

Plagiat et référencement : comment détecter les problèmes ? Quels sont les recours ? (1ère partie)

[Retour au sommaire de la lettre](#)

Plusieurs éditeurs de sites web se sont aperçus, dans le passé, que des webmasters inconvenants copiaient purement et simplement leur contenu en pensant bénéficier d'un bon référencement obtenu par le site "source" parfois après de longs mois, voire années, de travail. Ce type de copie illicite s'effectuant bien entendu sans demander la moindre autorisation. Comment détecter ces plagiat ? Comment en découvrir les auteurs ? Comment faire pour stopper ces agissements ? Cet article tente d'apporter quelques éclairages à ces questions cruciales...

Préambule

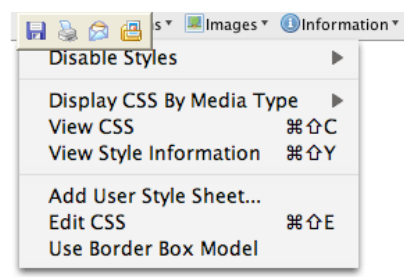
Il pourrait être facile de considérer Internet comme une zone de non-droit. Or, comme tout autre média, il est soumis au code de la propriété intellectuelle. Lorsqu'un site se lance dans le référencement, il peut lui arriver de s'inspirer de ce que font les concurrents. Pour autant, s'inspirer ne signifie pas copier exactement ce qu'ils font ! Car dans ce cas il s'agit de plagiat, ce qui est répréhensible et passible de poursuites judiciaires !

Outre le préjudice moral et financier, le plagiat peut engendrer un préjudice de visibilité au sein des moteurs de recherche pour votre site, entraînant au mieux une perte de résultats, au pire la disparition de votre site au profit de l'url incriminée (phénomène de "duplicate content").

Il suffit d'avoir une forte notoriété publique, un Page Rank avec un excellent indice (au-delà de 6) et une visibilité dans les moteurs de recherche sur des requêtes stratégiques qui renvoient plusieurs millions de résultats pour que vous vous attiriez certainement des convoitises ! Il faut alors redoubler d'attentions et discerner avant tout plusieurs points :

- Comment détecter le plagiat ?
- Comment découvrir les auteurs du plagiat ?
- Comment faire pour stopper les agissements des contrevenants ?

Avant d'entamer toute investigation, munissez-vous d'outils qui vont faciliter vos recherches. Un navigateur comme Firefox vous aidera notamment dans vos analyses. Ce navigateur très performant présente l'avantage d'avoir des composants très utiles pour le référencement. Certains permettent (entre autres) de désactiver Javascript ou les feuilles de Style d'une page web... Un véritable « couteau suisse » à l'usage des webmasters. Vous pourrez télécharger ce plug-in à l'adresse suivante : <http://chrispederick.com/work/webdeveloper/>



Avec cet outil, vous pourrez afficher tout contenu susceptible d'être caché et couper toute redirection utilisée avec le Javascript. Dès lors, une page « mise à nue » dévoile parfois bien des secrets...

Le cas échéant, vous aurez également besoin d'analyser l'arborescence complète d'un site web. Dans ces cas là, un logiciel « aspirateur de site » comme HT Tracks (<http://www.01net.com/windows/Internet/aspirateur/fiches/1911.html>) peut être très utile. Il va vous permettre de consulter et de créer la structure complète du site à analyser avec ses répertoires et sous-répertoires.

Comment détecter le plagiat ?

1ère solution : De nombreux outils statistiques comme Xiti (<http://www.xiti.com/>), eStat (<http://www.estat.com/>) ou Decisive Insight (<http://www.di.nextedia.com/>) et bien d'autres donnent de nombreuses informations sur la visibilité, le trafic, la transformation au niveau des actions (ventes, inscription à une newsletter...). Ils permettent de connaître les mots qui ont été utilisés par les internautes pour accéder à votre site internet *via* les moteurs de recherche. Faites attention à bien analyser en profondeur les requêtes utilisées par les internautes pour trouver votre site Internet !

