

# Google Panda

Comprendre, Analyser, Agir



par **Olivier Andrieu**

# Google Panda

Comprendre, Analyser, Agir

Olivier Andrieu

Editeur du site  
[www.abondance.com](http://www.abondance.com)

*Illustrations réalisées par Grifil ([www.grifil.com](http://www.grifil.com))*

ISBN : 978-2-212-85415-2

# Sommaire

<b>Introduction</b>	4
<b>Chapitre 1 : Comprendre</b>	6
Historique de Panda	7
Panda 1.0	7
Panda 2.0	10
Google Panda, c'est quoi ?	16
Un filtre de nettoyage	16
Un système basé sur les théories de <i>Machine Learning</i> ?	17
Un système automatique, sans <i>whitelist</i>	19
A priori, pas un filtre antispam classique	19
Le contenu de faible qualité	20
<b>Chapitre 2 : Analyser</b>	22
Les types de sites visés par Panda	23
Les fermes de contenu	23
Les agrégateurs de contenu	24
Les comparateurs de prix	24
Les forums et sites de questions/réponses	25
Le <i>duplicate content</i>	26
Les dommages collatéraux	27
Quelques méthodes de spam visées	27
Le <i>cloaking</i>	27
Le <i>content spinning</i>	28
Le <i>scraping</i>	30
L'indexation de pages de résultats de moteur interne	31
Les recommandations de Google	31
<b>Chapitre 3 : Agir</b>	33
Les 10 actions à mettre en place	34
<b>Conclusion</b>	44
<b>Webographie</b>	45
<b>Présentation de l'auteur</b>	50

# Introduction

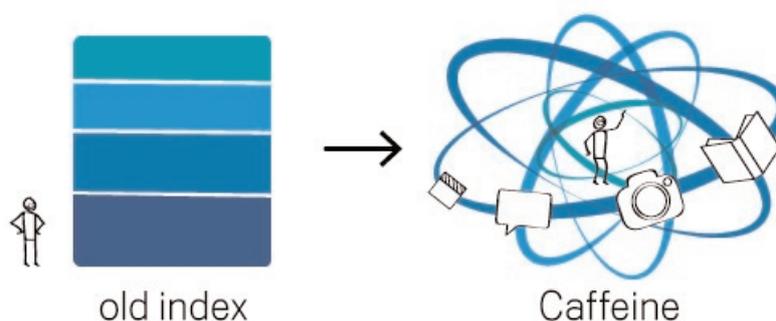
Depuis que Google existe, le moteur de recherche a fait évoluer son index et ses algorithmes de pertinence, pour coller au plus près à la réalité du marché du *search* et à la concurrence croissante de nouveaux acteurs, notamment avec l'avènement de Bing.

Le but de Google est resté, année après année, le même : rester ce que l'outil était rapidement devenu au début des années 2000 : le meilleur moteur de recherche de la planète Web !

Pour cela, les équipes d'ingénieurs de la firme de Mountain View travaillent quotidiennement d'arrache-pied pour s'adapter aux évolutions du Web. Il en résulte plus de 400 changements par an (plus d'un par jour !) de l'algorithme utilisé pour mesurer la pertinence d'une page par rapport à une requête donnée. Une fréquence d'innovation inégalée dans ce domaine jusqu'alors.

« Créer pour durer », voici le credo du moteur de recherche ciselé par Google. Et force est de constater que, jusqu'à maintenant, il y arrive plutôt bien...

Si la formule de classement des résultats évolue chaque jour, la structure d'indexation des pages du Web a également changé avec l'avènement en 2010 de **Caffeine**, une nouvelle façon de référencer la Toile, donnant à Google beaucoup plus de capacités quantitatives pour pouvoir « digérer » le flux croissant de pages web qui apparaissent chaque jour sur Internet.



*Caffeine, nouvelle structure de l'index de Google, offrant une meilleure et plus complète indexation des contenus web et multimédias.*

Source de l'illustration :

<http://googlewebmastercentral.blogspot.com/2010/06/our-new-search-index-caffeine.html>

Mais lorsqu'on « avale » plus de pages web qu'auparavant, le risque d'ingurgiter plus de spam (ou *spamdexing* : fraude aux moteurs de recherche) est également plus important et la digestion n'en est que plus difficile. C'est certainement ce phénomène qui a engendré la mise en place du « Panda update » chez Google... Car Panda n'est qu'une des nombreuses péripéties que l'algorithme du moteur connaît continuellement.

En effet, chaque année ou presque, parmi les nombreux changements effectués sur l'algorithme de Google, il en est un ou deux pour lesquels les *googlers* (nom des employés de la société) « appuient un peu plus sur les curseurs » (ou le champignon, selon que vous soyez geek ou mycophile) et les changements, en termes de positionnement dans les résultats et de trafic généré sur un site web, sont parfois profondément chamboulés. Lorsque ces modifications de l'algorithme sont plus « violentes » que la normale, la communauté des webmasters et des référenceurs donne communément un nom à ce nouvel algorithme, comme pour un cyclone...

Dans le passé, on a donc connu (et dû affronter...) les mises à jour Florida (décembre 2003), Vince (mars 2009), Mayday (mai 2010), etc. Certaines ont fait couler beaucoup d'encre, d'autres sont restées confinées au petit monde du SEO (*Search Engine Optimization*, acronyme anglais servant de plus en plus à dénommer le référencement naturel en France).

Et puis un jour est arrivé Panda, qui pourrait bien bouleverser pour les années qui viennent, la façon dont on envisage le référencement naturel, l'optimisation de sites web et la visibilité organique sur les moteurs de recherche...

Il nous semblait donc important de consacrer un petit guide d'explication à ce « phénomène » pour bien comprendre ce qu'il est – mais aussi ce qu'il n'est pas – et comment l'appréhender au mieux de vos intérêts... Et de ceux de Google, bien sûr ;-)...