

Domaine :	Recherche	Référencement
Niveau :	Pour tous	Avancé

Une affaire ayant opposé dernièrement les entreprises CNI et Luteciel, qui aurait pu nous amener de nombreuses informations si le dossier avait été mieux préparé par l'un des protagonistes, a cependant eu le mérite de mettre en avant la prise en compte du "duplicate content" par Google selon une vision judiciaire. Voici quelques éclaircissements à ce sujet...

Le référencement est un monde merveilleux qui met souvent en avant le meilleur de l'homme ou les pires comportements... Le tribunal de commerce de Paris a récemment eu à trancher un différend entre deux concurrents où l'un d'entre eux avait dupliqué le contenu du site de son concurrent pour que Google Panda le pénalise. Retour sur les faits...

Google Panda, les Conditions Générales de Google et le duplicate content

Abondance est probablement l'un des sites les plus renseignés et fournis sur le référencement en général et Google Panda en particulier. Pourtant, nous nous permettons de rappeler ici que Panda a notamment pour particularité de mettre en avant des sites de qualité et dont le contenu est de qualité. Cette notion de qualité est évidemment définie unilatéralement par Google qui a, par exemple, publié sur son blog 23 points afin de déterminer ce qu'est un site de qualité (<http://www.abondance.com/actualites/20110507-10869-google-23-questions-a-se-poser-pour-eviter-panda.html>).

Parmi ceux-ci, l'originalité du contenu est un élément important. En effet, un site reproduisant de manière identique ou quasi-identique un contenu d'un autre site est pénalisé par Google Panda. L'anecdote nous rapporte qu'au lancement de ce filtre de nettoyage par le moteur de recherche, plusieurs sites dont le contenu était original étaient moins bien « rankés » que ceux qui copiaient servilement le contenu...

L'organisation juridique de ces pratiques s'effectue par le biais des Conditions Générales de Google qui posent les règles générales d'utilisation de leurs services. La dernière version des Conditions Générales de Google est encore plus synthétique que les précédentes. En ce qui concerne les pratiques illicites, il est juste précisé dans ce document : « *N'utilisez pas nos Services de façon impropre. Ne tentez pas, par exemple, de produire des interférences avec nos Services ou d'y accéder en utilisant une méthode autre que l'interface et les instructions que nous mettons à votre disposition. Vous ne devez utiliser nos Services que dans le respect des lois en vigueur, y compris les lois et réglementations applicables concernant le contrôle des exportations et ré-exportations. Nous pouvons suspendre ou cesser la fourniture de nos Services si vous ne respectez pas les conditions ou règlements applicables, ou si nous examinons une suspicion d'utilisation impropre.* ». Aucune trace d'une référence à une quelconque interdiction à des pratiques de black SEO ou spamdexing ou autre pratique que le référencement sain révolse...

Ce document « juridique » peut être complété de certaines conditions spécifiques. Ainsi que le précisent les Conditions Générales, « *il se peut donc que des conditions additionnelles ou particulières à certains Services (p. ex. des conditions de limite d'âge) s'appliquent. Ces conditions additionnelles seront mises à votre disposition avec les Services concernés. Si vous choisissez d'utiliser ces Services, vous acceptez que ces conditions additionnelles fassent alors également partie de votre engagement contractuel avec nous* ».

Bien que l'identification de ces conditions additionnelles soit réellement compliquée (ce qui permettra à Google d'en invoquer (ou pas) facilement devant un tribunal, en cas de

besoin...), il est probable que les « conseils aux webmaster » (<http://support.google.com/webmasters/bin/answer.py?hl=fr&answer=35769>) en fassent partie. Ce document pose plusieurs principes relatifs à la qualité d'un site :

- « Concevez vos pages en pensant d'abord aux internautes et non aux moteurs de recherche.
- Ne trompez pas les utilisateurs.
- Évitez les "astuces" destinées à améliorer le classement sur les moteurs de recherche. Pour savoir si votre site Web respecte nos consignes, posez-vous simplement la question suivante : "Cela me dérangerait-il d'expliquer au propriétaire d'un site Web concurrent ou à un employé de Google quelles sont les solutions que j'ai adoptées ?". Vous pouvez également vous poser les questions suivantes : "Ces solutions sont-elles d'une aide quelconque pour les internautes ?" "Aurions-nous fait appel à ces techniques si les moteurs de recherche n'existaient pas ?"
- Pensez aux éléments qui rendent votre site Web unique et attrayant, et qui lui confèrent de la valeur. Faites-en sorte que votre site Web se distingue des sites concurrents dans votre secteur d'activité ».

