

**Lettre professionnelle
Recherche et Référencement**

Numéro 160

Juin 2014



SARL RESEAU ABONDANCE - Olivier Andrieu - Consultant -

3 rue des Châteaux - 67140 Heiligenstein -

Tél : 03 88 08 83 26 - Fax : 03 88 08 83 27 -

E-mail : olivier@abondance.com - Web : www.abondance.com

Recherche et Référencement

Numéro 160 - Juin 2014

-- Lettre d'actualité mensuelle sur la recherche d'information et le référencement de sites Web --

Au sommaire ce mois-ci :

- > [Spam de contenu : comment le reconnaître ? Comment le détecter ?](#) (5 pages)
- > [Les services de Content Discovery \(2ème partie\)](#) (8 pages)
- > [Comment sortir d'une pénalité algorithmique Google ?](#) (8 pages)
- > [Panda 4.0 : anatomie de la dernière version du filtre de Google](#) (11 pages)
- > [Le droit à l'oubli enfin respecté par Google ?](#) (4 pages)

- > [Revue d'outils moteurs et SEO](#) (4 pages)
- > [Revue d'URL](#) (7 pages)

Lettre parue le lundi 16 juin 2014 (prochaine parution : Mardi 15 juillet 2014)

Le contenu de cette lettre est accessible sur la zone "Abonnés" du site Abondance, à l'adresse : <http://abonnes.abondance.com/>

La lettre "Recherche & Référencement" paraît aux alentours du 15 de chaque mois (un seul numéro pour les mois de juillet-août)
Pour tout renseignement : © Olivier Andrieu, olivier@abondance.com

IMPORTANT : Pour continuer de vous délivrer la meilleure information possible, la lettre "Recherche & Référencement" doit être utilisée au mieux par ses abonnés, qui doivent notamment se plier aux règles suivantes :

- * Il est **interdit d'envoyer, par mail** ou autres systèmes de transmission, un exemplaire de la lettre à quelque format que ce soit, à une personne qui ne serait pas abonnée.
- * Il est **interdit de diffuser à l'intérieur d'une entreprise, à plusieurs adresses e-mails ou sur un intranet**, le contenu de la lettre "Recherche & Référencement" si vous n'avez pas souscrit un **abonnement "Entreprise"** (<http://lettres.abondance.com/rechref.html>) qui est fait pour cela.
- * Il est **interdit de transmettre vos login / mot de passe** à des personnes qui ne seraient pas abonnées. Pour information, les adresses IP des machines se connectant sont relevées et tout abus pourrait être sanctionné si ce phénomène ne pouvait être justifié par vos services techniques.
- * Il est **interdit d'utiliser à des fins privées ou publiques tout contenu** de cette lettre sans notre autorisation expresse et écrite : envoi à des clients, des prospects, voire reprise intégrale du contenu en changeant le nom de la lettre et de la source (eh, oui, ça arrive...), etc.

De nombreux abus ont été constatés ces dernières semaines (merci à ceux qui nous les ont signalés) aussi, nous nous permettons de vous rappeler ces quelques règles de "savoir lire" :-)...

Merci pour votre compréhension et bonne lecture.

Toute l'information contenue de cette lettre est © Abondance / Olivier Andrieu. Toute reproduction ou distribution des informations, données et textes de cette lettre est interdite sans l'autorisation expresse du site Abondance. Toute reproduction ou représentation, intégrale ou partielle de ce contenu, faite sans le consentement de l'auteur, serait illicite. La loi n'autorise, que les copies ou reproductions réservées à l'usage privé du copiste et non destinées à l'utilisation collective, d'une part, et, d'autre part, que les analyses et les courtes citations dans un but d'exemple et d'illustration.

Sommaire du numéro 160 - Juin 2014

Spam de contenu : comment le reconnaître ? Comment le détecter ?	4
Les services de Content Discovery (2ème partie)	9
Comment sortir d'une pénalité algorithmique Google ?	17
Panda 4.0 : anatomie de la dernière version du filtre de Google	25
Le droit à l'oubli enfin respecté par Google ?	36
Revue d'Outils Moteurs et SEO	40
Revue d'URL	44

Spam de contenu : comment le reconnaître ? Comment le détecter ?

[Retour au sommaire de la lettre](#)

Domaine :	Recherche	Référencement
Niveau :	Pour tous	Avancé

De nombreuses méthodologies sont utilisées par les moteurs de recherche actuels pour détecter le spam dans leur index et produire ainsi des résultats de la meilleure qualité possible. Parmi celles-ci, la notion de "classifieur" est fondamentale. Cet article détaille une étude et un algorithme détaillés il y a quelques années à cette fin, ainsi qu'un projet actuellement en cours et qui permettra d'avoir, d'ici quelques mois, plus d'informations précises à ce sujet, ainsi que des outils pour définir le niveau de qualité d'une page web. Passionnant...

Nous allons aborder aujourd'hui la notion de spam de contenu, ce que l'on appelle souvent "webspam" dans la littérature scientifique. Sans aucune surprise, la détection du webspam est un enjeu important pour les moteurs de recherche. L'objectif de cet article est de définir ce qu'est le spam de contenu et de donner les clés permettant de comprendre comment un moteur va le détecter en utilisant des outils de classification.

Nous décrirons ensuite l'étude du projet que nous réalisons actuellement, et qui a pour objectif de mettre en évidence les critères constitutifs du spam de contenu, pour ensuite créer un outil de qualification de contenu.

Qu'est-ce que le spam de contenu ?

La formule, amusante mais convenue, nous le dit : « *webspam, you know it when you see it* ». Dans le folklore scientifique, il n'y a pas donc pas de vraie définition formelle du spam de contenu. Au final, on s'aperçoit généralement que le spam de contenu se traduit par une page web de faible qualité, dont le texte est réalisé au kilomètre (automatiquement ou à la main) et qui est "moche" (template de faible qualité graphique). Mais ceci n'est pas une réelle définition, et pour le chercheur, c'est la qualification par des quality raters qui fera foi : une page est du spam si elle a été notée « spam » par des êtres humains.

Et là apparaît un petit drame, car les humains ne sont pas toujours d'accord entre eux. Bien sûr, pour les pages contenant du texte généré par des méthodes automatiques de base, tout le monde est d'accord sur la qualité du contenu. Mais il existe de nombreux cas très *borderline*, et souvent, seuls les initiés qui connaissent les astuces y verront du spam, tandis que l'internaute lambda n'y trouvera rien à redire. C'est par exemple le cas des sites qui vont faire une copie de Wikipedia en y rajoutant quelques publicités.

Bref, la détermination de ce qui est du spam de contenu est réalisé par le monitoring du comportement humain. Ce monitoring peut être explicite ou implicite. Explicite quand on demande à des "quality raters" de noter des pages web, ou implicite quand le moteur va regarder le comportement de ses utilisateurs (retour "post click", temps passé sur une page, etc.).

Que faire quand un ensemble de pages ont été qualifiées ?

Avoir des pages qualifiées c'est bien, mais ce n'est pas un aboutissement acceptable pour un moteur de recherche. En effet, il est impossible de faire qualifier toutes les pages du web. Pour pallier ce problème, le moteur n'a d'autre choix que d'extrapoler la qualification de tout son index à partir de la qualification d'un petit ensemble de pages web.

Est-ce qu'il est possible, et facile, d'extrapoler ? La réponse est deux fois « oui ». C'est possible grâce aux techniques mises au point en apprentissage (le fameux *Machine Learning* qu'on retrouve à toutes les sauces dans le discours de Google, Facebook et consorts). Et c'est raisonnablement facile en utilisant des techniques issues de la classification supervisée, comme par exemple l'algorithme C4.5 de Ross Quinlan.

Qu'est ce que la classification supervisée ? Le problème de la classification est celui de faire la correspondance entre des objets en cours d'analyse et des catégories dans lesquelles on cherche à ranger ces objets. Un programme qui réalise cette identification s'appelle un "classifieur". On parle de classification supervisée quand la méthode de classification est initialisée avec des données qualifiées.

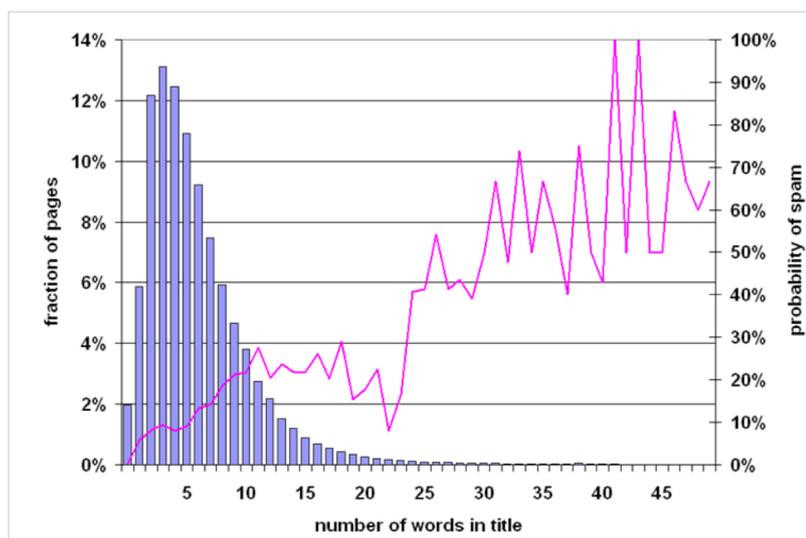
Pour revenir à notre question de base : une fois un ensemble de pages qualifié, on va calculer un certain nombre de statistiques sur des critères choisis de la page. Puis, grâce à l'algorithme C4.5 (ou un autre plus évolué), on va créer un classifieur. Dans notre cas, ce classifieur sera un arbre de décision. C'est-à-dire un ensemble de règles permettant de confronter dynamiquement une page en cours d'analyse avec certains critères (pas forcément tous, l'enjeu étant de diminuer le nombre moyen de critères utilisés pour chaque page), pour déterminer si une page est du spam ou du contenu de qualité.

Une fois ceci réalisé, on s'assurera que les résultats obtenus sont raisonnables, et notamment que l'on n'a pas trop de faux positifs. Un faux positif étant une page qui a été considérée comme du spam par le classifieur, alors qu'en fait elle n'en est pas.

Le cas d'école : l'étude de Ntoulas, Najork, Manasse et Fetterly

Nous allons maintenant vous parler d'une étude scientifique réalisée il y a plusieurs années (en 2006) et publiée (en partie) dans l'article de Ntoulas et al. (voir les sources à la fin de cet article). Cette étude est ancienne, et son seul intérêt de nos jours est qu'il s'agit d'un bel exemple méthodologique.

Les auteurs de cette étude ont utilisé un dataset de 18 000 pages qualifiées par des *quality raters*. Ils ont ensuite extrait les caractéristiques du dataset qui permettent selon eux de déterminer ce qui caractérise le spam et qui est détectable algorithmiquement. Il faut noter qu'ils ne se sont intéressés qu'à des critères *in text*. Par exemple, imaginons que l'on pense que le nombre de mots dans le <title> est discriminant pour le spam. Ce que l'on va faire pour tester cette hypothèse est de tracer la courbe suivante (issue de l'article de Ntoulas et al.) :

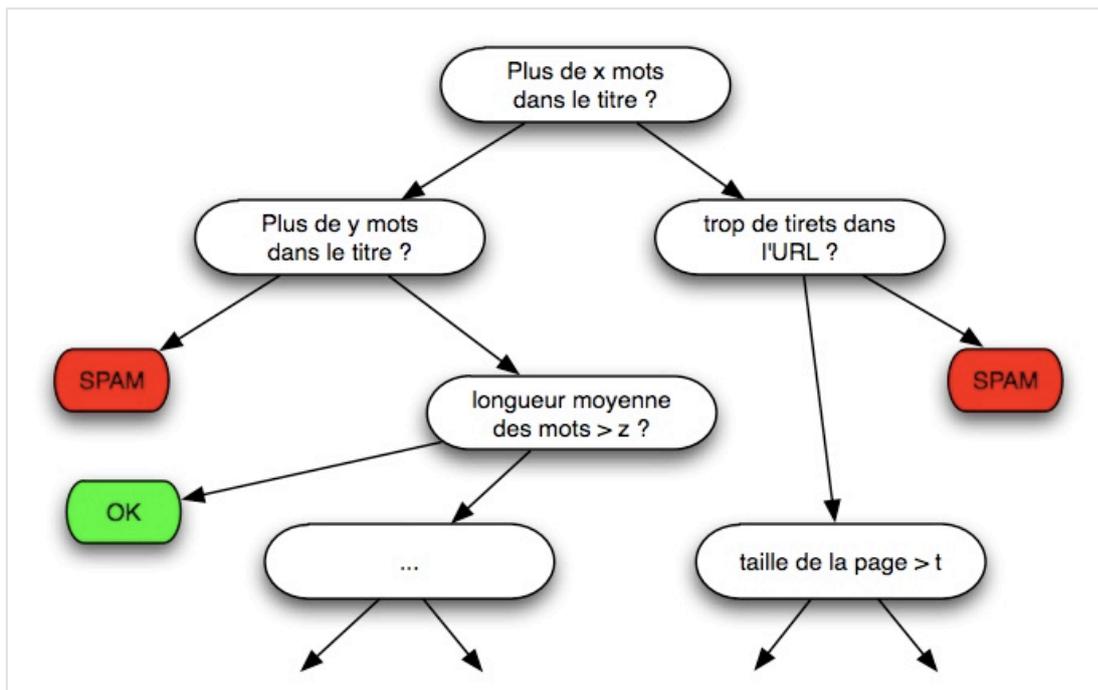


Cette courbe donne la proportion de pages qui contiennent un certain nombre de mots dans le <title>, corrélé (courbe violette) par la proportion de spam qui correspond. On voit en regardant cette figure que les pages qui ont un grand nombre de mots dans <title> ont une forte probabilité d'être du spam (par exemple probabilité 83% avec un <title> de 45 mots, alors que la probabilité d'être du spam est de 1% pour 5 mots dans le <title>).

Ntoulas et ses collègues ont ainsi étudié un grand nombre de critères, parmi lesquels :

- Taille de la page : quand un humain écrit, il fait une page ni trop courte, ni trop longue.
- Nombre de points, tirets et autres séparateurs dans l'URL : l'idée est qu'une page de spam a une URL générée automatiquement et qui contient le maximum de mots clés, avec des séparateurs entre eux.
- Longueur du nom de domaine : un long domaine est forcément louche car généralement on choisit un acronyme ou une courte suite de mots.
- Nombre de mots dans le titre : comme son nom l'indique.
- Fraction du texte qui est du texte d'ancre de lien : si tous les mots de la page sont des liens, on a sans doute une page de spam qui fait abusivement des liens.
- Longueur moyenne des mots : dans un texte standard, les mots apparaissent avec certaines probabilités, par conséquent les mots ont une longueur moyenne particulière. Un texte artificiellement généré est souvent déviant car la présence de certains mots est renforcée.
- Rapport de compression, rapport contenu/contenant : quelle est la taille de la page compressée par rapport à sa version normale ? Quelle est la proportion de code par rapport au réel contenu ?
- Vraisemblance d'indépendance : est ce que le texte de la page pourrait apparaître avec grande ou petite probabilité si il était généré par des singes tapant au hasard sur une machine à écrire ?

Il existe bien sûr de très nombreux autres critères, mais notre but ici n'est pas d'en faire une liste exhaustive. Une fois les critères mis au point, le travail est quasiment fini puisqu'il va suffire de passer les données obtenus dans l'algorithme C4.5 pour obtenir un arbre de décision efficace. Voici un exemple (totalement fantaisiste) d'arbre de décision (issu de notre blog) :



Avec le classifieur obtenu, Ntoulas et ses coauteurs annoncent un taux de reconnaissance du spam supérieur à 85%, et un taux de faux positif de 1,4%. 1,4%, cela a l'air d'être

très petit, mais c'est à remettre en perspective de la taille de l'index d'un moteur de recherche. Si l'index contient 20 000 milliards de pages de qualité, alors 280 milliards d'entre elles seront considérées (à tort !) comme du spam.

Et chère(e) lecteur/lectrice, il faut bien se rendre compte que Panda et Penguin sont deux filtres construits avec ce type de méthode ! Attention, nous avons écrit « avec ce type » car bien sûr, vu la quantité de données en jeu, Google doit utiliser des algorithmes plus efficaces pour générer un ou des arbres de décision. Et c'est incroyable, car dans les labos de Google, une équipe de chercheurs a écrit un article intitulé « *PLANET: Massively Parallel Learning of Tree Ensembles with MapReduce* » dont le premier auteur s'appelle Biswanath Panda !

2006, c'est vieux ! Et le futur ?

Vous en avez peut-être déjà entendu parler, mais nous sommes en cours de réalisation d'une étude similaire, à plus grande échelle et avec des critères réactualisés.

Cette étude sera intéressante pour les chercheurs en algorithmique du Web. En effet, une fois le jeu de données et les critères connus, on peut travailler sur des nouveaux algorithmes pour améliorer le domaine de la recherche d'information sur le web. Elle sera aussi intéressante, de manière évidente, pour les moteurs. Mais elle sera également une grande source d'information pour les référenceurs : avec les critères, vous savez ce que vous devez faire (enfin surtout, ne pas faire) pour ne pas être filtré. Avec le classifieur vous avez un outil qui vous dit si vous êtes dans la zone verte, orange ou rouge pour un moteur qui se préoccupe de la qualité perçue.

Par ailleurs, nous souhaitons obtenir des résultats qui vont au-delà du spam. En effet, la qualité d'une page web peut se noter de manière plus fine qu'au travers de la dualité spam/pas spam. Nous avons choisi d'avoir un niveau de qualité intermédiaire : "spam" versus "haute qualité" versus "faible qualité".

Enfin, nous espérons voir des nouvelles choses, car le web a beaucoup changé depuis 2006. Les webmasters ont de nouvelles manières de penser les sites web. Les filous ont des techniques nouvelles pour générer du contenu. Et enfin, les réseaux sociaux ont largement changé la donne au niveau de la publication des contenus sur le web.

