

**Lettre professionnelle
Recherche et Référencement**

Numéro 155

Janvier 2014



**LE SITE ABONDANCE
VOUS SOUHAITE
UNE BONNE ANNEE 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 155 - Janvier 2014

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

Au sommaire ce mois-ci :

- > [Comment les moteurs de recherche détectent-ils le content spinning ?](#) (5 pages)
- > [Content Gems, futur concurrent de Google Alertes ?](#) (7 pages)
- > [Fin du générateur de mots clés de Google : quels autres outils que le planificateur utiliser ?](#) (14 pages)
- > [WDF*IDF : la "formule magique" des référenceurs allemands](#) (9 pages)
- > [L'IP Tracking sous les feux de la critique](#) (3 pages)

- > [Revue d'outils moteurs et SEO](#) (5 pages)
- > [Revue d'URL](#) (9 pages)

Prochaine parution : Lundi 17 février 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 155 - Janvier 2014

Comment les moteurs de recherche détectent-ils le content spinning ?	4
Content Gems, futur concurrent de Google Alertes ?	9
Fin du générateur de mots clés de Google : quels autres outils que le planificateur utiliser ?	16
WDF*IDF : la "formule magique" des référenceurs allemands	30
L'IP Tracking sous les feux de la critique	39
Revue d'Outils Moteurs et SEO	42
Revue d'URL	46

Comment les moteurs de recherche détectent-ils le content spinning ?

[Retour au sommaire de la lettre](#)

Domaine :	Recherche	Référencement
Niveau :	Pour tous	Avancé

Le content spinning, déjà évoqué dans nos colonnes, est une méthode utilisée par de nombreux référenceurs pour obtenir des textes pertinents, vraisemblables et originaux pour leur site web. Mais il peut s'agir de méthodes dangereuses si elles sont détectées par les moteurs de recherche (Google a d'ailleurs clairement indiqué dans le passé que le spinning était interdit). Mais comment font ces moteurs pour détecter le content spinning ? Retour sur les travaux de trois chercheurs américains à ce sujet...

Chers lecteurs et lectrices de la lettre R&R, vous savez très certainement déjà ce qu'est le spin, car plusieurs articles vous en ont parlé par le passé (voir articles de la lettre d'avril et mai 2013). Nous n'allons donc pas vous faire l'affront de vous le (ré)expliquer en détail. Cependant, pour s'assurer que nous parlons très exactement de la même chose, nous allons dans un premier temps définir le vocabulaire que nous allons utiliser dans cet article.

Pour de nombreuses tâches liées à nos métiers de référenceurs ou webmaster, nous avons besoin de disposer de contenus textuels. Que ce soit pour faire de l'acquisition de liens et ainsi apporter de l'information contextuelle, ou tout simplement agrandir facilement la taille d'un site web, ou bien encore pour créer de la valeur ajoutée sur des pages à faible intérêt (listing, catégories), on a toujours besoin de contenus, parfois même de plus en plus de contenus.

Disposer de nombreux textes est donc un atout non négligeable dans une optique de référencement. Mais pas n'importe quels textes ! En effet, ceux-ci doivent pouvoir répondre à quelques exigences fondamentales :

- **être pertinents** pour les requêtes ciblées (n'oublions pas que l'enjeu final est d'obtenir un bon positionnement) ;
- **être vraisemblables** (sinon on parle de spam) ;
- **être originaux** (les textes dupliqués ou quasi-dupliqués peuvent être détectés et donc pénalisés, cf. lettre R&R de décembre 2013).

Pour obtenir ces textes de façon peu onéreuse et rapide, trois options sont généralement possibles :

- **Acheter du contenu offshore.** La qualité est variable et selon le pays où sont sous-traités les textes, on peut avoir des soucis de ciblage (certains termes ont des fréquences d'usage différents, certaines expressions n'existent pas ou au contraire sont sur-utilisées).
- **Copier le contenu d'un site tiers.** Nous l'avons vu, ce n'est pas une option si l'on souhaite éviter d'être repéré par les filtres de détection des contenus dupliqués ou quasi-dupliqués.
- **Utiliser des méthodes de réécriture de contenu.** Le *content spinning* est une technique de réécriture de contenu qui vise à remplacer des mots d'un texte ou à modifier la structure de ce texte (ou les deux à la fois) pour créer un texte différent du premier, mais qui porte la même information.

A l'heure actuelle, il n'existe pas d'algorithme, et par conséquent pas d'outil logiciel, capable de créer à partir d'un texte libre - écrit en langage naturel - ce que l'on appelle un **masterspin**. Un masterspin, c'est une expression qui permet de générer de nombreux textes différents, qu'on appelle des **spuns**, via un outil de génération automatique (ce que l'on désigne sous le terme de **spinner**).

Un masterspin ressemble beaucoup à une expression régulière. On peut y décrire des alternatives et des substitutions à l'aide d'opérateurs spécifiques. Pour cela on utilise les

acolades { et }, ainsi que le *pipe* |. Par exemple, l'expression suivante est un masterspin qui peut générer 4 phrases différentes :

La {loutre | pie} est {joyeuse | voleuse}.

En effet, ce masterspin définit les phrases suivantes : *la loutre est joyeuse, la loutre est voleuse, la pie est joyeuse et la pie est voleuse*. On peut construire des masterspins complexes en imbriquant les opérateurs. Ces masterspins permettent alors la génération de très nombreux textes, avec une variabilité très grande entre les textes.

On voit de suite que la qualité du processus de content spinning est directement dépendante de la qualité du masterspin qui a été mis en place pour la génération. Ecrire un bon masterspin est devenu un vrai métier, et ceux qui ne savent pas le faire et ne souhaitent pas faire appel à un professionnel pour cela n'obtiennent jamais de bon résultats !

Afin d'aider les plus allergiques aux masterspins, certains outils de content spinning proposent de prendre en entrée des textes en langage naturel et d'opérer de simples substitutions des mots par des synonymes pour ainsi créer des textes capables de passer aux travers des outils de détection de contenu dupliqué.

Même si ce type de fonctionnalité ne donne que des résultats médiocres, il devient alors possible à l'amateur le plus complet d'accéder à une méthode de génération automatique de contenu.

Naturellement, plus les techniques de content spinning sont utilisées, plus il devient indispensable pour les moteurs de recherche d'être capable de détecter leur utilisation. Zhang, Wang et Voelker (*voir sources à la fin de l'article*) annoncent ainsi que sur environ 430 000 pages de sites de type wiki qu'ils ont étudié, 52% contenaient des spuns.

Sur le plan de la détection, il n'y a pas eu beaucoup d'avancées ces dernières années. La plupart du temps, il n'est de toute façon pas nécessaire de déterminer si un texte est un spun pour pénaliser la page qui le contient. En effet, dans la grande majorité des cas, les pages sur lesquelles sont mis en place les spuns sont de toute façon de très faible qualité, et sont donc pénalisées pour d'autres raisons (profil de liens abusif, site ayant toutes les caractéristiques du spam, etc.).

Cependant, le problème de la détection des spuns sur le web est devenu un sujet de choix pour certaines équipes de recherche. D'abord pour la beauté du sport, car le sujet en lui-même est un vrai challenge algorithmique. Mais aussi car, au delà de la détection des spuns, se profile le problème de leur source.

Dans ce cas, c'est le moteur de recherche qui y a intérêt. S'il sait déterminer si un texte est un spun ou un véritable texte, s'il est capable de regrouper les spuns générés à partir d'un même spinner ou d'un même masterspin, voire même par un même référencier, il peut alors mettre en place des contre-mesures efficace pour pénaliser des réseaux, des outils et des personnes.

L'article scientifique de Zhang, Wang et Voelker est le premier, à notre connaissance, à s'intéresser aux problèmes de la détection de spuns. Même s'il ne vise que la détection des spuns obtenus par la seule substitution de synonymes, il a le mérite de montrer que la détection est possible pour les méthodes de spinning les plus basiques. Bien sûr, les techniques évoluées passent encore largement au travers des mailles du filet, ce qui fait que les bons spinneurs ont encore de beaux jours devant eux !

Dans la suite de cet article, nous allons vous exposer les travaux de Zhang, Wang et Voelker, ce qui vous permettra de bien comprendre comment on peut détecter du spun simple. Tous les trois sont en poste à l'Université de Californie à San Diego (UCSD).

Une cible : The Best Spinner

L'approche des trois chercheurs vise avant tout à détecter le contenu généré avec *The Best Spinner* (un logiciel du commerce vendu à l'adresse <http://thebestspinner.com> et qu'on notera TBS à partir de maintenant). TBS permet de générer, à partir d'un texte qu'on lui fournit, des centaines de variantes de ce texte, toutes différentes mais toutes globalement similaires sémantiquement. En pratique, il est considéré que TBS prend en entrée un texte et le transforme en *masterspin* en repérant des blocs de mots qui peuvent être remplacés par des groupes de mots de sens similaires (synonymes, formes conjuguées, etc.).

Pour comprendre le fonctionnement de TBS, les chercheurs ont mis en place du *reverse engineering* pur et dur. Plus précisément, ils ont découvert que TBS télécharge le dictionnaire des synonymes et le sauve localement dans un format crypté de manière très basique. Ils ont ensuite désassemblé le binaire pour comprendre le mécanisme de connexion au serveur qui fournit le dictionnaire, ce qui leur a permis d'en obtenir une version lisible. La version téléchargée en août 2013 contient 750 114 synonymes groupés sur 92 386 lignes. Chaque ligne contient un mot, et les synonymes de ce mot. Il faut noter que le dictionnaire n'est pas transitif : si le mot XXX a pour synonyme YYY, et que ZZZ a pour synonyme XXX, cela ne veut pas dire que TBS va considérer que ZZZ et YYY sont synonymes.

Des méthodes qui ne fonctionnent pas

Dans un premier temps, il est intéressant de voir quelles sont les méthodes qui ne permettent absolument pas de détecter les spuns, même lorsque la stratégie de génération est la plus basique qui soit (remplacement des mots par des synonymes). La première méthode sans résultat est, étonnamment, l'algorithme des *shingles* que nous avons abordé dans la précédente lettre R&R. Toutes les expériences menées sur ce sujet montrent que les textes issus du spin ont les mêmes distances de Jaccard entre eux que des textes qui ne sont pas issus du spin. Cela prouve une chose : le spin, qui est utilisé pour passer à travers les méthodes de détection de contenu dupliqué, remplit parfaitement son rôle d'origine.

La deuxième méthode qui ne fonctionne pas consiste à analyser la structure même des phrases, en faisant ce que l'on appelle l'étiquetage morpho-syntaxique du texte. Sous un nom qui paraît ésotérique se cache une technique simple : on va associer aux mots les informations grammaticales qui leurs sont associés, comme le genre, le nombre, le fait d'être un adjectif, un verbe, etc. (notons que l'étiquetage morpho-syntaxique est parfois compliqué, comme l'illustre la phrase « l'homme ferme la porte » : le terme « ferme » est-il un adjectif ou un nom commun ? S'agit-il d'un homme, qualifié de ferme, qui porte une femme, ou bien un homme qui fait l'action de ferme la porte ?).

On pourrait croire que cette technique va permettre de différencier les spuns car lorsque l'on spinne un *masterspin*, il va y avoir des mots qui vont être remplacés par d'autres avec un étiquetage morpho-syntaxique différent (car TBS ne fait pas cette opération d'étiquetage). Même si ceci est exact, cela ne permet absolument pas de différencier nettement les textes légitimes des spuns comme le montre les expériences réalisées par l'équipe de UCSD.

La question est donc toujours la même : comment faire pour détecter efficacement les spuns produits par TBS ?

Une méthode qui fonctionne, mais qui est spécifique à un outil

L'idée de la méthode efficace proposée par Zhang, Wang et Voelker est basée sur la connaissance du dictionnaire de synonymes dont nous vous avons parlé plus haut. Une fois que l'on connaît les mots et motifs qui sont susceptibles d'être remplacés et/ou de remplacer des parties du texte, on peut classer tous les mots des textes que l'on analyse dans deux catégories. La première est celles des **mutables**, il s'agit des mots qui sont

dans le dictionnaire utilisé par TBS. La seconde est celle des **immuables**. Les immuables sont les mots qui ne sont pas modifiés par l'outil de spinning. N'étant pas modifiés, ils sont des entités des textes idéales pour différencier les spuns des contenus légitimes.

D'un point de vue algorithmique, la méthode devient alors très simple. Elle sera similaire à la méthode des shingles mais sans shingles. A leur place, on va utiliser les immuables comme représentants des textes en cours d'analyse.

Pour résumer la méthode, voici comment procéder pour comparer deux textes et déterminer si il s'agit de deux spuns issus d'un même masterspin :

1. On calcule les ensembles des immuables de chaque texte. On note A l'ensemble des immuables de l'un, et B l'ensemble de ceux de l'autre.
2. On calcule l'intersection et l'union de A et B.
3. La similarité de spin entre A et B est donné par le coefficient de Jaccard de A et B pour les immuables, c'est-à-dire par la quantité fournie par l'opération :

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

Pour être tout à fait correct, la méthode employée par les trois chercheurs n'est pas réellement un calcul de coefficient de Jaccard, notamment parce que l'approche est gloutonne. Qu'est-ce que cela veut dire ? Que chaque texte est traité mot par mot, et donc que le calcul des immuables peut être en partie incorrect. Par exemple, si le texte est « *la loutre agile fait la sieste* » et que la réécriture se fait de deux manières possibles avec dans l'une « *la loutre* » qui devient « *la petite bête* » et dans l'autre « *la loutre agile* » devient « *le leste animal* », alors la méthode de détection ne verra que la première et considérera « *agile* » comme un immuable vu que « *loutre* » sera supprimé car présent dans le dictionnaire.

Cependant, malgré ce petit bémol, la technique donne selon les auteurs de très bons résultats. Le score obtenu dans les expériences présentées dans l'article par la méthode des immuables varie de 75% à 90% selon les cas, là où la méthode des shingles va de 20% à 30%, et celle de l'étiquetage morfo-syntaxique va de 20% à 40%.

Les principaux défauts de la méthode sont assez clairs. Nous avons déjà évoqué le fait qu'il ne s'agit pas d'un vrai coefficient de Jaccard. Cela pourrait être gênant dans certains cas pathologiques de textes contenant uniquement des mots qui sont dans des groupes de mots spinnables. Le mot « *loutre* » dans l'exemple précédent est l'illustration de ce type de mot. Mais l'autre défaut, qui est de taille, est qu'il faut connaître le dictionnaire utilisé par l'outil pour appliquer la méthode.

Dans le cas présenté ici, TBS ne protège pas bien son dictionnaire, ce qui fait qu'on peut facilement calculer les ensembles d'immuables. Mais rien ne permet de dire que cela sera ainsi dans le futur. Par ailleurs, TBS peut mettre en place une contre-mesure en modifiant substantiellement et souvent son dictionnaire. Les auteurs ont d'ailleurs étudié l'évolution dans le temps du dictionnaire et ont montré qu'entre deux versions du dictionnaire, 94% de son contenu était resté inchangé, ce qui permettait de conserver des résultats de détection très corrects.

On remarquera que dans le cas d'outils online, le dictionnaire devient inaccessible totalement, et le seul moyen de retrouver l'information pour la détection est de faire de l'attaque active sur l'outil. C'est-à-dire qu'il faut l'utiliser pour générer des spuns de textes bien choisis, pour découvrir le contenu du dictionnaire de synonymes avec un processus d'essai/erreur. La détection devient alors de fait particulièrement coûteuse.

Conclusion

L'article de Zhang, Wang et Voelker présente bien d'autres aspects de la détection, notamment pour l'accélérer quand il faut analyser des milliers de textes, choses que nous n'aborderons pas ici afin de ne pas complexifier notre propos.

Ce qu'il faut retenir : le spin de base est assez facile à détecter quand on détient le dictionnaire servant à faire la génération de textes. Ce dictionnaire est donc le nerf de la guerre et nous encourageons les référenceurs à créer le leur ! Nous encourageons aussi les créateurs d'outils à faire du spin plus évolué, et à protéger leurs bases en fournissant le service de manière déportée, online.

Sources

DSpin: Detecting Automatically Spun Content on the Web :

<http://www.cs.ucsd.edu/~voelker/pubs/dspin-ndss14.pdf>

Qing Zhang (<http://cseweb.ucsd.edu/~q5zhang/>), David Wang (<http://cseweb.ucsd.edu/users/dywang/>)

Proceedings of the Network and Distributed System Security Symposium (NDSS) :

<http://www.internetsociety.org/events/ndss-symposium-2014>

Geoffrey M. Voelker (<http://cseweb.ucsd.edu/~voelker/>), San Diego, CA, February 2014)

The Best Spinner :

<http://thebestspinner.com>

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.

Content Gems, futur concurrent de Google Alertes ?

[Retour au sommaire de la lettre](#)

Domaine :	Recherche	Référencement
Niveau :	Pour tous	Avancé

Le service Google Alertes, qui envoie des mails de surveillance du contenu du Web lorsque certains textes contiennent des mots clés donnés, est un incontournable d'un service de veille. Mais il existe d'autres services qui peuvent également se montrer intéressant dans cette optique. Voici notamment une description de ContentGems qui, malgré quelques défauts, vaut certainement la peine d'être testé...

Les services d'alerte par mots-clés dont Google Alertes (<http://www.google.fr/alerts>) est le représentant emblématique sont légion sur le web. En avril 2013 nous en testions d'ailleurs plus d'une vingtaine autour de 15 critères comparatifs (<http://socialcompare.com/fr/comparison/services-gratuits-d-alertes-par-mots-cles-1pp8mohd>). Cette étude nous permettait de conclure qu'il serait difficile de remplacer Google Alertes, s'il venait à disparaître, pour les raisons suivantes :

- La puissance des robots d'indexation de Google et la taille de son index qu'aucun autre acteur n'est prêt d'égaliser ;
- La possibilité de faire des requêtes "complexes" et de les transformer en alertes (par ex : [filetype:pdf intitle:"intelligence économique" site:gouv.fr](#) pour voir arriver des documents PDF titrant "intelligence économique" et publiés sur des sites du gouvernement français) ;
- Sa capacité à s'adapter à toutes les langues et le fait qu'il n'est pas centré sur un espace géographique spécifique (à savoir les Etats-Unis pour beaucoup de ces services).