Sans entrer dans le détail, les mêmes « conseils aux webmasters » précisent les pratiques considérées par Google comme répréhensibles. Ainsi que le précise ce document, Google est prêt à réagir s'ils détectent d'autres pratiques malhonnêtes qui ne sont pas répertoriées dans cette liste :

- « Évitez les techniques suivantes :
- Contenu généré automatiquement ;
 - Participer à des systèmes de lien ;
 - Techniques de dissimulation (cloaking) ;
 - Redirections trompeuses ;
 - Textes ou liens cachés ;
 - Pages satellite ;
 - Contenu détourné ;
 - Participer à des programmes d'affiliation qui n'apportent pas une valeur ajoutée suffisante ;
 - Multiplication de mots clés non pertinents sur les pages ;
 - Création de pages générant des opérations malveillantes, telles que des tentatives de phishing ou encore l'installation de virus, de chevaux de Troie ou d'autres programmes malveillants ;
 - Utilisation abusive des extraits enrichis balisés ;
 - Envoi de requêtes automatiques à Google. »

Ce document est complété par un autre document, les « Content Guidelines » (<http://support.google.com/webmasters/bin/topic.py?hl=fr&topic=2371375>) qui précise les instructions générales de Google en matière de contenu. A nouveau, nous pensons que ce document fait partie du champ contractuel et permet donc à Google d'opposer les « Content Guidelines » à toute personne devant un tribunal le cas échéant.

Ces « Content Guidelines » sont très précis quant au duplicate content. La page relative à cette pratique (<http://support.google.com/webmasters/bin/answer.py?hl=fr&hlrm=en&answer=66359>) précise :

- « Par contenu en double, on entend généralement des blocs de contenu importants, appartenant à un même domaine ou répartis sur plusieurs domaines, qui sont identiques ou sensiblement similaires. À l'origine, la plupart de ces contenus ne sont pas malveillants. Exemples de contenu non malveillant :
- forums de discussion pouvant générer à la fois des pages normales et des pages "raccourcies" destinées aux mobiles ;
 - articles en vente affichés ou liés via plusieurs URL distinctes ;
 - versions imprimables uniquement de pages Web.

Si votre site contient plusieurs pages dont le contenu est très similaire, vous pouvez indiquer à Google votre URL préférée de plusieurs façons. (C'est ce qu'on appelle la "canonicalisation".) ... Dans certains cas cependant, le contenu est délibérément dupliqué entre les domaines afin de manipuler le classement du site par les moteurs de recherche

ou d'augmenter le trafic. Ce type de pratique trompeuse peut avoir des effets négatifs sur la navigation de l'internaute qui voit quasiment le même contenu se répéter dans un ensemble de résultats de recherche.

Google s'efforce d'indexer et d'afficher des pages contenant des informations distinctes. Avec ce filtrage, si votre site propose, par exemple, une version "ordinaire" et une version "imprimable" de chaque article et qu'aucune de ces deux versions n'est bloquée par une balise Meta noindex, nous choisissons l'une des deux pour l'inclure dans la liste. Dans les rares cas où nous estimons que du contenu en double est peut-être affiché dans le but de manipuler nos classements et de tromper nos utilisateurs, nous apportons également les ajustements appropriés dans l'indexation et le classement des sites impliqués. En conséquence, le classement du site peut être affecté, ou le site peut être retiré définitivement de l'index Google, auquel cas il n'apparaîtra plus dans les résultats de recherche.

Les mesures suivantes vous permettent de résoudre les problèmes de contenu en double de manière proactive et de vous assurer que les visiteurs accèdent au contenu que vous souhaitez leur présenter :

- **Utilisez des redirections 301** : si vous avez restructuré votre site, utilisez des redirections 301 ("RedirectPermanent") dans votre fichier .htaccess pour rediriger efficacement les internautes, Googlebot et d'autres robots d'exploration. Pour ce faire, utilisez un fichier .htaccess dans Apache, et la console d'administration dans IIS.

- **Soyez cohérent** : soyez cohérent dans votre système de liens internes. Par exemple, n'établissez pas de lien vers <http://www.example.com/page/>, <http://www.example.com/page> ni <http://www.example.com/page/index.htm>.