Nos objectifs et notre méthodologie

Nos objectifs, annoncés dès le début, restent toujours les mêmes :

- Un jeu de données constitué d'environ 300 000 pages en langue anglaise, toutes taguées selon leur qualité (spam, haute qualité, faible qualité).
- Un jeu de données constitué d'environ 150 000 pages en langue française, toutes taguées selon leur qualité (spam, haute qualité, faible qualité).
- Deux rapports analysant ces jeux de données pour expliciter les critères constitutifs du spam, des contenus de haute qualité et des contenus de faible qualité.
- Deux classifieurs (un par langue) qui utilisent les critères pour décider automatiquement du niveau de qualité d'une page web fournie en entrée.

Notre méthodologie est similaire à celle de Ntoulas et al. Notre équipe (Jérôme Darbon, Thomas Largillier, Guillaume Peyronnet et Sylvain Peyronnet) est en cours de finalisation d'un crawl de plusieurs centaines de millions de pages sur le web pour créer une immense base de données d'URL. Le crawl a été (étonnamment) ce qui a été le plus complexe à mettre en place. Toutes les briques pour la suite ont été faciles à réaliser, mais le crawl a été une source constante de problèmes (matériels, logiciels, mais aussi conceptuels).

Une fois ce crawl terminé, nous tirerons uniformément des pages dans la base d'URL pour obtenir les deux *datasets* (anglais et français). Ensuite, nous utiliserons (à l'été 2014) le système déjà en place pour faire taguer par des êtres humains toutes les pages. On parle de 450 000 pages, qui doivent toutes être taguées entre 3 et 5 fois pour garantir les bonnes propriétés du résultat. Cela fait autour de 2 millions de tags à obtenir.

Ensuite, nous analyserons les *datasets*. Pour cela, il faudra lister tous les critères candidats, puis analyser toutes les pages au regard de chaque critère. En plus de tous les critères usuels, nous avons rassemblé des nouveaux critères candidats, avec l'aide de plusieurs acteurs des « méthodes agressives de référencement ». On trouve parmi ces nouveaux critères : le ratio entre nofollow et dofollow sur les liens sortants de chaque page, la position des liens sur la page (plus de liens dans le footer que plein texte par exemple), les signaux d'auteurs et les partages sociaux. On trouve aussi des choses plus ésotériques comme par exemple le résidu post-lemmatisation, qui caractérise la richesse de la langue du texte, ou encore l'analyse du TTFB (*Time To First Byte*, souvent assimilé à la vitesse d'affichage) de la page.

Une fois tous les critères analysés, la dernière tâche est de créer un classifieur avec les résultats obtenus. Cela demande une bonne puissance de calcul, du type de ce que fournit une grosse station de calcul. Encore une fois, nous irons plus loin que Ntoulas et ses collègues, notamment en testant plusieurs méthodes de classification supervisée. Notamment, nous utiliserons des méthodes qui pondèrent le coût des critères pour générer des arbres de décision « économiques ». En effet, il est préférable de tester d'abord les critères les moins coûteux à étudier, pour garder la puissance de calcul pour les pages les plus ambiguës.

Conclusion

Il n'existe pas vraiment de conclusion, car le meilleur est à venir, avec les résultats de cette étude, dont nous dévoilerons une partie (mais pas tout) ici même ;)

Sources

Quinlan, J. R. C4.5: *Programs for Machine Learning*. Morgan Kaufmann Publishers, 1993.
<https://www.google.fr/search?hl=fr&tbo=p&tbm=bks&q=isbn:1558602380>

Mehryar Mohri, Afshin Rostamizadeh, Ameet Talwalkar. *Foundations of Machine Learning*, The MIT Press, 2012.
<https://www.google.fr/search?hl=fr&tbo=p&tbm=bks&q=isbn:026201825X>

Alexandros Ntoulas, Marc Najork, Mark Manasse, Dennis Fetterly: *Detecting spam web pages through content analysis*. WWW 2006: 83-92. 2006.
<http://www2006.org/programme/item.php?id=3052>

Page de Biswanath Panda :
<http://research.google.com/pubs/author38399.html>

Biswanath Panda, Joshua S. Herbach, Sugato Basu, Roberto J. Bayardo. PLANET: *Massively Parallel Learning of Tree Ensembles with MapReduce*. Proceedings of the 35th International Conference on Very Large Data Bases (VLDB-2009).
<https://32mech.googlecode.com/files/vldb2009.pdf>

Sylvain Peyronnet, Professeur des Universités à l'Université de Caen Basse-Normandie (<http://sylvain.berbiqui.org/>) et **Guillaume Peyronnet**, gérant de Nalrem Médias (<http://www.gpeyronnet.fr/nalrem-medias.html>). Ensemble, ils font des formations (<http://www.peyronnet.eu/blog/masterclass-moteurs-seo/>) et essaient de battre les loutres à la pêche à la truite.

Les services de Content Discovery (2ème partie)

[Retour au sommaire de la lettre](#)

Domaine :	Recherche	Référencement
Niveau :	Pour tous	Avancé

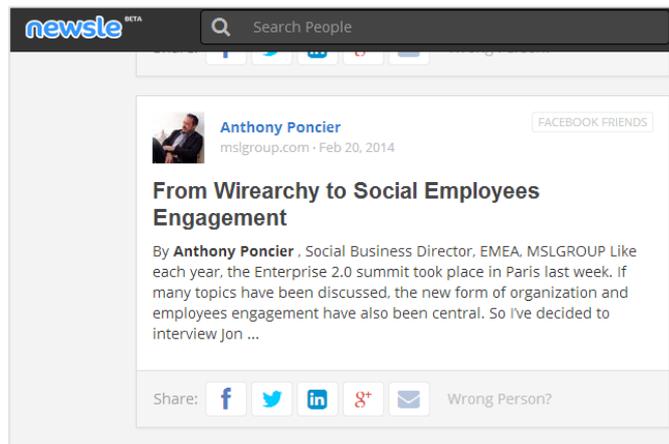
Nous sommes tous connectés à des réseaux sociaux comme Twitter ou Facebook, mais, parfois, leur utilisation à des fins de veille ou de community management est complexe, l'information étant souvent noyée dans la masse. C'est là qu'interviennent les outils de "content discovery" qui permettent, comme leur nom l'indique, de découvrir des contenus intéressants en traitant et filtrant les flux de données proposés par ces outils. Après l'exploration en profondeur de quatre d'entre eux le mois dernier, en voici trois nouveaux ce mois-ci, parmi les plus intéressants...

Depuis quelques années maintenant, l'animation de communautés sur le web ou « Community Management » (d'une page Facebook, d'un groupe LinkedIn, d'un compte Twitter) est devenue une profession à part entière. Si les interactions du community manager avec ses membres sous forme de réponses ou de discussions sont une des conditions de leur satisfaction, le fait d'apporter des contenus susceptibles de les intéresser sous forme d'articles, de vidéos ou de photos en est une autre. Encore faut-il identifier ce contenu. Les systèmes d'alertes par mots-clés de type Google Alertes jouent ce rôle depuis longtemps (pour d'autres solutions similaires voir ce tableau comparatif de 23 services : <http://socialcompare.com/fr/comparison/services-gratuits-d-alertes-par-mots-cles-1pp8mohd>) mais on voit arriver depuis quelques mois de nombreux autres services qui tentent d'« alimenter » le community manager (ou le veilleur bien sûr) en contenus pertinents en utilisant des techniques de repérage variées. Nous vous proposons d'en découvrir ici quelques-unes dans cette deuxième partie de notre article sur les outils de "content discovery"

Newsle

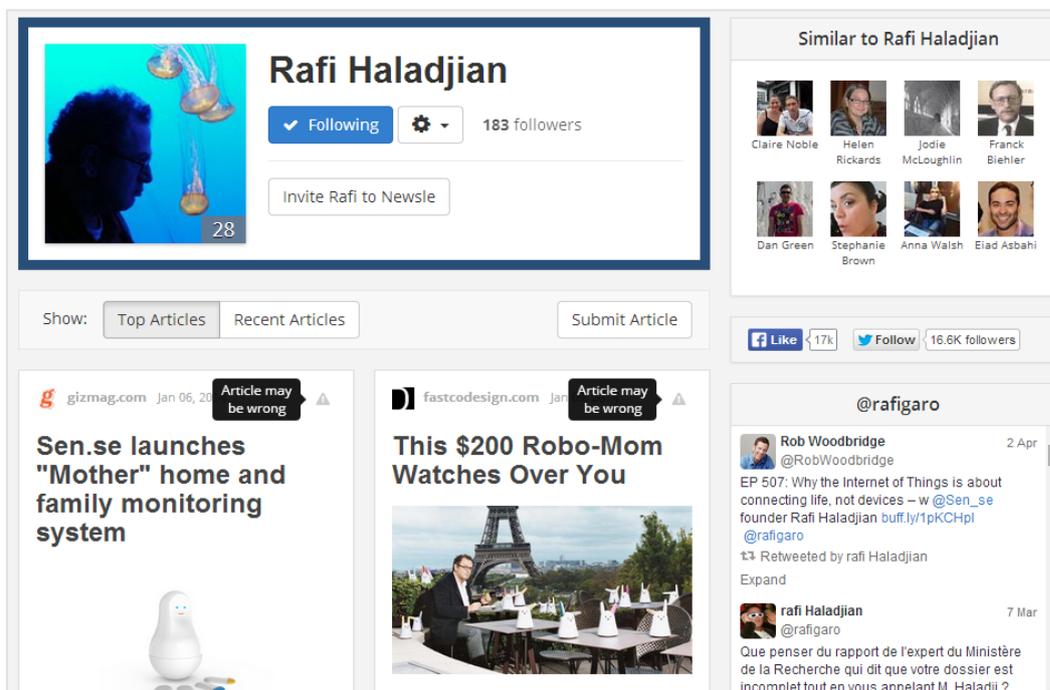
Newsle (<http://newsle.com/>) est un service qui a été lancé en janvier 2011, autant dire un dinosaure donc. A la différence des précédents, il n'intègre pas Twitter mais uniquement Facebook et LinkedIn. Une fois authentifié grâce à l'un de ces deux services, il faudra donc ajouter l'autre si l'on souhaite être complet.

A partir de là, le service va « scanner » des sources de presse et des blogs et détecter les noms de vos amis et contacts dont on parle pour les diffuser sur un dashboard de type « mur ». Il s'agit donc de « content discovery » d'un genre un peu particulier par rapport à ceux déjà étudiés puisqu'il est centré sur les personnes et non sur ce qu'elles partagent.



Il est possible de filtrer le flux en fonction de la source, LinkedIn ou Facebook donc, mais aussi sur des « Public figures », c'est-à-dire des personnalités (journalistes, politiques,...) identifiées comme tel par Newsle et qu'il met ainsi en avant.

Dans la partie droite de la page d'accueil, le service vous indique ceux qui, parmi vos contacts, sont les plus présents dans l'actualité. Il s'agit d'une fonctionnalité intéressante car en cliquant sur l'un d'eux, vous accédez à une page profil générée automatiquement par Newsle dans laquelle se trouvent agrégés les contenus (billets, articles, vidéos,...) dans lesquels ils sont cités.



On y trouve également les derniers tweets de la personne lorsque Newsle a pu identifier son compte. Ce profil est généré automatiquement sans que la personne soit elle-même inscrite au service et il est possible de s'y abonner (*Follow*) pour suivre les contenus qui la citeront à l'avenir. Il suffit que la personne existe en tant qu'ami ou contact d'un utilisateur de Newsle pour que le service génère sa page profil, ce qui ne va pas sans poser un certain nombre de questions quant à la maîtrise de son e-réputation personnelle et au respect de la vie privée. Ainsi, pour peu que vous soyez un minimum connectés (qui ne l'est pas parmi les lecteurs de cette lettre professionnelle ?), vous disposez probablement d'une page profil Newsle sans le savoir.

Dans le même ordre d'idées, il est possible, à partir de sa page d'accueil, de lancer une recherche de personnes. Vous obtenez en premier lieu vos contacts (*following*) mais également d'autres personnes connues du système dont vous pourrez éventuellement surveiller les profils. Newsle n'agit toutefois pas comme un moteur de recherche de profils puisqu'il n'affiche que 5 résultats.

Le « content discovery » n'est donc pas le seul objectif de ce service mais il peut être néanmoins configuré spécifiquement dans les « *settings* » via un système de newsletters.

Il est alors possible de recevoir 3 types d'alertes :

- Celles indiquant que vos contacts et amis sont cités dans l'actualité ;
- Celles indiquant que des journalistes identifiés par Newsle et que vous suivez ont publié un nouvel article ;
- Celles vous concernant.

Newsle Account Settings

Friend Alerts

Email me when my friends make the news

Email Frequency:

Daily Every 3 Days Weekly

Journalist Alerts

Email me when journalists I follow write new articles

Email Frequency:

Daily Every 3 Days Weekly

News About You

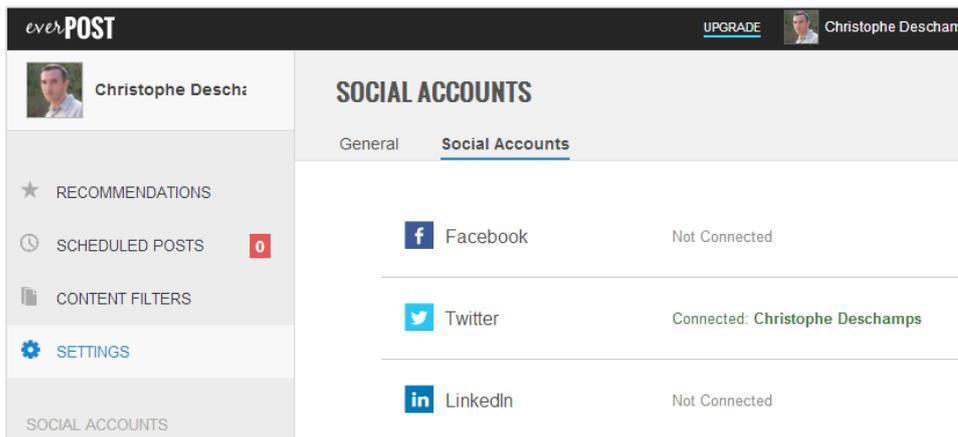
Email me when I make the news

[Advanced Settings](#) for news about me

Newsle est donc un service spécifique puisqu'il permet de faire remonter de l'information nominative et pourra en ce sens intéresser les veilleurs qui auraient à surveiller des personnes plus que des thématiques.

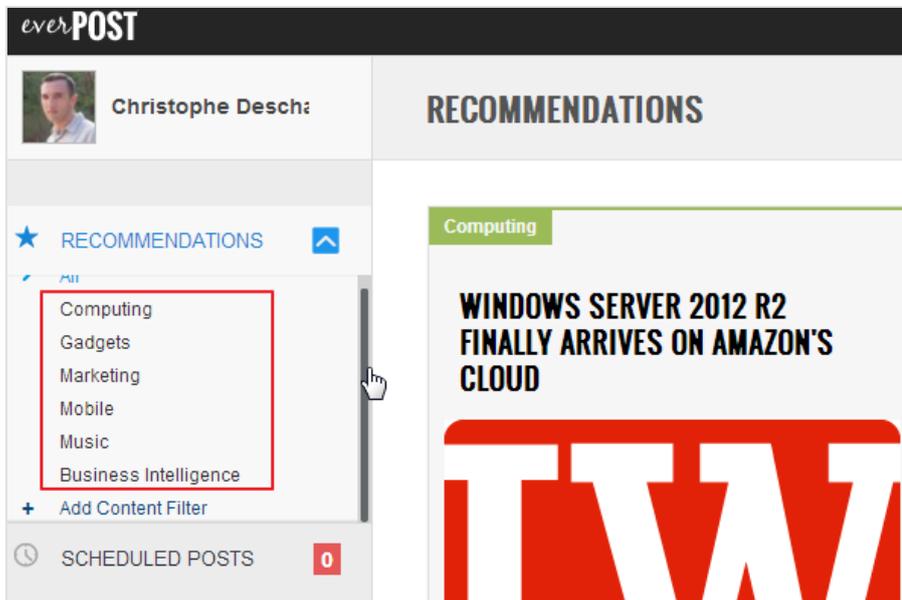
Everpost

Everpost (<http://everpost.co/>) est une startup israélienne qui a été lancée il y a un an. Sans surprise, on s'y connecte avec ses identifiants Twitter, Facebook ou LinkedIn et il est ensuite possible d'intégrer les deux services qui n'ont pas servi à l'authentification.



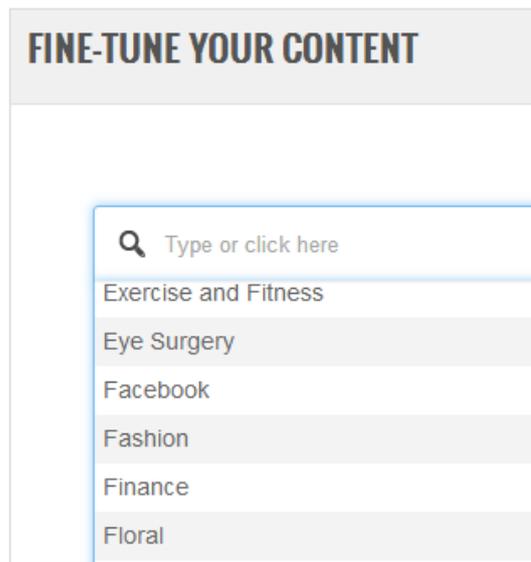
Everpost se veut un service de découverte de contenu et de curation. Il utilise donc vos comptes sociaux tant pour récupérer des items que pour vous permettre d'en programmer la rediffusion.

Sur la page d'accueil vous trouvez des recommandations qui ont été calculées après une analyse de vos profils (dont on n'aura pas le détail de fonctionnement). Ainsi dans notre cas Everpost identifie à partir de notre compte Twitter les 6 thématiques suivantes et propose des contenus relatifs :



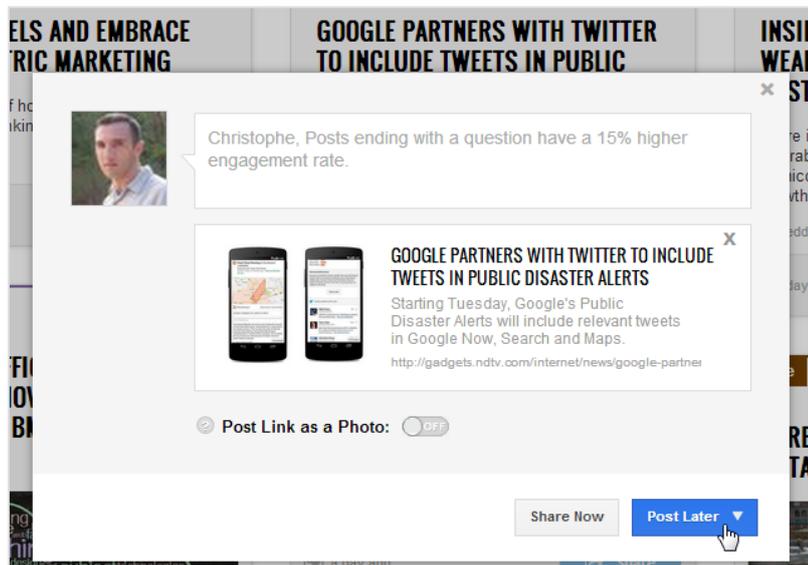
Ici donc, pas d'analyse des contenus partagés par vos contacts sur les réseaux sociaux, mais une mise en correspondance de thèmes avec des articles tirés de sites et blogs d'actualité.

