

Le référencement de sites web dynamiques (2ème partie)

[Retour au sommaire de la lettre](#)

Le référencement de sites web dynamiques est l'une des principales sources d'interrogations des webmasters actuels. Après avoir longtemps été un facteur totalement bloquant pour les moteurs de recherche, la situation s'est assouplie depuis quelques temps. Le mois dernier, nous avons étudié quels sont les types de pages qui sont compatibles avec les moteurs et ceux qui ne le sont pas, ainsi que les problèmes posés aux moteurs par les sites dynamiques. Ce mois-ci, nous traitons des sept différentes façons de traiter le problème proposées par les sociétés de référencement actuelles.

Nous avons vu, le mois dernier, que les sites web dynamiques sont parfois des obstacles pour les moteurs de recherche, mais que la situation a plutôt tendance à s'améliorer. Cependant, parfois, certaines pages représentent un réel obstacle pour les robots et il faut intervenir, mettre en œuvre une procédure spécifique, pour faire en sorte qu'un site soit indexé par les Google et consort.

Dans ce cas, il existe plusieurs façons de prendre en compte le fait qu'un site web soit dynamique. Les référenceurs professionnels proposent, la plupart du temps, plusieurs solutions, parfois très différentes. Ce sont ces diverses solutions que nous allons passer en revue dans ce deuxième article. Commençons par la technique la plus souvent pratiquée...

Solution 1 : la page satellite

Il s'agit là de la solution la plus souvent proposée sur le marché par les sociétés de référencement. Elle est simple dans sa conception et revient à dire : "Votre site est dynamique et peut présenter des blocages pour les moteurs ? Pas de problèmes, nous allons créer de nouvelles pages, spécialement construites pour le référencement et ce sont sur celles-ci que nous allons travailler."

Une page satellite (ou page alias, fantôme, doorway et bien d'autres noms, différents selon la société de référencement mais reposant tous sur la même philosophie) est une page HTML optimisée pour un mot clé ou une expression donnée et contenant une redirection - le plus souvent en Javascript - vers une page "réelle" du site.

La page satellite est donc prise en compte par le moteur qui, si elle est bien optimisée, la classe en bonne position sur le mot clé pour laquelle elle a été constituée. Lorsque l'internaute clique sur le lien, dans la page de résultat du moteur, il va vers la page satellite et la redirection automatique qu'elle contient le renvoie directement vers la page du site web final. Le tour est joué.

Avantages de la page satellite :

- Le site web en lui-même n'est pas touché, pas modifié, le client garde donc la totale maîtrise de son contenu et de ses pages web et sous-traite totalement l'aspect "référencement". Les pages satellites peuvent même être hébergées chez le référenceur, comme cela se passe souvent.
- La page satellite n'étant pas vue par l'internaute, le référenceur peut réellement travailler de façon pointue son contenu et son optimisation.

Inconvénients de la page satellite :

- Son contenu appartient parfois au référenceur, notamment s'il héberge lui-même les pages satellites de son client. Certains contrats peuvent ne pas être très clairs à ce sujet. Si le client change de référenceur, il lui faudra peut-être redémarrer le travail à zéro. Si vous désirez passer par ce type de prestation, assurez-vous bien que les pages satellites développées pour votre société restent votre propriété en fin de contrat...
- Il est toujours plus difficile d'obtenir un bon PageRank (la popularité selon Google) sur une page satellite que sur une page réelle du site. En effet, il est rare que les

sites qui pointent vers vous aient établi des liens vers vos pages satellites... Même si un maillage complexe et élaboré de différentes pages satellites entre elles permet de pallier le plus souvent ce problème, on ne sait pas vraiment jusqu'à quand cela va fonctionner...