- **Utilisez des domaines de premier niveau** : pour nous aider à présenter la version la plus appropriée d'un document, utilisez dans la mesure du possible des domaines de premier niveau pour gérer du contenu spécifique à un pays donné. Nous pouvons supposer que le site <http://www.example.de>, par exemple, présente du contenu destiné à l'Allemagne, par rapport à <http://www.example.com/de> ou <http://de.example.com>.

- **Soyez prudent en diffusant votre contenu** : si vous publiez votre contenu sur d'autres sites, Google affichera systématiquement la version jugée la plus appropriée pour les internautes pour chaque recherche donnée, qui pourra correspondre ou non à celle que vous préférez. Cependant, il est utile de s'assurer que chaque site sur lequel votre contenu est diffusé inclut un lien renvoyant vers votre article original. Vous pouvez également demander à ceux qui utilisent votre contenu diffusé d'utiliser la balise Meta noindex pour empêcher les moteurs de recherche d'indexer leur version du contenu.

- **Utilisez les Outils pour les webmasters** afin de nous indiquer le type d'indexation à privilégier pour votre site : vous pouvez spécifier à Google votre domaine favori (par exemple, <http://www.example.com> ou <http://example.com>).

- **Limitez les répétitions** : par exemple, au lieu d'inclure un avis de copyright au bas de chaque page, insérez un bref récapitulatif, puis établissez un lien vers une page plus détaillée. Par ailleurs, vous pouvez utiliser l'outil de gestion des paramètres pour définir la manière dont Google doit traiter les paramètres d'URL.

- **Évitez la publication de pages incomplètes** : les internautes n'apprécient pas les pages "vides", évitez donc les espaces réservés dans la mesure du possible. Par exemple, ne publiez pas de pages pour lesquelles vous ne disposez pas encore de contenu réel. Si vous créez des pages contenant des espaces réservés, utilisez la balise Meta noindex pour bloquer leur indexation.

- **Apprenez à maîtriser votre système de gestion du contenu** : vérifiez que vous maîtrisez l'affichage du contenu sur votre site Web. Les blogs, forums et systèmes associés affichent souvent le même contenu dans des formats divers. Par exemple, un message de blog peut apparaître sur la page d'accueil d'un blog, dans une page d'archives et dans une autre page avec le même libellé.

- **Limitez les contenus similaires** : si de nombreuses pages de votre site sont similaires, développez chacune d'entre elles ou regroupez-les pour n'en faire qu'une seule. Par exemple, si votre site de voyages présente des pages distinctes pour deux villes, mais que celles-ci comportent des informations identiques, regroupez les informations sur les deux villes sur une seule page ou développez chaque page afin qu'elles contiennent des informations bien différentes.

(...)

Le contenu en double n'entraîne pas de conséquences négatives particulières pour votre site sauf si l'objectif semble être de tromper et de manipuler les résultats des moteurs de recherche. En cas de problèmes de contenu en double, et si vous ne suivez pas les recommandations indiquées ci-dessus, nous nous chargeons de choisir la version du contenu à afficher dans nos résultats de recherche.

Cependant, si nous estimons qu'il s'agit de pratiques trompeuses et retirons en conséquence votre site de nos résultats de recherche, revoyez votre site. Si votre site a été retiré de nos résultats de recherche, consultez nos *Consignes aux webmasters* pour plus d'informations. Après avoir apporté les modifications nécessaires et vous être assuré que votre site est désormais conforme à nos recommandations, envoyez-le pour qu'il soit réexaminé.

Dans de rares cas, il est possible que notre algorithme choisisse une URL sur un site externe qui héberge votre contenu sans votre autorisation. Si vous pensez qu'un autre site ne respecte pas les règles relatives aux droits d'auteur, vous pouvez contacter l'hébergeur du site pour demander la suppression de votre contenu.».

Il est important de noter que les Internautes (et surtout les concurrents) peuvent « reporter » à Google toute pratique considérée comme illicite en envoyant un rapport de spam à l'adresse : <https://www.google.com/webmasters/tools/spamreport>.

En conséquence, Google a prévu, dans son champ contractuel opposable à tous, une partie relative au duplicate content qui lui permet de sanctionner tout site qui utilise cette pratique.

L'affaire CNI / Luteciel

Cette affaire aurait pu apporter plusieurs précisions aux professionnels du référencement que nous sommes, mais la société qui a porté le litige devant le tribunal n'a pas utilisée la meilleure voie et le tribunal n'a pas eu à statuer sur les points qui nous auraient intéressés... Les faits n'en sont pas moins intéressants... et cocasses.