Si vous rencontrez des mots clés suspects (mots clés incohérents par rapport à votre contenu, mots clés associant une marque concurrente avec un mot clé générique de type...), identifiez immédiatement la provenance. S'il s'agit de référencement naturel, analysez la provenance et faites une recherche sur le mot clé sur le moteur concerné pour connaître la page qui ressort dans les résultats.

Par exemple, éditeur d'un site de voyages, vous avez comme mot clé d'accès « maison de retraite » ou « partir en retraite » parmi les mots clés les plus générateurs de trafic en provenance de Google. Dès lors, il est possible d'analyser très rapidement la source du problème. Sur la requête « maison de retraite », vous chercherez les sites qui ressortent en recherche naturelle et en liens sponsorisés (LS). Rien d'anormal en ce qui concerne les LS mais dans la recherche naturelle, vous vous arrêtez sur l'un d'entre eux car vous vous apercevez que le site a entièrement recopié l'environnement de votre charte graphique : les menus sont rigoureusement identiques et le logo de votre marque a été contrefait. Mieux ! En étudiant le code source de la page, vous vous apercevez que le site incriminé a tout simplement intégré vos propres tags de votre outil statistique (Xiti ou autre) sur ses pages. CQFD. Il est temps de prendre contact avec le site concerné...

2ème solution : Il existe de nombreuses commandes utiles pour le référencement. La plus utilisée est la commande « site : » valable pour les principaux moteurs de recherche (Yahoo, Google, Live Search). Cette commande permet de connaître le nombre de pages qui ont été indexées par le moteur de recherche. Attention aux chiffres / statistiques que vous pourrez obtenir. Ils sont souvent fournis à titre indicatif mais donnent un bon aperçu de l'état d'indexation d'un site. Vérifiez avant tout que les urls ont été bien indexées et qu'un contenu (appelé snippet dans le langage du référencement) soit bien associé à chaque url.

Si vous constatez que certaines urls commencent à être filtrées par Google (certaines urls n'affichent plus de contenu), alors qu'elles étaient en pleine forme il y a encore quelques semaines, vous êtes en droit de vous poser quelques questions. Vous pouvez certainement être victime de *duplicate content* (contenu dupliqué). Le tout est de s'en assurer et d'en connaître la raison.

Prenez quelques exemples d'urls incriminées et testez leur contenu sur le moteur de recherche choisi. Pour ce faire, copiez un bloc de texte extrait de vos pages et collez le sur l'interface utilisateur du moteur tout en veillant bien à utiliser les guillemets. Le moteur retracera toutes les pages web contenant l'expression exacte. Analysez le résultat en regardant quels sont les noms de domaine qui ressortent. N'hésitez pas à vous rendre sur les sites « suspects » et à analyser leur contenu. Cette technique est aussi très utile lorsque l'internaute souhaite détecter si le contenu d'un site est original ou repris d'un site déjà existant.

Voici en exemple une expression présente sur la page d'accueil du site de Nextedia (<http://www.nextedia.com/>) : "En prenant chaque jour une part plus importante de la consommation média des français, Internet fait naître des modes de relation radicalement différents aux médias et à la publicité". En faisant une recherche sur Google (en ayant pris soin d'intégrer les guillemets), un site (autre que celui de Nextedia) ressort exactement sur cette expression.

Google

Rechercher dans : Web Pages francophones Pages : France

Rechercher "En prenant chaque jour une part plus importante de la consommation média des français, Internet fait naître des modes de relation radicalement différents aux médias et à la publicité.".

Images
Actualités
Groupes
plus >

Rechercher "En prenant chaque jour une part plus importante de la consommation média des français, Internet fait naître des modes de relation radicalement différents aux médias et à la publicité.".