Il est possible de personnaliser les recommandations en choisissant des « topics » à partir d'une liste existante.



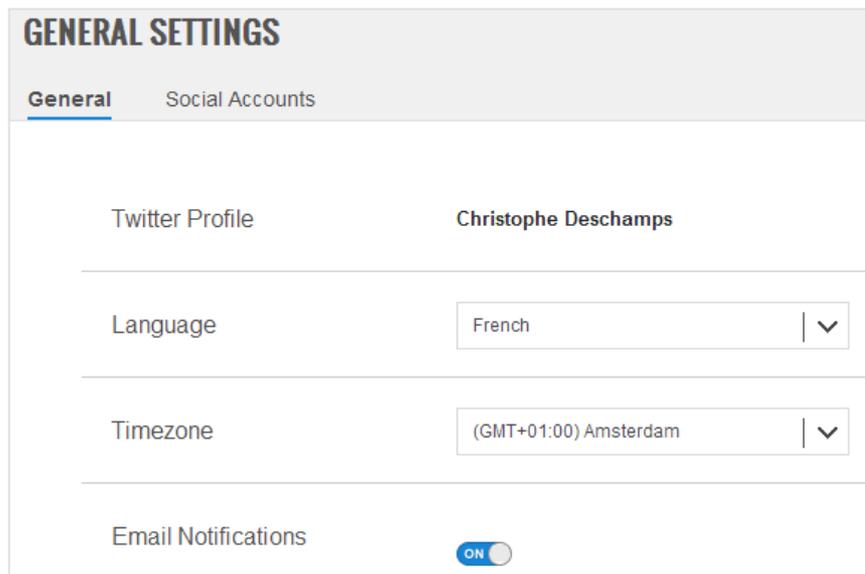
Mais impossible, en revanche, de faire de la détection de mots-clés.

Chaque recommandation peut ensuite être partagée directement ou programmée pour plus tard via une interface adaptée vous permettant d'ajouter votre commentaire à l'item.



Cette diffusion se fait bien entendu vers les services que vous avez choisi d'intégrer initialement.

La découverte de contenu peut se faire aussi via une notification par email que vous pouvez activer dans les fonctionnalités.



GENERAL SETTINGS

General Social Accounts

Twitter Profile **Christophe Deschamps**

Language French

Timezone ((GMT+01:00) Amsterdam)

Email Notifications **ON**

Au final un service simple qui ne nous semble pas apporter d'innovations suffisamment importantes en l'état pour lui permettre de se distinguer de services plus anciens. Notons tout de même que les versions payantes permettent la surveillance de mots-clés.

Trove

Trove (<http://trove.com/>) est justement l'un de ces outils anciens puisqu'il a été lancé en 2011 par le Washington Post en tant qu'agrégateur d'actualités issues de plus de 15 000 sources en ligne. Il a toutefois connu une renaissance en ce début d'année 2014 lorsque ce même journal a été racheté par Jeff Bezos, patron d'Amazon. A l'instar d'Everpost, il s'agit d'un service proposant à la fois du « content discovery » et de la curation.

Pour s'y connecter, on s'identifie via Twitter ou Facebook pour arriver sur une page où le service va générer des recommandations personnalisées basées sur :

- Les liens que vous tweetez ;
- Vos tweets favoris ;
- Les contenus qui passent sur votre timeline.



Home
Top news and the latest from Trove

Get updates from our editors:

Editors picked this in U.S. NEWS. 47 minutes ago
President Obama has not specified what time of assistance would be used, but he says the White House has not ruled anything out.

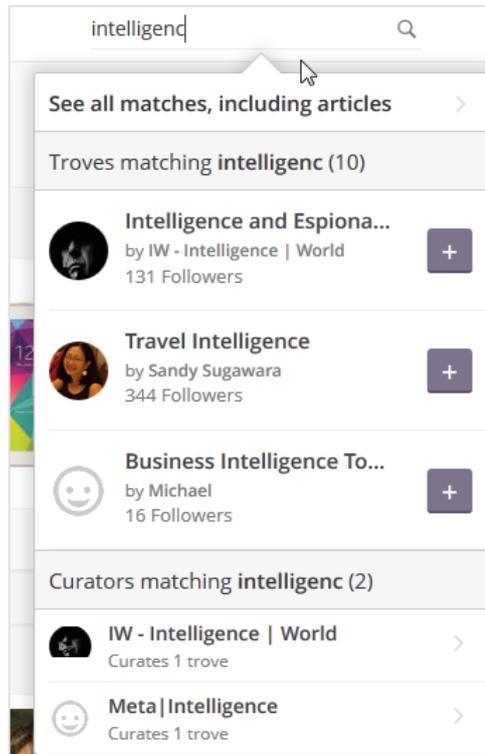
Obama: Iraq will need additional US assistance
WASHINGTON — President Barack Obama said Thursday Iraq will need more help from the United States as it seeks to push back a violent Islamic insurgency that has captured two key cities and is vowing to press toward Bagdad.

Generating recommendations based on:

- Links you've tweeted ✓
- Tweets you've favorited ✓
- Your Home Timeline ✓
- Checking for matches

In the News
Follow troves suggested by our editors.

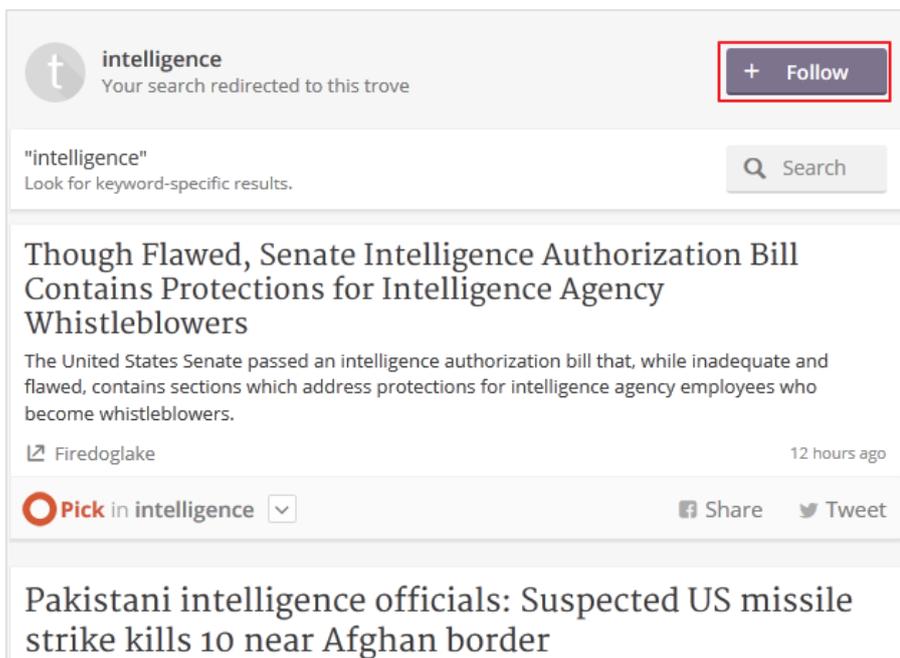
Pas de chance : durant nos tests, cette fonctionnalité semblait hors service. Ce sont ces sources auxquelles vous allez vous abonner et qui, dans la logique de ce service, vont ensuite servir à alimenter vos pages de curation baptisées « troves ». Nous avons néanmoins pu nous abonner à des flux génériques afin de poursuivre nos tests. Par ailleurs, il est possible d'ajouter manuellement des centres d'intérêt. Il faut pour cela utiliser la barre de recherche présente en haut de page.



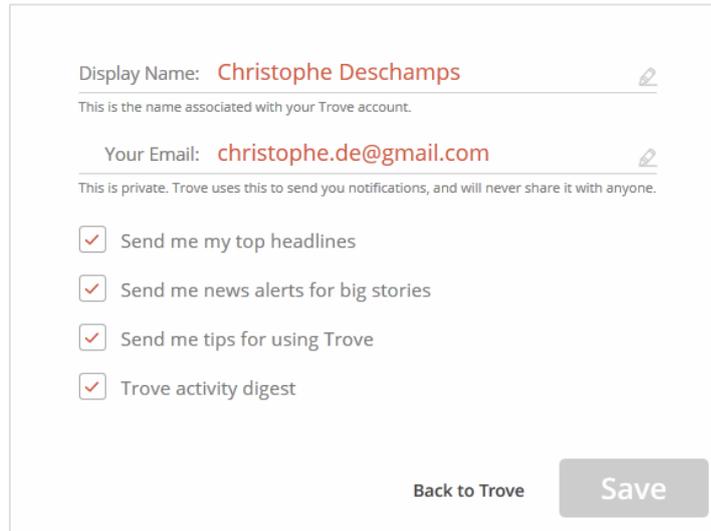
Deux types de recommandations sont proposés ici :

- Ajouter à sa page la page de curation d'un utilisateur de Trove afin de suivre ses découvertes.
- Visiter le profil des curateurs pour les ajouter comme amis dans le réseau social de Trove.

Ce moteur de recherche permet également de découvrir des articles de presse dans lesquels vos mots-clés sont inclus.



Par ailleurs, comme on le voit ci-dessus, chaque requête peut être intégrée à ses sources et vous fera donc remonter les articles à venir traitant de vos sujets, sur le modèle des alertes Google. Pour recevoir un email quotidien synthétisant ces nouveautés il suffit de se rendre dans les « Settings », en haut à droite de la page. Vous pourrez alors choisir quels types de contenus doivent ainsi remonter.



Display Name: **Christophe Deschamps** 

This is the name associated with your Trove account.

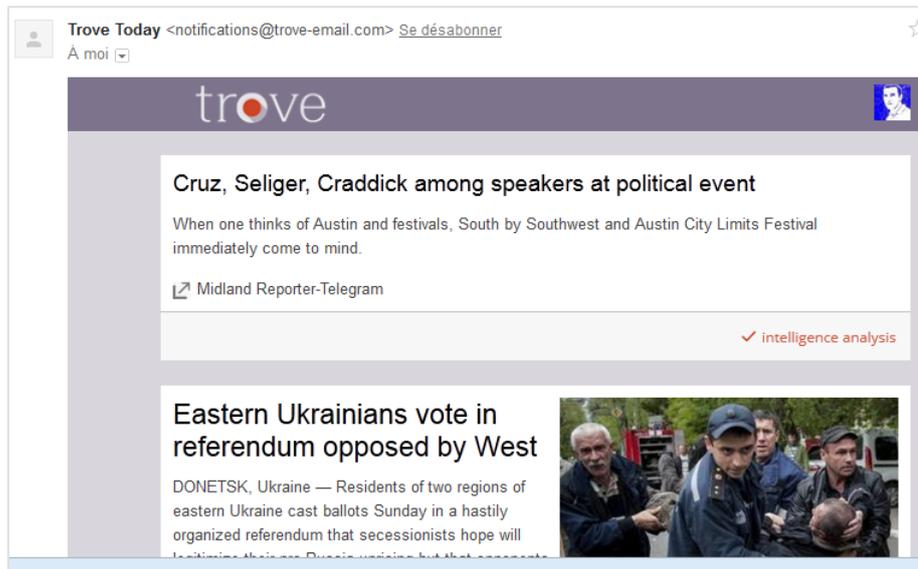
Your Email: **christophe.de@gmail.com** 

This is private. Trove uses this to send you notifications, and will never share it with anyone.

- Send me my top headlines
- Send me news alerts for big stories
- Send me tips for using Trove
- Trove activity digest

[Back to Trove](#) [Save](#)

Les synthèses reçues sont propres et bien formatées comme on peut le constater ci-dessous :



Malgré le désagrément initial (qui ne devrait pas durer) constaté sur Trove, il nous semble être un service très abouti et particulièrement efficace lorsqu'il s'agit de découvrir des contenus... en anglais.

Mais notre tour d'horizon des services de « content discovery » n'est pas terminé et peut-être aurons-nous de bonnes surprises le mois prochain !!

Christophe Deschamps
Consultant et formateur en gestion de l'information.
Responsable du blog Outils Froids (<http://www.outilsfroids.net/>)

Comment sortir d'une pénalité algorithmique Google ?

[Retour au sommaire de la lettre](#)

Domaine :	Recherche	Référencement
Niveau :	Pour tous	Avancé

Les pénalités algorithmiques qu'inflige Google peuvent faire peur : non seulement elles sont complexes à comprendre et à analyser, mais elles sont également difficiles à combattre et demandent parfois une patience infinie. Nous allons donc voir ici les différentes méthodes que l'on peut appliquer à un site web pour sortir des différentes pénalités du moment...

Peut-on sortir d'une pénalité algorithmique ?

C'est sans doute la première question à se poser : est-il réellement possible de sortir d'une pénalité algorithmique infligée par Google ? À cette question, les réponses sont multiples. Selon certains référenceurs et certaines agences, il serait quasiment impossible d'en sortir : une pénalité comme Google panda ou Google Penguin aurait un effet définitif sur le nom de domaine.

En réalité, il est tout à fait possible de sortir d'une pénalité algorithmique. La réelle problématique est qu'il faut mettre en place une procédure qui prend énormément de temps et d'énergie pour réussir. Et même lorsque vous allouez des moyens et un budget pour un tel travail, il est possible de ne sortir que partiellement de la pénalité et de ne jamais retrouver l'intégralité des positions et du trafic qui avaient été préalablement acquis par le site web.

Dans la lettre R&R du mois dernier (mai 2014), nous avons parlé des différents types de pénalité qui existent : <http://recherche-referencement.abondance.com/2014/05/comment-savoir-si-son-site-est-penalise.html>. La première étape consiste donc à savoir quelle est la pénalité que l'on doit combattre.

En effet, en fonction de l'algorithme qui vous touche, le travail de nettoyage et de désoptimisation sera complètement différent. Nous vous conseillons donc de relire l'article du mois dernier, mais aussi de consulter les changements d'algorithmes recensés par Moz sur cette page : <http://moz.com/google-algorithm-change>.

MOZ PRODUCTS LEARN COMMUNITY BLOGS ABOUT ? Q Login

Google Algorithm Change History

Each year, Google changes its search algorithm around 500-600 times. While most of these changes are minor, Google occasionally rolls out a "major" algorithmic update (such as Google Panda and Google Penguin) that affects search results in significant ways.

For search marketers, knowing the dates of these Google updates can help explain changes in rankings and organic website traffic, and ultimately improve search engine optimization. Below, we've listed the major algorithmic changes that have had the biggest impact on search.

1.5k Facebook 4,366 Twitter 2,215 LinkedIn 1,876 Share

2014 Updates

Panda 4.0 — May 19, 2014
Google confirmed a major Panda update that likely included both an algorithm update and a data refresh. Officially, about 7.5% of English-language queries were affected. While Matt Cutts said it began rolling out on 5/20, our data strongly suggests it started earlier.

[Google Begins Rolling Out Panda 4.0 Now \(SEL\)](#)

[Panda 4.0, Payday Loan 2.0 & eBay's Very Bad Day \(Moz\)](#)

Payday Loan 2.0 — May 16, 2014
Just prior to Panda 4.0, Google updated its "payday loan" algorithm, which targets especially spammy queries. The exact date of the roll-out was unclear (Google said "this past weekend" on 5/20), and the back-to-back updates made the details difficult to sort out.

[Official: Google Payday Loan Algorithm 2.0 Launched: Targets "Very Spammy Queries" \(SEL\)](#)

Unnamed Update — March 24, 2014

2014
2013
2012
2011
2010
2009
2008
2007
2006
2005
2004

Ce qu'il faut faire dans tout les cas de figure

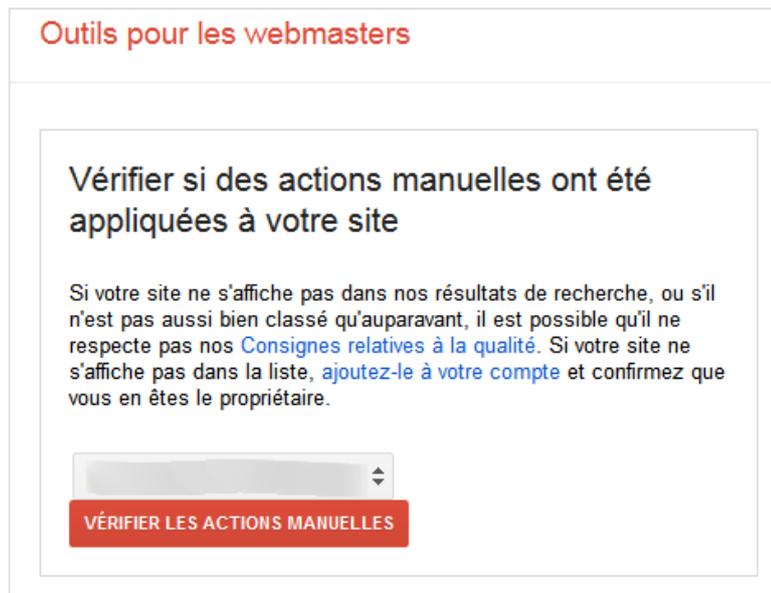
Quel que soit le filtre concerné, il existe plusieurs règles à suivre de manière systématique.

La première, c'est de faire continuer à vivre son site : ne vous arrêtez pas de publier des contenus ni d'être actif sur votre blog ou sur les réseaux sociaux. Continuez donc à créer de nouveaux contenus, à valider et à répondre aux commentaires et à animer votre présence sur les blogs, les forums et les réseaux sociaux.

