sur les recherches des visiteurs

whatUseek

Site de bonne qualité, surtout au niveau de la personnalisation de l'affichage des résultats, mais un peu juste pour les options de recherche.

Points positifs :

- facilité de mise en œuvre
- possibilité de gérer les sites avec mot de passe
- possibilité d'exclure des répertoires et des pages de l'indexation (robots.txt)
- possibilité de déclarer des synonymes pour faciliter la recherche
- possibilité d'exclure certains mots de la recherche
- possibilité de recherche par secteurs (groupe de pages)
- gestion des caractères accentués
- personnalisation complète de la page de résultats (paramétrage en HTML)
- contrôle du tri de présentation des résultats (très précis)
- affichage des lignes où les mots ont été trouvés, ceux-ci étant mis en gras
- affichage du texte optionnel (possibilité de ne laisser que les titres)
- statistiques sur les recherches des visiteurs

Points négatifs :

- anglais nécessaire pour tout personnaliser correctement
- bannière de publicité
- pas de possibilité de rajouter plusieurs sites et/ou plusieurs pages
- réindexation mensuelle ou manuelle (pas de planification)
- pas de conditions booléennes pour les recherches (ET, OU, NOT) (seulement ET)
- pas de possibilité de recherche des expressions

X-recherche

Moteur francophone intéressant grâce au nombre de pages gérées (1500) et pour sa simplicité. Même si le nombre d'options proposé n'est pas des plus conséquents, les principales sont là et il conviendra donc très bien pour la mise en place d'une solution simple. La recherche avec plusieurs mots clés semble néanmoins moins efficace qu'avec d'autres solutions.

Points positifs :

- facilité de mise en oeuvre
- francophone
- possibilité d'exclure des répertoires et des pages de l'indexation (robots.txt)
- conditions booléennes pour les recherches (ET, OU, NOT)
- gestion des caractères accentués
- option "mot exact"
- affichage du texte optionnel (possibilité de ne laisser que les titres)
- statistiques sur les recherches des visiteurs

Points négatifs :

- bannière de publicité
- pas de possibilité de rajouter plusieurs sites et/ou plusieurs pages
- pas de possibilité de gérer les sites avec mot de passe
- pas de possibilité de recherche par secteurs (groupe de pages)
- réindexation mensuelle ou manuelle (pas de planification)
- élimine les accents à la recherche (par exemple : recherche aussi bien "inséré" que "insere", "insère", etc... si on demande spécifiquement "inséré" ou "insere")
- pas de possibilité de recherche des expressions
- personnalisation (limitée) de la page de résultats
- efficacité moins intéressante pour les recherches avec plusieurs mots clés

Synthèse comparative des services en ligne de moteurs de recherche internes

Dans cette synthèse comparative, on n'a retenu que les cinq solutions les plus intéressantes, après une première sélection.

Voir le tableau donné en annexe 1.

Solution proposée

On propose en premier choix, le moteur Atomz qui est le meilleur du panel sur la plupart des points de comparaison (ses seuls inconvénients sont la langue anglophone pour l'administration et la limitation à 500 pages indexées). Le second choix se porterait sur Pico Search, dont l'efficacité, mesurée avec la pertinence des résultats des recherches, est un peu inférieure à celle de Atomz. Extense viendrait ensuite. Ecila a été placé en 4^{ème} choix en raison de son efficacité et de son indisponibilité actuelle. Enfin, Antinea qui pourrait être notre deuxième meilleure solution, a été placé en dernier à cause de problèmes de pérennité de la solution.

Bibliographie

Abondance : <http://www.abondance.com/ressources/tmcomp.html>

Le coin des Experts : <http://perso.wanadoo.fr/coin.des.experts>

Annexe 1 : tableau comparatif des services en ligne de moteurs de recherche internes

	Pertinence des résultats des recherches	Personnalisation de la page de résultat des recherches	Limitation en nombre de pages indexées	Réindexation	Langue pour l'administration de la solution	Logo/bannière de publicité	Référencement multidomaine	Version gratuite ou payante	Particularité
Atomz	+++++	Très avancée	500	Automatique paramétrable, ou manuelle	Anglais	Logo à insérer	Oui	Gratuit	
Pico Search	+++	Très avancée	1500	Mensuelle ou manuelle (pas de planification)	Anglais	Logo à insérer	Oui	Gratuit	
Extense	++	n.c.	50 en version gratuite	Tous les 15 jours (version gratuite) Automatique paramétrable, ou manuelle (version payante)	Français	Logo en version gratuite	Non	2 versions, une gratuite, une payante	
Ecila	++	Non	n.c.	Automatique, 1 fois par semaine	Français	Bannière en page de résultats	Non	Gratuit	Indisponible actuellement
Antinea	++++	Limitée	Aucune	Automatique, 2 fois par semaine	Français	Logo en page de résultats	Oui	Gratuit	Hébergé sur un serveur universitaire (problème de pérennité)