Certains se sont pourtant engouffrés dans cette voie. Par exemple, Intigi, une société qui s'est créée en 2011 et dont le service éponyme a été lancé mi-2012. Il a été rebaptisé en mars 2013 **ContentGems** (<https://contentgems.com/start>) et propose un système d'alertes par mots-clés innovant (sur le papier du moins) que nous avons jugé utile de tester pour vous.

La création d'un compte se fait facilement en donnant un pseudo, une adresse email et un mot de passe.

La première étape consiste ensuite à générer une requête (*create interest*) que l'on va créer dans le formulaire booléen proposé.

The screenshot shows the 'Create a New Interest' interface. At the top, there's a navigation bar with 'ContentGems', 'Your Interests', 'Sources', and 'Sharing'. Below that, the title 'Create a New Interest' is displayed. A search bar contains 'Intelligence économique'. There are three sections for keyword selection: 'Should contain these keywords' with three input fields (one containing 'veille stratégique', one containing 'veille économique', and one empty), 'Must contain ALL of these keywords' with one empty input field, and 'Must not contain ANY of these keywords' with one empty input field. At the bottom, there is a green 'Create Interest' button and a blue 'cancel' link.

Mauvaise surprise : les accents ne sont pas traités et lorsqu'on les supprime, le système ne donne aucun résultat. Il faut donc se résoudre à utiliser l'anglais. Bonne surprise en revanche, de nombreux opérateurs sont utilisables tels que les troncatures, les opérateurs de proximité, la pondération de mots-clés ou encore la possibilité de rechercher dans le champ titre uniquement.

Une fois la requête écrite, une fonctionnalité permet de choisir de faire remonter les résultats:

- les plus populaires des dernières 24 heures ;
- les plus populaires des sept derniers jours ;
- les plus pertinents des dernières 24 heures.

Malheureusement il n'est expliqué nulle part comment ces scores de pertinence et popularité sont calculés.

Il est également possible de personnaliser encore plus ces résultats en utilisant la fonction « Custom settings ».

This screenshot shows the 'Competitive intelligence' interest page. It features the same keyword selection form as the previous image. On the right side, there is a sidebar with four sorting options: 'Popular (24 hrs)' (highlighted with a red box), 'Popular (7 days)', 'Relevant (24 hrs)', and 'Custom Settings...'. At the top right of the page, there is a 'Delete...' button and a 'Help with interests' link. At the bottom, there is an RSS feed link: 'https://contentgems.com/feeds/v1/interests/CPIPIAnP40oWl.WV7 (Regenerate URL)'.

Ici l'on pourra :

- inclure ou exclure des groupes de sources personnalisées (bundles) ;
- choisir entre résultats populaires et pertinents ;

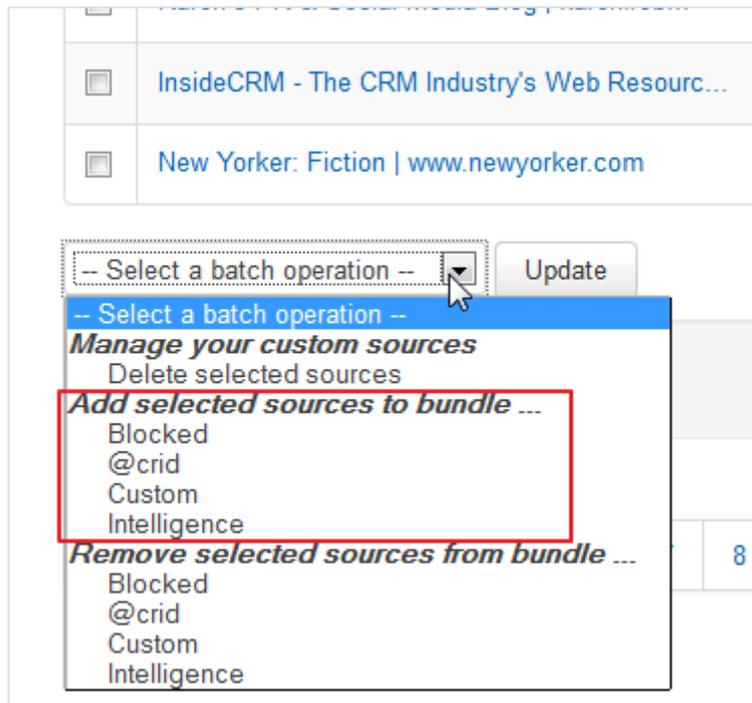
- choisir des résultats calculés sur une durée de 24h ou de 7 jours ;
- ne faire remonter que les résultats devant contenir des images ;
- ceux avec un nombre de mots minimum ;
- ceux avec un taux de pertinence à indiquer soi-même ;
- ceux avec un taux de popularité minimal à indiquer ;
- choisir des les afficher classés par date (antéchronologique par défaut), popularité ou pertinence.

Attention : il n'est possible de surveiller que deux thèmes dans la version gratuite. La seconde étape de la configuration de ContentGems consiste à choisir les sources à utiliser en se rendant dans l'onglet « Sources ». Par défaut, le service annonce monitorer plus de 200 000 sources disposant de flux RSS. Il est possible de les rechercher par mots-clés dans la fenêtre de droite « Filter sources ».

Name	Bundles	Last Checked
Marketwire - Breaking News Releases - New ...		less than a minute ago
PR Newswire: Wyoming www.prnswire.com		less than a minute ago
Technology: Computing guardian.co.uk w...		less than a minute ago
Earthquake News www.mercurynews.com		less than a minute ago
nymag.com: Food nymag.com		1 minute ago
Bit_Synthesis bitsynthesis.com		1 minute ago
XForms and Internet Applications interne...		1 minute ago
Fraser's Affiliate Marketing Blog www.af...		1 minute ago

On va par ailleurs pouvoir ajouter à partir de cette page les flux RSS que l'on souhaite, soit un par un, soit groupés sous forme d'un fichier OPML (attention, fonction limitée à 20 sources dans la version gratuite). L'intérêt sera alors de pouvoir mettre en place des alertes par mots-clés sur ses propres sources. On peut également dans la version payante intégrer des flux Twitter et l'on dispose *a minima* de son flux Twitter personnel gratuitement.

La fenêtre de droite permet également de filtrer les sources par groupes (Bundles) et de les trier en fonction de la dernière mise à jour ou du type de sources (RSS ou Twitter). La création de bundles se fait dans la fenêtre « *Manage Bundles* » en dessous du bloc « Filter Sources ». Il suffit d'ajouter un thème et de cliquer sur Add pour créer un dossier dans lequel on pourra ensuite ajouter des flux thématiques à suivre et qu'il sera possible de monitorer par mots-clés (cf. ci-dessus).

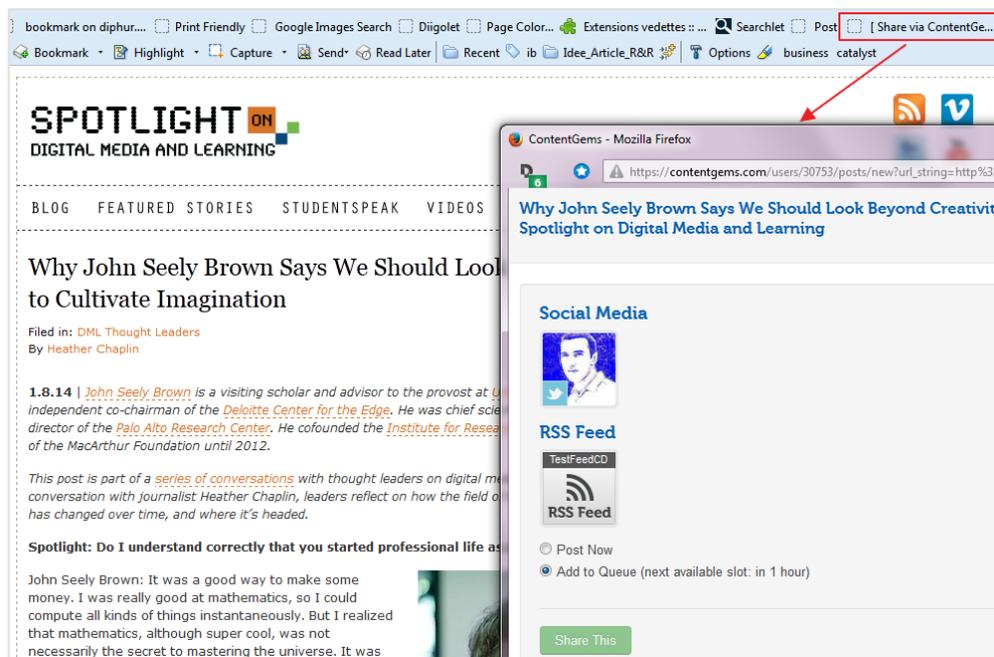


La troisième étape consiste à configurer les réseaux sociaux sur lesquels on souhaite pouvoir partager les remontées d'alertes générées par le service. Il est possible de générer son propre flux RSS et de connecter ses comptes :

- Twitter ;
- LinkedIn ;
- Facebook ;
- Wordpress ;
- Blogger ;
- Expression Engine (système de blogs) ;
- Business Catalyst (système de création de sites web d'Adobe).



A partir de cette même page, il est également possible d'installer dans son navigateur un bookmarklet ou une extension qui permettra de partager n'importe quel contenu trouvé en ligne avec l'un des comptes précédemment connectés.



Une fois ces configurations effectuées, ne reste plus qu'à en attendre le résultat sous forme d'un email qui vous parviendra une fois par jour.

A partir de cet email on peut directement partager un contenu vers un de ses comptes connectés mais aussi envoyer l'article vers un service personnel de stockage et facilitation de la lecture comme Instapaper ou Pocket, ou encore le connecter au service Buffer, qui va permettre d'en programmer de manière optimale la diffusion vers les réseaux sociaux.

ContentGems Daily Summary

What's New

Content Marketing [See all results](#)

A 4 constants of social media marketing

[HOLYKAW.ALLTOP.COM](#)— Long-time editor of SmartBrief on Social Media Jesse Stanchack shares, from his experience, the four constants he's noticed in social media marketing. Here's one: Empathy is everything. Organizations spend untold amounts of time trying to figure out how to communicate their brand to the world. They'd be better off spending that time trying to understand their audience.

[Share via ContentGems](#) · [Buffer](#) · [Instapaper](#) · [Pocket](#)

Higher Quality Content Best Practices

[SOCIALMEDIATODAY.COM](#)— As a content provider, over the past 8 years I've written blog posts for a variety of agencies and their clients. Combining this experience with my love of inbound marketing, I've come up with a few tips I think will help agencies and other writers provide quality content on behalf of their end clients.

[Share via ContentGems](#) · [Buffer](#) · [Instapaper](#) · [Pocket](#)

Ted Rubin

[TEDRUBIN.COM](#)— on Jan 12, 14 • by Ted Rubin • with No Comments Native content is not about pitching your product, it's about creating connection. Remember it is only native to the content consumer if it provides value to them and provides a real solution or makes an emotional connection. — Comments...

[Share via ContentGems](#) · [Buffer](#) · [Instapaper](#) · [Pocket](#)

Si l'on souhaite au préalable lire l'article en question (ce qui semble tout de même préférable), il suffit de cliquer sur son titre. On arrive alors sur une page « framée » dans laquelle on trouve une barre d'outils ContentGems permettant de partager l'article en question vers ses comptes connectés ainsi que par email ou vers le client de médias sociaux Hootsuite.

ContentGems [holykaw.alltop.com](#) Share via: [Email](#) [Buffer](#) [HootSuite](#) [ContentGems](#)



WORK HEALTH CULTURE INTERESTS TECH PEOPLE GOOD NEWS GEOS SPORTS A B C D E F G H I J K L M N O P Q R S T U V W X Y Z #

Hot Topics New Topics My Recent Topics MyAlltop Holy Kaw Search

Holy Kaw! All the topics that interest us

[Follow @guykawasaki](#)

4 constants of social media marketing

Posted Jan 11th, 2014 at 1:00 PM

ContentGems se veut donc un service complet allant de la détection de contenus à leur rediffusion vers de nombreux espaces en ligne, qu'il s'agisse de sites, blogs ou profils sur des médias sociaux. En ce sens, c'est donc un outil de curation qui, à la différence d'un Scoop.it, ne capture pas vos contenus dans une interface propriétaire, ce qui est particulièrement appréciable.

En revanche, si les résultats obtenus grâce aux alertes sont plutôt de qualité grâce aux possibilités de requêtes avancées, on regrettera, d'une part, la non-prise en compte des accents français et d'autre part un manque de précision dans la qualité des contenus à surveiller. Ici justement on aurait aimé une interface de création d'alertes aussi détaillée que celle de Scoop.it, permettant notamment de choisir ses sources par thèmes (Twitter, Facebook, Vidéos, images, actus,...).

Alors non, ContentGems n'est pas non plus le remplaçant direct de Google Alertes (mais qui pourrait l'être), en revanche il peut faire un bon travail lorsqu'il est associé à d'autres services du même type et l'étendue des possibilités qu'il offre mérite largement qu'on lui donne une chance et que l'on surveille ses évolutions.

Christophe Deschamps

Consultant et formateur en gestion de l'information.

Responsable du blog Outils Froids (<http://www.outilsfroids.net/>)

Fin du générateur de mots clés de Google : quels autres outils que le planificateur utiliser ?

[Retour au sommaire de la lettre](#)

Domaine :	Recherche	Référencement
Niveau :	Pour tous	Avancé

Dans le courant de l'année 2013, le générateur de mots clés Adwords a fermé ses portes pour laisser place à son remplaçant : le Keyword Ad Planer, plus contraignant et moins précis. Quels autres outils peut-on utiliser pour remplacer ce nouveau planificateur, le tout afin d'avoir une liste complète et pertinente de mots clés pour son référencement naturel ? Nous allons voir ici différents outils gratuits et payants à utiliser conjointement, certains issus directement de Google ou de Bing, d'autres de sociétés privées.

Le remplacement du générateur de mots clés Adwords au mois d'août dernier (<http://www.abondance.com/actualites/20130829-13060-loutil-de-planification-des-mots-cles-remplace-le-generateur-de-mots-cles-google-moins-de-possibilites-et-de-fonctionnalites-disponibles.html>) cause plusieurs soucis pour les référenceurs. D'une part, il faut avoir un compte Adwords pour pouvoir l'utiliser, mais surtout celui-ci est moins précis en proposant beaucoup moins d'options de filtrage des données.

Depuis l'annonce officielle de la fin du générateur de mots-clés, les autres outils permettant de dresser une liste d'expressions pertinentes pour son référencement naturel sont devenus de plus en plus utiles (source sur la fin du générateur : <https://support.google.com/adwords/answer/3141229>)

Nous allons donc voir ici tous les outils qu'un référenceur peut utiliser pour mener à bien cette tâche.

Les outils internes

Google Analytics et autres outils

Commençons par le plus simple : Google Analytics (ou n'importe quel autre outil de WebAnalytics). Malgré le fait que le (not provided) soit de plus en plus présent dans les statistiques de visites d'un site Internet, les outils de WebAnalytics restent pertinent pour améliorer et cibler des mots clés.

Pour les recherches des internautes qui ont généré des clics sur votre site Internet, et qui apparaissent dans vos statistiques, vous allez pouvoir les afficher et les trier.

Cela a plusieurs intérêts :

- vous allez pouvoir ajouter de nouveaux mots clés auxquels vous n'aviez pas pensé et pour lesquels vous êtes déjà présents sur Google. Il sera d'ailleurs toujours temps d'améliorer votre positionnement sur ces expressions.
- d'autre part, cela permet surtout de connaître le retour sur investissement de chaque mot clé de sa liste, et de savoir s'ils sont ou non rentables pour la société.

L'outil : <https://www.google.com/analytics/>

Les outils pour webmasters

Le deuxième outil que l'on va pouvoir utiliser est le centre webmaster Google et/ou Bing de son site.

De la même manière qu'avec un outil de WebAnalytics, les outils pour webmasters d'un site vont permettre de connaître les mots clés tapés par les internautes. Le gros avantage dans ce cas-là est que l'on va pouvoir connaître les mots clés pour lesquels le site est affiché, mais pour lesquels il n'y a pas forcément de clics affichés dans votre outil de Webanalytics (et ici, il n'y aura pas la problématique du not provided).

Là encore, le concept est identique : on va chercher à trouver tous les mots clés pertinents pour le référencement de son site Internet. On va notamment chercher à connaître tous les mots clés sur lesquels on est déjà positionné, et dont on pourrait améliorer le positionnement, mais aussi tous les mots clés sur lesquels le site apparaît, mais qui ne débouchent pas forcément sur des visites (cela peut notamment vous inciter à retravailler différemment vos contenus, vos balises Title ou vos meta "description").

Ici, voici un exemple avec les mots clés affichés dans le centre webmaster de Google :

Requête	Impressions	Clics ▲	CTR	Position moy.
☆ [blurred]	819	199	24 %	1,0
☆ [blurred]	152	127	84 %	1,0
☆ [blurred]	389	91	23 %	1,0
☆ [blurred]	113	33	29 %	1,0
☆ [blurred]	35	33	94 %	1,0
☆ [blurred]	136	32	24 %	1,0
☆ [blurred]	1 930	27	1 %	3,7
☆ [blurred]	413	27	7 %	2,7
☆ [blurred]	374	24	6 %	2,5
☆ [blurred]	382	16	4 %	11
☆ [blurred]	955	10	1 %	11
☆ [blurred]	48	10	21 %	1,0
☆ [blurred]	299	8	3 %	16
☆ [blurred]	23	8	35 %	1,0
☆ [blurred]	193	5	3 %	19

Les outils en ligne :

- Google : <https://www.google.com/webmasters/tools/home?hl=fr> ;
- Bing : <https://ssl.bing.com/webmaster/home/mysites>.