La deuxième, c'est de continuer son travail de référencement. Vous devez continuer vos campagnes de netlinking, l'optimisation de vos contenus ou encore le partenariat avec d'autres sites web. Attention cependant, gardez en mémoire que c'est le travail que vous aviez fait au préalable qui a sûrement provoqué cette pénalité. Vous devrez donc analyser ce qui a fonctionné et ce qui n'a pas fonctionné pour adapter votre méthodologie dans votre futur travail de référencement naturel.

Ce n'est qu'ensuite que vous allez pouvoir travailler vos contenus et liens actuels pour sortir de la pénalité. Pour cela, referez-vous aux sections suivantes dans cet article en fonction de l'algorithme qui vous concerne.

Sachez qu'il est inutile de faire une demande de réexamen dans le centre webmaster de Google disponible ici : <https://www.google.com/webmasters/tools/reconsideration?hl=fr>
La raison à cela est simple : vous ne pourrez faire une telle demande que lorsque vous êtes touché par une pénalité manuelle.



La différence entre certains algorithmes

Une des choses à savoir est que les grands algorithmes de Google s'appliquent à un instant T. Le problème est qu'il faudra donc attendre la prochaine mise à jour de l'algorithme pour éventuellement sortir de la pénalité, surtout dans le cas de Google Penguin. Vous pouvez donc vous retrouver facilement dans le cas suivant : vous avez nettoyé vos contenus et vos liens, mais le site reste pénalisé pendant plusieurs mois encore.

C'est d'ailleurs une des raisons pour lesquelles il ne faut pas hésiter à agir en profondeur du premier coup. Par exemple, lors du nettoyage des liens, il faut le faire de manière complète et ne pas se cantonner à supprimer, modifier ou désavouer quelques liens et espérer que la pénalité soit retirée.

De plus, il n'existe pas de délais liés aux pénalités algorithmiques. Contrairement aux pénalités manuelles, le filtre sera actif tant que les problématiques initiales n'auront pas été corrigées. Attendre sans rien faire n'est donc absolument pas une solution viable. Source : <http://www.abondance.com/actualites/20140522-13932-matt-cutts-les-delais-dexpiration-penalites-manuelles.html>

Voyons maintenant ce qu'il est possible de faire, algorithme par algorithme.

Google Panda

Google Panda est sans doute un des algorithmes les plus simples à traiter : il cible en effet les sites ayant un contenu de faible qualité : les textes courts, les contenus dupliqués, le spam, les contenus sur-optimisés ou encore le contenu « spinné » (c'est-à-dire le contenu généré de manière automatique).

Si votre site est touché par ce type de pénalité, la procédure est extrêmement simple : vous devez travailler la qualité de l'intégralité de vos contenus.



Spam et contenu dupliqué

La première étape consiste à supprimer purement et simplement tous les contenus de spam et tous les contenus spinnés. Ils n'auront aucun intérêt pour le référencement, et libre à vous ensuite de renvoyer un code 410 à Google (« ce contenu n'existe plus ») ou de rediriger ces pages vers l'accueil de votre site avec une redirection 301.

La deuxième étape du travail doit vous amener à modifier l'ensemble des contenus dupliqués. Vous devrez soit les supprimer, soit les réécrire intégralement afin d'avoir un contenu unique et pertinent. L'idée est que Google ne puisse jamais trouver un même contenu à différents endroits de votre site web (en priorité), ou éventuellement sur d'autres sites web.

La qualité du contenu

La troisième étape est relativement simple : il faut améliorer tous vos contenus de faible qualité. Si vous avez des contenus extrêmement courts ou peu pertinents pour l'internaute, il sera extrêmement intéressant de les étoffer et de les améliorer pour augmenter la pertinence globale de votre site web.

Cela peut aussi passer par un remaniement complet de la structure de votre site, de vos catégories et de vos thématiques, ou encore par une amélioration drastique de l'ergonomie du site ou de son temps de chargement. On parlera d'ailleurs souvent de la notion d'expérience utilisateur et de taux de rebond, même si rien ne prouve que Google peut efficacement mesurer ces deux notions lors d'une pénalité Google Panda.

Le contenu sur-optimisé

Enfin, la quatrième étape va consister à dés-optimiser tous les contenus qui n'étaient pensés que pour le référencement naturel. Relisez donc l'ensemble de vos contenus et posez vous la question de savoir si chaque contenu est trop optimisé pour le SEO.

Pour le savoir, la méthode est relativement simple. Il suffit de se poser la question suivante : est-ce que lorsque vous lisez chaque contenu celui-ci vous paraît cohérent, écrit en bon français, pertinent, et surtout non redondant. Si par exemple vous trouvez le même mot clé toutes les deux phrases, il y a de fortes chances que ce contenu soit beaucoup trop optimisé pour les moteurs de recherche. Tout doit être analysé : vos contenus textes, vos images, vos liens ou encore vos vidéos, mais aussi toutes les parties communes de votre site : le *footer*, la ou les *sidebar(s)*, le menu principal...

La clé pour sortir de Google panda est donc simple : produire du contenu « naturel », pertinent et unique.

La complexité avec certains types de contenus

Dans certains cas de figure, vous pourrez avoir beaucoup de difficultés à contrer ce type de pénalité. C'est notamment le cas avec tous les contenus qui sont générés de manière automatique, semi-automatique ou par les utilisateurs.

Par exemple, si vous avez un comparateur de prix, si vous avez un agrégateur de contenus ou encore un forum, vous aurez beaucoup moins la main sur la gestion des contenus publiés, sur le fait qu'il soit ou non optimisés ou suroptimisés, ou qu'ils soient ou non dupliqués. A vous de trouver, en fonction de votre thématique et de votre secteur d'activité, le meilleur moyen pour rendre uniques et pertinentes vos publications.

Google Penguin



Les critères on-page

Les critères liés aux contenus du site sont assez faibles avec Google Penguin. Cependant, dans certains cas, ils peuvent être impliqués dans cette pénalité.

Si vous êtes touché par cet algorithme, vous devez donc également travailler vos contenus. Appliquez donc exactement les mêmes conseils que pour Google Panda par rapport à la qualité de vos contenus et de vos liens internes.

Trouver et lister ces liens

Le cœur même de cette pénalité est lié aux liens qui pointent vers votre site web. Avant d'agir, il va donc falloir lister l'ensemble des backlinks que vous avez créé pour savoir lesquels peuvent ou non poser problème.

La première étape consiste donc à vous connecter au centre pour webmasters Google de votre site puis aller dans le menu « Trafic de recherche > Liens vers votre site » qui liste une partie des liens détectés par Google. Cela constitue une première base de travail.

Vue d'ensemble » Tous les domaines
1 000 domaines principaux comportant des liens vers des pages de votre site

Télécharger ce tableau Télécharger d'autres exemples de liens Télécharger les derniers liens

Afficher 25 lignes 1 à 25 sur 1 000 < >

Domaines	Liens ▲	Pages référencées par le biais de liens
forums-abondance.com	62 329	2
rccgaming.com	62 228	1
blogspot.com	23 955	43
wordpress.com	23 282	10
gportal.hu	19 203	1
abondance.com	17 557	8
mylivepage.com	8 743	1
lacoctelera.net	7 649	1
okoun.cz	7 606	2
ulitzer.com	7 179	1
stumbleupon.com	3 398	1
cassy.be	3 153	1
bloggingstocks.com	1 291	1
over-blog.com	1 170	2
benheck.com	1 094	1

Malheureusement, celle-ci est insuffisante car le centre pour webmasters n'affichera jamais l'intégralité des liens. Nous vous conseillons donc d'utiliser d'autres outils externes qui vont compléter cette liste. On peut notamment citer :

- SemRush <http://fr.semrush.com/fr/> ;
- Ahrefs : <https://ahrefs.com/> ;
- Majestic SEO : <http://www.majesticseo.com/> ;
- Open Site Explorer : <http://www.opensiteexplorer.org/> ;
- Linkdetox : <http://www.linkdetox.com/>.

Pour l'intégralité des liens, il va ensuite falloir les trier. Le plus simple est de lister dans un fichier Excel les informations suivantes pour mieux comprendre votre profil de liens et pouvoir agir en conséquence :

- l'ancre du lien ;
- la destination du lien ;
- la présence de l'attribut nofollow (ou pas) ;
- la date de création du lien ;

- la thématique du site où à été créé le lien ;
- la thématique du contenu où a été créé le lien ;
- le type de contenus où le lien est présent (forum, blog, annuaire...) ;
- le type de liens : une URL, un mot-clé de marque, une expression optimisée SEO, une expression générique ou vague (« cliquez ici », « voir ce site »...) ;
- l'emplacement du lien : footer, sidebar, dans le contenu principal... ;
- Un lien unique ou sitewide (c'est à dire un lien présent sur toutes les pages du site distant).

Suppression et modifications des liens

Une fois cette analyse réalisée, vous allez pouvoir commencer à travailler. De tous les retours d'expériences qui ont été faits par des référenceurs, il s'avère que la méthode la plus efficace pour sortir d'une pénalité Google Penguin est de nettoyer manuellement et en profondeur l'ensemble de votre profil de liens. Cela va consister à faire deux choses : modifier ou supprimer vos backlinks.

Pour savoir ce que vous devez faire ou non, la méthodologie est simple : est-ce que le site ou la page sur lequel le lien a été créé est un emplacement logique et naturel pour ce type de backlink. Si la réponse est oui, il est recommandé de conserver ce lien, et à la rigueur de simplement le désoptimiser si nécessaire :

- soit au niveau de l'ancre ;
- soit au niveau de l'attribut nofollow ;
- soit au niveau de la destination du lien.

Si à l'inverse la réponse est non, il va falloir supprimer directement ce lien. C'est sans doute cette partie-là du travail qui sera la plus complexe, puisque vous n'aurez pas forcément la possibilité de supprimer directement le lien, et que certains webmasters vont être difficiles à joindre ou pourraient refuser une telle suppression ou tout simplement ne pas vous répondre.

Remarque : pensez toujours à conserver la liste de vos actions car cela pourrait toujours servir à prouver votre bonne foi auprès de Google.

L'outil de désaveu des liens

L'outil de désaveu des liens est un outil proposé par Google dans le centre de Webmaster de Google : <https://www.google.com/webmasters/tools/disavow-links-main?pli=1>

Le concept est simple : vous listez tous les liens que vous considérez comme indésirables pour que Google ne les prenne plus en compte. Théoriquement, cela peut vous faire sortir d'une pénalité Penguin car vous désavouerez tous les liens nocifs et sur-optimisés.

Attention cependant, ce n'est pas une solution idéale : le plus efficace reste de loin le fait de modifier ou supprimer les liens qui posent problème. Source : <https://www.youtube.com/watch?v=j7Hjgt2Hn4>

Au niveau de la méthode, elle est assez simple : vous envoyez un fichier texte qui liste l'ensemble des URL qui posent problème (une URL par ligne). Vous pouvez aussi utiliser la syntaxe suivante pour désavouer tous les liens depuis un domaine ou un sous-domaine précis :

- Le domaine principal : **domain:site.com** ou **domain:www.site.com** ;
- Un sous-domaine : **domain:sousdomaine.site.com** ;

De base, le sous-domaine est considéré comme une entité à part entière. Il faudra donc les traiter séparément pour désavouer les liens qu'il contient (source : <http://webmasters.stackexchange.com/questions/57870/disavow-a-subdomain>).

Si jamais vous vous trompez sur certains liens, il faudra modifier le fichier, supprimer les URL en question puis le renvoyer. Sachez également que vous ne pouvez avoir qu'un seul fichier de désaveu par site.

Doit-on continuer à créer des liens ?

Vous pouvez tout à fait continuer vos campagnes de netlinking en même temps. Google ne vous pénalise pas en effet pour vos futurs liens, mais bien pour certains liens actuels. Il n'y a donc aucune raison d'arrêter de créer des liens vers votre site web.

Par contre, soyez encore plus vigilant, car Google surveille d'autant plus votre site. Faites en sorte que tous les futurs liens soient les plus naturels possible, en variant les ancres, en créant des liens en nofollow, en créant des liens depuis des sites ayant une thématique vraiment proche de votre propre site, et en faisant également varier la destination de chaque lien (page d'accueil, catégorie, article...).

Les autres algorithmes

Google Panda et Google Penguin ne sont pas les seuls algorithmes qui peuvent pénaliser votre site. Il existe par exemple Google Page Layout qui cible les sites dont les publicités sont trop présentes et ou le contenu principal a été trop mis en retrait.

Pour chacun de ces algorithmes (et des nouveaux qui seront mis en place des les prochaines années, comme PayDay Loan pour les requêtes très concurrentielles et très spammées), il suffit de se demander ce qui a réellement posé problème, puis de corriger intégralement la problématique.

Pour reprendre l'exemple de Google Page Layout, il faudra supprimer les publicités en trop, et recentrer tout le design et l'ergonomie du site sur le contenu principal.

La solution du changement de site

Le changement de domaine

Au départ, certains conseillaient très souvent de changer de domaine principal pour éliminer la pénalité. Le problème est qu'une redirection 301 transmet très souvent la pénalité : ce n'est pas une solution idéale.

D'ailleurs, selon certaines personnes chez Google, même sans 301, le site pourrait recevoir la pénalité car Google pourrait détecter qu'il s'agit d'un seul et même site : <http://www.abondance.com/actualites/20140226-13676-changer-le-nom-de-domaine-dun-site-ne-supprime-pas-une-eventuelle-penalite.html>

Créer un autre site

L'autre solution est de créer un nouveau site de A à Z. La pénalité va ainsi disparaître (à condition que vous ne repreniez pas le contenu de l'ancien site à l'identique), mais vous devrez recommencer à zéro l'intégralité de votre travail de référencement naturel, à savoir la création et l'optimisation des contenus, la création de liens...

C'est donc une solution à utiliser en dernier recours lorsque tout le reste n'aura pas fonctionné.

Conclusion

Sortir d'une pénalité algorithmique est donc tout à fait possible, mais c'est un travail long à mettre en place pour que cela puisse fonctionner.

Il faudra tout d'abord bien comprendre et analyser tout ce qui a posé problème. Plus vous comprendrez quelle pénalité vous a été infligée, plus vous serez capable d'agir en conséquence.

Ce n'est qu'ensuite que vous pourrez nettoyer, désoptimiser et améliorer tous vos contenus et vos liens. Ce travail prendra des semaines ou des mois entier, et il faudra alors patienter pour que la pénalité soit éventuellement levée totalement ou partiellement. Et ce temps-là risque certainement d'être le plus long... Une période qui sera mise à contribution pour vous poser les bonnes questions sur votre stratégie SEO...

Daniel Roch, *Consultant WordPress, Référencement et Webmarketing chez SeoMix (<http://www.seomix.fr/>)*

Panda 4.0 : anatomie de la dernière version du filtre de Google

[Retour au sommaire de la lettre](#)

Domaine :	Recherche	Référencement
Niveau :	Pour tous	Avancé

Google a lancé le 20 mai dernier la quatrième version de son filtre Panda qui vise à lutter contre le contenu de faible qualité. Au travers de l'analyse des sites gagnants et perdants après ce lancement et des brevets déposés par le moteur de recherche, on peut en savoir un petit plus sur l'évolution de cet algorithme...

La dernière version du filtre Panda de Google a été déployée dans le Monde le 20 mai 2014 (<http://www.abondance.com/actualites/20140521-13926-google-panda-4-0-en-cours-deploiement.html>), selon les informations divulguées par Google. Pour la première fois depuis de nombreux mois, Google a officiellement communiqué sur une mise à jour de Panda, ce qui a évidemment suscité la curiosité : qu'est-ce que cette mouture recèle d'important aux yeux de Google ? En quoi est-elle différente des autres mises à jour ? Nous allons essayer de répondre à ces questions, en faisant le point au passage sur ce que l'on sait des mises à jour Panda.

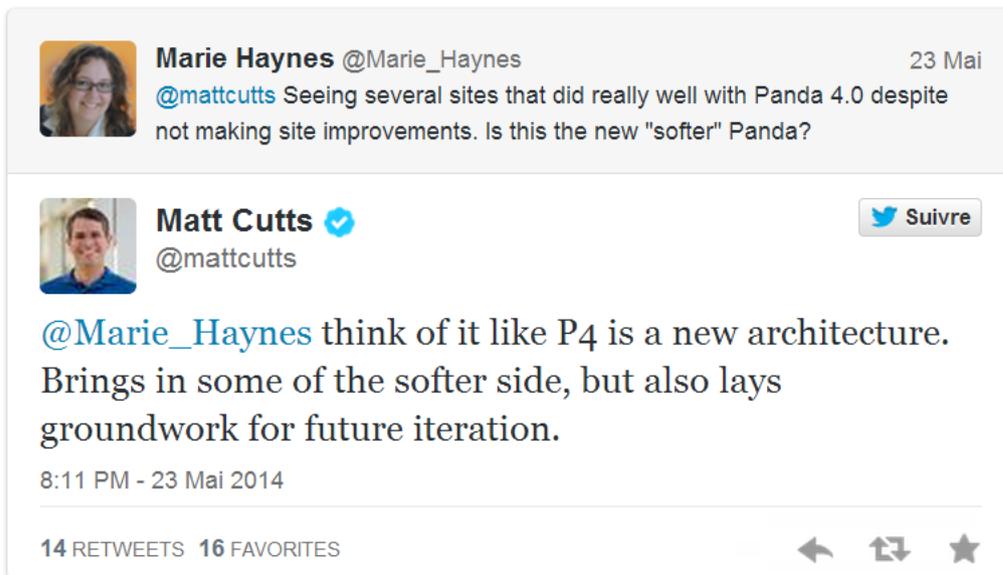
Panda 4.0 : une mise à jour atypique

Une mise à jour "officielle"

L'une des premières surprises apportées par cette mise à jour est le fait qu'elle a fait l'objet d'une annonce officielle de la part de Google. En effet, depuis de nombreux mois, les mises à jour (Google les appelle des itérations) des filtres Panda étaient censées se dérouler chaque mois, mais plus aucune communication officielle ne le confirmait. Cette fois-ci l'annonce a pris la forme d'un tweet de Matt Cutts, datant du mardi 20 mai 2014.



