

Pourquoi les pages de votre site ne sont-elles pas référencées par les moteurs ?

[Retour au sommaire de la lettre](#)

Votre site est mal référencé sur les moteurs de recherche, voire totalement absent de leur index ? Les causes - qui existent toujours - peuvent en être multiples. En voici les principales (5 raisons qui font que votre site peut ne pas être indexé et 6 pour qu'il soit mal indexé) avec, pour chacune d'entre elles, les remèdes à mettre en place pour obtenir de meilleurs résultats le plus rapidement possible...

Vous vous débâtez pour que votre site obtienne une meilleure présence dans les index des moteurs ? Vous voudriez qu'un plus grand nombre de vos pages soient indexées ? C'est une préoccupation assez normale pour un webmaster sain d'esprit :-) Nous allons essayer, dans cet article, de lister les raisons pour lesquelles un site est mal pris en compte par les différents moteurs de recherche en tentant d'y apporter les remèdes adéquats.

Attention : il ne s'agit pas ici d'essayer d'obtenir de meilleurs positionnements de vos pages en réaction à certains mots clés, mais uniquement de faire en sorte que le plus grand nombre possible de vos pages soient référencées, donc "trouvables" sur les moteurs. Nous ne parlerons donc pas ici d'optimisation du code HTML ou de ce type de pratiques visant à améliorer la "réactivité" de vos pages en regard des algorithmes de pertinence des outils de recherche... Chaque chose en son temps : avant d'être en tête de gondole, il faut déjà être dans les rayons... :-)

Nous partirons donc, dans cet article, du constat que votre site n'est pas, ou est mal, indexé par les moteurs. Par exemple : seules 10% de vos pages sont trouvées par les robots. Ou seule votre page d'accueil. Etc. Il y a certainement une raison à cela. Nous allons essayer de la trouver et de vous indiquer comment faire pour améliorer la situation.

Comment lister toutes les pages de son site indexées par les moteurs de recherche ?

Première chose à faire, bien évidemment, avant toute panique : lister toutes les pages de votre site qui sont bien prises en compte par les différents moteurs.

Pour info, le site **Soumission Manuelle**, du réseau Abondance, vous permet de faire ce type d'investigations en quelques clics :
<http://www.soumission-manuelle.com/>
:-)

Pour cela, on utilisera le plus souvent la syntaxe "site:", comprise par la majorité des moteurs actuels.

Google

La commande "site:votresite.com" vous donnera la liste des pages indexées par ce moteur. Exemple : [site:abondance.com](http://www.google.com/search?q=site:abondance.com).

Attention cependant si vous utilisez, comme le site Abondance, des sous-domaines du type annuaire.abondance.com, palmares.abondance.com, actu.abondance.com, etc. La requête "site:abondance.com" vous donnera accès à TOUTES les pages (et donc tous les sous-domaines) du site demandé. Si vous désirez lister les pages d'un sous-domaine précis, tapez : "site:actu.abondance.com" par exemple. **MAIS** la requête "site:www.abondance.com" ne vous listera pas (erreur classique) les pages des sous-domaines actu.abondance.com ou palmares.abondance.com. Uniquement celles présentes sous le nom de serveur www.abondance.com... Ne l'oubliez pas !

Yahoo! Search

Sur Yahoo! également, la syntaxe à utiliser est "site:" : site:abondance.com avec les mêmes restrictions et possibilités que celles évoquées ci-dessus.

MSN Search

"Même motif, même punition" pour MSN Search : site:abondance.com

Exalead

La syntaxe d'interrogation "site:" reste d'actualité sur Exalead : site:abondance.com

Ask

En revanche, contrairement à ce que dit ce moteur, Ask ne comprend pas la syntaxe "site:" seule mais uniquement accompagnée d'un mot clé de recherche comme abondance site:abondance.com. Ce "bug" était présent sur Google pendant longtemps avant qu'il ne soit corrigé par les équipes techniques du moteur. Ce n'est pas (encore) le cas sur Ask : si aucun mot représentatif n'est présent dans vos pages, il ne semble pas qu'il soit possible de faire une requête exhaustive de ce type sur votre site, des recherches comme * site:abondance.com ou ? site:abondance.com ne donnant aucun résultat. Dommage...