Les équipes internes

Parmi les outils internes que l'on peut utiliser pour dresser une liste de mots clés exhaustives, il faut penser aux équipes internes de la société. Lorsque l'on cherche à obtenir une liste pertinente d'expressions pour son référencement naturel, il est toujours utile d'interroger les personnes actives dans l'entreprise. On sera ainsi amené à contacter :

- les vendeurs ;
- le service après-vente ;
- les services marketing ;
- la direction ;
- les clients ;
- ...

Les sites des concurrents

Il ne faut pas oublier non plus que les concurrents directs sont sans doute l'une des meilleures sources d'inspiration lorsque l'on cherche à faire un audit de mots clés.

Rendez-vous sur leurs sites pour savoir comment ils présentent et affichent les différents produits et services que vous avez en commun. Regardez particulièrement la manière dont ils expriment et définissent chacun d'entre-eux. Pensez notamment à regarder les balises TITLE qu'ils utilisent, ou encore le nom de chacun de leurs contenus.

N'oubliez pas non plus de faire la même démarche auprès de vos concurrents indirects, c'est-à-dire tous les sites qui se positionnent sur vos mots clés, sans pour autant vendre vos produits et services. Parmi ceux que l'on retrouve souvent parmi les concurrents indirects, vous retrouverez sans doute les sites Wikipédia, YouTube, Doctissimo ou encore commentcamarche.net.

Les outils externes

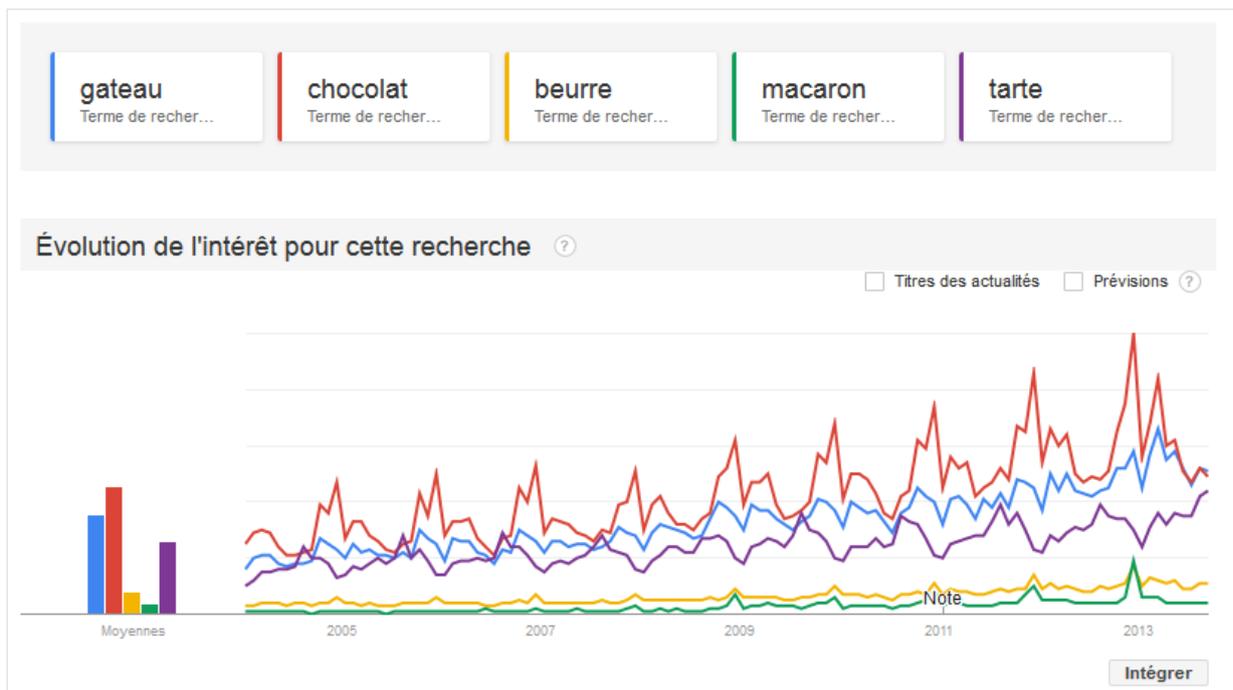
Heureusement, il existe d'autres outils gratuits ou payants qui vont nous permettre d'obtenir une analyse beaucoup plus pertinente de ces mots clés.

Google Trends

Parmi les outils qui deviennent de plus en plus intéressants, on peut noter l'outil Google Trends. Depuis sa fusion avec Google Insights, celui-ci donne des statistiques et informations intéressantes sur les mots clés les plus utilisés, tout en permettant de les comparer entre eux.

Il est en effet possible de comparer l'évolution de l'intérêt d'un mot clé par rapport à un autre. L'outil ne donne pas les statistiques précises mais permet de mieux comprendre l'évolution d'un secteur d'activité, ou la saisonnalité de celui-ci.

Comme le montre l'exemple suivant, vous pouvez comparer jusqu'à cinq expressions à la fois :



L'intérêt est de pouvoir affiner cette recherche grâce à de nombreux filtres qui se basent sur la date, sur le pays, sur le type de recherche effectuée, ou encore sur la catégorie d'activité associée au mot clé.

L'outil permet donc de confirmer ou d'infirmer la pertinence d'un mot clé, d'en connaître sa saisonnalité ou encore de connaître son impact et intérêt en fonction de la zone géographique ou de la langue utilisée.

Autre atout : l'outil va vous proposer une autre liste de mots clés associés à votre recherche, avec les termes les plus recherchés mais également avec ceux ayant enregistré la plus forte progression sur la période analysée. Voici un exemple avec une recherche sur le mot clé "chocolat" :

En progression	
recette thermomix	+140 %
salon du chocolat	+80 %
ganache chocolat	+70 %
glacage chocolat	+70 %
thermomix	+60 %
cupcake	+50 %
pain au chocolat	+50 %

Le site : <http://www.google.com/trends/?hl=fr>

Pour en savoir plus sur cet outil, vous pouvez également lire cet ancien article de la newsletter abondance : <http://recherche-referencement.abondance.com/2013/10/comment-utiliser-google-trends-dans-une.html>

Bing Keyword Tools

Le principal concurrent français de Google propose depuis 2012 un outil équivalent, à la seule condition d'avoir un compte Microsoft (et pas forcément un compte Bing Ads).

Ceux ayant l'habitude du générateur de mots clés de Google ne devraient pas être dépayés. Vous pouvez chercher un ou plusieurs mots clés, en indiquant différents filtres (pays, langue, date, ciblage, ...).

Rechercher des volumes d'expressions et de mots clés à l'aide de notre outil de recherche organique de mots clés. Les données affichées ici proviennent d'une recherche organique sur Bing.

abondance RECHERCHER
Historique

Pays / Région France Langue Tout Stricte oui

Date Les 30 derniers jours : 09/12/2013 - 07/01/2014

EXPORTER

MOTS CLÉS	IMPRESSIONS	TENDANCE
abondance	376	

Suggestions **1**

MOTS CLÉS	IMPRESSIONS	TENDANCE
piercing abondance	135	
abondance haute savoie	64	
la chapelle d abondance	36	
chapelle d abondance	51	

Attention cependant : l'outil utilise les données de Bing. Donc si vous ciblez la France, ne vous étonnez pas d'avoir des volumes de recherche beaucoup plus faibles que les données issues de Google.

Le site : <http://www.bing.com/toolbox/keywords>

Ahrefs

Ahrefs est un outil qui permet d'analyser en profondeur le netlinking d'un site Internet, de connaître son évolution, et de trouver de nouveaux liens. L'outil donne accès à une partie des données en version gratuite, et à la totalité sur la version payante.

Le côté pratique de Ahrefs est que l'on peut aussi s'en servir pour dresser la liste des mots clés d'un site Internet ou d'un secteur d'activité. Pour cela, vous devrez faire cette analyse sur votre propre site et surtout sur ceux de vos concurrents, puis il faudra se rendre dans différents menus.

Le plus pertinent est celui des ancres utilisées ("Anchor"), surtout lorsque vous allez analyser vos concurrents. Ce menu vous donne la liste complète des ancres utilisées pour faire des liens vers le site analysé. Vous allez ainsi pouvoir savoir sur quels mots clés les autres sites du secteur veulent se positionner :

1 - 50 of 6,019 phrases < 1 2 3 4 5 6 7 8 ... 121 > EXPORT ⚙️

	IF Referring Domains Count	Referring Domains with Anchor	I Referring Pages Count	Referring Pages with Anchor	I Anchor Text	
1.	782	27%	108,088	45%	abondance	DOMAINS ▼ SNIPPETS ▼
2.	389	14%	1,143	0.47%	http://www.abondance.com/	DOMAINS ▼ SNIPPETS ▼
3.	245	9%	918	0.38%	www.abondance.com	DOMAINS ▼ SNIPPETS ▼
4.	198	7%	13,493	6%	abondance.com	DOMAINS ▼ SNIPPETS ▼
5.	141	5%	431	0.18%	http://www.abondance.com	DOMAINS ▼ SNIPPETS ▼
6.	90	3%	482	0.20%	<a>noText	DOMAINS ▼ SNIPPETS ▼
7.	80	3%	2,714	1%	olivier andrieu	DOMAINS ▼ SNIPPETS ▼
8.	60	2%	1,883	0.78%	google ne crawle pas votre site si votre fichier robots.txt n'est pas disponible	DOMAINS ▼ SNIPPETS ▼

De la même manière, il est possible de classer toutes les pages d'un site selon la balise TITLE de la page. Là encore, rien de mieux pour savoir sur quels contenu le site d'un concurrent cherche à se positionner.

! Rank	!# Page Title	! Domains	! Backlinks	! Backlinks	! Redirects	! Date	! Size	! Int	! Ext
1.	13 www.abondance.com/ droit/audit.html redirection vers le site Abondance	1	2			33 h	199 B	0	0
2.	12 www.abondance.com/ actualites/20081223-... iPhone : Google s'adapte, Live Search se lance - Actualité Abondance					8 d	9.8 kB	39	33
3.	12 www.abondance.com/ applis/print_version-... iPhone : Google s'adapte, Live Search se lance - Actualité Abondance					5 d	1.3 kB	1	0
4.	12 www.abondance.com/ actualites/20110503-... iPad : les internautes cliquent moins sur le premier résultat Google ...					13 d	9.7 kB	37	27
5.	12 www.abondance.com/ applis/print_version-... iPad : les internautes cliquent moins sur le premier résultat Google ...					6 d	1.4 kB	1	0
6.	12 www.abondance.com/ actualites/20080702-... iGoogle va se socialiser - Actualité Abondance					5 d	9.6 kB	37	32
7.	12 www.abondance.com/ applis/print_version-... iGoogle va se socialiser - Actualité Abondance					17 d	1.3 kB	1	0
8.	12 www.abondance.com/ actualites/20070911-... iGoogle suit la coupe du monde de rugby - Actualité Abondance					5 d	9.7 kB	36	29

Enfin, la fonctionnalité la plus intéressante est de pouvoir créer des rapports SEO à la demande. Et ceux-ci sont réellement complets pour trouver tous les mots clés liés à un site. Le rapport classe et affiche dans différents sous-menu les statistiques et informations concernant :

- les balises TITLE ;
- les H1 ;
- les H2 ;
- les meta "description" ;
- les URL ;
- les liens.

Cela vous permettra donc d'étoffer en profondeur votre liste de mots clés.

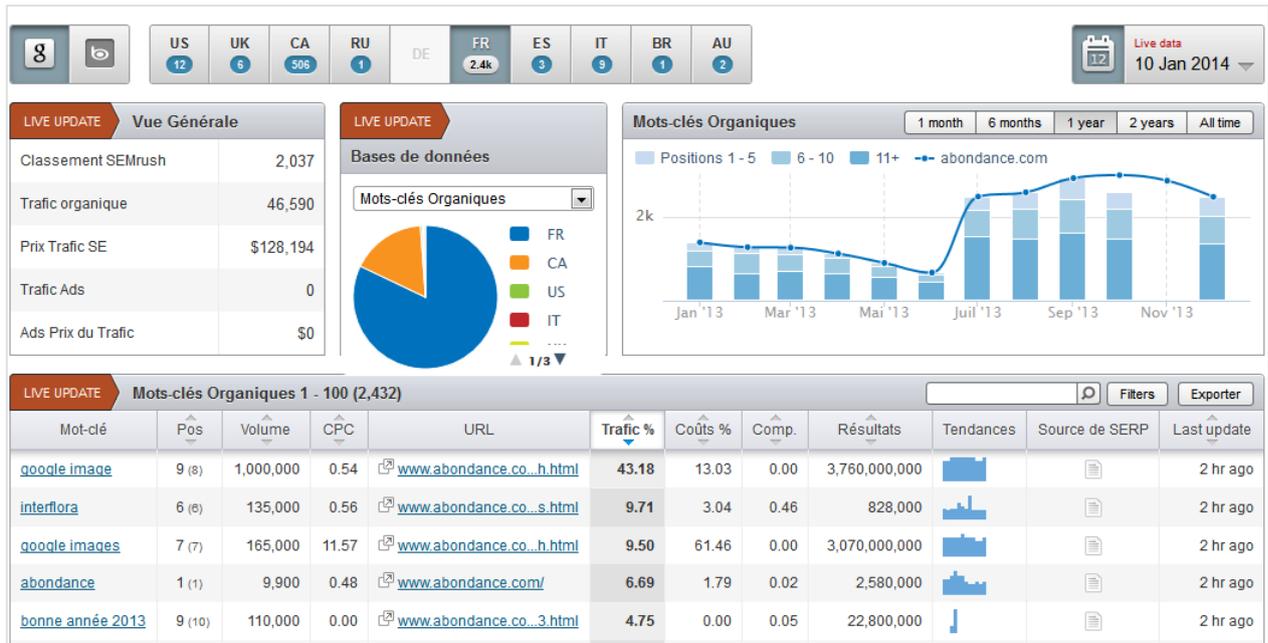
Le site : <https://ahrefs.com/>

SemRush

SemRush est un autre outil en ligne très intéressant pour le référencement naturel. Il permet en effet d'obtenir de nombreuses informations sur les mots clés d'un secteur d'activité ou d'un site, ainsi que des données en provenance d'Adwords et Adsense, les backlinks d'un site ou encore la liste des concurrents.

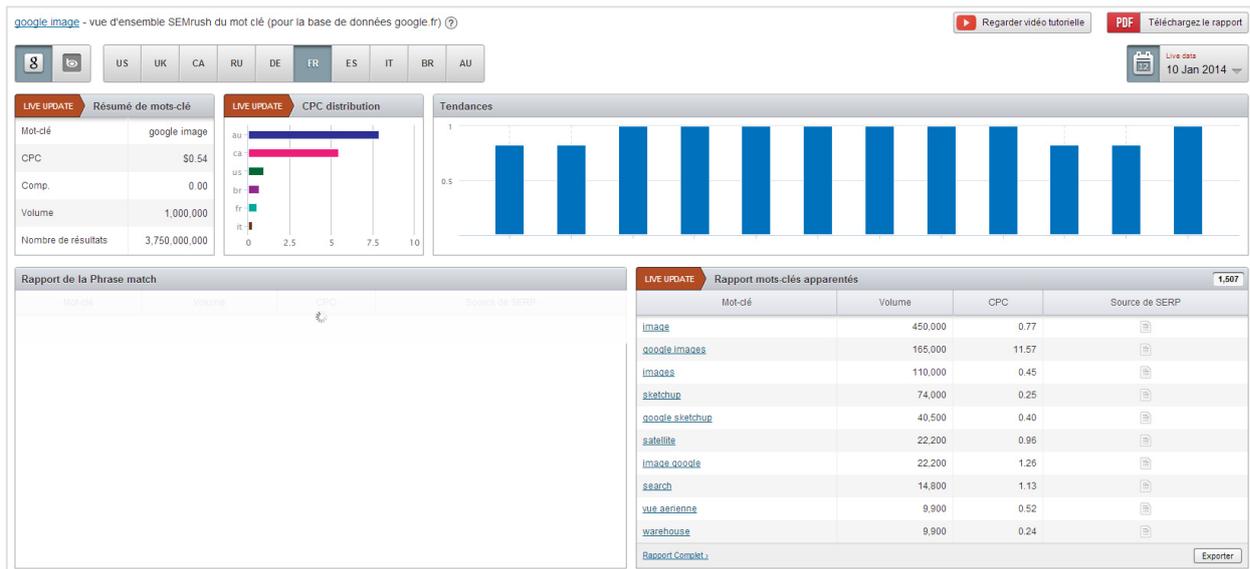
Là encore, l'outil n'est qu'en partie gratuit. Il faudra payer un abonnement pour obtenir toutes les données qui vous intéressent.

Dans notre cas de figure, c'est l'analyse de la Recherche Organique qui va nous intéresser le plus. Dans le premier menu, vous aurez accès à une série de sous-menus. Dans "Positionnement", SemRush dresse un état des mots clés pour lesquels le site apparaît, en affichant au passage sa position, mais surtout le volume de recherche et le CPC (coût par clic publicitaire) du mot clé, ce qui permet de savoir si cette expression est pertinente ou non.



Et si vous cliquez sur un mot clé, vous aurez accès à d'autres informations très pertinentes :

- les tendance de recherche ;
- les mots clés apparentés ;
- les sites positionnés sur cette requête ;
- Etc.



SemRush vous donnera aussi des statistiques sur l'évolution de positionnement du site concerné (notamment pour voir les mots clés sur lesquels il était positionné avant) ou encore la liste détaillées des noms de domaines concurrents.

Le site : <http://fr.semrush.com/>

OpenSiteExplorer

Voici encore un autre outil d'analyse du netlinking de site Internet. Là encore, pensez au fait que vous pouvez utiliser les informations des sites de vos concurrents, à savoir les ancres utilisées, les balises TITLE ou encore les URL de leurs contenus. Dans le premier menu, vous pouvez ainsi voir les sites qui font un lien vers le site analysé et l'ancre utilisée.