Images
Actualités
Groupes

Rechercher

Recherche avancée
Préférences

Résultats 1 - 4

Podemus => politique <http://podemus.com/podemus.com-le-media-du-...>
Podemus => politique [http://podemus.com/podemus.com-le-media-du-podcast-audio-et-video Wed, 21 Feb 2007 02:48:07 +0000 podemus.com.fr.podemus.com-le-...](http://podemus.com/podemus.com-le-media-du-podcast-audio-et-video-Wed,21Feb200702:48:07+0000podemus.com.fr.podemus.com-le-...)
fr.podemus.com/tag/?rss=politique - 31k - [En cache](#) - [Pages similaires](#)

WebTV Decideurs.tv
t t t. NEXTEDIA Gael Duval 200207. En prenant chaque jour une part plus importante de la consommation média des français, Internet fait naître des modes de relation radicalement différents aux médias et à la publicité.".
www.decideurs.tv/ - 36k - 21 fév 2007 - [En cache](#) - [Pages similaires](#)

WebTV Decideurs.tv
Accuracy met à la disposition des dirigeants - Direction Générale, Direction Financière, Direction de la Stratégie et du Développement, Direction Juridique. ...
www.decideurs.tv/permalink/2497/accuracy-frederic-duponchel-210207.aspx - 37k - 21 fév 2007 - [En cache](#) - [Pages similaires](#)
[[Autres résultats, domaine www.decideurs.tv](#)]

Nextedia - Groupe indépendant de communication interactive
PRESENTATION. En prenant chaque jour une part plus importante de la consommation média des français, Internet fait naître des modes de relation radicalement différents aux médias et à la publicité.".
www.nextedia.com/ - 13k - [En cache](#) - [Pages similaires](#)

Pour limiter les résultats aux pages les plus pertinentes (total : 4), Google a ignoré certaines pages à contenu similaire.
Si vous le souhaitez, vous pouvez [relancer la recherche en incluant les pages ignorées](#).

8 résultats (0,18 secondes)

Le site suspect est un site média ayant interrogé Nextedia sur les problématiques de e-marketing. L'alarme aura donc été de courte durée !

3ème solution : Vérifiez régulièrement le positionnement des sites présents en première page des résultats sur vos requêtes stratégiques au sein des moteurs de recherche (surtout si vous-même y êtes !). Plus vous serez en haut des résultats, plus vous allez attirer des envieux sur vos bons positionnements. Il est donc crucial de savoir qui se positionne sur les premières pages de résultats et surtout avec quel type de contenu..

Quel que soit le degré d'investigation, si vous apercevez un *snippet* (résumé textuel des pages dans les résultats des moteurs) familier, n'hésitez pas à vous rendre sur le site. Vérifiez si aucune redirection n'est présente depuis le nom de domaine et analysez scrupuleusement son code source. Grâce à Firefox, vous pourrez d'office supprimer les feuilles de style pour voir si ces dernières n'étaient pas utilisées pour cacher du contenu.

Afin de retrouver les copies de vos pages, vous pouvez rechercher, à l'aide d'un moteur de recherche (Google.fr par exemple), les chaînes de caractères suivantes :

- Le titre de l'article.
- Le nom de l'auteur.
- Le corps du texte.
- Des expressions types de votre article.

Pour cela, il est vraiment intéressant d'utiliser les options avancées de google http://www.google.fr/advanced_search?hl=fr qui offrent toute une panoplie de personnalisation au niveau de la recherche.

Google Recherche avancée [Conseils de recherche](#) | [À propos de Google](#)

Pages contenant tous les mots suivants [] 10 résultats [v] Recherche Google
cette expression exacte []
au moins un des mots suivants []
aucun des mots suivants []

Langue Résultats pour les pages écrites en Toutes les langues [v]
Format de fichier Seulement [v] Limiter les résultats au format de fichier Tous formats [v]
Date Lister les pages Web mises à jour pendant la période spécifiée Date indifférente [v]
Emplacement Pages dans lesquelles le ou les termes figurent n'importe où dans la page [v]
n'importe où dans la page [v]
dans le titre de la page [v]
dans le corps de la page [v]
dans l'adresse de la page [v]
dans les liens de la page [v]
fr, etc. [Plus de détails...](#)

Domaines Seulement [v] afficher les pages du site ou du domaine

Droits d'utilisation Afficher les résultats sans te

SafeSearch Aucun filtrage Filtrer [SafeSearch](#)