Avant toute investigation dans une voie expliquée dans la suite de cet article, n'hésitez pas à faire un "état des lieux" de la présence de vos pages dans les index des moteurs grâce à ces syntaxes de recherche... Combien voit-on de messages dans les forums du type "mon site a disparu de Google !" alors qu'il a simplement été rétrogradé, parfois temporairement, sur certains mots clés...

Pour plus d'infos à ce sujet, lire les articles suivants :
- La syntaxe site: sur les moteurs de recherche majeurs
(<http://abonnes.abondance.com/archives/2005-06/index.html>)

Différentes raisons de non-indexation de votre site par les moteurs

Votre site est refusé par les moteurs de recherche ? Il est pourtant mis en ligne depuis plusieurs semaines ? Il peut y avoir plusieurs raisons à cela... Voici la liste des principales d'entre elles (il peut bien sûr en exister bien d'autres mais ils nous semble avoir listé ci-dessous les principales)...

1. Blocage par un fichier Robots.txt ou des balises meta "robots"

Eh oui, cela peut paraître idiot, mais nous avons de multiples exemples de non-indexation de sites web dûes à la présence d'un fichier "robots.txt" empêchant les robots des moteurs d'indexer le site ou de balise meta "Robots" ayant un paramètre "noindex" dans les pages. Les raisons peuvent en être nombreuses :

- Mise en place de ces informations lors de tests puis oubli de les enlever.
- Mise en place d'un fichier "robots.txt" par l'hébergeur à l'insu de l'hébergé (cela arrive, malheureusement)...
- Mauvaise rédaction du fichier "Robots.txt". Il suffit d'un rien pour qu'une erreur dans ce fichier interdise l'accès à tous les moteurs.

Vérifiez donc vos fichiers "robots.txt" et vos balises meta "robots" afin d'éviter tout problème de ce type. Cela prend quelques secondes et peut éviter bien des désagréments !

De la même façon, si vous avez mis en place des redirections serveurs (301, 302) vérifiez la syntaxe par exemple des fichiers htaccess utilisés, afin d'être sûr que les moteurs les lisent de façon adéquate...

Pour plus d'infos à ce sujet, lire les articles suivants :
- Le fichier Robots.txt (<http://docs.abondance.com/robots.html>).
- Les balise meta "Robots" (http://docs.abondance.com/meta_1.html)

- Optimiser le temps d'entrée d'un site dans les index des moteurs de recherche (<http://abonnes.abondance.com/archives/2003-06/index.html>)
- Changer un nom de domaine et des intitulés d'urls en gardant son référencement (<http://abonnes.abondance.com/archives/2006-01/index.html>)
- Nom de domaine et référencement (<http://abonnes.abondance.com/archives/2005-12/index.html>)

2. Manque de "backlinks" vers votre site

Bien sûr, pour soumettre et proposer votre site web sur les moteurs de recherche, vous pouvez utiliser les formulaires de type "Add URL" (exemple pour Google : <http://www.google.fr/addurl/?hl=fr&continue=/addurl>). Mais la pratique nous indique clairement que la meilleure des voies pour obtenir une indexation rapide et efficace reste d'obtenir le plus vite possible un ou plusieurs liens ("backlinks") vers votre page d'accueil depuis une page populaire, c'est-à-dire disposant d'un PageRank d'au moins 6. Ne connaissez-vous personne qui détienne ce type de page et qui puisse vous "rendre ce service" ?

Si cela vous semble difficile, tentez de multiplier les "backlinks" émanant de pages à PageRank "le plus fort possible". Inscrivez-vous dans des annuaires. Etc. Bref, faites en sorte d'obtenir, dès les "premiers jours de vie" de votre site, des liens émanant de pages "les plus populaires possibles". Evitez cependant de faire n'importe quoi... Trop de liens obtenus trop rapidement peut aussi jouer dans l'effet "Sandbox" sur Google... Tablez donc sur une dizaine ou une vingtaine de liens "rapides" qui feront logiquement indexer votre site sans soucis en quelques jours. Encore une fois, ne soyez pas tenté par des sirènes commerciales comme l'achat de lien au *pro rata* du PageRank ou ce type d'offre qui commence à fleurir sur le Web... Faites les choses proprement, de façon loyale et honnête et vous n'aurez aucun souci par la suite...