Authority		Page Link Metrics		Page Social Metrics	
DOMAIN AUTHORITY	PAGE AUTHORITY	JUST-DISCOVERED	ESTABLISHED LINKS	FACEBOOK	TWITTER
70/100	75/100	2,793 32 Days	1,264 Root Domains 37,170 Total Links	Social metrics only available to Moz Analytics subscribers Learn more	

Inbound Links	Just-Discovered	Top Pages	Linking Domains	Anchor Text	Compare Link Metrics	Advanced Reports
---------------	-----------------	-----------	-----------------	-------------	----------------------	------------------

Show links from pages to and

1 - 50 inbound links from 1,264 domains First 1 2 3 4 Export to CSV

Title and URL of Linking Page	Link Anchor Text	Page Authority	Domain Authority
<p>► Dictionnaire en ligne collaboratif - Dico du Net</p> <p>www.dicodunet.com/</p>	(img alt) Abondance	78	74
<p>Abondance : référencement, SEO et moteurs de re...</p> <p>www.abondance.com/</p>	[No Anchor Text]	75	70
<p>Google Fight : make a fight with Googlefight</p> <p>www.googlefight.com/</p>	(img alt) Abondance	73	67
<p>SEO Mistakes: Spam in other languages</p> <p>www.matcutts.com/blog/seo-mistakes-spam-in-other-languages/</p>	Abondance	72	86

Avec la version payante, vous aurez aussi accès aux statistiques sociales (Facebook Twitter, ...), vous permettant donc d'analyser plus en profondeur ce qui est apprécié et recherché par les internautes dans votre secteur d'activité.

Le site : <http://www.opensiteexplorer.org/>

Majestic SEO

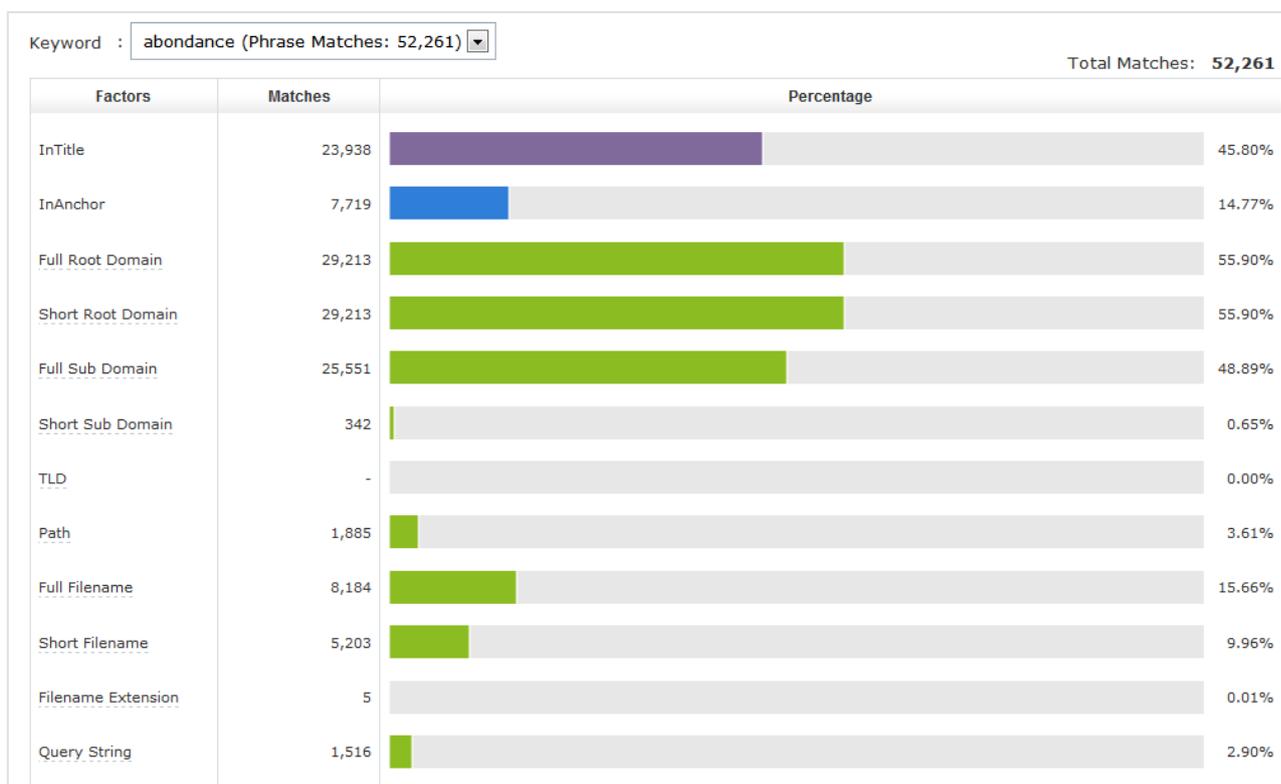
Encore un mastodonte dans le milieu du référencement naturel, l'outil en ligne Majestic SEO permet d'analyser soit une expression, soit un site. Dans les deux cas, cela permet également de faire une analyse de mots clés.

Lorsque vous cherchez un terme, l'outil ressort la liste des résultats sur ce terme. Ce qui est intéressant, c'est de pouvoir connaître comment est utilisé ce mot clé dans chaque résultat de Google (dans les URL, les balises Title, les titres de niveau H1, ...).

Search Results Ranking Factors Map Match Counts

Showing results 1 to 20 out of 1000 for "abondance" ★ Rate These Results

#	Item	Search Score					Citation Flow Trust Flow		Referring Domains External Backlinks		
		Total	InTitle	InAnchor	InURL	Other	Multiplier				
1	Abondance : référencement, SEO et mot... www.abondance.com/	13,187	77	10,239	2,355	516	1,000	61	66	1,914	257,665
2	Office de Tourisme d' Abondance : vill... www.abondance.org/	7,746	41	6,014	1,335	356	1,000	42	46	319	52,054
3	Le réseau Abondance , tous les outils ... www.reseau-abondance.com/	5,877	35	4,798	690	354	1,000	49	50	78	37,113
4	Forum Abondance : Forum de discussion... www.forums-abondance.com/	5,744	32	4,690	663	359	1,000	50	51	53	12,792
5	Piercing abondance - Vente de piercin... www.piercing-abondance.com/	4,132	75	2,845	816	396	1,000	40	51	521	119,874
6	La Boutique Abondance : Lettres d'inf... www.boutique-abondance.com/	3,596	28	2,565	626	377	1,000	49	50	42	36,068



Lorsque vous faites une recherche sur un site Internet, l'outil apportera le même type d'informations que le fait Ahrefs ou OpenSiteExplorer, notamment sur les ancrs utilisée.

Le site : <http://www.majesticseo.com/>

Yooda WebTools

Yooda WebTools est une extension du logiciel payant Yooda SeeUrunk Falcon. Cet outil en ligne gratuit (limité) ou payant va aussi permettre de pouvoir trouver des idées de mots clés. Il inclut en réalité 3 outils différents.

Yooda Keyword Suggest

Cet outil est composé en réalité de deux parties. La première permet de lister les mots clés suggérés en fonction du terme que vous aurez tapé, en donnant également le nombre de recherche en France, le niveau de concurrence et le nombre de résultats.

SEO Benchmark
 Keyword Suggest

Suggestion à partir de mot-clé adresse du site

Rechercher

abondance
Liste des mots-clés suggérés
[Exporter le tableau en CSV](#)

Proximité	Mots-clés suggérés (177)	Nb. rech. France	Concurrence	Nombre de résultats
-	abondance	9 900	<div style="width: 100%; height: 10px; background-color: green;"></div>	2 710 000
30%	val d abondance	720	<div style="width: 100%; height: 10px; background-color: green;"></div>	715 000
25%	vallée d abondance	140	<div style="width: 100%; height: 10px; background-color: green;"></div>	371 000
23%	abondance.com	210	<div style="width: 100%; height: 10px; background-color: #ccc;"></div>	8 260 000
22%	vallée de l abondance	20	<div style="width: 100%; height: 10px; background-color: #ccc;"></div>	1 690 000
21%	vallee d abondance	50	<div style="width: 100%; height: 10px; background-color: green;"></div>	87 100
21%	le val d abondance	50	<div style="width: 100%; height: 10px; background-color: green;"></div>	581 000
21%	ot abondance	50	<div style="width: 100%; height: 10px; background-color: green;"></div>	443 000
21%	station ski abondance	40	<div style="width: 100%; height: 10px; background-color: orange;"></div>	461 000
21%	la vallée d abondance	30	<div style="width: 100%; height: 10px; background-color: orange;"></div>	481 000

La seconde permet de suggérer des expressions en fonction de l'adresse d'un site Internet, avec les mêmes données.

Yooda SEO Benchmark

Enfin, le dernier outil de Yooda permet de générer un rapport SEO sur un site donné, ce qui permet d'obtenir entre autres les mots clés qui apportent du trafic au site, les meilleures pages ou encore les sites concurrents.

<http://www.abondance.com>
Mots-clés qui apportent du trafic au site
[Exporter le tableau en CSV](#)

Mot-clé	Trafic pour le site	Pos.	Page positionnée	Nb. rech. France	Concurrence	Nombre de résultats
abondance	3 643	1 ^{er}	/ (homepage)	9 900	<div style="width: 100%; height: 10px; background-color: green;"></div>	2 710 000
bonne année 2013	3 520	9 ^e	/actualites/20130102-12160-bonne...	110 000	<div style="width: 100%; height: 10px; background-color: green;"></div>	14 600 000
123 people	1 137	3 ^e	/actualites/20081028-9700-123peo...	12 100	<div style="width: 100%; height: 10px; background-color: green;"></div>	869 000 000
google page d accueil	579	9 ^e	/actualites/20071211-9268-nouvell...	18 100	<div style="width: 100%; height: 10px; background-color: green;"></div>	17 400 000
referencement	545	5 ^e	/ (homepage)	9 900	<div style="width: 100%; height: 10px; background-color: red;"></div>	3 350 000
referencement -amazon.com	545	5 ^e	/ (homepage)	9 900	<div style="width: 100%; height: 10px; background-color: red;"></div>	75 600 000
ask.com	387	9 ^e	/actualites/20100729-10420-ask-c...	12 100	<div style="width: 100%; height: 10px; background-color: green;"></div>	1 220 000 000
hummingbird	365	2 ^e	/actualites/20130927-13198-hum...	2 900	<div style="width: 100%; height: 10px; background-color: green;"></div>	9 050 000
référencement	363	5 ^e	/ (homepage)	6 600	<div style="width: 100%; height: 10px; background-color: red;"></div>	8 410 000
googleimage	317	10 ^e	/actualites/20131112-13359-nouvel...	9 900	<div style="width: 100%; height: 10px; background-color: #ccc;"></div>	5 590 000

Chaque mot clé est cliquable, ce qui vous donne alors les mots clés suggérés par rapport à ce nouveau mot clé.

Le site : <http://webtools.seeurank.com/keywordsuggestfree/>

Google

Le moteur de recherche Google est lui-même un très bon outil. Pour vos mots-clés, faites une recherche car cela vous apportera des informations de plusieurs types pour trouver de nouvelles expressions :

- les textes utilisés dans les annonces Adwords ;
- les contenus affichés dans les résultats ;
- les suggestions de recherche.

Recherches associées à **abondance**

[abondance définition](#) [abondance piercing](#)
[abondance fromage](#) [abondance vache](#)
[chapelle d'abondance](#) [abondance altitude](#)
[abondance demain](#) [abondance signification](#)



UberSuggest

Nous avons vu que Google affichait parfois des "recherches associées". L'outil UberSuggest va regrouper dans une interface toutes ces recherches complémentaires, mais va surtout proposer deux fonctionnalités pratiques :

- le téléchargement de ces données ;
- une amélioration de votre recherche initiale.

UberSuggest va en effet lister les 10 recherches associées à votre mot clé, mais aussi les 10 recherches associées à votre expression suivie de chaque lettre de l'alphabet ou de chaque chiffre. L'outil va donc nous lister les expressions connexes du type :

- abondance ;
- abondance a ;
- abondance b ;
- abondance c ;
- ...

↑ abondance + a

- ⊕ abondance altitude
- ⊕ abondance antonyme
- ⊕ abondance autos
- ⊕ abondance aop
- ⊕ abondance abbaye
- ⊕ abondance anglais
- ⊕ abondance argent
- ⊕ abondance alpes
- ⊕ abondance andrieu
- ⊕ abondance atomique

↑ abondance + b

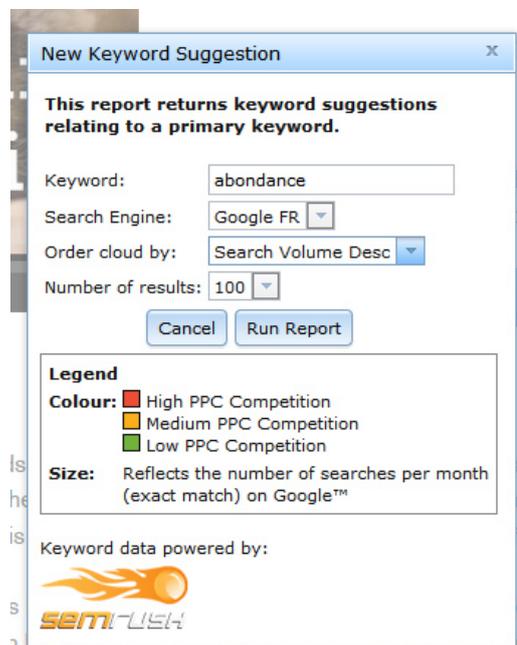
- ⊕ abondance bois
- ⊕ abondance blog
- ⊕ abondance bois la réole
- ⊕ abondance bovin
- ⊕ abondance boulogne billancourt
- ⊕ abondance bio
- ⊕ abondance bible
- ⊕ abondance berthoud
- ⊕ abondance boulogne
- ⊕ abondance bijoux

Le site : <http://ubersuggest.org/>

Keyword Eye

Là encore, Keyword Eye est à la fois un outil gratuit (l'inscription est obligatoire) ou payant pour avoir accès à toutes les données.

Une fois inscrit, cliquez en haut à gauche sur "New", puis déterminer le mot clé sur lequel va se baser votre recherche.



Vous noterez d'ailleurs que les données proviennent de SemRush. Dans notre exemple, nous cherchons les mots clés les plus tapés sur Google FR par rapport à l'expression "abondance" :



Keyword	Search Engine	Search Volume	PPC Comp.	PPC Comp. Score (%)
referencement	Google FR	12,100	High	82
abondance	Google FR	9,900	Low	2
fromage	Google FR	9,900	Low	6
référencement	Google FR	6,600	High	68
la chapelle d abondance	Google FR	5,400	Medium	40
moteurs de recherche	Google FR	5,400	Low	23
le grenier	Google FR	2,900	Low	2
la chapelle	Google FR	2,400	Low	1
moteur recherche	Google FR	2,400	Low	12
chapelle d abondance	Google FR	1,900	Low	28
recherche web	Google FR	1,300	Low	2
cours de conduite	Google FR	880	Medium	43
referencement site	Google FR	880	High	91

Seul défaut : l'outil sera redondant si vous utilisez déjà SemRush.

Le site : <http://www.keywordeye.com/>

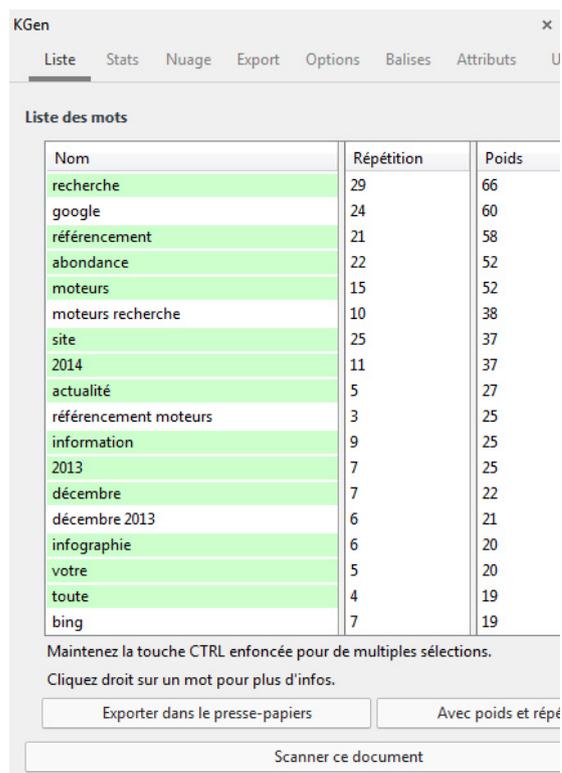
SelfSEO Typo Generator

Pour vos mots clés principaux, ceux qui génèrent beaucoup de recherches par mois, il est toujours intéressant de lister les fautes de frappes associées. Pour cela, utilisez le très bon outil SelfSEO Typo Generator.

Le site : http://www.selfseo.com/keyword_typo_generator.php

Kgen

Kgen est une extension Firefox qui vous permet d'analyser en profondeur une page. L'outil va en effet vous lister tous les mots clés de la page, le nombre d'utilisations et leur "poids" dans le document.



The screenshot shows the KGen application window with a menu bar (Liste, Stats, Nuage, Export, Options, Balises, Attributs) and a table titled "Liste des mots". The table has three columns: "Nom", "Répétition", and "Poids". Below the table are instructions and buttons: "Maintenez la touche CTRL enfoncée pour de multiples sélections.", "Cliquez droit sur un mot pour plus d'infos.", "Exporter dans le presse-papiers", "Avec poids et répé", and "Scanner ce document".

Nom	Répétition	Poids
recherche	29	66
google	24	60
référencement	21	58
abondance	22	52
moteurs	15	52
moteurs recherche	10	38
site	25	37
2014	11	37
actualité	5	27
référencement moteurs	3	25
information	9	25
2013	7	25
décembre	7	22
décembre 2013	6	21
infographie	6	20
votre	5	20
toute	4	19
bing	7	19

Le site : <https://addons.mozilla.org/fr/firefox/addon/kgen/>

Les autres outils

D'autres outils peuvent venir compléter votre recherche de mots clés. Parmi ceux-ci, on peut notamment citer :

- un outil de synonymes : <http://www.crisco.unicaen.fr/des/> ;
- un outil de recherche de mots clés sur les réseaux sociaux : <http://topsy.com/>.