Autre détail intéressant à remarquer, Matt Cutts appelle depuis dans sa communication la mise à jour Panda 4.0, au lieu de Panda 27 ou 28 (son numéro d'itération). On peut en déduire qu'il s'agit de la 4ème version de l'algorithme, et non d'un simple "recalcul". Matt Cutts a effectivement confirmé, toujours sur Twitter, qu'il s'agissait d'un changement d'architecture logicielle.



Une mise à jour dont l'impact s'est révélé important

Au total, selon les dires de Matt Cutts au site Seroundtable, 7,5% des requêtes en langue anglaise ont été impactées par les changements de résultats apportés par Panda. Les itérations impactent entre 1% et 3% des requêtes habituellement. La première version de Panda en février 2011 avait chamboulé les résultats sur 11,8% des requêtes. C'est donc l'une des mises à jour les plus importantes dans l'histoire de ce filtre. On ne connaît par contre l'impact sur les autres versions linguistiques de Google, en particulier en France. Mais les différents baromètres fournis par les outils ont montré des changements profonds dans les pages de résultats françaises également.

Déploiement de filtres en série

Le déploiement de Panda a été précédé au cours du week-end précédent (soit 48 heures avant seulement) par le déploiement d'un autre filtre antispam surnommé « Payday loan » dont c'était la deuxième mouture (son premier passage remonte au printemps 2013 et la 3.0 a été lancée à la mi-juin 2014). « Payday loan » vise à nettoyer le spam présent sur certaines requêtes particulièrement ciblées par les web spammers, parmi lesquelles la fameuse requête « Payday loan ».

Les mouvements dans les pages de résultats ont été globalement importants pendant tout le mois de mai, au point que les forums bruissaient d'annonces non confirmées du passage de Panda ou de Penguin pendant toute la période.

Une mise à jour plus "douce" ?

Ce qui a frappé les observateurs c'est que cette nouvelle version du filtre a cette fois-ci non seulement touché un grand nombre de nouveaux sites, mais qu'elle a été marquée par la remontée spectaculaire d'autres sites parfois pénalisés depuis longtemps par les versions précédentes.

Cette version est-elle plus "douce" (*softer*) comme l'avait annoncé Matt Cutts ? En fait il ne s'agit pas de la "*softer update*" annoncée par Matt Cutts il y'a quelques mois à l'occasion d'une conférence, et qui est censée être plus adaptée aux sites édités par de petites entreprises. Mais incontestablement, cette version semble plus "subtile" et plus "sélective", elle fait à la fois tomber des sites qui n'avaient pas encore été impactés, et remonter des sites qui avaient été touchés parfois depuis longtemps.

Quelles catégories de sites ont été les plus impactés par Panda 4.0 ?

Il s'est dit pas mal de bêtises sur la Toile à propos de Panda 4.0, comme à chaque fois que les gens analysent la liste des sites qui montent et qui descendent dans les classements après une mise à jour. Il faut effectivement garder à l'esprit que tous les jours, des sites montent et descendent dans les classements, et toutes les semaines, on observe une collection de hausses spectaculaires et de baisses drastiques. Après une mise à jour, des tendances se dégagent, mais des sites peuvent figurer dans le top 100 des sites qui montent ou baissent pour d'autres raisons que Panda 4.0.

Par exemple, beaucoup ont remarqué la chute sévère d'eBay sur les courbes de visibilité de Searchmetrics. eBay est-il pénalisé par Panda ? Possible, mais bien des éléments plaident pour une coïncidence (notamment le fait que la chute ait apparemment commencé bien avant le 20 mai 2014) et d'autres pénalités ou problèmes semblent bien mieux expliquer la présence d'eBay dans le top des perdants post Panda.

Les perdants et les gagnants sur GOOGLE USA

Voici la liste des sites gagnants et des sites perdants d'après la société Searchmetrics :

Google SERPs: Panda 4.0 Losers

The lists are ordered by absolute loss / gain, which is shown in the respective column as rounded percentage. Basis of the analysis are domains that at least had a value of 10000 SEO Visibility before Panda 4.0 and that show now a change of at least +/-20 %.

domain	percent
ask.com	> - 50%
ebay.com	> - 33%
biography.com	> - 33%
retailmenot.com	> - 33%
starpulse.com	> - 50%
history.com	> - 33%
isitdownrightnow.com	> - 50%
aceshowbiz.com	> - 75%
examiner.com	> - 50%
yellowpages.com	> - 20%
yourtango.com	> - 75%
dealcatcher.com	> - 50%
livescience.com	> - 50%
webopedia.com	> - 50%
xmarks.com	> - 50%
simplyrecipes.com	> - 33%
siteslike.com	> - 50%
digitaltrends.com	> - 50%
health.com	> - 50%
spoonful.com	> - 75%
songkick.com	> - 75%
realsimple.com	> - 33%
appbrain.com	> - 33%
thehollywoodgossip.com	> - 50%
dealspl.us	> - 33%
techtarget.com	> - 33%
gossipcop.com	> - 50%
rd.com	> - 75%
chow.com	> - 33%
doxo.com	> - 50%
heavy.com	> - 50%
csmonitor.com	> - 33%
parenting.com	> - 50%
globalpost.com	> - 75%
espnfc.com	> - 50%
serviceguidance.com	> - 50%
mnn.com	> - 75%
mystore411.com	> - 50%
urlm.co	> - 33%
delish.com	> - 50%
healthcentral.com	> - 33%
whatscookingamerica.net	> - 50%
columbia.edu	> - 20%
songlyrics.com	> - 20%
internetslang.com	> - 33%
ibiblio.org	> - 50%
webutation.info	> - 50%
cheapflights.com	> - 33%
mybanktracker.com	> - 50%

La liste révèle que les fondamentaux de Google Panda sont toujours les mêmes. On trouve les catégories de site habituelles : les annuairistes (yellowpages.com), les fermes de contenu ([livescience](http://livescience.com)), les agrégateurs de contenus, les faux sites de news sans valeur ajoutée ([thehollywoodgossip](http://thehollywoodgossip.com)) les sites de coupons ([dealcatcher](http://dealcatcher.com)), les agrégateurs de recettes, etc.

Dans la liste, comme le remarque Marcus Tober (le directeur métier de Searchmetrics), on remarque également des sites qui ont probablement été impactés par d'autres causes (au premier rang desquels on trouve eBay).

La liste des gagnants est moins claire, mais on remarque néanmoins que cette mise à jour a permis la remontée d'anciens sites Pandalisés.

Google SERPs: Panda 4.0 Winners

domain	percent
glassdoor.com	> 100%
emedicinehealth.com	> 500 %
medterms.com	> 500 %
yourdictionary.com	> 50%
shopstyle.com	> 250%
zimbio.com	> 500 %
myrecipes.com	> 250%
couponcabin.com	> 250%
buzzfeed.com	> 25%
consumeraffairs.com	> 100%
wordpress.com	> 20%
thinkexist.com	> 250%
onhealth.com	> 250%
alternativeto.net	> 100%
whosdatedwho.com	> 250%
reverso.net	> 50%
wikimedia.org	> 100%
dogtime.com	> 100%
findthebest.com	> 50%
eatingwell.com	> 100%
quotegarden.com	> 100%
goodhousekeeping.com	> 250%
everydayhealth.com	> 25%
simplyhired.com	> 100%
momswhothink.com	> 100%
similarsites.com	> 100%
southernliving.com	> 50%
theknot.com	> 25%

allaboutvision.com	> 100%
openculture.com	> 50%
babyzone.com	> 50%
tasteofhome.com	> 33%
gotquestions.org	> 100%
movie4k.to	> 50%
wmagazine.com	> 33%
ycharts.com	> 100%
historyplace.com	> 50%
rcn.com	> 100%
salary.com	> 50%
skepdic.com	> 100%
mediawiki.org	> 100%
oodle.com	> 100%
abbreviations.com	> 100%
homes.com	> 100%
spokeo.com	> 50%
hlnvtv.com	> 33%
sparkpeople.com	> 33%
hayneedle.com	> 50%
emedtv.com	> 100%

En ce qui concerne la France, les données de Searchmetrics ont été publiée par l'agence Search Foresight :

Domaines Perdants	Visibilité	Perte	Evolution	Domaines gagnants	Visibilité	Gain	Evolution
softonic.fr	727453	-235233	-24%	comprendrechoisir.com	143994	88615	160%
yelp.fr	93888	-57944	-38%	passeportsante.net	123112	64226	109%
idealo.fr	36790	-46850	-56%	recettes.de	136019	59386	77%
annuairepagesblanches.org	19743	-37198	-65%	arobase.org	56121	35745	175%
films-vf.com	14335	-33170	-70%	plurielles.fr	60167	35295	142%
118712.fr	55524	-29419	-35%	e-sante.fr	46133	34758	306%
bonial.fr	50554	-27458	-35%	babelio.com	69977	33280	91%
ebay.com	40454	-24284	-38%	annoncesjaunes.fr	52863	32608	161%
xmarks.com	33836	-24229	-42%	750g.com	109363	32516	42%
annuaire-inverse-france.com	38848	-23255	-37%	pratique.fr	51278	31097	154%
allrecipes.fr	54809	-22654	-29%	universalis.fr	40622	30662	308%
alldocteurs.fr	36190	-21505	-37%	sokrostream.com	47865	30497	176%
film-en-lignee.com	9755	-20366	-68%	malavida.com	64723	29273	83%
achetezfacile.com	27302	-19366	-41%	largus.fr	68366	28546	72%
ma-reduc.com	45799	-19254	-30%	paroles-musique.com	45379	25342	126%
chaticam.com	13307	-18490	-58%	leguide.com	115141	24906	28%
cine-resistances.fr	7158	-18449	-72%	puretrend.com	30180	23488	351%

phpnuke.org	21387	-14319	-40%	wat.tv	164025	23068	16%
midilibre.fr	40214	-14196	-26%	myskreen.com	41588	22492	118%
paroles2chansons.com	22029	-13975	-39%	ptitchef.com	37857	22475	146%
webutation.info	14619	-13848	-49%	dicocitations.com	40599	21712	115%
tonicmovies.com	16459	-13673	-45%	francetv.fr	86968	21462	33%
yakeo.com	23495	-13495	-36%	routard.com	191288	19722	11%
answers.com	13114	-13227	-50%	kelkoo.fr	42734	19432	83%
similarsearch.com	23202	-13033	-36%	radins.com	70673	18413	35%
1001actus.com	10556	-13017	-55%	memoclic.com	28505	17627	162%
jeuxdeheros.com	1197	-12663	-91%	fapvid.com	134968	17192	15%
ask.com	4379	-12528	-74%	purebreak.com	21332	17067	400%
agendaculturel.fr	14443	-11780	-45%	jcomjeune.com	28869	16509	134%
hugesex.tv	43298	-11619	-21%	uptodown.com	39634	15589	65%
me-desinscrire.fr	7627	-10417	-58%	notrefamille.com	71371	15397	28%
melycampus.fr	8563	-10358	-55%	cylex-france.fr	30943	15107	95%
nomao.com	7755	-9864	-56%	118000.fr	58356	14702	34%
perfectgirls.net	7459	-9863	-57%	cbanque.com	51918	13318	35%
jeu-gratuit.net	20249	-9849	-33%	aujardin.info	22851	13258	138%
espace-client.net	35749	-9826	-22%	dossierfamilial.com	20183	13074	184%
topwarez.fr	1956	-9675	-83%	canalmonde.fr	17009	12956	320%
appstv.com	6276	-9519	-60%	euronews.com	28654	12927	82%
cuponation.fr	5693	-9030	-61%	logiciel.net	21518	12810	147%
reponseatout.com	23431	-8989	-28%	franceinfo.fr	100330	12741	15%
dressfirst.fr	6065	-8853	-59%	enfant.com	19380	12050	164%
commerces.com	16212	-8596	-35%	lsa-conso.fr	47753	11929	33%
begeek.fr	20579	-8344	-29%	meilleurmobile.com	38961	11878	44%
seconnecter.net	27636	-8313	-23%	foiredustreaming.com	23158	11772	103%
publicsenat.fr	2050	-8212	-80%	shopstyle.fr	18314	11766	180%
informer.com	28450	-8139	-22%	girlsgogames.fr	124729	11739	10%
foot01.com	24265	-8112	-25%	supertoinette.com	23888	11283	90%
economiamatin.fr	5869	-8035	-58%	studyrama.com	17396	10956	170%
netguide.com	13986	-7566	-35%	openclassrooms.com	67198	10910	19%
voyages-sncf.mobi	8970	-7535	-46%	eurekasante.fr	43304	10848	33%
dinosauria.com	10593	-7480	-41%	dilandau.la	11103	10833	4012%
hotelhotel.com	8111	-7439	-48%	docteurclic.com	13922	10671	328%
code-2-reduction.fr	3562	-7302	-67%	aguea.com	12565	10489	505%
dilandau.eu	9233	-7086	-43%	applicationiphone.com	19043	10351	119%
forumpro.fr	14874	-6770	-31%	jamendo.com	23153	10345	81%
mon-compte.org	15075	-6693	-31%	aujourd'hui.fr	52880	10272	24%
siteslike.com	3104	-6454	-68%	portalux.com	14149	10197	258%
cuisine-libre.fr	7179	-6443	-47%	jobijoba.com	21076	9917	89%
rosny2.com	14966	-6424	-30%	blogdumoderateur.com	12563	9674	335%
hm.com	16103	-6419	-29%	videocourtesclic.com	37943	9658	34%
modele-tatouage.fr	8821	-6360	-42%	commeaucinema.com	34915	9645	38%
aftouch-cuisine.com	6885	-6359	-48%	maison-deco.com	15637	9626	160%
ca-nord-est.fr	23171	-6265	-21%	decideursenregion.fr	11212	9521	563%
frenchweb.fr	16336	-6248	-28%	petitfute.com	16337	9446	137%
anime-no-paradise.com	4058	-6115	-60%	inverse-annuaire.org	13425	9422	235%
ac-creteil.fr	22307	-6032	-21%	321auto.com	18324	9418	106%
audiko.net	1359	-5994	-82%	softonic.com	66606	9402	16%
digischool.fr	9442	-5903	-38%	motoservices.com	13676	9238	208%
locafilm.com	6952	-5833	-46%	fia-net.com	20088	9162	84%
annuairepagesblanches.org	13361	-5717	-30%	shopoon.fr	14792	8980	155%
unige.ch	11665	-5540	-32%	custplace.com	11165	8844	381%
centrerecommercial-auchan.fr	13773	-5445	-28%	provenceweb.fr	17610	8781	99%
mon-compte.net	7072	-5434	-43%	01amour.com	20512	8537	71%
misterfast.com	5911	-5302	-47%	unifrance.org	21722	8497	64%
arnaques-internet.info	8717	-5283	-38%	turf-fr.com	15020	8407	127%
tubeadulte.com	3816	-5251	-58%	divertissonsous.com	38307	8404	28%

chanel.com	4868	-5225	-52%	youmoviz.tv	16298	8285	103%
fil-info-france.com	9765	-5178	-35%	videospornofrancaises.com	23061	8140	55%
pornomure.com	5021	-5168	-51%	logitheque.com	47418	8076	21%
mailorama.fr	9638	-5163	-35%	lerepairedesmotards.com	11906	7929	199%
cuisineetvinsdefrance.com	19641	-5032	-20%	brandalley.fr	47681	7844	20%
telecharger-films.ws	1372	-4984	-78%	observatoiredefranchise.fr	12468	7820	168%
e-lyco.fr	14409	-4903	-25%	pornodingue.com	17949	7809	77%
kvideosxxl.com	712	-4888	-87%	el-annuaire.com	10966	7745	240%
youscribe.com	18927	-4860	-20%	ciao.fr	52577	7718	17%
okaidi.fr	2732	-4784	-64%	dictionnaire-juridique.com	16137	7713	92%
lesprixducoin.com	9916	-4784	-33%	hotmail.com	60156	7603	14%
conseil-general.com	8845	-4715	-35%	cosmovisions.com	9437	7528	394%
euria.org	4926	-4702	-49%	guide-sites-rencontres.fr	14595	7519	106%
pricededuc.com	7418	-4685	-39%	wamiz.com	51084	7518	17%
jolpress.com	3219	-4669	-59%	igraal.com	70165	7478	12%
pretachanger.fr	6559	-4619	-41%	be.com	24079	7467	45%
notices-pdf.com	9733	-4600	-32%	mcetv.fr	23081	7454	48%
yellowpages-fr.com	8613	-4573	-35%	jetelecharge.com	71648	7427	12%
annuaire.pro	2613	-4572	-64%	espacefrancais.com	9550	7407	346%
cinemavf.net	8001	-4508	-36%	streamingi.net	56642	7393	15%
ebay.co.uk	2246	-4494	-67%	concertandco.com	16851	7307	77%
tv-direct.fr	7078	-4446	-39%	perles-du-bon-coin.fr	25795	7122	38%
toutes-les-villes.com	16403	-4395	-21%	citation-du-jour.fr	12805	7114	125%
zataz.com	513	-4282	-89%	lespagesjaunesafrique.com	18655	6957	59%

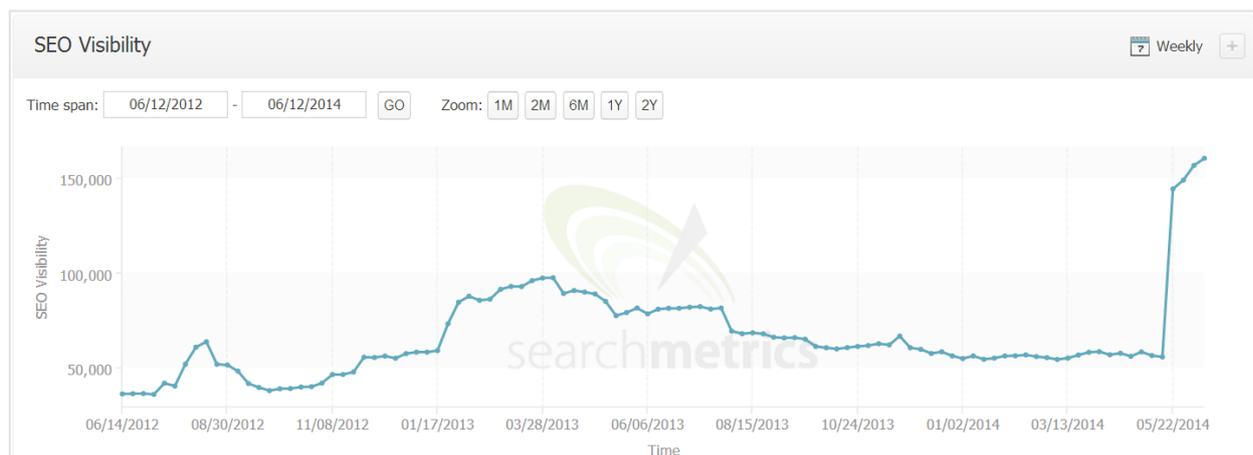
Une fois de plus, ce qui frappe c'est la sortie de pénalité spectaculaire d'un certain nombre de sites.