- Certains moteurs de recherche indiquent clairement dans leurs conditions générales d'utilisation qu'ils sont contre ces techniques (même s'ils ne les combattent pas toujours de façon assidue). Voir plus loin dans cet article.
- Certaines sociétés de référencement ou webmasters font, malheureusement, réellement n'importe quoi avec les pages satellites (mots clés n'ayant aucun rapport avec le contenu du site, liens cachés, etc.) actuellement, ce qui risque de discréditer cette méthode aux yeux des moteurs. La situation nous rappelle un peu celle que l'on a connu avec les balises Meta, qui étaient une superbe zone d'information pour renseigner les moteurs sur le contenu d'une page et une parfaite opportunité pour le référencement. Beaucoup de gens ont usé et abusé de cette possibilité et l'ont dénaturée. Résultat : les moteurs n'en tiennent plus compte ou si peu... En sera-t-il de même dans le futur avec les pages satellites ? La question, à notre avis, reste posée... On a, d'ailleurs, vu depuis environ un mois bon nombre de page satellites dégringolant dans les résultats de Google, certains référenceurs perdant du jour au lendemain plus de 20% de leur trafic. Google aurait-il commencé une "chasse aux pages satellite" ?

Ceci dit, il serait faux de dire que la technique de la page satellite ne fonctionne pas aujourd'hui. Elle représente même une véritable alternative pour pallier les problèmes techniques dûs au dynamisme d'un site. C'est un peu la "solution de repli", à nos yeux, s'il n'existe pas de réelle solution d'intervenir directement sur le site, son contenu et ses urls pour les optimiser, ce qui reste, à notre avis, toujours la meilleure solution de référencement.

Il faut également noter que, dans sa page d'information pour les webmasters et référenceurs (<http://www.google.com/webmasters/seo.html>), Google indique qu'il n'aime pas vraiment les pages satellites : *"If an SEO creates deceptive or misleading content on your behalf, such as doorway pages or "throwaway" domains, your site could be removed entirely from Google's index."* Un webmaster averti en vaut deux : terrain glissant...

A quoi reconnaît-on une page satellite dans une page de résultats ? A deux points rapidement identifiables :

- 1. Lorsque l'on clique sur le lien, on est redirigé vers une autre page (le titre de la page et son url changent).
- 2. Le mot clé pour lequel elle est optimisée se trouve souvent être le nom de la page en question. Exemple pour l'expression "audit conseil", on trouve dans les pages de Google des résultats du type :
www.siteweb.fr/audit_conseil.htm
www.siteweb.com/cabinet_conseil/audit_conseil.htm
www.siteweb.fr/fr/ingenierie/audit_conseil.htm
www.siteweb.fr/audit-conseil-immobilier.html

Le fait que le nom du fichier HTML corresponde aux mots de la requête est un premier indice de la présence d'une page satellite dans les résultats du moteur de recherche. Si vous cliquez sur le lien en question et que vous êtes redirigé vers un autre document, vous pouvez vous dire que vous avez mis le doigt (ou plutôt la souris) sur un telle page.

Essayez des mots clés porteurs, comme "assurance vie", "prêt bancaire" ou autre sur Google, et vous identifierez ainsi rapidement quelles pages sont des "satellites".

Comment lire le code HTML d'une page satellite ? Bonne question, car ce type de document contenant une redirection automatique et immédiate, il n'est pas facile, voire souvent impossible d'en lire le code. Il existe de multiples méthodes pour lire ce code aisément. En voici deux. La première :

- 1. Copiez l'url de la page satellite telle qu'identifiée dans les pages de résultats du moteur (exemple : http://www.siteweb.fr/audit_conseil.htm).
- 2. Allez sur le site Outiref (<http://www.outiref.com/>).
- 3. Dans la zone "Tapez l'adresse d'une page web :", collez l'adresse de la page satellite et cliquez sur le bouton "Valider".
- 4. Sur la page de résultats du site Outiref, cliquez sur le lien "voir code source".