Conclusion

Comme vous pouvez le constater, les outils qui permettent de se passer du Keyword Ad Planer de Google sont nombreux. Certains sont très complets, et chacun apportera des informations supplémentaire dans votre recherche et analyse de mots clés.

Attention cependant : la difficulté ne sera pas de lister les mots clés, mais bel et bien de les trier et filtrer pour ne conserver que ceux qui seront pertinents pour votre référencement naturel ! Bon travail !

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

WDF*IDF : la "formule magique" des référenceurs allemands

[Retour au sommaire de la lettre](#)

Domaine :	Recherche	Référencement
Niveau :	Pour tous	Avancé

*Le microcosme SEO allemand est en effervescence actuellement autour du concept de "WDF*IDF", qui n'est rien d'autre qu'un dérivé du TF*IDF bien connu depuis de nombreuses années et qui fait partie intégrante des algorithmes de pertinence de Google et consorts. Alors qu'en est-il ? Est-ce vraiment nouveau et faut-il s'y attarder ou n'est-ce que de la poudre aux yeux ?... Voici quelques éléments de réponse...*

Nota : certaines URL correspondant aux sources citées dans cet article se trouvent à la fin.

Depuis fin 2012, une "nouvelle" approche du référencement fait fureur au sein de la communauté SEO allemande : optimiser son contenu à l'aide de WDF*IDF. Les conférences sur le sujet et les billets de blogs se sont multipliés outre-Rhin, et la communauté anglo-saxonne commence à en parler. Mais de quoi s'agit-il ?

Nous allons voir que ce qu'ont découvert les allemands n'a rien de nouveau, que cette approche est déjà dépassée depuis longtemps et que cette méthode n'apporte pas grand-chose dans la pratique. C'est néanmoins l'occasion de reparler des méthodes de linguistique statistique utilisées par les moteurs de recherche. Bien connaître ces méthodes est un passage obligé pour être en mesure de comprendre les algorithmes d'un moteur de recherche, et pour mettre au point des techniques de référencement avancées.

Les origines de la formule WDF*IDF

La formule a été introduite par un référenceur allemand, Karl Kratz (*photo ci-contre*), en avril 2010, dans un billet de son blog à propos de la "densité de mots clés" (*Cf. bibliographie en fin d'article*). Mais c'est en réalité une conférence de Karl Kratz à Berlin fin 2012 qui a véritablement fait découvrir le concept à la communauté SEO allemande. Depuis lors, toutes les conférences SEO allemandes parlent de la formule et de ses applications.



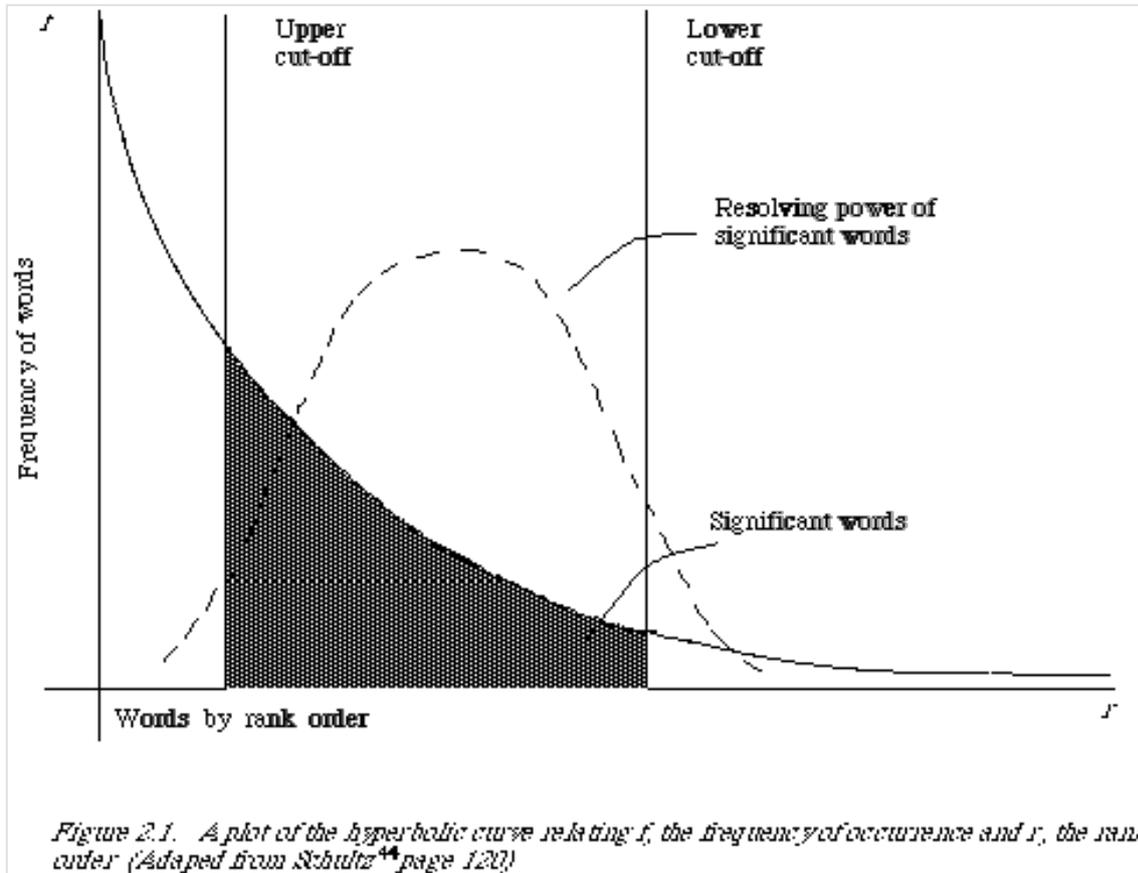
Mais qu'est-ce que la formule WDF*IDF ?

En réalité, cette formule est plus connue dans le monde anglophone et francophone sous l'appellation : TF*IDF.

Cet acronyme signifie : *Term Frequency * Inverse Document Frequency (Fréquence des termes * Inverse de la Fréquence dans les Documents)*. Certains auteurs préfèrent utiliser le terme WDF (*Within Document Frequency : Fréquence au sein du Document*) à la place de *Term Frequency*, pour mieux marquer la différence entre les deux termes de la formule.

Il s'agit d'une "vieille" formule, issue des travaux du linguiste et chercheur en informatique Gerald Salton à partir de la fin des années 60, et des perfectionnements introduits par Karen Spärck Jones et Stephen Robertson dans les années 70. La définition de l'*Inverse document frequency* par Karen Jones date de 1972, et la formule TF*IDF apparaît pour la première fois en 1975 dans un article de Gerald Salton.

L'objectif de cette formule est de calculer un "poids" pour un terme dans un document, afin de pouvoir mesurer facilement si un document était susceptible d'être intéressant à regarder lors d'une recherche par mots clés.



L'analyse de fréquence d'occurrence des termes dans un document suit classiquement une loi de Zipf. L'expérience prouve que les mots clés les plus fréquents ne sont pas spécifiques au document et à son sujet mais résultent de la syntaxe, de la présence de « mots vides » et de termes creux. On doit donc éliminer de la liste la « top tail » et la « long tail », le problème c'est que les limites entourant la middle tail (« upper » et « lower cut off » sur le graphe ci-dessus) ne sont pas faciles à déterminer : cette approche n'est pas automatisable.

Pour calculer le "poids", l'"importance" d'un terme dans un document, la première idée qui vient à l'esprit est de mesurer le nombre d'occurrences d'un terme dans le document. On s'aperçoit assez vite que ce critère ne vaut rien si on ne tient pas compte du nombre de termes présent dans le document. On calcule donc la fréquence d'un terme dans le document (TF, term frequency), qui est improprement appelée "densité de mots clés" dans le jargon des référenceurs.

Document 1

Son nom est célébré par le bocage **qui** frémit, et par le ruisseau **qui** murmure, les vents l'emportent jusqu'à l'arc céleste, l'arc de grâce et de consolation que sa main tendit dans les nuages.

Calcul de TF (fréquence du terme) sur le Document1 ci-dessus pour le terme « qui ».
Le texte comporte 38 termes, et 2 occurrences du mot « qui » :

$$tf_{1,1} = \frac{n_{1,1}}{\sum_k n_{k,1}} = \frac{2}{38}$$

Remarque : La somme des occurrences pour chaque terme différent est égal au nombre total des termes dans le document.

Mais les travaux de Spärck Jones ont prouvé très tôt (il y a 42 ans !) que la fréquence du terme était également un critère de poids sans intérêt si on ne tient pas compte de la rareté ou au contraire du caractère extrêmement répandu d'un terme dans l'ensemble des documents analysés (ce que l'on appelle dans le jargon des linguistes le "corpus"). En effet, plus un terme est rare, moins il a de chances d'apparaître souvent dans un document : il se verra donc attribuer systématiquement une fréquence faible (une "densité" faible), alors qu'une seule occurrence de ce terme rare devrait "qualifier" un document comme "intéressant" dans une recherche par mots clés. Il faut donc pondérer la fréquence des termes par le nombre de documents dans lequel le terme apparaît.

$$idf_t = \log \frac{N}{df_t}$$

La formule de l'inverse de la fréquence d'un terme dans le document est très simple à calculer : N représente le nombre total de documents analysés (toutes les pages web indexées pour un moteur). df_t représente le nombre de documents contenant le terme. La fréquence serait donnée par df_t / N : on prend l'inverse car IDF sert à pondérer la fréquence des termes. Et on utilise le logarithme de la valeur parce N/df peut donner des valeurs très grandes.

La formule complète de TF*IDF est la suivante :

$$w_{x,y} = tf_{x,y} \times \log \left(\frac{N}{df_x} \right)$$

TF-IDF

Term x within document y

$tf_{x,y}$ = frequency of x in y

df_x = number of documents containing x

N = total number of documents

Nous évitons à dessein le terme "pertinent" dans ce contexte, car ces calculs de poids n'ont pas grand chose à voir avec la pertinence du document pour un terme donné : en fait, on mesure en quelque sorte l'écart entre la probabilité théorique qu'un terme soit présent dans un document (mesuré par le terme IDF), et sa présence réelle (mesuré par la fréquence d'apparition du terme dans le document : WDF). Certains chercheurs parlent de mesure de la "spécificité du terme" ou de "potentiel de discrimination du terme" pour décrire ce que représente vraiment ce poids.

En quoi est-ce nouveau ?

En réalité, la formule WDF*IDF ne représente rien de nouveau : elle fait l'actualité en Allemagne car outre-Rhin, les SEO viennent juste de découvrir la linguistique statistique et les travaux de base sur la Recherche d'Informations (*Information Retrieval* en anglais). Qui plus est, le choix de l'appellation WDF*IDF (plutôt rare en dehors de l'Allemagne) leur a fait croire que le monde entier ignorait ces travaux (alors qu'une recherche sur TF*IDF risquait de faire remonter pas mal d'articles de SEO sur le sujet).

En France, la première présentation de ces techniques à la communauté SEO remonte à 2004 : l'auteur de cet article a eu l'honneur de présenter les travaux de Gerald Salton lors du séminaire Netmarketing organisé par Raphaël Richard en 2004. Depuis, le concept a été abordé de nombreuses fois, notamment lors des SEO Camp'us (en particulier en 2009 et en 2010).

nm 2004
 le forum de la performance internet

Les algorithmes sémantiques et linguistiques

MODELES VECTORIELS

Repose sur les espaces vectoriels de Salton

tf : term frequency
idf : inverse document frequency

*tf*idf : mesure de fréquence pondérée*

Mesures de similarité utilisées :
Dice, Jaccard, Cosinus, Coefficient de recouvrement
(nombreuses variantes)

$D_1 = (0.8, 0.3)$
 $D_2 = (0.2, 0.7)$
 $Q = (0.4, 0.8)$
 $\cos \alpha_1 = 0.74$
 $\cos \alpha_2 = 0.98$

NM Netmarketing 2004 2/3 décembre 2004
17

Une slide tirée de la conférence du Séminaire Netmarketing sur le thème « SEO et linguistique statistique, recherche d'information et traitement automatisé du langage.

La formule utilisée pour wdf*idf n'est que l'une des nombreuses variantes de la formule classique, et qui a été popularisée par Donna K. Harman, qui au sein de l'institut américain NIST (*National Institute of Standards and Technology*), a été la cheville ouvrière des ateliers TREC (Text Retrieval Conference) dont le principal objectif est d'évaluer la qualité des techniques déployées en *Information Retrieval*.

$$WDF(i) = \frac{\log_2(\text{Freq}(i,j)+1)}{\log_2(L)}$$

$$IDF_t = \log \left(1 + \frac{N_D}{f_t} \right)$$

*La formule de Donna K. Harman ci-dessus est l'une des nombreuses variantes de tf*idf inventées depuis les années 70.*

A quoi cela peut servir en théorie ?

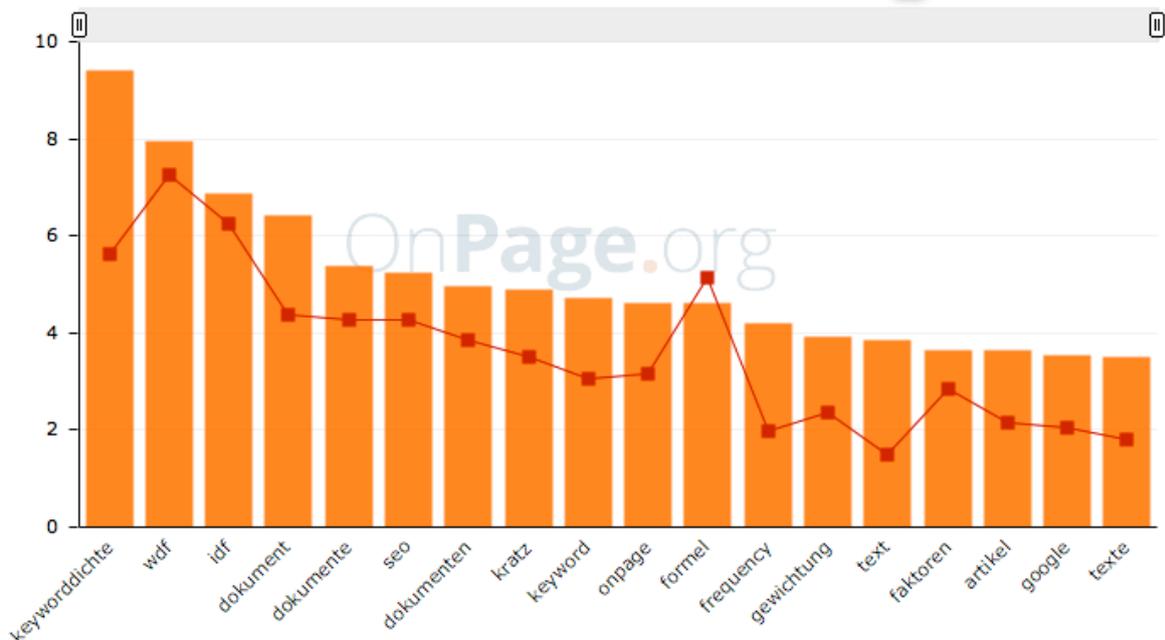
Connaître les poids des différents termes dans un document permet de :

- **Comparer des documents entre eux** pour savoir quel poids de termes caractérisent les premiers résultats fournis par un moteur sur une requête donnée. Cette retro-ingénierie permet d'estimer grossièrement la fréquence de terme idéale (et bien mieux que l'approche traditionnelle se basant sur la densité de mots clés).

- Calculer un "Cosinus de Salton" entre une requête et un document, pour voir si le classement en fonction de ce critère reflète les SERP. Si oui, on a un moyen simple de savoir si un changement dans le contenu textuel de la page peut permettre de remonter sur une requête.

Nous avons en son temps appelé cette approche l'"alignement sémantique" : un changement dans le contenu textuel permet en effet théoriquement de "rapprocher" le document de l'idéal attendu par l'algorithme du moteur (mais uniquement sur les critères de proximité sémantique, qui font partie des très nombreux signaux utilisés par les moteurs pour identifier les documents les plus pertinents sur une requête).

La méthode de Karl Kratz est très exactement une méthode d'alignement sémantique. Une méthode qui est apparue tellement révolutionnaire outre-Rhin que nombre de sociétés de création contenus travaillant pour le marché allemand se sont senties obligées d'annoncer fièrement qu'elles fabriquaient des textes optimisés selon la méthode WDF*IDF. La formule de Karl Kratz est donc devenue un argument marketing.



Une analyse des poids wdf*idf pour les termes d'une page web réalisée par l'outil Onpage.org

Que vaut la formule des SEO allemands ?

La formule de Donna Harman, comme toutes les autres variantes, a fait l'objet d'évaluation lors des conférences TREC successives. Il s'avère que l'approche qui a donné les meilleurs résultats et qui fait (presque) consensus aujourd'hui, est plutôt Okapi BM25 (voir bibliographie). Donc, quitte à vouloir calculer un TF*IDF, on peut faire mieux que la version allemande de WDF*IDF. Il existe d'ailleurs sur Internet des articles décrivant comment implémenter Okapi BM25 dans des moteurs de recherche open source comme Xapian, ou Lucène SolR (qui utilise par défaut une version basique de TF*IDF).

Une formule approchée et "corrigée" selon des critères très contestables

Néanmoins, analyser le poids des termes avec TF*IDF représente déjà un saut qualitatif énorme par rapport à une analyse de type "densité de mots clés". Et si la formule WDF*IDF de Donna Harman n'est pas la meilleure, c'est l'une des meilleures.