Un premier exemple avec un site d'infos sur la santé :



Pour cet exemple, on peut penser qu'un changement de critères est à l'oeuvre, car fondamentalement le site n'a pas changé de contenu à ce point.

Un deuxième exemple qui a été parfois considéré par une ferme de contenus (Google a-t-il changé ses critères ou le site a-t-il tiré des leçons du passé et s'est amélioré : probablement les deux ici).



Il semble, et Matt Cutts l'a confirmé, que cette mise à jour est plus "sélective", et a donc permis la remontée de nombreux sites longtemps pénalisés.

Mais on trouve également des exemples où le filtre devient plus sévère. Un exemple de site pénalisé pour finir : le cas d'un comparateur de prix :



Quoi de neuf dans cette mise à jour alors ?

Google, par l'intermédiaire de Matt Cutts, a révélé que le déploiement de Panda 4.0 était beaucoup plus rapide qu'avec les versions précédentes (qui prenaient jusqu'à 10 jours pour être prises en compte sur toutes les requêtes et tous les pays). C'est donc un changement d'architecture logicielle mais aussi un changement dans le fonctionnement du filtre.

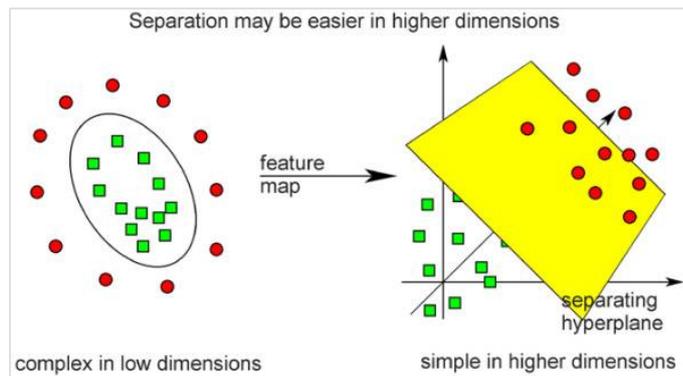
Mais pour mieux comprendre ce qui a changé, il faut faire un retour en arrière et faire le point sur ce que l'on sait sur le filtre Panda.

Panda, un "classifieur" de sites

Peu d'informations ont "fuité" sur la nature exacte de ce filtre, mais suffisamment néanmoins pour se faire une idée précise du processus mis en oeuvre. Panda cherche à "imiter" les choix que feraient des évaluateurs humains, à qui on confierait le soin de trier les sites entre d'un côté de la ligne ceux qui sont de mauvaise qualité, et n'ont rien à faire dans les pages de résultats, et les autres, qui répondent à des critères de qualité minimum. Amit Singhal, le responsable de l'algorithme chez Google, avait ainsi révélé une liste de questions qui permettaient de se faire une idée des critères qui pouvaient être

utilisés par des évaluateurs humains. La magie du système inventé par Navneet Panda, l'ingénieur qui a inventé le filtre (et qui a donné son nom à la première "bestiole" du bestiaire de Google), est qu'il ne cherche pas à classer les sites selon les critères utilisés par les humains, mais à trouver automatiquement une façon de prendre les mêmes décisions, à partir des données à sa disposition (différentes et parfois plus limitées).

Il semble que la fonction d'évaluation générée à chaque itération de Panda ressemble à ce qui est obtenu lorsque l'on utilise les *Support Vector Machines*. Les SVM sont capables de générer automatiquement, à partir d'un échantillon pré-trié par des humains, une fonction de classement (dans le jargon informatique on appelle cela un "classifieur", voir également l'article des frères Peyronnet à ce sujet dans cette même lettre) qui permet de séparer les sites en deux classes : "spam" et "non spam". Une fois ce classifieur créé, il peut servir à trier TOUS les sites, avec un taux de succès important.



Un schéma expliquant le principe des SVM. Une astuce géométrico-mathématique (un changement de dimension) permet de « simplifier » la définition d'une fonction permettant de classer des items. Il existe de nombreuses variantes de cette approche, on ne sait pas si les SVM sont réellement utilisés dans Panda: il existe juste une ressemblance entre la description du fonctionnement de l'algorithme de Panda par Amit Singhal dans un article de Wired en 2011, et ce type de schéma...

Néanmoins, l'exercice a ses limites, et les différentes versions de Panda ont montré que la première version était parfaitement capable de faire de mauvais choix, en "oubliant" de pénaliser certains sites et en frappant d'autres, parfois à tort. L'un des enjeux pour Google est d'améliorer la qualité de son classifieur pour éliminer en particulier les faux positifs et en même temps mieux discerner les cas "gris".

Dans un premier temps, ce "classifieur" a été déployé à la main. Les itérations suivantes ont été calculées avec une périodicité variable, avant que le déploiement ne soit automatisé. La séparation que fait Google entre les "versions" de Panda et les "itérations" vient du fait que l'on peut soit :

- refaire le tri sans changer le classifieur. Dans ce cas on prend juste en compte les nouvelles données collectées sur les sites ;
- régénérer le classifieur sans changer les fondamentaux (comme la liste des indicateurs utilisés, ou la forme de la fonction de tri) ;
- générer un classifieur totalement nouveau, exploitant de nouveaux critères et éventuellement différent en termes de complexité ou de nature de la fonction d'évaluation.

Clairement, Panda 4.0 semble appartenir à la 3ème catégorie.

Le brevet de Navneet Panda

Bill Slawski, du blog SEO by the Sea, a publié deux posts successifs à propos d'un brevet publié par l'ingénieur à qui on a attribué la paternité du filtre Panda : Navneet Panda.

Ce brevet, baptisé sommairement *Ranking search results*, décrit une méthode pour améliorer la qualité des pages de résultats de Google en déclassant certains types de sites.

Bill Slawski affirme qu'il s'agit DU brevet sur l'algorithme Panda. Nous pensons sincèrement qu'il s'agit probablement d'une des nombreuses stratégies que Navneet Panda a dû imaginer pour disposer des données permettant de distinguer certains sites de mauvaises qualité. Il nous semble logique de penser qu'il en existe d'autres, tant Panda semble répondre à de nombreux critères, beaucoup plus variés que ceux décrits dans ce brevet. Mais il s'agit néanmoins d'une confirmation éclatante de l'existence de deux critères souvent évoqués (au milieu d'autres) comme conduisant potentiellement à une pénalité Panda :

1. La mauvaise qualité des backlinks ;
2. L'absence de requêtes de référence associées au site.

Rappelons comme d'habitude que ce n'est pas parce que Google a déposé un brevet qu'il décrit un système réellement mis en production sur ses outils. L'algorithme de Google peut utiliser une version modifiée de ce système, comme ne pas l'utiliser du tout.

La mauvaise qualité des liens

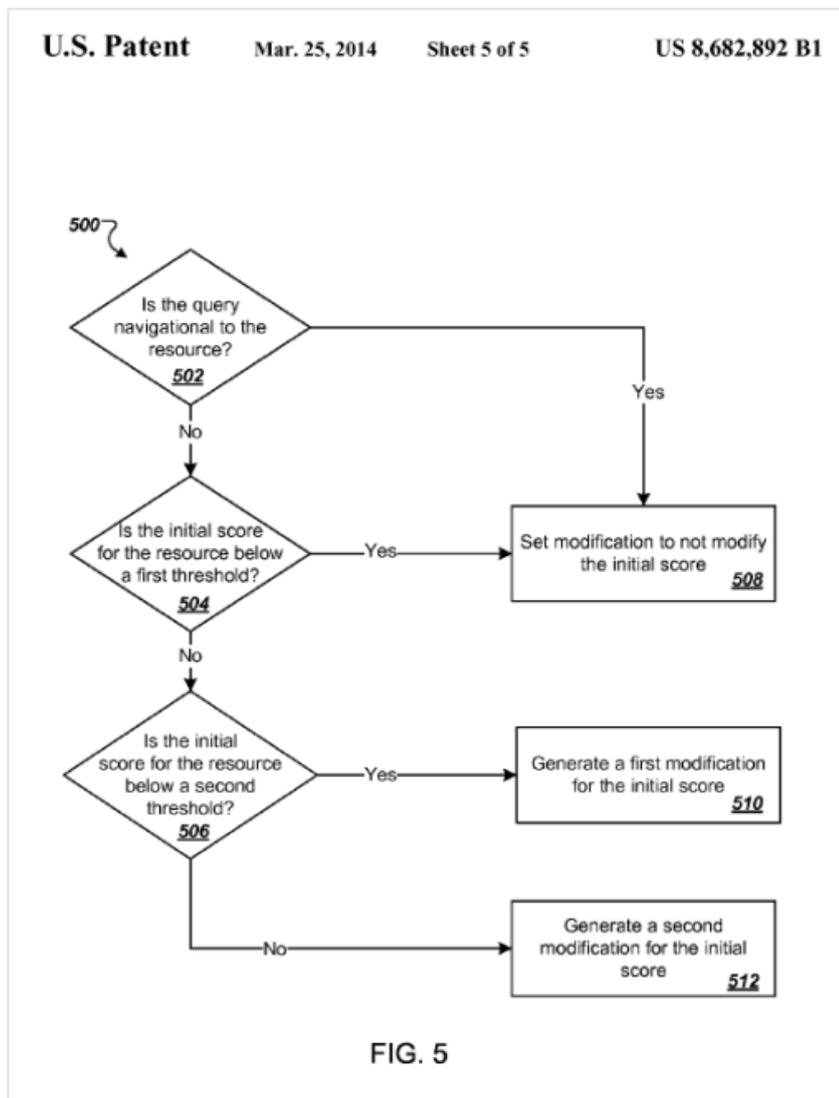
Le brevet décrit une méthode permettant de filtrer, parmi les backlinks, tous les liens que contrôle le webmaster, pour créer à la fin un indicateur dépendant du nombre de liens « indépendants » (non manipulés, donc indépendants vis-à-vis du webmaster/éditeur du site). Un grand nombre de liens indépendants protège donc de Panda. Beaucoup le subodoraient, et l'on retrouve ce critère dans les recommandations de plusieurs experts SEO.

Plus étonnant toutefois, on constate que le brevet évoque l'exploitation de « liens implicites », c'est-à-dire de simples mentions dans des textes contenus dans des pages web (sans balise `` : en écrivant `abondance.com`, je viens de faire un lien implicite vers le site d'Olivier Andrieu).

Les requêtes de référence

Une autre méthode décrite dans le brevet consiste à analyser à quel point un site contient des pages qui sont des réponses logiques sur des requêtes particulières (les requêtes de référence du site). Un indicateur est calculé qui tient compte pour un site donné, du nombre de requêtes de référence ciblées par le site.

Les deux indicateurs sont ensuite utilisés de manière combinée pour créer un indicateur composite censé aider à la détection de sites de mauvaise qualité (et à favoriser les autres).



Un diagramme illustrant le brevet, montrant comment l'indicateur composite est utilisé pour reranker les pages.

Panda 4.0 change-t-il les guidelines pour ne pas être pénalisé ?

On en sait donc de plus en plus sur la nature et le fonctionnement du filtre Panda. Mais le filtre est complexe par nature et tient compte de nombreux critères (dont peut-être ceux exposés dans le brevet évoqué ci-dessus).

Les bonnes pratiques pour ne pas être « Pandalisé » restent donc inchangées, mais on peut considérer que la dernière version du filtre est un peu plus sélective que les versions précédentes, et montre que l'on peut sortir dorénavant complètement d'une pénalité, même pour des sites qui semblaient spécialement ciblés par les versions précédentes de Panda. C'est l'aboutissement de trois ans d'améliorations successives chez Google.

Espérons que les prochaines versions seront encore plus « fines » dans leurs choix de victimes, car il n'y a rien de pire que d'être pénalisé injustement par un algorithme aveugle et mal conçu... ce qu'était probablement en partie la première version de Panda.

Le droit à l'oubli enfin respecté par Google ?

[Retour au sommaire de la lettre](#)

Domaine :	Recherche	Référencement
Niveau :	Pour tous	Avancé

Google a annoncé récemment la mise en place d'un formulaire de demande de droit à l'oubli, permettant de supprimer certains contenus de l'index du moteur de recherche. Cet article revient sur la décision de la Cour Suprême de l'Union européenne qui a provoqué cette action et analyse les conséquences d'une telle mise en oeuvre...

La cour suprême de l'Union européenne a dernièrement rendu un arrêt « important » d'un point de vue médiatique et opérationnel en rappelant une règle vieille de 19 ans qui n'était pas respectée pleinement jusque-là. Les juges de la Cour de Justice de l'Union européenne ont ainsi remis en avant le « droit à l'oubli » visé dans la directive européenne de 1995 pour l'appliquer aux moteurs de recherche dans le cadre des résultats naturels affichés. Google, le moteur concerné, a immédiatement accepté la décision et déployé une procédure permettant aux personnes de demander la suppression des données personnelles.

Qu'est ce que le droit à l'oubli ?

Le droit des données personnelles (la loi "Informatique et Libertés" en France et la Directive 95/46/CE relative à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données pour l'Union européenne) prévoit le droit à l'oubli de manière très explicite.

Par exemple, l'article 6 de la loi « informatique et libertés » prévoit qu'un « *traitement ne peut porter que sur des données à caractère personnel* » que si elles « *sont conservées sous une forme permettant l'identification des personnes concernées pendant une durée qui n'excède pas la durée nécessaire aux finalités pour lesquelles elles sont collectées et traitées* ». L'article 38 précise à ce titre que « *toute personne physique a le droit de s'opposer, pour des motifs légitimes, à ce que des données à caractère personnel la concernant fassent l'objet d'un traitement* ».

Ainsi, toute personne a le droit de demander, pour des raisons légitimes, que ses données personnelles ne soient plus traitées.

Ce droit à l'oubli a toujours été un fondement principal de la législation des données personnelles et trouve son origine dans les grands principes (français) du pardon et du droit à une seconde chance.

Ce droit a été mis en œuvre fréquemment par la justice française. Par exemple, dans une affaire traitée en référé, Yahoo! s'est vue reprochée de la part d'une personne, Mademoiselle J.N., l'indexation par son moteur de recherche naturelle, de sites pornographiques associés à son nom, diffusant notamment des photos de Mademoiselle J.N. dénudée (Tribunal de Grande Instance Paris, référé, 17 avril 2008, inédit). Le nouveau juge parisien a souligné que le traitement de nom et prénom entrainé dans le champ d'application de cette loi et que donc, Yahoo! était dans l'obligation de respecter les devoirs d'un responsable du traitement, à savoir accéder à la demande de suppression de traitement de données personnelles.

Pourquoi les moteurs ne respectaient pas strictement ce droit ?

En pratique, il était fréquent que les moteurs reçoivent des demandes de personnes physiques demandant la suppression de référencement naturel de sites ou contenus désavantageux voire violant la vie privée. La pratique mise en place consistait à ce que la personne concernée prenne d'abord contact avec le site violant la vie privée avant de contacter le moteur de recherche. Cette pratique, plus ou moins imposée par les moteurs, a ensuite été validée par les tribunaux français.

Mais, le problème était qu'une fois le contact pris était constaté comme infructueux, la personne se retournait vers les moteurs qui désindexait rarement. En effet et notamment chez Google, il était de politique d'attendre une décision de justice pour effectuer ces déréférencements. Or, un procès coûte et surtout dure. Ainsi, une personne devait d'abord passer par la case « jugement » qui peut durer jusqu'à 18 mois pour voir enfin sa requête acceptée. Si la vie privée était violée, il fallait attendre plusieurs mois encore...

Le cas Google Spain / AEPD-Gonzalez

Le cas qui a donné lieu à l'arrêt de la CJUE est intéressant car il ne s'agissait pas de la vie privée d'une personne physique (comme par exemple et souvent, des photos dénudées), mais d'un entrepreneur qui avait été dans l'obligation de vendre aux enchères un bien immobilier suite à une saisie en 1998. Or, lorsqu'un internaute introduisait son nom dans Google, la liste de résultats affichait des liens vers deux pages du quotidien de La Vanguardia, datées de janvier et mars 1998 qui annonçaient notamment la vente destinée à recouvrer les dettes de sécurité sociale dues par M. González.

Cet entrepreneur a demandé à la CNIL espagnole (AEPD) d'enjoindre le quotidien de désindexer cette information et d'enjoindre Google de faire de même soutenant que la saisie dont il avait fait l'objet avait été entièrement réglée depuis plusieurs années et que la mention de celle-ci était désormais dépourvue de toute pertinence. L'AEPD a demandé à Google de prendre les mesures nécessaires pour retirer les données de leur index. Google a immédiatement saisi la justice pour demander l'annulation de la décision de l'AEPD. C'est dans ce contexte que la juridiction espagnole a déféré une série de questions à la Cour de justice européenne.

Il n'a pas été très compliqué pour la Cour de découvrir le droit à l'oubli puisqu'il existe depuis 1995 et qu'il fait l'objet de nombreuses batailles entre les moteurs et la Commission depuis plus de 10 ans.

De manière plus précise, la Cour constate que Google est, dans certaines conditions, obligé de supprimer de la liste de résultats, affichée à la suite d'une recherche effectuée à partir du nom d'une personne, des liens vers des pages web, publiées par des tiers et contenant des informations relatives à cette personne. La Cour précise qu'une telle obligation peut exister également dans l'hypothèse où ce nom ou ces informations ne sont pas effacés préalablement ou simultanément de ces pages web, et ce, le cas échéant, même lorsque leur publication en elle-même sur lesdites pages est licite.

Cependant, la Cour constate qu'il y a lieu de rechercher un juste équilibre notamment entre un droit à l'information / accès à l'information propre à Internet et le respect de la vie privée et le droit à la protection de données à caractère personnel. La Cour relève à cet égard que cet équilibre peut toutefois dépendre, dans des cas particuliers, de la nature de l'information en question et de sa sensibilité pour la vie privée de la personne concernée ainsi que de l'intérêt du public à recevoir cette information, lequel peut varier, notamment, en fonction du rôle joué par cette personne dans la vie publique.