Et voilà ! Vous avez accès au code source de la page satellite. A vous de l'analyser (par copier/coller dans votre éditeur HTML préféré) et de comprendre pourquoi celle-ci est si bien classée ;-)

Autre solution sur Internet Explorer notamment : indiquer la mention "view-source:" dans la barre d'adresse de votre navigateur. Par exemple, à la place de http://www.siteweb.fr/audit_conseil.htm, vous allez indiquer view-source:http://www.siteweb.fr/audit_conseil.htm
Vous aurez directement accès au code source de la page en question.

Conclusion au sujet de la page satellite : il s'agit :

- D'une pratique courante, largement utilisée par les référenceurs professionnels. Certains n'ont même, en pratique, que celle-ci dans leur portefeuille :-)
- Qui fonctionne aujourd'hui mais pour laquelle on peut se poser des questions au sujet de la pérennité.
- Qui, à notre sens (mais le débat reste ouvert), peut être utilisée s'il n'existe pas de réelle solution d'intervenir directement sur le site, son contenu et ses urls pour les optimiser, ce qui reste, à notre avis, toujours la meilleure solution de référencement.

Le problème vient parfois, hélas, du client... Nous avons eu écho de référenceurs à qui le client aurait dit : "De toutes façons, il n'est absolument pas question que vous touchiez à notre site pour le référencer". Dans ce cas, la page satellite est effectivement une solution intéressante, même si on ne peut s'empêcher de penser que le client fait fausse route avec ce type d'affirmation... Un référenceur a même eu comme préambule de la part d'un prospect : "De toutes façons, si vous ne proposez pas de pages satellites, nous ne signerons pas avec vous"... Il y a encore du travail à faire en termes d'évangélisation dans le domaine du référencement... Les clients ne sont pas toujours très sérieux... ;-))

2. Le cloaking

La technique du cloaking (ou IP delivery) est, finalement, assez proche de celle de la page satellite. Imaginons que votre page d'accueil s'appelle index.html. Vous allez créer, dans un premier temps, une copie conforme de cette page. La première sera la page originelle du site, l'autre sera optimisée.

Les systèmes de cloaking sont ensuite installés sur le serveur, sous forme de logiciel spécifiques, et tentent d'identifier qui arrive sur vos pages :

- Soit c'est le robot d'un moteur (reconnaisable par son "user agent" et/ou son adresse IP, dont des bases de données sont aisément identifiables sur le Web) et, dans ce cas, le système lui fournit la page optimisée (l'équivalent de la page satellite).
- Soit c'est un internaute et, dans ce cas, le serveur lui envoie la page "normale".

L'analogie avec la page satellite n'est pas fautive puisque, finalement, seule la fonction de redirection est différente. Dans un cas, elle est mise en place au moment du clic (page satellite), dans l'autre cas, au moment de la visite du robot.

Que penser de cette technique ? Eh bien, un peu la même chose qu'avec les pages satellite. Elle constituerait, après tout, une bonne façon de pallier les problèmes techniques posés par les sites web dynamiques. Sauf que... La plupart des gros moteurs de recherche, dont Google, ont indiqué par le passé qu'ils refusaient ce type de solution, considérée comme du spam. Google est très clair à ce sujet sur son site (<http://www.google.com/webmasters/faq.html>) : "To preserve the accuracy

and quality of our search results, Google may permanently ban from our index any sites or site authors that engage in cloaking to distort their search rankings."

Avantages du cloaking :

- Peu complexe à mettre en œuvre, la plupart des infos et des outils sont disponibles en ligne.
- Mêmes avantages que pour la page satellite.

Inconvénients du cloaking :

- Les gros moteurs de recherche n'apprécient que modérément ce type de pratique. Risque de liste noire assez important.
- Mêmes inconvénients que pour la page satellite.

Il sera donc difficile de prendre en compte le cloaking dans le cadre du référencement d'un site web dynamique.