 Besser. **Aber schwieriger.**

Linkless Outranking
WDF*IDF

Wenn man es ein
bisschen genauer nimmt:



$$w_{i,j} = \frac{\text{lb}(TF_{i,j} + 1)}{\text{lb}(L_j - (TF_{i,j} * (tc - 1)))} * \text{ld}\left(1 + \frac{R_{1,j} \cup R_{2,j} \cup \dots \cup R_{n,j}}{N_i}\right)$$

Gewichtung
eines Terms i
im Dokument j

**Anzahl aller
Terme** im
Dokument j.

Häufigkeit
des Term i im
Dokument j.

**Anzahl der
Einzelterme** der
Term-Kombination

**Anzahl aller
Dokumente** in denen
der **Term i**
vorkommt.

**Vereinigungsmenge
aller Resultate** in der
Datenbank, die alle
Terme i_n enthalten,
welche im Dokument
j vorkommen.

© karl kratz online marketing 2012

19

*La formule de Donna K. Harman telle que Karl Kratz
la présente dans la langue de Goethe.*

Mais les choses se gâtent lorsqu'on regarde de près comment les référenceurs allemands l'utilisent... Le principal problème avec une formule de type TF*IDF, c'est de savoir comment on peut calculer le terme IDF (la fréquence des termes dans le document) alors qu'on ne connaît pas le corpus ! En effet, il faudrait, comme le ferait un moteur comme Google, crawler tout le web, et compter la fréquence d'apparition de chaque terme dans toutes les pages web.

Une première idée appliquée par nos cousins allemands consiste à remplacer la fréquence des documents par le nombre de résultats fournis par Google pour un terme donné. Mais ce n'est pas une fréquence et la formule ainsi modifiée donne des résultats absurdes.

La seconde solution consiste à calculer une fréquence approchée en extrayant tous les termes d'un document, d'interroger Google pour obtenir le nombre de résultats pour une requête sur chaque terme du document, et d'en déduire la fréquence IDF. En théorie, on obtient effectivement un poids approché, mais selon le type de document, on peut être très proche ou très éloigné de la vérité !!

On pourrait certes sérieusement améliorer la précision et la pertinence des calculs dès lors que l'analyse des corpus est possible. C'est notamment vrai si on analyse des pages techniques, et si on peut "aspirer" toutes les pages qui parlent du sujet, soit un sous ensemble limité du World Wide Web. On peut aussi se baser sur des statistiques disponibles pour des gros corpus web mis à la disposition des chercheurs en traitement automatisé du langage.

Ceux qui ont réellement essayé et notamment ceux qui ont développé les plugins ou les applications de calcul en ligne de WDF*IDF ont donc transformé la formule en WDF*P*IDF ou P est un coefficient de correction qui permet d'éviter que la hiérarchie des poids donnée par la formule sur un texte donné ne soit totalement absurde. Notons que le coefficient P est empirique et qu'il est croustillant de constater que les auteurs des

méthodes opérationnelles de calcul restent toujours très évasifs sur ce que cache le coefficient P et comment il est déterminé.



Ce que pense probablement Matt Cutts de la formule $wdf*p*idf...$

L'alignement sémantique à l'aide de TF*IDF a ses limites...

L'un des défauts la méthode est qu'elle ne tient aucun compte de trois réalités :

- La fréquence d'apparition des termes est fortement influencée par la syntaxe d'une langue et les habitudes de langage. Bref, on peut croire que certains termes sont importants dans un document, alors que la plupart ne sont pas réellement porteurs de sens. *A contrario*, on ne verra pas que certains termes rares ont une forte probabilité d'apparition dans un document parce qu'ils sont très fréquemment présents à côté d'autres termes (c'est à dire qu'ils ont une forte fréquence de cooccurrence).

- L'approche ne fait aucun cas de l'existence d'expressions toutes faites, comme dans "Président de la République" : les termes "président" et "république" sont analysés séparément alors qu'ils forment un tout.

- Enfin, certaines pages web sont « multi-sujets » comme des pages d'actualité ou des pages de catégories ou de rubriques. Dans ce cas, la présence d'un mot clé spécifique à un sujet donné ne garantit pas que la page soit pertinente pour ce terme. Ce cas est tellement fréquent que les moteurs en tiennent compte, et c'est pourquoi leurs algorithmes prennent en compte d'autres facteurs que $tf*idf$, $wdf*idf$, ou Okapi BM25.

Bref, une méthode d'alignement sémantique efficace nécessite d'aller plus loin :

- Comme utiliser LSA/LSI pour identifier les termes qui sont "corrélés" dans un document, afin de filtrer le bruit dans les fréquences de termes (WDF) et identifier les termes qui comptent d'un point de vue sémantique.

- Avoir une approche tenant compte des entités nommées et des syntagmes importants (approche par groupe de mots clés, ou n-grammes).

Le problème, c'est que toutes ces analyses demandent beaucoup de temps et de compétences, et que dans bien des cas, le jeu n'en vaudra pas la chandelle. Car la supériorité de cette méthode sur un bon brief éditorial et un bon contrôle qualité sur le travail des rédacteurs web n'est pas toujours assurée.

Cette méthode était valable en 2004 ! Nous sommes en 2014 ...

Qui plus est, Google a évolué, et nous sommes à peu près sûr qu'il utilise des critères bien plus sophistiqués que TF*IDF pour identifier les documents à faire remonter dans ses pages de résultats. Cela fait trois ans que l'on parle d'indexation de syntagmes (et donc la fin du paradigme en "sac de mots" où tous les termes sont analysés individuellement sans tenir compte de leur ordre et de ce que l'on trouve comme autres termes autour d'eux).

Et maintenant avec Hummingbird, nous en sommes à l'indexation de concepts, ce qui sonnera à terme le glas du paradigme précédent où le mot-clé était tout puissant !

Faut-il oublier ces techniques, et ne pas s'y intéresser ? Pas du tout, il est important qu'un référenceur de 2014 connaisse ces fondamentaux, pour comprendre comment les algorithmes des moteurs ont fonctionné jadis, car après tout, les versions actuelles ne diffèrent en réalité que par leur sophistication plus élevée. Et elles peuvent aider à faire du référencement avancé et à évaluer correctement le contenu d'une page et d'un site web. Mais il faut aussi garder en tête deux réalités : ces techniques sont datées, et elles ont leurs limites.

Et nous espérons que cet article vous permettra de distinguer le vrai du faux et surtout les argumentations marketing pseudo scientifiques et opportunistes quand, inmanquablement, la mode du WDF*IDF va commencer à influencer les méthodologies du monde du SEO francophone.

Bibliographie

Un article de Wikipedia sur TF*IDF :
<http://fr.wikipedia.org/wiki/TF-IDF>

L'article séminal de Karl Kratz (en allemand) :
<http://www.karlkraz.de/onlinemarketing-blog/seo-keyword-density/>

Le premier article sur le sujet dans l'univers SEO anglo-saxon :
<http://pagerankchecker.com/blog/seo/that-s-wdf-idf/>

Un autre article en anglais:
<http://sno.pe/b/seo-content-quality-and-mathematical-determination-with-wdfidf/>

Tf*idf : l'une des premières évocations en France devant la communauté SEO :
<http://www.webmaster-hub.com/publication/IMG/pdf/slides-semantique-netmarketing2004.pdf>

L'origine de la formule : extrait en ligne d'un livre de Donna K. Harman :
<http://www.ai.wu.ac.at/~koch/courses/wuw/archive/inf-sem-ws-00/zoechling/statistik.htm>

Okapi BM25 : la "meilleure" formule de calcul de poids de termes dans un document :
http://en.wikipedia.org/wiki/Okapi_BM25

Présentation au SEO Camp'us 2009 : TF*IDF, Cosinus de Salton et alignement sémantique :
http://www.seo-camp.org/wp-content/uploads/2009/02/apport_semantique.pdf

Calculateur de WDF*IDF pour Wordpress :
<http://codecanyon.net/item/wordpress-wdfidf-seo-calculator/5122990>

Calculateur online :
<http://onpage-tool.de/>

Et pour en savoir plus : Cours de Recherche d'Information par A. Rozenknop (pour les étudiants de l'Université de Paris XIII)

http://lipn.univ-paris13.fr/~rozenknop/Cours/MICR_REI/Seance6/modeles-RI-1.pdf

Bonne lecture à tous !!

Philippe YONNET , *Directeur de l'agence Search-Foresight / Groupe MyMedia.*
Président de l'association SEO Camp (<http://www.seo-camp.org/>)

L'IP Tracking sous les feux de la critique

[Retour au sommaire de la lettre](#)

Domaine :	Recherche	Référencement
Niveau :	Pour tous	Avancé

Depuis des années, la pratique de l'IP Tracking permet d'effectuer des opérations sur un site web en fonction du comportement de l'internaute, en gardant en mémoire son adresse IP. Mais certains abus, notamment dans le domaine de la réservation des billets d'avion, ont mis ces pratiques sous le feu de la critique. Qu'en est-il aujourd'hui et quelle incidence cela peut-il avoir à l'avenir sur le monde du SEO ?...

Depuis plusieurs années, la pratique controversée de l'IP tracking s'est développée sur de très nombreux sites. Elle consiste en une « optimisation » du contenu affiché en fonction des visites de l'internaute et ce, grâce à l'enregistrement et le traitement de son adresse IP. Cette pratique a surtout connu un succès important sur les sites marchands dont certains n'hésitaient pas à modifier leurs prix en fonction de l'historique et des recherches de l'internaute. Ce référencement de l'internaute a finalement alerté certaines associations qui, aujourd'hui, ont obtenu des autorités qu'elles se penchent sur la question.

Description de la pratique de l'IP tracking

L'IP tracking ou pistage d'adresse IP est une pratique qui consiste à enregistrer l'adresse IP du terminal utilisé par l'internaute, et conserver cette adresse IP pour un traitement spécifique. En tant que telle, cette pratique est assez courante et pratiquée depuis plus de 10 ans pour des raisons d'efficacité évidente.

Le problème surgit lorsque cette méthode est utilisée à des fins plus discutables. Par exemple, les services de réservation en ligne peuvent faire varier à leur avantage les prix d'un produit en fonction des recherches effectuées par l'internaute. Concrètement, le consommateur effectuant par exemple une recherche de billet d'avion, mais qui ne finalise pas la transaction immédiatement, verra les prix du billet augmenter et/ou le nombre de billets disponibles diminuer lors de ses prochaines visites. L'opérateur de réservation en ligne ayant ainsi gardé en mémoire l'adresse IP utilisée, reconnaît l'internaute et suscite l'acte d'achat du consommateur par une modification artificielle des prix.

Cette pratique pose des difficultés notamment au regard de la collecte des données personnelles des consommateurs, de la loyauté de l'information sur les prix pratiqués, et du risque de pratiques discriminantes à l'égard de certains consommateurs. En effet, un internaute achetant son billet *via* le terminal sur lequel il a effectué ses simulations d'achat préalables, paierait un prix plus élevé que les internautes qui achètent leurs billets immédiatement, pour un service pourtant identique.

Devant ce problème, les politiques se sont mobilisés pour obtenir la fin de ces pratiques. Etrangement, seuls les problèmes relatifs à la réservation de billets, voyages et nuits d'hôtel ont été concernés par les actions gouvernementales et administratives, laissant toutes les autres pratiques hors du champ des enquêtes et sanctions...

Les actions en cours

Au début de l'année 2013, plusieurs politiciens se sont publiquement émus de ces pratiques et ont souhaité obtenir à tout le moins des éclaircissements.

Une députée socialiste, Françoise Castex, accessoirement rapporteur pour la Commission des affaires juridiques du Parlement européen sur le rapport sur une approche globale de la protection des données à caractère personnel dans l'Union européenne, a dans un premier temps sollicité la Commission européenne.

En effet, la question des données personnelles est de compétence fédérale depuis plus de 20 ans et il est logique de s'adresser à l'échelon pertinent. La députée a donc très simplement (en 4 lignes) demandé à la Commission si la pratique de l'IP tracking était légale et si une enquête était prévue. La Commission n'a malheureusement apporté aucune véritable réponse (spécialité bruxelloise !) et a renvoyé la question aux autorités compétentes.

La députée a donc posé la même question à la CNIL le 24 avril 2013. A la différence de la Commission européenne (que personne ne contrôle), la CNIL a, pour sa part, l'obligation de répondre en apportant de vraies réponses. Dans ce cadre, la CNIL a donc confirmé diligenter une enquête et a précisé que le procédé de l'IP tracking « *se basant sur une collecte de données personnelles, dont l'adresse IP, suscite de nombreuses interrogations* ».

C'est dans ce cadre que l'enquête a débuté, en collaboration avec la Direction Générale de la concurrence, de la consommation et de la répression des fraudes (qui est en fait un organisme rompu aux enquêtes de terrain, notamment pour le contrôle de la législation afférente à la consommation, à l'inverse de la CNIL). Il est important de souligner que cette enquête est limitée à « *l'utilisation réelle de ces techniques [d'IP tracking] par les sites de vente de billetterie en ligne* ».

Au titre de cette enquête et de manière préliminaire, la CNIL s'interroge :

- « *d'une part, sur la licéité de la collecte de données notamment au regard de l'information préalable des personnes,*
- *d'autre part, sur la légitimité de la finalité poursuivie* ».

La CNIL se fonde sur l'article 10 de la loi « Informatique et Libertés » du 6 janvier 1978 qui interdit « *tout automatisme dans la prise de décision produisant des effets juridiques à l'égard d'une personne* » et considère que l'IP tracking « *est susceptible de porter préjudice à certains clients qui se verraient appliquer des tarifs moins avantageux en fonction de leur profil de navigation* ».

Enfin, selon la CNIL, « *l'IP Tracking doit également être appréhendé sur le fondement des pratiques commerciales déloyales* ». A ce titre la CNIL se fonde sur l'article L. 120-1 du Code de la consommation, aux termes duquel : « *Une pratique commerciale est déloyale lorsqu'elle est contraire aux exigences de la diligence professionnelle et **qu'elle altère, ou est susceptible d'altérer de manière substantielle, le comportement économique du consommateur** normalement informé et raisonnablement attentif et avisé, à l'égard d'un bien ou d'un service* ».

Pour l'heure, la CNIL n'a communiqué aucun résultat de l'action conjointe menée avec la DGCCRF. Toutefois, il est extrêmement probable que l'enquête conclut à une pratique illicite, principalement en violation de la législation sur la consommation et la loi informatique et libertés.

En fonction, soit la loi, soit la jurisprudence, interdira rapidement ces pratiques. Logiquement, cette interdiction portera uniquement sur la vente de billets / de nuits d'hôtel sur Internet.

En l'état, il semblerait que l'interdiction par la loi soit la plus plausible. En effet, le 9 janvier 2014, Jean-Vincent Placé (certes, député Vert, mais proche de Françoise Castex) a posé une question parlementaire au Gouvernement, qui n'a pas hésité à défendre d'une certaine manière l'IP tracking. Ainsi, le Ministre du Redressement Productif a précisé : « *Les constatations réalisées à ce jour montrent que l'intérêt économique des entreprises à mettre en œuvre des pratiques d'individualisation des propositions commerciales en fonction de données collectées sur les internautes est avéré (les bannières personnalisées ont un meilleur taux de conversion). Le processus de segmentation est techniquement*

abouti. À ce jour (NdR, au 9 janvier 2014), il n'a pas été démontré que la technique [d'IP tracking] est utilisée pour proposer des prix personnalisés. En tout état de cause, elle ne semble pas reposer sur la pratique de l'IP tracking telle qu'elle est dénoncée par la presse. L'adresse IP n'est qu'un des éléments collectés et ce n'est pas le plus pertinent, les entreprises se fondant notamment sur des informations stockées dans le navigateur telles les cookies ou témoins de connexion. Bien entendu, la CNIL et la DGCCRF vont maintenir collectivement leur vigilance en continuant à réaliser des enquêtes communes sur ces pratiques supposées ».

Cette réponse, étonnante, semble démontrer que l'IP tracking n'est pas encore banni en France...

En revanche, au niveau européen, la Commission des affaires juridiques du Parlement européen (qui reste l'autorité compétente, la France n'ayant pas compétence pour remettre en cause les décisions éventuelles prises au niveau de l'Union européenne) a adopté le 27 novembre 2013, un amendement au projet d'avis concernant l'application de la directive 2005/29/CE sur les pratiques commerciales déloyales, aux termes duquel il est demandé à la Commission européenne : « *d'enquêter sur la fréquence de cette pratique qui génère une concurrence déloyale et porte atteinte aux données personnelles des utilisateurs, et, le cas échéant, de proposer une législation adéquate pour protéger les consommateurs* ».

L'intérêt pour le SEO

De plus en plus de techniques SEO utilisent les adresses IP pour optimiser la navigation, donc le trafic et donc le référencement et le fonctionnement des moteurs de recherche. Ces pratiques devront être revues ou adaptées aux changements qui vont arriver en 2014, notamment du fait de cette lutte contre les pratiques illicites reposant sur l'IP tracking.

En tant que tel, le fait que les enquêtes et éventuelles futures poursuites soient limitées aux réservations de billets d'avion et de nuits d'hôtel restreint le champ d'application des futures nouvelles règles. Toutefois, il est probable qu'une brèche soit ouverte et que les futures règles s'appliquent en réalité à tout le monde e-marchand et "search".

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...