Une dernière question est posée à la Cour : peut-on demander la suppression de données « historiques » concernant une personne. La Cour répond qu'il convient, à ce moment, d'examiner si cette personne a un droit à ce que les informations en question relatives à sa personne ne soient plus, au stade actuel, liées à son nom par une liste de résultats qui est affichée à la suite d'une recherche effectuée à partir de son nom. Si tel est le cas, les liens vers des pages web contenant ces informations doivent être supprimés de cette liste

de résultats, à moins qu'il existe des raisons particulières, telles que le rôle joué par cette personne dans la vie publique, justifiant un intérêt prépondérant du public à avoir, dans le cadre d'une telle recherche, accès à ces informations.

La Cour précise que la personne concernée peut adresser de telles demandes directement à l'exploitant du moteur de recherche qui doit alors dûment examiner le bien-fondé de celles-ci. Lorsque le responsable du traitement ne donne pas suite à ces demandes, la personne concernée peut saisir l'autorité de contrôle ou l'autorité judiciaire pour que celles-ci effectuent les vérifications nécessaires et ordonnent à ce responsable des mesures précises en conséquence.

Cette jurisprudence nouvelle ne fait, finalement, que préciser un régime bien connu. L'intérêt particulier porte sur le fait qu'il s'agisse de la Cour la plus importante en Europe qui précise, de manière non discutable, un régime juridique. Google n'a plus aucun moyen judiciaire de contester cette décision.

C'est d'ailleurs à ce titre et pourquoi Google a immédiatement mis en œuvre sa politique d'accepter de telles demandes.

Comment Google met en œuvre la décision de la CJUE

Google a immédiatement réagi en mettant en ligne une page de demande de suppression de données personnelles (https://support.google.com/legal/contact/lr_eudpa?product=websearch). Bing a annoncé qu'il en fera de même très prochainement.

Selon un récent arrêt de la Cour de justice de l'Union européenne, certains utilisateurs ont le droit de demander aux moteurs de recherche de supprimer les résultats de recherche qui incluent leur nom, pour autant que lesdits résultats soient *inadéquats, pas ou plus pertinents ou excessifs au regard des finalités du traitement*.

Dans le cadre de l'application de cette décision, nous évaluerons chaque demande individuelle et tenterons de trouver un juste équilibre entre la protection de la vie privée des individus et le droit du public à accéder à ces informations et à les diffuser. Lors de l'évaluation de votre demande, nous vérifierons si les résultats comprennent des informations obsolètes vous concernant. Nous chercherons également à déterminer si ces informations présentent un intérêt public, par exemple, si elles concernent des escroqueries financières, une négligence professionnelle, des condamnations pénales ou une conduite publique adoptée par un fonctionnaire.

Si vous souhaitez soumettre une demande de suppression, veuillez remplir le formulaire ci-dessous. Notez que ce formulaire ne constitue qu'une première étape. Nous sommes impatients de collaborer étroitement avec les autorités de protection des données et d'autres organismes compétents en la matière au cours des prochains mois afin de perfectionner notre approche.

Demande de suppression de résultat de recherche au titre de la législation européenne relative à la protection des données

Vous aurez besoin d'une copie d'une pièce d'identité avec photo en cours de validité pour remplir ce formulaire. Les champs identifiés par un astérisque (*) doivent obligatoirement être renseignés.

 Nous nous efforçons de mettre en place dans les meilleurs délais notre procédure de demande de suppression au titre de la législation européenne relative à la protection des données. En attendant, veuillez remplir le formulaire ci-dessous. Nous vous tiendrons informé lorsque nous débuterons le traitement de votre demande. Merci de votre patience.

Sélectionnez le pays dont la législation s'applique à votre demande *

France ↕

Vos informations

Nom *

Nom complet associé aux résultats de recherche à supprimer

Votre nom 

Votre nom, s'il est différent

Votre relation avec la personne que vous représentez, si ce n'est vous-même ("conjoint" ou "avocat", par exemple)

Dès le premier paragraphe, Google précise - avec raison - que les demandes doivent être justifiées, pertinentes et soulignent les règles (rappelées par la Cour), de détermination d'un tel caractère.

Il est bien précisé que les demandes doivent émaner des personnes directement ou d'un représentant (comme un avocat), avec la copie d'une pièce d'identité ainsi que les URL incriminées.

Cette procédure ressemble désormais pour beaucoup à la procédure légale mise en place dans le cadre de l'article 6 de la LEN où toute personne peut demander le déréférencement d'un contenu « manifestement illicite ». Il est probable qu'en pratique, le service qui traite les demandes soit le même et que les procédures s'harmonisent totalement rapidement.

Ainsi, c'est bien les moteurs qui vont déterminer la recevabilité d'une demande de suppression des moteurs. En mettant en œuvre la décision de la CJUE, les moteurs (et surtout Google). Concrètement, personne ne connaît les critères à ce jour et il est probable que les services internes juridiques ne les aient pas totalement définis. Il est donc possible de voir, malgré la mise en place de cette procédure, de nombreuses personnes déférer devant la CNIL (ou équivalent ailleurs en Europe), les décisions des moteurs de ne pas déréférencer certains contenus.

Aujourd'hui (et selon le Guardian : <http://www.theguardian.com/technology/2014/jun/08/google-search-results-indicate-right-to-be-forgotten-censorship>), à la manière du message déjà diffusé en bas de page lors d'une demande légale, Google réfléchit à la possibilité d'indiquer aux internautes que des résultats ont été supprimés sur demande de l'utilisateur et d'en connaître la raison. Il est probable que Google publie également régulièrement les statistiques de telles demandes.

Seul l'avenir et la pratiques nous permettront de valider l'efficacité de ces mesures et la satisfaction des utilisateurs. En toutes hypothèses, à ce stade, le respect par Google et les moteurs en général est un très bon signe d'un meilleur respect de la vie privée des Internaute.

Alexandre Diehl

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

Revue d'Outils Moteurs et SEO

[Retour au sommaire de la lettre](#)

Domaine :	Recherche	Référencement
Niveau :	Pour tous	Avancé

Cette rubrique signale chaque mois 5 outils (site web, logiciel, plug-ins Firefox...) qui nous ont parus intéressants (ou parfois étonnants...) pour le référencement naturel ou la recherche d'information. Des outils pas obligatoirement très récents, mais parfois méconnus ou qui gagnent à être utilisés selon nous. Chaque outil est décrit en quelques phrases, sous la forme d'une fiche signalétique vous permettant de le tester au plus vite...

Nom : Pro Rank Tracker

URL : <https://proranktracker.com/>

Domaine : SEO

Type : Site web

Descriptif : Chaque mois, ou presque, nous vous proposons dans cette rubrique un outil permettant de consulter les classements de votre site sur Google pour vos mots clés favoris. Voici donc celui de juin 2015 :-)



Nom : Analyseur de page web

URL : <http://www.damien-hernandez.com/test/>

Domaine : SEO

Type : Site web

Descriptif : Analyseur / auditeur de site web : contenu, balises meta, vitesse, liens, etc. A tester !

The screenshot shows the homepage of 'Analyseur de page web'. At the top, the title 'Analyseur de page web' is prominently displayed. Below it, a subtitle reads 'Analyseur de page web est un outil gratuit de référencement qui vous aidera à analyser vos pages web'. A search bar contains the text 'example.com' and a green 'Analyse' button. Below the search bar, there are six feature cards arranged in two columns:

- Analyse du contenu:** Icon of a document. Text: 'Voir une analyse de contenu. Vérifiez le ratio texte/HTML, les balises titres (<h1> à <h6>) etc...'
- Test de vitesse:** Icon of a bar chart. Text: 'Accélérer le temps de chargement de votre site web concernant vos pages les plus lentes.'
- Balises méta:** Icon of a cursor pointing to a tag. Text: 'Obtenez la liste complète des méta tags de vos pages Web. Vérifiez le titre du site, les mots-clés, les balises META Open Graph (Og) et bien plus encore.'
- Obtenez des conseils:** Icon of a target symbol. Text: 'Cette application vous affiche automatiquement les vulnérabilités du site et propose des conseils.'
- Extracteur de liens:** Icon of a magnifying glass over a link. Text: 'Aperçu des liens de votre site web (ancres et URLs) et découvrez le pourcentage des liens externes et internes.'
- Evaluation du site:** Icon of a pencil. Text: 'Obtenez une évaluation complète de votre site Web totalement gratuite.'

Nom : Traffic Travis

URL : <http://www.traffictravis.com/>

Domaine : SEO

Type : Logiciel

Descriptif : Un autre analyseur / auditeur de site web : contenu, analyse des backlinks, positionnement, etc. Une suite SEO qui a l'air assez complète.

The screenshot shows the interface of the 'Traffic Travis' software. At the top left is the 'traffic trAVIS' logo. At the top right is a green 'Free Download' button with a download icon. Below the logo is a yellow banner with text: 'Mac Users: We recommend using AffilioTools, our new and improved web-based application for maintaining and monitoring your websites. Check it out here!'. The main heading reads 'Improve your website's search engine rankings.'. Below this is a dashboard with tabs for 'Dashboard', 'SEO', 'PPC', and 'Research'. The 'SEO' tab is active, showing a line graph with the following data points:

Date	Rankings
Jan 7	30
Jan 14	20
Jan 21	15
Jan 28	14
Feb 4	12
Feb 11	10
Feb 18	8
Feb 25	5

In the bottom right corner of the dashboard, there is a circular badge with the 'cnet' logo, five stars, and the text: 'An impressive collection of data, we highly recommend!'.

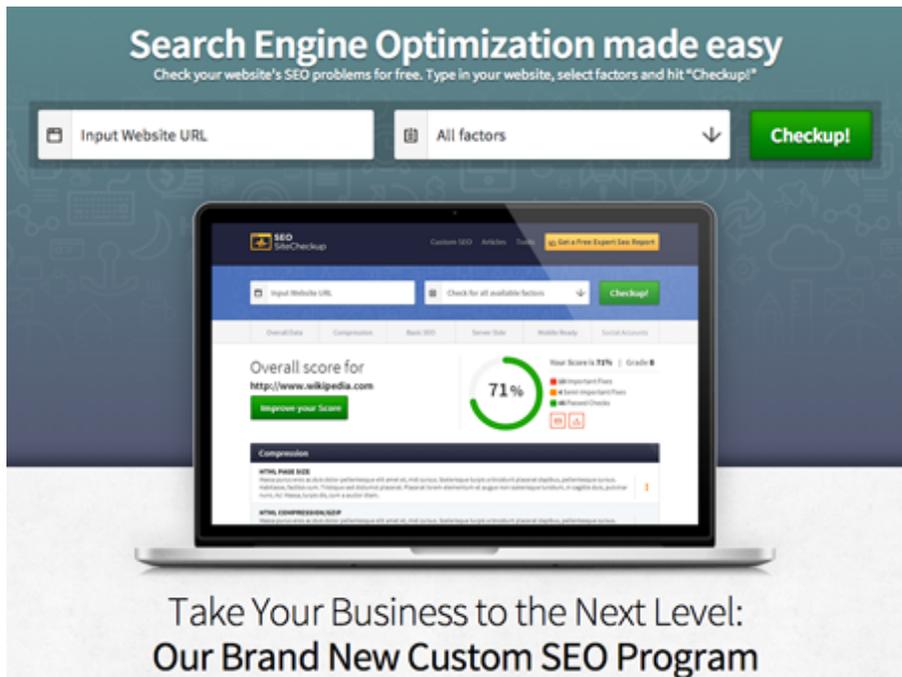
Nom : SEO Site Checkup

URL : <http://seositecheckup.com/>

Domaine : SEO

Type : Site web

Descriptif : C'est le mois des outils d'analyse SEO d'un site web qui semblent être à la mode.



Nom : Google Redirects Fixer

URL : <https://addons.mozilla.org/fr/firefox/addon/google-no-tracking-url/>

Domaine : SEO / Recherche / Veille

Type : Extension Firefox

Descriptif : Permet de supprimer l'URL de redirection rajoutée par Google dans ses résultats de recherche (exemple :

https://www.google.fr/url?sa=t&rct=j&q=&esrc=s&source=web&cd=2&cad=rja&uact=8&ved=0CDIQFjAB&url=http%3A%2F%2Fwww.test.com%2F&ei=B62VU7SjKvLP0AW864HoDw&usg=AFQjCNH21KLjC0CBkjon2DwD_CZ0HApLMw&sig2=FoH35jLxT8Yo7S2JeqxNtA)), très utile lorsqu'on fait souvent des copier/coller d'URL directement depuis les SERP...



Google Redirects Fixer 1.3.1 PAS DE REDÉMARRAGE
par [matagus](#)

Google uses a redirection link to track your clicks. This addon simply removes that redirection and turns every search result in its original link, saving your time and giving you more security and privacy.

Please carefully read the description!

[Télécharger maintenant](#)

Fonctionne avec Firefox 19.0 - 32.0 - [Voir d'autres versions](#)

Rubrique proposée par Olivier Andrieu, éditeur du site Abondance.com.

Revue d'URL

[Retour au sommaire de la lettre](#)

Domaine :	Recherche	Référencement
Niveau :	Pour tous	Avancé

Revue d'URL en français... ■ ■ ■

-> Espionnez mon cocon !

Afin d'illustrer ma formation sur le concept du cocon sémantique, j'ai intégré une zone de test sur mon site. Cela va permettre à ceux qui suivent la formation d'observer ce que je fais quasiment en temps réel. Du coup, vous pouvez aussi aller jeter un coup d'oeil pour suivre mes progrès...

<http://www.laurentbourrelly.com/blog/1653.php>

-> Caractéristiques de l'algo Google Panda 4.0

Le SEO Glenn Gabe a publié une analyse de 41 sites concernés par la sortie de Panda 4.0 vers le 20 mai 2014. Voici un résumé de ce qu'il a observé sur les sites ayant été pénalisés et sur ceux qui ont profité de cette mise à jour...

<http://www.webrankinfo.com/dossiers/conseils/caracteristiques-panda-4-0>

-> Google+ et référencement naturel : mythe ou réalité ?

Votre activité sur Google+ améliore-t-elle votre référencement ? Depuis le lancement de G+ en 2011, de nombreux référenceurs spéculent sur l'impact présumé de ce réseau social sur les classements des sites dans les résultats de recherche Google. Qu'en est-il réellement ? Tour d'horizon d'une question qui fait débat au sein de la communauté SEO...

<http://www.emarketinglicious.fr/social-media/google-referencement-naturel-mythe-realite>

-> Les 10 commandements de l'entretien d'embauche pour un SEO

Le 27 mai dernier se tenait une session de Speed Recruiting organisé par le SEO Camp à Paris. J'étais à la fois co-organisateur et recruteur, je ne suis donc pas forcément très objectif sur la réussite de cet événement mais, en tant que recruteur, j'ai pu découvrir une bonne dizaine de profils dont certains vraiment très intéressants (à suivre...). Toutefois, à force de faire passer des entretiens, j'ai quelques conseils que j'aimerais partager...

<http://www.woptimo.com/2014/06/les-10-commandements-de-lentretien-dembauche-pour-un-seo/>

-> L'authorship est-il réservé à vos pages web ?

Hello à tous, un article qui sera très court (du moins je pense). Je constate jour après jour que beaucoup utilisent l'authorship de manière automatique et non réfléchit en le mettant uniquement en place sur chaque page de leur site...

<http://www.visibilite-referencement.fr/blog/authorship-externe>

-> Le référencement c'est comme un rendez-vous chez le dentiste

Ici et là sur la toile, on retrouve différentes analogies à propos du référencement naturel. Aujourd'hui, nous vous proposons un parallèle singulier entre le référencement naturel et un rendez-vous chez le dentiste...

<http://www.pullseo.com/seo-vs-dentiste/>

-> Conseils référencement : 18,2% des clics vont sur le 1er résultat SEO

En effectuant ma veille d'informations de ce matin, mon logiciel de veille m'a rapatrié une infographie sympathique sur les raisons démontrant que le SEO est primordial pour le business en ligne. Ce qui m'a surtout attiré sur cette illustration, ce sont les statistiques et les pourcentages (qui datent de 2013, mais elles ne doivent pas avoir varier énormément entre temps)...

<http://autoveille.wordpress.com/2014/06/02/conseils-referencement-182-des-clics-vont-sur-le-1er-resultat-seo/>

-> SEO : comment bien désoptimiser

Désoptimiser : une nouvelle tendance du SEO... Qui l'eût cru ? Un métier qui porte dans son nom le terme d'"optimisation" en est donc arrivé à vouloir aujourd'hui brider cette optimisation, ou la cacher le mieux possible. Car elle a pu aller trop loin, vers la "suroptimisation"...

<http://www.journaldunet.com/solutions/seo-referencement/desoptimisation-nouvelle-tendance-du-seo/>

-> Rédiger des balises meta de niveau 3

Google remplace parfois certaines meta balises de votre site dans les SERPS. Ce phénomène s'est amplifié depuis quelques mois, ainsi qu'il a été remarqué sur Axe Net ou JV Web. Etrange, non ?...

<http://www.webmaster-referencement.fr/astuces-meta-balises.html>

-> Quand Google n'a plus confiance en toi

On accuse Panda et Pingouin d'avoir muté dans des pénalités de plus en plus déroutantes pour ceux qui suivent de près l'hécatombe...

<http://www.laurentbourrelly.com/blog/1637.php>

-> Attention au contenu dupliqué par les marketplaces !

Si vous possédez une boutique en ligne, vous vous êtes sans doute posé la question de tenter l'aventure des places de marché ou des comparateurs de prix. Le principe est de proposer vos produits sur des sites qui ont bien plus de visiteurs que le vôtre et donc de bénéficier de leur trafic. Mais attention, le problème du "duplicate content" guette. Et avec lui le risque de voir votre site pénalisé par Google...

<http://blog.axe-net.fr/contenu-duplique-marketplaces/>

-> Quand les moteurs de recherche ne suffisent plus

Dans la jungle d'informations qu'est le web, les simulateurs ou comparateurs permettent de hiérarchiser les solutions les plus adaptées aux besoins du consommateur. D'après le Figaro, ces outils, plébiscités par les internautes, connaissent un boum sur la toile, ils reposent néanmoins sur une relation de confiance...