Cependant, il semblerait que les moteurs de recherche nuancent, petit à petit, leur opinion sur le cloaking. En effet, on a vu Google, dans un passé récent, mettre en liste noire un site ayant utilisé ce type de pratique (<http://actu.abondance.com/2004-21/whenu.html>) mais également accepter "officiellement" qu'un autre le fasse (<http://searchenginewatch.com/sereport/article.php/3360681>). Le discours n'est donc pas totalement radical. En tout état de cause, l'idéal serait de contacter les moteurs de recherche pour leur exposer votre cas si vous désirez utiliser du cloaking dans le cadre de votre référencement. Le risque reste aujourd'hui élevé sinon...

Nota : Voir également, à ce sujet, l'article de la lettre R&R :

- Le cloaking : spam ou pas spam ? (mars et avril 2003)

<http://abonnes.abondance.com/archives/parthemes.html>

3. La recopie de site web

Autre solution de plus en plus proposée par les sociétés de référencement, notamment les plus importantes : la recopie pure et simple du site web de leur client sur leurs serveurs afin de le rendre "statique" (proposant des urls compatibles avec les contraintes des moteurs).

Le principe est simple : le référenceur effectue une "copie conforme", un "clone" de votre site sur son serveur. Ce sont alors ces pages, dont les urls auront été rendues "spider compatibles" et le contenu optimisé, qui seront prises en compte pour le référencement. Par ce biais, le référenceur rend statique un site dynamique.

Avantages de la recopie de sites web :

- Toutes les pages peuvent ainsi être rendues statiques et peuvent alors être optimisées pour le référencement, puisque le référenceur ne travaille que sur un "clone" du site réel.

Inconvénients de la recopie de sites web :

- La mise à jour des pages doit être prise en compte (par exemple les prix des articles ou leur description ou les nouveaux articles). Si le site est mis à jour quotidiennement, la "moulinette" peut vite ressembler à une "usine à gaz"...

- La solution peut s'avérer complexe à mettre en œuvre, donc onéreuse. Cette solution sera peut-être réservée aux sites importants, disposants de budgets de référencement conséquents.

Conclusion : une solution qui peut s'avérer intéressante mais qui n'est pas à la portée de toutes les bourses...

4. Création de pages de contenus

Autre solution proposée par les sociétés de référencement : la création de pages web, sans redirection, hébergée sur votre site ou sur le serveur du référenceur, sans redirection, proposant un contenu spécifique et, bien sûr, une url "propre" pour les moteurs.

L'idée, là encore, est simple : il s'agit de créer des pages, disposant d'une adresse statique, et qui, à l'instar des pages satellites, serviront en priorité au référencement de votre site. Mais, là où les pages satellites redirigent l'internaute vers une page réelle du site, la page de contenu propose un contenu réel, basé sur celui de votre site originel et respectant, donc, votre charte graphique.

Le référenceur travaille avec le client et crée des pages en "piochant" du contenu sur le site web de départ, contenu qui lui sert à proposer du code HTML optimisé (le premier paragraphe, par exemple, sera plus particulièrement mis en valeur avec les mots importants en gras, de grande taille et cliquables, etc.). Cette page proposera également, comme toutes les pages du site, une barre de navigation permettant au visiteur de cliquer sur les autres rubriques.

Avantages des pages de contenus :

- Bonne pérennité de la solution puisqu'il s'agit là de mettre à la disposition de l'internaute un contenu de qualité, adapté à la requête demandée. Difficile de penser que les moteurs voient ce type de possibilité d'un mauvais œil à l'avenir...

Inconvénients des pages de contenus :

- Cela demande un vrai travail éditorial dont la mise en œuvre peut s'avérer longue, voire onéreuse.
- Les options techniques retenues par le client pour la charte graphique de son site peut, dans certains cas, être un obstacle au référencement (Flash, menus en JavaScript) et augmenter le travail d'optimisation.
- Le client doit pouvoir contrôler du mieux possible le contenu mis en ligne par le référenceur.

La création de pages de contenus est une solution assez peu répandue, finalement, mais qui permet d'obtenir très souvent de bons résultats pour un rapport qualité/prix intéressant. Elle demande cependant un travail important et un contrôle strict de la part du client au sujet des informations mises en ligne par le référenceur.