Recherche de pages spécifiques

Similaires Rechercher les pages similaires à cette page [] Rechercher
par exemple : www.google.com/help.html

Liens Rechercher les pages liées à cette page [] Rechercher

Recherches sur un sujet précis

Nouveau ! [Google Recherche de codes](#) - Recherche de code source public

©2007 Google

4ème solution : Vos propres recherches peuvent prendre du temps et ne sont pas toujours efficaces. Cela nécessite donc de la patience de la part du Webmaster/responsable du site pour analyser, rechercher, vérifier. C'est à ce moment là que l'entraide et la solidarité sur Internet peuvent intervenir.

Tout l'intérêt réside alors à créer un espace d'échanges (forum de discussion, blog...), une communauté permettant aux internautes de communiquer/échanger avec vous.

Cet espace d'échanges permettra de fidéliser vos internautes. S'ils se sentent bien chez vous, ils vous le rendront et n'hésiteront pas à vous avertir s'ils ont constaté sur votre site le moindre problème et surtout s'il est victime de plagiat.

Exemple : fin 2006, un site d'information sur le référencement, très consulté, disposant d'un bon Page Rank (7) et ressortant sur des expressions clés de son secteur d'activité, voit le contenu d'une partie de son forum plagié. Un membre lance un topic en avertissant la communauté. Le webmaster est aussitôt prévenu et peut agir en conséquence.

5ème solution : utilisez un logiciel pour détecter le plagiat. Il en existe de nombreux sur le marché anglo-saxon mais malheureusement beaucoup moins sur le marché francophone. Nous pouvons citer au passage Copy Space disponible à l'adresse suivante : <http://www.copyscape.com>.

Comment connaître le plagieur ?

D'abord trouver une adresse mail/un contact sur le site. Ensuite lui envoyer un message.

Si aucun mail ne figure sur le site, vous pouvez identifier le propriétaire du nom de domaine (une recherche de type WHOIS permet rapidement de le trouver). Plusieurs outils sont à votre disposition sur le web dont Domain Tools disponible à l'adresse : <http://whois.domaintools.com/>

Ce site permet d'obtenir des informations précises sur la personne qui a réservé le nom de domaine et de connaître l'adresse complète (adresse postale, numéro de téléphone...) de la personne qui a enregistré le nom de domaine. Cette mine d'informations vous sera très utile pour prendre contact et entrer en communication avec le dépositaire du nom de domaine.

Que faire ensuite ?

Une entente à l'amiable est toujours possible. Un procès est souvent long et s'avère souvent coûteux. Dans la majorité des cas, la personne incriminée accepte de supprimer tout élément considéré comme illicite.

Si elle refuse, l'action en justice peut être envisagée. Il faut, avant tout, faire constater le problème par un huissier de justice avant d'entamer une procédure judiciaire. Attention, car la procédure peut être considérée comme nulle par le tribunal si l'huissier de justice ne prend pas en compte tout élément susceptible d'altérer ses constatations : il doit supprimer le cache, préciser si son ordinateur était connecté à un serveur proxy dépourvu de système de cache des pages visitées... Ces précautions lui permettront d'être le plus précis et le plus objectif possible lors du constat.

Intégrer du contenu dont on ne détient pas les droits peut constituer une contrefaçon passible d'engager la responsabilité civile du contrevenant. La contrefaçon peut aller jusqu'à 3 ans d'emprisonnement et 300 000 euros d'amende avec de lourdes pénalités par jour de retard. Sachez le, un webmaster averti en vaut toujours deux !

Comme vous avez pu le constater au cours de cet article, une reprise de contenu n'est pas forcément synonyme de plagiat. D'ailleurs, l'apparition des fils RSS a permis de développer cette pratique. Dès lors, existe-t-il donc une limite réelle entre plagiat et reprise de contenu ? Les sites partageant leur contenu avec d'autres sites via des fils RSS rencontrent-ils des problèmes de référencement ? C'est ce que nous développerons le mois prochain lors de la deuxième partie de ce dossier.

Damien Henckès, Consultant, Nextedia