<http://www.journaldunet.com/ebusiness/expert/57498/quand-les-moteurs-de-recherche-ne-suffisent-plus.shtml>

-> Il était une fois la balise H1

Sacrée balise H1 qui se rêvait sur un piédestal et qui ne fût finalement qu'un géant aux pieds d'argile. Tragédie grecque ou drame moderne ? A l'heure où tout un chacun essaie de se faire une place tout en haut des SERPs, la relative facilité à utiliser des raccourcis, à de beaux jours devant elle. C'est donc l'histoire d'une balise H1 qui se voyait déjà en haut de l'affiche...

<http://www.simplewebsite.fr/blog/referencement/il-etait-fois-balise-h1/>

-> Cocon Sémantique, le système complet !

Vous savez que le cocon sémantique est ma marque de fabrique. Sans doute parce que j'ai propagé et affiné ce concept depuis de nombreuses années. Finalement, j'ai décidé de lâcher ma méthode comme jamais auparavant, même lors des prestations de consulting...

<http://www.laurentbourrelly.com/blog/1648.php>

-> Google Panda a-t-il impacté les sites de communiqués de presse ?

Il y a une bonne dizaine de jours, nous vous indiquions qu'une mise à jour de Google Panda avait été lancée par Google. Intitulée Google Panda 4.0, cette mise à jour avait pour but de filtrer les contenus de faible qualité des résultats de recherche de Google. Et il semblerait que des sites de communiqués de presse aient également été touchés...

<http://www.referenceur.be/google-panda-impacte-sites-communiques-de-presse-12793.html>

-> Retour sur le WebCampDay d'Angers

Samedi 24 mai 2014 se tenait le Web Camp Day d'Angers (49), une journée axée autour de deux cycles de conférences en webmarketing et référencement soutenue par

l'association SEO Camp (présidée depuis peu par Hasni Khabeb qui nous avait fait l'honneur de sa présence !)...

<http://blog.internet-formation.fr/2014/05/retour-sur-le-webcampday-dangers-webcampday/>

-> 6 mensonges SEO et SEM qui ont la dent dure...

Je ne suis pas du genre à rédiger ce type d'articles que je considère parfois comme « racoleur » et je m'en excuse par avance... Il faut dire que les référenceurs ont pris conscience du pouvoir qui était le leur face aux clients et webmasters désemparés et en quête de visibilité. Par conséquent, cet article est à la fois une ode à la lutte contre les thèses si souvent défendues et pourtant démenties ou non prouvées...

<http://blog.internet-formation.fr/2014/05/6-mensonges-seo-et-sem-qui-ont-la-dent-dure/>

-> Comment Google nous impose son « Plus » ?

Et bien voilà, l'article touche à son début. Dans cet article, je vais tenter de revenir sur la stratégie mise en place par Google pour imposer sa plateforme. Cet article se veut la suite d'un de mes articles : La stratégie « social » de Google (25 juillet 2011). L'article est à lire pour mieux comprendre le contexte...

<http://www.nouslesgeeks.fr/2014/05/27/comment-google-nous-impose-son-plus/>

-> Non, vos visiteurs n'arrivent pas par votre page d'accueil !

Si parfois, je l'admets bien volontiers, certains de vos visiteurs découvrent votre site à partir de la page d'accueil. Mais si votre site est assez conséquent, si vous proposez plusieurs services ou produits, et si le travail de référencement a été correctement effectué, une bonne part de vos visiteurs ne découvrira pas votre site à partir de la homepage...

<http://blog.axe-net.fr/non-vos-visiteurs-narrivent-pas-par-votre-page-daccueil/>

-> Y a t'il une pénalité pour contenu dupliqué ?

J'entends encore tout et son contraire à propos du contenu dupliqué. C'est le méga marronnier du référencement, qui trône dans notre jardin depuis une éternité...

<http://www.laurentbourrelly.com/blog/1621.php>

-> 70 + Techniques Détaillées de Netlinking

J'avais déjà écrit un article sur ce même sujet, il y a deux ans. Beaucoup de gens l'avaient lu et avaient participé aux discussions qui ont pu suivre (j'y reviendrai probablement dans un de mes futurs posts). La rédaction d'un tel guide est aussi passionnante que longue...

<http://refeo.com/blog/seo/netlinking/>

-> Comment s'assurer qu'un projet SEO est réaliste

Vous avez un projet de référencement naturel en tête, un besoin de visibilité pour un futur site ou un site existant. La première question qui vient alors à l'esprit est de savoir si ce projet de référencement est à la fois réaliste et rentable. Cette question se pose à tous les porteurs de projets, mais aussi à tous les gestionnaires de sites et aux équipes dirigeantes. Pour tenter de répondre à cette question plus fréquente qu'il n'y paraît, je vous propose une réflexion structurée en plusieurs points qui peut être reprise pour n'importe quel cas concret...

<http://www.webmarketing-com.com/2014/05/21/28015-comment-sassurer-quun-projet-seo-realiste>

-> Le cocon sémantique, l'arme fatale du SEO ?

Je suis ravi et désolé à la fois que les notions « on site » de glissement sémantique et renforcement du maillage interne soient à la mode. Ravi parce que le cocon sémantique est mon dada. Désolé car il a fallu attendre l'arrivée des pénalités type Panda et Pingouin pour que l'importance cruciale de ce élément soit mis sur le devant de la scène...

<http://www.laurentbourrelly.com/blog/1631.php>

-> Comment Google a tué l'essence du Web

Lorsqu'internet a commencé à devenir un outil grand public les moteurs de recherche étaient nettement moins pertinents que les annuaires pour nous guider dans la découverte de nouveaux sites mais avaient en revanche beaucoup plus de réponses à nous fournir...

<http://www.webnotoriete.com/articles/comment-google-tue-lessence-du-web/>

-> Quels titres bénéficient du fonds pour la presse créé par Google ?

Créé en septembre dernier, le fonds d'aide à l'innovation numérique de la presse de Google a versé 16,4 millions d'euros aux médias français en 2013. Mais lesquels et pour quels projets ?...

<http://www.journaldunet.com/media/publishers/beneficiaires-du-fonds-google-aipg-pour-l-innovation-numerique-de-la-presse-en-2013.shtml>

-> L'évolution des parts de marché des moteurs de recherche

Comme tous les mois, ComScore a dévoilé les parts de marché des moteurs de recherche aux Etats-Unis. Il n'en ressort rien de très surprenant si ce n'est que Google continue de progresser alors que les sites de Yahoo continuent de chuter...

<http://www.referenceur.be/evolution-parts-de-marche-moteurs-de-recherche-12648.html>

-> La liste Gloubiboulga SEO 2014

Voici en exclusivité pour vous la liste Gloubiboulga SEO ! Elle présente 125 blogs et sites SEO à suivre pour une veille en référencement efficace. Téléchargez ce document !...

<http://www.webmaster-referencement.fr/liste-referencieurs-veille.html>

-> 5 raisons d'utiliser les microdatas sur votre site

Nous avons déjà présenté l'étendue des microdatas dans un article précédent. Depuis, de nouvelles propriétés sont apparues sur le site schema.org pour structurer encore plus d'éléments sur votre site...

<http://www.pullseo.com/pourquoi-utiliser-les-microdatas-sur-votre-site/>

-> Les graphes du web, structures en SEO, opus 2

Nous continuons sur les structures en SEO, et allons faire de la « géographie thématique ». Nous allons donc nous promener dans le(s) graphe(s) du web et terminer par une introduction au graphe de votre site. Une boussole et on y va !...

<http://www.referencement-naturel-white-hat.fr/les-graphes-du-web-structures-en-seo-opus-2/>

-> Combiner des mots-clés facilement avec Excel 2010

Que ce soit dans le cadre de la gestion de campagnes Adwords ou de l'optimisation de son référencement, le besoin de combiner des mots-clés revient assez régulièrement. Des outils dédiés existent pour cette tâche, mais il faut une licence dans le cas des logiciels, et une connexion pour les outils en ligne. La méthode que je vous propose utilise Excel, dans sa version 2010...

<http://www.sebastien-billard.fr/seo/post/Combiner-des-mots-cles-facilement-avec-Excel-2010-14-05-2014>

-> 10 choses impressionnantes que vous ne savez pas à propos de Google !

S'il existe de nombreux moteurs de recherche disponibles sur le web (dont certains n'ont sans doute pas la notoriété qu'ils méritent), il est difficile d'imaginer le web aujourd'hui sans penser à Google. Mais est-ce que vous savez tout sur ce moteur de recherche qui rythme nos journées depuis maintenant plus de 15 ans ?...

<http://www.referenceur.be/10-choses-impressionnantes-google-12609.html>

-> Guide des bonnes pratiques sur Google+

Après seulement 3 ans d'existence, Google+ a su convaincre plus de 359 millions d'utilisateurs actifs. Bien souvent sous-exploité, Google+ se révèle une formidable opportunité de travailler sa visibilité au moment où Facebook demande un budget de plus en plus conséquent pour maintenir une bonne visibilité...

<http://www.webmarketing-com.com/2014/05/19/28214-guide-bonnes-pratiques-google>

-> La Méta Keywords a un impact sur Baidu, Yandex et Naver

J'ai décidé de publier ce petit article sur le poids de la balise méta Keywords afin d'éclaircir un peu tout ce que l'on peut lire sur elle. En effectuant ma veille informationnelle avec mon outil de veille, j'ai constaté que de nombreux SEO estiment que la méta Keywords n'a pas de poids du tout pour l'ensemble des moteurs de recherche...

<http://autoveille.wordpress.com/2014/05/21/la-meta-keywords-a-un-impact-sur-baidu-yandex-et-naver/>

-> La balise de vos pages web - les fondamentaux

Référencement, réseaux sociaux, interopérabilité et performance web sont autant de sujets qui sont à prendre en compte à travers la balise head de vos pages web, découvrons ensemble les fondamentaux. (partie 1/2)...

<http://www.journaldunet.com/developpeur/expert/57640/la-balise--de-vos-pages-web---les-fondamentaux.shtml>

-> Quand l'outil de planification Adwords vous cache des mots-clés

L'outil de planification de Google Adwords est l'une des sources d'inspiration favorite des SEO. Un point trop rarement évoqué cependant est le fait que Google vous cache des mots-clés alors même que celui-ci possède bien des données à leur sujet...

<http://www.sebastien-billard.fr/seo/post/Quand-l-outil-de-planification-Adwords-vous-cache-des-mots-cles-11-06-2014>

-> Pourquoi faut-il désindexer certaines pages de son site Internet ?

Si vous ne faites pas attention, les moteurs de recherche analyseront l'intégralité de votre site et donc indexeront des centaines de milliers de pages sans intérêt. Il faut donc empêcher l'indexation de ses pages et quand il est trop tard, il faut les désindexer afin d'éviter une pénalité. Lors de mes prestations en référencement naturel, il m'arrive trop souvent de devoir désindexer des milliers de pages web...

<http://www.rc2i.net/conseil-en-referencement/supprimer-du-contenu/>

Revue d'URL en anglais...

-> The Hidden Power of Nofollow Links

A few years ago, while I was still on the client side of things, I received an email from a blogger I was working with. As part of our fledgling link building program, my company had been sending out free products in exchange for a review and link to our site. Oldest trick in the book, right? However, the blogger's email threw me off: she told me her policy was to nofollow links, and asked if this would be all right...

<http://moz.com/blog/the-hidden-power-of-nofollow-links>

-> Ten Timely Tips for Terrific Title Tags

Google's most recent results page layout changes meant more than just a search experience update. With the change came adjustments to how each individual result is displayed, which may almost certainly necessitate a refresh for a core search engine optimization (SEO) element: the venerable title tag...

<http://searchenginewatch.com/article/2348627/Ten-Timely-Tips-for-Terrific-Title-Tags>

-> Finding entity names in Google's Knowledge Graph

Most of us searchers and site owners and search engine optimizers are familiar with Google's Link Graph, and how Google uses the connections between websites to help in ranking pages on the Web. In part, Google looks at the relevance of the content of a page compared to a query that a searcher enters at the search engine...

<http://www.seobythesea.com/2014/06/entity-names-in-google/>

-> SEOs: Negative SEO On Google Much Easier Now

As you all know, negative SEO is not new, in fact, Google has said it is rare but possible since 2007. Sites as large as Expedia may have suffered from it and Google had to reword their documentation on the topic...

<http://www.seroundtable.com/google-negative-seo-easy-18655.html>

-> How Google might identify synonyms for entities using anchor text

When Google indexes the Web, it's often been convenient to think about the search engine running two different methods or approaches that seem to run in parallel. One of those involves the crawling and indexing and ranking of pages on the web (and images, videos, news, podcasts, and other documents)...

<http://www.seobythesea.com/2014/06/synonyms-for-entities/>

-> More Panda 4.0 Findings: Syndication, User Engagement, Indexation & Keyword Hoarding

Now that the dust has settled, it's clear that Panda 4.0 was a powerful update that impacted many sites across the web. And that impact has been fascinating to analyze...

<http://searchenginewatch.com/article/2348110/More-Panda-4.0-Findings-Syndication-User-Engagement-Indexation-Keyword-Hoarding>

-> A Recipe for a Tasty Semantic Search & SEO Salad

It seems that there's a lot of misunderstanding surrounding the way that search engine optimization (SEO) and semantic search play together. They aren't totally separate concepts!...

<http://searchenginewatch.com/article/2347859/A-Recipe-for-a-Tasty-Semantic-Search-SEO-Salad>

-> Backlinks = Rankings, Rankings = Traffic – Deal With It

Link building, done correctly, is hard work. It's laborious and filled with lots of rejection. In many ways, it's like telemarketing – nobody likes it, but it pays off. In fact, link building still works better than anything else to boost organic rankings...

<http://searchenginewatch.com/article/2347618/Backlinks-Rankings-Rankings-Traffic-Deal-With-It>

-> 4 Steps to Maximize Local Search Success

"Four out of five consumers say they want search ads to be customized to their city, ZIP code or immediate surroundings," according to new research from Google highlighting the importance of location...

<http://searchenginewatch.com/article/2346905/4-Steps-to-Maximize-Local-Search-Success>

-> Panda 4.0, Payday Loan 2.0 & eBay's Very Bad Day

After a period of relative quiet, MozCast detected a major "temperature" spike in Google's algorithm at some point on Monday, May 19th. This occurred after some historic lows, including the 3rd coldest day on record (May 11th)...

<http://moz.com/blog/panda-4-payday-loan-2-and-ebays-very-bad-day>

-> Negative SEO From Links — What Can You Do If You're Hit?

In this post, we cover the basics of negative search engine optimization (SEO) and look at what you can do to fight negative SEO link spam. This being a hot topic in the industry, I encourage readers to comment at the end of the article...

<http://searchengineland.com/negative-seo-links-can-191571>

-> Search Engine Indexation Analysis: A Simple Guide

Ever wondered about all the content you have floating around Google's organic search index? We're often consumed with analyzing the traffic we receive and the hopes of traffic we don't have yet to remember we are likely getting some impressions just outside the realm of mass search volume...

<http://searchenginewatch.com/article/2345548/Search-Engine-Indexation-Analysis-A-Simple-Guide>

-> Google: If Content Isn't Changed, Manual Actions Won't Be Removed

An interesting Google Webmaster Help thread has a webmaster who has a manual action, he said he removed the content, and even left the page blank but the manual action won't go away...

<http://www.seroundtable.com/google-manual-action-page-removal-18561.html>

-> The Ultimate Entrepreneur's SEO Checklist for 2014

For entrepreneurs looking to swim with the SEO sharks; think of this business rule: Always ask yourself how someone could preempt your products or service. How can they put you out of business? Is it price? Is it service? Is it SEO? Is it social?...

<http://searchenginewatch.com/article/2344630/The-Ultimate-Entrepreneurs-SEO-Checklist-for-2014>

-> Were You Really Hit By A Search Engine Penalty, Or Is It Something Else?

Often, I'll speak to prospective clients who think they've been hit by an algorithm update or a penalty. They've lost a lot of organic traffic, so they immediately think that a search engine change is to blame...

<http://searchengineland.com/really-hit-penalty-something-else-190618>

-> Webmaster: I've Tested Negative SEO Through Links & It Works

Last week we reported that most webmasters are claiming negative SEO is easier now than ever. If you look at the conversation there, you will see it is somewhat of a hot topic...

<http://www.seroundtable.com/negative-seo-steps-18674.html>

-> A Marketer's Guide To Using Regular Expressions In SEO

Regular expressions (regex) are one of the most powerful tools we have in our SEO arsenal, but they're incredibly intimidating! Here are some tips and tricks from one SEO to another that I hope will help you dip your toes into the powerful world of regex...

<http://searchengineland.com/using-regular-expressions-seo-193146>

-> 10 Questions To Ask When Creating A Link Building Campaign

Whether in a blog post or a conference presentation, link building experts often focus on specific tactics when giving advice. The message can sometimes come across as, "Here are some things — go do them, and you'll be set."...

<http://searchengineland.com/10-questions-ask-creating-link-building-campaign-193392>

-> My Site Is Boring. How Do I Get Links?

The topic of getting links to boring web sites or web sites that compete with dozens or hundreds of other web sites is not new. If you sell widgets and there are hundreds of other sites that sell the same widgets, then how do you drive up interest to get people to link to your site?...

<http://www.seroundtable.com/link-bait-boring-18681.html>

-> Your Google Algorithm Cheat Sheet: Panda, Penguin, and Hummingbird

If you're reading the Moz blog, then you probably have a decent understanding of Google and its algorithm changes. However, there is probably a good percentage of the Moz audience that is still confused about the effects that Panda, Penguin, and Hummingbird can have on your site. I did write a post last year about the main differences between Penguin and a Manual Unnatural Links Penalty, and if you haven't read that, it'll give you a good primer...

<http://moz.com/blog/google-algorithm-cheat-sheet-panda-penguin-hummingbird>

-> Bing Sitemaps Best Practices & Large Site Sitemaps Issues

Bing has posted on the Bing Webmaster blog their best practices on how to implement XML Sitemaps files for your site. The post includes a large topic on how the best practices apply to really large web sites, with many URLs...

<http://searchengineland.com/bing-sitemaps-best-practices-large-site-sitemaps-issues-193679>

Rubrique proposée par Olivier Andrieu, éditeur du site Abondance.com.

Merci pour votre lecture...

Pour toute suggestion au sujet de cette lettre professionnelle : olivier@abondance.com