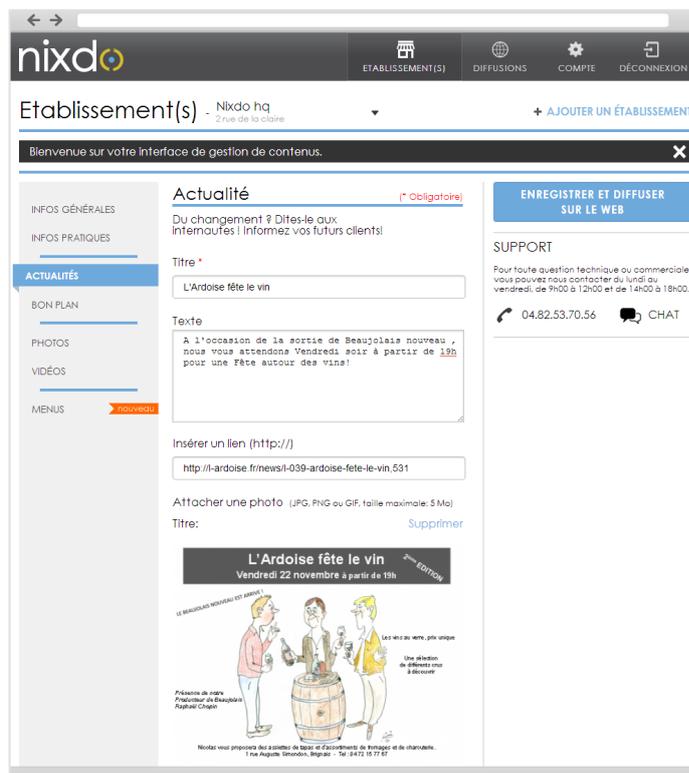
Nom : Nixdo

URL : <https://info.nixdo.com/>

Domaine : SEO local

Type : Site web

Descriptif : Extrait du dossier de presse : *Nixdo permet gratuitement aux commerçants de diffuser et synchroniser les informations concernant leur point de vente ou leur établissement sur les annuaires et sites Internet spécialisés dans la recherche locale, mais aussi aux agences web d'optimiser la présence de leurs clients sur ces supports web et leur Local SEO.*



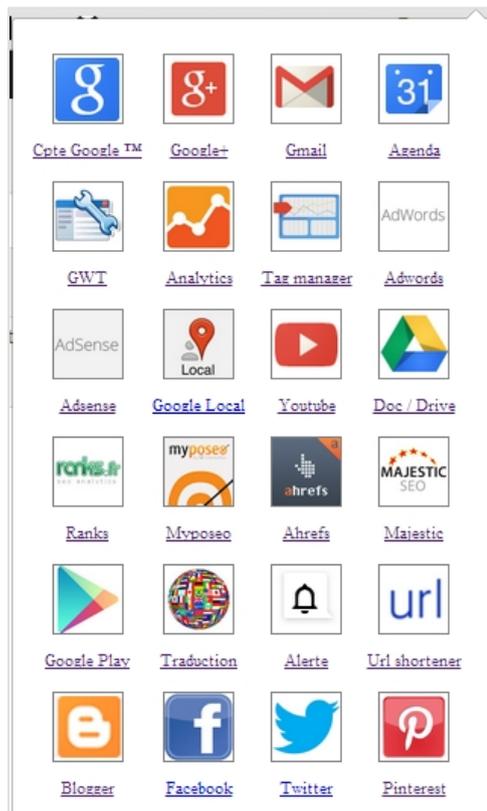
Nom : Boîte à outils SEO

URL : <https://chrome.google.com/webstore/detail/boite-a-outils-seo/iaekkhgjpmeadneinekafkemngdnbhnn?hl=fr>

Domaine : SEO

Type : Extension Chrome

Descriptif : Une petite extension chrome pour regrouper vos accès aux produits Google et autres. Description ici : <http://gain-de-temps.com/blog/extension-chrome-boite-a-outils-seo/>



Nom : Image SEO Tools

URL : <http://www.feedthebot.com/tools/alt/>

Domaine : SEO

Type : Site web

Descriptif : Un petite outil qui extrait les images d'une page web dont vous fournissez l'URL et qui audite chacune d'entre elle au travers de critères SEO.

image SEO tool

This tool will look at a webpage and examine how well the images on that page are following basic image SEO techniques.
Enter URL below:

Image name: logo2007.gif



✔ Above image has 9 words in the alt text.

ⓘ The ALT text of this image appears to be following the Google guidelines well, but you may want to review the text and make sure the ALT text is describing the image, not just repeating words.

✔ Above image has dimensions (height/width)

Alt Text: Abondance.com : Recherche d'information, referencement et promotion de sites Web

Width: 275
Height: 69

Nom : **Webloog**

URL : <http://www.webloog.net/>

Domaine : **SEO**

Type : **Réseau social**

Descriptif : Un réseau social dédié au petit monde du SEO. Le connaissiez-vous ?



Connexion à mon compte - Mot de passe oublié ?

Suivre 590 abonnés J'aime 356 +1 152

Accueil Inscription gratuite Derniers sites ajoutés Les membres Metrics Contact

Trouvez rapidement des blogs et des sites web dans votre thématique

Nom : **WebBoard**

URL : <http://www.cladx.com/e-outils/webboard.html>

Domaine : **SEO**

Type : **Curation**

Descriptif : Un site web qui regroupe les derniers articles traitant de SEO. Présentation un peu austère mais les infos sont utiles pour votre veille (et Abondance s'y trouve ;-))...

The screenshot displays the 'WebBoard - Actualité du référencement' website. It is divided into two main sections: 'Sur les forums' on the left and 'Sur les blogs SEO Francophone' on the right. The 'Sur les forums' section lists various posts with titles, dates, and authors. The 'Sur les blogs SEO Francophone' section features a list of articles, each with a profile picture, author name, article title, and publication date. The articles cover topics such as Google's algorithm, mobile user behavior, and SEO maintenance.

WebBoard - Actualité du référencement

[-< Retour à la liste des e-outils](#)

Sur les forums

- 🔗 **Dév d'un Bot PHP pour twitter**
13/01/2014 12:38 - Bigorno sur scripts-seo
- 📌 **Graphiste indépendant nouveau dans le bizX.**
13/01/2014 11:50 - cedok sur webxfrance
- 📌 **Nouveau membre du forum, présentation de site**
13/01/2014 11:11 - AudreyWebx sur webxfrance
- 📌 **Echanges de liens Voyage/High tech**
13/01/2014 11:01 - Ndco sur veille-seo
- 📌 **Actualité Google+ permet d'envoyer un mail à n'importe qui**
13/01/2014 11:00 - jeannotlapin sur bizpowa
- 📌 **Biz non X Affiliation ZANOX et Publicidees**
13/01/2014 10:56 - Allenpsy sur bizpowa
- 📌 **Graphiste indépendant à votre disposition.**
13/01/2014 10:45 - cedok sur webxfrance
- 📌 **Divers C'est à qui ce site ? : pornodependance.com**
13/01/2014 10:39 - fantasmix sur bizpowa
- 📌 **13/01: Adapter sa stratégie SEO, Les seins de Kate Upton, Barrilla: l'impression 3D**
13/01/2014 09:20 - AudreyWebx sur webxfrance
- 📌 **Actualité Affaire Hollande-Gayet: et le banditisme Corse ?**
13/01/2014 06:24 - fantasmix sur bizpowa

Sur les blogs SEO Francophone

- 📌 **Le comportement des internautes dans l'algo SEO de Google ?**
Par Olivier Duffez - [Voir tous ses articles](#) 13 Janvier 2014 à 9h47
- 📌 **Comment trouver l'ID numérique d'un profil ou page Google+**
Par Olivier Duffez - [Voir tous ses articles](#) 13 Janvier 2014 à 9h27
- 📌 **Infographie : 61% des internautes mobiles quittent un site non-optimisé**
Par Antoine Winants - [Voir tous ses articles](#) 13 Janvier 2014 à 9h15
- 📌 **SEO et site en maintenance, on raconte quoi à Googlebot?**
Par Aymeric Bouillat - [Voir tous ses articles](#) 13 Janvier 2014 à 8h00
- 📌 **20 Idées de Jeux-Concours Facebook Saisonniers pour votre Calendrier Editorial [Infographie]**
Par Isabelle Mathieu - [Voir tous ses articles](#) 13 Janvier 2014 à 6h50
- 📌 **A vos marques, Prêts ? Entrez ! de Rémy Bigot**
Par Alexandra Martin - [Voir tous ses articles](#) 12 Janvier 2014 à 18h12
- 📌 **SEO Campus 2014 c'est parti !**
Par Laurent Bourrelly - [Voir tous ses articles](#) 12 Janvier 2014 à 12h09
- 📌 **Google et le comportement de l'utilisateur**
Par AxeNet - [Voir tous ses articles](#) 12 Janvier 2014 à 3h00

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

Domaine :	Recherche	Référencement
Niveau :	Pour tous	Avancé



Le dossier du mois : bilan 2013, prévisions et bonnes résolutions 2014 !

Chaque début d'année est propice aux bilans sur les 12 mois écoulés, mais aussi aux prévisions sur ceux que nous allons vivre, ainsi qu'aux bonnes résolutions (mais sont-elles toujours tenues ? :-). Voici quelques articles à ce sujet parus ces dernières semaines au sujet du SEO...

-> 14 moyens de développer votre marque avec Google+ en 2014
Kate Bennison Alessi, chef du département Social, Strategy & Solutions chez Google pour la région SEEMEA (Sud et Est de l'Europe, Moyen Orient et Afrique), a vanté toutes les solutions proposées par Google pour que les marques se développent grâce à Google+...
<http://www.webrankinfo.com/dossiers/google-plus/conseils-developpement-marque>

-> 10 excuses pour ne rien changer à votre SEO en 2014
Il est amusant de voir que nos politiques nous proposent tous « le changement » à chaque élection. Parce que s'il y a bien une chose que la plupart des gens détestent, c'est justement le changement...
<http://blog.axe-net.fr/10-excuses-pour-ne-rien-changer-a-votre-seo-en-2014/>

-> Référencement : ce qui changera en 2014
Notamment avec le lancement de Hummingbird, le nouvel algorithme de recherche de Google, le monde du référencement a été riche en changements, des changements qu'il faudra mettre en application en 2014...
<http://www.linformatique.org/referencement-ce-qui-changera-en-2014/>

-> SEO: Quelles leçons à tirer pour 2014... et après?
2014 sera-t-elle plus bouleversante que 2013 ? Petit tour d'horizon des prédictions pour vous aider à adapter votre stratégie de référencement naturel...
<http://rentabilisez.com/2013/12/28/seo-queelles-lecons-a-tirer-pour-2014-et-apres/>

-> 4 choses que je vais faire pour mon SEO en 2014
Hey, c'est mon tour de faire des prédictions SEO pour l'année 2014 ! Je sais, vous avez déjà lu plein d'articles de Nostradamuseo qui vous ont déjà tout appris sur la façon d'être le meilleur référenceur l'année prochaine, c'est pour ça que je vais faire un article un peu différent...
<http://noviseo.fr/2013/12/seo-2014/>

-> Ces évolutions qui vont bouleverser les webmarketeurs en 2014
Cette année 2013 n'est pas encore terminée que nous pensons déjà à l'année 2014 qui se profile. Et étant donné les évolutions et annonces liées à la communication web, autant dire que cette nouvelle année va être chargée ! Petite sélection des annonces et évolutions dont il va falloir tenir compte...
<http://www.pullseo.com/ces-evolutions-qui-vont-bouleverser-les-webmarketeurs-en-2014/>

-> Les 5 résolutions 2014 de notre agence de communication...
Comme chaque début d'année, il est temps de prendre les bonnes résolutions qui, comme le paiement de certains clients, ne prendront effet qu'en 2024 (Private Joke) !
<http://targetweb.fr/resolutions-2014-agence-de-communication/>

-> Nos 7 prédictions pour 2014

C'est la période propice aux prédictions ! Nous avons donc choisi de regarder vers l'avant en recueillant, aux 4 coins de la planète, les prévisions éclairées de nos experts Digital analytics français, allemands, espagnols, anglais et brésiliens...

<http://blog.atinternet.com/fr/index.php/2013/12/31/tendances/nos-7-predictions-pour-2014/3685>

-> 7 résolutions à prendre en référencement en 2014

L'année 2013 vient à peine de se clôturer qu'il faut déjà se tourner vers 2014 et les nouvelles perspectives en référencement...

<http://www.referenceur.be/7-resolutions-referencement-2014-10824.html>

-> Les tendances du référencement naturel pour 2014

Les tendances de cette nouvelle année 2014 ne seront pas nouvelles. Elles sont en réalité la suite logique des évolutions que nous avons remarqué l'année dernière...

<http://www.genetvision.com/7-blog-genetvision/71-les-tendances-du-referecment-naturel-pour-2014.html>

-> Prévisions e-marketing 2014 !

2014 est déjà là avec son lot sans doute de surprises. Quelles seront les tendances web marketing pour les 365 jours qui viennent ? Je m'oblige à cet exercice qui pousse à aller chercher de l'info çà et là...

<http://www.mauricelarger.com/seo-et-publicite-pour-2014/>

-> Prévisions 2014

Qu'il est loin le temps de la parution du dernier article sur ce blog. Les excuses sont nombreuses et plus ou moins acceptables. Je vais toutefois reprendre mes dix doigts et essayer de vous livrer des prévisions pour cette nouvelle année, sujet qui m'est cher puisqu'il s'agit du dernier billet public de mon blog...

<http://www.jambonbuzz.com/focus/previsions-2014-2014-01-05>

-> 10 résolutions pour son site Web pour 2014

Habituellement, on prend les bonnes résolutions le jour du nouvel an. Il n'est cependant pas rare d'en ajouter dans les premiers jours de cette nouvelle année. Voici donc 10 résolutions que vous pouvez ajouter à votre liste pour améliorer votre visibilité sur le web...

<http://www.pullseo.com/10-resolutions-pour-son-site-web-pour-2014/>

-> Les bonnes résolutions SEO pour 2014

Très bonne année à tous, remplie de bonheur, santé, succès et argent (accessoirement :) !

<http://www.miss-seo-girl.com/les-bonnes-resolutions-seo-pour-2014/>

-> 6 Predictions for the Marketing World in 2014

It's time, once again, to continue my annual tradition of analyzing my predictions from 2013, and if I score high enough, predicting what will happen in 2014. I use this process because it keeps me honest, and keeps you informed as to my relative levels of accuracy (or failure)...

<http://moz.com/blog/6-predictions-for-the-marketing-world-in-2014>

-> The Shape of Things to Come: Google in 2014

We can't imagine the future without first understanding the past. In this post, I will present what I consider the most relevant events we experienced this year in search, and will try to paint a picture of things to come by answering this question: How will Google evolve now that it has acquired Wavii, Behav.io, PostRank, and Grapple, along with machine learning and neural computing technology?...

<http://moz.com/blog/search-marketing-in-2014>

-> How Link Building Changed in 2013

Every year, link building gets harder. We rename the process, we struggle to adapt to updates, we beg clients to let us stay as safe as possible, we argue about new factors that may outweigh it in Google's eyes, but still, we keep building links...

<http://searchenginewatch.com/article/2320362/How-Link-Building-Changed-in-2013>

-> Link Building 2013: Looking Back At This Year's Top Link Week Columns

With each passing year, search engine algorithms grow increasingly sophisticated in their ability to identify (and penalize) spammy links. Among other things, this year marked the fourth and fifth releases of Google's spam-filtering Penguin update (Penguin 2.0 and 2.1, respectively)...

<http://searchengineland.com/link-building-2013-looking-back-at-this-years-top-link-week-columns-180269>

-> Link Building the Right Way in 2014

Let's face it. Link building isn't sexy. Link building isn't fun. And in 2014, if done correctly, it certainly isn't cheap or easy...

<http://searchenginewatch.com/article/2320527/Link-Building-the-Right-Way-in-2014>

-> 5 Ways to Have Killer Content in 2014

If there's one thing we heard ad nauseam in 2013 (and 2012, 2011, 2010...), it was that "content is king." Frankly, this phrase has been said so many times that it has lost its meaning. What exactly does that mean?...

<http://searchenginewatch.com/article/2320358/5-Ways-to-Have-Killer-Content-in-2014>

-> Top 10 SEO New Year's Resolutions for a Successful 2014

It's been quite a year for SEO folk everywhere. As 2013 shuts its doors and 2014 beckons, it's an opportune time to reevaluate, reassess, and revisit SEO tactics...

<http://searchenginewatch.com/article/2320353/Top-10-SEO-New-Years-Resolutions-for-a-Successful-2014>

-> 12 SEO Experts Weigh In on SEO 2014 Predictions

SEO is a particularly dynamic field, and those of us working in these trenches like to share our thoughts about this ever-changing environment. I've taken the opportunity to interview a dozen well-regarded experts in the industry, and ask the two biggest SEO questions on marketers' minds for 2014. Their answers may help you prepare and ramp up for the coming year. We're interested in what you think, so please add your own predictions to the comments at the end of this piece.

<http://www.business2community.com/seo/12-seo-experts-weigh-seo-2014-predictions-0729815>

-> 14 Link Resolutions for 2014

If 2013's link building industry could be summed up in one word, that word would be "interesting." There were some big shakeups that other people have covered in various locations on the web, so I won't go into that here; but, what I will discuss is what your 14 resolutions for 2014 could be...

<http://searchengineland.com/14-link-resolutions-for-2014-180812>

-> 6 Essential Elements for Mobile SEO Success in 2014

Adopting a successful mobile SEO strategy in 2014 will undoubtedly give your organization a competitive advantage...

<http://searchenginewatch.com/article/2320360/6-Essential-Elements-for-Mobile-SEO-Success-in-2014>

-> Linking the Google Way in 2014

My previous column showed you the right way to build links in 2014. Now it's time to show you a specific linking tactic that may blow your mind. You're about to learn how to link "Google style."...

<http://searchenginewatch.com/article/2322390/Linking-the-Google-Way-in-2014>

Revue d'URL en français...

-> Désavouer les mauvais liens sans les supprimer, suffisant pour sortir d'une pénalité Pingouin ?

Le Googler John Mueller répond à la question "Pour ne plus être pénalisé par l'algorithme Pingouin, cela peut-il suffire de seulement désavouer tous les mauvais liens, sans les supprimer ?". Explications...

<http://www.webrankinfo.com/dossiers/strategies-de-liens/desavouer-sans-supprimer>

-> 70% des internautes recherchent des entreprises connues dans les SERPs
Search Engine Land et Survey Monkey viennent de réaliser une étude SEM (Search Engine Marketing) concernant les habitudes des consommateurs sur internet ainsi que les partis pris qu'ils peuvent avoir lorsqu'ils recherchent des produits sur le web, et plus précisément sur les moteurs de recherche...

<http://www.referenceur.be/etude-sem-70-internautes-recherchent-entreprises-connees-serps-10652.html>

-> Importance de l'âge des pages chez Google

Tout le monde sait combien Google aime les pages qui ont du bouchon, mais, régulièrement, je tombe sur des sites qui ont effacé des vieilles pages alors que des liens pointent encore dessus...