5. L'URL Rewriting

Il s'agit ici de la solution qui est certainement la plus efficace pour un site dynamique. Le but est de définir des règles de réécriture pour des adresses de pages web dynamiques, puisque ce sont elles qui, dans la plupart des cas, bloquent les moteurs.

Exemple : des urls du type :

<http://www.sitedynamique.com/prog.cgi?kw=motcle&langue=fr&zone=france&encodage=ISO-8859-1>

Pourront être réécrites en :

http://www.sitedynamique.com/prog.cgi_motcle_fr_france_ISO-8859-1

Cette adresse ne pose plus de problèmes à un moteur de recherche. Le tour est joué et le site web dynamique devient indexable sans soucis par les robots.

Mettre en place une politique d'URL Rewriting ne propose pas de difficultés majeures, même si cela peut prendre du temps et demande surtout beaucoup de rigueur organisationnelle. Mais, une fois que cela est fait, vous n'avez plus à vous en occuper, sauf modification majeure dans la structure de votre base de données.

Avantages de l'URL rewriting :

- Permet de rendre compatible avec les moteurs de recherche bon nombre de sites web dynamiques.
- Règles de réécriture d'adresses établies une fois pour toutes.
- Pris en compte par tous les moteurs de recherche.

Inconvénients de l'URL rewriting :

- La mise en place des règles de réécriture peut être assez longue et fastidieuse.
- Certaines configurations techniques (serveurs / solutions propriétaires) ne proposent pas de telle solution.

L'URL Rewriting étant une composante essentielle d'un bon référencement sur les moteurs de recherche de sites web dynamiques, nous lui consacrerons toute la troisième partie de cet article, le mois prochain. Un peu de patience, donc, pour en savoir plus à ce sujet...

6. Optimisation des pages non-dynamiques

Une solution que l'on oublie souvent consiste à optimiser uniquement les pages web d'un site ne représentant pas d'obstacle pour les moteurs : page d'accueil, plan du site, présentation de la société, etc. Toutes ces pages n'ont pas, *a priori*, besoin de passage de paramètres dans leurs adresses et il devrait être possible de faire en sorte qu'elles soient disponibles grâce à des URLs "propres" (sans "?" ni "&" ni identifiant de session, etc.). Si ce n'est pas le cas par défaut, peut-être cela vaut-il la chandelle de faire en sorte de simplifier les adresses de ces documents, ce qui les rendra immédiatement indexables par les moteurs.

Peut-être que, pour certains sites, vous pouvez faire l'impasse sur l'indexation des pages web purement dynamiques, c'est-à-dire affichant des informations issues d'une base de données, et de ne prendre en compte, pour le référencement, que des pages statiques du site. C'est, en tout cas, une solution très économique en temps... et en argent, même si elle n'est clairement pas la plus optimisée !

Avantages de l'optimisation des pages statiques :

- Peut être rapide et peu onéreux.

Inconvénients de l'optimisation des pages statiques :

- Il n'est pas toujours techniquement possible de rendre statique les pages "connexes" d'un site.

- Cette solution ne permet pas de référencer les pages utilisant des informations issues d'une base de données, comme un catalogue produits par exemple.

7. Offres de référencement payant et publicitaire

Dernière solution : prendre en compte les offres XML ("autrement appelées "référencement payant", "Trusted Feed", "Paid Inclusion", "Feed XML"...) de moteurs comme Voila/Wanadoo et Yahoo!/Overture, qui vous permettent, contre rémunération, d'indexer sans soucis vos pages web dynamiques.

Vous avez, alors, une garantie d'indexation et de rafraîchissement (toutes les 48 heures le plus souvent), mais pas de garantie de positionnement (même si cela reste à démontrer...). Ask Jeeves, qui proposait une solution de ce type, est en train de l'abandonner. Google, pour sa part, n'en propose pas et clame haut et fort qu'il ne mettra jamais ce type de solution à son catalogue.