Pour plus d'infos à ce sujet, lire les articles suivants :

- Comment intégrer l'index de Google en 3 jours (<http://abonnes.abondance.com/archives/2004-05/index.html>)
- Les échanges de liens "intelligents" (<http://abonnes.abondance.com/archives/2003-10/index.html>)

3. Utilisation de technologies bloquantes pour les robots

Il existe finalement assez peu de technologies qui soient réellement réhivitoires pour les robots et leur interdisent l'accès à vos pages. Mais cela peut arriver. Par exemple, l'obligation d'accepter les cookies, l'emploi d'identifiants de sessions, sur un site dynamique, dès la page d'accueil, peut poser des problèmes et provoquer une non-indexation totale du site. Certains tests d'entrée peuvent également causer souci, notamment si vous testez le navigateur utilisé par l'internaute sur la base du "User Agent" et que vous oubliez le cas "autre" correspondant aux robots qui ne sont pas des navigateurs... Mais, *a priori*, il existe plus de caractères freinants (voir plus loin dans cet article) pour les robots à l'heure actuelle que de technologies strictement bloquantes. Ceci dit, vérifiez quand même que rien, techniquement parlant, ne peut bloquer un robot sur votre site, on ne sait jamais...

Pour plus d'infos à ce sujet, lire les articles suivants :

- Le référencement de sites web dynamiques (1ère partie) (<http://abonnes.abondance.com/archives/2004-06/index.html>)
- Le référencement de sites web dynamiques (2ème partie) (<http://abonnes.abondance.com/archives/2004-07/index.html>)
- Le référencement de sites web dynamiques (3ème partie) (<http://abonnes.abondance.com/archives/2004-09/index.html>)
- Les moteurs de recherche lisent-ils le contenu des iframes ? (<http://abonnes.abondance.com/archives/2005-07/index.html>)
- Les moteurs de recherche lisent-ils les textes dans les menus déroulants ? (1ère partie) (<http://abonnes.abondance.com/archives/2005-05/index.html>)
- Les moteurs de recherche lisent-ils les textes dans les menus déroulants ? (2ème partie) (<http://abonnes.abondance.com/archives/2005-06/index.html>)

- Comment optimiser un site sous Domino / Lotus Notes pour les moteurs de recherche ? (<http://abonnes.abondance.com/archives/2005-06/index.html>)
- Comment les moteurs référencent-ils les pages sécurisées (https) ? (<http://abonnes.abondance.com/archives/2005-04/index.html>)

4. Mise en liste noire

C'est évidemment un cas radical et, surtout, très rare, quoi qu'en pensent beaucoup. Bon, d'un autre côté, si vous êtes en liste noire, c'est certainement que vous l'avez mérité :-). Vous savez donc ce qu'il vous reste à faire....

Pour plus d'infos à ce sujet, lire les articles suivants :
- Mon site est blacklisté. Que faire ? (<http://abonnes.abondance.com/archives/2005-11/index.html>)

5. Problème technique au niveau de l'hébergeur

Il peut arriver que, lorsque le robot arrive sur votre serveur pour indexer ses pages, votre site soit "down" pour des raisons diverses (problème chez l'hébergeur, maintenance, etc.). Si ce souci a pu générer des problèmes par le passé, il semblerait que les robots aient appris à mieux le gérer avec le temps... Lorsqu'un site est inatteignable pour des raisons techniques, son "aspiration" est très rapidement reprogrammée, quelques minutes ou quelques heures plus tard, afin de pallier ce problème. En tout état de cause, un site web n'est jamais, aujourd'hui, ôté d'un index à cause d'une non accessibilité du serveur pendant quelques minutes, voire quelques heures...
Autre point sur lequel on a finalement peu d'infos : si vous êtes sur un serveur mutualisé, donc partagé par plusieurs sites ne vous appartenant pas forcément, un moteur de recherche peut vous pénaliser en bloquant l'adresse IP d'un site ayant par trop spammé son index. Même si les moteurs s'en défendent, ce type de cas de figure peut arriver. Vérifiez donc, en utilisant des outils comme Whois.sc (<http://www.whois.sc/>) ou la recherche par adresse IP de MSN Search, quels sont les autres sites hébergés sur le même serveur que le vôtre et regardez s'ils ne sont pas potentiellement des sources de problèmes pour les moteurs...