La société CNI exploite le site internet « alibabuy.com ». Ce site est un moteur de recherche indexant différentes données provenant de voyagistes en ligne pour permettre aux internautes de comparer les offres des dits voyagistes.

La société Luteciel a créé le site internet « passager.com » lequel est un comparateur de prix de billets d'avions et d'hôtels.

Considérant que le site passager.com était une copie servile de son site alibabuy.com qui lui portait préjudice, CNI a assigné Luteciel devant le tribunal de commerce de Paris, en demandant de constater que les sociétés Luteciel et Viaticum se sont livrées et se livrent encore à des actes de concurrence déloyale caractérisés et ont engagé leur responsabilité, d'interdire à Luteciel l'exploitation du site passager.com tant que les contenus parasitaires et copies quasi serviles n'auront pas été supprimés et le versement d'une somme cumulée de 620.000€.

Luteciel s'est défendu en soulignant que CNI n'apporte pas la preuve de l'antériorité, et de la distinctivité des éléments du contenu du site « alibabuy.com » dont elle se prévaut. Pourtant, un constat d'huissier mettait en exergue que le site passager.com semble être une copie servile d'alibabuy.com. La copie du site alibabuy-com par passager.com semblait avoir pour objectif de nuire au référencement d'alibabuy.com par Google, compte tenu des « Content Guidelines » de Panda.

Il est important de souligner ici que CNI (ou son avocat ?) avait pris le parti de s'adresser au tribunal de commerce. Or, ce tribunal n'a pas le droit de statuer sur les problèmes de contrefaçon, copie, etc., seul le Tribunal de Grande Instance est compétent sur ce sujet. En conséquence, il était évidemment difficile de traiter le problème de la copie du site devant un juge... qui n'a pas le droit d'en entendre parler... Tout ce procès a donc été tourné principalement autour du comportement de référencement et non sur la copie elle-même du site.

C'est dans ce contexte que le tribunal a considéré que :

- CNI n'apporte pas d'éléments probants permettant d'établir une éventuelle confusion de la part des consommateurs entre les deux sites ;
- CNI indique elle-même à plusieurs reprises dans ses écritures que l'objectif de Luteciel n'est pas de développer une audience auprès des internautes mais de dégrader le référencement de ses concurrents pour rejeter la concurrence déloyale sur le terrain de la copie servile en elle-même.

En ce qui concerne le référencement et les règles de « duplicate content » de Google, le tribunal a relevé que :

- l'on doit donc considérer que **la mise en ligne d'un site internet, copie servile des codes informatiques d'un concurrent, dans le but de nuire à son référencement est un acte déloyal** même si son aspect visuel ne porte pas à confusion pour l'internaute moyen. Ceci est logique et n'est pas nouveau : le Tribunal de Grande Instance de Paris avait déjà, il y a 12 ans, conclu au même principe dans un jugement Panafieu.
- CNI n'a pas démontré que Luteciel a repris les codes informatiques du site alibabuy.com dans une proportion ne relevant pas de la simple similitude d'activité et dans une proportion suffisamment importante pour nuire à son référencement. En effet, CNI n'a pas pris le soin de contacter un réel expert en matière de référencement qui aurait pu rédiger un rapport de qualité démontrant la réalité des similitudes et surtout son caractère anormal au regard des contraintes de l'algorithme de Google ;
- Pis, CNI n'a même pas été en mesure de démontrer que le référencement de son site ait souffert des agissements décrits.

En conséquence, CNI n'a rien eu...

Il est dommage pour la communauté du référencement que CNI n'ait pas mieux préparé son dossier en allant devant un juge compétent et permettant d'apprécier les similarités et copies de sites et en faisant appel aux services d'un véritable expert du référencement. Nous aurions, aujourd'hui, un jugement probable riche d'enseignements et qui aurait apporté des précisions importantes aux professionnels. De plus, il est également désolant que Google n'ait pas été appelé à la cause, leurs conclusions auraient également été intéressantes pour tous les professionnels du référencement.

Parions que la nature humaine nous apportera très rapidement son lot de pratiques malveillantes et frauduleuses qui seront sanctionnés par de prochains jugements passionnants... Dont nous reparlerons donc très bientôt...

Alexandre Diehl

Avocat à la Cour, cabinet Lawint (<http://www.lawint.com/>)