- Voila/Wanadoo ne travaille pas (encore) en direct avec les éditeurs de sites et il vous faudra donc passer par une société de référencement "agrée" (la plupart des grandes sociétés de référencement françaises le sont).

- Pour ce qui est de Yahoo!/Overture, l'offre ne semble pour l'instant disponible qu'au travers des Etats-Unis (<http://www.content.overture.com/d/USm/ays/sm.jhtml>). Il vous faudra donc contacter directement Yahoo! outre-Atlantique ou passer par une société française ayant passé un contrat avec Yahoo! US. Une solution "régionalisée" en Europe de cette offre (baptisée "SiteMatch") est attendue à la rentrée mais, pour l'instant, rien ne dit qu'elle verra le jour. Wait and see...

Avantages de l'offre de référencement payant :

- Rapidité.
- Indexation d'à peu près n'importe quel type de page.

Inconvénients de l'offre de référencement payant :

- C'est payant :-)
- Cela peut s'avérer rapidement onéreux en fonction du nombre de pages à indexer.
- Les modèles économiques proposés (CPM, CPC) sont différents d'une offre à l'autre et ne permettent pas toujours de définir un budget précis sur une durée donnée.
- Une fois que vous ne payez plus, vous n'avez aucune garantie que les pages en question restent dans l'index.

Enfin, il vous restera toujours la solution des liens sponsorisés Espotting, Google et Overture, entre autres, pour mener à bien vos campagnes publicitaires. Nous ne citerons cette solution que rapidement dans le cadre de cet article, puisqu'il ne s'agit pas ici d'un pur référencement...

Conclusion

Nous avons vu, dans les pages précédentes, sept méthodes proposées par les sociétés de référencement pour mieux indexer les sites web dynamiques. Bien entendu, plusieurs d'entre elles peuvent être couplées, mixées. C'est d'ailleurs ce qui arrive souvent.

Mais il nous semble essentiel de dire, en conclusion de cette deuxième partie, que la solution idéale n'a pas changé à nos yeux, que ce soit pour des sites statiques ou dynamiques : il s'agit de l'optimisation de façon native, de votre site web lui-même. Donc, si vous désirez obtenir les

meilleurs résultats et la plus grande pérennité pour votre référencement, des solutions comme celle de l'URL rewriting, couplée à une bonne optimisation des pages web de votre site, nous semblent parmi les plus efficaces.

Dans le cas contraire, si vous ne désirez pas toucher à votre site (ce qui est, rappelons-le, bien dommage...) ou si vous ne pouvez pas le faire pour des raisons techniques, vous pourrez toujours mettre en place des "rustines" comme des pages satellites ou des pages de contenus ou recourir à la solution de copie du site sur des serveurs distants. Tout cela fonctionne, pas de soucis, mais coûte également parfois assez cher.

D'autre part, toutes les techniques utilisées aujourd'hui - pages satellites, cloaking - restent suspendues à l'épée de Damoclès des politiques anti-spam des moteurs de recherche. Jusqu'où peut-on aller dans l'optimisation d'une page satellite ? Où faut-il s'arrêter ? Quelles techniques sont autorisées et quelles autres sont interdites ? Le flou actuel, dans ce domaine, n'aide pas y s'y retrouver facilement et les moteurs entretiennent, malheureusement, parfois ce flou artistique.

Peut-être que la solution, une fois de plus, viendrait dans la prise en compte du référencement du site lors de sa conception et de sa création, et non pas, de façon trop tardive, lorsqu'il est en ligne comme l'immense majorité de sites actuels, ce qui est bien dommage... Car que de problèmes seraient évités si cela était le cas...

Notre troisième et dernier article, au sujet du référencement des sites web dynamiques, sera entièrement consacré à la technique dite d'"url rewriting", rapidement exposée dans l'article de ce mois.

Au mois prochain donc...