Pour plus d'infos à ce sujet, lire les articles suivants :
- Redirections : comment les gérer au mieux pour le référencement de votre site ? (<http://abonnes.abondance.com/archives/2005-10/index.html>)
- Question aux moteurs : l'incidence d'un changement d'hébergeur sur un référencement (<http://abonnes.abondance.com/archives/2003-04/index.html>)
- La syntaxe avancée de MSN Search (<http://abonnes.abondance.com/archives/2005-12/index.html>)

Globalement, il nous semble avoir fait le tour des différentes raisons pour lesquelles un site peut être **totalemment absent** de l'index d'un moteur de recherche. En tout cas des raisons les plus générales, car il peut bien entendu subsister, au cas par cas, des raisons très spécifiques impossibles à évoquer ici... Voyons maintenant ce qui peut faire en sorte qu'il soit **mal indexé**...

Différentes raisons de mauvaise indexation de votre site par les moteurs

Dans ce cas, votre site est présent dans l'index, mais avec parcimonie. Par exemple, uniquement votre page d'accueil ou quelques pages internes mais pas votre page d'accueil, etc. Cela peut-être dû à plusieurs raisons qui ne sont d'ailleurs pas antinomiques...

1. Problèmes d'incompatibilité sur les liens de votre page d'accueil

On le sait, les robots des moteurs ont du mal à suivre les liens écrits notamment en JavaScript, par exemple. Si votre site web présente ce "défaut" (au sens des robots en tout cas), il vous faut le contrer, par exemple, par l'adjonction d'un "plan du site" très détaillé et présentant des liens vers

un grand nombre de vos pages. Evidemment, le lien depuis la page d'accueil vers le plan du site doit être un lien "classique", donc non écrit en JavaScript, pour que les robots puissent le suivre. Il en sera de même pour les liens à l'intérieur de la page "Plan du site". Ce type de page est très important pour le référencement, pourtant beaucoup de webmasters le négligent. C'est bien dommage car il rend de nombreux services et facilite grandement l'indexation des pages... Autre solution : l'utilisation d'outils de type "Google SiteMaps" (à prendre en compte en complémentarité avec votre plan du site, puisque seul Google le propose) :

<http://www.google.com/webmasters/sitemaps/login>.

Nous parlons ici des liens de la page d'accueil mais il est évident que si tous les liens de votre site sont incompatibles avec les robots des moteurs, cela peut poser de nombreux problèmes futurs puisque vous perdez, dans ce cas, leur maîtrise ainsi que celle des textes qui les composent ("texte offshore"), ce qui est clairement pénalisant pour votre positionnement et votre trafic "moteur"... Bref, dans tous les cas, il est de toutes façons plus efficace de proposer aux robots des liens qu'ils comprennent... Si cela n'est pas possible pour de multiples raisons, optez pour un plan du site détaillé...

Pour plus d'infos à ce sujet, lire les articles suivants :

- Google et les liens JavaScript (<http://abonnes.abondance.com/archives/2005-01/index.html>)
- Comment rendre des liens JavaScript compatibles avec les spiders des moteurs (<http://abonnes.abondance.com/archives/2004-10/index.html>)
- Comment mieux référencer son site avec Google Sitemaps (1ère partie) (<http://abonnes.abondance.com/archives/2005-09/index.html>)
- Comment mieux référencer son site avec Google Sitemaps (2ème partie) (<http://abonnes.abondance.com/archives/2005-10/index.html>)

2. Sites en Flash

Si votre site est entièrement réalisé en Flash, seule la page d'accueil (la page HTML qui lance l'animation Flash) est en théorie indexable, même si les animations Flash deviennent, au fil du temps, de mieux en mieux référencées par les moteurs. Il se posera cependant très rapidement un problème de positionnement et de trafic émanant des moteurs car pour l'instant, il est extrêmement rare que l'on voie une animation Flash, même si elle est indexée, bien positionnée sur un mot clé donné... La seule solution quantitative est donc de développer un deuxième site, développé en HTML, à l'attention des robots et des internautes ne disposant pas du plug-in Flash... Et oui, ce format a des avantages et des inconvénients...

Pour plus d'infos à ce sujet, lire les articles suivants :

- Comment référencer un site web réalisé avec des frames ou en Flash (<http://www.abondance.com/abonnes/archives/acturech/0204.html>)

3. Problème de profondeur des liens

Nous avons vu, depuis un certain temps, subvenir pas mal de problèmes d'indexation sur des sites web pour lesquels, pour atteindre certaines pages internes, il fallait de nombreux clics de souris depuis la page d'accueil. Il semblerait que ce critère soit un frein pour l'indexation quantitative sur de nombreux moteurs. L'idéal, donc, serait de faire en sorte que chaque page de votre site web soit accessible en 3 clics maximum depuis la page d'accueil. Là encore, l'utilisation d'un "plan du site" peut être essentielle...