<http://www.moteurzine.com/2013/12/16/importance-de-lage-des-pages-chez-google/>

-> Les propositions de Google à Bruxelles favoriseraient ses propres services
Commandée par un regroupement d'entreprises du numérique dont Microsoft, une étude d'eye-tracking montre que le dispositif proposé par Google aggraverait sa position dominante...

<http://www.journaldunet.com/ebusiness/le-net/google-etude-icomp-1213.shtml>

-> Interview Daniel Roch, consultant WordPress

Rencontrez avec nous les professionnels du webmarketing. Aujourd'hui c'est Daniel Roch, consultant WordPress & SEO, qui se prête au jeu des questions-réponses...

<http://blog.yooda.com/2013/12/12/2510-interview-daniel-roch-consultant-wordpress/>

-> Comment dépenaliser votre site web – tutoriel

VRDCI est une société de référencement qui aime bien partager ses expériences. Cette fois-ci, VRDCI vous expose la démarche à suivre pour dépenaliser votre site web si Google a trouvé des liens artificiels pointant vers celui-ci...

<http://www.moteurzine.com/2013/12/12/comment-depenaliser-votre-site-web-tutoriel/>

-> Vous n'aviez pas cru en GG+ vous avez eu tort !!!

Bonjour, je reprends la plume plus vite que prévu. Je souhaitais garder ça pour mon petit livre, mais l'information mérite d'être relayée au plus vite pour que tout le monde puisse en profiter et s'adapter...

<http://www.visibilite-referencement.fr/blog/personnalisation-google-plus>

-> Google a-t-il un cerveau ?

Aujourd'hui, nous recevons Sylvain Peyronnet, professeur et chercheur à l'université de Caen Basse-Normandie (laboratoire GREYC)...

<http://www.referencement-naturel-white-hat.fr/google-t-il-un-cerveau/>

-> Brevet de Google ? Le CTR des backlinks

Depuis l'apparition de penguin, nous sommes nombreux à réfléchir continuellement à la question « capitale » : c'est quoi un lien de qualité ?...

<http://blog.axe-net.fr/ctr-des-backlinks/>

-> Automatisation et référencement : Un clash évident ? Le cas Scrapebox

L'association ou plutôt la confrontation entre le référencement « clean » (je vous laisse le soin d'interpréter les guillemets ici) et l'utilisation de logiciels d'automatisation est un un faux semblant qui a toujours la vie dure dans le domaine du SEO.

<http://www.webmarketing-com.com/2013/12/13/24919-automatisation-et-referencement-un-clash-evident-le-cas-scrapebox>

-> Le fichier sitemap.xml n'est pas du tout inutile

Bonjour à tous, un petit billet que j'avais envie d'écrire depuis un petit moment. La raison est simple, tout le monde dit à peu près tout et son contraire sur ce sujet. Du coup j'avais

envie de donner mon avis sur le véritable intérêt d'un sitemap.xml et redorer un peu son blason...

<http://www.visibilite-referencement.fr/blog/sitemap-utile-indexation>

-> Pourquoi le SMO est essentiel dans une stratégie SEO ?

C'est un article rédigé en début de semaine par Marcela De Vivo sur Search Engine Journal qui nous a donné envie de parler de l'importance de l'intégration du SMO (Social Media Optimization) au sein d'une stratégie SEO (Search Engine Optimization)...

<http://www.referenceur.be/pourquoi-social-media-optimization-smo-essentiel-strategie-seo-10698.html>

-> Améliorez votre visibilité sur google.fr grâce à Google+ !

Depuis le 17/12/2013, pour les internautes connectés à leur compte Google, les pages de résultats de recherche sur google.fr sont par défaut personnalisées en fonction de leur activité sociale sur Google+. Découvrez comment ce changement majeur va bouleverser le référencement Google en 2014...

<http://www.webrankinfo.com/dossiers/google-plus/booster-visibilite-serp>

-> Mesurer la densité de mots-clés

Après la préparation de mon nouvel article sur le sitemap.xml pour Veille Digitale, et le développement / mise à jour en continu de mon logiciel de veille et du site AUTOVEILLE, je continue avec la création d'un petit outil SEO pour extraire automatiquement des mots...

<http://autoveille.wordpress.com/2013/12/18/mesurer-la-densite-de-mots-cles-outil-seo/>

-> Moteur de recherche sur les thèmes WordPress

Aujourd'hui, quand je recherche un thème pour WordPress, j'utilise Google et je lis de nombreux articles présentant des thèmes...

<http://www.moteurzine.com/2013/12/23/moteur-de-recherche-sur-les-themes-wordpress/>

-> Retour sur le SEO Camp'us Lille 2013

Le 7 décembre dernier avait lieu le premier SEO Camp'us Lille. 8 intervenants, 5h de conférences, une centaine de participants venus des quatre coins du Nord et de la France : cette première édition a été un joli succès !...

<http://www.seo-camp.org/retour-sur-le-seo-campus-lille-2013-4889>

-> Google+, "le futur de Google"

Raillez, mauvaises langues : Google+ est loin d'être un échec, selon l'insensible géant. Ce serait même le futur de la recherche... et de tous ses services...

<http://www.zdnet.fr/actualites/google-le-futur-de-google-39796616.htm>

-> Intégrer l'Authorank sur un blog WordPress .com

J'espère que vous avez passé un bon Noël. En faisant ma veille SEO et le suivi des positions, je constate que Google a bien pris en compte ma dernière action technique : la mise en place de l'Authorank sur mon blog WordPress .com (oui, celui-ci) !...

<http://autoveille.wordpress.com/2013/12/26/integrer-lauthorank-sur-un-blog-wordpress-com-technique-seo/>

-> SEO et duplication

On a souvent tendance à parler de duplication dans le cadre d'optimisation au référencement, voici un tour complet sur les différents types de « duplicate »...

<http://www.alexseo.fr/conseils-referencement-naturel,seo-et-duplication.html>

-> Référencement pour applications mobiles : comment réussir ?

Il y a quelques mois je présentais au SMX Paris 2013 avec Alexandra Combeau les bases de l'ASO, le SEO pour les stores d'applications et le plan marketing mobile. Que s'est-il passé depuis 6 mois sur le référencement des applis mobiles ?...

<http://www.cafe-referencement.com/seo-technique/referencement-pour-applications-mobiles-comment-reussir-957>

-> 6 cordes faciles à tirer pour vos liens

Je sais, ce sont les "ficelles" qu'on tire dans l'expression et pas les "cordes", mais du coup ça collait plus avec ma superbe image, alors il va falloir vous y faire ! Aujourd'hui j'aimerais vous donner 6 petits trucs relativement simples à mettre en place et qui vont vous permettre de gagner quelques liens supplémentaires...

<http://noviseo.fr/2014/01/liens-quick-wins/>

-> Les Inrocks, les nuls en référencement

C'est toujours assez comiques de lire dans un journal supposé être intelligent un article sur le référencement plein de bêtises (pour ne pas dire autre chose)...

<http://www.moteurzine.com/2014/01/03/les-inrocks-les-nuls-en-referencement/>

-> Comment être certain de rédiger un bon titre

Excellente présentation dédiée à "l'art du titre" qui mérite un maximum de visibilité...

<http://oseox.fr/blog/index.php/1045-choix-titre>

-> Plus de spamco ? Eh non !

Il y a déjà un bon moment que je me pose la question de laisser en place les liens en signature de commentaires sur ce blog. Au mois d'août, j'avais déjà supprimé la possibilité de mettre des liens optimisés, cette fois je passe à ce que je considère comme l'étape ultime. Ma bonne résolution 2014 en fait...

<http://blog.axe-net.fr/plus-de-spamco-eh-non/>

-> Outils netlinking : 21 astuces que vous ignorez certainement

Il existe une multitude d'outils netlinking, certains dédiés à des tâches bien particulières comme la gestion des liens, la phase de recherche et de contact pour des partenariats, l'analyse d'un profil de liens etc. Certains de ces outils sont gratuits, d'autres payants...

<http://blog.resoneo.com/2014/01/outils-netlinking-21-astuces/>

-> Google pingouin : se sortir de la pénalité, ou s'en prémunir

Jusqu'à ce jour, je n'ai jamais eu la motivation d'écrire un billet sur pingouin. Sûrement la faute à un grand nombre de publications, overdose d'infographies, et pas envie de faire du recyclage pour faire du remplissage...

<http://www.htitipi.com/blog/google-pingouin.html>

-> Comment Rap Genius est redevenu visible sur Google 10 jours après sa pénalité

Fin décembre, nous vous expliquions sur ce blog que le site web spécialisé dans les paroles de chansons Rap Genius avait disparu du moteur de recherche de Google pour cause de manipulation de l'algorithme de Google via du spam et des systèmes de liens (pour ceux qui souhaitent se remémorer l'histoire, relisez cet article)...

<http://www.referenceur.be/rap-genius-visible-google-10-jours-apres-penalite-10860.html>

-> Les brevets qui aident Google et son Colibri à comprendre nos requêtes

Analyse des mots clés, de leur contexte, score de confiance envers de possibles synonymes... Les techniques d'analyse des requêtes brevetées par Google sont sophistiquées...

<http://www.journaldunet.com/solutions/seo-referencement/colibri-ou-google-hummingbird-explique-par-les-brevets.shtml>

-> Mettre en place une structure de site en silos

Comme nous l'avons déjà longuement développé lors d'un précédent article concernant l'optimisation de la structure d'un site, on peut clairement affirmer que la bonne évolution d'un site et de sa visibilité dans les moteurs sont en partie dues à son architecture...

<http://www.pole-position-seo.com/2014/01/structure-de-site-en-silo.html>

-> 6 cordes faciles à tirer pour vos liens

La semaine dernière, je vous ai présenté dans cet article trois des six cordes à tirer pour vos liens. En d'autres termes, trouver de nouvelles idées faciles à mettre en place pour récupérer de nouveaux liens vers votre site...

<http://noviseo.fr/2014/01/liens-quick-wins-2/>

-> SEO: le plus beau métier du monde ?

Allez, je vous avais promis un billet par jour cette semaine, nous sommes vendredi et donc on va terminer en légèreté et bonne humeur...

<http://www.visibilite-referencement.fr/blog/seo-metier-passionnant>

-> Google et le comportement de l'utilisateur

Il y a bien longtemps que Google tente d'obtenir un maximum d'information concernant le comportement des utilisateurs sur son moteur de recherche et sur les sites web, ceci afin d'améliorer ses résultats en étant moins dépendant des liens...

<http://blog.axe-net.fr/google-analyse-comportement-internaute/>

-> SEO et site en maintenance, on raconte quoi à Googlebot ?

Il est parfois nécessaire de mettre son site en pause quand des maintenances doivent être effectuées (upgrade d'un CMS, maintenance base de données, mise en ligne d'un site de préprod, etc.)...

<http://www.yapasdequoi.com/seo/3461-seo-et-site-en-maintenance-raconte-quoi-googlebot.html>

-> Pourquoi les entreprises ont-elles un train de retard sur le SEO ?

Le SEO est devenu d'une importance cruciale avec la concurrence de plus en plus rude que se livrent les entreprises pour être visible sur internet. Mais elles semblent avoir toujours un train de retard...

<http://www.journaldunet.com/solutions/expert/56231/pourquoi-les-entreprises-ont-elles-un-train-de-retard-sur-le-seo.shtml>

-> "Réussir son référencement web" : les bonnes feuilles

Spécialiste du SEO et éditeur du site Abondance.com, Olivier Andrieu publie chez Eyrolles une nouvelle version, actualisée et considérablement remaniée, de son ouvrage de référence "Réussir son référencement web"...

<http://www.journaldunet.com/solutions/seo-referencement/reussir-son-referencement-web-2014/>

-> Le Spam en Search : C'est quoi exactement ?

Déjà la deuxième semaine du mois de janvier 2014 ! Ce matin, en effectuant ma veille stratégique à l'aide de mon outil de veille automatique, je suis tombée sur une infographie définissant le spam en search. Je pense qu'il est très intéressant de la partager ici avec vous...

<http://autoveille.wordpress.com/2014/01/13/le-spam-en-search-cest-quoi-exactement-seo/>

-> J'ai vaincu la pénalité Google !

Voici un petit retour d'expérience sur la sortie d'une pénalité manuelle de Google, la sanction la plus redoutée des référenceurs...

<http://www.alexseo.fr/conseils-referencement-naturel,j-ai-vaincu-la-penalite-google.html>

Revue d'URL en anglais...

-> Old Sites vs. New Sites: Which Will Win in SEO?

General trust is moving from older big brands to new innovative companies, but Google still seems to prefer established websites. Should Google change the way it values authority or is the change in public opinion already taking its effect on SEO?...

<http://searchenginewatch.com/article/2318629/Old-Sites-vs.-New-Sites-Which-Will-Win-in-SEO>

-> Hit By Panda and Confused About Low-Quality Content? Run This Google Analytics Report Now

There has always been a lot of confusion surrounding Google Panda. And I totally get it. It's frustrating, scary, and an all-around horrible thing to happen to a website and business owner...

<http://searchenginewatch.com/article/2319712/Hit-By-Panda-and-Confused-About-Low-Quality-Content-Run-This-Google-Analytics-Report-Now>

-> Did Google Help or Hurt SEO in 2013?

Last year I wrote An Open Letter To Google. After a year of updates that smacked down site owners and took negative SEO and collateral damage to a new level, I felt some things needed to be said. Not to mention the lackluster advice they gave that relied on the whole web operating by the rulebook of Kevin Costner...

<http://searchenginewatch.com/article/2319945/Did-Google-Help-or-Hurt-SEO-in-2013>

-> I Am an Entity: Hacking the Knowledge Graph

For a long time Google has algorithmically led users towards web pages based on search strings, yet over the past few years, we've seen many changes which are leading to a more data-driven model of semantic search...

<http://moz.com/blog/i-am-an-entity-hacking-the-knowledge-graph>

-> Machine Learning for SEOs

Since the Panda and Penguin updates, the SEO community has been talking more and more about machine learning, and yet often the term still isn't well understood. We know that it is the "magic" behind Panda and Penguin, but how does it work? Why didn't they use it earlier? What does it have to do with the periodic "data refreshes" we see for both of these algorithms?...

<http://moz.com/blog/machine-learning-for-seos>

-> Use Google Keyword Research To Target The Long Tail In Local Search

There's something of a void where local business marketing is concerned. Local marketers understand they need to target keyword queries in SEO and PPC campaigns, but they assume they know what people search for, and they miss out on valuable referrals because of it. Here's a brief guide on local keyword research to close the gap...

<http://searchengineland.com/use-google-keyword-research-to-target-the-long-tail-in-local-search-179373>

-> 10 Essential WordPress Plugins to Improve SEO & Usability

After you've installed WordPress and covered the basics of improving WordPress for SEO and the user experience, it's time to go a bit deeper. Many great plugins are available to help your WordPress site when it comes to SEO, usability, and conversion. Here are 10 great plugins – most of which are available for free...

<http://searchenginewatch.com/article/2268750/10-Essential-WordPress-Plugins-to-Improve-SEO-Usability>

-> 3 Reasons Why Responsive Web Design is the Best Option For Your Mobile SEO Strategy

As smartphone and tablet adoption rapidly increases, so does the importance of mobile-friendly websites...

<http://searchenginewatch.com/article/2253965/3-Reasons-Why-Responsive-Web-Design-is-the-Best-Option-For-Your-Mobile-SEO-Strategy>

-> Google Not Indexing Your Sitemap URLs? Might Be A Canonical Issue

A Google Webmaster Help thread has a webmaster all upset that Google shows that they have indexed none of the URLs they submitted via the XML Sitemap file. Obviously, this can be concerning to any webmaster...

<http://www.seroundtable.com/google-sitemap-zero-17895.html>

-> Google: Don't Change Your URLs For SEO Purposes

Google's John Mueller responded to a question in the Google Webmaster Help forums about changing a site's URL structure completely...

<http://www.seroundtable.com/google-urls-seo-17889.html>

-> How to Use the Information Inside Google's VED Parameter

It's been two years now since Google announced that they'd be gradually withholding the search terms from referer headers, as secure (SSL) search was rolled out. This has meant that the keyword report in Analytics shows (not provided) for most visitors...

<http://moz.com/blog/inside-googles-ved-parameter>

-> Just Say No To SEO (Sometimes)

There are some SEO rules that you should break. While conventional SEO wisdom advises certain types of behavior, there are times when that advice should be ignored...

<http://searchengineland.com/just-say-no-to-seo-sometimes-180965>

-> Google's December Authorship Shake-up

Back in mid-December, the newly launched MozCast Feature Graph showed a significant short-term drop in the number of tracked searches displaying authorship mark-up. Here's a 30-day view of the data (from November 22, 2013 to December 21)...

<http://moz.com/blog/googles-december-authorship-shake-up>

-> Only 15% Say The Disavow Tool Benefits Their Rankings

The Google Disavow tool is a popular subject with all the link penalties floating around. But the big question is, does it actually help your site to rank better?...

<http://www.seroundtable.com/google-disavow-rankins-17946.html>

-> One-Line Google Reconsideration Request Worked

A webmaster at WebmasterWorld claimed that while some reconsideration requests he submit take five or more times, with tons of details sent to Google - one he sent just recently went through with a one-line request...

<http://www.seroundtable.com/one-line-reconsideration-request-17947.html>

-> How to Create a Prioritized SEO Action Plan

If you spend a few hours on any SEO blog, you'll learn all about how to review a website, identify SEO problems, and propose solutions...

<http://moz.com/blog/how-to-create-a-prioritized-seo-action-plan>

-> Content Strategy for an International SEO Campaign

A content strategy is important for any marketing effort. But developing a content strategy for international marketing efforts has some unique challenges. Here's how to craft a content strategy as part of an international SEO campaign...

<http://searchenginewatch.com/article/2322090/Content-Strategy-for-an-International-SEO-Campaign>

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

Merci pour votre lecture... Pour toute suggestion : olivier@abondance.com