Comprenons-nous bien : il ne s'agit pas là d'une "profondeur physique" dans les répertoires du site. Par exemple, une page ayant comme adresse www.votresite.com/rep1/rep2/rep3/rep4/rep5/index.html a autant de chance d'être indexée qu'un document accessible à l'url www.votresite.com/rep1/index.html... On parle ici du nombre de "clics" nécessaires au robot pour atteindre l'une ou l'autre page. Peu importe leur emplacement physique sur le serveur...

4. Absence de liens

Certaines pages de votre site peuvent ne pas être accessibles au travers de liens tout simplement. Exemple typique : l'utilisation d'un moteur de recherche interne pour accéder à vos "fiches produits".



The image shows a web interface for a real estate search. At the top, it says "VOTRE PROJET" and "Offre commerciale au 10/02/06 : 3257 lots". On the left is a photo of a woman holding a small house model. To the right are several filter options with checkboxes: "Appartement neuf", "Maison neuve", "Terrain", "Résidence principale", "Investissement", and "Résidence secondaire". Below these are input fields for "Votre budget", a dropdown for "Département" (set to "01 - Ain"), and a field for "Tous nos prix ?" with "code privilège" below it. A "VALIDER" button is at the bottom right. At the bottom left of the interface, there are two buttons: "Recherche DETAILLÉE" and "Recherche GÉOGRAPHIQUE".

Exemple-type d'interface d'accès à des fiches produits non compatible avec les robots des moteurs

Inconvénient : les robots ne savent pas faire des choix dans les menus déroulants, cliquer sur des boutons radios et valider un formulaire. Ils sont donc bloqués à ce niveau et ne viendront jamais indexer vos pages de contenu - qui sont pourtant certainement les plus importantes - si vous ne leur donnez pas un autre accès. La solution ? Toujours ce bon vieux "Plan du site" et les outils de type "Google SiteMaps"...

5. Problème de durée de vie et d'absence de PR, de mises à jour fréquentes et de backlinks

Il semblerait également que la plus ou moins forte indexation d'un site (en termes de nombre de pages) dépende, notamment sur Google, de plusieurs autres facteurs comme :

- Durée de vie : de façon assez logique, plus le site est "vieux", plus les robots du moteur "découvrent" de nouvelles pages et les indexent, au fur et à mesure.
- Backlinks et PageRank : plus le site dispose de backlinks (liens entrants), notamment vers ses pages internes, et plus le PageRank de la page d'accueil augmente, mieux le site est indexé quantitativement parlant.
- Délais de mise à jour : Il semblerait qu'un site web souvent mis à jour soit mieux indexé, toujours au niveau quantitatif, qu'un site rarement modifié... N'oubliez pas d'en tenir compte...

Pour plus d'infos à ce sujet, lire les articles suivants :

- La demande de brevet "Information Retrieval Based on Historical Data" de Google (<http://abonnes.abondance.com/archives/2005-05/index.html>)

6. Problème de "Duplicate Content"

Certaines pages de votre site peuvent être indexées par les moteurs mais ceux-ci, considérant leurs contenus comme trop proches, les classent en "duplicate content". Bref, ils considèrent que de nombreux documents, sur votre site, sont identiques. Souvent à tort parfois, d'ailleurs. Il faut donc faire en sorte que, notamment, le début de chaque page de votre site soit très différent, et notamment le trio "Title - Balise Meta Description - Balise Meta Keywords"...

Pour plus d'infos à ce sujet, lire les articles suivants :

- Comment éviter le "duplicate content" sur Google

(<http://abonnes.abondance.com/archives/2005-01/index.html>)

Conclusion

Nous avons fait, dans cet article, un tour qui nous semble le plus complet possible des raisons pour lesquelles un site peut être mal indexé voire absent des index des moteurs de recherche. Sachez qu'il faut cependant un peu de temps avant d'obtenir des résultats satisfaisants à ce niveau. Si l'indexation de la page d'accueil de votre site peut s'effectuer rapidement (quelques jours), le délai d'attente pour une meilleure exhaustivité quantitative se chiffre souvent en semaines voire en mois... Il faut savoir être patient dans le domaine du référencement...